A Method for Predicting Human Walking Patterns using Smartphone's Accelerometer Sensor

Zaid T. Alhalhouli

Computer and Information Technology Department
Tafila Technical University (TTU)
Tafila, 179 (66110)
Jordan

Abstract—**Recently, the techniques for monitoring and recognizing human walking patterns have become one of the most important research topics, especially in health applications related to fitness and disease progression. This paper aims at combining machine learning techniques with Smartphone sensors readings (i.e. accelerometer sensor) in order to develop a smart model capable of classifying walking patterns into different categories (fast, normal, slow, very slow or very fast) along with variable of gender, male or female and sensor place, waist, hand or leg. In this paper, we use several machine learning algorithms including: Neural Network, KNN, Random forest, and Tree to train and test extracted data from Smartphone sensors. The results indicate that Smartphone sensor can be exploited in developing a reliable model for identifying the human walking patterns based on accelerometer readings. In addition, results show that Random forest is the best performing classifiers with an accuracy of (92.3%) and (91.8%) when applied on waist datasets for both males and females respectively.**

Keywords—*Smartphone's; accelerometer sensor; walking patterns; machine learning classifiers*

I. INTRODUCTION

With the development of mobile technology over the past few years, mobile devices have become prevalent and nowadays equipped with different kinds of sensors, such as GPS, accelerometer, etc. [1]. The growing development and capabilities of smartphones' sensors have increased researchers' interest to utilize these capabilities in daily life for health care applications that monitor human some activities such as walking patterns for elderly or people with disabilities.

Contemporary Smartphones equipped with many sensors such as Accelerometer, Gyroscope, Magnetometer, and GPS. However, motion sensors are the best suitable for monitoring a device's movements, vibration, tilt, shake, rotation, or swing to identify movements' orientation along the three axes (X, Y, and Z) as shown in Fig. 1. These sensors can determine the phone's orientation if portrait or landscape, and whether the phone' screen is upward or downward. Moreover, the accelerometer sensor can detect how fast your phone is moving in any linear direction [2].

As shown in Fig. 1, when a device is held in its default orientation, X-axis is horizontal and points to the right; Y-axis is vertical and points up; and Z-axis is perpendicular to the face of the screen. Thereof, motion sensors can easily collect the data related to the user's movements and orientations. However, there is a limited ability to automatically support a decision based on large collected data. Therefore, there is a need for developing new data mining and machine learning techniques to make use of these data [2].

Artificial Intelligence provides many solutions based on machine learning techniques that allow the systems to learn and automatically improved according to experiences without explicit programming. Machine learning develops techniques to learn and access through observing to determine the pattern for making a good future decision. Thus, the purpose of machine learning to make the computer automatically decides without any need for human hand [3].

This paper aims at developing a smart model to accurately classify the walking patterns into several categories including: very fast, fast, normal, slow, or very slow based on several machine learning techniques. Therefore, the current study combines general-purpose machine learning techniques with smartphone sensor readings (e.g. acceleration sensor). Also, it involves several sensors that haven't been considered before for recognizing walking patterns (very slow, slow, normal, very fast) and attempts to determine which part of the human body can be the best fit holding smartphone device. And notable and obtain the accuracy according to previous studies related to this algorithm.
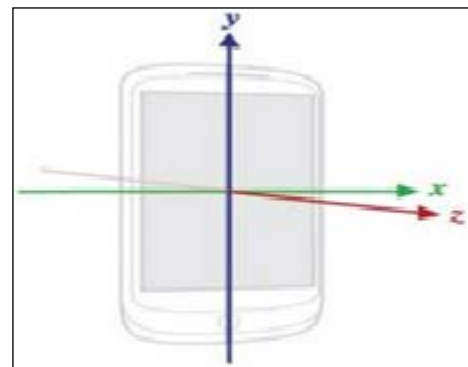

Fig. 1. Device Coordinates System.

II. RELATED WORK

Most of the existing approaches for walking pattern recognition relay on body-worn motion sensors such as accelerometer and gyroscope sensors [4-7]. Different approaches were developed to use foot mounted sensor [8], wearable accelerometer arrays mounted on several parts of the body such as the shank, sacrum, and thigh [5]; and wrist-worn sensors for accurate walking speed estimation [9]. However, these approaches could not efficiently serve general use while having different walking scenarios, in which the user holding his/her smartphone.

Different smartphone-based systems have been developed. Cox et al. developed a simple solution that can estimate the walking speed based on the integration of acceleration by using smartphone and machine learning techniques [10]. Cho et al. suggested to standardize the inertial sensor-based speed estimation using the GPS of the smartphone when the user is walking outdoors [11]. Park et al. investigated the normalized kernel methods on the collected accelerometer data to achieve higher accuracy of walking speed estimation [12].

Even if having an intensive research effort to exploit machine learning techniques to improve the walking speed patterns estimation accuracy, the extracting of effective features still challenging. Thereof, we leverage automated extraction of the most effective features using the deep convolutional neural network (DCNN) to maximize the walking speed estimation accuracy [13, 14].

Recently, deep learning techniques have been dramatically involved in many studies [13, 14, and 15]. However, they focused on the recognition of gait patterns rather than on the recognition of walking speed patterns. Gong et al. (2016) developed DCNN to perform gait assessment for multiple sclerosis patients based on the spectral and temporal associations among sensor data collected with several inertial body sensors [13]. Gadaleta and Rossi adopted the DCNN to recognize a target user based on the way of their walking utilizing the accelerometer and gyroscope data of smartphone [14]. Hannink et al. used the DCNN to estimate the stride length [15].

In this context, the research aims to classify and assess four supervised Machine Learning algorithms, which are Naïve Bayes (NB), KNN and Decision Tree (DT). The study shows the performance accuracy and capability of the experimented algorithms to provide a comparative analysis. Followings summarize the selected supervised algorithms:

- The k-nearest neighbours (KNN): KNN is a simple classification and regression algorithm that stores all the available cases and classifies new incoming cases based on a certain similarity measure. Conceptually, KNN is a simple algorithm; nevertheless, it is still capable of solving complex problems. The KNN algorithm is a type of instance-based learning or lazy learning, wherein the function is approximated only locally. All computation is ceased until classification [16].

- Artificial Neural Networks (ANNs): ANNs are networks inspired by biological neural networks. Neural networks are non-linear classifier which can model complex relationships between the inputs and the outputs. A neural network consists of a collection of processing units called neurons that work together in parallel to produce some output [17]. Each connection between neurons can transmit a signal to other neurons and each neuron calculates its output using the nonlinear function of the sum of all neuron's inputs [16].

- Decision Tree (DT): DT is a common learning method used in data mining. DT refers to a hierarchal and predictive model which uses the item's observation as branches to reach the item's target value in the leaf. DT is a tree with decision nodes, which have more than one branch and leaf nodes, which represent the decision [16].

- Random Forests (RF): RF is a classifier consisting of a collection of tree-structured classifiers. The random forest classifies a new object from an input vector by examining the input vector on each tree in the forest. Each tree casts a unit vote at the input vector by giving a classification. The forest selects the classification having the most votes overall the trees in the forest [18].

III. METHODS

This work adopts supervised learning approach to extract features vector that describes the walking patterns of human (fast, very fast, normal, slow, very slow) based on the collected data from fixed and predefined walking distance. Therefore, this paper attempts at determining the best combination of accelerometer sensor data, sensor axis(es) and learning algorithms to detect walking patterns. Fig. 2 defines the research steps.

Based on Fig. 2, the defined method's steps can be addressed as follows:

Step 1: Collecting raw data by using the accelerometer sensor to be stored in the smartphone file system.

Step 2: Retrieving stored sensor data to an excel file format for pre-processing task. Data will be clustered for each walking pattern attempts according to several variables including: sensor location, type of walk and gender as shown. All variables were coded to their corresponding numeric values as shown in Table I; for instance, values of 1, 0 and 2, which will be expressed as (102), mean that smartphone was placed at the waist of a male user walking normally. Afterwards, the data related to males were separately retrieved from the excel file, as well as the females' data excluding outliers' cases.

Step 3: Testing and training of revised data using several machine learning classifiers. In addition, accuracy evaluation applied to assess the performance of the involved classifiers using Orange software, which is an open-source software package released under GPL [19].
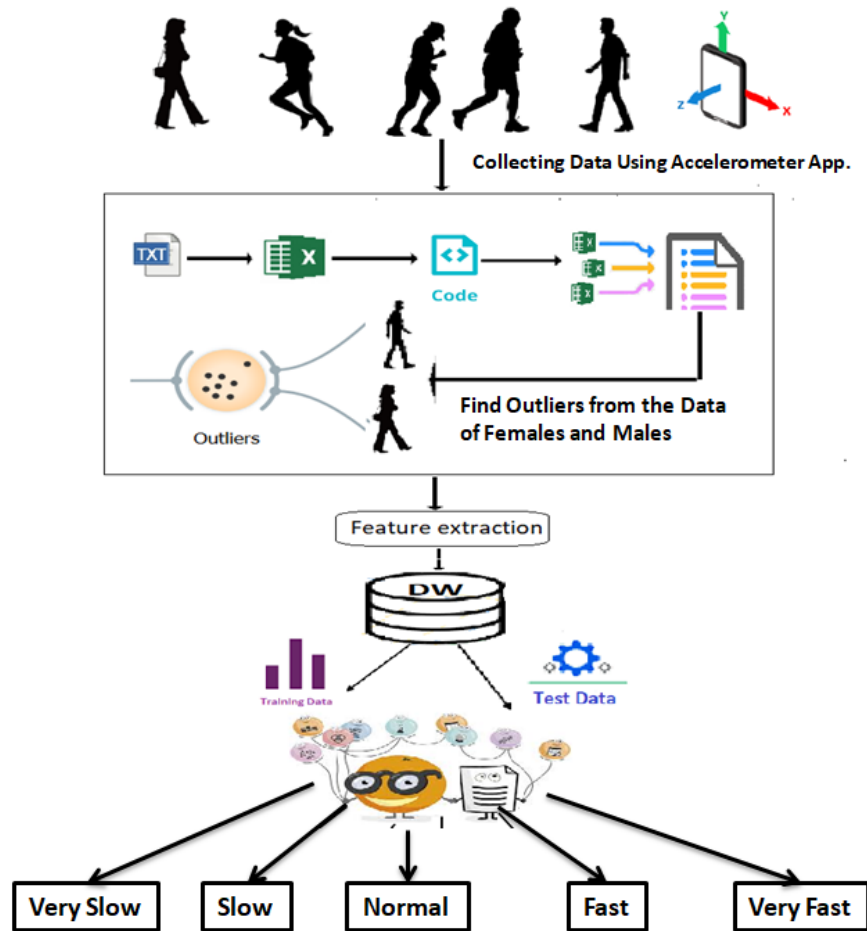
Fig. 2.   Research Steps

TABLE I.   FLAG MAPPING

Variable	Option	Code
Smart Phone Position Place	Waist	1
	Hand	2
	Leg	3
Gender	Male	0
	Female	1
Walk Pattern	Fast	1
	Normal	2
	Slow	3
	Very slow	4
	Very fast	5

IV. EVALUATION

To evaluate the proposed model, we compared the accuracy of several machine learning algorithms includes: Artificial Neural Networks (ANN), Random Forest (RF), K-Nearest Neighbors algorithm (KNN), and Tree. These algorithms are commonly used in similar studies due to their ability to process sensing data [15, 51, 20, and 24].

This paper investigates the accuracy evaluation of the used machine learning classifiers for each walk category on the confusion matrix as shown in equation 1, which ranges from 0.0 to 1.0 (Thang et al., 2012). Therefore, the closer of classifier accuracy is to 1.0, the better the prediction of the walks types. Fig. 3 shows the evaluation metric, using Orange software, to ensure the validly of the proposed model of walking patterns (fast, very fast, normal, slow or very slow).

$$Accuracy = \frac{TP+TN}{P+N} \qquad (1)$$

Where TP denotes True Positives, TN denotes True Negatives, P denotes positive cases and N denotes Negative cases.
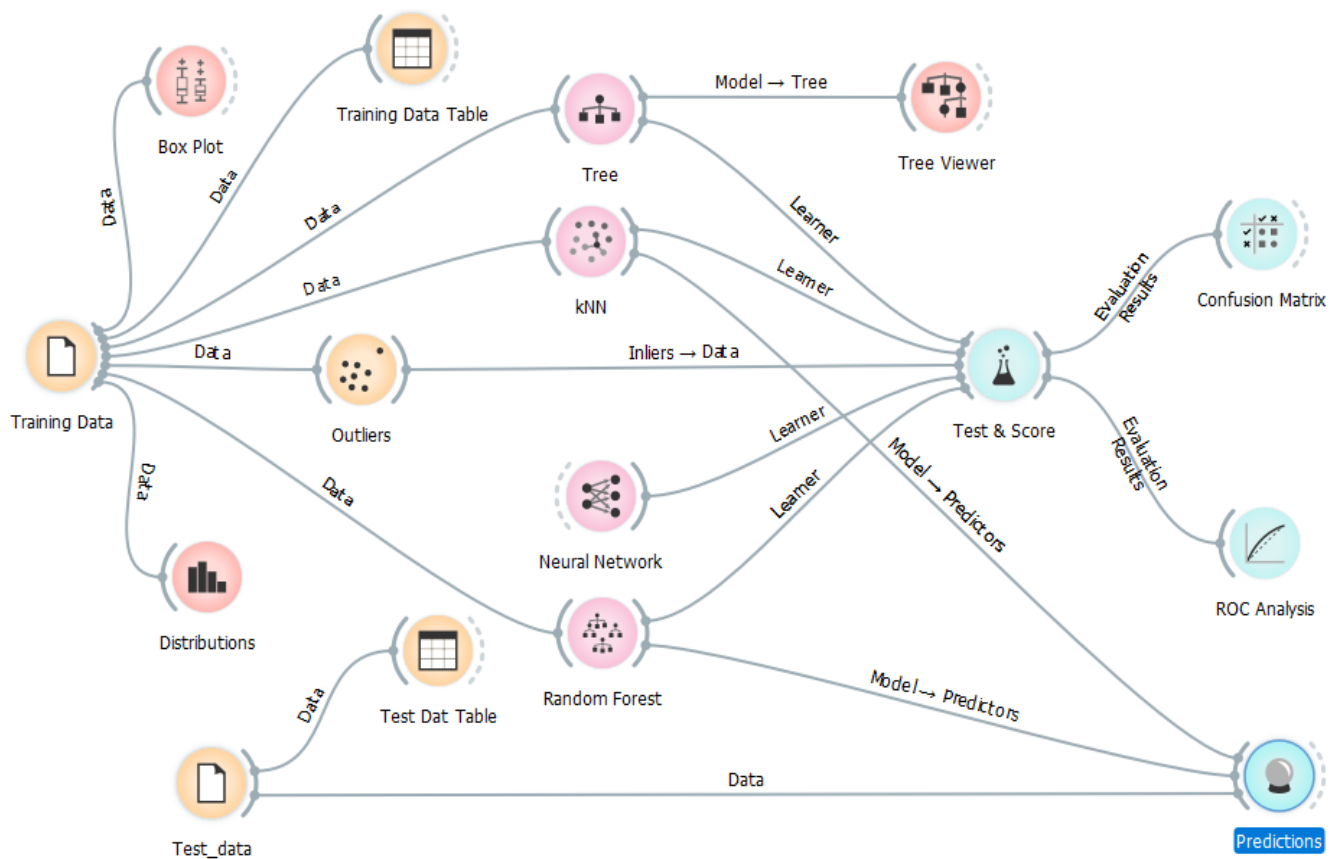
Fig. 3.   Evaluation Metric.

V.   DATA COLLECTION AND EXPERIMENTS

This section shows the real-world experiments conducted to collect the accelerometer sensor data for the walking patterns. In this experiment, a total of eight volunteer university students (four males and four females) with an average age of (21 years) participated for collecting data from smartphone accelerometer of type Huawei; however, different models used (Y6 Prime 2018, Y7 Prime 2018, Nova 3i). Each student asked to walk a certain distance in normal, fast, very fast, slow and very slow speeds in order to collect the data. During the walk, the smartphones placed on the hand, leg, and waist to ensure accurate reading data as shown in Fig. 4.
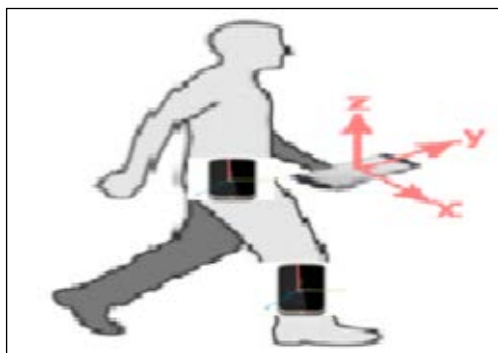
Fig. 4.   Sensors Layout.

For proceeding our evaluation procedures, we mapped the codes of the used variables as previously presented in Table I. Then all produced datasets are trained and tested using several machine learning classifiers to evaluate their accuracy including Artificial Neural Networks (ANN), Random Forest (RF), K-Nearest Neighbors algorithm (KNN), and Tree.

VI.   RESULTS

As previously described, the resultant dataset from all experiments was trained and tested using several machine learning classifiers. Table II shows a sample of the produced dataset representing the extracted features and variables flags.

TABLE II.   SAMPLE OF OUR DATA SET

1	X (m/s)	Y (m/s)	Z (m/s)	Flag
2	-0.755092	-7.313296	0.199615	101
3	-0.058055	-7.719409	-0.482911	101
4	1.220921	-9.970569	0.677047	101
5	3.900410	-12.573337	-1.679872	101
6	0.740468	-15.148089	-0.233811	101
7	-4.812709	-12.647845	-0.614732	101
8	-6.990199	-8.847034	-0.442063	101
9	-3.629466	-10.212100	-1.258363	101
10	1.647435	-15.811786	-1.579438	101

Afterwards, we find and compare the accuracy for each classifier in terms of smartphone replacement; hand; waist or leg. Results prove that the classification accuracy increased when the sensor is placed at the waist. This can be justified as the waist is the steadiest part of the human body when abnormal movements presented. Therefore, the waist readings were approved in this study for validation purpose.

The classification results based on males' waist dataset show that the Random Forest algorithm maintains the highest accuracy level of (92.1%).Similarly, for classifying females' waist dataset, the Random Forest algorithm achieved the highest accuracy level of (91.8%) as well. Table III shows the results.

As shown in Table III, Random Forest is the best performing algorithm when applied to both males and females' waist datasets. However, a very small difference between them, which can be attributed to the different number of trained records and the physiological differences between males and females.

In order to determine the best performing classifier, we separately obtained the confusion matrix of applying all classifiers on waist datasets for males and females to measure their performance. In the confusion matrix, each row represents the instances in an actual class; while each column represents the instance in a predicted class or vice versa. Confusion matrix summarizes the results of the testing algorithm and provides a report of the number of True Positive

(TP), False Positives (FP), True Negatives (TN), and False Negatives (FN). Table IV describes the mapped codes for the used variables.

Flags in Table III used to find the confusion matrix as shown in Fig. 5 and 6, respectively.

Based on Fig. 5, results show that the Random Forest is the best performing algorithm when classifying the females' data comes from smartphones placed at their waists. When the walking patterns are fast, it has been found that 889 out of 1414 cases were correctly classified; while when are normal, 2678 out of 3128 cases were correctly classified; and 5063 out of 5368 cases were correctly classified when walks are slow; and 1226 out of 1463 cases were correctly classified when walks are very slow; finally 7674 out of 7717 were correctly classified when walks are very fast.

Similarly, as for the male confusion matrix, results show that the Random Forest is the best performing algorithm when classifying the males' data come from smartphones placed at their waists. When the walking patterns are fast, it has been found that 1066 out of 1586 cases were correctly classified; while when they are normal, 2640 out of 3133 cases were correctly classified; and 4719 out of 5033 cases were correctly classified when walks are slow; and 1438 out of 1781 cases were correctly classified when walks are very slow; finally 9950 out of 10021were correctly classified when walks are very fast.

TABLE III. RESULT OF CLASSIFICATION

Model	Accuracy Male			Accuracy Female		
	Hand	*Waist*	*Leg*	*Hand*	*Waist*	*Leg*
Random Forest	0.661	0.923	0.615	0.7355	0.918	0.662
Neural Network	0.650	0.913	0.608	0.6833	0.912	0.642
KNN	0.647	0.919	0.594	0.7364	0.915	0.648
Tree	0.620	0.896	0.579	0.7149	0.886	0.605

TABLE IV. CODE DESCRIPTION OF USED DATASET

Flag	Description
111	Waist; female; fast
112	Waist; female; normal
113	Waist; female; slow
114	Waist; female; very slow
115	Waist; female; very fast
101	Waist; male; fast
102	Waist; male; normal
103	Waist; male; slow
104	Waist; male; very slow
105	Waist; male; very fast

Confusion matrix for Random Forest (showing number of instances)

	Predicted	111	112	113	114	115	Σ
Actual	111	853	358	123	80	0	1414
	112	416	2432	258	22	0	3128
	113	142	208	4903	69	46	5368
	114	83	37	83	1166	94	1463
	115	0	2	33	45	7637	7717
	Σ	1494	3037	5400	1382	7777	19090

Confusion matrix for kNN (showing number of instances)

	Predicted	111	112	113	114	115	Σ
Actual	111	779	435	138	62	0	1414
	112	150	2678	282	18	0	3128
	113	78	183	5066	36	5	5368
	114	67	23	97	1160	116	1463
	115	0	0	2	19	7696	7717
	Σ	1074	3319	5585	1295	7817	19090

Confusion matrix for Random Forest (showing number of instances)

	Predicted	111	112	113	114	115	Σ
Actual	111	889	349	122	54	0	1414
	112	224	2678	213	13	0	3128
	113	86	189	5063	30	0	5368
	114	62	27	82	1226	66	1463
	115	0	0	10	33	7674	7717
	Σ	1261	3243	5490	1356	7740	19090

Confusion matrix for kNN (showing number of instances)

	Predicted	111	112	113	114	115	Σ
Actual	111	888	354	121	51	0	1414
	112	227	2645	245	10	1	3128
	113	106	170	5067	21	4	5368
	114	54	21	113	1185	90	1463
	115	0	0	1	6	7710	7717
	Σ	1275	3190	5547	1273	7805	19090

Fig. 5. Confusion Matrix for Females.

Confusion matrix for kNN (showing number of instances)

	Predicted	101	102	103	104	105	Σ
Actual	101	1025	343	144	74	0	1586
	102	200	2630	276	23	4	3133
	103	82	108	4760	15	68	5033
	104	109	70	113	1357	132	1781
	105	0	0	17	18	9986	10021
	Σ	1416	3151	5310	1487	10190	21554

Confusion matrix for Random Forest (showing number of instances)

	Predicted	101	102	103	104	105	Σ
Actual	101	1066	328	125	67	0	1586
	102	219	2640	245	28	1	3133
	103	85	159	4719	29	41	5033
	104	107	58	88	1438	90	1781
	105	0	1	30	40	9950	10021
	Σ	1477	3186	5207	1602	10082	21554

Confusion matrix for Neural Network (showing number of instances)

	Predicted	101	102	103	104	105	Σ
Actual	101	965	386	165	70	0	1586
	102	186	2561	359	26	1	3133
	103	96	150	4707	27	53	5033
	104	125	52	94	1282	228	1781
	105	0	0	30	46	9945	10021
	Σ	1372	3149	5355	1451	10227	21554

Confusion matrix for Tree (showing number of instances)

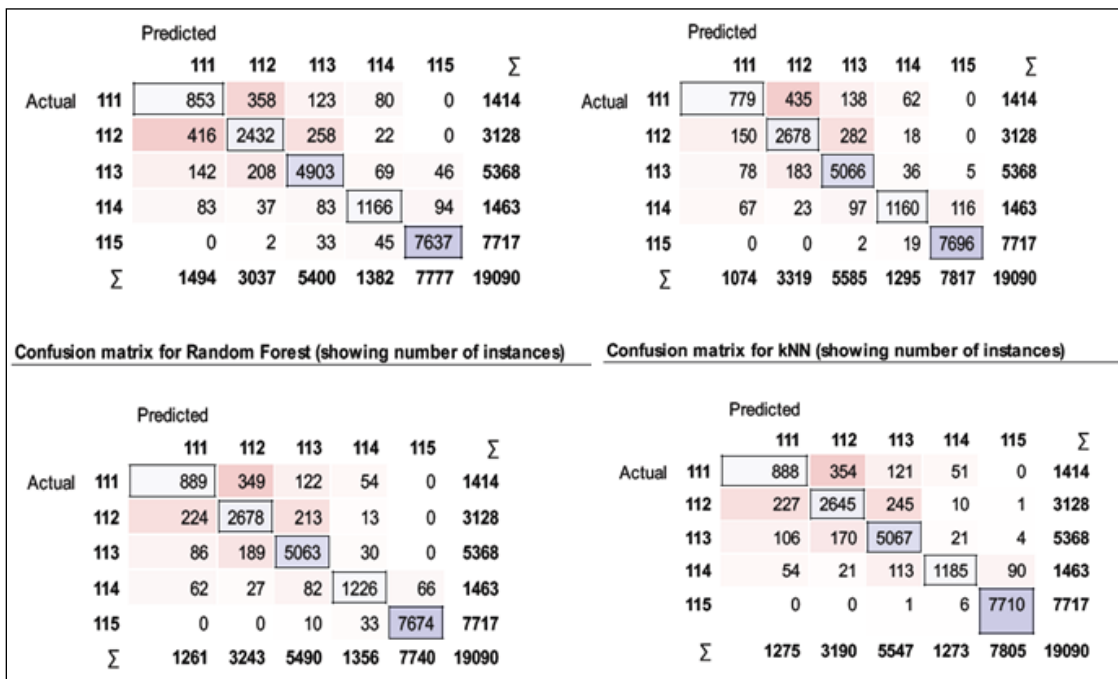|	Predicted	101	102	103	104	105	Σ
Actual	101	1048	308	137	93	0	1586
	102	377	2466	247	41	2	3133
	103	156	239	4469	81	88	5033
	104	117	74	92	1374	124	1781
	105	0	2	79	61	9879	10021
	Σ	1698	3089	5024	1650	10093	21554

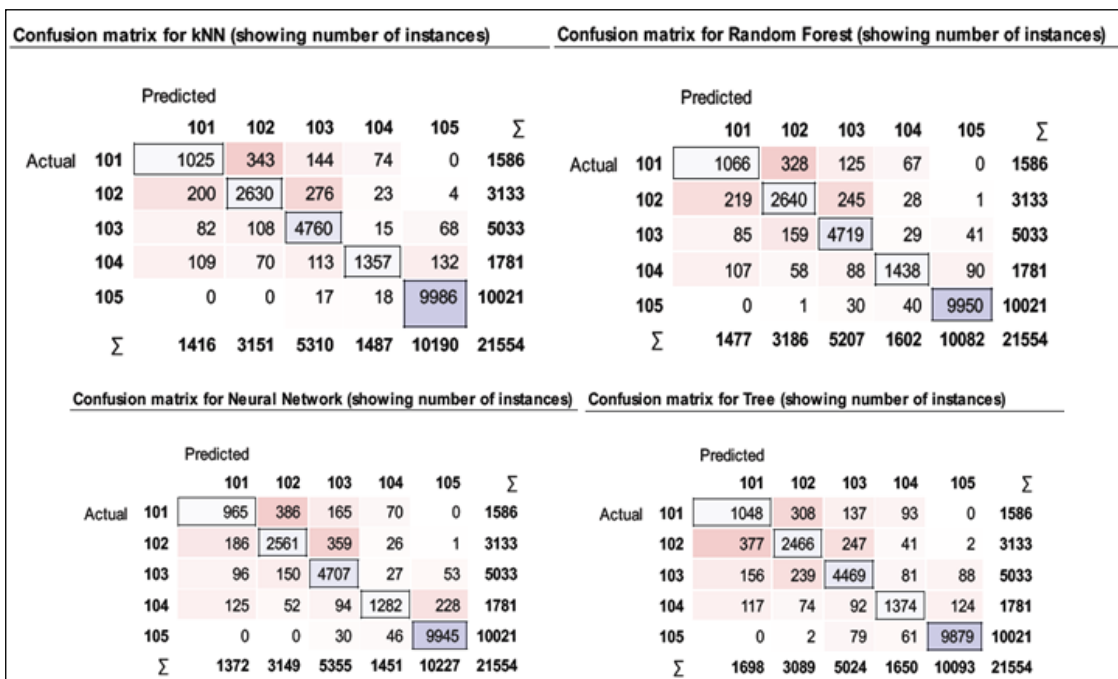Fig. 6. Confusion Matrix for Males.

VII. VALIDATION

We validate our model by displaying the Box plot distribution of the Y (m/s2) readings related to males' and females' waist datasets, since these datasets achieved the highest rate of correctly classified cases when applying Random Forest classifier. Results confirmed that there was no significant discrepancy between the mixed data of males and females as shown in Fig. 7, which in turn shows the validity of our experiments.

Furthermore, the model was validated by applying the two most accurate classifiers, Random Forest and KNN on an external Excel file, which contains unflagged data, attributes of the waist; Female; and slow. Results show that the Random Forest algorithm made the correct prediction and returned (113) flag code, which represents waist; Female; and slow respectively as shown in Fig. 8.
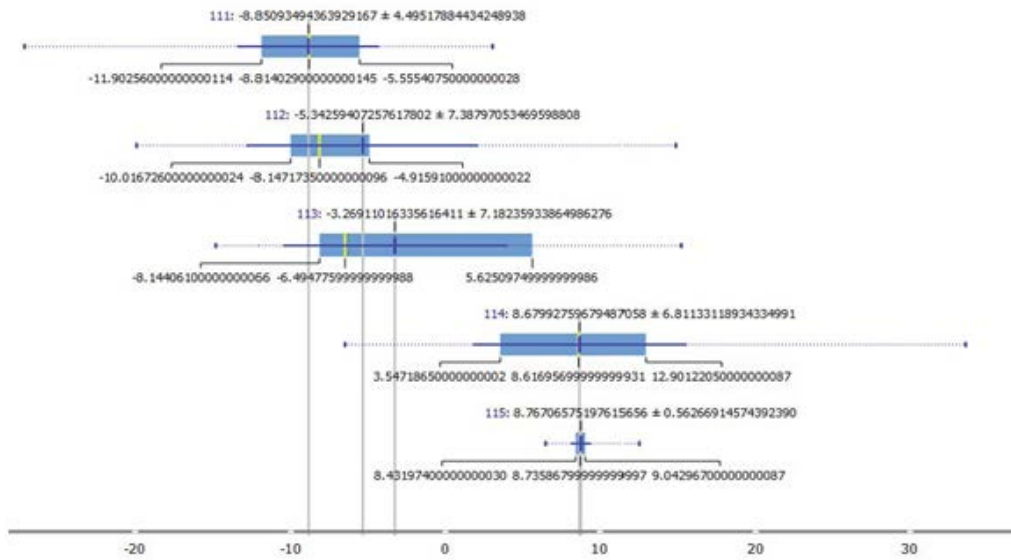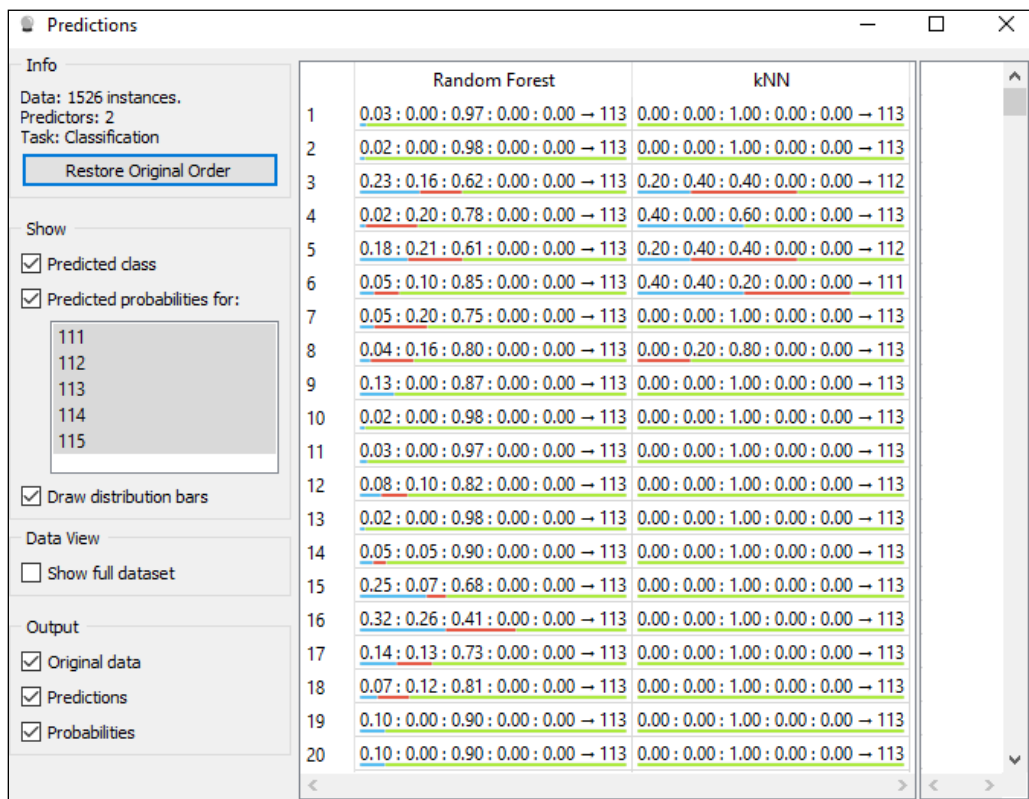
Fig. 7. Box Plot for Attribute ' Y (m/s2).

Fig. 8. Result of Test.

VIII. DISCUSSION

In this paper, we proposed a method to accurately predict all patterns of human walking including very slow, slow, normal, very fast. Therefore, it combines general-purpose machine learning techniques with smartphone sensor readings (e.g. acceleration sensor) to develop a smart model capable of classifying and predicting the walking patterns into very fast, fast, normal, slow, or very slow. In addition, it determines the best part for placing the sensor on the human body (Hand, Waist, and Leg). Thus, we provide a distinguished study by using several sensors simultaneously placed at different human body parts to collect sensing data to be trained and tested by accurate classifiers.

To achieve the study aim, we involved several variables including sensor location, gender during the experiments to identify the walking pattern classes as an activity. A total of 8

students, 4 males and 4 females, participated in our experiments and performed different walking scenarios while three smartphones similarly oriented and placed at waist; hand; and leg of every individual simultaneously. The results after processing the involved datasets indicated that the Random Forest (RF) is the best performing classifier in terms of accuracy when classifying both males or female's waist datasets; however, males dataset proves a higher performance with an accuracy of (92.3%) and (91.8%) for females' dataset. In addition, results indicated that the waist can be the best steady human body part for placing smartphone sensors to recognize walking patterns.

Finally, when comparing our work with previous literature, we can find several methods that can be used to determine the walking patterns [1, 21-23]. However, Tang & Phoha (2016) [1] found that KNN is the best which; while our study indicates that Random forest is the best performing classifier. Additionally, Thang & et al. (2012) [22] adopted SVM classifiers to identify the user's gender based on biometric gait with an accuracy of (92.7%). Also, Gupta& et al (2014) [23] conducted a similar study using the Mean shift clustering algorithm with an accuracy level of (95%).

IX. Conclusion and Future Work

In this paper, we developed a method to collect sensing data and accurately classify all human walking patterns including very slow, slow, normal, very fast. The evaluation results of current methods involving the application of four classifiers (K-Nearest Neighbor, Random Forest, Tree, and Neural Networks) indicate that the Random Forest is the best performing classifier. Random Forest achieves a higher accuracy level when applied on waist datasets for both males and females compared to other classifiers. However, the researchers have a plan to improve resultant accuracy and expand the research domain to include more samples of people using different methods and environments settings, such as stairs and rectum.

References

[1] Tang, C., &Phoha, V. V. (2016, September). An empirical evaluation of activities and classifiers for user identification on smartphones. In 2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (Btas) (pp. 1-8). IEEE.

[2] Liu, M. (2013). A study of mobile sensing using smartphones. International Journal of Distributed Sensor Networks, 9(3), 272916.

[3] Bishop, C. M. (2006). Pattern recognition and machine learning. springer.

[4] Hu, J. S., Sun, K. C., & Cheng, C. Y. (2012, December). A model-based human walking speed estimation using body acceleration data. In 2012 IEEE International Conference on Robotics and Biomimetics (ROBIO) (pp. 1985-1990). IEEE.

[5] McGinnis, R. S., Mahadevan, N., Moon, Y., Seagers, K., Sheth, N., Wright Jr, J. A., ... &Pindado, J. A. (2017). A machine learning approach for gait speed estimation using skin-mounted wearable sensors: From healthy controls to individuals with multiple sclerosis. PloS one, 12(6), e0178366.

[6] HERREN, R., SPARTI, A., AMINIAN, K., & SCHUTZ, Y. (1999). The prediction of speed and incline in outdoor running in humans using accelerometry. Medicine & Science in Sports & Exercise, 31(7), 1053-1059.

[7] Li, Q., Young, M., Naing, V., &Donelan, J. M. (2010). Walking speed estimation using a shank-mounted inertial measurement unit. Journal of biomechanics, 43(8), 1640-1643.

[8] Mannini, A., & Sabatini, A. M. (2010). Machine learning methods for classifying human physical activity from on-body accelerometers. Sensors, 10(2), 1154-1175.

[9] Zihajehzadeh, S., & Park, E. J. (2016). Regression model-based walking speed estimation using wrist-worn inertial sensor. PloS one, 11(10).

[10] Cox, J., Cao, Y., Chen, G., He, J., & Xiao, D. (2014, December). Smartphone-based walking speed estimation for stroke mitigation. In 2014 IEEE International Symposium on Multimedia (pp. 328-332). IEEE.

[11] Cho, D. K., Mun, M., Lee, U., Kaiser, W. J., &Gerla, M. (2010, March). Autogait: A mobile platform that accurately estimates the distance walked. In 2010 IEEE International Conference on Pervasive Computing and Communications (PerCom) (pp. 116-124). IEEE.

[12] Park, J. G., Patel, A., Curtis, D., Teller, S., &Ledlie, J. (2012, September). Online pose classification and walking speed estimation using handheld devices. In Proceedings of the 2012 ACM Conference on Ubiquitous Computing (pp. 113-122). ACM.

[13] Gong, J., Goldman, M. D., &Lach, J. (2016, October). Deepmotion: a deep convolutional neural network on inertial body sensors for gait assessment in multiple sclerosis. In 2016 IEEE Wireless Health (WH) (pp. 1-8). IEEE.

[14] Gadaleta, M., & Rossi, M. (2018). Idnet: Smartphone-based gait recognition with convolutional neural networks. Pattern Recognition, 74, 25-37.

[15] Hannink, J., Kautz, T., Pasluosta, C. F., Barth, J., Schülein, S., Gaßmann, K. G., ... &Eskofier, B. M. (2016). Stride length estimation with deep learning. arXiv preprint arXiv:1609.03321.

[16] Hammouri, A., Hammad, M., Alnabhan, M., &Alsarayrah, F. (2018). Software bug prediction using machine learning approach. IJACSA) International Journal of Advanced Computer Science and Applications, 9(2).

[17] Misirli, A.T. and Bener, A.B., (2014, June). A mapping study on bayesian networks for software quality prediction. In Proceedings of the 3rd International Workshop on Realizing Artificial Intelligence Synergies in Software Engineering (pp. 7-11).

[18] Guo, L., Ma, Y., Cukic, B., & Singh, H. (2004, November). Robust prediction of fault-proneness by random forests. In 15th international symposium on software reliability engineering (pp. 417-428). IEEE.

[19] Orange Tools Website. [online] https://orange.biolab.si/download/# windows. (Accessed 21 January 2020).

[20] Domingos, P. (2015) The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World. Basic Books.

[21] Fawcett, T., (2006). An introduction to ROC analysis. Pattern recognition letters, 27(8), pp.861-874.

[22] Thang, H. M., Viet, V. Q., Thuc, N. D., & Choi, D. (2012, November). Gait identification using accelerometer on mobile phone. In 2012 International Conference on Control, Automation and Information Sciences (ICCAIS) (pp. 344-348). IEEE.

[23] Gupta, J. P., Polytool, D., Singh, N., &Aemwal, V. B. (2014). Analysis of gait pattern to recognize the human activities. IJIMAI, 2(7), 7-16.

[24] Abbadi M., Abadleh A., Alja'afreh S., Halhouli Z. (2019). Machine Learning-Based Approach for Detecting Driver Behavior Using Smartphone Sensors. IJSTR, 8(12), 1057-1060.