Classification of Gait Parameters in Stroke with Peripheral Neuropathy (PN) by using k-Nearest Neighbors (kNN) Algorithm

N. Anang, Faculty of Electrical Engineering, UiTM Shah Alam, Selangor, R. Jailani, Faculty of Electrical Engineering, UiTM Shah Alam, Selangor, N. Mustafah, Faculty of Medicine, UiTM Sungai Buloh, Selangor, H. Manaf, Faculty of Health Sciences, UiTM Puncak Alam, Selangor

***Abstract*—This paper presents the gait pattern classification between 3 groups which are control, stroke only and stroke with Peripheral Neuropathy (SPN) using k-Nearest Neighbors (kNN) algorithm. Control group has been used as a reference or baseline in order to see the difference in the gait pattern. The model able to classify patients into their respective group based on the gait parameters collected. Furthermore, the findings also will help them to monitor patient's performances in rehabilitation program from time to time. 29 subjects has been recruited (9 SPN, 10 stroke subjects and 10 control subjects) with range of age between 40 to 65 years old. Additionally, all subjects must be able to walk freely without any cane or mechanical aid during walking. Vicon® Nexus Plug-in-Gait has been used to compute the kinematic gait parameters. From the results, it is found that there are 9 significant differences in kinematic angles and spatio-temporal data. The classification model developed has been successfully discriminate three different groups with 83.33% accuracy.**

***Index Terms*—kNN, stroke, stroke with peripheral neuropathy (PN), gait pattern**

I. Introduction

Stroke is define by World Health Organization as rapidly developing clinical signs of focal (or global) disturbance of cerebral function, lasting more than 24 hours or leading to death, with no apparent cause other than that of vascular origin. Common stroke complications include motor deficits such as hemiparesis, imbalance, incoordination of the upper and lower limbs, insensate, and cognitive impairment [1]. Brunnstrom described recovery pattern of stroke patients in seven stages from flaccidity, spasticity present, spasticity increase, spasticity decrease, spasticity wanes, coordination reappear, normal function returns. Spasticity limits mobility as it will cause muscle tightness that can decrease volitional movement [2]. As a result, spasticity can cause impairments in motor functions [3]. In short, delay in the motor activity onset, abnormal timing, and wrong sequences of motor activity among stroke survivors may result in impairment of postural control and gait deficit. Previous research study found that in individuals after diagnosed with stroke could be observed to loss of strength rather than loss of control [4]. Apart from that, the risk factors of stroke includes high blood pressure (BP), high blood cholesterol, diabetic mellitus (DM), smoking, disorders of blood rhythm, kidney disease, family genetics, lack of physical activities, and also imbalance diets [5].

Peripheral neuropathy (PN) affects sensory, motor, and autonomic components of the peripheral nervous system [6]. There are many problem arises due to PN disease such as loss of limb, low blood pressure, digestive problems, sexual dysfunction, and also increased or decreased sweating. These complications occur when there is no consultation from doctors and PN is left untreated. Clinically, thorough history will provide clues on causes of peripheral neuropathy. Neurological examination assessing sensory, motor and autonomic function including pain, light touch, vibration, pressure, proprioception, motor strength and ankle reflex [5],[6] will provide extend of PN. Electrodiagnostic tests which include sensory, motor nerve conduction, F response, H reflex and needle electromyography (EMG) will categorize PN into demyelinating or axonal.

Gait is a manner on how humans walk. Meanwhile, walking is an everyday activity in daily life [8], [9]. In order to provide propulsion, walking is a repetitive action in moving both legs alternately, without having both feet touch the ground at the same time [9], [10]. One complete gait cycle can be described as the time interval between two successive events in one of the repetitive events during walking [9]. Gait analysis can be both applied in analysing not only normality but also

This manuscript is submitted on 14 March 2018 and accepted on 4th July 2018. An acknowledgement is accorded to the Ministry of Education, Malaysia for the funds received through the Niche Research Grant Scheme (NRGS), Project file: 600-RMI/NRGS 5/3 (9/2013), National Stroke Association Malaysia (NASAM) and to the Research Management Institute (RMI), and Human Motion Gait Analysis Laboratory, IRMI Premier Laboratory, Universiti Teknologi MARA (UiTM) Shah Alam.

N. Anang, Faculty of Electrical Engineering, Universiti Teknologi MARA (UiTM) Shah Alam, (email: nurhazwanianang@gmail.com)

R. Jailani, Faculty of Electrical Engineering, Universiti Teknologi MARA (UiTM) Shah Alam, (email: rozita8115@gmail.com)

N. Mustafah, Faculty of Medicine, Universiti Teknologi MARA (UiTM) Sungai Buloh, (email: nadiamm81@yahoo.com)

H. Manaf, Faculty of Health Sciences, Universiti Teknologi MARA (UiTM) Puncak Alam, (email: haidzirmanaf@yahoo.com)

abnormality in motion [11], [12]. For example, when the person starts to walk by landing his right foot, and to complete his gait cycle is during the next landing of the same foot. Data from a healthy subject or control group were used as the reference data in gait study. This reference data or sometimes named as normative data were then compared with subjects under the gait study. By comparing the control group, the differences in gait pattern are observed and the value between its deviations from the analysis will bring meaningful information to the given study [13], [14].

kNN is a simple yet robust nonlinear classifier. The development of kNN algorithm was first introduced by Dasarathy in the year of 1991 [15]. Among gait classifier, k nearest neighbors algorithm (kNN) is found to provide a reliable result in detection, classification and categorization as a non-parametric method. Two phases involved in kNN which are training and classification. kNN is one of supervised learning classifier. It works by finding the k nearest instance in training set and after that, kNN do the prediction using kNN Euclidean rules. All data in training set was labeled and being used to classify unlabeled data in the testing set. Data separation used in this study is 80:20 which is 80% of the data will be used for training and 20% of the data is for testing or classification. The distance between this data and all training data is computed and the nearest distance is selected to represent the number of k [16]. There are several distance measures must be followed when dealing with kNN. Some of them are euclidean, city block and correlation [17]. A study showed that distance measure especially euclidean affects the accuracy of kNN classification system [17].

In gait analysis, there are several researches carried out by suing kNN as a classifier [18]–[22]. For example, this research [18] has proven that every man has their unique style of walking and also they found that people identification using dynamic gait features is still perceivable with better recognition rate with different factors such as footwear, clothing, carrying condition and walking speed. Moreover, study about human identification from low-resolution video using height and stride parameters of walking gait was done by [19] and results found that kNN also manage to classify human based on walking pattern. This study [20] proposes a set of new gait signature metrics for recognizing different walking patterns in human gait by using kNN classification method. They uses lightweight wireless accelerometer sensor to measure the variability of acceleration in lateral, vertical and anterior/posterior directions for features extraction [20]. Gait analysis also applicable for gender identification. One related study did investigation on gender identification by using kNN, and the results gives high precision [21].

II. METHODOLOGY

This study is divided into several parts which are data collection, pre-processing and model development.

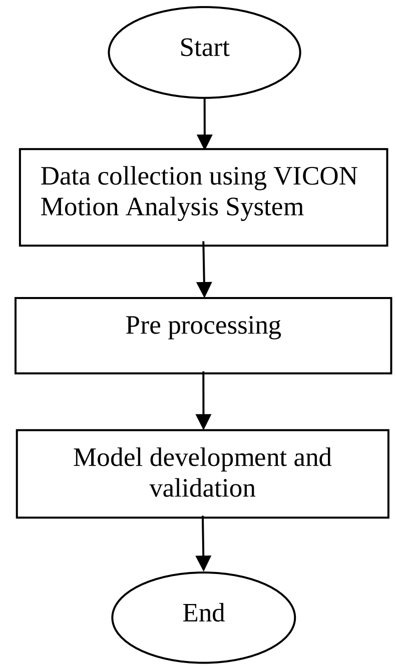

Fig. 1.Flowchart on the methodology if this study

Fig. 1 shows the flowchart of the methodology from the beginning until the end of the study. Four main parts involved in this study; there are data collection, pre-processing, model development and validation.

A. *Data Collection*

Data collection for this study will be conducted at Human Motion Gait Analysis Laboratory, IRMI Premier Laboratory, Universiti Teknologi MARA (UiTM) Shah Alam. The criteria of each subjects were observed by the researcher qualitatively in order to ensure that all the subjects had fulfilled the requirement before signing the consent form. The inclusion and exclusion criteria are as in Table I and Table II. Consent form was need to be signed prior to enrolment into this study. The parent or guardian was given information for this study and all the procedures were explained during briefing session. This study was approved by the Universiti Teknologi MARA (UiTM) Shah Alam research ethics committee and each of the subject’s guardian need to sign an informed consent form to allow subjects to participate in the study.

TABLE I
INCLUSION CRITERIA FOR SUBJECTS

Subjects	Criteria
1. Chronic stroke with PN 2. Chronic stroke	1) Post-stroke 6 – 12 months *(Based on medical record and clinical assessment by medical doctors).* 2) Age 45 – 60 years old. 3) Able to understand English and/or Malay language and consented for the study 4) Able to stand and walk 5 m without aid.
1. Control	1) Age 45 – 60 years old 2) Able to understand English and/or Malay language and consented for the study

TABLE II
EXCLUSION CRITERIA FOR SUBJECTS

Subjects	Criteria
1. Chronic stroke with PN 2. Chronic stroke 3. Control	1) Underlying other neurological problems i.e. other brain pathology, spinal disease, anterior horn pathology, nerve root pathology, neuromuscular junction pathologyand muscle pathology 2) Underlying musculoskeletal problems i.e. low back pain, knee and hip osteoarthritis, hip or knee pain, plantar fasciitis, Achilles tendonitis, deformities of the foot due to the accident or congenital abnormalities 3) Alcoholic 4) Underlying foot ulcer

Full body kinematic data were recorded using the Vicon Nexus 3D motion analysis system (Vicon Motion System, Oxford, UK) at a sampling rate of 100 Hz. Individual participant's anthropometry measurements were entered into the Plug-in-Gait (PiG) modelling software (Vicon, Oxford metrics, Ltd). This research is only focusing on the lower limbs. A total of 16 retro reflective spherical markers were placed bilaterally on the following anatomical landmarks: right/left anterior superior iliac spine, right/left posterior superior iliac spine, right/left thigh, right/left knee, right/left tibial wand marker, right/left ankle, right/left toe, and right/left heel. All markers are positioned either directly on the skin or tight-fitting spandex tights. Participants are instructed to walk at a self-selected speed within 5m walkway. Eight infrared cameras with a sampling frequency of 100 Hz will record the 3D trajectories of each marker as the participant walk. The 3D gait data are collected and processed by the Vicon Clinical Manager (version 1.36) software [23]. Five walking trials are administered until at least three gait cycles were performed. This system is known as an optical tracking system of movement analysis because multiple infrared cameras are going to use to track the motion of all limb segments during the tasks [24].

B. Pre processing

Data obtained from Vicon software at previous step during walking trials will undergo pre-processing techniques. The kinematic data which are angles of hip, knee and ankle were then normalized to 100% gait cycle [25]. One gait cycle is determined with two occurrences of the same foot strike. Besides, the angle for hip, knee and ankle for three groups are analyzed and the comparisons between three groups are made. Furthermore, spatio-temporal gait parameters are determined by taking the average value of cadence, walking speed, stride time, stride length, step time and step length. After that, all data has been undergoing statistical analysis using SPSS Inc. (2015). Since the distribution of all data was not normally distributed, Kruskal-wallis test were applied in order to compare the difference between three groups and the p-values were monitored. [26].

C. Model Development and Validation

After data analysis is done, the most significant parameters obtained during pre-processing technique will be extracted to kNN classifiers for classification purpose. The classifier works by selecting the nearest neighbor for k samples according to the rule specified [27]. After that, the distance between the query vector and all the objects in the training set is computed. Here, the distance algorithm plays a role. Then, the distances for all objects in the training set are sorted. Next, the nearest neighbor based on the k-th minimum distance is determined. All the classes of the training set for the sorted values belong to k are collected and gathered. Lastly, the majority of the nearest neighbors as prediction values is used. This research uses distance metric which specify on Euclidean (EU) distance metric [17]. The kNN classifier is varied by different values of k. The variation value of k is started with 1 until 10 and the optimum value of k is chosen based on the performance measures.

III. RESULTS AND DISCUSSION

A total of 29 subjects (10 control, 10 stroke and 9 stroke with PN) participated in this study. The subjects range between 40 to 65 years of age. The analyses of kinematic and spatio-temporal parameters between three groups are illustrated in numerical analysis by using Kruskal-Wallis test. This study is focusing on kinematic parameters at the lower limbs which are at hip, knee and ankle (Table III). Whereas, the data for spatio-temporal parameters are cadence, walking speed, stride time, stride length, step time and step length are included (Table IV). Besides, during numerical analysis, p-values for each parameter have been monitored to observe the difference between all groups.

Table III shows the numerical analysis of kinematic joint angles using Kruskal-Wallis test. This test has been used to compare 3 groups of subjects. Since the distribution of the data is not normally distributed, Kruskal-Wallis test is been selected to check on the differences between both groups [26]. The p-value from the test statistics has been observed. If the p-value is less than 0.05, then the data is considered to have significant difference between the groups. The data are classified into 3 parts which are average, maximum and minimum angle parameters. From average and minimum angle parameters, it can be observed that the mean values at hip, knee and ankle for control subjects lies in the middle between stroke and SPN subjects. Moreover, at maximum angles, control subjects have greater mean values at hip and knee but, smaller degree at maximum ankle angle. From the test statistics shown in Table I, stroke and SPN subjects does not have normal walking style as control subjects. Thus, from the results obtained, there are 4 significant different observed at average knee (0.034°), maximum hip (0.004°) and knee (0.000°), and minimum ankle (0.025°) between the groups. The findings of this study shows that stroke and SDPN subjects having complexity while walking that leads to abnormal walking pattern as compared to control or healthy people.

TABLE III
NUMERICAL ANALYSIS OF KINEMATIC GAIT PARAMETERS USING KRUSKAL-WALLIS TEST

Angles	Joints	Mean Value			p-value (Kruskal-Wallis test)
		Control	Stroke	SPN	
Average	Hip	15.00	13.20	17.00	0.624
	Knee	16.70	18.70	9.00	0.034*
	Ankle	16.00	18.90	9.56	0.052
Maximum	Hip	20.40	16.30	7.56	0.004*
	Knee	21.60	17.20	5.22	0.000*
	Ankle	15.00	17.30	12.44	0.463
Minimum	Hip	13.20	13.50	18.67	0.297
	Knee	16.50	17.60	10.44	0.148
	Ankle	11.00	20.80	13.00	0.025*

Significant value* ($p<0.05$)

Table IV shows the results of Kruskal-Wallis test for spatio-temporal parameters. It can be seen that there are five significant different obtained which happened at cadence, walking speed, stride time, step time, and stride length. The p-value for cadence is 0.000. Meanwhile, the p-values for walking speed, stride time, step time, and stride length results in 0.001, 0.000, 0.000, and 0.022 respectively. Altogether these 5 parameters conclude that subjects diagnosed with stroke, and stroke with PN has difficulties in walking and complexity in movement.

TABLE IV
NUMERICAL ANALYSIS OF SPATIO-TEMPORAL GAIT PARAMETERS USING KRUSKAL-WALLIS TEST

Spatio-temporal parameters	Mean values			p-value (Kruskal-Wallis test)
	Control	Stroke	SPN	
Cadence	23.35	13.45	7.44	0.000*
Walking speed	22.80	12.80	8.78	0.001*
Stride time	6.65	16.55	22.56	0.000*
Step time	6.45	17.80	21.39	0.000*
Step length	19.20	15.00	10.33	0.077
Stride length	20.50	14.20	9.78	0.022*

Significant value* ($p<0.05$)

The most significant parameters obtained from both kinematic and spatio-temporal parameters will be extracted to kNN classifier as the input for classification purpose. As a result, there were 9 input parameters which are average knee, maximum hip, maximum knee, minimum ankle, cadence, walking speed, stride time, step time and stride length were used for kNN classification model. The classification performances are computed based on accuracy, precision, specificity and sensitivity[28]. The value of k is varies from 1 to 10 and it can be observed that the most accuracy obtained when k is equal to 1. Table V and Table VI shows the confusion matrix for kNN training and testing performance respectively.

A total of 29 data was used in kNN classification model. Since the amount of data is very limited, 29 data was randomly divided into 80-20 ratio of data splitting where, 80 % out of 29 data used for training and 20 % out of 29 data used for testing purpose. As a conclusion, 23 data was used for training and 6 data was used for testing purpose. From the results in confusion matrix shown in Table V, 10 data were correctly classified into group 1, 6 data were accurately classified as group 2 and 7 data were rightly classified as group 3. To conclude, for training performance there was no data incorrectly classified for group 1, 2 and 3. Thus, the accuracy for training performance is 100.00 %. Moreover, for testing performance which illustrated in Table VI, there is no data classified for group 1. Meanwhile for group 2, there were

3 data were correctly classified and 1 data was wrongly classified as group 1. In group 3, there were 2 data correctly classified into the respective group and no data incorrectly classified into group 1 and 2.

TABLE V
CONFUSION MATRIX FOR KNN TRAINING PERFORMANCES

		Predicted group		
		Group 1	Group 2	Group 3
Actual group	Group 1	10	0	0
	Group 2	0	6	0
	Group 3	0	0	7

TABLE VI
CONFUSION MATRIX FOR KNN TESTING PERFORMANCES

		Predicted group		
		Group 1	Group 2	Group 3
Actual group	Group 1	0	0	0
	Group 2	1	3	0
	Group 3	0	0	2

Table VII shows the kNN classification performance measure for three groups which are SPN, stroke, and control group. The performance displayed for testing data includes accuracy, specificity, sensitivity and precision. Based on the table above, specificity and precision gives the same values which are 100.00 %, whereas sensitivity gives 75.00 %. The accuracy of the kNN classification model is 83.33 %.

TABLE VII
KNN CLASSIFICATION PERFORMANCE MEASURE

Parameter	Value (%)
Accuracy	83.33 %
Specificity	100.00 %
Sensitivity	75.00 %
Precision	100.00 %

Fig. 1 shows a line graph of number of k against training and testing accuracy of the classification model. From the graph, the best accuracy performances happened when k=1 and k=2. It is noticed that the percentage of training and testing performances for both k=1 and k=2 are 100% and 83.33% respectively. However, due to the factor of cost and time consuming, k=1 was selected instead of k=2. Ideally, the accuracy performance of training data supposedly higher than testing data since the percentage of training sample is higher than percentage of testing sample. Thus, from the graph, this kNN result proved it as all accuracies for training are higher than testing accuracies for k=1 until k=10. On the other note, the results also support that the developed model has successfully recognize the pattern of data.

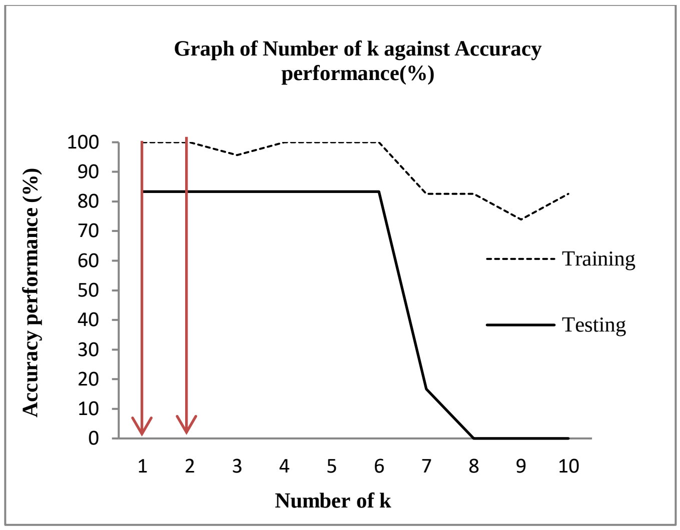

Fig. 2. Graph of number of k against accuracy performances for training and testing data

IV. CONCLUSION

The difference in kinematic angle and spatio-temporal data for control subject, stroke and SPN subjects has been illustrated numerical analysis by using Kruskal-Wallis test. The results shows that there are 9 parameters that has the most significant different which are cadence, walking speed, stride time, step time, step length, average knee, maximum hip, maximum knee and minimum ankle. Thus, these parameters have been used as input parameter for further analysis in classification model. This study applies kNN classification method for classification purpose. These findings are useful as the model can classify patients into their respective group based on the gait parameters collected. From this study, the accuracy obtained is 83.33 % which indicate that there are different in walking behavior between SDPN and stroke subjects although physically their walking patterns are almost the same. In future work, the number of subjects needs to be more in order to improve the classification model performances.

V. Acknowledgement

An acknowledgement is accorded to the Ministry of Education, Malaysia for the funds received through the Niche Research Grant Scheme (NRGS), Project file: 600-RMI/NRGS 5/3 (9/2013), National Stroke Association Malaysia (NASAM) and to the Research Management Institute (RMI), and Human Motion Gait Analysis Laboratory, IRMI Premier Laboratory, Universiti Teknologi MARA (UiTM) Shah Alam.

VI. References

[1] H. Fujimoto *et al.*, "Cortical changes underlying balance recovery in patients with hemiplegic stroke," *Neuroimage*, vol. 85, pp. 547–554, 2014.

[2] A. Martin, S. Abogunrin, H. Kurth, and J. Dinet, "Epidemiological, humanistic, and economic burden of illness of lower limb spasticity in adults: A systematic review," *Neuropsychiatr. Dis. Treat.*, vol. 10, pp. 111–122, 2014.

[3] R. M. Isa and Z. Sharif, "Assessment of Acute Ischemic Stroke Brainwave using Relative Power Ratio," pp. 8–10, 2013.

[4] Louise Ada Colleen Canning Tiffany Dwyer, "Effect of muscle length on strength and dexterity after stroke," *Clin. Rehabil.*, vol. 14, pp. 55–61, 2000.

[5] D. Mozaffarian *et al.*, *Heart disease and stroke statistics-2015 update : A report from the American Heart Association*, vol. 131, no. 4. 2015.

[6] A. J. M. Boulton, F. A. Gries, and J. A. Jervell, "Guidelines for the diagnosis and outpatient management of diabetic peripheral neuropathy," *Diabet. Med.*, vol. 15, no. 6, pp. 508–514, 1998.

[7] A. J. M. Boulton, "Management of Diabetic Peripheral Neuropathy," vol. 23, no. 1, pp. 9–15, 2005.

[8] B. Zhang, S. Jiang, K. Yan, and D. Wei, "Human Walking Analysis , Evaluation and Classification Based on Motion Capture System," *Heal. Manag. – Differ. Approaches Solut.*, pp. 361–398, 2011.

[9] M. W. Whittle, *Gait analysis: an introduction*. 2003.

[10] R. Bartlett, *Introduction to Sports Biomechanics*. .

[11] and S. E. W. D. A. Winter, A. E. Patla, J. S. Frank, "Biomechanical walking pattern changes in the fit and healthy elderly," *Phys. Ther.*, vol. 70, no. 340–347, 1990.

[12] B. K. Naveen Rohila Naresh Chauhan, "Abnormal Gait Recognition," *Int. J. Comput. Sci. Eng.*, vol. Vol. 02, no. No. 05, 2010, pp. 1544–1551, 2010.

[13] and M. B. I. Stancic, T. G. Supuk, "New Kinematic Parameters for Quantifying Irregularities in the Human and Humanoid Robot Gait," *Int. J. Adv. Robot. Syst.*, vol. 9, pp. 1–9, 2012.

[14] and K. M. R. Senden, B. Grimm, I. Heyligers, H. Savelberg, "Acceleration-based gait test for healthy subjects: reliability and reference data," *Gait Posture*, vol. 30, pp. 192–196, 2009.

[15] and Y. S. K. Aminian, P. Robert, E. Jequier, "Estimation of speed and incline of walking using neural network," *Instrum. Meas. IEEE Trans.*, vol. 44, pp. 743–746, 1995.

[16] E. E.-Q. Ismail Hmeidi , Bilal Hawashin, "Performance of KNN and SVM classifiers on full word Arabic articles," *Adv. Eng. Informatics*, vol. 22, no. 1, pp. 106–111, 2008.

[17] I. Saini, D. Singh, and A. Khosla, "QRS detection using K-Nearest Neighbor algorithm (KNN) and evaluation on standard ECG databases," *J. Adv. Res.*, vol. 4, no. 4, pp. 331–344, 2013.

[18] I. Bouchrika and M. Nixon, "Exploratory Factor Analysis of Gait Recognition," *Autom. Face Gesture Recognit. 2008 FG08 8th IEEE Int. Conf.*, pp. 1–6, 2008.

[19] R. Cutler and L. S. Davis, "Biometric Authentication," no. May 2017, p. 2005, 2005.

[20] Sangil Choi, Ik-Hyun Youn, R. LeMay, S. Burns, and Jong-Hoon Youn, "Biometric gait recognition based on wireless acceleration sensor using k-nearest neighbor classification," *2014 Int. Conf. Comput. Netw. Commun.*, no. February 2014, pp. 1091–1095, 2014.

[21] L. . Sudha and D. R Bhavani, "Gait based Gender Identification using Statistical Pattern Classifiers," *Int. J. Comput. Appl.*, vol. 40, no. 8, pp. 30–35, 2012.

[22] H. Josiński, A. Świtoński, K. Jędrasiak, and D. Kostrzewa, "Human Identification Based on Gait Motion Capture Data," vol. I, no. May 2017, pp. 14–17, 2012.

[23] N. K. Zakaria, R. Jailani, and N. M. Tahir, "Comparison kinematic angles between genders in children," *ISCAIE 2015 - 2015 IEEE Symp. Comput. Appl. Ind. Electron.*, pp. 181–185, 2015.

[24] N. . Zakaria, "GENDER DIFFERENCE IN NORMAL CHILDREN USING ANN," 2015.

[25] N. K. Zakaria, R. Jailani, N. M. Tahir, E. Engineering, and U. T. Mara, "Gender differences in gait features of healthy children," vol. 7, pp. 1–6, 2015.

[26] J. Pallant, *Spss Survival Manual*. McGraw-Hill Education (UK), 2013, 2013.

[27] S. Bhaduri, A. Khasnobish, R. Bose, and D. N. Tibarewala, "Classification of Lower Limb Motor Imagery Using K Nearest Neighbor and Naive-Bayesian Classifier," 2016.

[28] M. S. and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manag.*, vol. 45, pp. 427–437, 2009.