Activity With Gender Recognition Using Accelerometer and Gyroscope

Ahmed Sharshar[1] , Ahmed El-ghareeb[2] and Walid Gomaa

Cyber-Physical Systems Lab, Department of Computer Science and Engineering, Egypt-Japan University of Science and Technology, Borg El-Arab, Alexandria, Egypt
Faculty of Engineering Alexandria University, Alexandria, Egypt

Abstract— **Recently, the use of the inertia measurement units (IMU), especially the gyroscope and accelerometer sensors, has increased in the human activity recognition (HAR) due to the increasing use of smartwatches and smartphones. In addition to the high quality and efficiency result in by these sensors,they can capture the data of the body dynamic motion as function of time, then the stream of data is analyzed and processed to classify and predict the action being done, the gender, the health status and many other characteristics. Gender and activity recognition have been deeply studied recently, using various ways to recognize either of them through many interfaces, like voice, image, or inertia measurement motion data. All these types of classifications are crucial in many applications such as recommendation systems, speech recognition, sports tracking, security and most importantly in healthcare. In this research , we present a model based on MoVi dataset to predict gender with activity and see how every activity reflect on gender , using only two IMU sensors on right and left hand and explore the efficiency on using autocorrelation function as a feature extractor along with random forest as classifier..**

I. INTRODUCTION

Human activity and gender recognition have developed greatly with the evolution of the use of the inertia measurement units (IMU) over the years, it is available in smart phones and smart watches. Due to its importance in many applications, great work has been done in every field with the use of the accelerometer and gyroscope sensors. However, unfortunately , there is no such previous work that tried to predict both gender with activity at the same time, Despite its distinct benefits as it gives more information about the user and predict more accurate. Gender with activity recognition can gain high importance to get more knowledge about the user. In this research, we will introduce a new way for classifying activity with gender based on MoVi dataset. Using autocorrelation Function as feature extractor and compare between the two classifying methods to categorize between four activities for both genders.

Datasets collected from IMU sensors are not widely available, and what is available doesn't have the diversity of data we need. however, MoVi dataset contains data for about 21 activities for both genders 30 males and 60 females. every actor has 18 IMU sensors in different places on the body. Each sensor is composed of a 3-axis gyroscope, a 3-axis accelerometer, and a 3-axis magnetometer working with 120 fps. In addition to the global acceleration data, the IMU suit provides 3D displacements, speed, quaternions, and rotational speed for each joint. We worked only for 4 activities, walk, run, bye and clapping for both genders with 80 different volunteers.

There are many approaches to use for classification. the most recently approach is deep neural networks. However, such a method performs well when there is a huge data. Such a size of data is not always available specially when it's IMU sensors data, as such is time consuming and hard to collect. Also, it needs many volunteers with many sensors that can't be always available. Even if such a dataset existed, it can be formed out of very few seconds for each activity like MoVi dataset.

For these reason , we introduce feature extraction method as it's faster and more accurate than using raw data In addition, it did not require such huge amount of data, then the process should including the loading of the raw data into the feature extractor at first, then the classifier. Autocorrelation Function is performing well as a preprocessing tool , specially if it is implemented in deep learning processes. This paper shall report its effects and performance on machine learning algorithms , such as Random Forest. Random Forest classifier is a well known model that features both of high accuracy and a high computation performance, with low resources demanding. Thus, the output of the whole process is a highly reliable methods with expected good experimental results.

In this work we will focus on how answering the question:
1) what is the impact of using accelerometer and gyroscope in the right- and left-hand sensors to predict the activity with the gender?
The rest of this paper is organized as follows. Section 1 is an introduction, Section 2 is literature review where we present the related work in, Section 3 is a about dataset and its properties, Section 4 is about data pre-processing, feature extraction and used classifier models, Section 5 is about the experiment, in Section 6 we presented the results of the experiment and finally Section 7 is our conclusion and our future work.

II. LITERATURE REVIEW

In [1], Using autocorrelation as a feature extraction technique for real-time human activity recognition on a smartphone by three types of sensory signals: acceleration, angular velocity, and rotation displacement; where each is a tri-axial signal. The data was collected from multiple-independent objects for 14 activities of different ages and gender. These data were analyzed using autocorrelation function with certain lag (usually from 10 -15) and took these values as features of the original data. Then using random forest classifier for training and testing, the accuracy of both accelerometer and gyroscope that performed about 80%.
Authors in [2] used an ensemble learning algorithm based on Random Forest which integrated many individual Random Forest, and using a dataset of 19 different activities, they could reach an accuracy of around 93.4%. The training time is significantly small compared to other classification methods.
Authors in [3] used autocorrelation functions with random forest with 500 trees on a dataset that contains 19 different activities using 3-axis of accelerometer, magnetometer, and gyroscope, the results were up to 90.89%. These previous works show good results in using the random forest as a classifier in activity recognition.

The authors in [4] proposed a reliable model for gender recognition based on inertial data of 3 axis accelerometer and gyroscope signals from wearable IMU units that were placed at eight different parts of the human body from 20 different objects (10 males and 10 females). Using wavelet transform as feature extraction and random forest classifier. The highest accuracy obtained by combining the waist sensor with the left cube sensor was about 96.74%. However, using the Right-hand and left-hand sensor, the accuracy was around 89.56% for gender classification only.

Work done in [5] proposed a novel local descriptor method for feature extraction. Which is a multi-kernel local diamond pattern (MK-LDP)? The proposed MK-LDP extracts 2560 features from the raw sensors signals, RFINCA (Relief and neighborhood component analysis (NCA)) selects 512 most meaningful ones from the extracted 2560 features, then these 512 selected features are used as an input of the Support vector machine (SVM) for HAR. The dataset contains 19 activities, where the sensors are connected to chest level, right wrist, left wrist, right leg, and left leg on 4 men and 4 women. These sensors are 3-axis accelerometer, gyroscope, and magnetometer. The training was on 3 cases. These cases are gender classification, activity classification, both gender, and activity classification. The achieved best accuracy rates are 99.47%, 99.71%, and 99.36% respectively.

Here we are trying to make reliable work with the possibility of its application in practical life and to use only 2 IMU sensors in both right and left hand only with a 3-axis accelerometer, gyroscope, and magnetometer. Then, implement this method on smartwatches. Also, we examine the efficiency of using autocorrelation function as a feature extraction method along with random forest as a classifier in determining both activities with gender classification..

III. MoVi DATASET

We used MoVi dataset [6] in this work; an open dataset that is available online that contains two types of data, IMU sensors data and optical motion capture (MoCap). Here we used IMU data composed of 90 participants, 60 of them are females and 30 are males, all wearing natural clothes, performing a collection of 20 predefined everyday actions and sports movements. These actions are (1) Walking, (2) Jogging, (3) Running in place, (4) Side gallop, (5) Crawling, (6) Vertical jumping, (7) Jumping jacks, (8) Kicking, (9) Stretching, (10) Cross arms, (11) Sitting down on a chair, (11) Crossing legs while sitting, (13) Pointing, (14) Clapping hands, (15) Scratching one's head, (16) Throwing and catching, (17) Waving, (18) Pretending to take a picture, (19) Pretending to talk on the phone, (20) Pretending to check one's watch. In each of the five sequences, the actors additionally performed one self-chosen motion (21).

TABLE I
SUBJECTS OVERVIEW OF MOVI DATASET (MEAN(SD))

-	Male	Female
Height (m)	1.65 (0.08)	1.75 (0.06)
Weight (kg)	60.35 (8.03)	72.3 (10.98)
BMI (kg/m2)	22.16 (3.02)	23.6 (3.24)
Age (y)	20.47 (3.59)	23.6 (3.61)

The data were taken form 18 units of IMU distributed over 18 different places on the body as shown in Figure 1. In this work, we were interested in Right and left hand sensors which exist in the same locations as smart watches. These sensors are synchronized with both activities performed and recorded as videos for each object and its MATLAB readings..

Fig. 1. Sensors Placements On Subjects' Bodies

For the IMU system, we used the Noiton Neuron Edition V2 which comes as a suit attached with 18 IMU sensor. Each sensor is composed of a 3-axis gyroscope, a 3-axis accelerometer, and a 3-axis magnetometer working with 120 fps. Besides from the readings from the axis gyroscope, accelerometer, and magnetometer, the IMU suit provides other parameters, these parameters are acceleration, 3D displacements, speed, quaternions, and angular velocity for each joint. In this work, we are interested in acceleration and angular velocity parameters.

TABLE II
SUMMARY OF IMU DATA

Brand and model	Noitom, Neuron Edition V2
Number of sensors	18 Neurons
Sensor	9-axis IMU of 3-axis gyroscope, 3-axis accelerometer, and 3-axis magnetomete
Frame rate	120 HZ
File type	BVH and calculation files

IV. METHODOLOGY

A. *PREPROCESSING*

Based on the dataset collected in [6], we configured a preprocessing infrastructure with a hierarchical architecture. We retrieved the data

- presented as .mat files - and since each file contains different oriented data, we picked the section of data representing the samples' videos: $S1\_Syncronized$. After which, we began our first stage of preprocessing, which is the data segmentation. We segmented the data upon the four activities (walking, running, clapping, waving) chosen from the 21 activities each subject performs. We defined the begging and ending of each activity from the videos of the subjects, and segmented the data from the data file. Each segment is the whole activity in duration. Indeed, the segmentation has a reference of the right and hand sensor modules, in order to implement this whole sought process on a smart watch for all user - whether they are right or left handed.

The second stage is sampling. Each data segment varies in length from the other segments for the other activities performed by the other subjects. We sampled each data segment by 60 time period at the begging, based on the 120 Hz sampling rate , the 60 time periods are equal to 0.5 secs. Thus, we had all the data samples are uniform in length. Thus, we managed to perform the training and testing on uniform dataset smoothly

Furthermore, and by seeking better and more reliable results, we performed an overlapping operation on the data samples. We extended the 60 time periods to 180 time periods samples, each sample is 30 time periods away from the one preceding it.

B. FEATURE EXTRACTION

Based on the preprocessing stages and their output, the dataset has been enlarged tremendously. The files became in scale thousands, and each sample is a 180x32 matrix, 180 row representing time steps, and 32 columns representing the features of both hands. Thus, the data belong to the domain R^{5760}. In order to reduce the complexity of the training and testing processes , we used the ACF (Autocorrelation Function). The Auto correlation Function reduced each feature column into a single scalar value. Thus, we have more valuable input to the classifier instead of using the raw data.
The ACF is used to get the Peasron correlation between an input signal x , a delayed copy of that signal x_t by a specified lag h, the mean of the sample signals $\bar{x}$, and T as the length of the input signal x - See equations 1 and 2. According to the ACF implemented by Gomaa et al. [1], we determined how much each input signal is dependent on its delayed copies. Such conclusion is fruitful when it comes to periodic signals generated by periodic activities, like waving or clapping.

$$acf_x(h) = \frac{\gamma_x h}{\gamma_x h} \tag{1}$$

$$\gamma_x(h) = \frac{1}{T}\sum_{t=1}^{T-h}(x_{t+h} - \bar{x})(x_t - \bar{x}) \tag{2}$$

As mentioned earlier, the ACF reduces the columns representing the features into just one scalar value for each of the sample signal x and the delayed signals x_t. This process is repeated h times , where h is the determined lag. Based on the algorithm implemented in [1], the feature extractor results in an array with length of $K(N+1) - D$, as N represents the lags. And with number of segments equal to M, we would have a feature matrix of $M \times K(N+1)$ - see Algorithm 1: -

Algorithm 1: Auto-Correlation Function

```
Input : Array x, N ∈ R
Output : Array x̂ ∈ R^(M×KN)
initialize x̂ with zeroes, i → 0;
for s ∈ x do
    for h ← 0 to N do
        for  k ← 0 to K − 1 do
            x̂_{i,Kh+k} = acf_k(h) (From Eq. 2)
        end
    end
    i ← i + 1
end
```

C. SUPERVISED LEARNING

After the preprocessing and data preparation, we set a Random Forest classifier for the data samples.

Random Forest [7] is a widely used supervising classifier that is known for its performance and accuracy. The Random Forest uses decision trees in order to label the input sample, and such process is done by the trees decision taking that is held by voting. Furthermore, the Random Forest is based on ensamble learning, which clarifies its performance and reliability , specially against overfitting. Random Forest uses multiple layered trees in order to take a specific decision, which explains how does the Random Forest algorithm share the ensamble training features. In addition, the multi-layered trees of the algorithm include are implemented with the feature of bootstrap aggregation, which improves the accuracy and the stability of the learning process, especially the regression and statistical classification. Such feature allows the classifier to split the data randomly before the training processes, which carries out a performance and makes the classifier less prone to overfitting.

V. EXPERIMENT SETUP

To investigate the effectiveness of our methodology shown in section (4), we should evaluate the performance on the dataset MoVi, introduced in section (3) with Random Forest classifier.

A. DATASET

The data, as shown in section (3), contains many activities for both genders, and for different 18 places in the body So, we extracted only four activities: walking, running, clapping and waving for all the available subjects. Generally, there are 80 different subjects ; 29 subjects are males and 51 are females. As mentioned earlied, the extracted and preprocessed data are from two IMU units only, right and left hand, the reason for this is that some volunteers are left-handed and others are right-handed, so we want to capture all important data from both units.
The data from [6] is composed of nine components Accelerometer-X, Accelerometer-Y, Accelerometer-Z, Gyroscope-X, Gyroscope-Y, and Gyroscope-Z, magnetometer-X, magnetometer-Y, magnetometer-Z, these reading are represented as 5 parameters, acceleration, 3D displacements, speed, quaternions, and angular velocity. Here we worked on the acceleration and the angular velocity only, these two parameters can represent the 3-axis Gyroscope and 3- axis Accelerometer which both are found in smart watches and mobile phones [8]. The dataset was split into fixed-size samples. Each sample represents a 1.5-second signal with overlapping samples with 0.25 seconds and its label indicating the activity and indicating whether the subject is a male or a female. The sampling rate was at 120 Hz, in other words, each sample had 180 sensor readings represented as Matlab rows and 12 columns represent the 3-axis acceleration and 3-axis angular velocity for both right and left hand. Multiple scatter plots of the data are represented in Figures 2,3, and 4 :

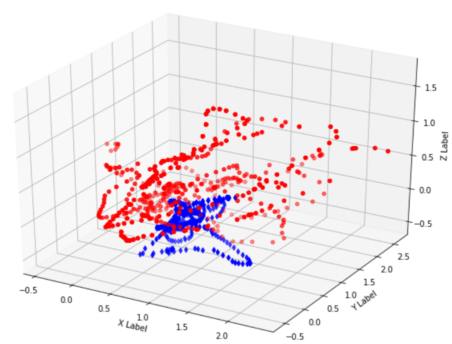

Fig. 2. 20 Samples Representing Data As Genders Only

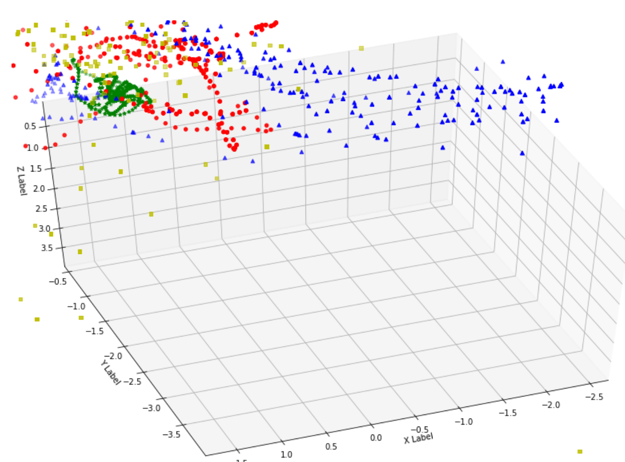

Fig. 3. 5 Samples Representing Data As Activities Only

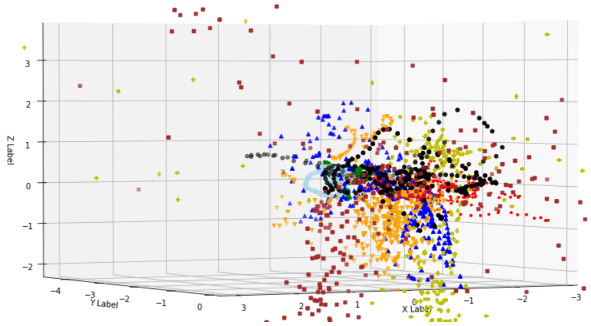

Fig. 4. 5 Samples Representing Data As All Combinations

TABLE III
SAMPLES INFORMATION

Attribute	Value
Number of activities	4
Number of subjects	80 (29 males and 51 females)
Total number of samples	4610

B. MODEL

In our model, the procedure was as following: we started by loading the data files automatically into the classifier, since the files are named by the activity and the gender related to them, then we applied feature extraction as shown in Section 4.2 (choosing the number of columns K to be 6 and the number of lags N to be 15). After that, we split the features matrix and the corresponding labels into training and test sets (70% for training and 30% for test). we ran the experiment 10 times (each time in different part of the data with the same ratio, 70% for test and 30% for training) and took the average metrics of accuracy, precision, recall, and F-score.
As for the classifier, we used Random forest, as in [9], with 8 classes (for 4 activities for both genders); we set the number of decision trees in the forest to 500 trees.

VI. RESULT AND DISCUSSION

In this section, we will represent the results of the experiment setup we have explained in section 5. We performed the experiment 10 times and took the mean of all results using 10 lags in Auto-correlation function as feature extractor and 500 trees of Random forest classifier.
Firstly, we checked the results from the activities only, 4 classes for the 4 activities: walking, running, clapping and waving with no differences between the genders. The results were around 93.5%, as we see in Table 4, accuracy is 93.52%, precision is 93.75%, recall is 93.52% and F_score is 93.45%. As we see in Figure 5 of the confusion matrix of this experiment, the classifier can predict walking , clapping, and running activities accurately and precisely, with a good accuracy, however, the accuracy of the waving activity is slightly low and the classifier mistakes the waving with walking. It can be explained through two reasons. The first reason, the waving activity is very short in length, so its data files cannot be satisfied for the classifier to learn exactly its features and that we used sensors from two hands only and the walking can be differentiated from waving if we had added the leg sensors. Nevertheless, the overall accuracy is still good using left-and right-hand sensors only.
Secondly, we tried to recognize gender by only using 4 activities, so we made 2 classes where one is for males and the other is for females, every classifier contains all the 4 activities for each gender. The overall results were around 85%, as we see in Table 4, the accuracy is 84.50%, precision is 85.24%, recall 84.50% and F_score is 83.93%. In Figure 6, based the confusion matrix of this experiment, the classifier can recognize females with a 95% accuracy; however, it cannot recognize males satisfactorily (with a 67% accuracy only). The reason for this is the nature of the dataset, as we mentioned in section (3), the dataset contains 80 different subjects; on the other hand, out of 80, 51 of the volunteers are females, this makes the data biased toward females which makes it easier for the classifier to recognize them.
Lastly, we tried to recognize gender along with activity, so we made 8 classes, each activity gets 2 classes for each gender. The results were around 85%, as we see in Table 4, the accuracy is 82.77%, precision is 84.07%, recall 82.77% and F_score is 82.34%. Also, in Figure 7 of the confusion matrix of this experiment, the overall accuracy of females activities is much higher than males activities , also we observed that the

activities of clapping and walking have different features that make them have higher accuracies than other activities, unlike waving which seems to not have a big difference between men and women. The overall accuracy is still sufficient.

TABLE IV

SUMMARY OF THE RESULTS

-	Activities	Gender	Gender with Activities
accuracy	93.52%	84.50%	82.77%
precession	93.75%	85.24%	84.06%
recall	93.52%	84.50%	82.77%
F_score	93.45%	83.92%	82.34%

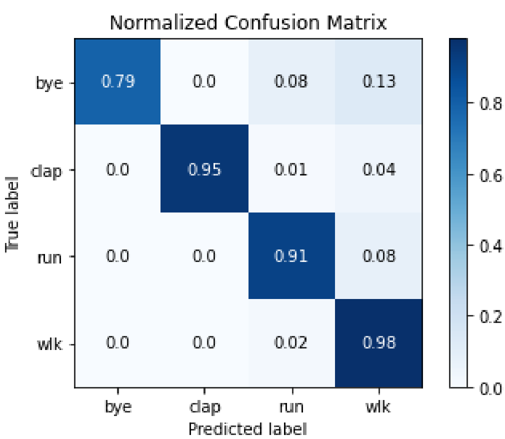

Fig. 5. Normalized Confusion Matrix For Activities Only

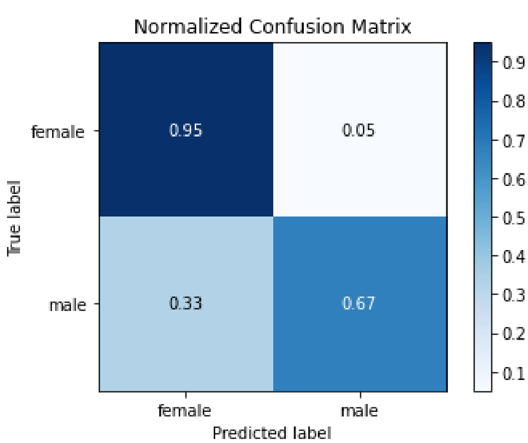

Fig. 6. Normalized Confusion Matrix For Gender Only

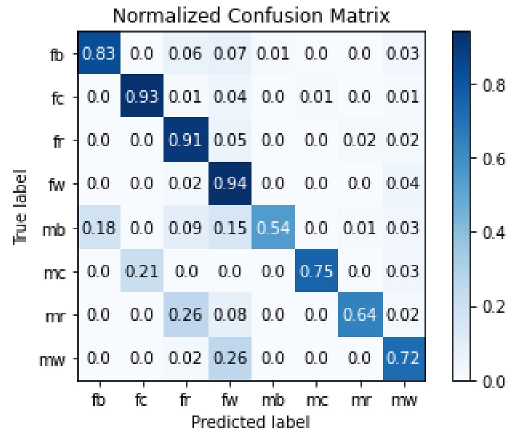

Fig. 7. Normalized Confusion Matrix For All Combinations

From these results, we can remark that using sensors from smart watches only can be sufficient for activities with gender recognition with good accuracies, low resources, and high performance. And we assume that if this dataset was not biased in terms of gender, the results would be better.

VII. CONCLUSIONS AND FUTURE WORK

In this work, we introduced a new classification method to predict both activity with gender for humans. Such method can be crucial for recognition systems and healthcare services. We used MoVi dataset to extract 4 activities of Walking, Running, Clapping and Waving for both genders from 80 different subjects. These data are extracted from only 2 IMU sensors placed on Right and left hands only. Each sensor uses 3-axis acceleration, gyroscope and magnetometer. we used Auto Correlation function as feature extractor and Random Forest as classifier. We made 3 different experiments, one to predict activity, one to predict gender, and one to predict gender with activity. All the 3 experiments showed good results with high computational performance, as this method proved to not demand high resources. Thus, the results are promising and encouraging for its implementation in the meantime technological devices, such as smart watches.

In the future, we seek to explore another classifier with a high potential of achieving better results. We target the Convolution Neural Network (CNN) for deep learning and Support Vector Machine (SVM) besides the Random Forest. Also, we seek to increase number of activities and the number of the used sensors to see the effect of increasing one or both of them on the performance and the resulted in metrics.

References

[1] W. Gomaa, R. Elbasiony and S. Ashry, "ADL Classification Based on Autocorrelation Function of Inertial Signals," 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), Cancun, 2017, pp. 833-837, doi: 10.1109/ICMLA.2017.00-53.

[2] Z. Feng, L. Mo and M. Li, "A Random Forest-based ensemble method for activity recognition," 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, 2015, pp. 5074-5077, doi: 10.1109/EMBC.2015.7319532.

[3] M. G. Abdu-Aguye and W. Gomaa, "Competitive Feature Extraction for Activity Recognition based on Wavelet Transforms and Adaptive Pooling," 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 2019, pp. 1-8, doi: 10.1109/IJCNN.2019.8852299.

[4] Mostafa, Abeer & Barghash, Toka & Assaf, Asmaa & Gomaa, Walid. (2020). Multi-sensor Gait Analysis for Gender Recognition. 629-636. 10.5220/0009792006290636.

[5] T. Tuncer, F. Ertam, S. Dogan and A. Subasi, "An automated daily sport activities and gender recognition method based on novel multi-kernel local diamond pattern using sensor signals," in IEEE Transactions on Instrumentation and Measurement, doi: 10.1109/TIM.2020.3003395.

[6] Saeed Ghorbani, Kimia Mahdaviani, Anne Thaler, Konrad Kording, Douglas James Cook, Gunnar Blohm, and Nikolaus F. Troje. MoVi: A Large Multipurpose Motion and Video Dataset, 2020.

[7] L. Breiman, “Random forests,” Machine learning, vol. 45, no. 1, pp. 5– 32, 2001.

[8] A. Jain and V. Kanhangad, "Investigating gender recognition in smartphones using accelerometer and gyroscope sensor readings," 2016 International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT), New Delhi, 2016, pp. 597-602, doi: 10.1109/ICCTICT.2016.7514649.

[9] Feshier, Osama & Nafea, Yousef & Hesham, Ahmed & Gomaa, Walid. (2020). Gait-Based Person Identification Using Multiple Inertial Sensors. 10.5220/0009791506210628.