Hybrid Architecture for Gender Recognition Using Smartphone Motion Sensors

Erhan Davarci, Emin Anarim
Electrical and Electronics Engineering Department, Bogazici University, Bebek, Istanbul 34342, Turkey
{erhan.davarci,anarim}@boun.edu.tr

***Abstract*—Motion sensor data in smart devices can be used as a side-channel to capture user's behavioral biometrics. In this paper, we investigate the feasibility of using smartphone motion sensors to detect the gender of the user. The main idea behind our study is based on behavioral differences of male and female users in touching and holding the smartphones. In order to implement our method, we collect data from 100 subjects while they are performing different activities like sitting, standing and walking. Our experiments point out that tapping behaviors are very discriminative for gender recognition but their significance decreases while users are walking. To address this issue, we also implement user activity detection and propose a hybrid model to predict gender in different user activities. In this context, user activity is firstly detected and then gender is predicted correspondingly. Consequently, we show that gender recognition is implicitly performed with a success rate of 85% in sitting and standing activities, whereas 83% success rate is acquired in walking scenario.**

I. Introduction

Modern smart devices especially smartphones are equipped with built-in motion sensors like accelerometer, gyroscope, and orientation sensors to monitor the movement and orientation of the device. Although these sensors provide higher user experience on smart devices, they can also be used as side-channels. Recent studies like [1], [2], [3], [4] have signified that different information about users and their inputs can be obtained by using device motion sensor data as a side-channel. In particular, the results of the conducted studies indicate two significant facts: First, motion sensors may carry side-channel information and they can be used to capture user sensitive data like PIN or password. Second, behavioral biometrics of a user can be obtained from the smartphone motion sensors for user identification.

In this study, we focus on how smartphone touching and holding behaviors differ for male and female users. We extract different feature sets by analyzing correlation between accelerometer sensor output and the user's tap behavior to reveal the discriminative patterns for gender. For this purpose, we implement the Android application, *BalloonLogger*, which collects accelerometer sensor data while users playing balloon popping game in different activities.

After data collection, we derive different features by using signal processing techniques. We also investigate the effect of user motion on gender recognition and indicate that user motion introduces new patterns on sensor data and affects the importance of features. Therefore, we also implement an activity detection to handle different user activities. Finally, our proposed algorithm yields the boolean answer whether the user is a male or a female.

Actually, identification of gender has various usage scenarios. For instance, user interaction with the device can be improved by automatically customizing screen or applications on the device. Gender information can also be used to recommend new features or applications and to provide more relevant search results to users. In addition, authentication mechanisms can benefit from gender identification as a secondary layer of authentication to improve security.

The major contributions of our paper are: First, we introduce the possibility that user gender can be recognized by evaluating the accelerometer sensor readings as a side-channel. For this purpose, we collect new data from users while they are playing our balloon popping game in different activities. Second, we analyze the effects of different feature sets and examine the success of our model for different user activities. Third, we implement user activity detection system and propose a hybrid architecture to implicitly recognize gender in different activity scenarios which are not addressed in previous studies.

The rest of this paper is organized as follows: In Section II, we review the related work. Section III describes the details of our scheme and presents the evaluated results. In Section IV, we introduce our proposed method for gender recognition and also investigate activity detection. Finally, we conclude this paper in Section V.

II. Related Work

Actually, motion sensors on smart devices can be exploited as a side-channel for different purpose like inferring user inputs, user identification or authentication and obtaining user behavioral biometrics. The authors showed that user inputs on touch screen can be inferred by utilizing accelerometer and gyroscope sensors in [5], [6]. Researchers have also investigated the possibility of user identification by using motion sensors and introduced the notion of touch behavioral biometrics [7], [1].

Furthermore, there are some studies focused on user biometrics like age group and gender by using motion sensors. For instance, motion sensors were used to determine user age group in [8], [9], [10]. On the other hand, gender recognition from keystroke dynamics data and touchscreen swipes was investigated in [11]. Based on their limited data set, the authors showed that the best results are 64.76% for the keystroke data and 57.16% for the swipes data. In [12], an approach for

ISBN: 978-9-0827-9706-0

gender recognition using touchscreen gestures was proposed. A two-dimensional attribute maps were obtained by forming images in feature extraction, and classification accuracy of 92.96% was achieved when all the gestures are combined. However, their dataset included only specific gestures and they did not consider tap events or user mobility.

Moreover, gender classification using human gait cycle based on accelerometer signals was performed in [13]. The best accuracy rates are 68.2% and 65% for different walking sequences. In [14], the authors introduced a method to predict gender with activity. They used dataset collected from smartphone kept in the user's pocket. They obtained accuracy rates up to 95% with different classifiers. However, the age range in their dataset is narrow and there is no detail about no overlap of user's samples between training and test sets. [15] also analyzed gender recognition using user gait data and stated the best accuracy of 96.3% using the bagged tree classifier with cross validation. However, cross validation may not reflect actual results because train and test sets can include data from the same user. Another issue in these studies is that they did not consider the situation where users are not mobile.

III. Methodology

A. Data Acquisition

For data collection, we develop the Android application *BalloonLogger* and perform an experiment with smartphones to collect new data from users. This application logs accelerometer data in x, y and z directions while users touch the screen. We acquire data from users in both static and dynamic scenarios. While users are sitting or standing position in *Static Scenario*, they are walking in *Dynamic Scenario*. When the application is launched, a balloon through different colors and sizes appears on a random position of the touch screen. As a user pops the balloon by touching it on the screen, next one is appeared with a random color, size and position. A total of 50 balloons are created in this way during the experiment. Data collection process takes around 5 minutes for each user. Users are not guided about how they hold, touch the device or their walking speed while playing *BalloonLogger*.

In this experiment, we collect data from 50 male (average age=30.1, min=17, max=57) and 50 female (average age=36.3, min=18, max=57) users. In our data set, there are 5206 taps data for *Static Scenario* and 5219 taps for *Dynamic Scenario*. In the analysis of sensor readings, the sampling rate of the sensor data is critical. Low sampling rate cannot give enough information to process sensor data. Therefore, we use popular Android smartphones like Samsung and LG during our experiment and they have all 100 Hz sampling rate.

B. Feature Extraction

Accelerometer measures the acceleration force on all three dimensions, i.e., it measures the velocity change rate due to the external force applied on smartphones by users [4]. In order to better observe the applied force and corresponding changes in sensor data, we use $AccSum$ term which is 2-norm of acceleration vector and computed as:

$$AccSum = |A|^2 = A_x^2 + A_y^2 + A_z^2, \tag{1}$$

where A_x, A_y, A_z are acceleration values in x, y, z dimensions, respectively. $AccSum$ is directly proportional to applied force due to user touches or movements.

Action_Down, *Action_Hold* and *Action_Up* are three consecutive phases occur during the tap event. When the user taps on the touchscreen, the smartphone will move downward and this is *Action_Down* phase. When the *Action_Down* finished; the smartphone will stop, *Action_Hold* phase. After, the user removes his finger and the hand holding the smartphone will cause the smartphone return to its starting position in *Action_Up* phase. When we analyze $AccSum$ values during the tap event, we observe the similar pattern in tap locations. This unique pattern in a tap event is shown in Fig. 1.

Therefore, we firstly extract features from $AccSum$ based on this unique pattern in tap events. The following 6 features are extracted for each tap event after normalizing $AccSum$:

1) P_1: The peak value at the end of *Action_Down*.
2) P_2: The minimum value in *Action_Down*.
3) P_3: Magnitude difference between P_1 and P_2 values.
4) P_4: Time difference between P_1 and P_2.
5) P_5: Time difference between starting points of *Action_Down* and *Action_Up* phases.
6) P_6: Variance of $AccSum$ values throughout *Action_Up*.

These features in tap events are highly correlated with the user's touch and hold behavior. In fact, P_1 measures the magnitude of tapping finger, P_2 measures the magnitude of force due to the hand holding the smartphone, P_3 and P_4 measure the change rate, P_5 determines the tap time and P_6 measures the fluctuation.

Then, we extract some statistical features about x, y, z axes and $AccSum$ for tap events: *mean* (μ), *standard deviation* (σ), *variance* (σ^2), *root-mean-square deviation (RMSD), energy, skewness, kurtosis and signal magnitude area (SMA)*. Most features such as *mean*, *standard deviation* and *variance* are self-explanatory. *RMSD* measures the differences between signal values and mean, *energy* computes signal energy, *skewness*

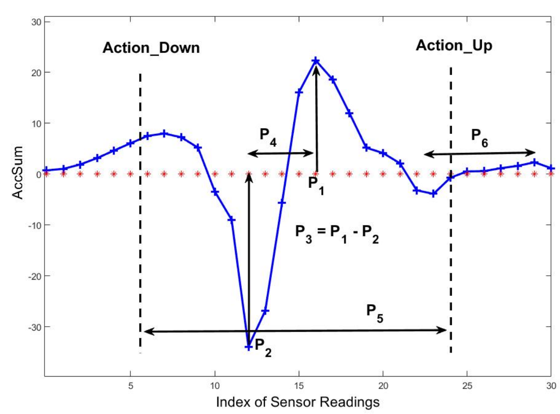


Fig. 1: The unique pattern in a tap event.

is a measure of symmetry, *kurtosis* is a measure of whether the data are heavy-tailed relative to a normal distribution, *SMA* measures the fluctuation degree of the signal. They are calculated as:

$$RMSD = \sqrt{(\frac{1}{n})\sum_{t=1}^{n}(s_t - \mu)^2} \tag{2}$$

$$Energy = \sum_{t=1}^{n}|s_t|^2 \tag{3}$$

$$Skewness = \frac{\sum_{t=1}^{n}(s_t - \mu)^3}{n \times \sigma^3} \tag{4}$$

$$Kurtosis = \frac{\sum_{t=1}^{n}(s_t - \mu)^4}{n \times \sigma^4} \tag{5}$$

$$SMA = \frac{1}{n}\sum_{t=1}^{n}|s_t| \tag{6}$$

In addition to these features, we also examine the distribution of each tap event. All tap events do not have the same number of samples even for different taps of the same user. Therefore, we firstly transform data of each tap to frequency domain and then retransform it to the time domain with reduced set of coefficients to reconstruct approximation of the original signal. This operation preserves general shape of the signal and eliminates noisy high frequency components. For this purpose, the Discrete Cosine Transform (DCT) of n-length data with transform length n and Inverse Discrete Cosine Transform (IDCT) with transform length m where $n \geq m$ are computed, respectively. By setting $m = 10$, we fix all tap events to 10 samples in reconstruction so we can use each of these 10 values as a feature. Correspondingly, one-dimensional N points DCT and IDCT equations are given below.

$$X_c[k] = \sum_{n=0}^{N-1} x[n]cos\frac{\pi(2n+1)k}{2N} \quad k = 0 : N-1 \tag{7}$$

$$x[n] = \frac{1}{N}X[0] + \frac{2}{N}\sum_{k=1}^{N-1} X[k]cos\frac{\pi(2n+1)k}{2N} \quad n = 1 : N-1 \tag{8}$$

In order to classify extracted tap features, we use Random Forest classifier with 10-fold cross validation. The accuracy rates for both static and dynamic scenarios are shown in Table I. We obtain 84.6% success rate for *Static Scenario* in which users are sitting or standing. In other words, we can determine a given tap whether belongs to male or female user with 84.6% accuracy. However, we have only 72.4% success rate for *Dynamic Scenario* in which users are walking. Therefore, these results point out that when we use only tap event features, their importance decreases when users are walking. This is because when users are mobile, the reactions of smartphones due to tap event are suppressed by user movement so tap features will become less distinctive in terms of gender recognition.

TABLE I: Gender Classification with Tap Features Only

User Activity	Accuracy Rate
Static Scenario (users are sitting or standing)	84.6%
Dynamic Scenario (users are walking)	72.4%

Up to now, all extracted features are tap features and belong to tap events. We detect tap locations in raw sensor readings and then extract 48 features for each tap: 6 specific features (P_1 to P_6) and 10 IDCT features from $AccSum$, other 32 statistical features for x, y, z axes and $AccSum$. Now, we also consider sensor data other than tap locations and extract features from them. In this manner, we first transform raw sensor data into time windows and then extract features for each time window. We test different window sizes and choose 100 sample points as an optimum window size which corresponds to one second length sensor data.

After, the same statistical features about three axes and $AccSum$ are extracted: 8 features/axis x 4 (x, y, z and $AccSum$) = 32 features for each time window. In addition, we apply DCT and IDCT operations on time windows and extract 20 features for each time window by setting $m = 20$. In total, there are 52 features extracted for each time window sample and we call these features as motion features.

The effects of tap and motion features for static and dynamic scenarios are given in Table II. By analyzing the results, we can conclude that in *Static Scenario* tap features are very distinguishing for gender recognition but when users are mobile, the effects of tap events are suppressed so motion features can be used to detect gender.

TABLE II: The Effects of Different Features on Gender Classification

	Accuracy Rate	
Features	**Static Scenario**	**Dynamic Scenario**
Only tap features	**84.6%**	72.4%
Only motion features	72.1%	**82.2%**
Average of tap and motion features	75.5%	79.8%

C. Classification

In this section, performances of different classification algorithms are compared by using $Python3$ environment. 6 different classification methods are experimented: Logistic Regression, k-Nearest Neighbor, Support Vector Classifier (SVC), Random Forest Tree, Decision Tree, and Artificial Neural Network (ANN). In classification, tap features are used for *Static Scenario* and motion features obtained from time windows are used for *Dynamic Scenario*.

Additionally, we implement Information Gain Based Feature Selection method on our feature sets to identify and remove the irrelevant or less important features in order to achieve better results. In this method, we calculate the information gain (also called entropy) for each feature and ignore features that have a lower gain.

Although k-fold cross validation (CV) is commonly used in age-group and gender detection studies, it can be misleading with its results because data sets may include lots of data samples for each user. In this manner, we implement and perform leave-one-user-out cross-validation (LOUOCV) as an evaluation method. In LOUOCV, the classifier is trained with all but one user data and this is repeated for all users. Table III shows both 10-fold CV and LOUOCV accuracy results for different classifiers. We also use $GridSearchCV$ to tune hyperparameters of our classifiers. SVC *(with kernel=rbf, C=100, gamma=1)* gives the highest accuracy rates of 86% and 84% for static and dynamic scenarios, respectively.

TABLE III: Results for Different Classification Methods

	10-fold CV Accuracy		**LOUOCV Accuracy**	
Classification Method	**Static Scenario**	**Dynamic Scenario**	**Static Scenario**	**Dynamic Scenario**
Logistic Regression	75.3%	80.6%	75.0%	79.0%
k-NN with k=10	85.7%	83.7%	80.0%	81.0%
SVC	**89.7%**	**86.3%**	**86.0%**	**84.0%**
RandomForest Tree	87.6%	85.1%	81.0%	83.0%
Decision Tree	81.7%	80.7%	81.0%	80.0%
ANN	86.6%	84.4%	84.0%	82.0%

IV. Proposed Hybrid Architecture for Gender Recognition

A. Main Framework

In this section, we propose new architecture for gender recognition using motion sensor data which is shown in Fig. 2. Our experiments show that tap features are very distinguishing for gender classification while users are standing still. On the other hand, when users are mobile, their importance decreases because of movement effect. Therefore, we should consider all sensor data and extract motion features from them while users are walking. In order to do this, we should first detect user activity then use suitable feature set, tap or motion features, to recognize gender. In prediction phase, we can use corresponding pre-trained model according to user activity.

For gender recognition, there are generally two groups of studies use motion sensors in the literature. In the first group, touch behaviors of users are used for gender recognition but they are not successful while users are mobile because of walking effects on touch reactions. On the other hand, data about user mobility like gait data are commonly used in the second group of studies. However, they are not viable while users are not mobile. Therefore, our proposed method is a good candidate to solve these limitations in related studies.

B. Activity Detection

For activity detection, we transform raw sensor data into time windows by setting window size as 200 sample points

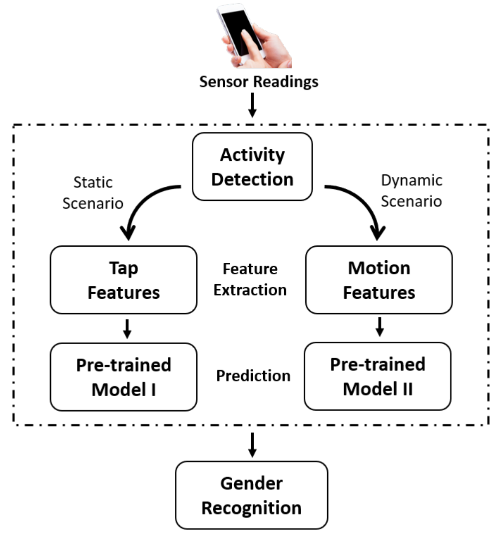


Fig. 2: Proposed architecture for gender recognition.

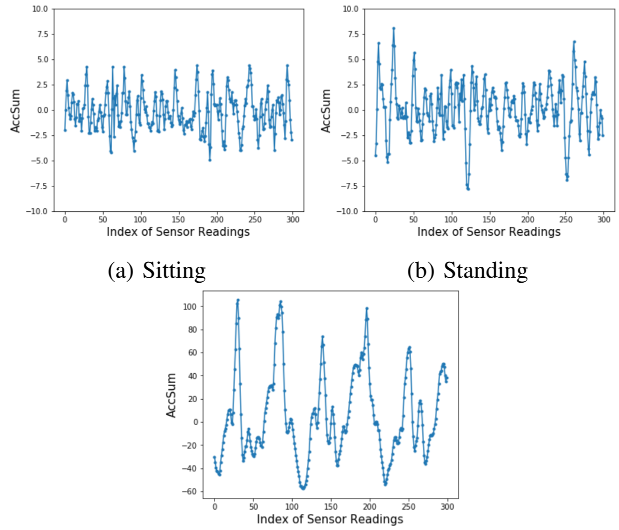
corresponding to two seconds of raw sensor data and perform 75% overlap between windows. 32 statistical and 20 IDCT features in section III-B are extract for each time window. After, we firstly check if our data are clustered by performing Hopkins test which measures the cluster tendency of data, (9) gives its formula. H value close to 1 tends to indicate the data are highly clustered and 0 means data are uniformly distributed. Since we obtain H value of 0.95, we can confirm that our data are highly clustered. Then, we also use Silhouette Algorithm to determine optimum number of clusters and obtain the highest Silhouette Score with $k = 2$.

$$H = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i} \tag{9}$$

Moreover, examples of $AccSum$ due to external forces created by different activities are shown in Fig. 3. Actually, $AccSum$ values are relatively constant in sitting and standing activities, whereas they are highly variable and the range of values is greater in walking activity due to motion effect. Therefore, we can handle this detection issue as two-class classification task: non-motion and motion. We classify sitting and standing activities as non-motion, and walking activity as motion scenario. In classification, we test different classifiers with LOUOCV and their results are exhibited in Table IV.

TABLE IV: Results for Activity Detection

Classification Method	**Accuracy Rate**
Logistic Regression	93.0%
k-NN with k=10	91.0%
SVC	98.0%
RandomForest Tree	91.5%
Decision Tree	90.0%
ANN	**99.0%**



(a) Sitting (b) Standing

(c) Walking

Fig. 3: Example $AccSum$ graphs for different activities.

By using all sensor data of users in balloon popping game, we obtain the highest accuracy of 99% with Artificial Neural Network. We use $Adam$ as optimization algorithm for training the neural network and tune its hyperparameters as follows: *Activation function=ReLU,* β_1*=0.9,* β_2*=0.999, learning rate* (α)*=0.001, hidden layer sizes=(50,100,100,100,50).*

When our activity detection algorithm is implemented, success rates of gender recognition using different length of sensor data are depicted in Table V. We take 5, 10 and 20 seconds sensor data from collected data of users and assume there is at least one tap event per second. Note that, if we take 10 seconds sensor data, the final success rate for gender recognition becomes 85% in *Static Scenario* and 83% in *Dynamic Scenario*.

TABLE V: Gender Recognition Results with Activity Detection

	Accuracy Rate	
Length of Sensor Data	**Static Scenario**	**Dynamic Scenario**
5-sec window	82.5%	81.0%
10-sec window	85.0%	83.0%
20-sec window	85.0%	83.0%

V. Conclusion

In this study, we have investigated the feasibility of detecting user gender by analyzing accelerometer sensor data on smartphones. In particular, we collected new dataset from 100 users while they are playing balloon popping game in both static and dynamic scenarios. Our experiment demonstrated that tap features are very distinguishing for gender recognition while users are sitting or standing (*Static Scenario*) whereas we used motion features while users are walking (*Dynamic Scenario*) because the reactions of smartphones caused by tap events are suppressed by motion effect. In the proposed approach, we first detect user activity with a success rate of 99%, then we recognize user gender using suitable feature sets and classifiers. The results of the paper show that our model can recognize the user gender successfully up to 85% in *Static Scenario*, and 83% in *Dynamic Scenario*. Actually, we conclude that identification of user gender can be used to improve user interaction with smart device and it can also be used as secondary authentication mechanism in addition to password or fingerprint. Furthermore, our proposed method can be improved by considering different user activities and expanding our dataset with more users. These subjects will be studied as future works.

References

[1] C. Bo, L. Zhang, X.-Y. Li, Q. Huang, and Y. Wang, "Silentsense: Silent User Identification via Touch and Movement Behavioral Biometrics," in *Proceedings of the 19th Annual International Conference on Mobile Computing & Networking*. ACM, 2013, pp. 187–190.

[2] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," *SIGKDD Explor. Newsl.*, vol. 12, no. 2, pp. 74–82, Mar. 2011. [Online]. Available: http://doi.acm.org/10.1145/1964897.1964918

[3] A. J. Aviv, "Side Channels Enabled by Smartphone Interaction," Ph.D. dissertation, Pennsylvania State University, 2012.

[4] Z. Xu, K. Bai, and S. Zhu, "Taplogger: Inferring user inputs on smartphone touchscreens using on-board motion sensors," in *Proceedings of the Fifth ACM Conference on Security and Privacy in Wireless and Mobile Networks*. ACM, 2012, pp. 113–124.

[5] E. Owusu, J. Han, S. Das, A. Perrig, and J. Zhang, "ACCessory: Password Inference Using Accelerometers on Smartphones," in *Proceedings of the Twelfth Workshop on Mobile Computing Systems & Applications*, ser. HotMobile '12. ACM, 2012, pp. 9:1–9:6. [Online]. Available: http://doi.acm.org/10.1145/2162081.2162095

[6] E. Miluzzo, A. Varshavsky, S. Balakrishnan, and R. R. Choudhury, "Tapprints: Your Finger Taps Have Fingerprints," in *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys '12. New York, NY, USA: ACM, 2012, pp. 323–336.

[7] A. De Luca, A. Hang, F. Brudy, C. Lindner, and H. Hussmann, "Touch me once and i know it's you!: Implicit authentication based on touch screen patterns," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2012, pp. 987–996.

[8] T. Nguyen, A. Roy, and N. Memon, "Kid on the phone toward automatic detection of children on mobile devices," *Computers and Security*, vol. 84, pp. 334–348, 2019.

[9] E. Davarci, B. Soysal, I. Erguler, S. O. Aydin, O. Dincer, and E. Anarim, "Age group detection using smartphone motion sensors," in *2017 25th European Signal Processing Conference (EUSIPCO)*, 2017, pp. 2201–2205.

[10] E. Davarcı, B. Soysal, İ. Ergüler, and E. Anarim, "Side channel analysis on android smartphones," in *2016 24th Signal Processing and Communication Application Conference (SIU)*. IEEE, 2016, pp. 553–556.

[11] M. Antal and G. Nemes, "Gender recognition from mobile biometric data," in *2016 IEEE 11th International Symposium on Applied Computational Intelligence and Informatics (SACI)*, May 2016, pp. 243–248.

[12] A. Jain and V. Kanhangad, "Gender recognition in smartphones using touchscreen gestures," *Pattern Recognition Letters*, vol. 125, pp. 604–611, 2019.

[13] S. I. Gillani, M. A. Azam, and M. Ehatisham-ul-Haq, "Age estimation and gender classification based on human gait analysis," in *2020 International Conference on Emerging Trends in Smart Technologies (ICETST)*, 2020, pp. 1–6.

[14] A. Sharshar, A. Fayez, Y. Ashraf, and W. Gomaa, "Activity with gender recognition using accelerometer and gyroscope," in *2021 15th International Conference on Ubiquitous Information Management and Communication (IMCOM)*. IEEE, 2021, pp. 1–7.

[15] T. Meena and K. Sarawadekar, "Gender recognition using in-built inertial sensors of smartphone," in *2020 IEEE REGION 10 CONFERENCE (TENCON)*. IEEE, 2020, pp. 462–467.