One Lead ECG Based Personal Identification with Feature Subspace Ensembles

Hugo Silva[1], Hugo Gamboa[2], and Ana Fred[3]

[1] Instituto de Telecomunicações, Lisbon, Portugal
hugo.silva@lx.it.pt
[2] Escola Superior de Tecnologia de Setúbal, Campus do IPS,
Setúbal, Portugal
hgamboa@est.ips.pt
[3] Instituto de Telecomunicações,
Instituto Superior Técnico, Lisbon, Portugal
afred@lx.it.pt

Abstract. In this paper we present results on real data, focusing on personal identification based on one lead ECG, using a reduced number of heartbeat waveforms. A wide range of features can be used to characterize the ECG signal trace with application to personal identification. We apply feature selection (FS) to the problem with the dual purpose of improving the recognition rate and reducing data dimensionality. A feature subspace ensemble method (FSE) is described which uses an association between FS and parallel classifier combination techniques to overcome some FS difficulties. With this approach, the discriminative information provided by multiple feature subspaces, determined by means of FS, contributes to the global classification system decision leading to improved classification performance. Furthermore, by considering more than one heartbeat waveform in the decision process through sequential classifier combination, higher recognition rates were obtained.

1 Introduction

Fiducial points of the electrocardiographic (ECG) signal, are typically used in clinical applications for diagnostics and evaluation of the cardiac system function [1][2][3]. These points have well characterized reference values, and deviations from those may express multiple anomalies.

The ECG provides a visualization of the electrical activity of the cardiac muscle fibres; as measured from the body surface, the ECG signal is directly related to the physiology of each individual. These measurements are influenced by physiologic factors which include: skin conductivity, genetic singularities, position, shape and size of the heart. Regardless of what factors originate differences in the measurement, the fact that the ECG contains physiologic dependant singularities potentiates its application to personal identification.

Recent research work has been devoted to the characterization of ECG features, unique to an individual, with clear evidence that accurate ECG based

P. Perner (Ed.): MLDM 2007, LNAI 4571, pp. 770–783, 2007.

personal identification is possible [4][5][6]. As a behavioral biometric technique the ECG is very appealing: *it is a non-invasive technique*; *it is not easily replicated or circumvented*; and *it requires the subject to be physiologically active*, among other characteristics.

A wide range of features can be used to characterize the ECG signal trace with application to personal identification [1][7][8][9][3], and a question arises: *for a given feature set, which features are truly relevant for the decision process, and which can be discarded.* The reasons why addressing this question is of paramount importance include: (a) *the curse of dimensionality problem* [10]; and (b) *the fact that some features may misguide the decision process* [11][12].

In pattern recognition, this can be addressed through *feature selection* (FS). Considering a d-dimensional feature representation space (FRS), $F=\{f_1, \cdots, f_d\}$, feature selection consists of determining which subspace $F^* \subset F$, if any, contains the features $f_j \in F$ with most relevant discriminative information [13]. For this purpose, a variety of methods has been proposed [14][15][16].

This paper presents results on real data, for the application of one lead ECG data to personal identification. Previous approaches to the problem [4][5][6], also using real data, have shown the potential of ECG data for subject identification through contingency matrix analysis. In our approach, we study the potential of subject identification using a reduced number of heartbeat waveforms, with the purpose of real-time analysis. We focus on studying the classification performance provided on one hand by a single heartbeat waveform, and on the other hand by multiple heartbeat waveforms. FS and classifier combination techniques are applied to the problem to improve the recognition rates, with positive results when compared to the cases where no FS is performed.

An overview of our *feature subspace ensemble* (FSE) approach is presented: a parallel classifier combination method, in which a global decision is produced by combination of the individual decisions of multiple classifiers, designed using subspaces of the original feature representation space F, obtained by means of FS [17]. Each considered feature subspace contributes to the global decision as a result of the classifier combination process. This allows us to overcome one of the difficulties associated with FS: *retrieval of relevant discriminative information contained in discarded features.* FSE was applied to the problem, and proved to be more effective than a single classifier trained on a single FRS, both for the cases where the original space F, and FS determined subspaces were used.

We evaluate the recognition rate of a single heartbeat waveform for different sizes of the training and validation data, in order to determine the minimum number of patterns necessary to achieve maximum recognition rates. With the same purpose, sequential classifier combination is also employed, to determine how the recognition rate evolves by using a reduced number of heartbeat waveforms for personal identification instead of a single one.

The rest of the paper is organized as follows: Section 2 describes the feature subspace ensemble parallel classifier combination approach. Section 3 details our one lead ECG based personal identification setup and evaluation conditions.

Section 4 presents results for the one lead ECG based personal identification problem. Finally, section 5 summarizes results and presents the main conclusions.

2 Feature Subspace Ensembles

Feature selection is an important tool in classification system design. The classification process is essentially a mapping $F \rightarrow W$, of the original FRS, F, into a set $W = \{w_1, \cdots, w_c\}$ of c categories. FS consists on determining a subspace $F^* \subset F$, containing only the features $f_j \in F$ with the most relevant discriminative information, with the threefold aim of: (a) *improving the discriminative capacitive*; (b) *reducing computational demands*; and (c) *removing redundant or superfluous information* [13]. For this purpose, numerous methods and frameworks have been suggested [18][19][20][14]. In this section, we overview FS and some of the difficulties arising from its usage, and describe a feature subspace ensemble (FSE) method, designed to overcome some of those problems.

Typically, FS methods fall into one of three generic classes: *filter methods*, which are based on the discriminative information provided by individual or groups of features from the original FRS; *wrapper methods*, which are based on the performance of a learning machine; and *embedded methods*, in which the feature subspaces are a consequence of the classifier training process. In general, FS methods are based on the optimization of a feature subspace evaluation criteria, which measures the relevance of F^* in terms of discriminative potential, and usually only suboptimal solutions are guaranteed.

Let $S(A, J, X)$ denote a *feature selection context* (FSC), defined as the FS parameters comprehended by the feature selection algorithm A, the feature subspace evaluation criteria J, and the training data X, through which a given F^* is determined.

As a result of FS, some features from the original FRS are discarded during the process and not incorporated in F^*. Although interesting results are achieved through FS [21][14][15], some difficulties often arise: (a) *solution overfitting to a particular feature selection context* (FSC); (b) *suboptimality of the obtained solutions*; (c) *solution diversity with respect to the FSC*; and (d) *loss of relevant discriminative information contained in features* $f_j \in F \setminus F^*$.

Thus, we devised a more effective method which uses parallel classifier combination rules [12][22], to combine the decisions of multiple, individual classifiers C_r; each designed using its own subspace $F_r^* \subset F$, obtained by means of feature selection in different FSCs. A related approach proposed in [23], uses the combined decision of classifiers constructed on sequentially selected features sets, forcing the full coverage of the original FRS, F.

Let $\mathcal{S} = \{S_1, \cdots, S_p\}$ be a set of p features selection contexts, differing in any combination of the parameters A_r, J_r, or X_r, $(0 < r \leq p)$. In our *feature subspace ensemble* (FSE) approach [17], a set of p feature subspaces $\mathcal{F}^* = \{F_1^*, \cdots, F_p^*\}$ is determined using each FSC, $S_r \in \mathcal{S}$ (thus the term feature subspace ensemble). Using each feature subspace $F_r^* \in \mathcal{F}$, a classifier C_r is designed, forming a set $\mathcal{C} = \{C_1, \cdots, C_p\}$ of p classifiers. For the classification of a given pattern x_i, each individual classifier $C_r \in \mathcal{C}$ produces a decisions $\hat{w}_{C_r}$, and in the end all decisions are combined by a classifier combination strategy [12][24][22][25][26], in order to produce a global decision $\hat{w}_{x_i}$. Figure 1 illustrates the described approach.

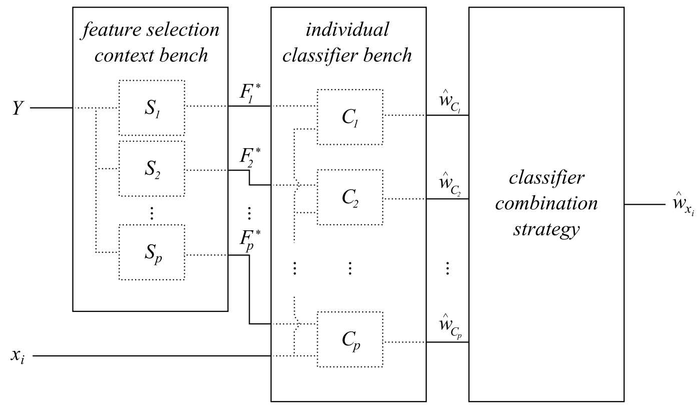


Fig. 1. Feature subspace ensemble (FSE) system. A set $\mathcal{C} = \{C_1, \cdots, C_p\}$ of p classifiers is trained using individual feature subspaces F_r^* obtained for some variation S_r of the FSC. Each classifier $C_r \in \mathcal{C}$ produces an individual decision $\hat{w}_{C_r}$. All individual decisions are combined using a classifier combination strategy to produce a global decision $\hat{w}_{x_i}$.

Figure 2 condenses the results of 50 FS runs on the SAT benchmark data from the UCI machine learning repository [27]; in a given run r, the feature selection context S_r, $(0 < r \leq 50)$, is composed by fixed A_r and J_r (that is, the same type in all runs), and randomly selecting 50% of the available patterns in each run to create X_r. A_r is a sequential forward search (SFS) wrapper framework (later described in section 3.3); J_r is the classification performance of a 1-NN decision rule using X_r as training data to classify the remaining 50% of the available patterns (used as validation set).

In the context of figure 2, in a FRS of dimension $d = 36$ features, the mean feature subspace size was approximately 23 features; an horizontal line indicates the histogram mean. As shown, only a few features are consistently selected in most feature subspaces over all runs, and there is full coverage of the original FRS. This means that there is a great diversity of subspaces with relevant discriminative information, and in a single FS run some of the discarded features may still contain useful information.

Through parallel classifier combination we incorporate in the global decision relevant discriminative information contained in each particular feature subspace, eventually recovering relevant features discarded as a result of a single FS run (e.g., due to a particularly inadequate or misleading FSC). This way, the classification system becomes less sensitive to misleading feature subspaces; the combined decisions of individual classifiers is capable of overcoming inaccurate

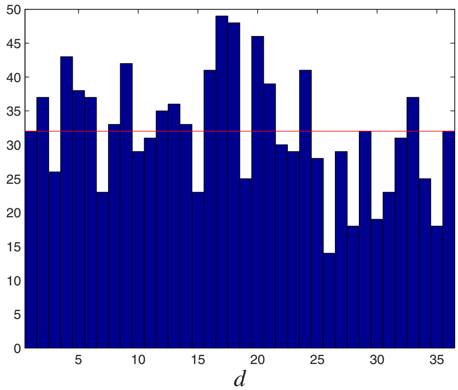


Fig. 2. Illustrative histogram of 50 FS runs on the SAT benchmark data from the UCI machine learning repository. The FSC context S_r of a given run r was established as follows: A_r and J_r are fixed for all runs, A_r being a wrapper sequential forward search framework, and J_r being the classification performance of a 1-NN decision rule trained on X_r to classify a validation set; X_r is randomly selected from the available set of patterns in each run. The horizontal axis corresponds to each of the dimensions of the FRS; the vertical axis corresponds to the number of times a given dimension d was selected. An horizontal line indicates the histogram mean.

decisions resulting from low quality feature subspaces, provided that a sufficient number of feature subspaces exists, that leads to accurate decisions.

In section 3.3 we present a FSE implementation, which we have applied to the ECG based personal identification problem. Comparative results show that, a single classifier designed using a single feature subspace obtained by means of FS, outperforms the case where the original feature representation space F is used, that is, when no feature selection is performed. With feature subspace ensembles further improvements were obtained, outperforming the classification performance of both cases.

3 One Lead ECG Based Personal Identification

3.1 Data Acquisition

Unlike previous work, where ECG recordings were performed at rest [28][6], and in stress potentiating tasks [4], we present preliminary results on real data acquired during a cognitive activity. Twenty six subjects, 18 males and 8 females, between the ages of 18 and 31 years, willingly participated in individual sessions (one per subject), during the course of which their ECG signal was recorded.

In each individual session the subject was asked to complete a concentration task on a computer, designed for an average completion time of 10 minutes. The subject interacted with the computer in a sitting position, using only the mouse as input device. No posture or motion restrictions during the activity were imposed, however, the ECG acquisition was part of a wider multi-modal physiological signal acquisition experiment; therefore due to the placement of other measurement apparatus in the subjects passive hand[1] it was suggested to the subject to reduce the movements of the passive hand to the indispensable minimum.

Fig. 3. Illustration of one grid of digits from the concentration task that each subject was asked to complete, and during which the ECG signal acquisition was performed

The task consisted on the presentation of two grids with 800 digits, similar to the one illustrated in figure 3, with the goal of identifying every consecutive pair of digits that added 10. Each grid was traversed in a line wise manner, from the top left to right bottom corner. The task was designed to induce saturation, having the following constraints: in order to be able to move from a current line to the next, the current line would have to be fully traversed; once a new line was moved into, the previously traversed ones could not be accessed. An horizontal bar and a cursor followed the mouse movement along the horizontal axis; the horizontal bar informed the subject of the point until which the current line had been traversed, and the cursor highlighted the pair of consecutive numbers over which the mouse was hovering at a given point. Whenever the user identified a consecutive pair of numbers matching the goal and highlighted by the cursor, he would mark it with a mouse click, and although it was not possible to return to previously visited lines, within the same line the markings could be revised.

A one lead surface mount ECG placement on the V_2 precordial derivation [1][3] was used. Facing the subject, the V_2 derivation is located on the fourth intercostal space over the mid clavicular line, at the right of the sternum. Prior to sensor placement, the area was prepared with abrasive gel and conductive paste was used on the electrodes to improve conductivity.

[1] we define active hand as the one used to control the input device; passive hand as the free hand.

3.2 Signal Processing and Feature Extraction

The acquired ECG signals were band-pass filtered in the passing band $2 - 30Hz$ with a zero-phase forward and reverse scheme [29], to remove high frequency powerline noise and low frequency baseline wander artifacts from the signal.

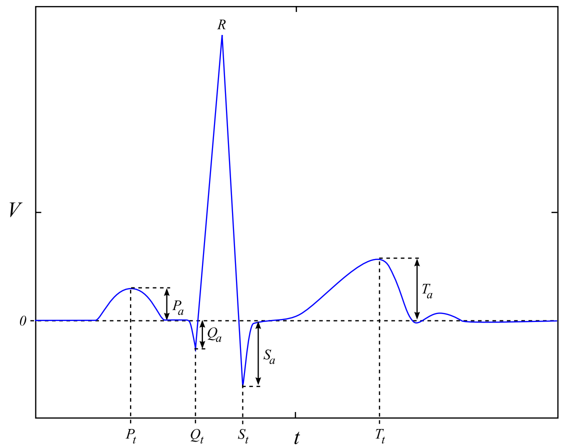


Fig. 4. Measured features from the ECG heartbeat waveform

Each heartbeat waveform was sequentially segmented from the full recording, and after this all waveforms were aligned by their R peaks. From the resulting collection of ECG heartbeat waveforms, the mean wave for groups of 10 heartbeat waveforms (without overlapping), was computed to minimize the effect of outliers. A labeled database was compiled, in which each pattern corresponds to a mean wave.

For each mean waveform, 8 latency and amplitude features were extracted, along with a sub-sampling of the waveform itself. This resulted in a feature representation space F of dimension $d = 53$, with 4 latency features, 4 amplitude features measured at selected points (figure 4), and 45 amplitude values measured at the sub-sampled points. No time limit was imposed to complete the task, and therefore the heartbeat wave form collection of each subject in the database was truncated at approximately 6 minutes[2] to ensure uniform class distribution.

3.3 Feature Selection and Classification

The ECG mean wave database is used for evaluation purposes; 50 data selection runs were performed, where in each run r three mutually exclusive sets X_r, Y_r and Z_r, of randomly selected patterns from the full recording are created. Also a feature subspace F_r^* is determined using the individual feature subspace

[2] Corresponding to the fastest completion time over all subjects.

selection framework described next. As a result, 50 feature subspaces will be available as a result of the performed data selection runs. X_r is created with 22.5% of the available patterns and used as training set both for FS and classifier design; Y_r is created with another 22.5% of the available patterns and used as validation set in FS; and the remaining 55% of the available patterns were used to create Z_r, which served as testing set for classification performance assessment.

For our experiments we have employed a wrapper FS framework [16], with a heuristic sequential forward search (SFS) method [30]. SFS is a state space search method, which starts from an initial state $F^*_{t=0} = \emptyset$ and iteratively evolves by constructing at each step all possible super-spaces $F_{t+1} = F^*_t \cup \{f_j \in F \setminus F^*_t\}$, adding each of the features $f_j \in F \setminus F^*_t$ to the optimal subspace F^*_t obtained at the previous step. J is used to evaluate each of the resulting super-spaces F_{t+1}, and F^*_{t+1} is selected as the set which optimizes J. If $J(F^*_{t+1}) < J(F^*_t)$ the search is stopped[3], and $F^*_r = F^*_t$ is considered to be the feature subspace with most relevant discriminative information for a given FSC r. Although conceptually simple, wrapper SFS feature selection has proven to hold comparable results in benchmark data when compared to other (more complex) methods [31][32].

The feature subspace evaluation criteria J in wrapper methods is the optimization of the classification performance of a learning machine. In our implementation, J is trained with X_r, and the recognition error over Y_r is used for feature subspace evaluation; therefore F^*_r is determined as the feature subspace that provides higher recognition rate over the validation set Y_r. Using all feature subspaces computed through SFS during the 50 data selection runs, a feature subspace ensemble $\mathcal{F} = \{F^*_1, \cdots, F^*_{50}\}$ was created, and used for classification performance evaluation of the FSE method.

For classification, we use the k-NN decision rule with an Euclidean neighborhood metric [12]. A 1-NN neighborhood was adopted, since it is a particular case of the k-NN rule where $\hat{w}_{x_i}$ for a given pattern x_i is assigned as the category of the closest pattern from the training set X_r. The same type of classifier is used for feature subspace evaluation criteria J, and for classification performance assessment.

Two types of classification performance analysis were performed. On one hand, we evaluated the recognition rate of a single heartbeat waveform for different sizes of the training and validation data, in order to determine the minimum number of patterns necessary to achieve maximum recognition rates. On the other hand, to determine how the recognition rate evolves by using more than one heartbeat waveforms in personal identification instead of a single one, we evaluated the classification performance achieved by combination of the individual decisions of a reduced set of heartbeat waveforms.

Therefore, additionally to the FSE parallel classifier combination method, sequential classifier combination was also employed. A simple majority voting strategy was adopted as classifier combination rule in both cases [33][24][34].

[3] $J(F)$ denotes the usage of J in the evaluation of a given feature subspace F.

4 Results

In this section we present results for the one lead ECG based personal identification. We evaluate the classification performance of a single classifier designed using a single feature subspace both for the cases where no feature selection is performed, and for FS selected feature subspaces. Our feature subspace ensemble method, described in section 2 is also applied to the problem.

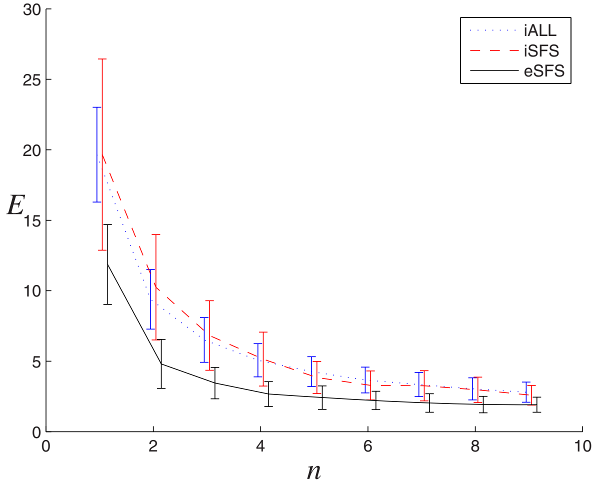
Figure 5 illustrates the evolution of the mean recognition error of a single heartbeat waveform (figure 5(a)), and feature subspace size (figure 5(b)), computed over 50 runs according to the methodology described in section 3.3. To determine the minimum number n of patterns necessary to achieve the maximum recognition rate, we experimented training and validation sets (X_r, and Y_r respectively) of different sizes, ranging from a single training and validation pattern $n = 1$ (1 mean heartbeat waveform), to the full set of $n = 9$ patterns (which as described in section 3.3, corresponds to 22.5% of the available patterns in each run). As we can observe the error rate is fairly similar with (curve $iSFS$) and without SFS feature selection (curve $iALL$), although feature selection leads to more compact feature spaces, as illustrated in figure 5(b). An improved recognition rate was achieved with the application of FSE to the problem (curve $eSFS$).

We can observe that even using a single pattern per subject in the training and validation sets, the average recognition error rate is approximately 19.65% using all features, and 19.66% with SFS selected feature subspaces. In this case, the feature subspace ensemble method reduced the recognition error rate to approximately 11.86%. By increasing the number n of patterns in the training and validation sets, the recognition error rate is highly decreased. The minimal values are reached when the whole set of training and validation data is used, with a recognition error rate of 2.80% using all features and 2.58% with SFS selected feature subspaces. In this case, the FSE method further improved the average recognition error rate to 1.91%.

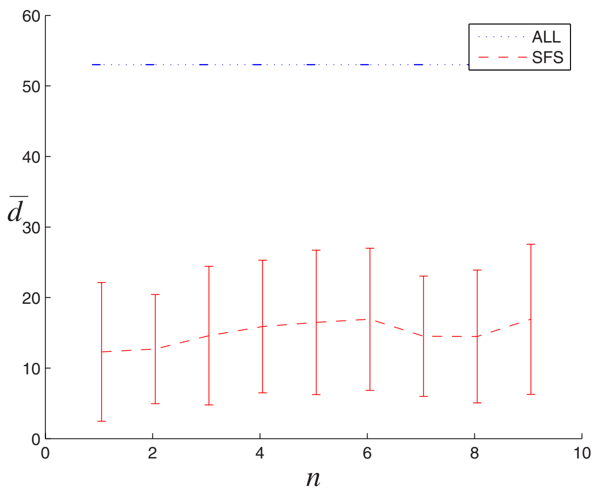
Figure 6 illustrates the feature histogram for the SFS selected feature subspaces over all runs, when the full set of training and validation patterns is used. The mean subspace size is 19.62 features; as we can observe, there is a high feature subspace diversity, and there are several relevant features that not all FSCs lead to. This explains why feature subspace ensembles consistently improved the recognition error rate. We can also observe the presence of irrelevant features, which FS discards or are rarely selected. From figure 5(a) we can see that these, although irrelevant are not misleading the classifier designed using the original feature representation space F, since the recognition error rate is only marginally superior to the results obtained for the classifier design using a single SFS selected feature subspace.

With FSE, a single mean heartbeat waveform, which in our case corresponds to approximately 7 seconds of signal acquisition[4] (since each pattern corresponds

[4] this calculation was performed taking as a reference an average normal resting heart rate of 70 beats per minute [3].



(a) Mean classification error and standard deviation intervals



(b) Mean feature subspace size and standard deviation intervals

Fig. 5. Mean recognition error of a single ECG heartbeat waveform (figure 5(a)), and feature subspace size (figure 5(b)). n: number of patterns used for the training and validation sets (X_r and Y_r); E: mean classification error; $\bar{d}$: mean subspace size; *all*: no feature selection; *sfs*: wrapper sequential forward search; the i prefix denotes the curves for individual classifier and subspace cases, and the e prefix denotes the curves for the feature subspace ensemble method.

to the mean wave of a group of 10 heartbeat waveforms), provides 98.09% recognition accuracy, using a training set of 9 patterns (that is, 63 seconds).

Maintaining the methodology described in section 3.3, we also evaluated the recognition rate of personal identification using more than one heartbeat

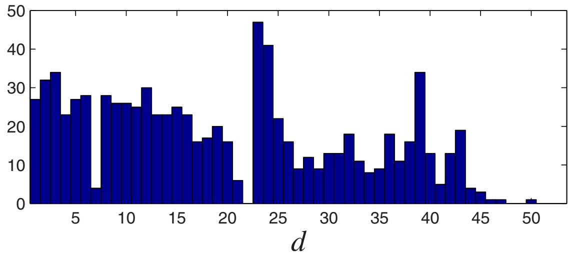


Fig. 6. Histogram for the SFS selected feature subspaces over all runs when the full set of training and validation patterns is used. The horizontal axis corresponds to each dimension of the FRS; the vertical axis corresponds to the number of times a given dimension d was selected. The horizontal line indicates the histogram mean.

waveform. The classification performance obtained for reduced sets of $h = 3, \cdots, 8$ heartbeat waveforms was evaluated, and sequential classifier combination through majority voting was used as classifier combination strategy. Figure 7 illustrates these results. It is important to recall that the FS step was performed to optimize the recognition rate of a single heartbeat waveform (as described in section 3.3). Nonetheless, as we can observe, considering a reduced set of heartbeat waveforms greatly improves the recognition accuracy. The highest recognition rate (99.97%), was obtained by majority voting the individual FSE

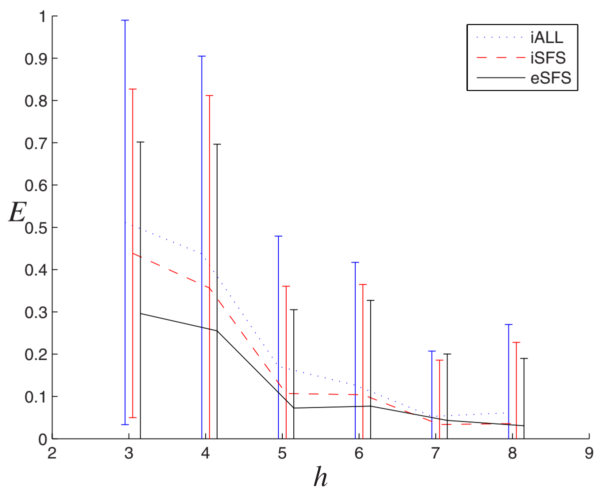


Fig. 7. Mean recognition error and standard deviation intervals for subject identification from sets with a reduced number h of ECG heartbeat waveforms. h: size of the set of heartbeats waveforms; E: mean classification error; *all*: no feature selection; *sfs*: wrapper sequential forward search; the i prefix denotes the curves for individual classifier and subspace cases, and the e prefix denotes the curves for the feature subspace ensemble method.

decisions for a group of 8 heartbeat waveforms (the equivalent to 56 seconds of signal acquisition according to the adopted methodology).

5 Conclusions

In this paper we addressed a real data problem of ECG based personal identification using a single, and reduced sets of heartbeat waveforms described in a feature representation space of dimension $d = 53$ measured features. We evaluated the classification performance of a single classifier using the original FRS, and recurred to feature selection to improve the recognition rate and reduce data dimensionality.

We introduced the concept of feature selection context (FSC): the conditions under which a given feature subspace is obtained; and described the generic feature subspace ensemble (FSE) approach: a parallel classifier combination method which uses an association between FS and classifier combination techniques [17]. FSE was designed to overcome some of the difficulties resulting from FS, namely: *FSC overfitting*; *suboptimality of FS methods*; and *recovery of relevant discriminative information contained in features discarded by FS*.

An instantiation of the FSE method using a wrapper heuristic sequential forward search (SFS) framework, 1-NN classifier and the majority voting classifier combination rule, was applied to the ECG based personal identification problem providing higher recognition rates than the single classifier designed using a single FRS cases (both the original FRS, and FS selected through feature subspaces).

Preliminary results have shown that the ECG can be used to identify individuals, particularly useful as a behavioral biometric technique. High recognition rates were achieved using a single heartbeat waveform, and we were able to further improve the results by using sequential classifier combination techniques to combine the individual decisions of a reduced set of heartbeat waveforms. It is important to enhance that, in each evaluation run, a random selection of the patterns was performed from the full recording. This indicates robustness of the ECG signal, since the task during which the signal was acquired was designed to induce saturation.

Through FSE, using a set of 9 training patterns we achieved a personal identification rate of 98.09% from a single heartbeat waveform pattern (which according to the adopted methodology corresponds to 7 seconds of signal acquisition). Using sequential classifier combination in conjunction with FSE, combining the individual decisions from FSE over a reduced set of heartbeat waveforms to produce a global decision, further improved the recognition rates. We were able to achieve a 99.97% subject recognition rate by combining the individual decisions of 8 heartbeat waveforms, which according to the adopted methodology corresponds to 56 seconds of signal acquisition.

FS targets dimensionality reduction and better discriminative ability, by selecting from the original FRS only the features with relevant discriminative information for a given FSC. Classifier combination strategies target the decision refinement, by taking into account multiple individual decisions in order to provide for a global decision. FSE has the potential to combine the advantages of both FS and classifier combination, since through FS reduced dimensionality is achieved; and through classifier combination, the classification system becomes less sensitive to misleading feature subspaces due to particularly inadequate FSCs.

Ongoing and future work includes further validation of the obtained results by including longer databases and a higher number of recordings per individual.

Acknowledgments

This work was partially supported by the Portuguese Foundation for Science and Technology (FCT), Portuguese Ministry of Science and Technology, and FEDER, under grant POSI/EEA-SRI/61924/2004, and by the Institute for Systems and Technologies of Information, Control and Communication (INSTICC) and the Instituto de Telecomunicações (IT), Pólo de Lisboa, under grant P260.

References

1. Chung, E.: Pocketguide to ECG Diagnosis. Blackwell Publishing Professional, Malden (December 2000)
2. Dubin, D.: Rapid Interpretation of EKG's, 6th edn. Cover Publishing Company, Cover (2000)
3. Lipman, B.: Ecg Assessment and Interpretation. F.A. Davis (February 1994)
4. Israel, S., Irvine, J., Cheng, A., Wiederhold, M., Wiederhold, B.: Ecg to identify individuals. Pattern Recognition 38(1), 133–142 (2005)
5. Shen, T., Tompkins, W., Hu, Y.: One-lead ecg for identity verification. In: Proceedings of the Second Joint EMBS/BMES Conference, pp. 62–63 (2002)
6. Biel, L., Petterson, O., Phillipson, L., Wide, P.: Ecg analysis: A new approach in human identification. IEEE Transactions on Instrumentation and Measurement 50(3), 808–812 (2001)
7. Liang, H.: Ecg feature elements identification for cardiologist expert diagnosis. In: ICPR 2004. Proceedings of the 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 1-7, pp. 3845–3848. IEEE Computer Society Press, Los Alamitos (2005)
8. Kunzmann, U., von Wagner, G., Schochlin, J.A.B.: Parameter extraction of ecg signals in real-time. Biomedizinische Technik. Biomedical engineering 47, 875–878 (2002)
9. Duskalov, I., Dotsinsky, I., Christov, I.: Developments in ecg acquisition, preprocessing, parameter measurement, and recording. IEEE Engineering in Medicine and Biology Magazine 17(2), 50–58 (1998)
10. Bellman, R.: Adaptive Control Processes. Princeton University Press, Princeton (1961)
11. Heijden, V., Duin, R., Ridder, D., Tax, D.: Classification, parameter estimation and state estimation - an engineering approach using MATLAB. John Wiley & Sons, Chichester (2004)
12. Duda, R., Hart, P., Stork, D.: Pattern classification, 2nd edn. John Wiley & Sons, Chichester (2001)
13. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. Journal of Machine Learning Research 3, 1157–1182 (2003)

14. Kudo, M., Sklansky, J.: Comparison of algorithms that select features for pattern classifiers. Pattern Recognition 33, 25–41 (2000)
15. Jain, A., Zongker, D.: Feature selection: Evaluation, application, and small sample performance. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(2), 153–158 (1997)
16. Kohavi, R., John, G.: Wrappers for feature subset selection. Artificial Intelligence 97(1-2), 273–324 (1997)
17. Silva, H., Fred, A.: Feature subspace ensembles: A parallel classifier combination scheme using feature selection. In: MCS 2007. LNCS, vol. 4472, Springer, Heidelberg (to appear, 2007)
18. Forman, G.: A pitfall and solution in multi-class feature selection for text classification. In: ICML '04. Proceedings of the 21st International Conference on Machine Learning, pp. 38–46. ACM Press, New York (2004)
19. Molina, L., Belanche, L., Nebot, A.: Feature selection algorithms: A survey and experimental evaluation. lsi technical report lsi-02-62-r (2002)
20. Kittler, J., Pudil, P., Somol, P.: Advances in statistical feature selection. In: Singh, S., Murshed, N., Kropatsch, W.G. (eds.) ICAPR 2001. LNCS, vol. 2013, pp. 425–434. Springer, Heidelberg (2001)
21. Molina, L., Belanche, L., Nebot, A.: Feature selection algorithms: a survey and experimental evaluation. In: ICDM 2002. Proceedings. IEEE International Conference on Data Mining, pp. 306–313. IEEE Computer Society Press, Los Alamitos (2002)
22. Duin, R., Tax, D.: Experiments with classifier combining rules. In: Kittler, J., Roli, F. (eds.) MCS 2000. LNCS, vol. 1857, Springer, Heidelberg (2000)
23. Skurichina, M., Duin, R.: Combining feature subsets in feature selection. In: Oza, N.C., Polikar, R., Kittler, J., Roli, F. (eds.) MCS 2005. LNCS, vol. 3541, pp. 165–175. Springer, Heidelberg (2005)
24. Fred, A.: Finding consistent clusters in data partitions. In: Kittler, J., Roli, F. (eds.) MCS 2001. LNCS, vol. 2096, pp. 309–318. Springer, Heidelberg (2001)
25. Lam, L.: Classifier combinations: Implementation and theoretical issues. In: Kittler, J., Roli, F. (eds.) MCS 2000. LNCS, vol. 1857, pp. 78–86. Springer, Heidelberg (2000)
26. Kittler, J., Hatef, M., Duin, R., Matas, J.: On combining classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence 20, 226–239 (1998)
27. Newman, D., Hettich, D., Blake, C., Merz, C.: UCI repository of machine learning databases (1998)
28. Shen, T., Tompkins, W.: Biometric statistical study of one-lead ecg features and body mass index (bmi). In: Proceedings of 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 1162–1165. IEEE Computer Society Press, Los Alamitos (2005)
29. Gustafsson, F.: Determining the initial states in forward-backward filtering. IEEE Transactions on Signal Processing 44(4), 988–992 (1996)
30. Russell, S., Norvig, P.: Artificial Intelligence: A Modern Approach, 2nd edn. Prentice-Hall, Englewood Cliffs (2002)
31. Silva, H.: Feature selection in pattern recognition systems. Master's thesis, Universidade Técnica de Lisboa, Instituto Superior Técnico (2007)
32. Reunanen, J.: Overfitting in making comparisons between variable selection methods. Journal of Machine Learning Research 3, 1371–1382 (2003)
33. Tax, D., Duin, R.: Using two-class classifiers for multiclass classification. In: International Conference on Pattern Recognition, Quebec, Canada (2002)
34. Lam, L., Suen, S.: Application of majority voting to pattern recognition: An analysis of its behavior and performance. IEEE Transactions on Systems, Man, and Cybernetics 27, 553–568 (1997)