Feature Selection for Non-stationary Data: Application to Human Recognition Using Medical Biometrics

Majid Komeili, Wael Louis, Narges Armanfard, Dimitrios Hatzinakos

Abstract—ECG and TEOAE are among the physiological signals that have attracted significant interest in biometric community due to their inherent robustness to replay and falsification attacks. However, they are time-dependent signals and this makes them hard to deal with in across-session human recognition scenario where only one session is available for enrollment. This paper presents a novel feature selection method to address this issue. It is based on an auxiliary dataset with multiple sessions where it selects a subset of features that are more persistent across different sessions. It uses local information in terms of sample margins while enforcing an across-session measure. This makes it a perfect fit for aforementioned biometric recognition problem. Comprehensive experiments on ECG and TEOAE variability due to time lapse and body posture are done. Performance of the proposed method is compared against seven state-of-the-art feature selection algorithms as well as another six approaches in the area of ECG and TEOAE biometric recognition. Experimental results demonstrate that the proposed method performs noticeably better than other algorithms.

Index Terms—Feature selection, Electrocardiogram, Transient Evoked Otoacoustic Emission, Biometric.

I. INTRODUCTION

BIOMETRIC systems have been deployed around the world and have been used extensively in the past decade. Automatic identity verification is becoming more appealing in many aspects such as financial transactions, tele-medicine and access control. Physiological characteristics such as fingerprint, iris and facial structure and behavioral characteristics such as voice, gait and signature have been used for identity recognition. However, the choice of right biometric depends on characteristics of application environment and usually is a trade-off between factors such as performance, ease of collection, user acceptability and deployment cost.

With the wide deployment of biometric systems, the potential of fooling or spoofing this technology is widely admitted. For example, fingerprints may be left behind whenever you touch a glass surface like phone screen and fake fingerprints can be recreated by some simple procedures using plastic molds and gelatin. Even iris images can be captured from few meters distance. Not to mention face images [1], [2] which can be captured from a longer distance. Voice trait

M. Komeili, W. Louis and D. Hatzinakos are with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario, Canada.
E-mail: {mkomeili, wlouis, dimitris}@ece.utoronto.ca
N. Armanfard is with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, Canada.
E-mail: armanfn@mcmaster.ca

is also vulnerable against replay attacks using pre-recorded voice playbacks. Such vulnerabilities have been motivating researchers towards novel biometric characteristics such as Electrocardiogram (ECG) and Transient Evoked Otoacoustic Emission (TEOAE) that are inherently robust to replay and falsification attacks.

ECG is related to variation of electrical activity of heart. TEOAE is an acoustic response generated from cochlea. It is related to inner ear structure and can be recorded using an earphone with built-in microphone. ECG and TEOAE have some advantages over conventional biometrics. ECG is a vital signal and presence of the ECG signal automatically ensures the liveness. Moreover, it is difficult to obtain someone's ECG without his or her permission. ECG is relatively an inexpensive technology because many mature data acquisition systems for collecting ECG signal is already developed in clinical applications. It is also a natural choice for human identification in medical care and tele-medicine application where ECG signal is collected primarily for diagnosis. Moreover, ECG is a continuous signal which allow a continuous authentication.

TEOAE is an acoustic response which is generated from within the inner ear after a low level transient click stimulus. TEOAE has the following key characteristics: It can be recorded non-invasively. Speaker and microphone can be easily integrated into a headset or earphone connected to a smart phone to be used for access control or personalization of service (see Fig. 1). TEOAE presences in almost every living individual (99%+) [3]. So, it's measurability is always guaranteed. Also, it might be useful in identification of newborn infants where performance of conventional methods based on face, fingerprint and iris degrade [4]–[6]. Clinically, TEOAE is simple, non-invasive yet effective in detecting hearing problems in newborn babies who are too young to cooperate in conventional hearing tests [7], and it has been extensively used in infant screening programs around the world [8].

The main difficulty in using such signals as biometric is that they are non-stationary. Through this paper by non-stationary we mean that signals varies over different recording sessions or postures. One of the most challenging scenarios is the case that there is only one recording session for enrollment while testing happens in a different session. In fact, in this case, signals' nonstationarity does not elicit during enrollment session and hence cannot be captured from only one recording session.

The proposed method aims to build "templates" from samples of only one enrollment session such that the templates are robust against signal variations over different sessions. Realiz-

ing this goal requires extra information in addition to what can be extracted from the enrollment session during "enrollment" phase. The proposed approach, during a "development" phase, produces this information from an "auxiliary" dataset consists of recordings of some general subjects different from the biometric system users where recordings from multiple sessions are available for each member of the auxiliary dataset. This information is produced in the context of feature selection. The process of selecting a subset of informative features to be used for model construction is known as feature selection. Our feature selection criteria selects features that are more persistent across different sessions. The desired persistent features are selected from the auxiliary dataset. These features are then used as a prior knowledge in building the users templates.

The proposed feature selection criteria is formulated as a logistic regression optimization problem that can be efficiently solved through gradient descent. It makes no assumption regarding the data distribution. The proposed method allows us to accommodate across session variations. None of the previous methods [9]–[21] have the advantage of handling across session variations and injecting a prior knowledge for building subjects templates that have only one enrollment session. We refer to the proposed algorithm as Multisession Feature Selection (MSFS).

Another key distinguishing characteristic of this paper is that our ECG signals are collected from thumbs using dry electrodes (see Fig. 1), contrasting with most of the previous studies that are based on signals collected from chest using adhesive gel electrodes (e.g. PTB and MIT databases [22]). Recording ECG from fingertips is much more realistic for biometric application compare to recording from chest area in which the user has to undress for electrode placement. Adhesive gel electrodes have a very good signal to noise ratio because of their low electrode/skin impedance. In comparison, dry electrodes have lower signal quality due to unstable electrode/skin interface. Therefore, signals recorded from fingertips using dry electrodes are noisier and harder to deal with [23]. Moreover, unlike the previous ECG datasets [24], [25], [26], [27], [12], [28], [29], our dataset allows us to consider challenges related to ECG variation across different postures. In this study, in addition to analyzing ECG signals recorded at different times (scheduled at least one week apart), we consider across posture variations. Furthermore, we demonstrate the effectiveness of the proposed method on our TEOAE dataset collected in separate recording sessions (scheduled at least one week apart). Note that all the previously existing TEOAE datasets (e.g. [30] and [31]) were recorded in the same session which is of little biometric value because such datasets do not allow to verify if the recognition algorithm does the job in a different session. In those datasets, both training and testing samples were recorded in the same session by either recording one signal and then dividing it into two parts (one for training and one for testing), or reinserting the earphone and capturing a second recording. An earlier version of this paper was appeared in [32] which is based on the idea of performing feature selection on an auxiliary dataset. However, the feature selection used in [32] treats features individually
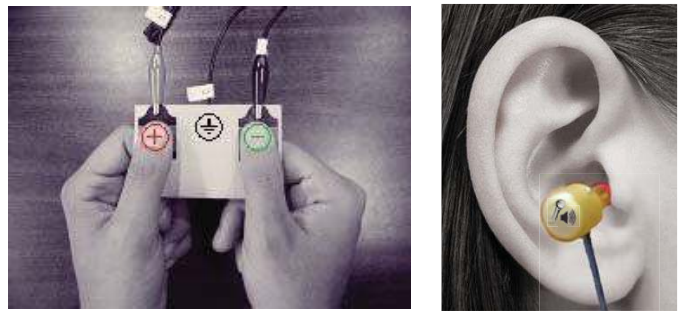

Fig. 1. Recording fingertip ECG and TEOAE signals.

and does not consider the interaction between features. In this paper, the idea of selecting features from a muti-session auxiliary dataset is realized in a more rigorous framework that does not suffer from aforementioned limitation. Further, in this paper we add comprehensive experiments on TEOAE and ECG (multi-posture) biometrics including comparison of the proposed feature selection algorithm with 7 existing feature selection methods and comparison of the whole system with 6 existing biometric recognition methods for ECG and TEOAE. In addition, a speed up version of the proposed method is presented that saves the CPU time by approximately 4 times.

The remaining portion of this paper is organized as follows: Section II briefly reviews recent works in the three areas of ECG and TEOAE biometrics and feature selection. An overview of the proposed approach is presented in Section III. Details of the proposed feature selection algorithm are presented in Section IV. In section V, an approach to speed up the proposed algorithm is presented. Experimental results are presented in section VI and section VII concludes the paper.

II. RELATED WORK

A. Electrocardiogram

Prior works on ECG biometric can be categorized as either fiducial points dependent or independent. The former is based on some characteristic points (known as fiducial points) such as peaks, slope, radius of curvature, onset and end of P, R and T waves. Detection of such characteristic points is however a challenging process due to noise. Therefore, following [33], we do not consider fiducial dependent approaches [25], [34]–[43]. The latter category considers ECG signal as a set of heartbeats or just a time series without segmenting it to heartbeats [9]–[21]. For example, in [11] autocorrelation features are extracted from overlapping windows and passed to linear discriminant analysis (LDA) [44] and nearest neighborhood for classification. Odinaka [16] employs STFT features and log-likelihood ratio for classification. It is basically a generative approach which models each subject with an axis aligned Gaussian distribution. Wang [21] uses sparse coefficients of an over-complete dictionary followed by max-pooling process to build templates and use nearest neighborhood for classification.

However, most of these methods are based on signals collected from chest area (e.g. PTB and MIT databases [22]) or lower rib cage [16] and only a few works have been done

based on ECG signals collected from fingertips [12], [45]. This paper is based on ECG signals collected from fingertips which is far more convenient for the user because it eliminates the need for the user to undress for electrode placement (See Fig. 1). More importantly, none of the previous works provide a theoretical development for handling the "single session enrollment, across session testing" scenario which is focus of this paper.

B. Transient evoked otoacoustic emission

The first study on feasibility of using TEOAE for human recognition was done by Swabey et al. [46] which uses power spectrum features and Euclidean distance for classification. In another work [30] by the same group, maximum likelihood estimation (MLE) is used to estimate probability density function (PDF) of inter-class and intra-class distance distributions. In [47] pair-wised distances in cepstrum-transform domain is used. A summary of above works is presented in [31]. The main drawback of these works is that due to difficulty of recording in separate sessions, both testing and training recordings were collected in the same session or with only removing and refitting the earphone which limits its application for biometric purpose.

To address this limitation, the first multisession dataset of TEOAE signals was collected under a setup suitable for biometric evaluation [48] and bivariate empirical mode decomposition (BEMD) were employed for signal decomposition. Later, [49] used continuous wavelet transform (CWT) for signal representation. The main drawback of these methods is that they rely on expert's knowledge to engineer features which were often chosen for their empirical performance, rather than a specific theoretical property related to across session variations.

C. Feature selection

Feature selections approaches can be divided into two categories: supervised and unsupervised. Supervised approaches use label information to guide the selection process whereas unsupervised approaches aims to describe structure of data in some feature space in the absence of label information [50], [51]. In this study, we focus on the supervised approaches because in the biometric recognition application class labels are provided.

Supervised feature selection methods can be roughly categorized into filters, wrappers and embedded methods. Wrapper methods, such as sequential forward selection and sequential backward selection [52], select a feature subset (among all possible feature subsets) that gives the best performance [53]. A model for each candidate feature subset is trained and then tested and its performance is used to guide the feature selection process. Various searching techniques have been proposed in the literature. For example, in [54] particle swarm optimization is used as a search technique. However, wrappers are computationally very intensive specially if the chosen model is complex. The other drawback of such algorithms is the high risk of over-fitting because they are tuned to a specific model. While wrapper methods use a model, i.e. a specific

classification algorithm, for evaluation of a feature subset, filter methods use a criterion function. The third category, embedded methods, embeds feature selection in classification [55], [56]. However, such integration does not fit in our proposed approach because in the proposed approach feature selection and classification have to be performed on separate set of data. Data for feature selection have to be collected from multiple sessions but data for classifier training have to be collected from a single session to comply with "single session enrollment, across-session test" scenario. More details on the proposed method will be provided in the next section.

On the other side, filter methods are relatively fast and do not suffer from aforementioned limitations. Different criterion functions have been proposed in the literature. Some feature selection approaches are based on mutual information and usually use some heuristics to handle the relevance-redundancy trade-off [57]–[61]. In [61] a minimum redundancy maximum relevance criterion is used which is based on mutual information. In [62], a greedy algorithm is presented that adds features to the current subset according to the selected and remaining features relevances with the labels. There are some other approaches which are based on evolutionary algorithms. In [63] feature selection is realized by differential evolution to reduce the search space. Some other feature selection algorithms are based on a maximum margin criteria [64], [65], [66], [67], [68], [69]. These methods are sample-based where the "margin" is defined as the difference between distance to the nearest same class sample (near-hit) and the nearest sample from opposite classes (near-miss). Relief [66] selects features that are statistically relevant to the target. The drawback of this method is that nearest miss and nearest hit samples are computed in the original space. This was addressed in Simba algorithm [65] through reevaluation of the margins. However, its objective function is not convex and suffers from many local minima. Later, in [64] a local margin-based feature selection approach was presented in which uses a local learning approach to decompose a complex nonlinear problem into a set of locally linear problems within a large margin framework. Nevertheless, none of the aforementioned feature selection methods are designed to accommodate multisession data. Indeed, these are unable to benefit from the session label information provided in addition to the class labels in our problem.

III. OVERVIEW

In this study we consider a "single session enrollment, across session test" scenario in which enrollment (classifier design) have to be done using one session and testing have to be done in another session. To alleviate the undesired effects of signal variation between enrollment and test sessions, in the proposed approach, classifiers are designed using a subset of more persistent features. Such features are selected by looking at a separate dataset with multiple sessions referred as auxiliary dataset. Therefore, each sample in the auxiliary dataset has two labels: class label and session label. There is no overlap between subjects in auxiliary dataset and actual biometric system users that are to be enrolled in the biometric

system. Details of the proposed feature selection method is presented in the next section.

IV. PROPOSED FEATURE SELECTION METHOD

Let $\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)}, s^{(i)})\}_{i=1}^{M} \subset \mathbb{R}^J \times \mathcal{Y} \times \mathcal{S}$ be an auxiliary dataset consists of M samples belong to N subjects and for each subject S different sessions are available. $\mathbf{x}^{(i)}$ is a J dimensional feature vector, $\mathcal{Y} = \{Y_1, \ldots, Y_N\}$ is the set of all class labels and $y^{(i)} \in \mathcal{Y}$ is the class label of the i-th sample $\mathbf{x}^{(i)}$. $\mathcal{S} = \{S_1, \ldots, S_S\}$ is the set of all session labels and $s^{(i)} \in \mathcal{S}$ is the session label of $x^{(i)}$. For example, $\mathcal{S}$ may indicate recording sessions on S different weeks, i.e., week 1, week 2, ..., week S.

Assume that $\mathbf{x}^{(i)}$ is excluded from $\mathcal{D}$ and we want to classify it while the corresponded subject, i.e. $y(i)$, is enrolled using samples of only one session. However, there are S possible options for enrollment session and we want $\mathbf{x}^{(i)}$ to be correctly classified using either of them. Let adopt a classification rule such that given session $s \in \mathcal{S}$ as enrollment, $\mathbf{x}^{(i)}$ is correctly classified if and only if its corresponding margin $\ell_s^{(i)}$ is positive where $\ell_s^{(i)}$ is defined as:

$$\ell_s^{(i)} = \mathbf{1}^\mathsf{T} \mathbf{d}_s^{(i)} \tag{1}$$

where $\mathbf{d}_s^{(i)} = \mathbf{d}_{NM}^{(i)} - \mathbf{d}_{NH_s}^{(i)}$ and $(\cdot)^\mathsf{T}$ is transpose operator. $\mathbf{d}_{NM}^{(i)}$ and $\mathbf{d}_{NH_s}^{(i)}$ are absolute difference vectors determined as follows:

$$\mathbf{d}_{NM}^{(i)} = \left| \mathbf{x}^{(i)} - \mathrm{NM}(\mathbf{x}^{(i)}) \right|,$$
$$\mathbf{d}_{NH_s}^{(i)} = \left| \mathbf{x}^{(i)} - \mathrm{NH}_s(\mathbf{x}^{(i)}) \right|, \tag{2}$$

where $|.|$ is element-wise absolute operator, $\mathrm{NM}(\mathbf{x}^{(i)})$ is the nearest neighbor of $\mathbf{x}^{(i)}$ with a different class label (nearest miss) and $\mathrm{NH}_s(\mathbf{x}^{(i)})$ is the nearest neighbor of $\mathbf{x}^{(i)}$ from session s with the same class label as $\mathbf{x}^{(i)}$ (nearest hit). Note that equations (1) and (2) imply l_1 distance. While other options are possible (e.g. Euclidean distance by substituting $|.|$ with $(.)^2$ in (2)), throughout this paper we use l_1 distance. A basic definition of margin as the difference between distance to nearest miss and nearest hit was previously used in [66], [65] and [64], but it does not deal with the multiple session issue – Hence this is not appropriate for our problem.

A positive margin implies that the query sample $\mathbf{x}^{(i)}$ will be correctly classified. We realize the goal of having less misclassification error through weighting features by a nonnegative vector $\mathbf{f}$ and measuring margins in the corresponding weighted space specified by $\mathbf{f}$. A larger margin provides even more room for the query sample $\mathbf{x}^{(i)}$ to wander in the weighted space and still be correctly classified, i.e. a better generalization on unseen data. Therefore, we seek to maximize the margins in the "weighted" space. Equation (1) defines the margin in the original space. Similarly, in the weighted space, margin of sample $\mathbf{x}^{(i)}$ with respect to session s can be computed as follows:

$$\ell_s^{(i)}(\mathbf{f}) = \mathbf{f}^\mathsf{T} \mathbf{d}_s^{(i)}. \tag{3}$$

In fact $\mathbf{f}$ could be considered as a weighting vector that assigns higher weights to the relevant features where relevant features are those that results in larger margins. At the outset,

$\mathbf{f}$ is unknown. Therefore, determining nearest miss and nearest hit in the weighted space defined by $\mathbf{f}$ is a challenging issue. To overcome this issue, we use an iterative approach for computing $\mathbf{f}$, where at each iteration $\mathbf{f}$ is determined based on the distances in the weighted space defined at the previous iteration. Following [64], the problem of learning feature weights is formulated as a logistic regression problem and solved via gradient decent. However, determining $\mathrm{NM}(\mathbf{x}^{(i)})$ and $\mathrm{NH}_s(\mathbf{x}^{(i)})$ in the presence of irrelevant features, may not be accurate. To address this issue, margin is estimated as the expectation of $\ell_s^{(i)}(\mathbf{f})$ over all possible candidates for $\mathrm{NM}(\mathbf{x}^{(i)})$ and $\mathrm{NH}_s(\mathbf{x}^{(i)})$ as follows:

$$\bar{\ell}_s^{(i)}(\mathbf{f}) = \mathbf{f}^\mathsf{T} \bar{\mathbf{d}}_s^{(i)} \tag{4}$$

where $\bar{\mathbf{d}}_s^{(i)} = \bar{\mathbf{d}}_{NM}^{(i)} - \bar{\mathbf{d}}_{NH_s}^{(i)}$; and $\bar{\mathbf{d}}_{NM}^{(i)}$ and $\bar{\mathbf{d}}_{NH_s}^{(i)}$ are:

$$\bar{\mathbf{d}}_{NM}^{(i)} = \mathbf{D}_{NM}^{(i)} \mathbf{p}_{NM}^{(i)\mathsf{T}}, \tag{5}$$

$$\bar{\mathbf{d}}_{NH_s}^{(i)} = \mathbf{D}_{NH_s}^{(i)} \mathbf{p}_{NH_s}^{(i)\mathsf{T}}. \tag{6}$$

$\mathbf{D}_{NM}^{(i)}$ and $\mathbf{D}_{NH_s}^{(i)}$ are matrices whose columns are absolute difference vectors with respect to $\mathbf{x}^{(i)}$:

$$\mathbf{D}_{NM}^{(i)} = \left(\left| \mathbf{x}^{(i)} - \mathbf{x}^{(\mathcal{M}^i(1))} \right|, \quad \ldots, \quad \left| \mathbf{x}^{(i)} - \mathbf{x}^{(\mathcal{M}^i(n))} \right| \right), \tag{7}$$

$$\mathbf{D}_{NH_s}^{(i)} = \left(\left| \mathbf{x}^{(i)} - \mathbf{x}^{(\mathcal{H}_s^i(1))} \right|, \quad \ldots, \quad \left| \mathbf{x}^{(i)} - \mathbf{x}^{(\mathcal{H}_s^i(m))} \right| \right). \tag{8}$$

$\mathcal{M}^i$ and $\mathcal{H}_s^i$ with cardinality of n and m denote set of all possible candidates for $\mathrm{NM}(\mathbf{x}^{(i)})$ and $\mathrm{NH}_s(\mathbf{x}^{(i)})$ respectively and are defined as:

$$\mathcal{M}^i = \left\{ j \in \{1, \ldots, M\} \mid y^{(j)} \neq y^{(i)} \right\}, \tag{9}$$

$$\mathcal{H}_s^i = \left\{ j \in \{1, \ldots, M\} \mid y^{(j)} = y^{(i)}, s^{(j)} = s, j \neq i \right\}. \tag{10}$$

$\mathbf{p}_{NM}^{(i)}$ ($\mathbf{p}_{NH_s}^{(i)}$) in eq. (5) (eq. (6)) is an n-dimensional (m-dimensional) row vector indicates the probability of samples in $\mathcal{M}^i$ ($\mathcal{H}_s^i$) being $\mathrm{NM}(\mathbf{x}^{(i)})$ ($\mathrm{NH}_s(\mathbf{x}^{(i)})$). Within the weighted space, samples situated closer to $\mathbf{x}^{(i)}$ are more probable to be the nearest sample. So the probabilities are determined as follows:

$$\mathbf{p}_{NM}^{(i)} = \exp\left(\frac{-\mathbf{f}^\mathsf{T} \mathbf{D}_{NM}^{(i)}}{\sigma} \right), \tag{11}$$

$$\mathbf{p}_{NH_s}^{(i)} = \exp\left(\frac{-\mathbf{f}^\mathsf{T} \mathbf{D}_{NH_s}^{(i)}}{\sigma} \right), \tag{12}$$

where σ is a user settable parameter. $\mathbf{p}_{NM}^{(i)}$ and $\mathbf{p}_{NH_s}^{(i)}$ are then normalized to sum to one to be the probabilities utilized in (5) and (6).

Having $M \times S$ margins of the form $\bar{\ell}_s^{(i)}(\mathbf{f}) = \mathbf{f}^\mathsf{T} \bar{\mathbf{d}}_s^{(i)}$, $s = 1, \ldots, S$, $i = 1, \ldots, M$, it is desired to maximize all margins. Considering a logistic regression formulation, the optimization problem can be expressed as follows:

$$\max_{\mathbf{f}} \sum_{i=1}^{M} \sum_{s=1}^{S} \mathcal{G}\left(\mathbf{f}^\mathsf{T} \bar{\mathbf{d}}_s^{(i)} \right), \quad \text{s.t.} \quad \mathbf{f} \geq 0, \tag{13}$$

Input: $\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)}, s^{(i)})\}_{i=1}^{M}, \sigma, \lambda$
Output: $\{\mathbf{f}\}$
1   Initialization: Set $\mathbf{f} = (1, \ldots, 1)^{\mathsf{T}}$;
2   **repeat**
3     $\mathbf{f}_{\text{prev.}} = \mathbf{f}$;
4   **for** $i \leftarrow 1$ **to** M **do**
5   Compute $\mathbf{D}_{NM}^{(i)}$ as in (7);
6   Compute $\mathbf{p}_{NM}^{(i)}$ using $\mathbf{f}_{prev.}$ as in (11);
7   Compute $\bar{\mathbf{d}}_{NM}^{(i)}$ as in (5);
8   **for** $s \leftarrow 1$ **to** S **do**
9   Compute $\mathbf{D}_{NH_s}^{(i)}$ as in (8);
10   Compute $\mathbf{p}_{NH_s}^{(i)}$ using $\mathbf{f}_{prev.}$ as in (12);
11   Compute $\bar{\mathbf{d}}_{NH_s}^{(i)}$ as in (6);
12   Compute $\bar{\mathbf{d}}_s^{(i)} = \bar{\mathbf{d}}_{NM}^{(i)} - \bar{\mathbf{d}}_{NH_s}^{(i)}$;
13   **end**
14   **end**
15   Compute $\mathbf{a}$ through solving (17);
16   Set $f_j = a_j^2$, $1 \leq j \leq J$
17 **until** $\|\mathbf{f} - \mathbf{f}_{prev.}\|_2 < \varepsilon$;

Algorithm 1: Pseudo code of the proposed MSFS feature selection algorithm.

where $\mathcal{G}(\cdot)$ is a logistic function.

$$\mathcal{G}(b) = \log \left(\frac{1}{1 + \exp(-b)} \right) \qquad (14)$$

$\mathcal{G}$ is a strictly increasing function, therefore maximizing $\mathcal{G}\left(\mathbf{f}^{\mathsf{T}}\bar{\mathbf{d}}_s^{(i)}\right)$ indeed implies maximizing $\mathbf{f}^{\mathsf{T}}\bar{\mathbf{d}}_s^{(i)}$. However, $\mathcal{G}$ is useful because it can take an input that can vary from negative to positive infinity whereas the output always ranges between 0 and 1. Then (13) can be simplified as follows:

$$\min_{\mathbf{f}} \sum_{i=1}^{M} \sum_{s=1}^{S} \log\left(1 + \exp\left(-\mathbf{f}^{\mathsf{T}}\bar{\mathbf{d}}_s^{(i)}\right)\right), \quad \text{s.t.} \quad \mathbf{f} \geq 0, \quad (15)$$

A regularization term may also be added that has been shown to be useful in many applications to control over-fitting [70]. Hence, the problem can be formulated as follows:

$$\min_{\mathbf{f}} \sum_{i=1}^{M} \sum_{s=1}^{S} \log\left(1 + \exp\left(-\mathbf{f}^{\mathsf{T}}\bar{\mathbf{d}}_s^{(i)}\right)\right) + \lambda\|\mathbf{f}\|_1, \text{ s.t. } \mathbf{f} \geq 0,$$
$$(16)$$

where λ is a user settable parameter. Equation (16) is a constrained convex optimization problem with respect to $\mathbf{f}$. We then replace the vector $\mathbf{f}$ with a new vector $\mathbf{a}$ such that $f_j = a_j^2$, $1 \leq j \leq J$ and reformulate the problem in an unconstrained form as follows:

$$\min_{\mathbf{a}} \sum_{i=1}^{M} \sum_{s=1}^{S} \log\left(1 + \exp\left(-\sum_{j=1}^{J} a_j^2 \bar{d}_s^{(i)}(j)\right)\right) + \lambda\|\mathbf{a}\|_2^2$$
$$(17)$$

where $\bar{d}_s^{(i)}(j)$ is the j-th element of vector $\bar{\mathbf{d}}_s^{(i)}$. Therefore (17) can be solved via gradient decent with step size of τ and

updated as follows:

$$\mathbf{a} \leftarrow \mathbf{a} - \tau \mathbf{\Delta}$$

$$\mathbf{\Delta} = \mathbf{a} \otimes \left(\lambda\mathbf{1} - \sum_{i=1}^{M} \sum_{s=1}^{S} \frac{\exp\left(-\sum_j a_j^2 \bar{d}_s^{(i)}(j)\right)}{1 + \exp\left(-\sum_j a_j^2 \bar{d}_s^{(i)}(j)\right)} \bar{\mathbf{d}}_s^{(i)} \right)$$
$$(18)$$

where $\otimes$ is Hadamard operator. We utilized fminbnd function in MATLAB to determine τ using the line search algorithm. Features weight vector is initialized to $\mathbf{1}$, so that all features have the same weight at the beginning and then will be updated using (18) until a stopping criteria is satisfied. In this study, the algorithm stops when the difference between the weights in two successive iterations is less than a threshold ε which is set to 0.01. Pseudo code of the proposed MSFS feature selection method is presented in Algorithm 1.

The proposed method considers the "single session enrollment, across-session test" scenario and indeed, our objective function is directly related to the classification error under this scenario. This is an interesting characteristic of the proposed method that makes it a perfect choice for human recognition using medical biometrics such as ECG and TEOAE which are known to be non-stationary across different sessions compare to many other conventional biometrics. Therefore it is superior to all previously existing feature selection algorithms in the literature.

V. SPEEDING UP USING RANDOM PROJECTION

The most time consuming part of the proposed algorithm is computing distances to determine the nearest miss and nearest hit for every sample. In practice, about 2% of the CPU time is spent on solving (15) using gradient descent and the rest of CPU time is spend mainly on computing NM and NH_s which requires computing distance between $\mathbf{x}^{(i)}$ and all samples in $\mathcal{M}^i$ and $\mathcal{H}_s^i$, $s = 1, ..., S$ which is problematic if both M (number of samples) and J (dimensionality) have high values. However, majority of samples have very small probability of being the nearest neighbor because they are too far from $\mathbf{x}^{(i)}$. To speed up the process of determining NM and NH_s, we use Random Projection to project the data onto a lower K-dimensional space and determine probabilities of being the nearest neighbor in the lower dimensional space and pick a subset of top samples, i.e. high probable samples. Only selected samples are then used to determine nearest neighbor in the J-dimensional space scaled by $\mathbf{f}$ which results in a faster processing time.

The idea of random projection is based on the Johnson-Lindenstrauss lemma [71] which states that if points in a vector space are projected onto a randomly selected subspace of suitably high dimension, then the distances between points are approximately preserved. Computation of random projection is simple and involves only forming a random $J \times K$ matrix $\mathbf{R}$ ($K \ll J$) which its elements are often generated from Gaussian distribution. Later, Achlioptas [72] showed that $\mathbf{R}$ can be generated by a much simpler distribution such that its elements can be drawn from $\{\sqrt{3}, 0, -\sqrt{3}\}$ with probability of $\frac{1}{6}$, $\frac{2}{3}$ and $\frac{1}{6}$ respectively.

Samples in the weighted space are projected to a K-dimensional space:

$$\hat{\mathbf{X}}_{\mathbf{f}} = \mathbf{R}^\mathsf{T}\mathbf{X}_{\mathbf{f}}, \tag{19}$$

where $\mathbf{X}_{\mathbf{f}} = \left(\mathbf{x}_{\mathbf{f}}^{(1)}, \ldots, \mathbf{x}_{\mathbf{f}}^{(M)}\right)$ is data matrix in the weighted space, i.e. $\mathbf{x}_{\mathbf{f}}^{(j)} = \mathbf{f} \otimes \mathbf{x}^{(j)}$, $j = 1, \ldots, M$. Computing matrices of absolute differences are much faster in the K-dimensional space because $K \ll J$:

$$\hat{\mathbf{D}}_{NM}^{(i)} = \left(\left|\hat{\mathbf{x}}_{\mathbf{f}}^{(i)} - \hat{\mathbf{x}}_{\mathbf{f}}^{(\mathcal{M}^i(1))}\right|, \ldots, \left|\hat{\mathbf{x}}_{\mathbf{f}}^{(i)} - \hat{\mathbf{x}}_{\mathbf{f}}^{(\mathcal{M}^i(n))}\right|\right) \tag{20}$$

$$\hat{\mathbf{D}}_{NH_s}^{(i)} = \left(\left|\hat{\mathbf{x}}_{\mathbf{f}}^{(i)} - \hat{\mathbf{x}}_{\mathbf{f}}^{(\mathcal{H}_s^i(1))}\right|, \ldots, \left|\hat{\mathbf{x}}_{\mathbf{f}}^{(i)} - \hat{\mathbf{x}}_{\mathbf{f}}^{(\mathcal{H}_s^i(m))}\right|\right) \tag{21}$$

where $\hat{\mathbf{x}}_{\mathbf{f}}^{(j)}$ is projection of $\mathbf{x}_{\mathbf{f}}^{(j)}$ using $\mathbf{R}$. Then probabilities of being the nearest neighbor can be determined as follows:

$$\hat{\mathbf{p}}_{NM}^{(i)} = \exp\left(\frac{-\mathbf{1}^\mathsf{T}\hat{\mathbf{D}}_{NM}^{(i)}}{\hat{\sigma}}\right) \tag{22}$$

$$\hat{\mathbf{p}}_{NH_s}^{(i)} = \exp\left(\frac{-\mathbf{1}^\mathsf{T}\hat{\mathbf{D}}_{NH_s}^{(i)}}{\hat{\sigma}}\right) \tag{23}$$

where $\hat{\sigma}$ has the same role as σ and is defined as $\sigma \times \left(\frac{K}{J}\right)$ to account for change in dimensionality. $\hat{\mathbf{p}}_{NM}^{(i)}$ and $\hat{\mathbf{p}}_{NH_s}^{(i)}$ are then normalized to sum to one. We form shortened lists of candidates for $NM(\mathbf{x}^{(i)})$ and $NH_s(\mathbf{x}^{(i)})$ denoted by $\hat{\mathcal{M}}^i$ and $\hat{\mathcal{H}}_s^i$ where $\hat{\mathcal{M}}^i \subset \mathcal{M}^i$ and $\hat{\mathcal{H}}_s^i \subset \mathcal{H}_s^i$. To this end, we sort members of $\hat{\mathcal{M}}^i$ and $\hat{\mathcal{H}}_s^i$ in descending order according to their corresponding probabilities $\hat{\mathbf{p}}_{NM}^{(i)}$ and $\hat{\mathbf{p}}_{NH_s}^{(i)}$. Then we pick the top t members of $\mathcal{M}^i$ ($\mathcal{H}_s^i$) to form $\hat{\mathcal{M}}^i$ ($\hat{\mathcal{H}}_s^i$). t is defined such that the summation of their corresponding probabilities are not less than a user-settable parameter γ. In all our experiments, γ is fixed at the value of 0.95. We use $\hat{\mathcal{M}}^i$ ($\hat{\mathcal{H}}_s^i$) instead of $\mathcal{M}^i$ in (7) ($\mathcal{H}_s^i$ in (8)) to determine $\mathbf{D}_{NM}^{(i)}$ ($\mathbf{D}_{NH_s}^{(i)}$) which indeed has less number of columns and is faster to compute. Then $\mathbf{p}_{NM}^{(i)}$ and $\mathbf{p}_{NH_s}^{(i)}$ are computed using (11) and (12) based on the resulting $\mathbf{D}_{NM}^{(i)}$ and $\mathbf{D}_{NH_s}^{(i)}$. Other steps are the same as the MSFS. Pseudo code of the proposed algorithm based on random projection (hereafter MSFS-RP) is presented in Algorithm 2.

The time spent on projection in (19) is very small and can be ignored because in each iteration $\hat{\mathbf{X}}(\mathbf{f})$ is computed only once. The time spent on computing $\mathbf{D}_{NM}^{(i)}$, $\mathbf{D}_{NH_s}^{(i)}$, $\hat{\mathbf{D}}_{NM}^{(i)}$ and $\hat{\mathbf{D}}_{NH_s}^{(i)}$ are dominant because they must be computed for every sample. Assume that the total number of samples in $\hat{\mathcal{M}}^i$ and $\hat{\mathcal{H}}_s^i$ is L. In MSFS algorithm, computing $\mathbf{D}_{NM}^{(i)}$ and $\mathbf{D}_{NH_s}^{(i)}$ is of the order $O(MJ)$. In MSFS-RP algorithm, $\hat{\mathbf{D}}_{NM}^{(i)}$ and $\hat{\mathbf{D}}_{NH_s}^{(i)}$ are computed in the lower dimensional space which is of the order $O(MK)$. In addition, $\mathbf{D}_{NM}^{(i)}$ and $\mathbf{D}_{NH_s}^{(i)}$ are to be computed on $\hat{\mathcal{M}}^i$ and $\hat{\mathcal{H}}_s^i$ which is of the order $O(LJ)$. Therefore the complexity is of the order $O(MK+LJ)$. As will be discussed in section VI-G, $MK \ll LJ$ and this complexity for MSFS-RP is approximately $O(LJ)$. Therefore, the CPU time of MSFS-RP is approximately $\frac{L}{M}$ of the CPU time of MSFS.

Input: $\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)}, s^{(i)})\}_{i=1}^M$, σ, λ, K, γ
Output: $\{\mathbf{f}\}$

1   Initialization: Set $\mathbf{f} = (1, \ldots, 1)^\mathsf{T}$;
2   **repeat**
3     $\mathbf{f}_{\text{prev.}} = \mathbf{f}$;
4   Compute $\hat{\mathbf{X}}_{\mathbf{f}}$ using $\mathbf{f}_{prev.}$ as in (19);
5   **for** $i \leftarrow 1$ **to** M **do**
6   Compute $\hat{\mathbf{D}}_{NM}^{(i)}$ as in (20);
7   Compute $\hat{\mathbf{p}}_{NM}^{(i)}$ as in (22);
8   Form $\hat{\mathcal{M}}^i \subset \mathcal{M}^i$ according to $\hat{\mathbf{p}}_{NM}^{(i)}$;
9   Compute $\mathbf{D}_{NM}^{(i)}$ as in (7) using $\hat{\mathcal{M}}^i$;
10   Compute $\mathbf{p}_{NM}^{(i)}$ using $\mathbf{f}_{prev.}$ as in (11);
11   Compute $\bar{\mathbf{d}}_{NM}^{(i)}$ as in (5);
12   **for** $s \leftarrow 1$ **to** S **do**
13   Compute $\hat{\mathbf{D}}_{NH_s}^{(i)}$ as in (21);
14   Compute $\hat{\mathbf{p}}_{NH_s}^{(i)}$ as in (23);
15   Form $\hat{\mathcal{H}}_s^i \subset \mathcal{H}_s^i$ using $\hat{\mathbf{p}}_{NH_s}^{(i)}$;
16   Compute $\mathbf{D}_{NH_s}^{(i)}$ as in (8) using $\hat{\mathcal{H}}_s^i$;
17   Compute $\mathbf{p}_{NH_s}^{(i)}$ using $\mathbf{f}_{prev.}$ as in (12);
18   Compute $\bar{\mathbf{d}}_{NH_s}^{(i)}$ as in (6);
19   Compute $\bar{\mathbf{d}}_s^{(i)} = \bar{\mathbf{d}}_{NM}^{(i)} - \bar{\mathbf{d}}_{NH_s}^{(i)}$;
20   **end**
21   **end**
22   Compute $\mathbf{a}$ through solving (17);
23   Set $f_j = a_j^2$, $1 \leq j \leq J$
24 **until** $\|\mathbf{f} - \mathbf{f}_{prev.}\|_2 < \varepsilon$;

Algorithm 2: Pseudo code of the proposed feature selection algorithm using random projection (MSFS-RP).

VI. EXPERIMENTAL RESULTS

In this section, we perform several experiments on synthetic and real-world datasets to demonstrate the effectiveness of the proposed method. The first two experiments evaluate the effectiveness of the proposed method to cope with variations of ECG and TEOAE across different sessions where our feature selection algorithm aims to select a subset of features which are more steady across sessions. In the third experiment, we use the proposed feature selection algorithm to select a subset of features which are more steady across different body postures. We consider the case that subjects are enrolled in sitting (standing) posture but tested in standing (sitting) posture. The experiments are designed with respect to two objectives:

- Comparing performance of the proposed method against state-of-the-art methods in the area of ECG biometric recognition (AC/LDA [11], Odinaka [16] and Wang [21]) and TEOAE biometric recognition (CWT/LDA [49], BEMD [48] and PDF Estimation [30]).
- Comparing performance of the proposed method against existing feature selection algorithms in the literature. To this end, performance of the proposed method is compared against seven state-of-the-art feature selection algorithms including locally linear feature selection (LLFS) [64], iterative search margin based algorithm (Simba) [65], differential evolution feature selection (DEFS) [63], joint mutual information (JMI) [57], interaction capping
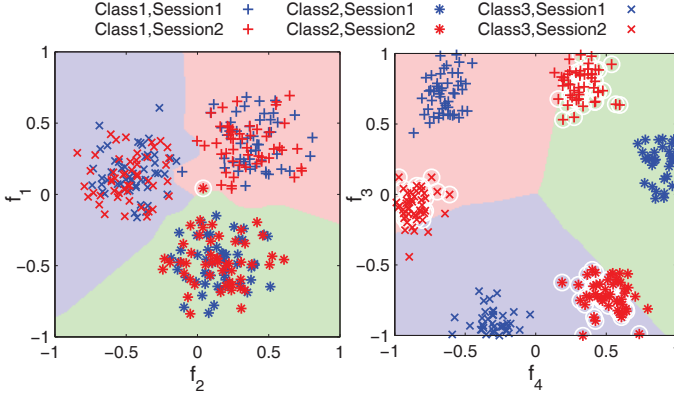
Fig. 2. Visualization of across-session training and testing on the synthetic dataset consists of four features, i.e., f_1 to f_4 with 5000 additional irrelevant features. Blue and red samples were used for training and testing respectively. Classification result using f_1 and f_2 is shown on the left and result for f_3 and f_4 is shown on the right. Classification boundaries are shown with different background colors.

(ICAP) [58], Markov random field (MRF) [73] and conditional infomax feature extraction (CIFE) [59].

Note that this paper has supplementary materials available at http://ieeexplore.ieee.org that include extra experiments about demonstrating how margin improves over successive iterations and also the effect of number of sessions used in feature selection.

A. Experimental Setup

1) Synthetic Dataset– Illustration of Multisession Concept: We use a synthetic dataset to illustrate the concept of feature selection for multisession data and the misclassification that may happen when enrollment and testing are from different sessions (as described in the beginning of section IV). This dataset consists of samples of 3 subjects and for each subject 2 sessions are available where each session includes 50 samples. Illustration of this dataset in the space defined by its four relevant features $\{f_1, f_2\}$ and $\{f_3, f_4\}$ is shown in Fig. 2 where two different sessions are shown in red and blue. In addition to the four relevant features, to have a more challenging problem, dataset is contaminated with 5000 irrelevant features drawn randomly from a Gaussian distribution with zero mean and unit variance. One may say that $\{f_1, f_2\}$ are better than $\{f_3, f_4\}$ because in the space defined by f_1 and f_2 samples of different sessions overlap but in the case of $\{f_3, f_4\}$ such an overlap does not exist. Moreover, if we take the first session (blue) as enrollment and the second session (red) for testing, it can be seen that in the case of using $\{f_1, f_2\}$, almost all samples will be correctly classified. But, using $\{f_3, f_4\}$ results in misclassification of many samples. Misclassified samples are marked by a circle around them. We will later use this "toy" example to demonstrate the effectiveness of the proposed feature selection algorithm to select a subset of features which are more steady across sessions (f_1 and f_2 in this case).

2) Real-world Datasets: Performance of the proposed method is also demonstrated using three experiments on real-world datasets. In the first experiment, we use ECG database collected in our lab (BioSec) in University of Toronto [74].

There are 82 subjects that have 2 or more ECG recordings in sitting posture. 46 out of 82 subjects have exactly 5 sessions. Follow-up sessions are collected over a six months period. Unlike many previous ECG databases, signals were collected from fingertips as shown in Fig. 1. We use aforementioned 46 subjects as auxiliary dataset and the remaining 36 subjects with 2-4 recording sessions (on average 4.28 sessions) are used for enrollment and testing. Hence, there is no overlap between these two sets.

The purpose of having an auxiliary dataset is to select a subset of more persistent features. Therefore, it is important that the auxiliary dataset have as many sessions as possible. According to the above decomposition scheme, 5 sessions per subject in auxiliary dataset is available. It also divides the available 82 subjects into two approximately equal sets of 46 and 36 subjects, so that the size of both sets are in the range of the size of multisession datasets reported in literature [24], [25], [26], [27], [12], [28], [29].

In the second experiment, we use TEOAE dataset collected in our lab (BioSec) in University of Toronto [74]. Unlike previous TEOAE datasets collected primarily for diagnostic purpose (e.g. [30] and [31]), our dataset is collected solely for biometric purpose. Therefore, optimal test conditions like a clinical setup in a sound proof room is avoided and signals were collected in a regular office where there were people talking and entering or leaving the office. Signals from 54 subjects are collected for two sessions scheduled at least one week apart. Following [75], 30 subjects are randomly chosen as auxiliary dataset and the remaining subjects are used for enrollment and testing.

In the third experiment, we use the ECG database collected in our lab (BioSec) in University of Toronto [74]. There are 79 subjects that have recordings in both sitting and standing postures. Subjects in the auxiliary dataset are the same as those in the first experiment. All the remaining subjects (33 subjects) are used for enrollment and testing.

3) Performance measure: In general a biometric system can be operated in two modes: Identification and verification. In our experiments, we consider verification mode of operation. It is a one-to-one matching process. In addition to a query sample, user provides an identity claim and the system aims to answer the following question: is the user who she/he claims to be? The biometric system accepts or rejects a claimed identity based on matching the query sample with template of the claimed identity.

Two types of error have to be considered: false rejection and false acceptance. The former occurs when a genuine identity request is mistakenly rejected by the system and the latter occurs when an intruder is mistakenly accepted by the system. Performance of a biometric system is often assessed based on false acceptance rate (FAR) and false rejection rate (FRR) defined as follows.

$$FRR = \frac{n_{FN}}{n_G} \tag{24}$$

$$FAR = \frac{n_{FP}}{n_I} \tag{25}$$

where n_G and n_I are total number of genuine and impostor attempts, respectively. n_{FP} is the number of falsely authen-

ticated attempts and n_{FN} is the number of falsely rejected genuine attempts. A widely-used performance measure is equal error rate (EER) which is the operating point where FAR and FRR are equal. Considering the "single session enrollment, across-session test" scenario described earlier, we pick a session for enrollment and use the rest of the available sessions for testing. That is, testing session can be any session in the past or future. For simplicity assume that there are P subjects and for each subject S sessions are available. For each of S sessions, we train P SVM classifiers in one-vs-all configuration. That gives a total of $P \times S$ templates. To simulate the positive trials (genuine attempts), each template is tested against $S - 1$ opposite sessions (same class), hence the term "across-session testing". This gives a total of $P \times S \times (S-1)$ positive trials. To simulate the negative trials (intruder attempts), each template is tested against $P-1$ other subjects (opposite class) in $S - 1$ opposite sessions. That is $(P-1) \times (S-1)$ intruder attempts against each template. This gives a total of $P \times S \times (P-1) \times (S-1)$ negative trials. Note that we refer to the number of subjects by P as oppose to N to highlight that EER is computed using a separate dataset that does not overlap the auxiliary dataset. Auxiliary dataset is dedicated to feature selection task.

We use linear SVM classifier because it has been successfully employed for ECG biometric recognition in recent works [76]–[78]. For a fair comparison, auxiliary datasets, enrollment and testing sets are common for all algorithms. Algorithms were implemented in MATLAB and executed on a desktop with an Intel core i7-3770 CPU and 16GB RAM. MATLAB's default settings for linear SVM classifier is used. More specifically, the penalty parameter is set to its default value i.e. 1. The number of selected features ranges from 1 to 1000 with a step size of 50. For each algorithm the appropriate number of selected features were determined using a bootstrapping algorithm [64] on the auxiliary dataset. For this purpose, each algorithm is run 10 times where at each run two subsets of equal size were randomly selected from the auxiliary dataset. One subset is used for feature selection and the other one is used for validation. The appropriate value for the number of features is the one that gives the minimum error on average over 10 runs. We use it for the final feature selection over entire auxiliary dataset.

4) Preprocessing: For ECG signals, following [37] and [11], a fourth order band-pass Butterworth filter between 0.5-40 Hz was employed for preprocessing. All signals are visually inspected before filtering. Each recording is segmented to its heartbeats using the method in [79] and [80] and aligned with respect to R peaks. For each session, Euclidean distance between segments and median is computed and segments whose distance are bigger than a threshold are considered as outlier and removed. Normalization is done by converting each feature to its z-score, i.e., we subtract the mean and divide by standard deviation so that each feature has zero mean and unit variance. Regarding the TEOAE signals, considering that Vivosonic Integrity System [81] used for signal recording has a built-in noise cancellation, no further preprocessing were needed. We only converted features to their z-score as described above.

5) Features' Description: There has been a variety of representations suggested in the literature to represent medical signals. For example, STFT has been used in [16] for ECG recognition. Similarly, CWT has been used in [19], Autocorrelation has been used in [11] and some statistical quantities from different frequency bands has been used in [82]. To have a comprehensive feature pool and consider the potential advantages of each of these features, we form a pool containing all features and select a subset of features using a feature selection algorithm.

Features of the feature pool are as follows: Continuous wavelet transform (CWT) with Daubechies 5 as mother wavelet is computed on 1 second window centered at R peak location. STFT with Hamming window of length 16 with step size of 13 is also computed on 1 second window centered at R peak location. Autocorrelation is computed on windows of length 6 seconds with 80 lags. We also considered six frequency bands: 8-13Hz, 13-18Hz, 18-25Hz, 25-30Hz, 30-35Hz, 35-50Hz and for each band, mean of power, standard deviation of power, maximum amplitude, standard deviation of amplitude, kurtosis and skewness were computed. Maximum, standard deviation, kurtosis and skewness are also computed from the signal itself. The signal amplitude, i.e. a one second window centered around R peak, was also considered. Aforementioned feature vectors were concatenated to form the final feature vector of the length 7278.

For TEOAE signals, we formed a similar feature pool with the addition of cepstrum features suggested in [47] for TEOAE which results in a feature vector of the length 7522. We only considered recordings from left ear which has been shown to be more challenging (compare to right ear) for biometric recognition [49] due to the ear asymmetries [83].

B. Experiments on across-session ECG recognition

In this section, performance of the proposed feature selection method is demonstrated on ECG signals when enrollment and testing are from different sessions. We use feature selection to find a subset of more steady features across different sessions. Length of each session is 2-5 minutes but we randomly selected 16 heartbeats to represent each session. This is a more challenging problem compare to the previous study in [45] which uses the whole session. The interested reader is referred to [16] and [33] for further information on effect of number of heartbeats. In [16] various number of heartbeats, i.e. 8, 16, 32, 64, 128 and 256, are investigated. According to their results significant improvement (%2) is achieved by increasing the number of heartbeats from 8 to 16; but further increasing it to 32 only slightly improves the EER by %0.2. Higher number of heartbeats has similar trend – e.g., increasing the number of heartbeats from 32 to 256 only improves EER by %0.5 that is not worth the extra heartbeats utilized. Furthermore, in [33], 8, 16 and 32 heartbeats are considered. According to their results EER is improved by %1 when the number of heartbeats is increased from 8 to 16; but further increasing the number of heartbeats does not improve the EER. Therefore, we choose 16 heartbeats. For each heartbeat, a window of the length one second centered
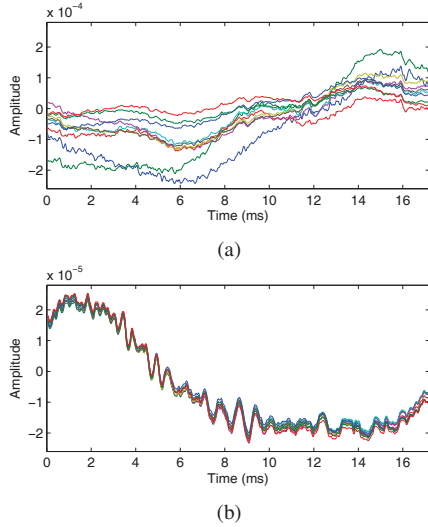
Fig. 3. Stabilization of TEOAE signal over time. (a) The first 10 samples. (b) The last 10 samples.

at R peak is used for feature extraction. Considering that segments may have some overlap, the effective length of each session in average is about 11 seconds. Each experiment is repeated 5 times and the average EER and standard deviations are reported. We compare the performance of the proposed method against three existing methods for ECG biometric recognition: AC/LDA [11], Odinaka [16] and Wang [21]. Table I shows EER and standard deviation for these comparison methods. It can be seen that the proposed MSFS method has significantly better performance than the other methods. We have carefully implemented these methods because they are not publicly available. In AC/LDA method, the number of autocorrelation lags is set to 30 and window length is set to 6 seconds as suggested in [11]. The Wang method has two parameters, dictionary size and length of segments. These are set to 2000 and 64 respectively as suggested in [21].

TABLE I. Across-session ECG experiment: EER (in percent) and standard deviation (in percent) for different methods are compared. Standard deviations are presented in parentheses.

AC/LDA[11]	Odinaka [16]	Wang [21]	MSFS
14.4(1.1)	18.4(0.7)	20.3(1)	**6.9(0.3)**

C. Experiments on across-session TEOAE recognition

In this section, generalization of the proposed method on other medical biometrics is demonstrated by applying it on TEOAE biometric. Similar to across-session ECG experiment, feature selection is used to find a subset of more steady features across different sessions. Duration of each session varies among the subjects and depends on how fast the response is stabilized for that subject. Vivosonic Integrity system [81] stops recording whenever Whole Wave Reproducibility (WWR) measure exceeds 90% or saturated at a specific level. The average length of each recording is about 39 seconds. In average, there are 122 samples for the left ear and 113 samples for right ear per session. Length of each sample is 17.2ms. Note that the actual raw samples of the

TABLE II. Across-session TEOAE experiment: EER (in percent) and standard deviation (in percent) for different methods are compared. Standard deviations are presented in parentheses.

CWT/LDA [49]	BEMD[48]	PDF Est. [30]	MSFS
25.2(1.8)	22.6(1)	28.4(1.7)	**3.3(1.2)**

length 17.2ms are not available as output. Samples undergo an internally implemented noise cancellation process which implies averaging over blocks of raw samples to produce output samples. Stimulus (click) interval is set to 21.12ms and as a result of internal noise cancellation process, the amount of time required to produce one sample at the output is about 338ms. Fig. 3 shows the first and last 10 samples for one subject. Note that responses are less noisier as the time proceeds due to the aforementioned noise cancellation. Unlike [49] that has used the last 10 samples of each session (requires the entire session), we randomly pick 10 samples from the first quarter of each session which in average reduce the length of each session to about 10 seconds. Each experiment is repeated 5 times and the average EER and standard deviations are reported. We compare the performance of the proposed method against three state-of-the-art methods in the area of TEOAE biometric recognition including CWT/LDA [49], BEMD [48] and PDF Estimation [30]. Results are presented in Table II which demonstrate significant improvement compare to other methods. We have carefully implemented these methods because they are not publicly available. In CWT/LDA method, Daubechies 5 is used as mother wavelet and CWT scale is set to 7 as suggested in [49]. In BEMD method, following [49] the second most refined level of intrinsic mode function (IMF) is used because it provides the best results.

D. Experiments on across-posture ECG recognition

In this section, we demonstrate the generalization of the proposed method by applying it to signals collected in different postures. In the previous experiments, feature selection was used to cope with long-term variation of ECG and TEOAE over different sessions. However, such variations can also happen in short-term due to different body postures. Here, we use feature selection to cope with variation of ECG due to different body postures. We only consider sitting and standing postures because this is the most challenging case as indicated in the previous study in [45]. Similar to section VI-B, 16 heartbeats were randomly selected to represent each session. The proposed method is compared against three state-of-the-art methods and results are presented in Table III. These results demonstrate the effectiveness of the proposed method in handling ECG variation due to different postures.

TABLE III. Across-posture ECG experiment: EER (in percent) and standard deviation (in percent) for different methods are compared. Standard deviations are presented in parentheses.

AC/LDA [11]	Odinaka[16]	Wang [21]	MSFS
12.2(1.4)	13.7(1.1)	18.3(1.6)	**3.7(0.9)**

E. Comparison with prior art in feature selection

The results presented in Table I, II and III shows significant improvement in the "single session enrollment, across session
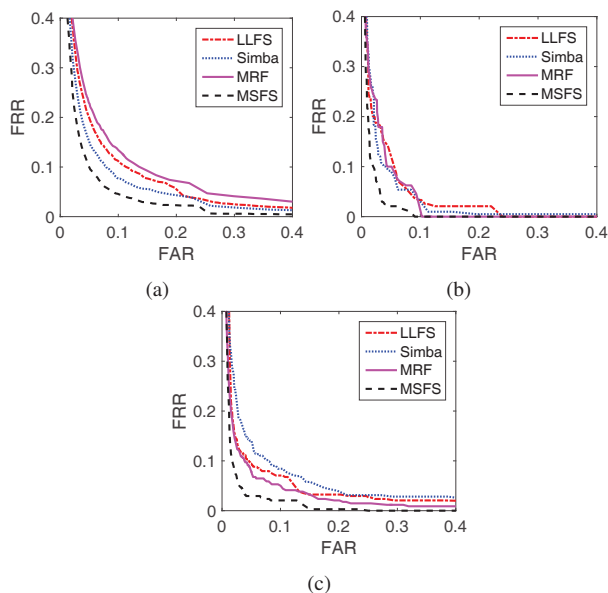
(a)

(b)

(c)

Fig. 4. ROC curves for three experiments. (a) Across-session ECG. (b)Across-session TEOAE. (c)Across-posture ECG.

Fig. 5. Sensitivity of weights to parameter λ.

Fig. 6. Sensitivity of weights to parameter σ.

test" scenario. This demonstrates that the idea of utilizing an auxiliary multisession dataset is indeed effective in dealing with nonstationarity across sessions. Since this idea is realized through a novel features selection algorithm, the reader may be interested to compare MSFS algorithm with existing feature selection algorithms in the literature. To address this issue, performance of the proposed feature selection algorithm is also compared with seven feature selection algorithms in the literature and the results for all three experiments are presented in Table IV. The best result at each experiment is in bold. The last row is the average EER computed over all three experiments (i.e., row 1 to 3). This indicates that the proposed algorithm performs significantly better than the other comparison algorithms. For reference, EER for SVM without feature selection is also presented in Table IV. The code for our comparison feature selection algorithms are all available on the respective author's websites and their parameters are set to their default values. To have a fair comparison, parameters of the proposed method are also set to default values and are discussed in Section VI-F. In addition, receiver operating characteristic (ROC) curves for comparison algorithms which exhibited the best performance as well as the proposed MSFS method are shown in Fig. 4. The number of selected features for these methods (i.e. LLFS, Simba, MRF and MSFS) in across-session ECG experiment are respectively 50, 50, 200

and 50. For across sessions TEOAE experiment, these are 250, 350, 250 and 250 and for across posture ECG experiment, these are 150, 200, 300 and 100 respectively.

F. Effect of parameters σ and λ

The proposed feature selection algorithm has two parameters: σ and λ. Parameter σ controls the local behavior of the proposed feature selection algorithm. A large value of σ increases the effect of far samples in estimating the nearest neighbor. The second parameter, λ, controls robustness against over-fitting. Fig. 5 and 6 show the feature weights computed by the proposed feature selection algorithm on the synthetic dataset for different values of σ and λ. It can be seen that the feature weights do not change significantly over a relatively wide range of values for σ and λ. The more stable features f_1 and f_2 indeed have greater weights compare to f_3 and f_4 that indeed suffer from across session variation (see Fig. 2). These results also assert that the proposed algorithm assigns very small weights to irrelevant features $\{f_5, \ldots, f_{5004}\}$. Generally these parameters can be determined using cross validation. However, in our experiments on real world datasets we set these parameters to 2 and 6 respectively – i.e., default values and they are fixed during all experiments.

G. CPU time and speeding up using Random Projection

In this section, CPU time of the proposed algorithm is investigated. Fig. 7 shows the CPU time for different number

TABLE IV. Comparison of different feature selection algorithms: EER (in percent) and standard deviation (in percent) are reported for three experiments. Standard deviations are presented in parentheses. First and second rows are corresponded to across-session ECG and TEOAE experiments, respectively. Third row is corresponded to across-posture ECG experiment. Last row shows the average EER over the three experiments. Results without feature selection (No FS) is also appended.

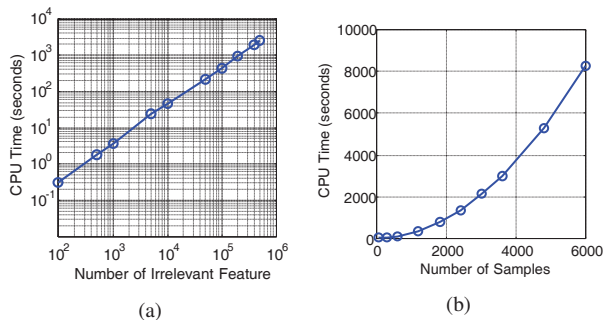| Experiment	LLFS [64]	Simba [65]	DEFS [63]	JMI [57]	ICAP [58]	MRF [73]	CIFE [59]	No FS (no feature selection)	MSFS
ECG (Across-session)	10.6(0.8)	8.8(0.6)	10.1(0.5)	13.2(0.6)	12.6(0.2)	11.5(1)	12.4(0.2)	13.5(1)	**6.9(0.3)**
TEOAE (Across-session)	6.6(1.7)	6(1.7)	7.5(1.9)	8.9(1.4)	7.8(1.2)	6.8(1.9)	7.9(1.6)	8.8(1.3)	**3.3(1.2)**
ECG (Across-Posture)	7.9(1.3)	9(1.2)	8.3(1.1)	8.3(0.9)	7.9(0.8)	6.5(1.1)	7.3(1.2)	9.9(1.1)	**3.7(0.9)**
Average	8.37	7.93	8.63	10.13	9.43	8.27	9.20	10.73	**4.63**

Fig. 7. The CPU time (in seconds) taken by the proposed algorithm to perform feature selection on the synthetic dataset where the parameters σ and λ are set to 1. (a) CPU time vs number of features. The figure shows linear complexity of the proposed method with respect to feature dimensionality. Number of samples is set to 300. (b) CPU time vs number of samples. The figure shows the proposed algorithm has quadratic time with respect to number of samples. Dimensionality is set to 5000.

TABLE VI. CPU time (in seconds) for our comparison feature selection algorithms. The symbols (*) and (†) denote algorithms are respectively implemented in MATLAB and C. The total time is reported in the last row. For convenience, CPU time of the proposed MSFS-RP method is repeated in the last column.

Experiment	LLFS* [64]	Simba* [65]	DEFS* [63]	JMI† [57]	ICAP† [58]	MRF† [73]	CIFE† [59]	MSFS-RP*
ECG (Across-session)	4108	1556	1256	400	769	110	754	790
TEOAE (Across-session)	126	47	304	69	147	3	139	65
ECG (Across-Posture)	682	260	694	188	350	18	331	151
Total time	4916	1863	2254	657	1266	131	1224	1006

of samples and dimensions on the synthetic dataset. This demonstrates that the proposed algorithm can be applied to high dimensional data with thousands of samples. More specifically, complexity of the proposed feature selection algorithm is linear in number of features and quadratic in number of samples. These results show that the proposed feature selection algorithm can be applied to large datasets where both number of samples and features are large.

Regarding the real-world datasets, CPU time of the proposed feature selection algorithms is presented in Table V. EER for both MSFS and MSFS-RP algorithms are also provided for comparison. As may be seen, MSFS-RP speeds up on average about 4 times compare to MSFS algorithm while its effect on EER can be neglected. Here, dimensionality is reduced from ~ 7500 to 500 (i.e., $K = 500$) by random projection. We realized that on average only 21% of samples were selected by the criteria described in section V to estimate the nearest miss and nearest hit for each sample. This leads to less complexity in computing $\bar{\mathbf{d}}_s^{(i)}$ and absolute difference matrices in (7) and (8) (see section V).

TABLE V. CPU time (in seconds) of the MSFS and MSFS-RP algorithms. EER (in percent) is also presented in parentheses.

Experiment	MSFS-RP	MSFS
ECG (Across-session) $M = 3675, J = 7278$	**790(7.6)**	4445(6.9)
TEOAE (Across-session) $M = 600, J = 7522$	**65(3.4)**	134(3.3)
ECG (Across-Posture) $M = 1472, J = 7278$	**151(3.9)**	710(3.7)

We also report the CPU time for our comparison feature selection algorithms in Table VI. The proposed method is implemented in MATLAB that is known to be much slower than many other programming languages especially C. Codes for LLFS, Simba and DEFS are written in MATLAB and JMI, ICAP, MRF and CIFE are written in C. It can be seen that the proposed MSFS-RP algorithm on average is faster than all MATLAB-implemented methods LLFS, Simba and DEFS as well as the two C implemented methods ICAP and CIFE. Although JMI algorithm is implemented in C, it still runs slower than MSFS-RP in 2 out of 3 experiments. The only method that is faster than the proposed method is MRF. This

can mainly be because of its faster programming language (i.e. C) since the complexity of our method and MRF are the same – i.e. quadratic in number of samples and linear in number of features [73]. Note that this is the development phase time which is performed off-line. On the other hand, enrollment and test phases involve only training and testing using SVM. This is much faster than the process in development phase, since no feature selection is required. In our experiments, the test phase is performed in a fraction of a second.

Compare to MSFS algorithm, MSFS-RP reduces the runtime by incorporating only a portion of samples in computing $\bar{\mathbf{d}}_{NM}^{(i)}$ and $\bar{\mathbf{d}}_{NH_s}^{(i)}$ (see equation (5) and (6)). Figure 8 shows the error in computing $\bar{\mathbf{d}}_{NM}^{(i)}$ on TEOAE dataset. Error is defined as $\frac{\|\bar{\mathbf{d}}_{ORG} - \bar{\mathbf{d}}_{RP}\|_2}{\|\bar{\mathbf{d}}_{ORG}\|_2}$ where $\bar{\mathbf{d}}_{ORG}$ corresponds to MSFS algorithm and $\bar{\mathbf{d}}_{RP}$ corresponds to MSFS-RP algorithm. As may be seen, the error is very small – i.e., less that 5%. This can explain why the MSFS-RP method has similar performance compare to the MSFS method because MSFS-RP and MSFS mainly differs in the way that $\bar{\mathbf{d}}_{NM}^{(i)}$ and $\bar{\mathbf{d}}_{NH_s}^{(i)}$ are computed and in fact other steps of both algorithms are the same.

We also compare the MSFS and MSFS-RP algorithms in terms of the number of features they have in common. Fig. 9 shows the percentage of common features between two methods for different values of $\gamma \in \{75\%, 85\%, 95\%, 99\%\}$ in TEOAE dataset. γ controls the number of selected samples. Smaller values of γ lead to selecting a smaller number of samples. Note that at around 300 selected features (where the proposed algorithm usually gives its best performance on this dataset), MSFS-RP with $\gamma = 0.95$ or 0.99 gives a feature subset that has almost all features in common with MSFS (about 98%). However, further decreasing γ significantly decreases the percentage of common features. We set γ to 0.95 (i.e., its default value) without tuning and it is fixed during all
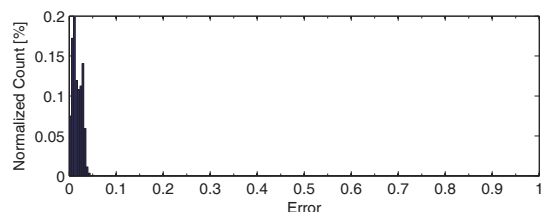

Fig. 8. Approximation error in computing $\bar{\mathbf{d}}_{NM}^{(i)}$ in (5) for TEOAE dataset. Error is defined as $\frac{\|\bar{\mathbf{d}}_{ORG} - \bar{\mathbf{d}}_{RP}\|_2}{\|\bar{\mathbf{d}}_{ORG}\|_2}$ where $\bar{\mathbf{d}}_{ORG}$ corresponds to MSFS method and $\bar{\mathbf{d}}_{RP}$ corresponds to MSFS-RP method.
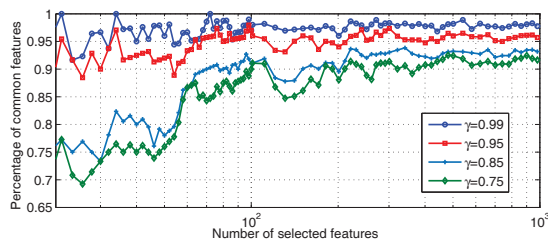
Fig. 9. Percentage of common features shared between MSFS and MSFS-RP versus the number of selected features in TEOAE dataset for $\gamma \in \{75\%, 85\%, 95\%, 99\%\}$.

experiments.

While MSFS-RP algorithm benefits from less complexity due to dealing with smaller subsets of samples, determining these subsets adds some extra time to the algorithm compare to MSFS algorithm. This implies that the dimensionality of low-dimensional space, K, should be small compare to the original dimensionality L. We randomly pick 1000 pairs of samples and compute the l_1 distance between members of each pair before and after being projected into a K-dimensional spaces. After normalizing by $\left(\frac{K}{J}\right)$, the root-mean-square error between the distances before and after projection is computed. We run this procedure 30 times and compute the average error over all runs. Fig. 10 shows the average error for different values of K. As may be seen, K can be reduced while its effect on error is unnoticeable. But if we keep reducing K beyond approximately 500, error will significantly increase. Therefore, we set K to 500 (i.e., its default value) without tuning and it is fixed during all experiments.

VII. CONCLUSION

In this paper an effective and practical framework for human recognition using medical biometrics such as ECG and TEOAE is presented. These biometrics are easy to collect. ECG can be collected from fingertips and TEOAE can be collected using a small earphone with built-in microphone. More importantly, they are robust to replay and falsification attacks due to the inherent liveness requirement and difficulty of stealing or fabrication. However, non-stationary nature of these signals make them hard to deal with especially when only one session is available for enrollment and testing is done in a separate session.

We propose a feature selection method that selects features which are more persistent across sessions. A multisession
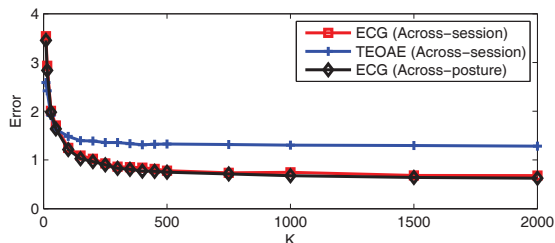
auxiliary dataset of some general subjects is used for feature selection. Selected features are then used for building templates of biometric system users. The proposed approach overcomes the weakness of previously existing feature selection algorithms that ignore the fact that the input data belong to different sessions. Moreover, we embedded an across session criteria in the feature selection process that perfectly fits the aforementioned single session enrollment, across session testing scenario. The proposed feature selection algorithm is computationally efficient, i.e. linear in number of features and quadratic in number of samples and we have demonstrated that it can be applied to large datasets with thousands of features and samples. The proposed method has been compared with six state-of-the-art methods for ECG and TEOAE biometric recognition. In addition, it has been compared with seven existing feature selection algorithms in the literature. Experimental results demonstrate the effectiveness of the proposed method to cope with variation of such signals due to time lapse and body posture.

REFERENCES

[1] Y. Xu, X. Fang, X. Li, J. Yang, J. You, H. Liu, and S. Teng, "Data uncertainty in face recognition," *Cybernetics, IEEE Transactions on*, vol. 44, no. 10, pp. 1950–1961, 2014.
[2] P. Moutafis and I. A. Kakadiaris, "Can we do better in unimodal biometric systems? a rank-based score normalization framework," *IEEE transactions on cybernetics*, vol. 45, no. 12, pp. 2654–2667, 2015.
[3] J. W. Hall, *Handbook of otoacoustic emissions*. Cengage Learning, 2000.
[4] G. Schumacher, "Fingerprint recognition for children," Publications Office of the European Union, Tech. Rep., 2013.
[5] S. Bharadwaj, H. S. Bhatt, R. Singh, M. Vatsa, and S. K. Singh, "Face recognition for newborns: A preliminary study," in *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*. IEEE, 2010, pp. 1–6.
[6] S. Tiwari, A. Singh, and S. K. Singh, "Intelligent method for face recognition of infant," *International Journal of Computer Applications*, vol. 52, no. 4, 2012.
[7] W. Eiserman and L. Shisler, "Identifying hearing loss in young children: Technology replaces the bell." *Zero to Three (J)*, vol. 30, no. 5, pp. 24–28, 2010.
[8] R. F. Cunningham, "Otoacoustic emissions: beyond newborn hearing screening," *Audiology Online*, 2011.
[9] G. G. Molina, F. Bruekers, C. Presura, M. Damstra, and M. van der Veen, "Morphological synthesis of ecg signals for person authentication," in *Proceedings of 15th European Signal Processing Conference*, 2007.
[10] K. N. Plataniotis, D. Hatzinakos, and J. K. Lee, "Ecg biometric recognition without fiducial detection," in *Biometric Consortium Conference, 2006 Biometrics Symposium: Special Session on Research at the*. IEEE, 2006, pp. 1–6.
[11] F. Agrafioti and D. Hatzinakos, "Ecg based recognition using second order statistics," in *Communication Networks and Services Research Conference, 2008. CNSR 2008. 6th Annual*. IEEE, 2008, pp. 82–87.
[12] A. D. Chan, M. M. Hamdy, A. Badre, and V. Badee, "Wavelet distance measure for person identification using electrocardiograms," *Instrumentation and Measurement, IEEE Transactions on*, vol. 57, no. 2, pp. 248–253, 2008.
[13] J. M. Irvine, S. A. Israel, W. Todd Scruggs, and W. J. Worek, "eigenpulse: Robust human identification from cardiovascular function," *Pattern Recognition*, vol. 41, no. 11, pp. 3427–3435, 2008.
[14] S. Z. Fatemian and D. Hatzinakos, "A new ecg feature extractor for biometric recognition," in *Digital Signal Processing, 2009 16th International Conference on*. IEEE, 2009, pp. 1–6.

Fig. 10. Average error in the distance between members of a pair of samples versus K. Error is averaged over 30 runs where in each run errors of 1000 pairs are computed. It can be seen that the error significantly increases when K is smaller than approximately 500.

[15] S.-C. Fang and H.-L. Chan, "Human identification by quantifying similarity and dissimilarity in electrocardiogram phase space," *Pattern Recognition*, vol. 42, no. 9, pp. 1824–1831, 2009.

[16] I. Odinaka, P.-H. Lai, A. D. Kaplan, J. A. O'Sullivan, E. J. Sirevaag, S. D. Kristjansson, A. K. Sheffield, and J. W. Rohrbaugh, "Ecg biometrics: A robust short-time frequency analysis," in *Information Forensics and Security (WIFS), 2010 IEEE International Workshop on*. IEEE, 2010, pp. 1–6.

[17] C. Ye, M. T. Coimbra, and B. V. Kumar, "Investigation of human identification using two-lead electrocardiogram (ecg) signals," in *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*. IEEE, 2010, pp. 1–8.

[18] M. Li and S. Narayanan, "Robust ecg biometrics by fusing temporal and cepstral information," in *Pattern Recognition (ICPR), 2010 20th International Conference on*. IEEE, 2010, pp. 1326–1329.

[19] S. Pathoumvanh, S. Airphaiboon, B. Prapochanung, and T. Leauhatong, "Ecg analysis for person identification," in *Biomedical Engineering International Conference (BMEiCON), 2013 6th*. IEEE, 2013, pp. 1–4.

[20] Z. Zhao, L. Yang, D. Chen, and Y. Luo, "A human ecg identification system based on ensemble empirical mode decomposition," *Sensors*, vol. 13, no. 5, pp. 6832–6864, 2013.

[21] J. Wang, M. She, S. Nahavandi, and A. Kouzani, "Human identification from ecg signals via sparse representation of local segments," *Signal Processing Letters, IEEE*, vol. 20, no. 10, pp. 937–940, 2013.

[22] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[23] J. Ottenbacher, L. Jatoba, U. Großmann, W. Stork, and K. Müller-Glaser, "Ecg electrodes for a context-aware cardiac permanent monitoring system," in *World Congress on Medical Physics and Biomedical Engineering 2006*. Springer, 2007, pp. 672–675.

[24] M. Homer, J. M. Irvine, and S. Wendelken, "A model-based approach to human identification using ecg," in *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2009, pp. 730 625–730 625.

[25] L. Biel, O. Pettersson, L. Philipson, and P. Wide, "Ecg analysis: a new approach in human identification," *Instrumentation and Measurement, IEEE Transactions on*, vol. 50, no. 3, pp. 808–812, 2001.

[26] J. Yao and Y. Wan, "A wavelet method for biometric identification using wearable ecg sensors," in *Medical Devices and Biosensors, 2008. ISSS-MDBS 2008. 5th International Summer School and Symposium on*. IEEE, 2008, pp. 297–300.

[27] F. Agrafioti and D. Hatzinakos, "Signal validation for cardiac biometrics," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. IEEE, 2010, pp. 1734–1737.

[28] D. Jang, S. Wendelken, and J. M. Irvine, "Robust human identification using ecg: eigenpulse revisited," in *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2010, pp. 76 670M–76 670M.

[29] G. Wübbeler, M. Stavridis, D. Kreiseler, R.-D. Bousseljot, and C. Elster, "Verification of humans using the electrocardiogram," *Pattern Recognition Letters*, vol. 28, no. 10, pp. 1172–1175, 2007.

[30] M. A. Swabey, P. Chambers, M. E. Lutman, N. M. White, J. Chad, A. Brown, and S. Beeby, "The biometric potential of transient otoacoustic emissions," *International Journal of Biometrics*, vol. 1, no. 3, pp. 349–364, 2009.

[31] N. J. Grabham, M. A. Swabey, P. Chambers, M. E. Lutman, N. M. White, J. E. Chad, and S. P. Beeby, "An evaluation of otoacoustic emissions as a biometric," *Information Forensics and Security, IEEE Transactions on*, vol. 8, no. 1, pp. 174–183, 2013.

[32] M. Komeili, N. Armanfard, D. Hatzinakos, and A. Venetsanopoulos, "Feature selection from multisession electrocardiogram signals for identity verification," in *Electrical and Computer Engineering (CCECE), 2015 IEEE 28th Canadian Conference on*. IEEE, 2015, pp. 603–608.

[33] I. Odinaka, P.-H. Lai, A. D. Kaplan, J. A. O'Sullivan, E. J. Sirevaag, and J. W. Rohrbaugh, "Ecg biometric recognition: A comparative analysis," *Information Forensics and Security, IEEE Transactions on*, vol. 7, no. 6, pp. 1812–1824, 2012.

[34] S. A. Israel, J. M. Irvine, A. Cheng, M. D. Wiederhold, and B. K. Wiederhold, "Ecg to identify individuals," *Pattern recognition*, vol. 38, no. 1, pp. 133–142, 2005.

[35] K.-S. Kim, T.-H. Yoon, J.-W. Lee, D.-J. Kim, and H.-S. Koo, "A robust human identification by normalized time-domain features of electrocardiogram," in *Engineering in Medicine and Biology Society,*

[36] 2005. IEEE-EMBS 2005. 27th Annual International Conference of the. IEEE, 2006, pp. 1114–1117.

[36] S. Saechia, J. Koseeyaporn, and P. Wardkein, "Human identification system based ecg signal," in *TENCON 2005 2005 IEEE Region 10*. IEEE, 2005, pp. 1–4.

[37] Y. Wang, F. Agrafioti, D. Hatzinakos, and K. N. Plataniotis, "Analysis of human electrocardiogram for biometric recognition," *EURASIP journal on Advances in Signal Processing*, vol. 2008, p. 19, 2008.

[38] J. M. Irvine and S. A. Israel, "A sequential procedure for individual identity verification using ecg," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, p. 3, 2009.

[39] C.-M. Ting and S.-H. Salleh, "Ecg based personal identification using extended kalman filter," in *Information Sciences Signal Processing and their Applications (ISSPA), 2010 10th International Conference on*. IEEE, 2010, pp. 774–777.

[40] N. Venkatesh and S. Jayaraman, "Human electrocardiogram for biometrics using dtw and flda," in *Pattern Recognition (ICPR), 2010 20th International Conference on*. IEEE, 2010, pp. 3838–3841.

[41] M. M. Tawfik, H. Selim, and T. Kamal, "Human identification using time normalized qt signal and the qrs complex of the ecg," in *Communication Systems Networks and Digital Signal Processing (CSNDSP), 2010 7th International Symposium on*. IEEE, 2010, pp. 755–759.

[42] M. Tantawi, K. Revett, A. Salem, and M. F. Tolba, "Fiducial feature reduction analysis for electrocardiogram (ecg) based biometric recognition," *Journal of Intelligent Information Systems*, vol. 40, no. 1, pp. 17–39, 2013.

[43] A. Kaveh and W. Chung, "Temporal and spectral features of single lead ecg for human identification," in *Biometric Measurements and Systems for Security and Medical Applications (BIOMS), 2013 IEEE Workshop on*. IEEE, 2013, pp. 17–21.

[44] J. Chai, H. Liu, and Z. Bao, "Generalized re-weighting local sampling mean discriminant analysis," *Pattern recognition*, vol. 43, no. 10, pp. 3422–3432, 2010.

[45] S. Wahabi, S. Pouryayevali, S. Hari, and D. Hatzinakos, "On evaluating ecg biometric systems: Session-dependence and body posture," *Information Forensics and Security, IEEE Transactions on*, vol. 9, no. 11, pp. 2002–2013, 2014.

[46] M. A. Swabey, S. P. Beeby, A. D. Brown, and J. E. Chad, "Using otoacoustic emissions as a biometric," in *Biometric Authentication*. Springer, 2004, pp. 600–606.

[47] P. Chambers, N. J. Grabham, M. A. Swabey, M. E. Lutman, N. M. White, J. E. Chad, and S. P. Beeby, "A comparison of verification in the temporal and cepstrum-transformed domains of transient evoked otoacoustic emissions for biometric identification," *International Journal of Biometrics*, vol. 3, no. 3, pp. 246–264, 2011.

[48] J. Gao, F. Agrafioti, S. Wang, and D. Hatzinakos, "Transient otoacoustic emissions for biometric recognition." in *ICASSP*, 2012, pp. 2249–2252.

[49] Y. Liu and D. Hatzinakos, "Earprint: Transient evoked otoacoustic emission for biometrics," *Information Forensics and Security, IEEE Transactions on*, vol. 9, no. 12, pp. 2291–2301, 2014.

[50] C. Hou, F. Nie, X. Li, D. Yi, and Y. Wu, "Joint embedding learning and sparse regression: A framework for unsupervised feature selection," *Cybernetics, IEEE Transactions on*, vol. 44, no. 6, pp. 793–804, 2014.

[51] M. Banerjee and N. R. Pal, "Unsupervised feature selection with controlled redundancy (ufescor)," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 12, pp. 3390–3403, 2015.

[52] E. Alpaydin, *Introduction to machine learning*. MIT press, 2004.

[53] G. Karakaya, S. Galelli, S. D. Ahipaşaoğlu, and R. Taormina, "Identifying (quasi) equally informative subsets in feature selection problems for classification: A max-relevance min-redundancy approach," *IEEE transactions on cybernetics*, vol. 46, no. 6, pp. 1424–1437, 2016.

[54] B. Xue, M. Zhang, and W. N. Browne, "Particle swarm optimization for feature selection in classification: A multi-objective approach," *Cybernetics, IEEE Transactions on*, vol. 43, no. 6, pp. 1656–1671, 2013.

[55] J. C. H. Hernandez, B. Duval, and J.-K. Hao, "A genetic embedded approach for gene selection and classification of microarray data," in *European Conference on Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics*. Springer, 2007, pp. 90–101.

[56] K. Nag and N. R. Pal, "A multiobjective genetic programming-based ensemble for simultaneous feature selection and classification," *IEEE transactions on cybernetics*, vol. 46, no. 2, pp. 499–510, 2016.

[57] H. H. Yang and J. E. Moody, "Data visualization and feature selection: New algorithms for nongaussian data." in *NIPS*, vol. 99. Citeseer, 1999, pp. 687–693.

[58] A. Jakulin, "Machine learning based on attribute interactions," Ph.D. dissertation, Univerza v Ljubljani, 2005.

[59] D. Lin and X. Tang, "Conditional infomax learning: An integrated framework for feature extraction and fusion," in *Computer Vision ECCV 2006*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2006, vol. 3951, pp. 68–82.

[60] P. Meyer and G. Bontempi, "On the use of variable complementarity for feature selection in cancer classification," in *Applications of Evolutionary Computing*. Springer Berlin / Heidelberg, 2006, vol. 3907, pp. 91–102.

[61] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 8, pp. 1226–1238, 2005.

[62] Y. Liu, F. Tang, and Z. Zeng, "Feature selection based on dependency margin." *IEEE transactions on cybernetics*, vol. 45, no. 6, p. 1209, 2015.

[63] A. Al-Ani, A. Alsukker, and R. N. Khushaba, "Feature subset selection using differential evolution and a wheel based search strategy," *Swarm and Evolutionary Computation*, vol. 9, pp. 15–26, 2013.

[64] Y. Sun, S. Todorovic, and S. Goodison, "Local-learning-based feature selection for high-dimensional data analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 9, pp. 1610–1626, 2010.

[65] R. Gilad-Bachrach, A. Navot, and N. Tishby, "Margin based feature selection-theory and algorithms," in *Proceedings of the twenty-first international conference on Machine learning*. ACM, 2004, p. 43.

[66] K. Kira and L. A. Rendell, "A practical approach to feature selection," in *Proceedings of the ninth international workshop on Machine learning*. Morgan Kaufmann Publishers Inc., 1992, pp. 249–256.

[67] Y. Sun, "Iterative relief for feature weighting: algorithms, theories, and applications," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 6, pp. 1035–1051, 2007.

[68] B. Chen, H. Liu, J. Chai, and Z. Bao, "Large margin feature weighting method via linear programming," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 21, no. 10, pp. 1475–1488, 2009.

[69] B. Liu, B. Fang, X. Liu, J. Chen, Z. Huang, and X. He, "Large margin subspace learning for feature selection," *Pattern Recognition*, vol. 46, no. 10, pp. 2798–2806, 2013.

[70] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.

[71] W. B. Johnson and J. Lindenstrauss, "Extensions of lipschitz mappings into a hilbert space," *Contemporary mathematics*, vol. 26, no. 189-206, p. 1, 1984.

[72] D. Achlioptas, "Database-friendly random projections," in *Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*. ACM, 2001, pp. 274–281.

[73] Q. Cheng, H. Zhou, and J. Cheng, "The fisher-markov selector: fast selecting maximally separable feature subset for multiclass classification with applications to high-dimensional data," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 6, pp. 1217–1233, 2011.

[74] Medical biometric databases. [Online]. Available: http://www.comm.utoronto.ca/ biometrics/databases

[75] Y. Liu and D. Hatzinakos, "Human acoustic fingerprints: A novel biometric modality for mobile security," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 3784–3788.

[76] H. P. Da Silva, A. Fred, A. Lourenco, and A. K. Jain, "Finger ecg signal for user authentication: Usability and performance," in *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*. IEEE, 2013, pp. 1–8.

[77] P. S. Raj and D. Hatzinakos, "Feasibility of single-arm single-lead ecg biometrics," in *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European*. IEEE, 2014, pp. 2525–2529.

[78] P. S. Raj, S. Sonowal, and D. Hatzinakos, "Non-negative sparse coding based scalable access control using fingertip ecg," in *Biometrics (IJCB), 2014 IEEE International Joint Conference on*. IEEE, 2014, pp. 1–6.

[79] J. Pan and W. J. Tompkins, "A real-time qrs detection algorithm," *Biomedical Engineering, IEEE Transactions on*, no. 3, pp. 230–236, 1985.

[80] H. Sedghamiz. (2013) An online algorithm for r,s and t wave detection. [Online]. Available: http:// www.mathworks.com /matlabcentral /fileexchange /45404-ecg-q-r-s-wave-online-detecto

[81] Vivosonic inc. [Online]. Available: www.vivosonic.com

[82] N. Armanfard, M. Komeili, J. P. Reilly, and L. Pino, "Vigilance lapse identification using sparse eeg electrode arrays," in *Electrical and Computer Engineering (CCECE), 2016 IEEE 29th Canadian Conference on*. IEEE, 2016.

[83] M. S. Robinette, "Clinical observations with transient evoked otoacoustic emissions with adults," *Seminars Hearing*, vol. 13, no. 1, p. 2326, 1992.

Majid Komeili received the M.Sc. degree in Electrical and Computer Engineering from Tarbiat Modares University, Tehran, Iran, in 2008. He is currently pursuing the Ph.D. degree with the University of Toronto, Toronto, ON, Canada. Since 2012, he has been a Research Assistant and a Teacher Assistant with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto. His current research interests include signal, image and video processing, specifically physiological signal analysis, machine learning, machine vision, video surveillance, and document image processing.

Wael Louis received B.Eng degree (with honours) in Electrical Engineering, Ryerson University, Canada, in 2008 and M.A.Sc. and Ph.D. degrees from The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Canada in 2010 and 2017, respectively. Wael Louis is a recipient (twice) of Alexander Graham Bell Canada Graduate Scholarship for both M.A.Sc. and Ph.D. levels. He has also received the Ontario Graduate Scholarship. His research interests lie in the areas of computer vision, signal and image processing, and biometric systems.

Narges Armanfard received the M.Sc. degree in Electrical and Computer Engineering from Tarbiat Modares University, Tehran, Iran, in 2008, and the Ph.D. degree in electrical and computer engineering from McMaster University, Hamilton, ON, Canada, in 2016. She is currently a Post-Doctoral Fellow with Toronto Rehabilitation Institute, University of Toronto, Toronto, ON, Canada. Her current research interests include signal, image and video processing, specifically physiological signal analysis, machine learning, machine vision, video surveillance, and document image processing.

Dimitris Hatzinakos received the Diploma degree from the University of Thessaloniki, Greece, in 1983, the M.A.Sc degree from the University of Ottawa, Canada, in 1986 and the Ph.D. degree from Northeastern University, Boston, MA, in 1990, all in Electrical Engineering. In September 1990 he joined the Department of Electrical and Computer Engineering, University of Toronto, where now he holds the rank of Professor with tenure. He served as Chair of the Communications Group of the Department during the period July 1999 to June 2004. Between 2004-2014, he was the holder of the Bell Canada Chair in Multimedia, at the University of Toronto. Since 2007 he has been the director and the chair of the management committee of the Identity, Privacy and Security Institute (IPSI) at the University of Toronto. His research interests are in the areas of Multimedia Signal Processing, Multimedia Security, Multimedia Communications and Biometric Systems. He is author/co-author of more than 300 papers in technical journals and conference proceedings and he has contributed to 17 books in his areas of interest.