ORIGINAL ARTICLE

Adam or Eve? Automatic users' gender classification via gestures analysis on touch devices

Alfonso Guarino[1] · Nicola Lettieri[2] · Delfina Malandrino[1] · Rocco Zaccagnino[1] · Carmine Capo[1]

Abstract

Gender classification of mobile devices' users has drawn a great deal of attention for its applications in healthcare, smart spaces, biometric-based access control systems and customization of user interface (UI). Previous works have shown that authentication systems can be more effective when considering soft biometric traits such as the gender, while others highlighted the significance of this trait for enhancing UIs. This paper presents a novel machine learning-based approach to gender classification leveraging the only touch gestures information derived from smartphones' APIs. To identify the most useful gesture and combination thereof for gender classification, we have considered two strategies: *single-view* learning, analyzing, one at a time, datasets relating to a single type of gesture, and *multi-view* learning, analyzing together datasets describing different types of gestures. This is one of the first works to apply such a strategy for gender recognition via gestures analysis on mobile devices. The methods have been evaluated on a large dataset of gestures collected through a mobile application, which includes not only scrolls, swipes, and taps but also pinch-to-zooms and drag-and-drops which are mostly overlooked in the literature. Conversely to the previous literature, we have also provided experiments of the solution in different scenarios, thus proposing a more comprehensive evaluation. The experimental results show that *scroll down* is the most useful gesture and *random forest* is the most convenient classifier for gender classification. Based on the (combination of) gestures taken into account, we have obtained F1-score up to 0.89 in validation and 0.85 in testing phase. Furthermore, the multi-view approach is recommended when dealing with unknown devices and combinations of gestures can be effectively adopted, building on the requirements of the system our solution is built-into. Solutions proposed turn out to be both an opportunity for gender-aware technologies and a potential risk deriving from unwanted gender classification.

Keywords Touch gesture · Automatic gender classification · Mobile devices · Machine learning · Authentication systems

1 Introduction

Nowadays, smartphones are nearly ubiquitous and commonly used to perform daily tasks such as banking, messages, photos, as well as browsing, connecting with others through social media, and storing sensitive data. Therefore, it is essential that these devices perform reliable user authentication to prevent impostor accesses. The results reported in the literature [9, 34, 38, 57, 60] indicate that the authentication performance can be improved by augmenting traditional biometric traits with soft biometric traits. Jain et al. [38] have shown improved performance when soft biometric traits, such as gender, have been incorporated into user authentication that employs face and fingerprint as primary features. Park et al. [57] achieved

✉ Alfonso Guarino
  alguarino@unisa.it

  Nicola Lettieri
  n.lettieri@inapp.org

  Delfina Malandrino
  dmalandrino@unisa.it

  Rocco Zaccagnino
  rzaccagnino@unisa.it

  Carmine Capo
  carmine.pc94@gmail.com

[1] University of Salerno, Fisciano, Italy

[2] National Institute for Public Policy Analysis (INAPP), Rome, Italy

improved performance when soft biometric traits such as gender and ethnicity are included in face recognition. Similarly, Idrus et al. [34] have shown performance enhancement when soft biometrics such as gender and age have been combined with the behavioral biometric traits. More recently, Ranjan et al. [60] have proved that a system including face detection, landmark localization, pose estimation, and gender recognition has superior performance, compared to a load of previous and different models, in recognizing users. Chai et al. [9] demonstrated the feasibility of boosting palmprint identification with gender information using convolutional neural networks.

Moreover, the gender information has seeped into the human–computer interaction research field for a long time. Several papers show the benefit derived from gender recognition (e.g., [22, 50, 75]) as well as, in more general terms, others highlight the significance of diversity-aware user interfaces (UI) and systems (e.g., [42, 77]). Back in 2000, D. Passig et al. [58] found that there is a significant difference in the level of satisfaction between boys and girls depending on the UI's design. Later on, B. Park et al. [56] suggested that customized UIs should include gender among other factors such as culture. In recent times, F. Batarseh et al. [4] highlighted that UI's colors could be customized based on gender. T. Ling et al. [47] found that gender plays a vital role in the perception of mobile devices' UIs within learning systems and consequently can affect students performance. Very recently, S. Sohail et al. [68] discovered that there is a significant difference between males and females in how they perceive gaming environments with different typographic factors. This was also proved by A. Jamil et al. [40] when analyzing other aspects of gaming UIs.

A software capable of adapting its UI according to the gender of the user could be very useful in scenarios where the actual device user cannot be known in advance: think about, for instance, a set of mobile devices randomly provided to employees (or students) of a company (in a school / university laboratory) at the entrance.

In this paper, we focus on gender classification based on machine learning and the analysis of different gestures datasets (see Fig. 1 for a visual abstract). In particular, we consider the usefulness of touch gestures on mobile devices in soft biometrics. Such gestures are the primary way to control these devices and the applications running on them.

1.1 The proposed approach

We collected the gestures datasets using mobile devices with touchscreen through an Android app, forcing users to perform specific touch gestures. The idea is that, the gestures collected carry with them also behavioral data during a user's interaction with the smartphone. We are interested
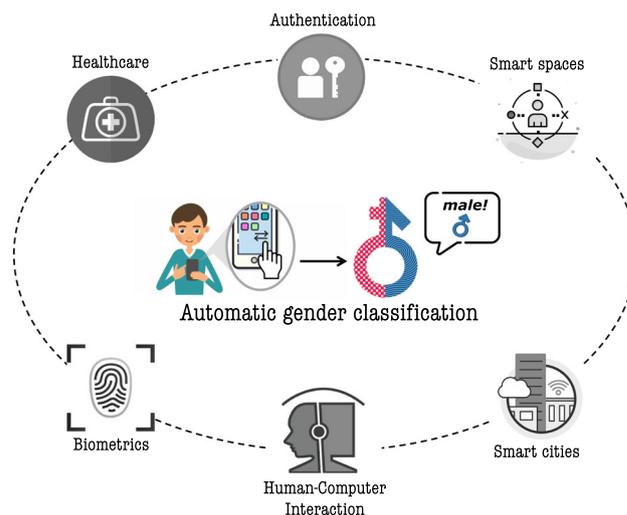
Fig. 1 Visual abstract of our proposal. The solution presented here (colored) classifies users' gender leveraging on touch gestures. It can be used to improve authentication performance, to enhance human–computer interaction, or as part of healthcare and smart spaces (in grey-scale). The developed app is just an instance of a variety of applications ones might use to collect gestures data in order to improve recognition tasks. At this stage of the project, we have not fully formalized the implementation details of every use case (hence, the use cases are in grey-scale)

in simpler gestures like *swipe* (left/right), *scroll* (up/down), *tap*, and more complex ones like *pinch-to-zoom* and *drag-and-drop* not considered in previous works. We do not make use of the smartphone's accelerometer and gyroscope data. Once collected the datasets, larger in both users and gestures compared to the literature, we derived features capable of effectively describing the fine-grained nature of gestures performed, such as length, curvature, finger's pressure and dimension, and velocity. Then, to identify the most useful gesture for the classification task, we performed classification measurements of single touch gestures (*single-view*) using leave-one-user-out cross-validation (LOUO-CV). We further perform experiments heading to enhance the previously made classification by combining touch gestures of different kind together, adapting the approach of *multi-view* learning [44, 46, 59, 69]. We have obtained that *scroll down* is the most useful gesture for gender classification, *random forest* is the most convenient classifier to address this problem. Furthermore, the multi-view approach is recommended when dealing with unknown devices and different combinations of gestures can be effectively adopted—building on the requirements of the authentication (or other kinds of) system our solution is built-into.

1.2 Literature's deficiency filled

We highlight that our proposal fills the following gaps in the literature:

- studying pinch-to-zoom and drag-and-drop gestures (which are among the most commonly performed gestures [3, 29, 55]);
- applying *multi-view* learning strategies to the gender recognition problem via gestures analysis on mobile devices. Such a strategy has proved to be effective in different contexts (e.g., [8, 73]);
- proposing a more robust evaluation of the methods with LOUO-CV;
- evaluate the proposal in different scenarios, i.e., with different mobile devices and never seen users (that did not participate in the data collection phase).

As we will see in Sect. 3, such aspects are mostly overlooked in the literature and not thoroughly explored.

1.3 Our contributions

The primary contributions of our work can be summarized as follows:

- Designing an approach for automatic gender classification based on machine learning and analysis of *only* gestures on touch devices; this, as highlighted in the literature, has a low impact on energy consumption of mobile devices with respect to approaches using gyroscope and accelerometer data stream;
- In-depth analysis of a large set of handcrafted features representing users' touch gestures;
- Considering, conversely to previous literature, complex gestures, i.e., pinch-to-zoom (turned out to be very useful) and drag-and-drop.
- Experimenting different learning approaches to the gender classification problem, that is single-view and multi-view learning; compared to previous works, this paper performs a more comprehensive evaluation of such techniques; to the best of our knowledge, this is one of the first works to apply multi-view learning strategies for gender recognition via gestures analysis on mobile devices.
- In-depth analysis of solution's performance in different scenarios, i.e., unknown users, unknown devices; to the best of our knowledge, this is one of the first work providing the assessment of the solution with the already seen users with entirely new devices, and never seen users.
- Discussion of the perspectives entailed by this kind of solution, potentiality and risks of its application in real-world.

1.4 Organization

The rest of the paper is structured as follows. Section 2 provides details about single-view and multi-view learning approaches and about the leave-one-user-out cross-validation. Section 3 is devoted to illustrate most recent related works and differences with the present one. Section 4 presents our solution for gender classification detailing the different approaches evaluated. Section 5 summarizes and compares results achieved, shows further experiments to better assess the generalization capabilities of the proposal, and highlights potentialities and risks of the gender classification approach here presented. Lastly, we provide some concluding remarks in Sect. 6.

2 Background

In this section, we dwell on *single-view* and *multi-view learning* approaches (Sect. 2.1) and on the cross-validation technique used here, that is LOUO-CV (Sect. 2.2).

2.1 Single- and multi-view learning approaches

2.1.1 Single-view learning

refers to the traditional approach of machine learning where a classifier is fit on a single dataset (Fig. 2a). In this case, the classifier has just one view-point (hence, a single-view) on the data. For example, a classifier that is fit on face images for skin-cancer prediction, only "knows" the patients' face images (single-view). Perhaps, the prediction performance can be enhanced by accounting/combining multiple sources of information, e.g., data from blood analysis, hence the *multi-view learning* [44, 46, 59, 69]. Therefore, in our case, we exploit both single-view and multi-view learning, where in single-view we consider only one type of gesture at a time, while in multi-view we combine in different ways gestures of different kind. As we will see more in detail in the following, combinations can be performed in different ways, among which we find: *(i)* concatenating samples of single-views, *(ii)* using classifiers (or statistical methods) for feature selection in single-views and then concatenating most important ones, and *(iii)* using single classifiers for single-views and exploiting their classification results for fitting another classifier for the final prediction.

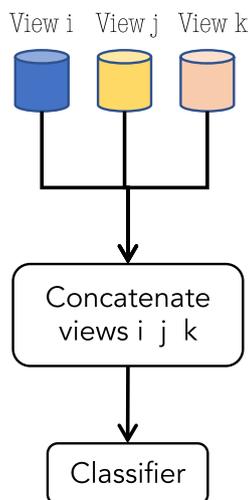Concerning the single-view approach, we consider the following classifiers:

- Random forest (RF): it is a supervised classification model which consists of an ensemble of methods based on bagging.
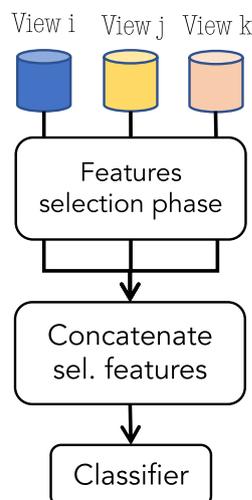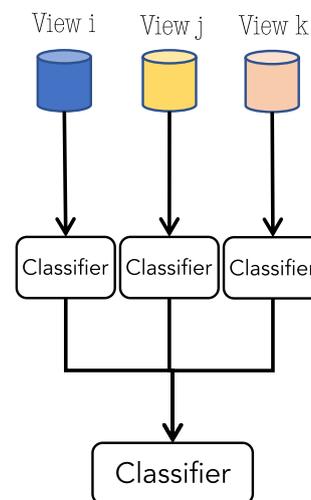
Single-view

View i

Multi-view

View i   View j   View k

View i   View j   View k

View i   View j   View k
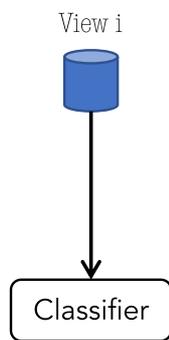
Fig. 2 a Single-view learning and (**b, c, d**) multi-view learning. **b** early integration, **c** intermediate integration, **d** late integration

– Support vector machine (SVM): it is a supervised learning model with associated learning algorithms; an SVM model is a representation of the examples as points in space, mapped so that the examples of the separate classes are divided by a clear gap.
– K-Nearest neighbors (KNN): it is a nonparametric method relying only on the most basic assumption underlying all prediction, i.e., that observations with similar characteristics will tend to have similar outcomes. Nearest neighbor methods assign a predicted value to a new observation based on the plurality or mean (sometimes weighted) of its k-nearest neighbors in the training set.
– Multilayer Perceptron (MLP): it is a feedforward artificial neural network which exploits a supervised learning technique called backpropagation for training.

Concerning multi-view learning, there are several techniques presented in the literature. Specifically, we adopt *early*, *intermediate* and *late integration* methods. The "early integration" (Fig. 2b) consists of concatenating the features associated with different gestures (single-views) performed by the same participants; in this way, each combination (concatenation of two or more single-view features-vector sample) represents one sample in the dataset; this method has the downside of considering large space features-vectors. The "intermediate integration" (Fig. 2c) consists in performing a features selection for each gesture [44, 45] (single-view), and then concatenating the features selected for each single-view; the advantages of such a technique are: *(i)* the heterogeneous nature of the gestures' features can be better exploited through the

single-views separation, *(ii)* the size of the output (and, therefore, the sample to analyze in subsequent phases) is reduced, and *(iii)* the separate extraction of significant features for different gestures implements the *divide-et-impera* principle, reducing the complexity of the tasks. Lastly, with the "late integration" (Fig. 2d) we train a classifier for each gesture (single-view) and then we use the outputs obtained by these models as input for a new model exploited for the final decision [23]. This method has the advantage to be easily implemented in parallel because each classifier is independently fit on a single-view, but, as a downside, it does not account interactions that could exist among single-views.

Multi-view learning has been exploited in a number of papers, particularly in the health domain. In more detail, just to mention some of the most recent ones, early integration has been used in radiomics [8] and for cancer prediction [73]. Intermediate and late integration has been used for predicting neurodegeneration [23]. Moreover, all the techniques have been applied for users' age-group classification [76]. Even if not new, the approach has never been applied before for gender classification based on the analysis of touch gestures as we do here.

2.2 Leave-one-user-out cross-validation

k-fold cross-validation (k-fold CV) is a resampling procedure used to evaluate machine learning models. The procedure has a single parameter called k that refers to the number of groups that a given dataset is to be split into. Cross-validation is primarily used to estimate the skill of a

machine learning model on unseen data. As clearly explained in [39] *"this approach involves randomly dividing the set of observations into k groups, or folds, of approximately equal size. The first fold is treated as a validation set, and the method is fit on the remaining k-1 folds"*. When k is equal to the size of the dataset (i.e., the number of samples), the literature refers to the so-called leave-one-out cross-validation (LOO-CV), which takes one only sample of the dataset for the validation and all samples minus one for fitting the model. Although k-fold cross-validation (as well as LOO-CV) is a well-established method, its usage is not suitable for the task of gender classification. For this purpose, an alternative cross-validation method is recommended: the leave-one-user-out cross-validation (LOUO-CV), a variant of LOO-CV. In this validation method, the classifier is trained with all but one user data, and this is repeated for all users. This method allows us to understand the generalization capability of a model, which is recognizing samples from users that were unseen during the training phase. The different validation methods are depicted in Fig. 3. Here, it is possible to spot the differences between the methods. In k-fold CV and LOO-CV, the random split could put samples generated by the user U_i in both the datasets used for fitting and the one used for validation. In the case of LOUO-CV, this does not happen, given the split is user-based. Therefore, in fitting set there are all the samples generated by all the users except a user U_i, while in the validation set there are all the samples generated by U_i. In other words, suppose we put samples of the same user in both training and testing datasets. In that case, the classifier could exhibit untruthful performance (a very high accuracy due to the uniqueness of a user in performing a specific gesture), which, however, do not reflect the actual capabilities when deployed and facing new users instead of previously seen ones. LOUO-CV has been used in several studies. Hemminki et al. [30] studied the recognition of a user's transportation based on GPS and accelerometer data. Tao et al [70] investigated

surgical gesture classification using sparse hidden-Markov models based on motion data. They compared LOO-CV and LOUO-CV evaluation methods for several datasets and reported a considerable decrease for the latter, more compliant with a real-world usage of the system. The same strategy has been applied for the same purposes by [1]. Antal et al. [2] propose the LOUO-CV for gender recognition through the analysis of touch gestures. Cornelius et al. [12] proposed a novel method for recognizing whether sensors are on the same body. Craley et al. [14] employ the LOUO-CV to evaluate a finger tracking system based on a tracker ring. More recently, Chen et al. [10] adopted the evaluation method for estimating the gameplay engagement.

As with other researches, evaluations here have been performed by using LOUO-CV.

3 Related work

In this section, we discuss prior works showing contact points with the one presented here, highlighting key differences. Given our aim is to classify users' gender by using touchscreen gestures data, *(i)* we first describe papers in the field of gender classification as a whole (Sect. 3.1), *(ii)* then papers employing touch gestures as soft biometric trait (Sect. 3.2), and *(iii)* lastly the closest project exploiting gestures for gender classification (Sect. 3.3).

3.1 Gender classification in general

Gender classification is a research area that attracted many scientists during the years. This task has been investigated in relation to several types of biometric data such as speech [78], face [60, 67], gait [35, 36], and even EEG [31]. More specifically, the authors in [60] have developed an algorithm that, given a photo, simultaneously performs face detection, landmarks localization, pose estimation and gender recognition using deep convolutional neural networks. Smith et al. [67] proposed a method based on transfer learning for both gender classification and age prediction leveraging on face images. Instead, Jain et al. [36] presented an approach for gender classification using users' gait information tracked leveraging on accelerometer and gyroscope sensors of a smartphone. They built a bootstrap aggregating classifier such sensors features for classification of the gender. The proposed approach's performance was evaluated on datasets collected using two different smartphones containing a total of 654 samples. The proposed approach achieved classification accuracy from 88.46 to 91.78% based on the activity user performed (walking, running, and so on).
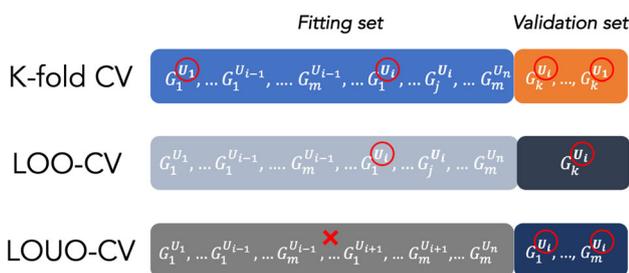
Fig. 3 Different approaches for cross-validation: k-fold cross-validation (k-fold CV), leave-one-out cross-validation (LOO-CV), and the one used in this work, i.e., leave-one-user-out cross-validation (LOUO-CV). The dotted circles highlight the user who generated the sample

3.2 Touch gestures as (soft) biometric trait

Touch gestures performed on smartphones have been used as (soft) biometric trait for several purposes (see [33] for a comprehensive overview). For example, several researchers employed gestures for age-group classification. In more details, in the work by [74], a method is proposed based on the concatenation of seven or more consecutive taps to recognize very young children (6 years old or less) from adults. Nguyen et al. [72] highlight that there is a risk for children under 12 years to be easily recognized (up to 0.99 ROC AUC) with respect to adults (more than 24 years old) by analyzing scroll, swipes, taps and other sensors altogether. Both works did not assure that the cross-validation was performed by splitting by users instead of samples' labels; therefore, the evaluation can be considered inadequate. Cheng et al. [11] proposed *iCare*, a system that can identify child users automatically and seamlessly when users operate smartphones. *iCare* records the touch behaviors and extracts hand geometry, finger information, and hand stability features (by means of accelerometer and gyroscope) that capture the age information. They conducted experiments on 100 people including 62 children and 38 adults. Results have shown that *iCare* can achieve 96.6% accuracy for child identification using only a single swipe on the screen, and the accuracy becomes 98.3% with three consecutive swipes. Lastly, Zaccagnino et al. [76] exploited touch gestures (scroll, swipe, tap, drag-and-drop, pinch-to-zoom) to lay the foundation of a safeguarding architecture for underages (age $\leq$ 16 according to the EU GDPR) on the phone (e.g., limiting harmful content displayed or preventing illegal contacts).

Other authors argued that touch gestures have the potential to identify users correctly. Specifically, Masood et al. [51] developed an algorithm based on entropy that quantifies the uniqueness of touch gestures, finding that it is possible to correctly re-identify participants in their trial. The results showed that writing samples (using the finger to write on a touchpad) could reveal 73.7% of information, and left swipes can reveal up to 68.6% of information of an individual. Instead, Rzecki et al [62] proposed a computational intelligence method which proved that long gestures (a single connected movement of a finger over the touchscreen) led to a very high person identification rate (up to 99.29%). They found that support vector machine and random forest were the most effective classifiers for this task. A summary of these works is available in Table 1.

3.3 Gender classification based on touch gestures

There exists not that much research on the use of touch gestures for gender classification. [26] were among the first ones to perform gender classification using touch gestures on smartphones. They report gender recognition accuracies of 87.32 to 91.63% using keystroke dynamics on their GREYC dataset. This evaluation can be considered inadequate since they used fivefold cross-validation; therefore, data from the same person were present both in the training and testing phases. Fairhurst et al. [21], besides user identity classification, performed gender classification on the same GREYC dataset. Again they report results based on tenfold cross-validation. Antal et al. [2] exploited keystroke dynamics and touchscreen swipes for gender recognition employing LOUO-CV and using random forest classifier. The best results were 64.76% accuracy for the keystroke dataset and 57.16% for the swipes dataset. More recently, Jain et al. [37] included in the analysis the sensors data (gyroscope and accelerometer) in addition to swipes. Concerning the gestures, they adopted GIST descriptor-based features extracted from two-dimensional maps of the touch gesture attributes, focusing on the length and curvature. Finally, a k-nearest neighbor classifier recognizes the user's gender. They evaluated their approach with fivefold cross-validation (user-based) on a set of 2268 gestures, finding accuracy of 92.96% when combining all the data sources (sensors and multiple gestures). None of these works made available the collected data (raw or preprocessed) allowing other researchers more relevant comparisons. A summary of these works is available in Table 1.

3.3.1 This work

Compared to the works mentioned above, we perform a more comprehensive evaluation of different classifiers both on single-view and multi-view approaches. We remark that the different integration techniques we adopt here have never been used for gender classification through touch gestures. We perform a more in-depth analysis of the hand-crafted features computed to represent touch gestures. Furthermore, we do not consider any sensor data (gyroscope and accelerometer are widely used in other approaches) which results into an energy gain, that for mobile devices represents a key concern [20, 25, 52, 79].

The dataset used here is wider in both samples and users (more than 9,500 samples and 147 users, respectively). Besides, our dataset contains complex touch gestures such as drag&drop and pinch-to-zoom that have never been used before for gender classification. Differently from [37],

Table 1 A summary and comparison of the existing methods for (soft) biometrics on smartphones using touch gestures

Study	Trait	Modality	Methodology	# Users in dataset	Main remarks
[21]	Gender	Keystrokes	Analysis of latency between a keystroke and successive one	133	(i) mainly based on timing; (ii) evaluated on a large dataset comparing different classifiers; (iii) validation not user-based. **Best performance**: 3% error rate
[26]	Gender	Keystrokes	Analysis of latency between a keystroke and successive one	133	(i) mainly based on timing; (ii) evaluated on a large dataset with SVM only; (iii) validation not user-based. **Best performance**: 92% accuracy
[74]	Age	Taps	Bayes rule classifier analyzes 7+ consecutive taps	119	(i) Young children can be effectively recognized from adults; (ii) the user need to provide many gestures; (iii) only one classifier evaluated. **Best performance**: 99% accuracy
[72]	Age	Swipes, Scrolls, Taps, and accelerometer and gyroscope	Random forest classifier analyzes 8+ consecutive gestures and 5 seconds sensors	50	(i) younger children (3-8 years) and also older ones (9-12 years) are correctly classified; (ii) does not include complex gestures such as pinch-to-zoom; (iii) validation not user-based, and performed on a small dataset evaluating one classifier. **Best performance**: 99% accuracy
[2]	Gender	Keystrokes, Swipes	LOUO-CV with random forest considering features like finger pressure and dimension, length of swipes, and latency between keystrokes	98	(i) showing features importance for each gesture; (ii) evaluation done without combining gestures and with the random forest only; (iii) the method is more effective with some users. **Best performance**: 61% accuracy
[62]	Identity	Touch gesture-based passwords	Analysis of long gestures with features transformation based on PCA	50	(i) the method is very effective regardless the classifier employed; (ii) using very long gestures; (iii) evaluated on a small dataset. **Best performance**: 99% accuracy
[51]	Identity	Swipes, Scrolls, Taps, Hand-writing	Using a *measure of uniqueness* of users' gestures based on entropy and cosine similarity metric	89	(i) combination of scroll, swipe and hand-writing is the most useful for recognizing users; (ii) users were free to perform any number of gestures; (iii) hand-writing is a not very common interaction. **Best performance**: 98.5% of information about users
[37]	Gender	Swipes, and Accelerometer, Gyroscope and Orientation sensors	K-Nearest Neighbors classifies multiple swipes' and sensors' data all together	126	(i) use of user-based cross-validation; (ii) validated on a large dataset; (iii) does not exhibit performance for each combination of swipes' and sensors' data considered. **Best performance**: 93% accuracy
[11]	Age	Swipes, Accelerometer, Gyroscope, and Orientation sensor	Tree-based classifier recognize children on the phone using sensors and touch interactions	100	(i) dataset collected without any constrains; (ii) experiments based on subgroups split by gender and race; (iii) validation method not clear. **Best performance**: 96% accuracy with consecutive swipes
[76]	Age	Swipes, Scrolls, Taps, Pinch-to-zooms, Drag-and-drops, Writing	Random forest analyzes a combination of scroll, swipe and pinch-to-zoom for age-group classification	147	(i) evaluated on different smartphones model; (ii) evaluated on a large dataset; (iii) not very effective with teenagers. **Best performance**: 0.92 ROC AUC
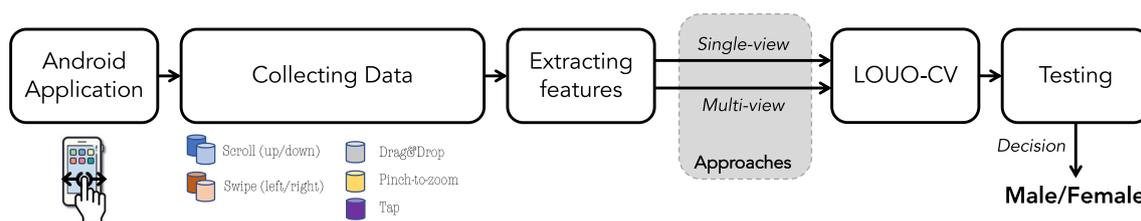
Fig. 4 Block diagram of the proposed approach for gender classification based on the analysis of touch gestures performed on mobile devices

where results are presented based on the number of gestures combined and not disclosing the exact combinations, we exhibit the results of every single-view and multi-view learning approach adopted. In addition, their work is based on the integrated analysis of gestures and sensors (e.g., accelerometer and gyroscope) which are not available on every mobile device on the market (some low-tier devices are not equipped with gyroscope). Lastly, none (except [76] for smartphones) of the works mentioned above have performed an evaluation of the proposed method with different mobile devices as we have done for smartphones and tablets.

4 Our method for gender classification

Figure 4 shows the block diagram of our proposal for users' gender classification through the analysis of gestures performed on touchscreen devices. We have developed an Android application to collect biometric data of users (Sect. 4.1). Such data are split into different datasets, one for each touch gestures considered (Sect. 4.2), that is scroll down (ScD), scroll up (ScU), swipe left (SwL), swipe right (SwR), tap (T), drag&drop (DD), pinch-to-zoom (P2Z). We then extract features from these gestures, such as x-y coordinates, pressure and dimension of the finger, velocity, and so on (Sect. 4.3). Next, we adopt single-view and multi-view learning approaches for twofold objectives: *(a)* with single-views we consider only one kind of gesture dataset at a time aiming at understanding the most useful touch gesture among the considered ones (Sect. 4.4.1); *(b)* with multi-views we consider different ways of combining different gestures aiming at understanding both whether combination of gestures improves the classification performance compared to the single-view approach and if so the best combination of gestures (Sect. 4.4.2). For both approaches, we envision a LOUO-CV (made with 80% of the datasets) and testing phase (made with the remaining 20% of datasets).

All the experiments have been run on a machine equipped with 2.8 GHz Intel i7 quad-core (Turbo Boost up to 3.8GHz) with 6MB shared cache L3 (model 7700HQ "Kaby Lake"), and 16GB 2133 MHz LPDDR3 RAM.

4.1 Android application

In order to collect data, we implemented an Android application that allows to capture and analyze user interactions with the smartphone. The app includes several games; each of them is thought to force the user in performing a specific touch gesture. We are interested in the following gestures: scroll (up/down), swipe (left/right), tap, drag&drop, pinch-to-zoom. Thus, we define the set of gestures $G = \{ScD, ScU, SwL, SwR, DD, P2Z\}$. We have employed the Android APIs `onScroll`, `onFling`, `onTouchEvent`, `onDrag`, `onTouch`. Thus we have developed five games:

- Game 1 (Fig. 5a) collecting data about ScD, and ScU;
- Game 2 (Fig. 5b) collecting data about SwL, and SwR;
- Game 3 (Fig. 5c) collecting data about T;
- Game 4 (Fig. 5d) collecting data about DD;
- Game 5 (Fig. 5e) collecting data about P2Z.

We remark that, in general terms, we needed a method to collect users data; it should have been the most attractive possible in order to include a large number of participants. Therefore, we have opted for a game-based app to make it more pleasant and joyful for users. In this way, we have gathered more users in our study. Our interest is not toward gaming ability, the app does not put users in any competition. The games do not show any score to the user, hence they do not aim to trigger any gaming performance. Lastly, observe that every user has just played once.

4.2 Collecting data

In this phase, we have gathered data from 147 participants. Before starting the collection phase, we explained to participants what they were expected to do in the study. For children, their parents completed written parental permission forms. All participants had to sign consent forms. We explained that we did not collect personal information during the experiments except for username, device ID, age, and gender. Raw data associated with gestures performed by users are tracked and saved for the subsequent analysis. We also explained that these data were kept confidential and used only for the period of the experimentation.

(a) Scroll   **(b)** Swipe   **(c)** Tap   **(d)** Drag&drop

(e) Pinch-to-zoom
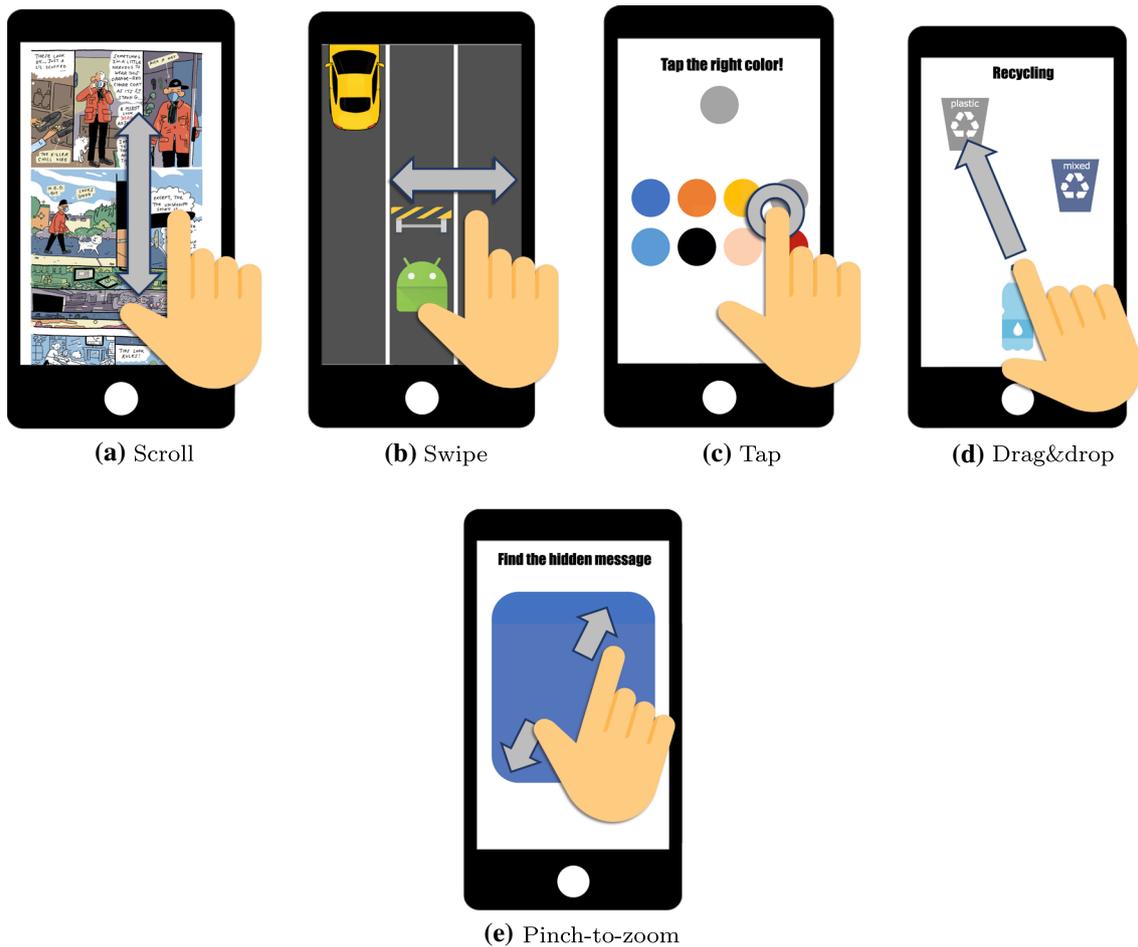
Fig. 5 Games developed to collect the biometric datasets. The games require users to perform specific touch gestures

Table 2 Description of the raw data captured through the Android app for the different touch gestures and corresponding dataset size. m = male, f = female

Game	Gesture (m/f)	Information
1	ScD (1569/ 1025)	*start time, start point, end point, duration, touch dimension, touch pressure, velocity along x and y axis*
	ScU (407/ 266)	
2	SwL (605/ 400)	*start time, start point, end point, duration, touch dimension, touch pressure, velocity along axis*
	SwR (588/ 384)	
3	T (1780/ 1162)	*start time, start point, end point, touch dimension, touch pressure*
4	DD (445/260)	*start time, start point, end point, duration, touch dimension, touch pressure, velocity along x and y axis*
5	P2Z (659/ 431)	*start time, finger1 start point, finger1 start dimension, finger1 start pressure, finger2 start point, finger2 start dimension, finger2 start pressure, length, fingers motion points, fingers motion velocity along x and y axis*
Total	9981 (6053/3928)	

The models of smartphones used for the experiments were a HTC Desire 820, LG Nexus 5X, and ASUS Zen-Fone 2.

We collected 9,981 touch gestures from 89 male and 52 female participants. The age varies in the range from 7 to 59 years. 34% of participants were under 16 years old. 49% of them were in the range 17-26. The data captured for each gesture (relying on the Android Touch API) and the sizes of the different datasets are available in Table 2. See https://bit.ly/3pyNpno for an excerpt of the data collected and https://bit.ly/3IZ6Lvz for all the data.

4.3 Extracting features

In this section, we highlight the features extracted for each gestures dataset. On the one hand, these features are taken from the Android Touch API; examples include: the number of fragments the gesture is composed of (frag number), the duration of the gesture (duration), the coordinates of the initial point of the gesture (x_s, y_s), and so on. On the other hand, additional features are engineered considering "geometric" properties of the gesture. For instance, based on x-y coordinates' analysis, we are interested in the gesture's length. Let (x_s, y_s) the coordinates of the start point, (x_e, y_e) the coordinates of the end point, then $length = \sqrt{(x_e - x_s)^2 + (y_e - y_s)^2}$. Accounting the time ($t_s$ start time, and t_e end time) together with the coordinates, we can compute the velocity: $vel = \frac{length}{t_e - t_s}$. Other features we are interested in are duration, pressure and dimension of the finger. For ScD/ScU we also compute features regarding the *turning points*. Given a ScD/ScU gesture, the turning point (x_{tp}, y_{tp}) is the point where the gesture

changes direction respect the *x*-axis. We consider the acceleration concerning the turning point (see Fig. 6). It is captured by two features, i.e., the acceleration of the ScD/ScU gesture from (x_s, y_s) to (y_{tp}, y_{tp}) and the acceleration of the ScD/ScU gesture from (x_{tp}, y_{tp}) to (x_e, y_e).

Likewise, we included information about the finger dimension and pressure in correspondence of (x_{tp}, y_{tp}) (indicated as *mid dimension* and *mid pressure* in Table 3).

As anticipated, specific Touch APIs allow developers to catch the fragments composing the gesture. For such gestures, we considered pressure, dimension and velocity in the following way: *(i)* the value concerning the whole gesture running, *(ii)* the maximum, minimum, mean value, and *(iii)* the values in correspondence of the quartiles of the gesture (the value at the start, 25%, 50%, 75% and at the end of the gesture running).

Lastly, concerning the P2Z we consider both fingers (finger1, finger2) coordinates, pressure, dimension.
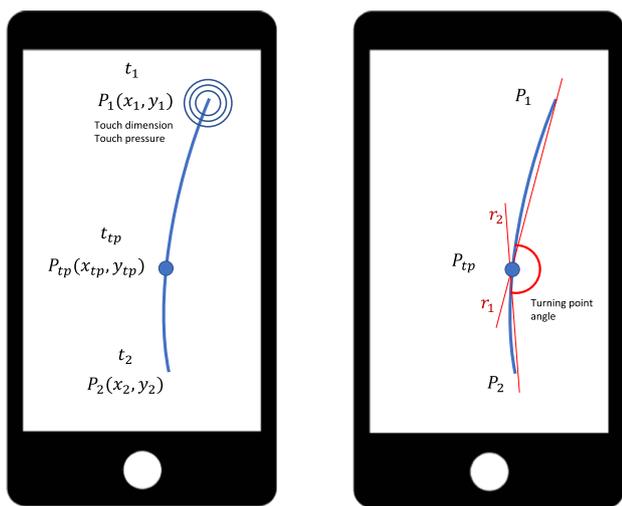
All the features extracted for each gesture are summarized in Table 3. Every feature has real values ($\in \mathbb{R}$), with the exception of *frag number* in ScD/ScU, T, and P2Z which has natural values ($\in \mathbb{N}$).

4.4 Evaluation

We evaluated the performance of random forest (RF), support vector machine (SVM), multilayer perceptron (MLP), and K-nearest neighbors (KNN) for the gender classification task. Features' values of samples in the various datasets have been scaled in the range [0, 1] prior to be analyzed in the subsequent phases. For the validation phase (LOUO-CV) we reserved roughly 80% of the available users (we used the samples of 117 users), while the remaining 20% was left for the testing phase (we used the samples of 30 users). The split was stratified. For every approach (single-view and multi-view) evaluated, during the LOUO-CV we found the best setting of main parameters for each classifier, with respect to the optimization of F1-score. We searched the parameters with a grid strategy, i.e., looking for the best ones among specific boundaries. For example, we bounded the search of *n_estimators* for RF from 20 to 500 (ten by ten step). The list of tuned parameters is shown in Table 4.

For the evaluation phase, we employed the scikit-learn library for Python.

In the following, we report the results obtained on the whole dataset collected (involving all the users). However, we remark that in intermediate phases during the data collection, we experimented our approaches on different ages range: 0–10, 11–20, 21–30, 31–40, 41–50, 51–60. The aim was having preliminary feedback from our solutions. We found that classifiers generally exhibited good

(a) Turning point coordinates.

(b) Turning point angle.

Fig. 6 Scroll (ScD, ScU) turning point (*tp*)

Table 3 Sketch of the features extracted for each touch gesture. Vel = velocity, dim = dimension. For further details please refer to https://bit.ly/3pyNpno

Gesture	Features
ScD/ScU	**50 features.** *frag number, duration, start point (x_s, y_s), end point (x_e, y_e), length, total vel, area, turning point (x_{tp}, y_{tp}), turning point angle, acceleration $_{s,tp}$, acceleration $_{tp,e}$, start dim, mid dim, end dim, min dim, max dim, mean dim, 25% dim, 50% dim, 75% dim, start pressure, mid pressure, end pressure, min pressure, max pressure, mean pressure, 25% pressure, 50% pressure, 75% pressure, start vel x, mid vel x, end vel x, min vel x, max vel x, mean vel x, 25% vel x, 50% vel x, 75% vel x, start vel y, mid vel y, end vel y, min vel y, max vel y, mean vel y, 25% vel y, 50% vel y, 75% vel y*
SwL/ SwR	**13 features.** *duration, start point (x_s, y_s), end point (x_e, y_e), dim, pressure, length, vel, acceleration, area, X vel, Y vel*
T	**29 features.** *frag number, start point (x_s, y_s), end point (x_e, y_e), x shift, y shift, x_{min}, x_{max}, y_{min}, y_{max}, start dim, mid dim, end dim, min dim, max dim, mean dim, 25% dim, 50% dim, 75% dim, start pressure, mid pressure, end pressure, min pressure, max pressure, mean pressure, 25% pressure, 50% pressure, 75% pressure*
DD	**13 features.** *duration, start point (x_s, y_s), end point (x_e, y_e), dim, pressure, length, vel, acceleration, area, x vel, y vel*
P2Z	**48 features.** *finger1 start point (x_1, y_1), finger1 dim, finger1 pressure, finger1 start point (x_2, y_2), finger2 dim, finger2 pressure, frag number, length, area, pinch grade, start dim, mid dim, end dim, min dim, max dim, mean dim, 25% dim, 50% dim, 75% dim, start pressure, mid pressure, end pressure, min pressure, max pressure, mean pressure, 25% pressure, 50% pressure, 75% pressure, start vel x, mid vel x, end vel x, min vel x, max vel x, mean vel x, 25% vel x, 50% vel x, 75% vel x, start vel y, mid vel y, end vel y, min vel y, max vel y, mean vel y, 25% vel y, 50% vel y, 75% vel y*

Table 4 Parameters tuned for each classifier evaluated (both in single-view and multi-view). *in* = input layer size. MLP *solver* was set to *lbfgs*

Classifier	Parameters	Range	Step
RF	*n_estimators*	$[20, 500]$	10
SVM	*C*	$[10^{-3}, 10^3]$	10
	kernel	$\{linear,\ rbf,\ sigmoid,\ poly\}$	–
MLP	*hidden_layer_size*	$[\frac{1}{10}, \frac{9}{10}] \times in$	$\frac{1}{20} \times in$
	activation	$\{identity,\ logistic,\ tanh,\ relu\}$	–
KNN	*n_neighbors*	$[3, 10]$	1
	weights	$\{uniform,\ distance\}$	–

performance with every gesture when testing with samples generated by users of similar age. In particular, the best results were in the ranges 21-30, and 31-40 with F1-score up to 0.94. The worst results were obtained in the range 0-10 and 51-60. We highlight our goal was a solution coping with the wider age range possible. Therefore, we argued it was more correct to report only the results obtained on the whole dataset.

4.4.1 Single-view approach

The objective of this approach is to understand which gesture is the most useful for gender classification. In Table 5, we report the results obtained when considering one gesture at a time. For every $g \in G$ we show the F1-score both in validation and testing phases; the *italic* font is used for emphasizing best results on a gesture-basis, and the **bold** font is adopted for highlighting the best result

Table 5 Performance comparison of different classifiers over the considered touch gestures in the single-view approach. For each classifier, the table shows F1-score for both validation and testing phases (validation/testing). *Time (ms)* shows the time elapsed on average for a single `fit()` in milliseconds

Gesture	RF	SVM	MLP	KNN
ScD	**0.89/0.85**	0.88/0.84	0.85/0.81	0.84/0.80
ScU	0.85/0.82	*0.87/0.83*	*0.87/0.83*	0.85/0.80
SwL	*0.77/0.72*	0.75/0.71	0.70/0.63	0.68/0.63
SwR	*0.80/0.82*	0.76/0.79	0.72/0.68	0.70/0.67
T	*0.72/0.65*	0.70/0.64	0.67/0.63	0.66/0.63
DD	*0.73/0.67*	0.72/0.66	0.69/0.65	0.69/0.65
P2Z	0.80/0.77	*0.81/77*	0.77/0.70	0.75/0.70
Time (ms)	60	106	90	37

obtained, that is the most useful gesture when analyzed with the most promising classifier.

We observe that RF and SVM are the most effective classifiers in the gender recognition task. Concerning the different gesture datasets, the results show ScD is the most useful gesture for classifying users' gender (from 0.82 to 0.89 F1-score in validation), followed by ScU, SwR, P2Z and SwL, respectively. RF when analyzing ScD exhibits F1-score of 0.89 in validation and 0.85 in testing. Lastly, to understand what is the best classifier between SVM and RF we perform a statistical test. We first assess the normality distribution of data with Shapiro–Wilk test [64] with a significance level of $\alpha = 0.05$, obtaining $p - value = 0.43$. Since $p - value > \alpha$, we accept the null hypothesis, that is we assume the data is normally distributed. Therefore, we can exploit the t-student test [41, 62]. We assume this difference between RF and SVM is zero (this is the null hypothesis) with the significance level .05 and check if we can reject this hypothesis. There are seven results of classifications, so the degree of freedom is six. RF's $M = 0.75$, $SD = 0.03$, and it is the same for SVM. We can calculate that t-Student test result is $t = 0$, and obtained $p = .5$. The hypothesis cannot be rejected, so from statistical point of view, none of these classifiers has significantly better accuracy than the other. The only time of training (RF is about 1.7 times faster than SVM on average in training) gives the advantage of RF method (grey background in Table 5) for the subsequent steps.

4.4.2 Multi-view approach

The objective of this approach is to understand which combination of gestures is the most useful for gender classification. Combinations we consider are without repetitions. Here, we experiment early, intermediate and late integration techniques (see Sect. 2.1):

– *Early*: we simply concatenate gestures. Let $g_1, g_2 \in G$ be two touch gestures, with $g_1 \neq g_2$ we concatenate the features of g_1 and g_2. For example, if g_1=ScU and g_2=SwL, the technique generates samples of 50+13 features. These are input for the final classifier.

– *Intermediate*: we use a feature ranking technique to only keep the most discriminating features for each $g \in G$, then we concatenate such features' sub-spaces for the final classification. RF has been chosen for the final decision due to the very good results obtained with the single-view approach and to its reduced training time.

– *Late*: we fit the best classifier (obtained in the single-view approach) for each gesture $g \in G$, then we concatenate these classifiers' results in a RF for the final decision. We select RF due to its reduced training time.

Concatenations we refer to are user-based, i.e., we concatenate gestures performed by the same user. We have performed this evaluation on pairs and triples of gestures. Pairs (*PA*) and triples (*TR*) analyzed are the following ones:

$$PA = \left(\binom{\{SwL, SwR, ScD, ScU, T, DD, P2Z\}}{2} \right)$$

$$TR = \left(\binom{\{SwL, SwR, ScD, ScU, T, DD, P2Z\}}{3} \right)$$

Results of these multi-view learning strategies are reported in Table 6. We highlight the best results using **bold** font, and the most convenient strategy with the grey background.

Early integration. The best results are obtained when combining ScD+SwL (50+13 features) and ScD+P2Z (50+48 features) with F1-score of 0.86 in validation, and from 0.80 to 0.84 in testing. The triple ScD+SwL+P2Z is on par with these results but considers more features (50+13+48).

Intermediate integration. Since features used are in the range [0, 1] and classes $\in \{male, female\}$, we adopt features selection method for each $g \in G$ by analyzing variations, i.e., computing the ANOVA F-measure [24] with `f_classif` method[1]. Figures 7 and 13 show F-measure computed for all features of each gesture. The most significant ones ($score > 0.2$) are colored in blue. We observe that for ScD and ScU the most significant features are those related to the *velocity* and *length*. Concerning the SwL and SwR, the *duration* and the *length* are among the most important characteristics. For T the most important features are those accounting *pressure*, while for DD *duration* and *velocity* are among the most useful. Lastly, for P2Z *length* and *area* are the most distinctive features. The best results have been obtained when combining ScD+ScU (17+23 features), ScD+SwL+P2Z (17+7+19 features), ScD+SwR+P2Z (17+10+19 features), ScD+ScU+SwL (17+23+7 features), ScD+ScU+P2Z (17+23+19 features) with F1-score of 0.86 in validation and from 0.84 to 0.85 in testing.

4.4.3 Late integration

For this strategy, we employ the `StackingClassifier` method. The best results have been achieved when

---

[1] An F-statistic, or F-test, is a class of statistical tests that calculate the ratio between variances values, such as the variance from two different samples or the explained and unexplained variance by a statistical test, like ANOVA. The ANOVA method is a type of F-statistic referred as an ANOVA f-test. Importantly, ANOVA is used when one variable is numeric and one is categorical, such as numerical input variables (features of the gestures) and a classification target variable in a classification task (male or female?). The results of this test can be used for feature selection [5, 32, 65] where those features that are independent of the target variable can be removed from the dataset.

Table 6 Performance comparison of different multi-view learning strategies over the considered combination of touch gestures. For each combination, the table shows F1-score for both validation and testing phases (validation/testing). *Time (ms)* shows the time elapsed on average for a single `fit()` in milliseconds

Combination	Early	Intermediate	Late
SwL+SwR	0.83/0.81	0.82/0.82	0.83/0.82
SwL+ScU	0.83/0.80	0.82/0.80	0.82/0.81
SwL+P2Z	0.84/0.82	0.82/0.82	0.83/0.82
SwL+T	0.81/0.79	0.81/0.82	0.80/0.80
SwL+DD	0.82/0.80	0.82/0.82	0.80/0.80
SwR+ScU	0.83/0.80	0.82/0.80	0.82/0.81
SwR+P2Z	0.83/0.80	0.83/0.81	0.83/0.82
SwR+T	0.80/0.80	0.82/0.82	0.81/0.80
SwR+DD	0.81/0.80	0.81/0.81	0.81/0.82
ScU+P2Z	0.83/0.81	0.82/0.80	0.82/0.82
ScU+T	0.81/0.79	0.81/0.82	0.80/0.80
ScU+DD	0.80/0.80	0.81/0.82	0.81/0.80
ScD+SwL	**0.86/0.80**	0.82/0.81	0.84/0.82
ScD+SwR	0.84/0.80	0.83/0.81	0.85/0.81
ScD+P2Z	**0.86/0.84**	0.84/0.84	0.82/0.82
ScD+ScU	0.85/0.83	**0.86/0.85**	0.85/0.84
ScD+T	0.83/0.81	0.83/0.82	0.82/0.79
ScD+DD	0.83/0.81	0.83/0.82	0.82/0.80
T+DD	0.78/0.79	0.80/0.80	0.78/0.77
P2Z+DD	0.81/0.81	0.82/0.80	0.80/0.80
P2Z+T	0.81/0.81	0.82/0.81	0.80/0.81
ScD+SwL+P2Z	**0.86/0.85**	**0.86/0.84**	0.83/0.83
ScD+SwL+DD	0.82/0.82	0.83/0.83	0.81/0.81
ScD+SwL+T	0.83/0.82	0.83/0.83	0.81/0.81
ScD+SwR+P2Z	0.84/0.84	**0.86/0.85**	0.83/0.82
ScD+P2Z+T	0.83/0.82	0.84/0.83	0.82/0.81
ScD+P2Z+DD	0.83/0.82	0.84/0.83	0.82/0.81
ScD+SwR+T	0.83/0.84	0.85/0.83	0.82/0.82
ScD+SwR+DD	0.83/0.83	0.83/0.83	0.81/0.81
ScD+ScU+SwL	0.85/0.82	**0.86/0.84**	0.84/0.83
ScD+ScU+SwR	0.85/0.83	0.85/0.85	0.84/0.83
ScD+ScU+P2Z	0.85/0.83	**0.86/0.84**	0.85/0.85
ScD+ScU+T	0.85/0.83	0.85/0.85	0.84/0.83
ScD+ScU+DD	0.84/0.82	0.84/0.83	0.83/0.82
ScD+T+DD	0.82/0.81	0.83/0.81	0.81/0.80
ScD+SwR+SwL	0.84/0.82	0.85/0.83	0.84/0.83
ScU+SwR+SwL	0.83/0.82	0.83/0.82	0.82/0.82
ScU+SwR+P2Z	0.84/0.82	0.83/0.83	0.84/0.82
ScU+SwL+P2Z	0.84/0.82	0.83/0.83	0.84/0.82
ScU+SwL+T	0.82/0.81	0.83/0.80	0.83/0.81
ScU+SwL+DD	0.82/0.80	0.83/0.80	0.81/0.80
ScU+SwR+T	0.84/0.81	0.84/0.81	0.81/0.80
ScU+SwR+DD	0.83/0.80	0.84/0.81	0.81/0.80
ScU+P2Z+T	0.84/0.82	0.84/0.83	0.82/0.81
ScU+P2Z+DD	0.84/0.82	0.85/0.83	0.83/0.81

Table 6 (continued)

Combination	Early	Intermediate	Late
ScU+T+DD	0.82/0.80	0.83/0.80	0.81/0.80
SwL+SwR+P2Z	0.82/0.81	0.83/0.80	0.83/0.81
SwL+SwR+T	0.82/0.80	0.82/0.81	0.81/0.80
SwL+SwR+DD	0.82/0.80	0.82/0.81	0.81/0.80
SwL+T+DD	0.82/0.80	0.82/0.81	0.81/0.80
SwR+T+DD	0.83/0.81	0.83/0.81	0.81/0.80
P2Z+T+DD	0.84/0.82	0.84/0.83	0.82/0.81
SwL+P2Z+T	0.83/0.82	0.84/0.83	0.82/0.81
SwL+P2Z+DD	0.84/0.81	0.84/0.82	0.83/0.81
SwR+P2Z+T	0.84/0.81	0.84/0.84	0.81/0.82
SwR+P2Z+DD	0.83/0.82	0.84/0.83	0.82/0.81
Time (ms)	3228	1860	2147

combining ScD+ScU and ScD+ScU+P2Z with F1-score of 0.85 in validation and from 0.84 to 0.85 in testing.

4.4.4 Overall

As one can imagine, the combinations including ScD were the most effective, with F1-score almost always higher than other combinations whatever the integration strategy adopted. However, we aim to understand what is the most effective strategy of multi-view learning. Since the Shapiro–Wilk test ($\alpha = .05$) was violated ($p - value = 0.0005 < \alpha$), we perform the nonparametric Kruskal–Wallis H test [53] to check if there is a significant difference among the strategies. We obtained $H = 11.918$ and $p - value = .00258$. The result is significant at $p < .05$; therefore, we select the intermediate integration strategy. This strategy has also a reduced training time (on average 1.74 times faster than the early integration and 1.15 times faster than the late integration).

5 Discussion

This section presents an in-depth analysis of the results obtained in the evaluation (Sect. 5.1) and potentiality and risks arising for users when this kind of solution is adopted in real-world applications/frameworks (Sect. 5.2).

5.1 Results

Best approach. From Table 5, we observed that, with the single-view approach, ScD is the most useful gesture for gender classification with F1-score of 0.89 in validation

Fig. 7 Scroll down (ScD)—Features selection performed computing ANOVA F-measure. In blue color the selected features

Feature selection for ScD
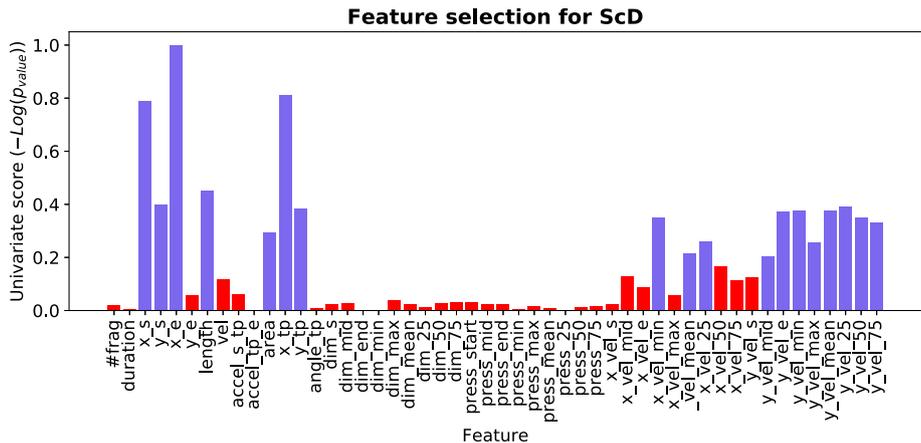
Fig. 8 Scroll up (ScU)—Features selection performed computing ANOVA F-measure. In blue color the selected features

Feature selection for ScU
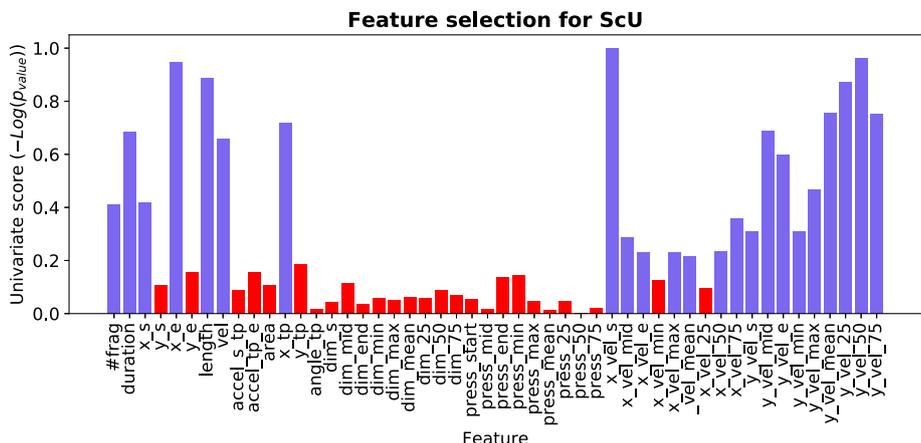
Fig. 9 Swipe left (SwL)—Features selection performed computing ANOVA F-measure. In blue color the selected features
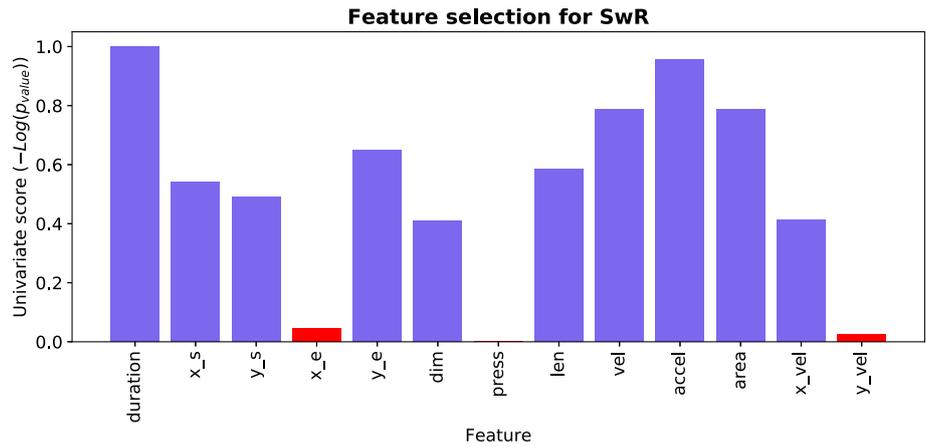

Feature selection for SwL
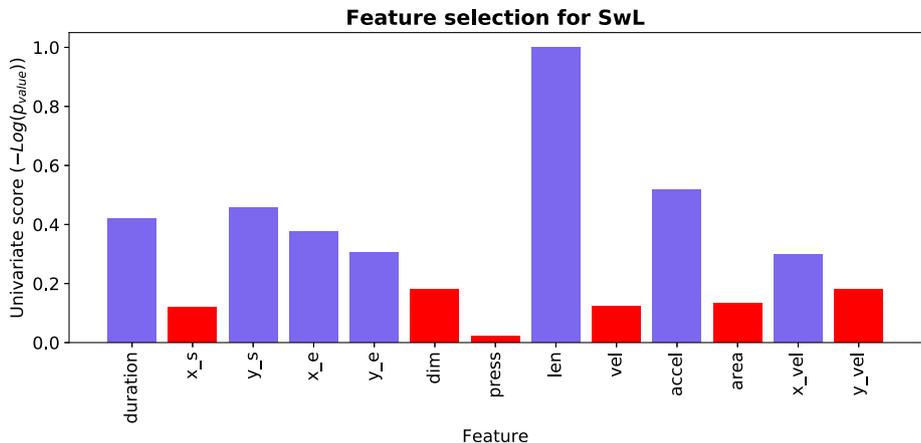
(LOUO-CV). When combining different gestures with the multi-view approach, overall, we do not find performance improvement against such score (see Fig. 14). This is due to the large size of the ScD dataset (the biggest among the gestures considered) which allows RF a broader learning.

Yet, we observe that when considering SwL+SwR, the multi-view approach exhibits an improvement in performance—up to 0.83 F1-score—with respect to the single-

view approach on SwL and SwR—with 0.77 and 0.80 F1-score, respectively—(see Fig. 15). This holds true for combinations including P2Z and SwL or SwR, e.g., SwL+P2Z.

In addition, we emphasize we have tried with combinations including four/five gestures without any performance improvement against the triples or pairs.
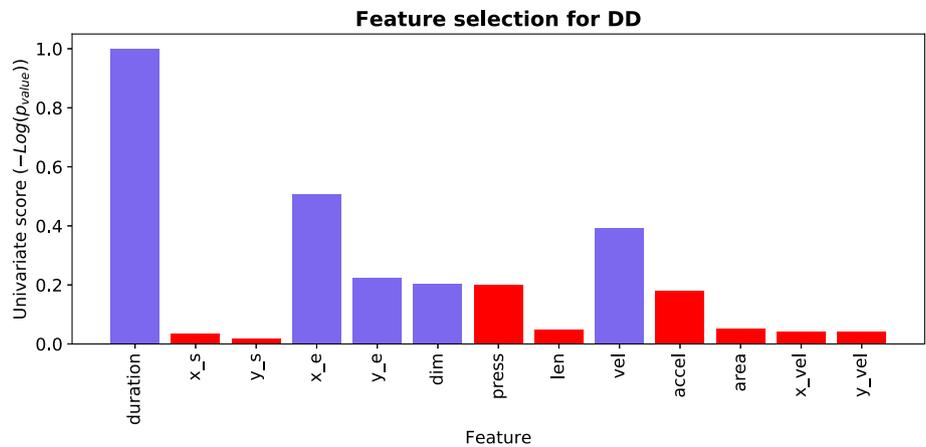
Fig. 10 Swipe right (SwR)—Features selection performed computing ANOVA F-measure. In blue color the selected features

Fig. 11 Tap (T)—Features selection performed computing ANOVA F-measure. In blue color the selected features
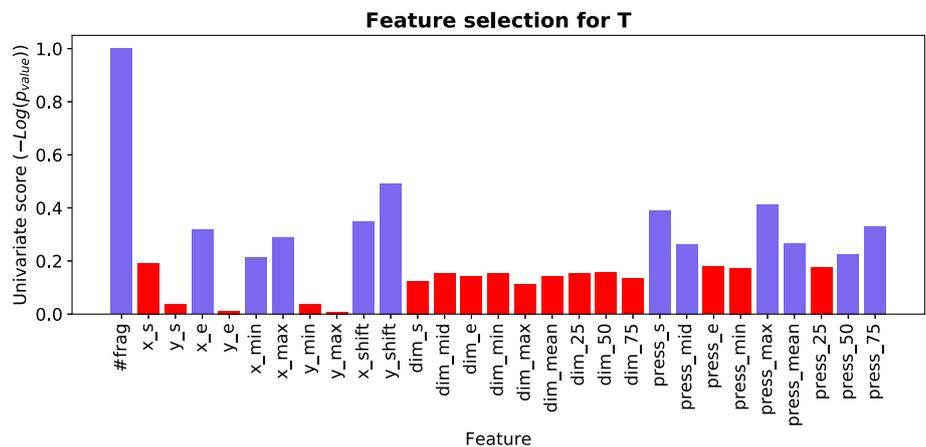
Fig. 12 Drag&Drop (DD)—Features selection performed computing ANOVA F-measure. In blue color the selected features


We conclude that based on the environment / framework where gender classification is needed to improve authentication (as well as enhancing other interactions), the framework can make use of the solutions here proposed trying to smooth as much as possible the user experience. If the framework already prompts users with swipe activities, the authentication can be improved by multi-view swipes without developing an *ad-hoc* interface for catching scrolls. This holds true for all the most useful touch gestures (and combinations of such) considered.

Generalization capability of our solution. Previous works in the literature pointed out that the pressure is not obtainable with every smartphone model available on the market. Some of them always return 0 as pressure value.

Fig. 13 Pinch-to-zoom (P2Z)—Features selection performed computing ANOVA F-measure. In blue color the selected features
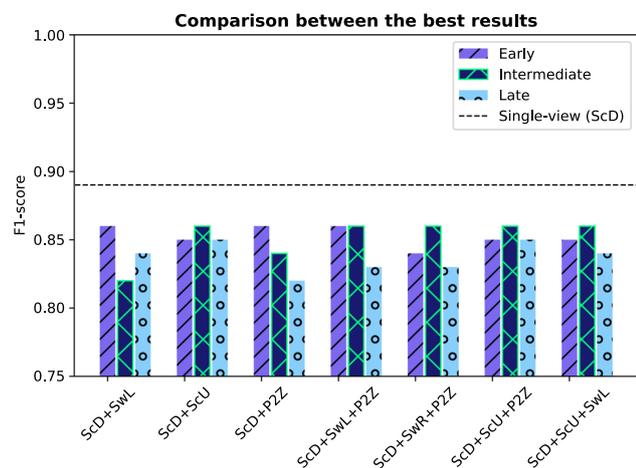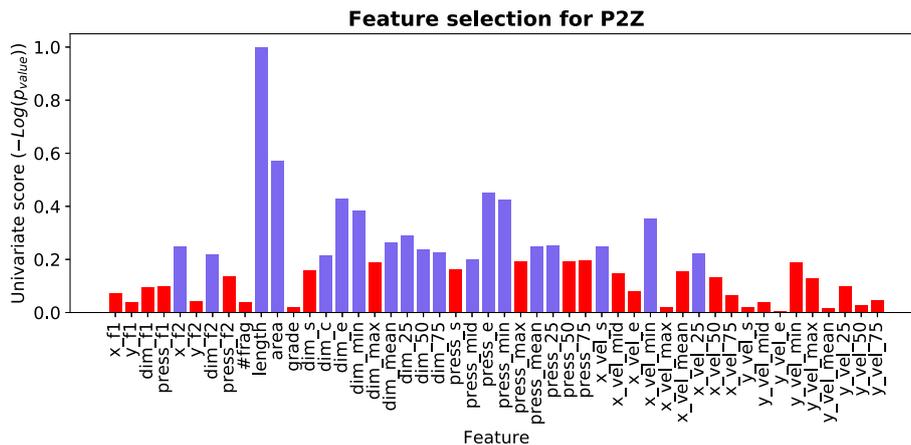

Fig. 14 Comparison between best results obtained in single-view learning (ScD), represented as a dashed threshold, against those achieved with the different strategies of multi-view learning (bars)

For this reason, we have performed an ablation test by dropping every pressure-related feature. In Fig. 16, we show the results obtained with the single-view approach (RF in particular) when the pressure is eliminated against considering all features. As partially confirmed by the feature importance evaluation in Sect. 4.4.2, dropping the pressure-related features does not have a big impact on the classifier' performance[2]. Now, to study the difference between the performance with and without pressure-related features, we perform a statistical test. Once checked that the data were normally distributed with the Shapiro–Wilk test, we evaluated the significance of the difference leveraging on the t-Student test, obtaining that with significance $< .05$ there is $t = 0.80$, while the p-value$= .22$. Therefore, the result is not significant at $p < .05$.

---

[2] In Sect. A, we provide the results of an experiment involving only pressure-related features

Likewise, with the multi-view approach the results without the pressure-related features do not show a statistically significant difference, ranging from 0.82 to 0.83 F1-score in LOUO-CV. We further inspected the generalization capability of our solution. We evaluated our best approaches on other smartphones, i.e., Samsung Galaxy S7 Edge and Samsung Galaxy S8. They have been used by 5 new participants (3 females, 2 males, age between 12 and 64 years), and 3 returning participants (2 males, 1 females). The returning participants (ASp) used only the Samsung Galaxy S8 which does not return the pressure value; new participants (NSp) used both devices. For this evaluation we only asked participants to play games for obtaining ScD, ScU, SwL, SwR, P2Z (see Fig. 5 in Sect. 4.1), the most useful gestures. A summary of the data collected is available in Table 7. The objective of this evaluation is to answer the following questions: *(a) "Does our solution for gender classification correctly classify never seen users on new devices?", (b) "Does our solution for gender classification correctly classify previously seen users on a different device?".* We evaluate the best solutions found in Sects. 4.4.1–4.4.2, that is RF for single-view and intermediate integration for multi-view. Results of this evaluation are in Table 8.

We notice that there are differences in performance between the results achieved in Sect. 4.4 with single-view and multi-view approaches (testing) and those got in the current experiment (see Fig. 17). Such differences are higher concerning new users on new devices, with whom our solutions exhibit lower F1-score. As we could expect, instead, when analyzing already seen participants we obtain high F1-score. If ASp have used the same devices, we would have obtained results comparable to those by [2] with the tenfold cross-validation non-user-based, that is more than 0.90 F1-score. In our case, such participants used new devices, hence the not perfect scores (but still very high). However, the differences with never seen participants get razor-thinner with the multi-view learning
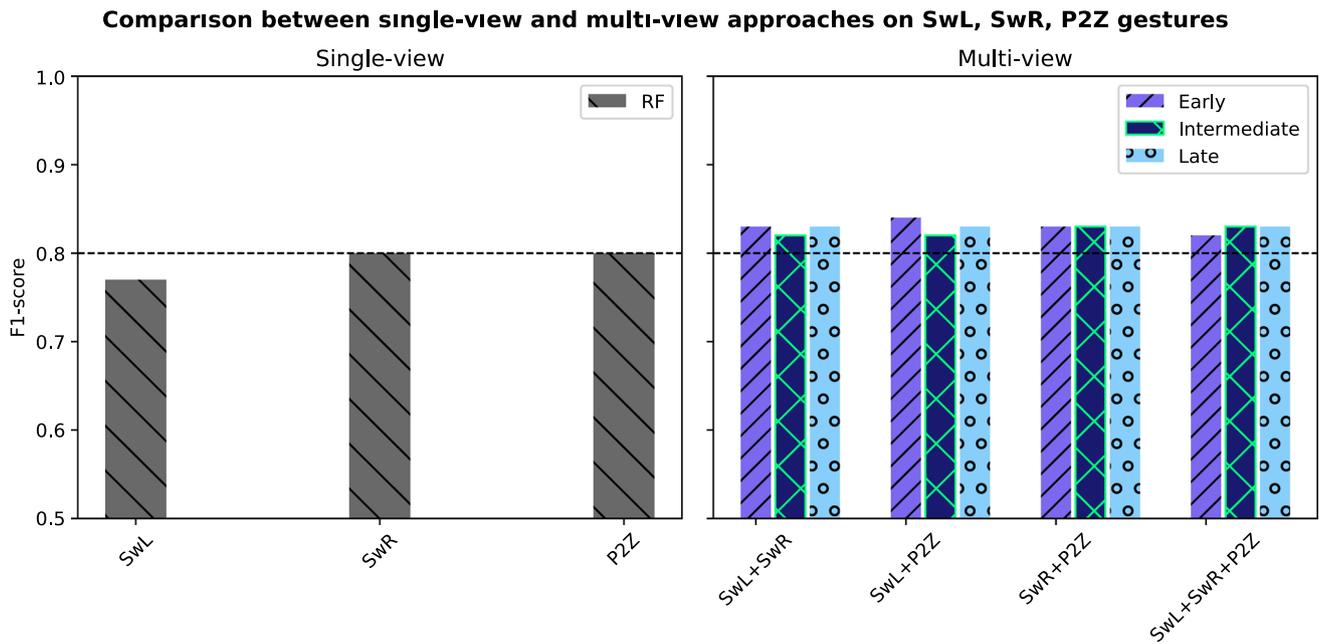
Fig. 15 Comparison between single-view approach on SwL, SwR, and P2Z against multi-view approach combining in pairs and triples such gestures. The dashed line represents the maximum score achieved in single-view (i.e., a threshold to better grasp multi-view's results)
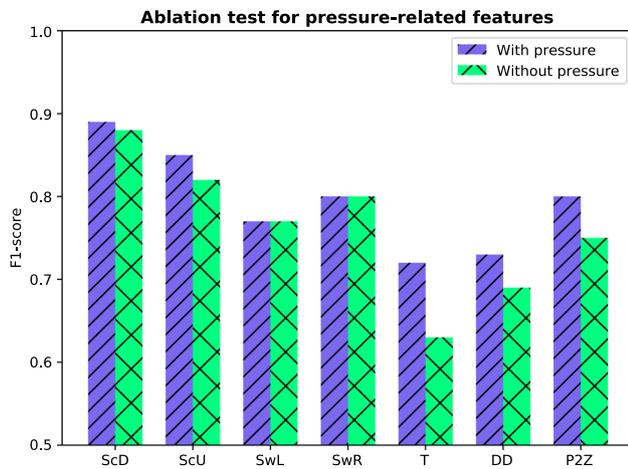
Fig. 16 The features ablation test: results in LOUO-CV with RF over the gesture datasets with and without the pressure-related features

Table 7 Collected gestures for the evaluation of generalization capability on different smartphone models. NSp = never seen participants, ASp = already seen participants. m = male, f = female

Gesture	#Samples	
	NSp (m/f)	ASp (m/f)
ScD	63/81	34/16
ScU	20/28	10/5
SwL	28/42	14/7
SwR	28/42	14/7
P2Z	32/42	16/8
Total	(406) 171/235	(131) 88/43

approach. Vice versa, we obtain much higher scores with already seen users. This means that the smartphones' hardware contributes to the performance and values returned by Touch APIs (apart the aforementioned pressure), but such contribution gets negligible as more gestures we consider for the classification.

Lastly, in order to fully answer to "*question (b)*" we have performed an experiment involving a different kind of mobile device: a tablet, model Samsung Tab A7, instead of smartphone. We have asked the same participants above mentioned to play again our Android app and provide further gestures (collected gestures are summarized in

Table 9). Next, we evaluated the effectiveness of our proposal on the tablet.

Results of the application of single-view and multi-view learning (intermediate integration) on the best combinations are reported in Table 10. We notice that there is a drop in the performance of our solution when applied on a tablet (see Fig. 18 for a better assessment). The main reason is the screen size that, on tablets, enables longer scrolls (also using the forefinger instead of the thumb) and broader pinch-to-zooms. Consequently, they are composed of more fragments, and often they are performed faster. With regard to length, just to make an example, we have $mean_{phone}(ScD) = 304.31$, $SD_{phone}(ScD) = 132.11$ against

Table 8 Most useful gestures (and combinations of such) for gender classification applied for understanding generalization capability of our solution on different smartphone models. NSp = never seen participants, ASp = already seen participants. S-V = single-view, M-V = multi-view

Approach	Gesture	F1-score NSp/ASp
S-V RF	ScD	0.82/0.88
	ScU	0.78/0.85
	SwL	0.67/0.81
	SwR	0.72/0.82
	P2Z	0.74/0.79
M-V Interm.	ScD+SwL	0.79/0.85
	ScD+P2Z	0.82/0.86
	ScD+ScU	0.83/0.87
	ScD+SwL+P2Z	0.84/0.89
	ScD+SwR+P2Z	0.83/0.89
	ScD+ScU+SwL	0.81/0.88
	ScD+ScU+P2Z	0.82/0.87

$mean_{tablet}(ScD) = 441.18$, $SD_{tablet}(ScD) = 158.84$, and $mean_{phone}(P2Z) = 680.98$, $SD_{phone}(P2Z) = 239.81$ against $mean_{tablet}(P2Z) = 867.58$, $SD_{tablet}(P2Z) = 282.77$. Thus, if we want to deploy our method on tablets, we need proper data—captured through tablets—to fit machine learning models on. Conversely, the performance drop is minor when accounting SwL and SwR: interestingly, the swipe

gesture is performed similarly on both smartphones and tablets. This has spillovers on multi-view approaches that exhibit better results when considering a swipe in the gesture combination.

Overall, we recommend for environments/frameworks, interested in the non-intrusive gender classification, the use of multi-view approach when dealing with unknown users and unknown devices. Vice versa, they can rely on the single-view approach which is clearly faster.

5.2 Limitations and future research.

Even if the dataset used here is broader than those used in previous works, evaluating the real effectiveness of the proposal would require a larger number of participants with very diverse smartphones (as well as tablets). Also, the dataset does not include elder people—to whom large part of the research in smart spaces and healthcare is reserved [18, 19]—for whom the effectiveness of the proposal will be evaluated in the next future.

The energy efficiency of the proposed solution is intuitively higher with respect to proposals available in the literature. Yet, we have not measured it. For such analysis, we would need *ad-hoc* instruments and evaluation phases as did in [13, 15]. This will be the goal of future steps of this project. Furthermore, we have only considered here combination of heterogeneous gestures. Next, we aim to study combination of gestures of the same kind and/or consecutive ones to account for the order. In order to enable such a study, we need to develop a new game that
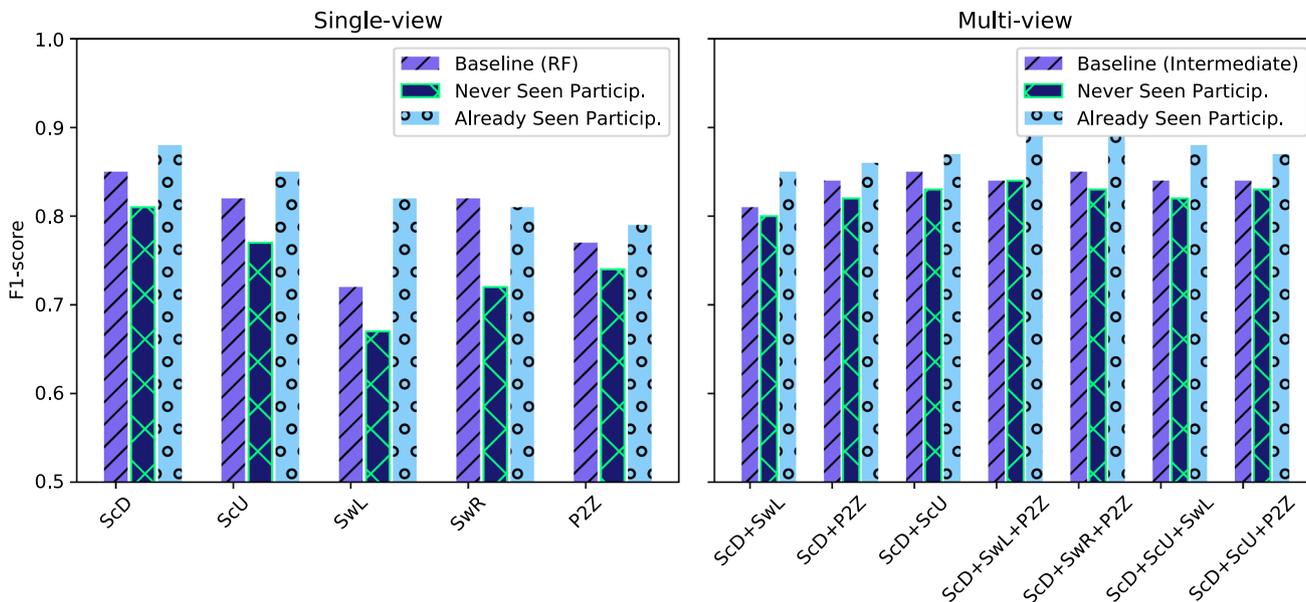
Fig. 17 Comparison of performance achieved when dealing with never seen participants on different smartphones, and already seen participants on different smartphones against those obtained in Sect. 4.4 by best approaches (baseline)

Table 9 Collected gestures for the evaluation of generalization capability on different kind of mobile device, i.e., tablet Samsung Tab A7. NSp = never seen participants, ASp = already seen participants. m = male, f = female

Gesture	#Samples	
	NSp (m/f)	ASp (m/f)
ScD	73/94	40/21
ScU	32/40	14/8
SwL	27/43	13/8
SwR	29/40	15/7
P2Z	41/49	21/12
Total	(468) 202/266	(159) 103/56

Table 10 Most useful gestures (and combinations of such) for gender classification applied for understanding generalization capability of our solution on tablets. NSp = never seen participants, ASp = already seen participants. S-V = single-view, M-V = multi-view

Approach	Gesture	F1-score NSp/ASp
S-V RF	ScD	0.79/0.82
	ScU	0.74/0.80
	SwL	0.66/0.80
	SwR	0.71/0.81
	P2Z	0.72/0.77
M-V Interm.	ScD+SwL	0.77/0.80
	ScD+P2Z	0.79/0.80
	ScD+ScU	0.80/0.81
	ScD+SwL+P2Z	0.82/0.84
	ScD+SwR+P2Z	0.82/0.85
	ScD+ScU+SwL	0.78/0.83
	ScD+ScU+P2Z	0.80/0.81

forces the user to perform diverse kinds of gestures in a row. In this work, we have manually engineered the features suitable to describe the considered touch gestures; on the one hand, this allowed us to identify the significant features (being more explicable) and comply with the latest directive of EU Commission on biometric systems[3] but, on the other hand, we have overlooked other approaches to gender classification based on touch gestures. In this vein, we will evaluate (and then compare) methods based on representation learning, using the stream of data directly

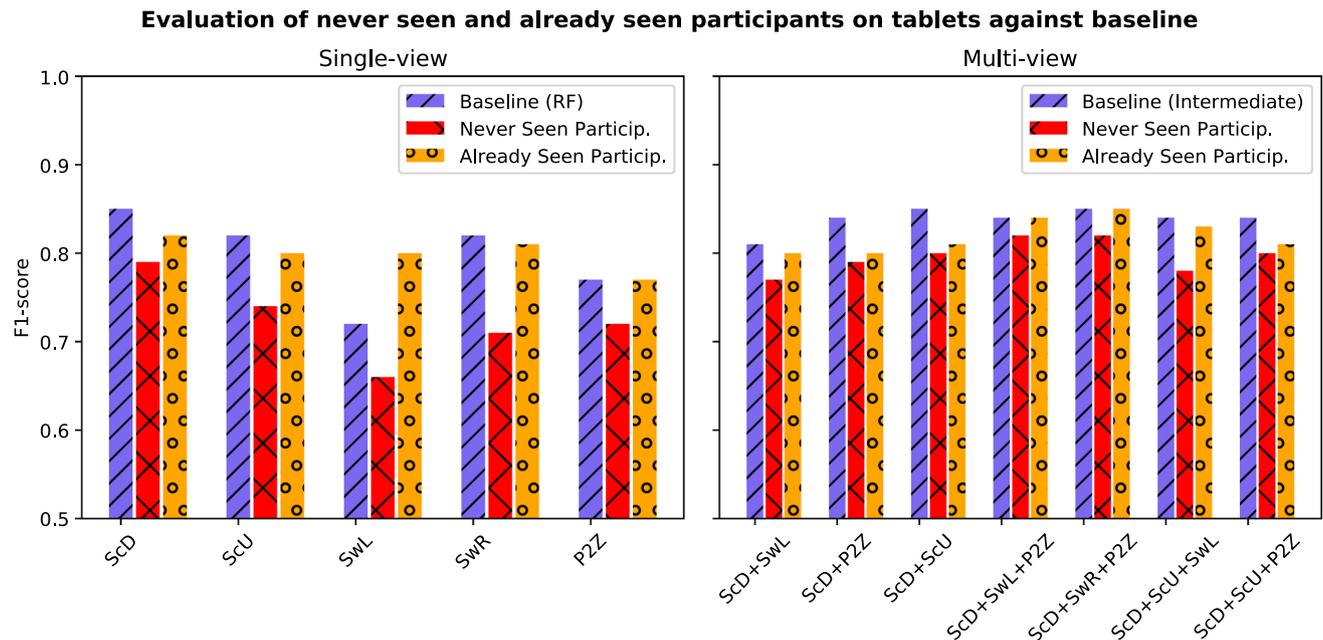provided by Touch APIs or visual representation of gestures.

Lastly, part of our effort will be put toward the development of a "mature" application improving our current prototype. However, datasets employed for the experiments are available at https://bit.ly/3IZ6Lvz.

### 5.3 Automatic gender classification: key for heaven, key for hell

Back in the days, the Nobel prized Richard P. Feynman explored the capacity of science to be a catalyst for both good and evil, stating that *"To every man is given the key to the gates of heaven; the same key opens the gates of hell."* These potentialities and risks apply particularly well today for automatic classification, like the one proposed here for gender. Indeed, in all the contexts in which automatic decisions/classifications occur, designers, engineers and all the persons behind the project must ensure that the solution has been designed and implemented in a way that certifies both its effectiveness and legitimacy, so that the results are beneficial and/or benign [16]. That is to say, we should ensure that it is an effective means for achieving some policy goal while remaining procedurally fair. In this regard, the perspective in enhancing authentication systems with soft biometric traits such as gender are quite promising. It is not by chance the growing literature about gender-aware systems [6, 7, 66, 71] fostering inclusion, enhancing user experience and human–computer interactions. For example, intelligent systems in a smart space can be customized based on gender information to provide an enhanced user experience. Nevertheless, these kinds of automatic classification of personal traits, as highlighted by Danaher et al. [16] can be problematic. If not required for beneficial or benign goals, obtaining the gender information should be unfeasible. If exploited by malware apps, unwanted software, or attackers of every type, automatic gender classification could clearly undermine the users' fundamental rights. The very possibility of stealthy exploit users' gender for shaping individuals' conception of the world, opinions, and values demand a deeper reflection. Moreover, the issue becomes even more important when we consider teenagers or kids on the phone[4].

For what concerns malware, there are several examples of malware detection systems based on the smartphones' permissions analysis [43, 49]. If such malware exploit an approach like the one proposed here to obtain the gender information, they will not request any particular permission becoming a severe threat for users' privacy and security. Conversely to [37] where the malware must declare the ACTIVITY_RECOGNITION permission to use gyroscope

---

[3] 2021/0106 (COD). The EU Commission defines biometric systems as *High-risk AI* and foster the development of solution enabling *human oversight*.

[4] https://bit.ly/3aQ4DZ4.

Evaluation of never seen and already seen participants on tablets against baseline

Fig. 18 Comparison of performance achieved when dealing with never seen participants on tablets, and already seen participants on tablets against those obtained in Sect. 4.4 by best approaches (Baseline)
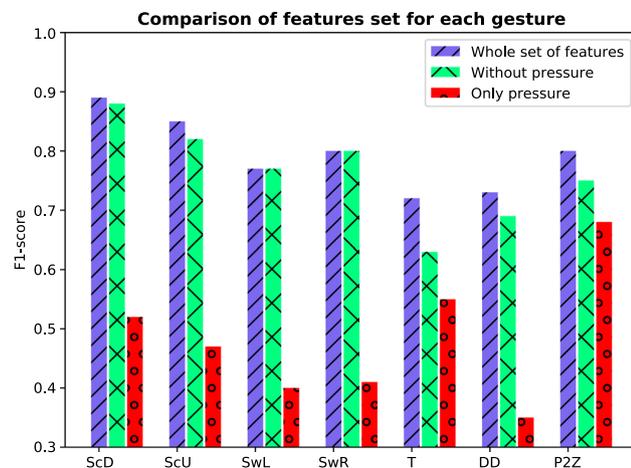
Fig. 19 Results obtained with each gestures on varying the features set used. In slate blue, we have the whole set of features engineered, in green we have the whole set without pressure-related features, and in red we have the set with only pressure-related features

and accelerometer, our solution does not demand permissions because touch gestures are implicitly enabled to control the device and every app. In fact, antivirus or antimalware software on VirusTotal[5] do not flag our app as malware.

Furthermore, there are a series of significant social and ethical concerns about automatic gender classification that are not yet fully explored. Among these, there are those

connected to gender as an *identity* descriptor [61, 63] more than *sex*, as we use here. People with diverse gender identities, including those identifying as transgender or gender nonbinary, are particularly concerned that these systems could miscategorize them [28]. People who express their gender differently from stereotypical male and female norms already experience discrimination and harm resulting from being miscategorized or misunderstood [54]. One of the participants in the recent study by Hamidi et al. [28] described how they would feel hurt if a *"million-dollar piece of software developed by however many people"* decided that they are not who they themselves believe they are. It emerges that the future of our project should involve *(a)* collaboration with vulnerable communities potentially harmed by this kind of automatic gender classification, *(b)* evaluating how miscategorization of individuals impacts the systems that make use of it, i.e., the emerging observable behavior, and *(c)* taking into account mechanisms to support minorities in systems; performance is improved by automatic gender classification.

6 Conclusion

This paper has proposed a novel machine learning-based solution for users' gender classification relying on touch gestures information gathered with smartphones. Extensive experiments with two approaches, i.e., single-view and

---

multi-view learning (early, intermediate, and late integration) and with different scenarios (unknown users, unknown devices), demonstrate the feasibility of our solution (from F1-score of 0.65 up to 0.89 based on the experiment and scenario). The gender information captured in this non-intrusive way can be used to improve authentication system's performance as well as for healthcare, smart spaces, and UIs' customization. Moreover, we shed light on the potentiality and risks connected with the use of this kind of automatic gender classification. We plan to develop a framework that utilizes the gender information for improving the performance of a biometric-based user authentication system in smart spaces, and to evaluate our solution (in terms of acceptance, affordance, experience, and so on) with final users (e.g., as done in [17, 27, 48]). Therefore, the approach here presented is part of a broader project that will also confront with problems arising in the minorities, in communities of people that do not categorize themselves between male and female, and will take into account whether transgender or gender nonbinary persons feel harmed and how by our system. Data used here will be made available through the project's official page to help researchers develop and compare their solutions against ours.

A Are pressure-related features accurate?

In this section, we report the results obtained by using the solely pressure-related features of gestures for gender recognition. Figure 19 depicts the results obtained. Overall, this experiment confirms that pressure-related features are not very significant for the gender recognition. In more detail, ScD and ScU with SwR, SwL and DD showed the worse results, from 0.35 to 0.52 F1-score in LOUO-CV. Instead, T and P2Z exhibit higher F1-score in this experiment, with 0.55 and 0.68, respectively (in LOUO-CV).

Declarations

Conflict of Interest The authors declare that they have no conflict of interest.

References

1. Ahmidi N, Tao L, Sefati S, Gao Y, Lea C, Haro BB, Zappella L, Khudanpur S, Vidal R, Hager GD (2017) A dataset and benchmarks for segmentation and recognition of gestures in robotic surgery. IEEE Trans Biomed Eng 64(9):2025–2041
2. Antal M, Nemes G (2016) Gender recognition from mobile biometric data. In: 2016 IEEE 11th international symposium on applied computational intelligence and informatics (SACI), pp. 243–248. IEEE
3. Anthony L, Brown Q, Tate B, Nias J, Brewer R, Irwin G (2014) Designing smarter touch-based interfaces for educational contexts. Pers Ubiquit Comput 18(6):1471–1483
4. Batarseh FA, Pithadia J (2017) Context-aware user interfaces for intelligent emergency applications. In: International and Interdisciplinary Conference on Modeling and Using Context, pp. 359–369. Springer
5. Bihl TJ, Bauer KW, Temple MA, Ramsey B (2015) Dimensional reduction analysis for physical layer device fingerprints with application to zigbee and z-wave devices. In: MILCOM 2015-2015 IEEE Military Communications Conference, pp. 360–365. IEEE
6. Bleja J, Langer H, Grossmann U, Mörz E (2020) Smart cities for everyone–age and gender as potential exclusion factors. In: 2020 IEEE European Technology and Engineering Management Summit (E-TEMS), pp. 1–5. IEEE
7. Breslin S, Wadhwa B (2018) Gender and human-computer interaction. The Wiley Handbook of Human Comput Interact 1:71–87
8. Cao H, Bernard S, Sabourin R, Heutte L (2019) Random forest dissimilarity based multi-view learning for radiomics application. Pattern Recogn 88:185–197
9. Chai T, Prasad S, Wang S (2019) Boosting palmprint identification with gender information using deepnet. Futur Gener Comput Syst 99:41–53
10. Chen X, Niu L, Veeraraghavan A, Sabharwal A (2019) Faceengage: Robust estimation of gameplay engagement from user-contributed (youtube) videos. IEEE Transactions on Affective Computing pp. 1–1 . https://doi.org/10.1109/TAFFC.2019.2945014
11. Cheng Y, Ji X, Li X, Zhang T, Malebary S, Qu X, Xu W (2020) Identifying child users via touchscreen interactions. ACM Trans Sensor Netw (TOSN) 16(4):1–25
12. Cornelius CT, Kotz DF (2012) Recognizing whether sensors are on the same body. Pervasive Mob Comput 8(6):822–836
13. Cozza F, Guarino A, Isernia F, Malandrino D, Rapuano A, Schiavone R, Zaccagnino R (2020) Hybrid and lightweight detection of third party tracking: Design, implementation, and evaluation. Comput Netw 167:106993
14. Craley J, Murray TS, Mendat DR, Andreou AG (2017) Action recognition using micro-doppler signatures and a recurrent neural network. In: 2017 51st Annual Conference on Information Sciences and Systems (CISS), pp. 1–5. IEEE
15. D'Ambrosio S, Pasquale SD, Iannone G, Malandrino D, Negro A, Patimo G, Scarano V, Spinelli R, Zaccagnino R (2017) Privacy as a proxy for green web browsing: Methodology and experimentation. Comput Netw 126:81–99. https://doi.org/10.1016/j.comnet.2017.07.003

16. Danaher J, Hogan MJ, Noone C, Kennedy R, Behan A, De Paor A, Felzmann H, Haklay M, Khoo SM, Morison J et al (2017) Algorithmic governance: Developing a research agenda through the power of collective intelligence. Big Data & Soc 4(2):2053951717726554

17. De Prisco R, Guarino A, Lettieri N, Malandrino D, Zaccagnino R (2021) Providing music service in ambient intelligence: experiments with gym users. Exp Syst Appl 177:114951

18. Dobkin BH (2017) A rehabilitation-internet-of-things in the home to augment motor skills and exercise training. Neurorehabil Neural Repair 31(3):217–227

19. Ellis T, Rochester L (2018) Mobilizing parkinson's disease: the future of exercise. J Parkinsons Dis 8(s1):S95–S100

20. Elsts A, Twomey N, McConville R, Craddock I (2020) Energy-efficient activity recognition framework using wearable accelerometers. J Netw Comput Appl 168:102770

21. Fairhurst M, Da Costa-Abreu M (2011) Using keystroke dynamics for gender identification in social network environment. In: 4th International Conference on Imaging for Crime Detection and Prevention 2011 (ICDP 2011), pp. 1–6 . https://doi.org/10.1049/ic.2011.0124

22. Fayyaz M, Yasmin M, Sharif M, Raza M (2021) J-ldfr: joint low-level and deep neural network feature representations for pedestrian gender classification. Neural Comput Appl 33:361–391

23. Fratello M, Caiazzo G, Trojsi F, Russo A, Tedeschi G, Tagliaferri R, Esposito F (2017) Multi-view ensemble classification of brain connectivity images for neurodegeneration type discrimination. Neuroinformatics 15(2):199–213

24. Freedman DA (2009) Statistical models: theory and practice. cambridge university press

25. García-Martín E, Rodrigues CF, Riley G, Grahn H (2019) Estimation of energy consumption in machine learning. J Parallel and Distribut Comput 134:75–88

26. Giot R, Rosenberger C (2012) A new soft biometric approach for keystroke dynamics based on gender recognition. Int J Inf Technol Manage 11(1–2):35–49

27. Guarino A, Lettieri N, Malandrino D, Zaccagnino R (2021) A machine learning-based approach to identify unlawful practices in online terms of service: analysis, implementation and evaluation. Neural Computing and Applications pp. 1–19

28. Hamidi F, Scheuerman MK, Branham SM (2018) Gender recognition or gender reductionism? the social implications of embedded gender recognition systems. In: Proceedings of the 2018 chi conference on human factors in computing systems, pp. 1–13

29. Hamza Z, Salivia G (2015) Study of touch gestures performance in touch devices by young children. Int J Recent and Innovation Trends in Comput Commun 3(3):1395–1400

30. Hemminki S, Nurmi P, Tarkoma S (2013) Accelerometer-based transportation mode detection on smartphones. In: Proceedings of the 11th ACM conference on embedded networked sensor systems, pp. 1–14

31. Hu J (2018) An approach to eeg-based gender recognition using entropy measurement methods. Knowl-Based Syst 140:134–141

32. Huberty CJ, Wisenbaker JM (1992) Variable importance in multivariate group comparisons. J Educ Stat 17(1):75–91

33. Ibrahim TM, Alarood AA, Chiroma H, Al-garadi MA, Rana N, Muhammad AN, Abubakar A, Haruna K, Gabralla LA et al (2019) Recent advances in mobile touch screen security authentication methods: A systematic literature review. Comput Security 85:1–24

34. Idrus SZS, Cherrier E, Rosenberger C, Mondal S, Bours P (2015) Keystroke dynamics performance enhancement with soft biometrics. In: IEEE International Conference on Identity, Security and Behavior Analysis (ISBA 2015), pp. 1–7. IEEE

35. Isaac ER, Elias S, Rajagopalan S, Easwarakumar K (2019) Multiview gait-based gender classification through pose-based voting. Pattern Recogn Lett 126:41–50

36. Jain A, Kanhangad V (2018) Gender classification in smartphones using gait information. Expert Syst Appl 93:257–266

37. Jain A, Kanhangad V (2019) Gender recognition in smartphones using touchscreen gestures. Pattern Recogn Lett 125:604–611

38. Jain AK, Nandakumar K, Lu X, Park U (2004) Integrating faces, fingerprints, and soft biometric traits for user recognition. In: International Workshop on Biometric Authentication, pp. 259–269. Springer

39. James G, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning, vol. 112. Springer

40. Jamil A, Faisal CMN, Habib MA, Jabbar S, Ahmad H (2020) Analyzing the impact of age and gender on user interaction in gaming environment. In: International Conference on Innovative Computing and Communications, pp. 721–729. Springer

41. Japkowicz N, Shah M (2011) Evaluating learning algorithms: a classification perspective. Cambridge University Press

42. Jin Y, Tintarev N, Verbert K (2018) Effects of individual traits on diversity-aware music recommender user interfaces. In: Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization, pp. 291–299

43. Li J, Sun L, Yan Q, Li Z, Srisa-An W, Ye H (2018) Significant permission identification for machine-learning-based android malware detection. IEEE Trans Industr Inf 14(7):3216–3225

44. Li Y, Ngom A (2015) Data integration in machine learning. In: 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 1665–1671. IEEE

45. Li Y, Wu FX, Ngom A (2016) A review on machine learning principles for multi-view biological data integration. Brief Bioinform 19(2):325–340

46. Li Y, Wu FX, Ngom A (2018) A review on machine learning principles for multi-view biological data integration. Brief Bioinform 19(2):325–340

47. Ling TC, Alam T, Hussin AA (2020) Investigation of the role of individual's gender in the design of graphical user interface for mobile learning device. In: Journal of Physics: Conference Series, vol. 1529, p. 032021. IOP Publishing

48. Liu L (2021) The artistic design of user interaction experience for mobile systems based on context-awareness and machine learning. Neural Computing and Applications pp. 1–11

49. Mahindru A, Singh P (2017) Dynamic permissions based android malware detection using machine learning techniques. In: Proceedings of the 10th innovations in software engineering conference, pp. 202–210

50. Mallouh AA, Qawaqneh Z, Barkana BD (2018) New transformed features generated by deep bottleneck extractor and a gmm-ubm classifier for speaker age and gender classification. Neural Comput Appl 30(8):2581–2593

51. Masood R, Zhao BZH, Asghar HJ, Kaafar MA (2018) Touch and you're trapp (ck) ed: Quantifying the uniqueness of touch gestures for tracking. Proceed Privacy Enhancing Technol 2018(2):122–142

52. McIntosh A, Hassan S, Hindle A (2019) What can android mobile app developers do about the energy consumption of machine learning? Empir Softw Eng 24(2):562–601

53. McKight P, Najab J (2010) Kruskal-wallis test. corsini encyclopedia of psychology

54. McLemore KA (2015) Experiences with misgendering: Identity misclassification of transgender spectrum individuals. Self and Identity 14(1):51–74

55. Nacher V, Jaen J, Navarro E, Catala A, González P (2015) Multi-touch gestures for pre-kindergarten children. Int J Hum Comput Stud 73:37–51

56. Park B, Song S, Kim J, Park W, Jang H (2007) User customization methods based on mental models: modular ui optimized for customizing in handheld device. In: International Conference on Human-Computer Interaction, pp. 445–451. Springer

57. Park U, Jain AK (2010) Face matching and retrieval using soft biometrics. IEEE Trans Inf Forensics Secur 5(3):406–415

58. Passig D, Levin H (2000) Gender preferences for multimedia interfaces. J Comput Assist Learn 16(1):64–71

59. Pavlidis P, Weston J, Cai J, Noble WS (2002) Learning gene functional classifications from multiple data types. J Comput Biol 9(2):401–411

60. Ranjan R, Patel VM, Chellappa R (2017) Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. IEEE Trans Pattern Anal Mach Intell 41(1):121–135

61. Repo J (2015) The biopolitics of gender. Oxford University Press

62. Rzecki K, Pławiak P, Niedźwiecki M, Sośnicki T, Leśkow J, Ciesielski M (2017) Person recognition based on touch screen gestures using computational intelligence methods. Inf Sci 415:70–84

63. Satchell C (2010) Women are people too: The problem of designing for gender. In: ACM Conference on Human Factors in Computing Systems (CHI 2010)

64. Shapiro SS, Wilk MB (1965) An Analysis of Variance Test for Normality (Complete Samples). Biometrika 52(3/4):591–611

65. Sheikhan M, Bejani M, Gharavian D (2013) Modular neural-svm scheme for speech emotion recognition using anova feature selection method. Neural Comput Appl 23(1):215–227

66. Singh YJ (2020) Is smart mobility also gender-smart? J Gend Stud 29(7):832–846

67. Smith P, Chen C (2018) Transfer learning with deep cnns for gender recognition and age estimation. In: 2018 IEEE International Conference on Big Data (Big Data), pp. 2564–2571. IEEE

68. Sohail S, Syed AM, Jamil A (2020) The influence of gender on performance in gaming environment with different typographic factors. In: https://doi.org/10.2139/ssrn.3680706, pp. 1–5

69. Sun S (2013) A survey of multi-view machine learning. Neural Comput Appl 23(7):2031–2038

70. Tao L, Elhamifar E, Khudanpur S, Hager GD, Vidal R (2012) Sparse hidden markov models for surgical gesture classification and skill evaluation. In: International conference on information processing in computer-assisted interventions, pp. 167–177. Springer

71. Thakare C, Chaurasia NK, Rathod D, Joshi G, Gudadhe S (2021) Gender aware cnn for speech emotion recognition. In: Health Informatics: A Computational Perspective in Healthcare, pp. 367–377. Springer

72. Toan N, Aditi R, Nasir M (2019) Kid on the phone! toward automatic detection of children on mobile devices. Comput Secur 84:334–348

73. Tong L, Wu H, Wang MD (2021) Integrating multi-omics data by learning modality invariant representations for improved prediction of overall survival of cancer. Methods 189:74–85. https://doi.org/10.1016/j.ymeth.2020.07.008.https://www.sciencedirect.com/science/article/pii/S1046202320300232. Machine learning for the analysis of multi-omics data

74. Vatavu RD, Anthony L, Brown Q (2015) Child or adult? Inferring Smartphone users' age group from touch measurements alone. In: IFIP Conference on Human-Computer Interaction, pp. 1–9. Springer

75. Xue G, Liu S, Gong D, Ma Y (2021) Atp-densenet: a hybrid deep learning-based gender identification of handwriting. Neural Comput Appl 33(10):4611–4622

76. Zaccagnino R, Capo C, Guarino A, Lettieri N, Malandrino D (2021) Techno-regulation and intelligent safeguards. Multimed Tools and Appl pp. 1–22

77. Zaman F, Khan I, Khusro S (2020) Towards the design of context-aware adaptive user interfaces to minimize drivers' distractions. Mob Inf Syst 2020:8858886. https://doi.org/10.1155/2020/8858886

78. Zhao H, Wang P (2019) A short review of age and gender recognition based on speech. In: 2019 IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing,(HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS), pp. 183–185. IEEE

79. Zheng L, Wu D, Ruan X, Weng S, Peng A, Tang B, Lu H, Shi H, Zheng H (2017) A novel energy-efficient approach for human activity recognition. Sensors 17(9):2064