Comparison of Dimensionality Reduction Methods for Road Surface Identification System

Gonzalo Safont[1(✉)], Addisson Salazar[1], Alberto Rodríguez[2], and Luis Vergara[1]

[1] Institute of Telecommunications and Multimedia Applications, Universitat Politècnica de València, Valencia, Spain
gonsaar@upvnet.upv.es, {asalazar,lvergara}@dcom.upv.es
[2] Departamento de Ingeniería de Comunicaciones, Universidad Miguel Hernández de Elche, Elche, Spain
arodriguezm@umh.es

Abstract. Road surface identification is attracting more attention in recent years as part of the development of autonomous vehicle technologies. Most works consider multiple sensors and many features in order to produce a more reliable and robust result. However, on-board limitations and generalization concerns dictate the need for dimensionality reduction methods. This work considers four dimensionality reduction methods: principal component analysis, sequential feature selection, ReliefF, and a novel feature ranking method. These methods are used on data obtained from a modified passenger car with four types of sensors. Results were obtained using three classifiers (linear discriminant analysis, support vector machines, and random forests) and a late fusion method based on alpha integration, reaching up to 96.10% accuracy. The considered dimensionality reduction methods were able to reduce the number of features required for classification greatly and increased classification performance. Furthermore, the proposed method was faster than ReliefF and sequential feature selection and yielded similar improvements.

Keywords: Classification · Decision fusion · Feature selection · Road surface identification · Self-driving vehicles

1 Introduction

Road surface identification (RSI) is one of the key parts in the development of autonomous or semi-autonomous vehicle technologies. Driving decisions and adjustments could be made depending on the road the vehicle is traversing, improving traffic flow, assisting the driver, and improving safety and driving experience. Existing studies on RSI can be roughly classified in three groups, depending on the road surfaces considered: (i) road roughness profiling and other road maintenance works; (ii) detecting hazardous weather conditions (sleet, snow, rain…); (iii) detecting different types of terrain (grass, asphalt, stone…).

Road roughness profiles and related studies are typically concerned with cost-effective solutions for supervising and planning road maintenance works. For instance,

K. Arai et al. (Eds.): SAI 2020, AISC 1229, pp. 554–563, 2020.
https://doi.org/10.1007/978-3-030-52246-9_40

Tudon-Martinez et al. estimated road roughness profiles from multisensor data as a cost-effective alternative to existing laser-based solutions [1], and Park et al. detected road damage and anomalies using deep ensemble networks [2]. Conversely, studies on hazardous weather conditions are typically concerned with increasing the safety of the vehicle occupants. For instance, Alonso et al. identified wet road conditions using tire/road noise to prevent *aquaplaning* [3], and Zhao, Wu and Chen employed video feeds to detect four hazardous weather conditions (wet road, snow, ice, and water) [4]. Finally, studies that detect the type of terrain have goals that depend on the considered classes. For instance, Masino et al. classified five types of pavement using the sound in the tire cavity for the purpose of estimating road traffic using support vector machines and post-processing [5], and Bystrov et al. have considered sonar, ultrasound, and radar to perform classification of the road surface in front of the car for the purposes of autonomous car technology [6–8]. Both studies reached over 90% accuracy in their respective problems.

This work is related to the third category: detecting different types of terrain. It presents a combination of four different sensors for road surface identification: accelerometers, microphones, speed signals, and the torque and position of the steering wheel. The last two sensors are already included in the vehicle's electric power steering (EPS) system, while the former two were specially fitted for this work. The number of features extracted from these sensors is relatively large compared to the number of available samples. Therefore, dimensionality reduction was carried out to identify significant features and eliminate irrelevant or redundant features. There are two main reasons for dimensionality reduction. Firstly, we would like to reduce the number of features for computational reasons: faster evaluation times, lower memory consumption, lower implementation cost, and so on. This is particularly important in autonomous vehicle technologies, where on-board systems are constrained. Secondly, reducing the number of features might improve performance or reduce the variability of the result. While hierarchical and knowledge discovery methods could perform a similar function [9–12], they tend to require large computational or temporal costs, which defeats the first reason.

This work compares several dimensionality reduction methods for a road surface identification system. The following methods are considered: principal component analysis (PCA, [13]); sequential feature selection [19]; and the ReliefF feature ranking method [20]. Furthermore, a novel feature ranking method is also proposed. The effect of these dimensionality reduction methods is considered on three different classifiers: linear discriminant analysis (LDA), support vector machines (SVM), and random forests (RDF). Furthermore, late fusion of the three classifiers using alpha integration [21–23] is also considered.

2 Dimensionality Reduction Methods

Given the high dimensionality of the data in this work, dimensionality reduction was performed on the extracted features before classification. Dimensionality reduction methods are typically classified in two categories: feature extraction, where new features are derived from the original ones; and feature selection, where one or more of the original features are selected and the rest are discarded [19]. In turn, feature selection is typically approached in one of two ways: ranking features according to some criterion and selecting the top Q features (feature ranking); or selecting a subset of features that keep or improve classification performance (subset selection) [19]. Subset selection algorithms can automatically determine the number of selected features, while feature ranking algorithms need to rely on a user-determined threshold (or some other method) to set the number of selected features.

2.1 Principal Component Analysis

PCA is a linear transformation of the data that maximizes the uncorrelation of the transformed components. The original features are projected onto an orthogonal space where the projections (components) are sorted by the amount of variance of the original data they explain. In many applications, the first few components contain most of the original variance, thus the rest can be discarded to reduce the dimensionality of the problem. PCA has been heavily used as a feature extraction method and to investigate data structure [13]. However, unlike feature selection methods, PCA requires all the original features to compute the projected components. The consideration of more advanced feature extraction methods such as stochastic PCA or independent component analysis [14–18] is outside the scope of this work.

2.2 Sequential Feature Selection

Sequential feature selection (SFS, [19]) is a feature subset selection method that performs a greedy search of the optimal subset of features. There are two main types of SFS: forward, where the subset of selected features is iteratively grown from the empty set; and backward, where the subset of features is iteratively reduced from the full set. Essentially, at each iteration, the method determines the effect of adding one of the unselected features to the subset (forward SFS) or removing one of the selected features from the subset (backward SFS) using a wrapped classification method. For each iteration, forward SFS adds the feature that would increase performance the most if added to the current subset. Conversely, backward SFS removes the feature that would increase performance the most if removed from the current subset. In both cases, iterations continue until performance cannot be improved any more. In a trained system, discarded features can be omitted from the feature extraction stage, thus lightening the load of the system. In this work, we considered forward SFS, with each of the classifiers as the wrapped method.

2.3 Feature Ranking

Feature ranking methods assign a score or importance to each feature and then rank them by score [19]. This ranked list can then be used to perform feature selection, e.g., by choosing the Q best-ranked features. As with feature subset selection, the features that are not selected are never used and, in a trained system, could be omitted from the feature extraction stage.

ReliefF. ReliefF is a feature selection method that considers the interactions between features, returning a score that can be used later for feature ranking [20]. The feature scores computed by ReliefF are based on the distances between nearest neighbors. Essentially, the distances between nearest neighbors of the same class decrease the score of the feature, and the distances between nearest neighbors of different classes increase the score of the feature.

Proposed Feature Ranking Method. In this work, we propose a feature ranking method based on simple classifiers. The method is robust, quick to compute, and does not require a lot of memory. For binary classification problems, the score of the ith feature is computed as the informedness [24] of a simple classifier fit to the feature:

$$informedness = specificity + recall - 1 \tag{1}$$

The simple one-variable classifier used for feature ranking is shown in Algorithm 1. The algorithm maximizes the informedness of the result when splitting the values using two thresholds; any value x is assigned to class 1 if $x_1 < x \leq x_2$ and class 0 otherwise. The search for the pair of values that optimize the informedness of the classification is done in only one pass through the set of values. This method is fast to compute and only considers the order of the values of the input variable, while disregarding the actual values. Thus, the method is robust with respect to isolated outliers, extreme values, and any data transformation that does not affect the order of the values of the features, such as: centering, scaling, exponentiation (when negative values of x are possible, only for odd powers), and logarithmic transformation.

Algorithm 1. Simple one-feature classifier based on the optimization of the informedness.

0 Given input feature values x_n and binary labels y_n ($y_n = 1$ if n belongs to class j, and 0 otherwise), $n = 1 \ldots N$

1 Sort the values x_n in ascending order, then arrange the binary labels y_n in the same order; call the sorted values x_n' and the sorted labels y_n'

2 For $n = 1 \ldots N$

3 Compute the probability of detection P_d and the probability of false alarm P_f for values up to x_n':

$$P_d(x \leq x_n') = \frac{\sum_{i=1}^{n} y_i'}{\sum_{i=1}^{N} y_i'}$$

$$P_f(x \leq x_n') = \frac{\sum_{i=1}^{n} 1 - y_i'}{\sum_{i=1}^{N} 1 - y_i'}$$

4 Compute informedness as $I(x \leq x_n') = P_d(x \leq x_n') - P_f(x \leq x_n')$

5 Find the indices of the maximum and minimum values of informedness,

$$n_1 = \underset{n}{\operatorname{argmin}}\, I(x \leq x_n')$$

$$n_2 = \underset{n}{\operatorname{argmax}}\, I(x \leq x_n')$$

where $x_1 = x_{n_1}'$ and $x_2 = x_{n_2}'$

6 If $x_1 < x_2$, the score of feature x is the informedness within the range $(x_1, x_2]$:

$$I(x_1 < x \leq x_2) = I(x_{n_1}' \leq x) - I(x_{n_2}' \leq x)$$

7 Else, the score of feature x is the informedness outside the range $(x_2, x_1]$:

$$I((x \leq x_2) \cup (x_1 < x)) = I(x_{n_1}' \leq x) - I(x_{n_2}' \leq x)$$

For multiclass problems such as the one shown in this work, the problem with K classes is first divided into K binary 1-vs-all problems. Then, the score of the ith feature is obtained as the average of its scores for each of the K binary problems.

3 Experiments

The dimensionality reduction methods described in Sect. 2 were used on a set of data from a road surface identification problem. These data were obtained using a specially converted passenger car coursing over three different road surfaces: cobblestones, smooth flat asphalt, and stripes. Ten channels were recorded:

- Three channels from an accelerometer on the intermediate shaft (X, Y, and Z directions).
- Three mono sound channels from microphones, two located close to the driver's head and one on the upper side of the electric power steering (EPS) system column.
- Two channels with the speed of the left and right wheels of the car.
- The torque and position of the steering wheel.

These four sensors were sampled at a rate of 48 kHz. A grand total of 63 files were taken, each with a different combination of vehicle speed and road surface, and the features shown in Table 1 were extracted from each channel in epochs of size 1.5 s, with an overlap of 1.4 s between windows. This resulted in a total of 8,309 samples and 560 features available for classification. The high dimensionality of the data and the restrictions of the problem (on-board car systems) justify the necessity of dimensionality reduction methods.

We considered the following classifiers because of their widespread application in machine learning problems: linear discriminant analysis (LDA), support vector machines with linear kernel (SVM), and random forests with 50 trees (RDF). While SVM and RDF might be computationally expensive to train, the trained model is fast to evaluate, making them appropriate for the task. LDA included a regularization term equal to $\gamma = 0.01$, in order to compensate for the large amount of features. Aside from the classifiers, we also considered late fusion using separated score integration (SSI, [21–23]). SSI is a method based on alpha integration that optimally combines the scores from several classifiers into one response. The parameters of SSI were obtained by the last mean squares criterion [23].

To test the effect of the considered dimensionality reduction methods on performance, we performed a series of Monte Carlo experiments that followed the procedure shown in Fig. 1. First, the files were randomly split in three subsets: training (50% of the files), validation (25% of the files), and testing (the remaining 25% of the files). For training, all the features shown in Table 1 were extracted. Then, one of the dimensionality reduction methods described in Sect. 2 was trained using said features, and the features remaining after reduction were used to train the LDA, SVM and RDF classifiers. During validation, the results of training were used to modify the feature extraction stage by omitting discarded features, and then the scores of LDA, SVM and RDF were obtained on the remaining features. SSI was trained using the scores of the classifiers on the validation stage. Finally, for testing, the modified extracted features were used to obtain results for LDA, SVM and RDF, and the scores of said classifiers were fused using SSI. This experiment was repeated for 100 iterations for each dimensionality reduction method, and the average and standard deviation of the results were obtained.

As explained in Sect. 2, PCA and feature ranking methods cannot automatically determine the number of selected features. To determine the optimal number of selected

Table 1. Features extracted from data x in epochs of length Δ.

Feature	Definition		
Third-order autocorrelation	$\frac{1}{\Delta-2}\sum_{n=3}^{\Delta} x(n)\cdot x(n-1)\cdot x(n-2)$		(2)
Time reversibility	$\left(\frac{1}{\Delta}\sum_{n=1}^{\Delta} x^2(n)\right)^{-3/2} \frac{1}{\Delta-1}\sum_{n=2}^{\Delta}\left(x(n)-x(n-1)\right)^3$		(3)
Average power	$\frac{1}{\Delta}\sum_{n=1}^{\Delta} x^2(n)$		(4)
Centroid frequency	$\frac{f_S}{\Delta}\frac{\sum_{f=1}^{\Delta} f\left\|X(f)\right\|^2}{\sum_{f=1}^{\Delta}\left\|X(f)\right\|^2}$	where $X(f)$ is the direct Fourier transform of x within the epoch taken at Δ points, and f_S is the sampling rate	(5)
Maximum frequency	$\frac{f_S}{\Delta}\left(\arg\max_f \left\|X(f)\right\|\right)$		(6)
Spectral contrast	$\frac{\max_f \left\|X(f)\right\|}{\min_f \left\|X(f)\right\|}, f_{o1} \le f \le f_{o2}$	where f_{o1}, f_{o2} are respectively the start and end indices of the oth octaves, taking 440Hz as the reference (440Hz is the end limit of the 4th octave) [25]. 10 octaves were considered.	(7)
Spectral slope	Trend a of the model $\log\left\|X(f)\right\| = a\log f + b$		(8)
Spectral flatness	$\frac{\left(\prod_{f=f_{o1}}^{f_{o2}}\left\|X(f)\right\|\right)^{1/(f_{02}-f_{01})}}{\frac{1}{f_{o2}-f_{o1}}\sum_{f=f_{o1}}^{f_{o2}}\left\|X(f)\right\|}$	where f_{o1}, f_{o2} are respectively the start and end indices of the oth quarter of octave [25]	(9)

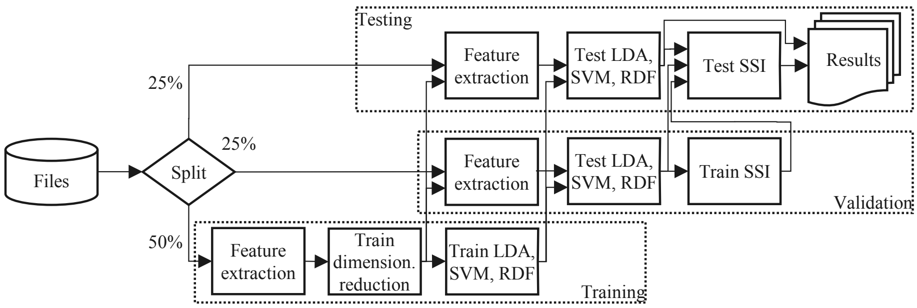

Fig. 1. Diagram of each iteration of the proposed Monte Carlo experiments.

features, the number of selected features was changed from 2 to 560 (all features) and the result of each classifier was verified over the validation subset for 100 iterations. SFS was not included in this experiment because the optimal number of selected features belonging in the subset is automatically decided by the method. Likewise, SSI was also not included in the experiment because it will be used to combine the optimal scores of LDA, SVM, and RDF, regardless of the features used for those classifiers.

The results of this training are shown in Fig. 2. For LDA, ReliefF yielded the best performance with 15 features, the proposed method yielded a very similar result with 10 features, and PCA yielded the worst result with 325 features. For SVM, ReliefF yielded the best result with 250 features, PCA yielded the next best result with 6 features, and the proposed method yielded the worst result with 5 features. Finally, for RDF, the proposed method and ReliefF yielded the best result with 560 features (i.e., no feature selection), and PCA yielded the worst result with 5 features. This might have owed to the subset selection performed in RDF, which might have interfered with the variable selection procedure.

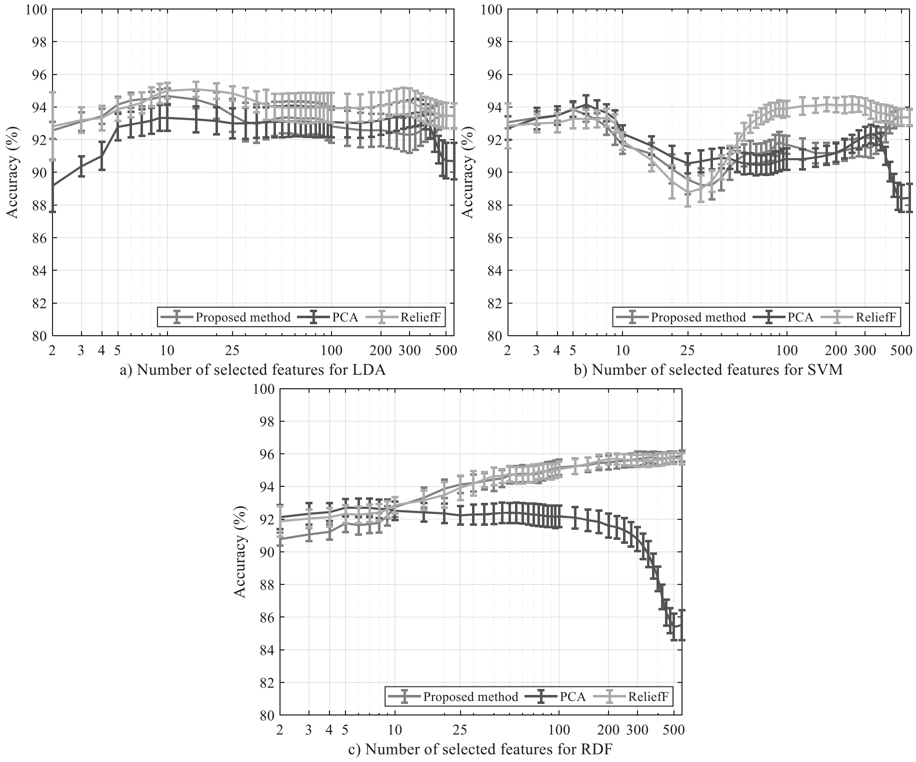

Fig. 2. Performance of each of the considered classifiers with respect to the number of features selected using PCA, ReliefF, and the proposed feature ranking method.

Overall, PCA yielded the lowest performance, and both feature ranking methods (ReliefF and the proposed method) yielded similar performances. PCA, ReliefF and the proposed method showed different trends and results for different classifiers, which suggested that the optimal number of remaining features would depend on the subsequent stages in the classification. Thus, in the following, we used a different number of remaining features for each dimensionality reduction method on each classifier.

After selecting the optimal number of features for each dimensionality reduction method and each classifier, we repeated the experiment and added the results for SFS and SSI. The results of this experiment are shown in Table 2.

Table 2. Results of the proposed RSI system with different dimensionality reduction methods.

Dimensionality reduction method	Classifier	# features	Accuracy (%)	
			Average	Std. error
Sequential feature selection	LDA	4	94.81	0.29
	SVM	4	93.64	0.49
	RDF	4	92.82	0.66
	SSI	n/a	96.03	0.20
PCA	LDA	325	93.36	0.69
	SVM	6	93.35	0.61
	RDF	5	92.12	0.69
	SSI	n/a	94.13	1.44
ReliefF	LDA	15	95.22	0.28
	SVM	250	94.46	0.56
	RDF	560	95.53	0.37
	SSI	n/a	96.10	0.71
Proposed feature ranking method	LDA	10	94.50	0.35
	SVM	5	93.10	0.46
	RDF	560	95.48	0.36
	SSI	n/a	96.02	0.28

In accordance with the results in Fig. 2, PCA yielded the worst overall result. ReliefF yielded the best result, and the proposed feature ranking method and SFS yielded similar results. With respect to the computation times, PCA and the proposed method took an average of 0.28 s to compute, ReliefF took an average of 53.33 s to compute, and SFS took an average of 50 min to compute. These calculations were performed in Matlab R2016b, running on a Windows 7 machine with an Intel Xeon E3 CPU and 16 GB of RAM. Thus, the proposed feature ranking method was much faster than the other considered feature selection methods, and just as fast as PCA.

4 Conclusion

This work has considered the effect of several dimensionality reduction methods on a road surface identification system using a real dataset with three types of road surfaces. Four dimensionality reduction methods were considered: one feature extraction method, PCA; one subset feature selection method, SFS; and two feature ranking methods, ReliefF and a novel method proposed in this work. The results of these dimensionality reduction methods were then tested using three single classifiers (LDA, SVM and RDF) and a late fusion method based on alpha integration (SSI).

The considered dimensionality reduction methods were able to significantly reduce the number of features required for classification and improve classification performance, reaching a maximum classification accuracy of 96.10%. Out of the original 560 features, most of the combinations of dimensionality reduction method and classifier were able to use 15 features or less (2.68% of the features).

Furthermore, the proposed feature ranking method yielded competitive results to successful dimensionality reduction techniques, such as ReliefF and forward sequential feature selection, and outperform the classical PCA technique. In terms of computational cost, the proposed feature ranking method was just as fast as PCA and considerably faster than ReliefF and SFS. This proves the potential of the proposed method for dimensionality reduction.

Acknowledgment. This work was supported by Spanish Administration and European Union grant TEC2017-84743-P, and Generalitat Valenciana under grant PROMETEO/2019/109.

References

1. Tudon-Martinez, J.C., Fergani, S., Sename, O., Martinez, J.J., Morales-Menendez, R., Dugard, L.: Adaptive road profile estimation in semiactive car suspensions. IEEE Trans. Control Syst. Technol. **23**(6), 2293–2305 (2015)
2. Park, J., Min, K., Kim, H., Lee, W., Cho, G., Huh, K.: Road surface classification using a deep ensemble network with sensor feature selection. Sensors **18**, (2018). Article no. 4342
3. Alonso, J., López, J.M., Pavón, I., Recuero, M., Asensio, C., Arcas, G., Bravo, A.: On-board wet road surface identification using tyre/road noise and support vector machines. Appl. Acoust. **76**, 407–415 (2014)
4. Zhao, J., Wu, H., and Chen, L.: Road surface state recognition based on SVM optimization and image segmentation processing. J. Adv. Transp. **2017** (2017). Article no. 6458495
5. Masino, J., Pinay, J., Reischl, M., Gauterin, F.: Road surface prediction from acoustical measurements in the tire cavity using support vector machine. Appl. Acoust. **125**, 41–48 (2017)
6. Bystrov, A., Hoare, E., Tran, T.Y., Clarke, N., Gashinova, M., Cherniakov, M.: Automotive surface identification system. In: IEEE International Conference on Vehicular Electronics and Safety (ICVES), Vienna, Austria, pp. 115–120 (2017)
7. Bystrov, A., Hoare, E., Tran, T.Y., Clarke, N., Gashinova, M., Cherniakov, M.: Road surface classification using automotive ultrasonic sensor. Procedia Eng. **168**, 19–22 (2016)
8. Bystrov, A., Abbas, M., Hoare, E., Tran, T.Y., Clarke, N., Gashinova, M., Cherniakov, M.: Analysis of classification algorithms applied to road surface recognition. In: 2015 IEEE Radar Conference (RadarCon), Piscataway, NJ, USA, pp. 907–911 (2015)
9. Igual, J., Salazar, A., Safont, G., Vergara, L.: Semi-supervised Bayesian classification of materials with impact-echo signals. Sensors **15**(5), 11528–11550 (2015)
10. Salazar, A., Igual, J., Vergara, L., Serrano, A.: Learning hierarchies from ICA mixtures. In: IEEE International Joint Conference on Artificial Neural Networks, Orlando, FL, USA, pp. 2271–2276 (2007)

11. Salazar, A., Gosalbez, J., Bosch, I., Miralles, R., Vergara, L.: A case study of knowledge discovery on academic achievement, student desertion and student retention. In: IEEE ITRE 2004 - 2nd International Conference on Information Technology: Research and Education, London, United Kingdom, pp. 150–154 (2004)
12. Salazar, A., Igual, J., Safont, G., Vergara, L., Vidal, A.: Image applications of agglomerative clustering using mixtures of non-Gaussian distributions. In: International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, pp. 459–463 (2015)
13. Jolliffe, I.T.: Principal Component Analysis. Springer, New York (2002)
14. Shamir, O.: A stochastic PCA and SVD algorithm with an exponential convergence rate. In: International Conference on Machine Learning, Lille, France, pp. 144–155 (2015)
15. Llinares, R., Igual, J., Salazar, A., Camacho, A.: Semi-blind source extraction of atrial activity by combining statistical and spectral features. Digit. Signal Process. Rev. J. **21**(2), 391–403 (2011)
16. Safont, G., Salazar, A., Rodriguez, A., Vergara, L.: On Recovering missing ground penetrating radar traces by statistical interpolation methods. Remote Sens. **6**(8), 7546–7565 (2014)
17. Safont, G., Salazar, A., Vergara, L., Gomez, E., Villanueva, V.: Probabilistic distance for mixtures of independent component analyzers. IEEE Trans. Neural Netw. Learn. Syst. **29**(4), 1161–1173 (2018)
18. Safont, G., Salazar, A., Vergara, L., Gómez, E., Villanueva, V.: Multichannel dynamic modeling of non-Gaussian mixtures. Pattern Recognit. **93**, 312–323 (2019)
19. Lui, H., Motoda, H. (eds.): Computational Methods of Feature Selection. CRC Press, Boca Ratón (2007)
20. Kononenko, I., Šimec, E., Robnik-Šikonja, M.: Overcoming the myopia of inductive learning algorithms with RELIEFF. Appl. Intell. **7**(1), 39–55 (1997)
21. Amari, S.: Information Geometry and its Applications. Springer, Berlin (2016)
22. Soriano, A., Vergara, L., Bouziane, A., Salazar, A.: Fusion of scores in a detection context based on alpha-integration. Neural Comput. **27**, 1983–2010 (2015)
23. Safont, G., Salazar, A., Vergara, L.: Multiclass alpha integration of scores from multiple classifiers. Neural Comput. **31**(4), 806–825 (2019)
24. Powers, D.M.W.: Evaluation: from precision, recall and F-measure to ROC, informedness, markedness & correlation. J. Mach. Learn. Technol. **2**(1), 37–63 (2011)
25. Peeters, G.: A large set of audio features for sound description (similarity and classification) in the CUIDADO project. CUIDADO IST Project Report 54 (2004)