Non-invasive Multi-modal Human Identification System Combining ECG, GSR, and Airflow Biosignals

Carmen Camara[1] · Pedro Peris-Lopez[1] · Juan E. Tapiador[1] · Guillermo Suarez-Tangil[1]

Abstract A huge amount of data can be collected through a wide variety of sensor technologies. Data mining techniques are often useful for the analysis of gathered data. This paper studies the use of three wearable sensors that monitor the electrocardiogram, airflow, and galvanic skin response of a subject with the purpose of designing an efficient multi-modal human identification system. The proposed system, based on the rotation forest ensemble algorithm, offers a high accuracy (99.6 % true acceptance rate and just 0.1 % false positive rate). For its evaluation, the proposed system was testing against the characteristics commonly demanded in a biometric system, including universality, uniqueness, permanence, and acceptance. Finally, a proof-of-concept implementation of the system is demonstrated on a smartphone and its performance is evaluated in terms of processing speed and power consumption. The identification of a sample is extremely efficient, taking around 200 ms and consuming just a few millijoules. It is thus feasible to use the proposed system on a regular smartphone for user identification.

✉ Carmen Camara
macamara@pa.uc3m.es

Pedro Peris-Lopez
pperis@inf.uc3m.es

Juan E. Tapiador
jestevez@inf.uc3m.es

Guillermo Suarez-Tangil
guillermo.suarez.tangil@uc3m.es

[1] Department of Computer Science, Carlos III University of Madrid (UC3 M), Avda. de la Universidad 30, Leganés 28911, Madrid, Spain

1 Introduction

As a result of recent advances in hardware technology, there is a substantial diversity of inexpensive sensors on the market. These devices have become computationally powerful and tiny, and can be integrated into devices such as smartphones. Medical devices in or around the human body have sensors for monitoring physiological parameters. Activity-tracking devices such as smart-bands are equipped with sensors that can monitor vital parameters such as the heart rate or body temperature. Physiological parameters are mainly monitored for the treatment of ailments or the tracking of daily activities in order to lead a long and healthy life [1]. Vital signals can be employed for security applications, such as biometrics, key establishment protocols, and entropy generators [2].

The present study focuses on the biometric use of such sensors and applies signal processing (SP) and artificial intelligence (AI) techniques. Biometrics refers to the automatic recognition of individuals based on measurable characteristics. These characteristics can be based on the physiology of a person (e.g., fingerprints, face, and iris). User behavior (e.g., keystroke timing or gait features) can also be used to identify a particular user. The National Institute of Standards and Technology (NIST) splits biometric characteristics into physiological and behavioral categories [3].

A wide set of physiological and behavioral characteristics can be used in biometrics. The choice of characteristics is guided by several properties required by a biometric system: (1) every person should have the

biometric characteristics (universality); (2) individuals within the target population can be distinguished from the biometric characteristics; that is, two individuals cannot be the same in terms of the characteristics (uniqueness); (3) the biometric characteristics should be invariant over time (permanence); (4) the acquisition of the biometric characteristics should be easy achieved using sensors, allowing subsequent processing and feature extraction (collectability or measurability); (5) users within the relevant population should not have objections to the use of the biometric characteristics (acceptability); (6) the accuracy level and recognition speed of the system must be adequate (performance); and (7) it should be difficult to counterfeit the biometric characteristics in order to fool the system or impersonate a user (resistance to circumvention). The proposed human identification system is tested in terms of these seven properties (see Sect. 3.2).

The goal of a biometric system can be identification (authentication) or verification. In a verification system, the user claims an identity and the system checks whether the claim is valid. The biometric system performs a one-to-one comparison: (1) the user provides an identifier (e.g., user name or ID number) so that the system can locate the corresponding stored template; (2) the template of the claimed user is retrieved; (3) a one-to-one comparison is executed and the classifier returns a probability of matching. In an identification system (the type of system proposed here), the system does not know the link between the template and the user. That is, the classifier does one-to-many comparisons and the user is successfully identified if one of those comparisons is within a previously established threshold.

A vast number of biometric systems are based on just one biometric characteristic. For instance, iris recognition, fingerprint identification, or DNA (deoxyribonucleic acid) are well-established biometric solutions [4]. Recently, the use of biosignals has attracted much attention and new biometric schemes, such as those based on electrocardiograms (ECGs) [5], electroencephalograms (EEGs) [6], or photoplethysmography (PPM) [7], have emerged. To overcome the limitations of uni-modal biometric systems (environmental variations, sensitivity to aging, and non-guarantee of perfect recognition), multi-modal systems combine the strengths of several biometric characteristics. For instance, multi-modal biometric systems based on voice and leap movements [8] or fingerprints and veins [9] have been recently proposed. In the context of biosignals, the combined used of ECG and EEG is one of the most used approaches [10, 11].

The distinctiveness of many of these biosignals among individuals can be due to a number of factors. In the case of the ECG, the electrical currents generated by the heart as it beats spread both within the heart and throughout the user's body. Thus, many unique signal features depend on anatomic features such as the difference in the position and size of the heart and the remaining physical conditions of the body [12]. The same is applicable to other electrical phenomena such as the galvanic skin response (GSR), which is used in the present study along with airflow and the ECG. In addition, the joint use of various biosignals will likely lead to more accurate biometric systems. This paper explores this assumption and reports experimental results for a prototype platform.

The general architecture of a biometric system based on machine learning techniques is described below. First, the raw signals are acquired by the sensors and prepared for analysis. This preprocessing phase may include normalization, re-sampling, and smoothing. Once preprocessed, the signals are characterized (feature extraction; see Sect. 2.3) by a set of numerical or nominal parameters that represent relevant information found in the signal. In the training phase, a set of samples are used to build the classifier, where each sample is composed of a list of features and a label corresponding to an individual. In the testing phase, all of the above-mentioned steps are repeated with the new samples and the classification model is used to determine the subject's label for an input sample. In our particular case, as described in Sect. 2.2, an ensemble classifier, which yields high performance in terms of accuracy and diversity, is used.

2 Materials and Methods

2.1 Dataset

The dataset was collected using an e-Health Sensor Platform over an Arduino Uno board (ATmega328 microcontroller). This is an inexpensive prototyping platform for biometric and medical applications in which body monitoring is needed. The platform is essentially a printed circuit board with a variety of body sensors that can be easily attached to an Arduino or a Raspeberry Pi board. In addition, the e-Health Sensor Platform comes with a C++ API that allows one to continuously read all sensors and send the information out through any of the available radio interfaces (Wi-Fi, 3G, GPRS, Bluetooth, 802.15.4, or ZigBee). Three sensors were connected to the board: an ECG sensor, a nasal airflow sensor, and a GSR sensor. Figure 1 shows how these external sensors are deployed for a user and connected to the Arduino board.

The ECG monitors and records the electrical activity of the heart. In particular, current flows within the body and potential differences occur as a consequence of the electrical activity of the cells. The ECG represents the evolution of these potential differences over time. A set of
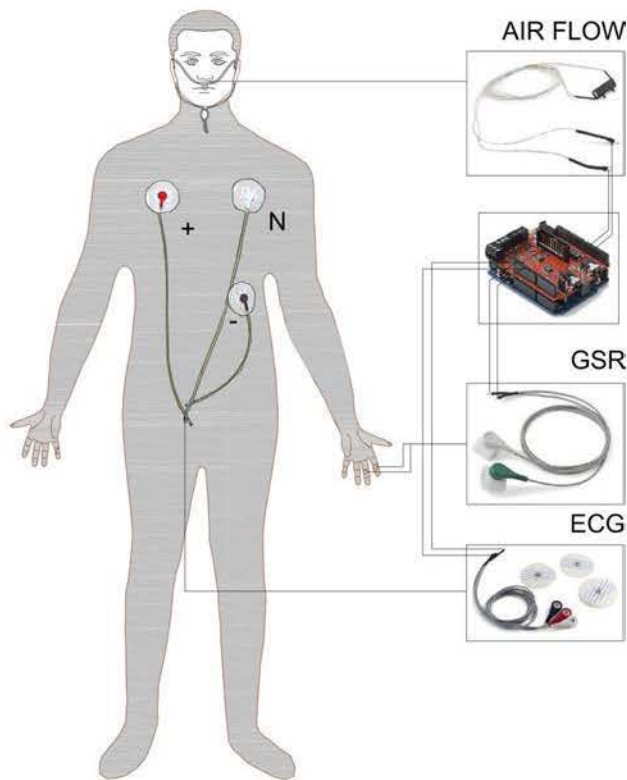
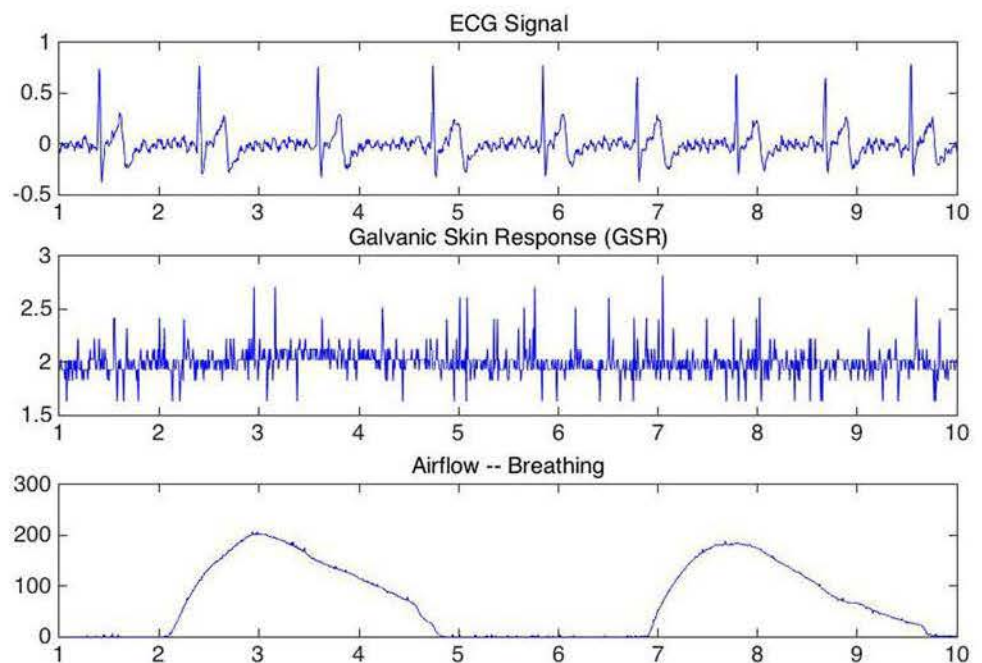Fig. 1 Acquisition of EEG, airflow, and GSR biosignals

analog value of between 0 and 5 V, which represents the ECG waveform. For diagnosis and a complete ECG (allowing interpretation of specific areas of the heart), there exists a standardized system with 10 electrodes to produce 12-lead ECG.

The upper airway is composed of the mouth and the two parallel nasal passages of the nose. This is the entry to the respiratory system and during inspiration and expiration all respired air passed through it. Interested readers are urged to consult [13], where the authors review the mechanisms of airflow in the human nasal airways. Tools for assessing airflow that measure parameters directly or indirectly have been developed [14]. The present study uses the e-Health airflow sensor for measuring the breathing rate for an individual. The device is connected through the analog input and returns a value of between 0 and 1024.

Skin conductance measures the electrical conductance of the skin. GSR, electrodermal response (EDR), and skin conductance response (SCR) are used to express the same concept. The measured level depends on the amount of moisture on the skin, varying according to the sweating rate. Skin conductance can be used as an indicator of the psychological or physiological arousal as the sympathetic nervous system controls the sweat glands. The present experiments used the e-Health GSR sensor, which measures the electrical conductance between two points. Note that GSR sensors can be considered as a sort of ohmmeter [15]. Lower values are obtained for people with drier skin.

The experimentation was conducted with six volunteers who were subjected to biosignal recording for periods of

electrodes over the skin is used for data acquisition. This study used a three-electrode system, which is often used for health monitoring. The e-Health ECG sensor returns an

Fig. 2 Examples of ECG, airflow, and GSR biosignals

around 1 min during several sessions on different days. The number of samples is not exactly the same for all users since the low-cost sensors used did not work properly during certain short intervals and we manually discarded those portions of the signals. Figure 2 shows an example of the three studied biosignals. These raw signals need to be preprocessed prior to their usage in the identification system. In particular, for each raw signal, the constant component (DC value) is eliminated, the signal is passed through a filter, and finally the signal shape is smoothed [16]. As discussed above, this preprocessing phase is required when dealing with signals gathered through a sensor, as the samples are often noisy.

2.2 Classification Algorithms

In inductive machine learning, a set of rules is inferred from a training set composed of input instances. After that, the built model can be used to classify new instances (testing set). There is a wide variety of algorithms and their selection is critical for some applications. Classifiers are often grouped by their function or form, for example, decision tree learning algorithms, kernel methods, or ensemble approaches. This paper focuses on ensembles due to their classification power. We first provide an overview of ensemble classifiers and then discuss in detail the particular algorithm (rotation forest) used in our experiments.

2.2.1 Ensemble Classification

An ensemble classifier assigns weights to individual classifiers and combines them [17–19]. The goal is that the ensemble outperforms the individual classifiers [20]. The two classic approaches for building ensembles are bagging (also called bootstrap aggregating) and boosting.

In bagging [21], each classifier in the ensemble is trained on a sample of instances from the training set. Sampling is done uniformly and with replacement. The pseudo-code of bagging training is given in Algorithm 1. One of the advantages of this method is that the training of the classifiers can be executed in parallel. After training, if a new instance has to be classified, each classifier returns a class and the one with the highest number of votes (majority voting) is the output of the bagged classifier.

Algorithm 1 Bagging Training [21]

Require: CA (a classification algorithm), K (number of iterations), X (original training set), N (the sample size)

$k \leftarrow 1$

repeat

    $X_t \leftarrow$ a sample of N instances from X with replacement.

    Construct a classifier C_t of sort CA and use X_t as the training set.

    $k \leftarrow k + 1$

until $k > K$

Algorithm 2 Boosting Training [22]

Require: CA (a classification algorithm), X (original training set), N (the sample size for the first classifier)

Ensure: C_1, C_2, C_3

    $X_1 \leftarrow$ Randomly selects N instances from X without replacement.

    $C_1 \leftarrow CA(X_1)$

    $X_2 \leftarrow$ Randomly selects instances from $X - X_1$ without replacement such that half of them are classified incorrectly and half are classified correctly by C_1.

    $C_2 \leftarrow CA(X_2)$

    $X_3 \leftarrow$ is composed of instances in $X - X_1 - X_2$ on which C_1 and C_2 disagree.

    $C_3 \leftarrow CA(X_3)$

Boosting is a family of methods [22], for example AdaBoost [23], whose goal is to improve the performance of a weak classifier. The algorithm iteratively executes a weak classifier on training data with different distributions. Once a classifier is trained, the data are re-weighted. Instances that are misclassi-fied gain weight and those correctly classified lose weight. Therefore, each weak classifier puts more emphasis on the instances that were incorrectly classified by the previous clas-sifier. Algorithm 2 shows the pseudo-code for boosting train-ing. After training, for the classification of a new instance, each classifier returns a class. Finally, the class with the highest number of votes is the output provided by the ensemble.

Diversity between the base classifiers forming the ensemble is a desirable property in order to increase its accuracy. It also contributes to reduce the variance error without increasing the bias error. The classifiers should thus be as different as possible, which means a low correlation between their errors [24]. Despite their differences, all the classifiers have to perform well with the training set. Various strategies can be followed for creating diversity in ensembles. In one study [25], the following taxonomy was introduced (the categories are not mutually exclusive and can be combined):

- Manipulating the classification algorithm. For instance, each classifier of the ensemble is trained using different parameters or employing different starting points in the hypothesis space.
- Manipulating the training sample. In the training phase, each classifier is trained with a different dataset. Re-sampling and partitioning are possible strategies.
- Changing the target attribute representation. In this case, the target attribute is transformed into a set of simple representations of the target attribute. This is achieved through aggregation or more complicated

functions. Each base classifier uses one of the mentioned representations.
- Partitioning the search space. The search space is divided and each base classier focuses on one of these partitions. Divide-and-conquer and feature subset are two popular approaches.
- Hybridization. Several types of classification algorithm or ensemble strategy are combined.

The present study uses the rotation forest (RF) algorithm, which partitions the search space by applying feature extraction to subsets of features. RF ensembles outperform boosting (i.e., AdaBoost) and decision-tree-based ensembles (i.e., random forest), and offer higher diversity than that of ensembles based on bagging. The following section provides a brief description of the RF algorithm.

2.2.2 Rotation Forest

RF is a classifier ensemble method based on feature extraction. Accuracy and diversity are the main features required of the built-in classifiers [26]. Decision trees are used as the base classifier since they are very accurate while also being sensitive to the rotation of the feature axes. For achieving diversity, after the feature set has been randomly divided into subsets, principal component analysis (PCA) is applied to each subset and the whole set of components is employed to preserve variability in the data.

Algorithm 3 shows the pseudo-code for RF. The variables X (an $N \times n$ matrix) and Y (an $N \times 1$ matrix) represent the objects and the labels of the training set, respectively. Each element y_j of Y belongs to one of the class labels (i.e., $\{\omega_1, \ldots, \omega_c\}$). There are L classifiers in the ensemble $\{C_1, \ldots, C_L\}$. F represents the feature set.

Algorithm 3 Rotation Forest Training [26]

Require: CA (a classification algorithm), X (original training set), L (the number of classifiers in the ensemble), K (the number of subsets)

  for $i \leftarrow 1$ to L **do**

    Split the feature set into K subsets $F_{i,j}$ (for j=1 ...K)

    for $j \leftarrow 1$ to K **do**

      Let $X_{i,j}$ be the dataset for the features in $F_{i,j}$.

      Eliminate from $X_{i,j}$ a random subset of classes.

      Select a bootstrap sample from $X_{i,j}$ with 3/4 of the number of objects in $X_{i,j}$. The new set is named $X'_{i,j}$.

      Apply PCA to $X'_{i,j}$ to obtain the coefficients in matrix $C_{i,j}$.

    end for

    Arrange $C_{i,j}$ for $j = 1$ to K in a rotation matrix as in Equation 1.

    Construct a new matrix XR_i^a rearranging the columns of R_i in order to match the order of features in F.

  end for

  Build classifier C_i using (XR_i^a, Y) training set.

To build the training set for the classifier C_i, the feature set F is split into K disjoint subsets $F_{i,j}$ of size M. For every subset $X_{i,j}$, we randomly select a non-empty subset of classes and then draw a bootstrap sample using 75 % of the original sample. We then run PCA using only the features in $F_{i,j}$ and the selected subset of classes $X'_{i,j}$. The coefficients of the principal components $a_{i,j}^{(1)}, \ldots, a_{i,j}^{(M)}$ of size $M \times 1$ are used to generate rotation matrix R_i:

$$
R_i = \begin{bmatrix}
a_{i,1}^{(1)}, a_{i,1}^{(2)}, \ldots, a_{i,1}^{(M_1)} & [0] & \cdot\ \cdot\ \cdot & [0] \\
[0] & a_{i,2}^{(1)}, a_{i,2}^{(2)}, \ldots, a_{i,2}^{(M_2)} & \cdot\ \cdot\ \cdot & [0] \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
[0] & [0] & \cdot\ \cdot\ \cdot & a_{i,K}^{(1)}, a_{i,K}^{(2)}, \ldots, a_{i,K}^{(M_K)}
\end{bmatrix}
\tag{1}
$$

Finally, the columns of R_i are rearranged in order to make them correspond with the original features. This new matrix of size $N \times n$, named R_i^a, is used to build the classier C_i using (XR_i^a, Y) as the training set.

Once the set of base classifiers has been trained, for the classification of a new instance x, we compute the average confidence for each class across the L classifiers. Let $d_{i,j}(xR_i^a)$ be the probability assigned by the classifier C_I to the hypothesis that x comes from class ω_j. For each class $(j = 1, \ldots, c)$, we compute the confidence $\mu_j(x)$ as shown in Eq. 2. The class with largest confidence is assigned to x.

$$
\mu_j(x) = \frac{1}{L} \sum_{i=1}^{L} d_{i,j}(xR_i^a)
\tag{2}
$$

2.3 Feature Extraction

The input data to the system consists of continuous signals with high dimensionality (vectors of $1 \times M$ samples, where $M \gg 1$). Their representation via a set of numerical parameters is crucial for automated analysis (in our particular case, human identification). For this, a set of features is carefully extracted from each signal, aiming at preserving the relevant information of the input data. Therefore, the dimensionality of the data is reduced via the feature extraction, yielding an N dimensional vector $(N < M)$. Each component of the vector is termed a feature.

The ECG shows the electrical conduction of the heart. Electrical potentials generated by the polarization and depolarization of cardiac tissue through each beat are transformed into a waveform. In particular, each individual beat is represented in the ECG as the PQRST complex. P

and T waves represent atrial depolarization and repolarization, respectively. The depolarization of the right and left ventricles is represented by the QSR complex.

As done in another study [27], each PQRST complex is characterized by a wide set of features. Regarding amplitude, the absolute peak values of each forming wave ($\{\Lambda_P, \Lambda_Q, \Lambda_R, \Lambda_S, \Lambda_T\}$) and relative values to the R peak ($\{\Theta_{RP}, \Theta_{RQ}, \Theta_{RS}, \Theta_{RT}, \Theta_{PQ}, \Theta_{QS}, \Theta_{TS}\}$) are used. The time intervals between the peaks of waves P, Q, R, S, and T are also used. The employed intervals ($\Delta_{PQ}, \Delta_{PR}, \Delta_{QR}, \Delta_{QS}, \Delta_{QT}, \Delta_{RS}, \Delta_{RT}, \Delta_{ST}$) are shown in Fig. 3. Finally, three angles [28] ($\{\angle Q, \angle R, \angle S\}$) that constitute the angular displacement between the peaks of the PQRST complex are used.

The respiratory rate is a common physiological parameter used in hospitals to check the patient status. A nasal thermocouple sensor that assesses thermal airflow changes and nasal temperature air was employed in this study. Two features were extracted from this signal, namely the average value (Ψ_{Avg}) during a PQRST complex (hearth beat) and the instantaneous value at the R peak of the ECG signal (Ψ_R).

GSR is used to measure the physiological arousal of a subject since the sympathetic nervous system controls skin sweating. GSR sensors sense the electrical conductance of the skin. Higher values represent greater moisture on the skin. Two values were extracted to characterize this signal, namely the average value of the signal during a PQRST complex (Υ_{Avg}) and the instantaneous value at the R peak of the ECG signal (Υ_R).

Prior to the study of how ECG, airflow, and GSR signals can be used for human identification, we analyzed the correlation of the three signals through their features. The correlation matrix is displayed at the top of Fig. 4. We computed the p-values. Figure 4 (bottom) shows which correlations are significant at a confidence level of 95 %. It is clear that there is a strong correlation between ECG features. The ECG features are correlated with the airflow and GSR as expected, since human activity is reflected through the different body signals. Regarding airflow and
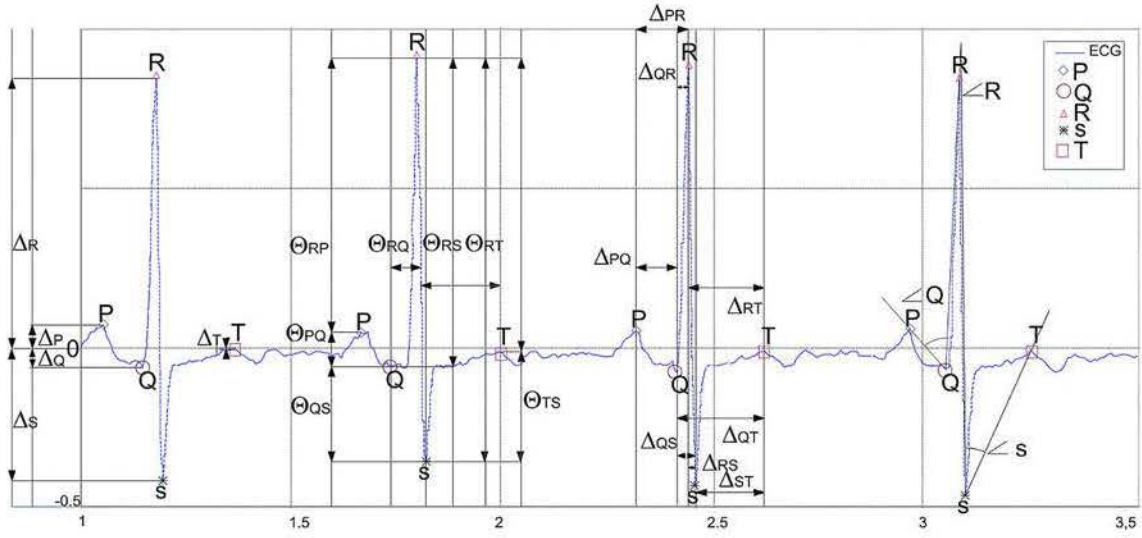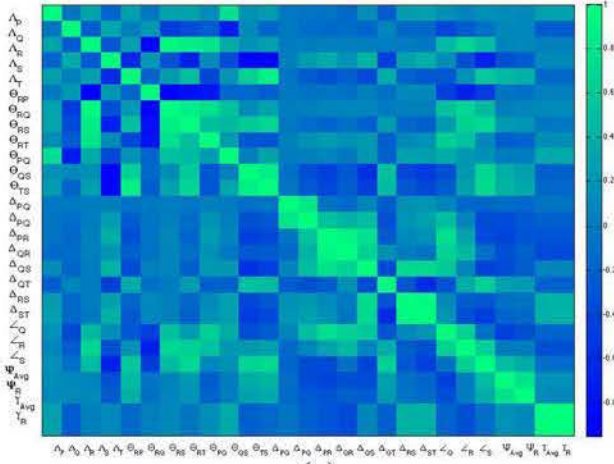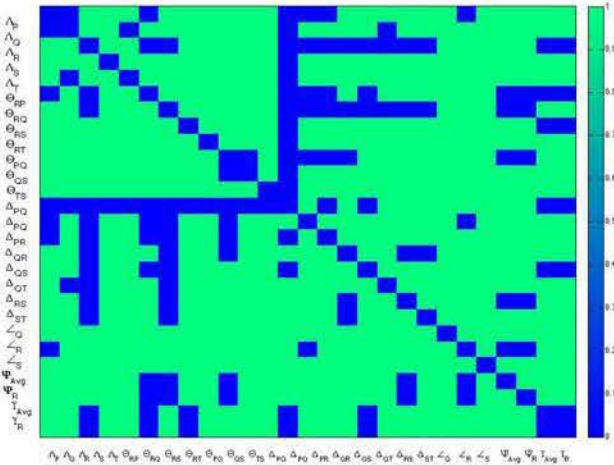
Fig. 3 ECG feature extraction

(a)

(b)

Fig. 4 Correlation between vital signals (ECG, airflow, and GSR).
a Correlation matrix and **b** significant correlation ($p < 0.01$)

GSR, both signals present a high correlation through the average value during a PQSRT complex. The average value and the instantaneous value at the R peak for the airflow signal are correlated. This is not the case with GSR since this signal is linked to the sympathetic nervous system and the signal fluctuates smoothly when the subject is at rest; there is no a peak in the GSR signal at each heart beat. That is, the GSR average value and the instantaneous value at the R peak are not correlated.

3 Results and Discussion

3.1 Performance System

Biosignals have been shown to an effective tool for human identification. ECG signals have been extensively studied for this purpose [6]. The present study tested the use of the ECG, airflow, and GSR signals for human identification. In particular, it was checked whether the hypothesis that the combined use of vital signals allows a human identification system to offer higher performance in comparison to the use of individual signals is true. If performance is measured in terms of the fraction of correctly classified instances (β), our hypothesis can be mathematically expressed as:

$$\begin{cases} H_0 : \beta_{\{ECG, \text{ Airflow and GSR}\}} > max\{\beta_{ECG}, \beta_{Airflow}, \beta_{GSR}\} \\ H_1 : \beta_{\{ECG, \text{ Airflow and GSR}\}} < max\{\beta_{ECG}, \beta_{Airflow}, \beta_{GSR}\} \end{cases}$$

$$(3)$$

The procedure followed for the analysis of the different signals is that given in Sect. 1. Firstly, the raw data are acquired through the sensors (in our case, several signals

Table 1 Independent biosignals

(a) Airflow—conf. matrix

	S1	S2	S3	S4	S5	S6
S1	**18**	0	0	5	5	0
S2	3	**11**	20	7	0	0
S3	0	13	**30**	1	0	0
S4	9	5	0	**19**	3	3
S5	7	0	0	1	**37**	0
S6	8	5	3	5	9	**6**

(b) Airflow—perform

	Sensitivity	Specificity	F-Measure
S1	0.643	0.868	0.493
S2	0.268	0.880	0.293
S3	0.682	0.878	0.619
S4	0.487	0.902	0.494
S5	0.822	0.910	0.925
S6	0.167	0.985	0.588
Weighted avg.	0.519	0.904	0.496

(c) GSR—conf. matrix

	S1	S2	S3	S4	S5	S6
S1	**22**	0	1	4	0	1
S2	0	**39**	1	0	1	0
S3	2	7	**17**	8	6	4
S4	11	0	1	**27**	0	0
S5	10	2	7	1	**23**	2
S6	0	0	0	0	0	**36**

(d) GSR—perform

	Sensitivity	Specificity	F-Measure
S1	0.786	0.888	0.603
S2	0.951	0.953	0.876
S3	0.386	0.947	0.479
S4	0.692	0.933	0.684
S5	0.511	0.963	0.613
S6	1.000	0.964	0.911
Weighted avg.	0.704	0.944	0.691

(e) ECG—conf. matrix

	S1	S2	S3	S4	S5	S6
S1	**27**	0	0	0	1	0
S2	0	**39**	1	0	1	0
S3	1	0	**41**	0	2	0
S4	0	0	0	**39**	0	0
S5	2	0	1	0	**42**	1
S6	0	0	1	0	1	**34**

(f) ECG—perform

	Sensitivity	Specificity	F-Measure
S1	0.964	0.985	0.931
S2	0.951	1.000	0.975
S3	0.932	0.989	0.943
S4	0.974	1.000	0.987
S5	0.933	0.968	0.903
S6	0.944	0.995	0.958
Weighted avg.	0.948	0.989	0.949

S1–S6 individuals who were subjected to biosignal recording; *Sensitivity* true positive rate; *Specificity* true negative rate; *F-Measure* test's accuracy

are simultaneously captured). After that, the signals are preprocessed (i.e., filtering, DC component elimination, smoothing, etc.). At this step, sets of representative features are extracted from the continuous signals (feature extraction is explained in detail in Sect. 2.3). Then, the feature vectors are passed through a classifier (training and testing phases). In the testing phase, the classifier (the model is built in the training phase) outputs a label corresponding to one of the existing subjects in the system.

For the classification algorithm, we selected the RF classifier due to its desirable diversity and high performance (see Sect. 2.2.2 for a detailed discussion on the algorithm). In particular, the attribute selection is done through PCA and the classification algorithm is C4.5 (the confidence factor and the number of instances per leaf are

set to 0.2 and 2.0, respectively). The dataset is split into two parts: 66 % for training and 34 % for testing.

Before testing our hypothesis H_0, the three signals were analyzed individually to check the suitability of each biosignal for identification purposes. Whereas ECG signals have been extensively studied, to the best of our knowledge, this is the first time that airflow and GSR are used for human identification. From the results shown in Table 1, it is evident that individually ECG outperforms airflow and GSR. Airflow offers poor results, only slightly better than chance. GSR could be used for authentication purpose, although the number of misclassified instances is moderately high.

Despite of the poor results with the airflow signal, this signal is correlated with the ECG, which is a truly effective

Table 2 Combined biosignals

(a) Airflow and GSR—conf. matrix

	S1	S2	S3	S4	S5	S6
S1	**17**	0	0	5	5	1
S2	0	**38**	3	0	0	0
S3	0	12	**27**	4	0	1
S4	4	0	0	**35**	0	0
S5	4	1	0	0	**38**	2
S6	0	0	0	0	0	**36**

(b) Airflow and GSR—perform

	Sensitivity	Specificity	F-Measure
S1	0.607	0.961	0.642
S2	0.927	0.932	0.826
S3	0.614	0.984	0.730
S4	0.897	0.954	0.843
S5	0.844	0.973	0.864
S6	1.000	0.980	0.947
Weighted avg.	0.820	0.964	0.815

(c) ECG and airflow—conf. matrix

	S1	S2	S3	S4	S5	S6
S1	**27**	0	0	0	1	0
S2	0	**41**	0	0	0	0
S3	0	0	**43**	0	0	1
S4	0	0	0	**39**	0	0
S5	1	0	0	0	**43**	1
S6	0	0	0	0	2	**34**

(d) ECG and Airflow—perform.

	Sensitivity	Specificity	F-Measure
S1	0.964	0.995	0.964
S2	1.000	1.000	1.000
S3	0.977	1.000	0.989
S4	1.000	1.000	1.000
S5	0.956	0.984	0.945
S6	0.944	0.990	0.944
Weighted avg.	0.974	0.995	0.974

(e) ECG and GSR—Conf. Matrix

	S1	S2	S3	S4	S5	S6
S1	**27**	0	0	0	1	0
S2	0	**41**	0	0	0	0
S3	1	0	**43**	0	0	0
S4	0	0	0	**39**	0	0
S5	0	0	1	0	**44**	0
S6	0	0	2	0	0	**36**

(f) ECG and GSR—perform

	Sensitivity	Specificity	F-Measure
S1	0.964	0.995	0.964
S2	1.000	1.000	1.000
S3	0.977	0.995	0.977
S4	1.000	1.000	1.000
S5	0.978	0.995	0.978
S6	1.000	1.000	1.000
Weighted avg.	0.987	0.997	0.987

(g) ECG, airflow and GSR—Conf. Matrix

	S1	S2	S3	S4	S5	S6
S1	**27**	0	0	0	1	0
S2	0	**41**	0	0	0	0
S3	0	0	**44**	0	0	0
S4	0	0	0	**39**	0	0
S5	0	0	0	0	**45**	0
S6	0	0	0	0	0	**36**

(h) ECG, airflow and GSR—perform

|	Sensitivity	Specificity	F-Measure
S1	0.964	1.000	0.982
S2	1.000	1.000	1.000
S3	1.000	1.000	1.000
S4	1.000	1.000	1.000
S5	1.000	0.995	0.989
S6	1.000	1.000	1.000
Weighted avg.	0.996	0.999	0.996

S1–S6 individuals who were subjected to biosignal recording; *Sensitivity* true positive rate; *Specificity* true negative rate; *F-Measure* test's accuracy

signal for the identification problem. Airflow and GSR are correlated through the average value during a PQRST complex. Motivated by this, we analyzed the results when both signals are combined. The number of correctly classified instances is higher that those obtained individually by each signal. In fact, these results are similar to those in the first works that used ECG signals [29] for human identification, and those in recent works based on EEG signals for mental task classification [30] (around 80 % of

instances correctly classified). The top of Table 2 shows the confusion and performance matricies.

Finally, we tested the combined use of the three biosignals. ECG, airflow, and GSR are correlated to each other and they are linked with the cardiovascular, respiratory, and sympathetic nervous systems, respectively. As show at the bottom of Table 2, the number of misclassified instances is tiny. For an overview, in Fig. 5 we summarize the performance results (i.e., corrected classified instances

Fig. 5 System performance
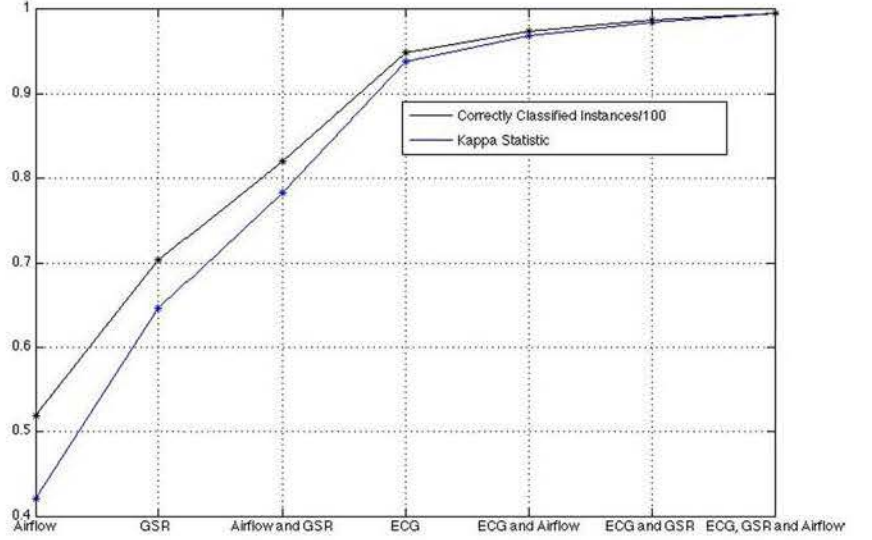
and kappa statistic) achieved with each of the studied combinations. From this, we accept hypothesis H_0 and conclude that the combined used of ECG, airflow, and GSR outperforms the individual use of these signals. Mathematically:

$$\beta_{\{ECG,\,Airflow\,and\,GRS\}} > \max\{\beta_{ECG}, \beta_{Airflow}, \beta_{GSR}\} \quad (4)$$

The true acceptance rate is almost 100 % (99.57 %) and the false acceptance rate is close to 0 % (0.1 %). The high value of the kappa statistic (0.9948) shows the strong relation between the labels and the used features. In other words, the features extracted from the three biosignals are reliable and effective for building a human identification system.

3.2 Evaluation

The proposed system was assessed from two perspectives. We evaluated the characteristics of the biometric system and tested the feasibility of supporting the identification system on a smartphone that acts as a personal server.

A biometric system should have the following seven characteristics [3]:

- *Universality* The biometric characteristic must be universally applicable. In the proposed solution, we use three vital signals that everyone posses. In fact, these signals imply that the individual is alive.
- *Uniqueness* The subjects within the target population can be unequivocally identified through the chosen characteristics. In this regard, we showed how the individual use of the three biosignals entails that certain samples are misclassified. Nevertheless, as shown in Table 2, the subjects are unequivocally identified; the

number of misclassified samples is almost 0. The samples of each patient contain data of several sessions in which the placement of the sensors is subject to small variations. Due to the high performance and the extremely low number of misclassified instances obtained (see Table 2; Fig. 5), there are no indications that the sensor placement variations have a relevant impact on system performance.

- *Permanence* The used characteristic should be invariant over time. In our particular case, the chosen biosignals are not fully invariant and there are slight changes during the subject's life. For instance, slight differences can been observed in ECG signals after a five-year period [31]. This does not imply a problem to our identification problem. It just means that the used classifier has to be updated (training phase execution and model building) after several years. Password-based systems (e.g., mail or cloud services) are more demanding, often requiring users to update their password at least one per year [32].
- *Collectability* This is linked with the acquisition devices employed. In our system, each signal is gathered through a low-cost sensor that is place over the skin of the subject. Therefore, signal acquisition is practical and affordable. In our particular case, the three sensors are connected to the Arduino board, which controls the process.
- *Acceptability* Users should not have objections with the use of the biometric characteristics. We cannot make any strong claims about this matter since the system has not been deployed. Nevertheless, from our experience with the dataset acquisition, we can state that users felt comfortable in this respect due to two main reasons: (1)

Table 3 Biosignal-based authentication proposals

System	Correctly classified instances
Proposed system	99.6 %
ECG [5]	86–100 % (single day data acquisition)
EEG [33]	72–80 %
Pulse-Response [3]	88–100 % (small data set)
EEG and ECG [11]	97.9 % (linear boundary)
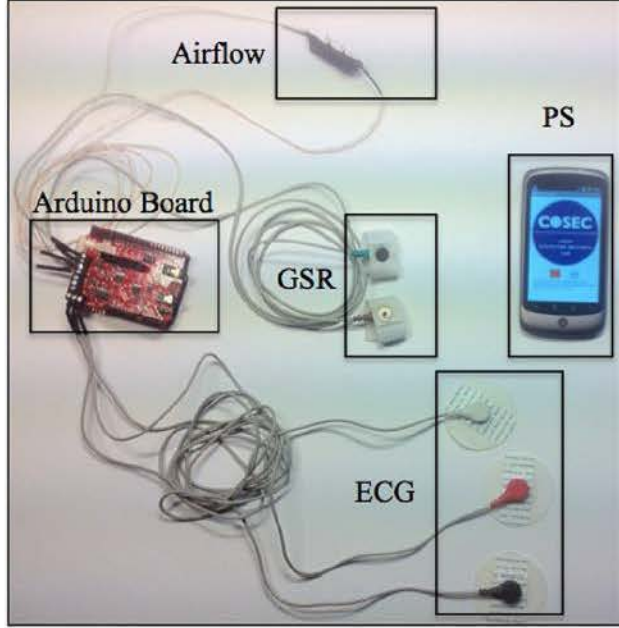
Fig. 6 Proof-of-concept identification system

the combined used of three biosignals generate trust with system users; and (2) the employed sensors are not invasive and are only placed over the skin of the subject.

- *Performance* This refers to the accuracy level of the system. The proposed system offers high performance: 99.57 % of instances were correctly classified and the precision was almost one (0.996), which translates to almost perfect sensitivity (i.e., true positive rate) and

specificity (true negative rate). For completeness, and to facilitate comparison with other approaches, the F-measure (the weighted harmonic mean of precision and recall) is also provided. Table 3 compares our system with other systems. In terms of performance, our system and those based on ECG and EEG offer performance close to 100 % [11]. This study showed how the combined use of ECG, airflow, and GSR improves the results obtained using these biosignals individually. This is the first time that GSR and airflow were tested for human identification.

- *Resistance to Circumvention* From a security point of view, this property is crucial. It means that it is difficult to fool the system or impersonate another user in the database. The ECG features are almost unforgeable [34] and the addition of the other two vital signals renders the proposed system more robust.

As a proof of concept, we tested the proposed biometric system on a Google Nexus One smartphone. The biosignals were measured via three sensors wired to the Arduino board. The Arduino transmits (via Bluetooth) the acquired signals to the Personal Server, which is an Android app designed for this purpose. To implement the classifiers at the core of the app, we used a stripped version of the Weka library for Android devices, which offers reasonably optimized implementations of the most common machine learning algorithms. Figure 6 shows all the devices involved in the prototype.

Apart from the system accuracy level, we tested the recognition speed and the power consumption of the classification task, which is the most demanding work [35]; transmission, preprocessing, and feature extraction were excluded in this calculation. We split the process into two well differentiated tasks. The first one, training, is only done for setting up the system prior to its utilization. This task is occasionally repeated, for instance, as a consequence of the aging of the biometric characteristics. The second one, identification (execution of the classifier; 1-to-many comparisons), is performed for testing the identity of

Table 4 Energy consumption in Joules (J) and execution time in seconds (s) and miliseconds (ms) for training and testing tasks, respectively, using RF. *A*, *G*, and *E* represent airflow, GSR, and ECG, respectively

	A	G	E	A&G	E&A&G
Time					
RF					
Train	22 s	19 s	102 s	34 s	113 s
Test	80 ms	56 ms	88 ms	70 ms	138 ms
Energy					
RF					
Train	6.835 J	5.530 J	38.51 J	11.115 J	38.85 J
Test	0.005 J	0.008 J	0.015 J	0.010 J	0.015 J

a user. Table 4 summarizes the results of our simulation for both tasks.

Energy consumption was measured by applying a battery of tests involving the tasks and algorithms mentioned above. Each test was an app containing a specific biometric system, such as the ECG or airflow and the GSR. The apps were loaded into the device sequentially, so only one execution was run at a time.

The device was previously instrumented with AppScope [36], an energy-metering framework based on monitoring the kernel activity for Android. AppScope collects usage information from the monitored device and estimates the consumption of each running application using an energy model given by DevScope [37]. AppScope provides the amount of energy consumed by an app in the form of several time series, each one associated with a component of the device (CPU, Wi-Fi, cellular, display, etc.). We restricted our measures to the CPU for computations, as the remaining components are not typically involved in a biometric system. Figure 7 shows an example of the energy consumed by the device during the execution of one of our tests.

The results show that both training and testing using biosignals are quite fast. Training a biometric model requires on average only a few seconds and, in the worst scenario, less than two minutes. During testing, the evaluation of a sample (instance derived from the three biosignals) requires less than 200 ms for all the cases. This means instantaneousness feedback for the user. The energy consumption is very small, especially considering that the frequency of the training is rather low and the testing task is extremely efficient; testing is thousand times less power demanding than training. For instance, the Google Nexus One battery has a capacity of 5.18 W h (1.8×10^4 J). Thus, the worst scenario (38.85 J; training and using the three biosignals) only represents 0.21 % of the overall battery capacity.
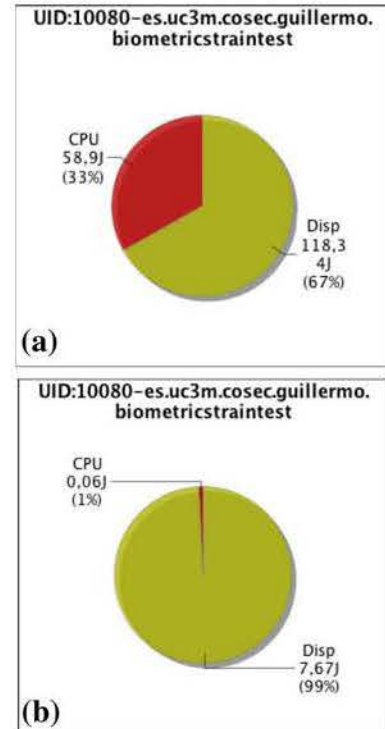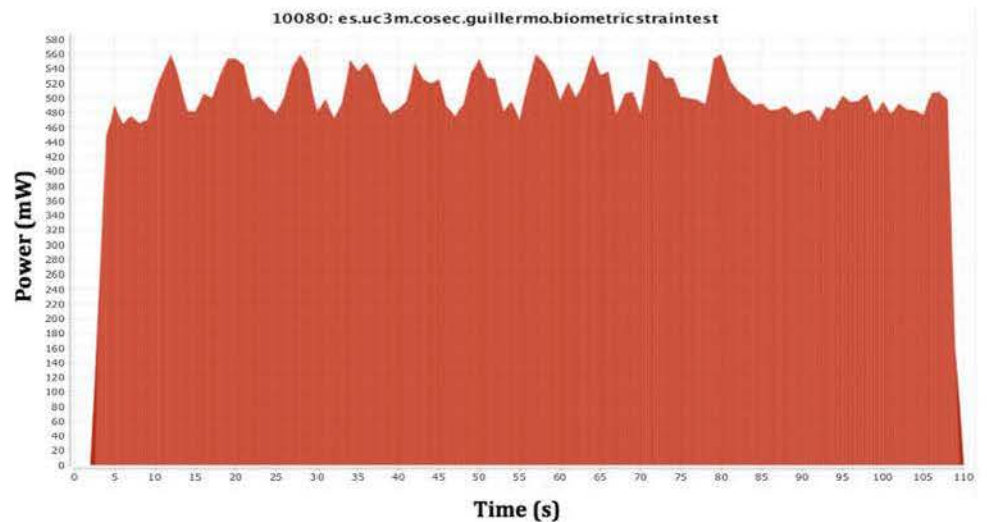
(a)

(b)

Fig. 8 Comparison between consumption of CPU (*red*) and display (*yellow*) during execution of RF with ECG, airflow, and GSR biosignals during a given time span. **a** Training for 14 s and **b** testing for 115 ms

Finally, for comparison, we measured the energy consumed by the display of the device during the same amount of time needed to train and test the human identification system. Figure 8 shows that both training and testing consume a small fraction of the overall energy. Training consumes about three times less than the display. Furthermore, the energy required to determine to which user a set of given biosignals belongs is negligible.

Fig. 7 Power consumed by proposed system during training task using RF with ECG, airflow, and GSR biosignals

4 Conclusion

This study determined whether uncommon vital parameters can be effective for biometrics and, in particular, for the identification problem. Apart of the well-known ECG signal, we explored the use of other vital parameters that can be gathered through non-invasive techniques. More precisely, our experiments used airflow and GSR in combination with the ECG. This is the first time that these parameters are used for security.

We tested the combined use of these three biosignals. Note that the combination of several signals does not guarantee improved performance when compared to that of the individual signals. Our experimental results suggest that the joint use of ECG, airflow, and GSR outperforms any other combination of these signals. The proposed system offers an accuracy of close to 100 % and seems effective for human identification. Many previous proposals did not checked the feasibility of the system in a real environment. In our case, as a proof of concept, we tested the proposed human identification system on a smartphone. The results in terms of processing speed and power consumption demonstrate its applicability in real-world applications.

As a result of the diversity of existing sensors, future work will test other biosignals for human identification and will address some of the limitations of our current research. For example, while the ECG and GSR sensors are small and easy to deploy, the airflow device is more cumbersome to wear, so results with easier-to-acquire biosignals would be welcome. A deeper study with a larger sample size should be also conducted to validate our results at a large scale. Furthermore, signals can be passively measured or else obtained as a response after a stimulus, which could have some security applications. Apart from biometric identification, biosignals can be useful for other security applications such as key establishment or random number generation.

References

1. Banaee, H., Ahmed, M. U., & Loutfi, A. (2013). Data mining for wearable sensors in health monitoring systems: A review of recent trends and challenges. *Sensors, 13*, 17472–17500.
2. Rostami, M., Juels, A., & Koushanfar, F. (2013). Heart-to-heart (H2H): authentication for implanted medical devices. In *Proceedings of the ACM SIGSAC conference on computer & communications security* (pp. 1099–1112).
3. Rasmussen, K. B., Roeschlin, M., Martinovic, I., & Tsudik, G. (2014). Authentication using pulse-response biometrics. In *Proceedings of the network and distributed system security symposium (NDSS)*.
4. Eng, A., & Wahsheh, L. (2013). Look into my eyes: A survey of biometric security. In *Proceedings of the tenth international conference on information technology: New generations* (pp. 422–427).
5. Odinaka, I., Po-Hsiang, L., Kaplan, A. D., O'Sullivan, J. A., Sirevaag, E. J., & Rohrbaugh, J. W. (2012). ECG biometric recognition: A comparative analysis. *IEEE Transactions on Information Forensics and Security, 7*, 1812–1824.
6. Khalifa, W., Salem, A., Roushdy, M., & Revett, K. (2012). A survey of EEG based user authentication schemes. In *Proceedings of the 8th international conference on informatics and systems* (pp. BIO-55–BIO-60).
7. Spachos, P., Gao, J., & Hatzinakos, D. (2011).Feasibility study of photoplethysmographic signals for biometric identification. In *Proceedings of the 17th international conference on digital signal processing (DSP)* (pp. 1–5).
8. Ichino, M., Sakano, H., & Komatsu, N. (2006). Multimodal biometrics of lip movements and voice using kernel fisher discriminant analysis. In *Proceedings of the international conference on control, automation, robotics and vision* (pp. 1–6).
9. Jani, R., & Agrawal, N. (2013). Proposed framework for enhancing security in fingerprint and finger-vein multimodal biometric recognition. In *Proceedings of the international conference on machine intelligence and research advancement* (pp. 440–444).
10. Revett, K., Deravi, F., & Sirlantzis, K. (2010). Biosignals for user authentication—Towards cognitive biometrics. In *Proceedings of the international conference on emerging security technologies (EST)* (pp. 71–76).
11. Riera, A., Dunne, S., Cester, I., & Ruffini, G. (2008). STAR-FAST: A wireless wearable EEG/ECG biometric system based on the ENOBIO Sensor. In *Proceedings of the international workshop on wearable micro and nanosystems for personalised health* (pp. 1–4).
12. Simon, B. P., & Eswaran, C. (1997). An ECG classifier de- signed using modified decision based neural networks. *Computer and Biomedical Research, 30*, 257–272.
13. Doorly, D. J., Taylor, D. J., & Schroter, R. C. (2008). Mechanics of airflow in the human nasal airways. *Respiratory Physiology & Neurobiology, 163*, 100–110.
14. Chaaban, M., & Corey, J. P. (2011). Assessing Nasal Air Flow. *Proceedings of the American Thoracic Society, 8*, 70–78.
15. Ogorevc, J., Gersak, G., Novak, D., & Drnovsek, J. (2013). Metrological evaluation of skin conductance measurements. *Measurement, 46*, 2993–3001.
16. Clifford, G. D., Azuaje, F., & McSharry, P. (2006). *Advanced methods and tools for ECG data analysis*. Norwood, MA: Artech House Inc.
17. Abellán, J., & Mantas, C. J. (2014). Improving experimental studies about ensembles of classifiers for bankruptcy prediction and credit scoring. *Expert Systems with Applications, 41*, 3825–3830.
18. Rahman, A., & Verma, B. (2013). Effect of ensemble classifier composition on offline cursive character recognition. *Information Processing and Management, 49*, 852–864.
19. Ozçift, A. (2011). Random forests ensemble classifier trained with data resampling strategy to improve cardiac arrhythmia diagnosis. *Computers in Biology and Medicine, 41*, 265–271.
20. Yin, X. C., Huang, K., Hao, H. W., Iqbal, K., & Wang, Z. B. (2014). A novel classifier ensemble method with sparsity and diversity. *Neurocomputing, 134*, 214–221.

21. Breiman, L. (1996). Bagging predictors. *Machine Learning, 24,* 123–140.

22. Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences, 55,* 119–139.

23. Cao, Y., Miao, Q. G., Liu, J. C., & Gao, L. (2013). Advance and prospects of adaboost algorithm. *Acta Automatica Sinica, 39,* 745–758.

24. Brown, G., Wyatt, J. L., Harris, R., & Yao, X. (2005). Diversity creation methods: A survey and categorization. *Information Fusion, 6,* 5–20.

25. Rokach, L. (2010). *Pattern Classification Using Ensemble Methods*. River Edge: World Scientific.

26. Rodriguez, J. J., Kuncheva, L. I., & Alonso, C. J. (2006). Rotation forest: A new classifier ensemble method. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 28,* 1619–1630.

27. Pal, S., & Mitra, M. (2011). ECG based biometric authentication: a novel data modeling approach. In *Proceedings of the international conference on image information processing (ICIIP)* (pp. 1–4).

28. Singh, Y., & Singh, S. (2012). Evaluation of electrocardiogram for biometric authentication. *Journal of Information Security, 3,* 39–48.

29. Shen, T. W., Tompkins, W. J., & Hu, Y. H. (2002). One-lead ECG for identity verification. In *Proceedings of the 24th annual conference and the annual fall meeting of the biomedical engineering society EMBS/BMES conference* (Vol. 1, pp. 62–63).

30. Anderson, C. W., & Bratman, J. A. (2008). Translating thoughts into actions by finding patterns in brainwaves. In *Proceedings of the fourteenth Yale workshop on adaptive and learning systems* (pp. 1–6).

31. Carmona, N., Rua-Seoane, J., Elorza, J., Saenz de Pipaon, E., Palacios, C., & Bragard, J. (2013). Aging of ECG characteristics over a five year period. In *Proceedings of the conference on computing in cardiology (CinC)* (pp. 1031–1034).

32. Schneier, B. (2010). Changing Passwords. https://www.schneier.com/blog/archives/2010/11/changing_passwo.html.

33. Singh, Y. N., Singh, S. K., & Ray, A. K. (2012). Bioelectrical Signals as emerging biometrics: Issues and challenges. *ISRN Signal Processing, 2012,* 1–13.

34. Tantawi, M. M., Revett, K., Tolba, M. F., & Salem, A. (2012). On the use of the electrocardiogram for biometric authentication. In *Proceedings of the 8th international conference on informatics and systems* (pp. BIO-48–BIO-54).

35. Suarez-Tangil, G., Tapiador, J. E., Peris-Lopez, P., & Pastrana, S. (2015). Power-aware anomaly detection in smartphones: An analysis of on-platform versus externalized operation. *Pervasive and Mobile Computing, 18,* 137–151.

36. Yoon, C., Kim, D., Jung, W., Kang, C., & Cha, H. (2012). Appscope: Application energy metering framework for android smartphone using kernel activity monitoring. In *Proceedings of the USENIX annual technical conference* (p. 36).

37. Jung, W., Kang, C., Yoon, C., Kim, D., & Cha, H. (2012). DevScope: A nonintrusive and online power analysis tool for smartphone hardware components. In *Proceedings of the eighth IEEE/ACM/IFIP international conference on Hardware/software codesign and system synthesis* (pp. 353–362).