applied sciences

MDPI

Article

Experimental Assessment of Feature Extraction Techniques Applied to the Identification of Properties of Common Objects, Using a Radar System

José Francisco Díez-Pastor, Pedro Latorre-Carmona *, José Luis Garrido-Labrador, José Miguel Ramírez-Sanz and Juan J. Rodríguez

Depto. de Ingeniería Informática, Universidad de Burgos, Avda. Cantabria s/n, 09006 Burgos, Spain; jfdpastor@ubu.es (J.F.D.-P.); jlgarrido@ubu.es (J.L.G.-L.); jmrsanz@ubu.es (J.M.R.-S.); jjrodriguez@ubu.es (J.J.R.)
* Correspondence: plcarmona@ubu.es

Abstract: Radar technology has evolved considerably in the last few decades. There are many areas where radar systems are applied, including air traffic control in airports, ocean surveillance, and research systems, to cite a few. Other types of sensors have recently appeared, which allow tracking sub-millimeter motion with high speed and accuracy rates. These millimeter-wave radars are giving rise to myriad new applications, from the recognition of the material close objects are made, to the recognition of hand gestures. They have also been recently used to identify how a person interacts with digital devices through the physical environment (Tangible User Interfaces, TUIs). In this case, the radar is used to detect the orientation, movement, or distance from the objects to the user's hands or the digital device. This paper presents a thoughtful comparative analysis of different feature extraction techniques and classification strategies applied on a series of datasets that cover problems such as the identification of materials, element counting, or determining the orientation and distance of objects to the sensor. The results outperform previous works using these datasets, especially when the accuracy was lowest, showing the benefits feature extraction techniques have on classification performance.

Keywords: radar signal; feature extraction; classification; stacking; tangible user interfaces

check for updates

Citation: Díez-Pastor, J.F.; Latorre-Carmona, P.; Garrido-Labrador, J.L.; Ramírez-Sanz, J.M.; Rodríguez, J.J. Experimental Assessment of Feature Extraction Techniques Applied to the Identification of Properties of Common Objects, Using a Radar System. *Appl. Sci.* **2021**, *11*, 6745. https://doi.org/10.3390/app11156745

Academic Editor: Francesco Bianconi

Received: 9 June 2021
Accepted: 21 July 2021
Published: 22 July 2021

1. Introduction

Radar sensing has been classically used in an extensive range of applications, due to its ability to operate under all-weather and scene illumination geometry independence acquisition conditions. This would be a critical advantage, under specific circumstances, when compared, for instance, to optical sensing. Advances in radar hardware and software technology have made it possible to reliably detect and track objects, under competitive classification accuracy conditions, in underwater, air, and ground environments [1,2]. However, this framework seems to have been applied only to relatively big objects in specific scenarios, i.e., airplanes, ships, or submarines [3,4].

Radar technology has also found an important field of application in remote sensing data classification and health monitoring for environmental preservation purposes, usually combined in a common optical remote sensing framework [5,6].

Apart from these research fields, radar sensors are also being applied nowadays in other areas, as in action and gesture recognition or in autonomous driving, to cite a few cases. In particular, there is an increasing interest in the use of radar technology in human gesture and action recognition because it is aimed at solving the problem of low recognition accuracy that vision based systems may have. Not only gestures [7] but also more complex actions are aimed at by the new radar acquisition technology and classification strategies [8]. Even daily or ordinary activities (e.g., cooking, eating, and resting) may be classified using this type of acquisition technology [9]. Usually, human action recognition is made using sensors whose location is fixed, but new approaches using unmanned aerial vehicles (UAVs) are starting to appear [10].

Another area where radar technology is being applied is related to object and people classification for systems aimed at helping create a safer driving environment, or even in autonomous driving [11,12].

Some of the newest radar sensing applications do not try to obtain information from objects over long, mid, or short distances (in the range of meters), but just within a few centimeters. A sensor developed by Google, called Soli, was announced in 2015 and obtained great media interest. It is a millimeter-wave radar, and it has found several research applications. In [13], this sensor is used to identify up to 26 types of materials and 10 body parts from several participants. In [14], it is used to classify and distinguish five common types of materials, namely aluminum, ceramic, plastic, wood, and water, regardless of their different sizes and thicknesses. A hand gesture recognition system, which successfully distinguishes 10 gestures, is proposed in [15]. This type of radar sensor has even been used for face verification [16] or to differentiate between blood samples of disparate glucose concentrations in the range of 0.5 to 3.5 mg/mL [17]. However, the number of research studies focused on the application of this technology for daily object material classification and other interactions seems to still be scarce [18]. A Tangible User Interface (TUI) allows the user to interact with digital information through real actions. This type of user interfaces is known to be more usable and easier to understand, especially for elderly people [19], since, although not everyone knows how to operate a keyboard or mouse, everyone is familiar with grasping or moving common objects. The idea behind TUIs is that there is a direct link between the digital system and the way the physical objects are manipulated.

Miniature milimeter-wave radar sensing has been proposed to enhance the interactions by identifying materials or estimating the orientation or distance at which the physical elements are located [18]. Millimeter-wave radar technology could be considered a cost effective option for material identification, counting objects/items, or estimating their position, even when they might be partially covered or occluded. This property is a great advantage when compared to other sensors such as cameras.

The aim of this paper is to apply a diverse group of classification strategies and feature extraction techniques on a series of datasets of different nature, acquired by a portable radar sensor, aiming at validating this technology as a good candidate to be used in TUI sensing problems. The classification strategies include Random Forests [20] and Support Vector Machines [21]. Moreover, different classifiers are combined in ensembles, using Stacked Generalization [22].

The feature extraction techniques applied in this paper are the basic aggregation features (e.g., averages and root mean squares) previously used by Yeo et al. [18] and two techniques originally introduced for time series classification: ROCKET [23] and TSFRESH [24]. The datasets used in this study, obtained using a radar system, are not time series. Nevertheless, as in time series, the features are arranged and therefore their order is important. For instance, functions on the values in an interval (e.g., the first half of the series or channel) can be used. Hence, methods proposed for times series classification can also be used for this kind of data. It is common practice to use time series classification methods (including multivariate time series) for datasets where the feature order is important even though the features do not represent different times [25,26], for instance, from spectroscopy [27,28]. Image contours or outlines, such as arrowheads [29], leaves [30], or fish species [31], can be represented as time series.

The rest of the paper is organized as follows. Section 2 describes the group of datasets, as well as the corresponding feature extraction and classification algorithms used. Section 3 discusses the results. These results are validated and analyzed using average rankings and post hoc tests such as the Nemenyi test, as well as by using Bayesian Signed-Rank Test, all of them in order to determine which combination of feature vector and classification method is best, and whether this improvement (i.e., difference in accuracy performance) is statistically significant or not. Section 4 presents the conclusions and discusses potential future lines of research.

2. Datasets, Feature Extraction, and Classification Methods

Radar sensing uses an electromagnetic signal that hits an object. This signal might be directly reflected or scattered/absorbed by it, therefore giving an overview of the properties of the material the object is made of or of the distance or position the object is at. The type of radar whose data form the repository we used is a Frequency-Modulated Continuous-Wave (FMCW) radar, and this type of radar system has shown its potential to be used for detection/recognition purposes.

Soli [18] is a mono-static, multi-channel (8 channels = 2 transmitters × 4 receivers) radar device, operating in the 57–64 GHz range (center frequency of 60 GHz), using an FMCW principle, where the radar transmits and receives information on a continuous basis. When an object is placed on the top or nearby the Soli sensor, the energy transmitted from the sensor is absorbed and scattered by the object, which varies depending on its distance, thickness, shape, density, internal composition, and surface properties, commonly described as the Radar Cross Section (RCS). As a result, the signals that are reflected back represent rich information about the contributions from a range of surface and internal properties.

Nevertheless, the complex mixing nature of the signals obtained by the radar sensor (due to these above-mentioned multiple factors that are contributing to the final, detected, and signal) make them have non-smooth and complex shapes, which contribute with an additional complexity degree to a potential signal classification strategy.

The aim of Section 2 is to present a coherent explanation of the different feature extraction techniques and classification strategies applied on a series of datasets obtained by Yeo et al. [18], using the Soli radar sensor. It is composed of the following subsections: Section 2.1 gives a detailed overview about the different types of datasets used in our classification performance analysis framework. Section 2.2 explains the feature extraction methods that are applied. Feature selection methods are considered taking into account their number in relation to the number of instances of each dataset (the so-called curse of dimensionality). Section 2.3 describes the different classification strategies applied.

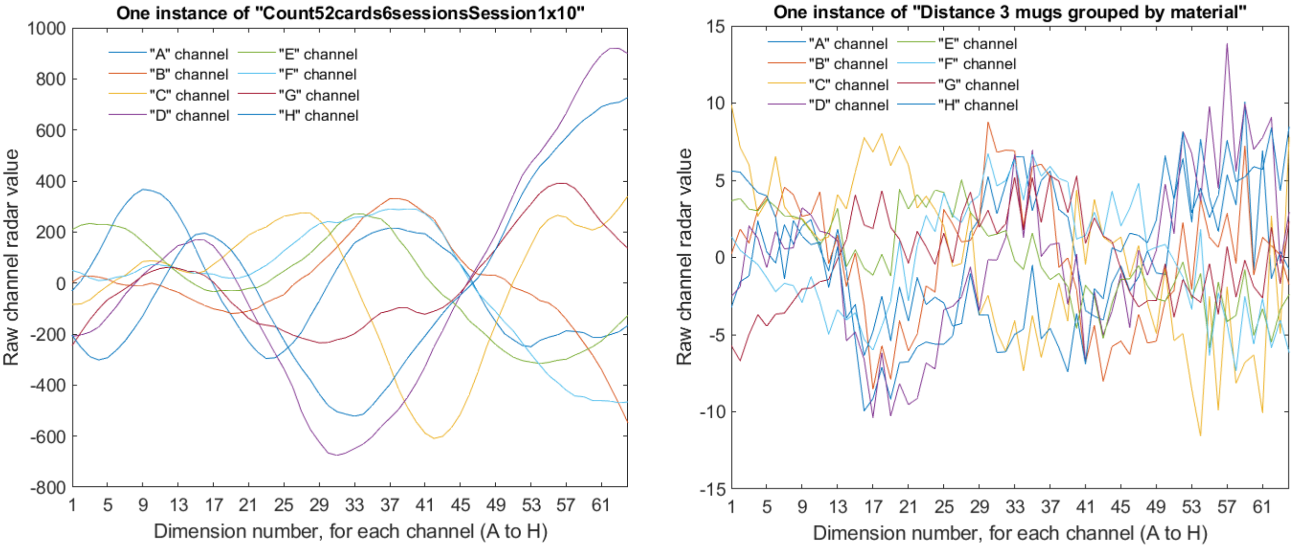
2.1. Radar Datasets

A previous study [18] showed that Soli could be used to identify materials, estimate number of objects, their orientation, or the the distance to the sensor. Yeo et al. [18] created a series of supervised classification datasets, formed by on the one hand the signal acquired by this sensor, and the class of the material on the other hand. The type of material, distance, and other features were also included. The following is a brief schematic description about the different types of categories these datasets are formed by:

- Material identification (of the material an object is made of), from a limited list of materials.
- Object identification, from a series of objects in the same category (i.e., a credit card of a specific bank from the rest of the credit cards).
- Counting the number of elements that might be piled up on the surface of the sensor.
- Distance estimation, from an object to the sensor.
- Order identification, of the items in a battery of objects.
- Flipping identification, where the orientation of an object may be inferred.
- Movement, where the angular position of the object or changes in the movement of one of them in relation to other(s) are obtained.

The complete group of datasets can be found at: https://github.com/tcboy88/solinteractiondata (accessed on 18 January 2021). As stated above, the radar chip has eight channels that acquire a series of raw signals. Each of these channels has 64 data points. This information is converted into a 512 (=64 × 8) feature vector and collectively saved in a CSV file. Therefore, the original dataset dimensionality value is always constant and equal to 512. As recommended by Yeo et al. [18], the easiest way to use the data was considering the Weka Graphical User Interface (GUI), converting them to Attribute-Relation File Format (i.e., .arff) files. This final group of datasets is formed by 34 files.

Figure 1a,b shows the eight 64D signals obtained by the radar sensor for two instances, each from a different dataset. We can see the complex and irregular shape of the signals detected by the sensor.



(**a**) (**b**)

Figure 1. Plots of the eight 64D signals (one 64D signal per channel) for two instances, each of a different dataset.

Table 1 shows the total number of samples and classes for each one of the files. These datasets show different acquisition conditions over a series of distinct objects, including playing cards, poker chips, and Lego blocks. For particular details and a deeper description about the acquisition conditions for the creation of these sets, the reader is referred to Section 6 in [18].

Table 1. Columns show radar datasets used, including the problem type (C, Counting; M, Material Identification; D, Distance estimation; O, Order identification; F, Flipping identification (Up/Down)); I, Object identification; R, Movement, rotation; P, Movement, position), the total number of samples, and the number of classes for each dataset.

	Dataset	**Problem**	**# of Samples**	**# of Classes**
1	Count + Order Lego	C	106	11
2	Count 20 chips NO case x30 sorted	C	629	21
3	Count 20 chips WITH case x30 sorted	C	630	21
4	Count 20 papers x10	C	210	21
5	Distance 3 mugs 10 distances	M + D	93	31
6	Distance 3 mugs grouped by material	M	93	4
7	Distance 7 slotting	D	60	8
8	Flip 10 creditcards NO case	F + I	220	21
9	Flip 10 creditcards WITH case	F + I	220	21
10	Flip 52 cards	F	105	3
11	Identify 10 creditcards NO case	I	110	11
12	Identify 10 creditcards WITH case	I	110	11
13	Identify 12 printed designs	I	130	13
14	Identify 12 touch on numpad	I	780	13
15	Identify 5 colors × 20 chips	I	240	6
16	Identify 6 tagged plastic cards	I	70	7
17	Identify 6 users by palm	I	210	7
18	Identify 6 users by touch behavior sorted	I	780	7

Table 1. *Cont.*

	Dataset	Problem	# of Samples	# of Classes
19	Identify 7 dominoes	I	80	8
20	Identify 9 touch on half-sphere	I	600	10
21	Order 3 coasters NO case	O	480	16
22	Order 3 creditcards NO case sorted	O	164	16
23	Order 4 creditcards NO case sorted	O	672	65
24	Order 4 creditcards WITH case sorted	O	390	65
25	Rotation interval half numeric	R	120	12
26	Rotation interval one numeric	R	120	12
27	Slide inside-out on desk surface	P	110	11
28	Slide outside-in ruler	P	110	11
29	session1 × 10	C	530	53
30	session2 × 10	C	530	53
31	session3 × 10	C	530	53
32	session4 × 10	C	530	53
33	session5 × 10	C	530	53
34	session6 × 10	C	530	53

2.2. Feature Extraction

The most straightforward way to apply a classification strategy on this dataset would be to use the complete 512-dimensional vector (raw data) as the feature vector. However, raw data often contain noise, redundancies, or irrelevant information; thus, there are different feature selection/extraction techniques that could be applied on such a feature space and subsequently be used instead or added to this feature vector. The following additional feature extraction techniques are considered in our case:

- **Basic aggregation features**: A series of basic features are extracted. They are defined as aggregations at different levels:
  - Along the signals of the eight channels: Average (AVG) and the average of the absolute values (ABS). There are 64 values in each case and 128 (=64 × 2) values in total.
  - For each channel (×8): Absolute mean square (AMS) and root mean square (RMS). There are 8 for each channel and 16 values in total.
  - At a global level: Maximum, minimum, mean, AVG, average of absolute values (ABS), and root means square (RMS). There are five values in total.
- **ROCKET** (RandOm Convolutional KErnel Transform [23]): It is a method that achieves state-of-the-art classification accuracy in several benchmarks, but it only requires a fraction of the training time used by other existing standard methods. Other methods for time series classification focus only on a single type of representation such as shape, frequency, or signal variance. Convolutional kernels are able to represent multiple characteristics of different types simultaneously.
- **TSFRESH** (Time Series FeatuRe Extraction based on Scalable Hypothesis tests): This methodology aims to avoid the time-consuming process of meaningful feature identification and extraction, from time series data. It consists of an algorithm and a Python package. The Python package implements the extraction of 794 features using multiple characterization methods, each of them executed with various sets of parameters. This framework also includes feature selection to identify those features that are statistically significant. The algorithm is described in [32]. The software package is presented in [24]. The total number of features is therefore 6352 (=8 × 794).

In this framework, each feature vector is individually analyzed in relation to its significance for predicting the target class label. As a result, a vector of p-values is obtained. This vector is assessed against the Benjamini–Yekutieli procedure [33], which allows the method to decide which features to keep. This selection strategy considers the application of a threshold, whose default value is 0.05. However, this selection is so restrictive in some cases that most of the features are discarded. An iterative method is applied that considers the threshold values $\{0.05, 0.1, 0.2, 0.5\}$ and selects the first one that keeps at least a 10% of the total number of attributes.

2.3. Classification Methods

Two main blocks of classification strategies were considered to be applied on the datasets, summarized as the use of: (a) a Support Vector Machine and a Random Forest classifier, being these two classifiers those that were used in [18] (presented in Section 2.3.1); (b) a Stacked Generalization approach (a particular type of ensemble machine learning algorithm) (Section 2.3.2), where we try to take advantage of the diversity that different classification methods may give, in order to help improve classification accuracy, when using them in a coordinated/simultaneous way.

2.3.1. Single Type Classifiers

Two classification methodologies were applied: (1) Support Vector Machines (SVM); (2) Random Forests (RFs). SVM is a widely used classification method, in different and varied areas of research, partly because of its capability and good behavior when dealing with problems with a small number of samples in (relation to) high-dimensional feature spaces. Originally developed and applied in linearly separable problems, it is aimed at obtaining the hyperplane whose distance to the two groups of data points (called margin), representing the two classes, was maximal. SVM was generalized later to deal with nonlinearly separable problems using the so-called (transformational) Kernel trick [21]. A mathematical transformation function is applied to map the nonlinear separable dataset into a higher dimensional space where the samples can be linearly separated using an hyperplane.

Under this mathematical framework, two parameters emerge (C, γ). The optimal value of these parameters is problem dependent. A Grid Search strategy was applied to assess their optimal values. The parameter C search interval and step size was $2^{-5}, 2^{-3}, \ldots, 2^{15}$, and the corresponding γ interval and step size was 2^{-15}, $2^{-13}, \ldots$, 2^{3}. Whenever the best best pair of parameters was obtained in the interval limits, the Grid Search was automatically extended by a factor of two. This procedure follows the guidelines given in [34].

Ensemble learning methods are based on the idea of using multiple learning algorithms to obtain better predictive performance than one could obtain from any of them, separately. The idea behind ensemble learning is the way an expert committee works in real life, i.e., it is usually easier to properly predict something when the prediction is made by more than one expert, and a consensus is obtained from them. Ensembles are combinations of several classifiers, which are often called base classifiers. There are several types of ensembles, which may be divided into two large groups: (a) homogeneous ensembles, where all the base classifiers are built using the same algorithm (but with different versions of the dataset or different training parameters); (b) heterogeneous ensembles, where the base classifiers are built using different algorithms.

Diversity is a key property in the search for an optimal ensemble strategy performance, since there is no benefit when combining base classifiers that always obtain the same predictions. There are several techniques to induce diversity in homogeneous ensembles. In Bagging [35], for instance, each classifier is trained with a variant of the training dataset, which uses different random samples of the training set. Random Forests [20] are ensembles of Decision Trees [36]. In this method, the diversity during the training process is enforced by combining the sampling of the training set, as Bagging does, with the random selection of subsets of attributes in each node of the tree. This way, in each node, the splits only consider the selected subset of attributes. Later, on the prediction stage, each base classifier predicts a class, and the class selected the most (the mode) is the final prediction of the ensemble. RFs are used to correct the tendency of the decision trees to overfit. The main parameter of an RF is its size (i.e., the number of trees that are generated into the ensemble). In our study, 100 decision trees were used because it is a usual and sufficient number [37].

2.3.2. Stacked Generalization

Stacked Generalization is an ensemble method where a new model learns how to best combine the predictions from multiple existing models. In this approach, any learning algorithm could be used to combine them.

In particular, the generation of classifiers that are accurate and diverse is only the first part in an ensemble classifier generation process. The second part (as important as the former one) is the method used to obtain the ensemble outputs by combining the outputs of the base classifiers. Two of the approaches used to combine the outputs of the base classifiers are: (a) majority voting; (b) average of probabilities. Alternatively, there are methods that may be able to learn the so-called combination rules. These methods (called meta-classifiers) are particularly useful when the base classifiers do not have the same success rate (among them) when classifying instances. This may happen when the base classifiers are generated using different training sets or different training algorithms.

Stacked Generalization (also called Stacking) [22] builds a classifier that takes as inputs values, the output values of the base classifiers, and learns to map these values of the base classifiers into the correct final output value. In other words, no voting strategy is applied in order to combine the predictions of the base classifiers. In this case, a meta-classifier is used. The base classifiers are trained with the training set and the meta-classifier is trained with the predictions of the base classifiers. Figure 2 shows a scheme of one of the stacking approaches used.

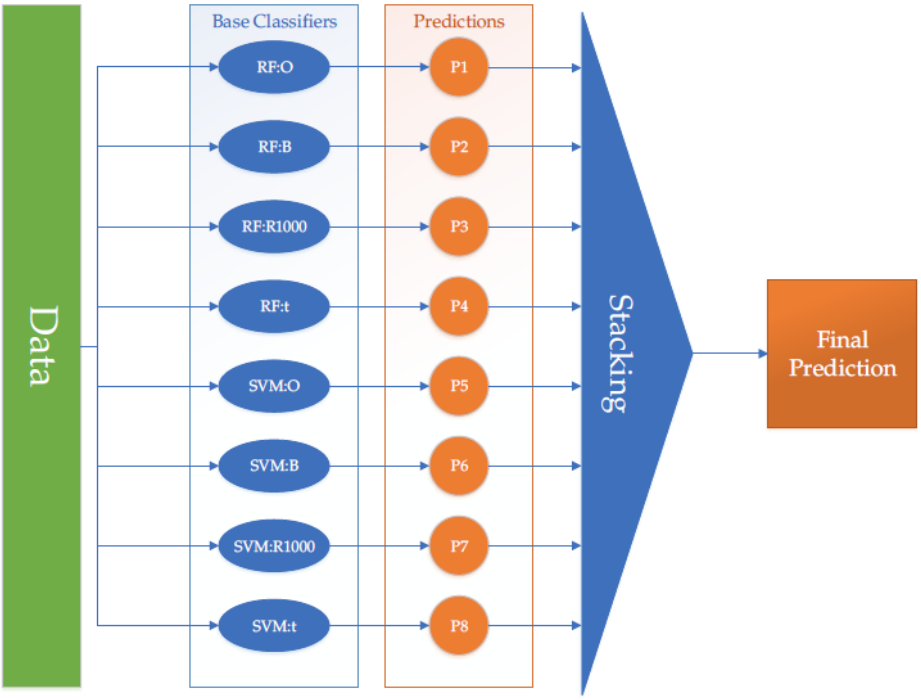
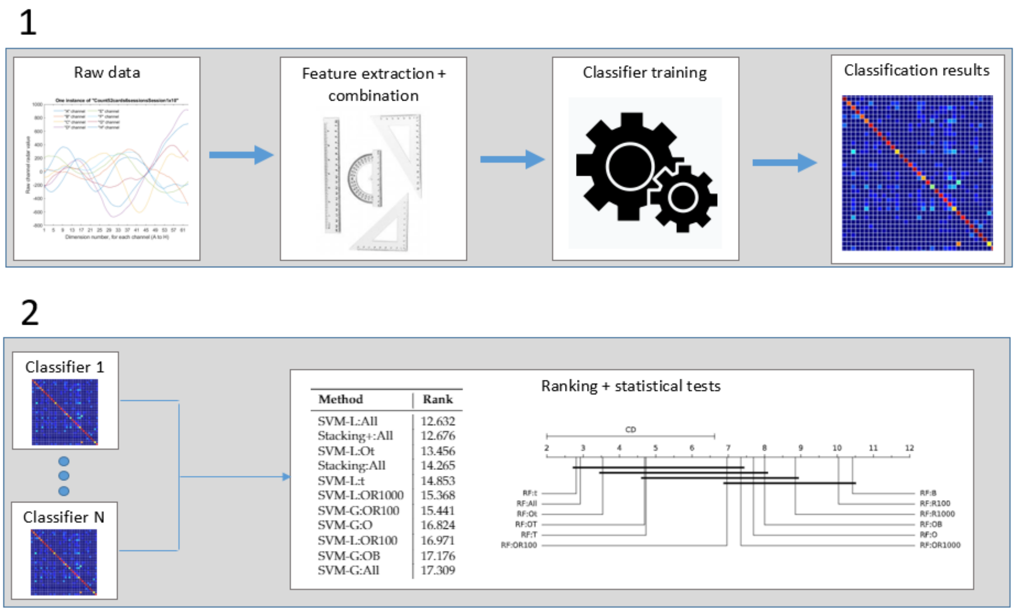


Figure 2. Diagram of the Stacked Generalization approach.

The predictions of the classifiers are obtained from a different partition set than the one used for training. This is achieved by dividing the training set into several partitions. Therefore, Stacking can also be seen as a sophisticated form of attribute extraction. Stacking base classifiers are usually trained using different algorithms. Another strategy would be to use different views, i.e., subsets of attributes obtained by each feature extraction method. This type of Stacking strategy is often called multi-view stacking and has been successfully used when applied to other (but somehow similar) problems [38–40].

Figure 3 shows a flow diagram of the general data processing strategy followed in the paper. The figure shows the processing chain divided into two parts. The first of them shows that different types of features and feature combination strategies are obtained from the raw data values, and different classifiers are trained and used to obtain the classification accuracy results for the different datasets. In the second part, these classification accuracy results, for each dataset, are ranked, and the statistical significance of their differences are obtained in order to assess which method is better, and whether the differences among classification strategies are statistically significant or not.



Method	Rank
SVM-L:All	12.632
Stacking+:All	12.676
SVM-L:Ot	13.456
Stacking:All	14.265
SVM-L:t	14.853
SVM-L:OR1000	15.368
SVM-G:OR100	15.441
SVM-G:O	16.824
SVM-L:OR100	16.971
SVM-G:OB	17.176
SVM-G:All	17.309

Figure 3. Flow chart describing the general data processing pipeline, divided into two steps.

3. Results and Discussion

Accuracy (for the Random Forest, SVM Linear with default parameters (rhe same configuration as the one used in [18]), and the optimized Gaussian SVM classifiers) was assessed using different combinations of the attributes explained in Section 2.2. Table 2 shows these combination strategies: The symbol '&' means that the referred attributes are concatenated. Stacking was applied using two different configurations, called as follows:

- **Stacking:All**: Eight base classifiers were assembled (four RF classifiers and four SVM linear classifiers), one for each one of the four extracted feature sets (raw features (O), basic aggregation features (B), ROCKET with 1000 kernels (R1000), and TSFRESH with feature selection (t)).
- **Stacking+:All**: Ten base classifiers were assembled: the eight base classifiers described in the previous case, a linear SVM, and an RF, trained in both cases with the concatenation of all the attributes.

In all the results that follow, we use SVM-L (SVM with Linear Kernel) to refer to SVM with default parameters and SVM-G to refer to Grid search-optimized SVM with RBF Kernel. We simply use RF for the Random Forest classifier. Features used for training the classifiers use the same abbreviation as that shown in Table 2.

In order to compare the performance of the different classification methods, we might use the average accuracy of each one of the pairs (Classifier:Feature Set) evaluated throughout the 34 datasets shown in Table 1. Nevertheless, when comparing multiple methods on multiple datasets, an alternative to (and sometimes more appropriate way than) comparing average accuracies is to use average ranks [41]. Average ranks are computed in the following way: For a given dataset, the methods (in this case, a method is a pair formed by Classifier and Feature set) are sorted from best to worst. The best method receives rank = 1, the second best receives rank = 2, etc. In the case of a tie, average ranks are assigned. For instance, if two methods tie for the top rank, they both receive rank = 1.5. The average ranks across all the datasets are then computed for each method.

Table 2. Combination of features used, including their total number. In the case of variable-size feature extraction methods, the minimum, maximum, and average values, respectively, are also given (in parenthesis).

Features	Abbreviation	N° of Attributes
Raw Original Data	(**O**)	512
Basic aggregation features	(**B**)	149
ROCKET with 100 kernels	(**R100**)	200
ROCKET with 1000 kernels	(**R1000**)	2000
TSFRESH without feature selection	(**T**)	6352
TSFRESH with feature selection	(**t**)	Variable (885/3567/3235)
Original & Basic	(**OB**)	661
Original & ROCKET (100)	(**OR100**)	712
Original & ROCKET (1000)	(**OR1000**)	2512
Original & TSFRESH (without feat. sel.)	(**OT**)	6864
Original & TSFRESH (with feat. sel.)	(**Ot**)	Variable (1397/4079/3747)
Original, basic, ROCKET (1000) and TSFRESH (with feat. sel.)	(**All**)	Variable (3546/6228/5896)

Post hoc tests were applied in order to identify statistically significant differences among the performance results. Some of these tests are strict in the conclusions that might be obtained from them. We found that the classification accuracy results from some of the 34 datasets are substantially high, for a considerable number of methods. Therefore, aiming at inferring the classifiers and attributes that work best with the hardest datasets (i.e., those that are more interesting), the statistical comparison of the methods was carried out twice: first using all the datasets and then using the subset formed by the difficult ones (the division between easy and difficult datasets was determined based on the performance of a baseline classifier).

Post hoc tests based on mean-ranks are commonly used, but their application has been questioned recently [42]. Hence, the results are also compared with the Bayesian Signed-Rank Test [43].

3.1. Results Corresponding to All the Datasets

Table 3 shows a selection of the results, for a subset of the pairs (Classifier:Feature Set). Given the large number of pairs, it is not possible to include all the methods in a single table (Tables A1–A4 (in the Appendix A) show the complete set of results for the RF, SVM-L, and SVM-G classifiers and the Stacking strategy, respectively, considering all the datasets). The pairs in the subset were selected so that for all data sets there was some method with the highest accuracy. In some datasets, many methods share the best accuracy, so it is not possible to include all of them in the subset. Therefore, the subset of pairs was further reduced according to the average accuracy across all the datasets. Moreover, the pairs with the feature set OB were also included because it was used by Yeo et al. [18].

Table 3. Results for a subset of the classifiers with different feature sets.

	RF **:OB**	**RF** **:Ot**	**SVM-L** **:R1000**	**SVM-L** **:t**	**SVM-L** **:OB**	**SVM-L** **:OR1000**	**SVM-L** **:Ot**	**SVM-L** **:All**	**SVM-G** **:R1000**	**SVM-G** **:OB**	**SVM-G** **:OR100**	**Stack** **:All**	**Stack+** **:All**
Count + Order Lego	96.09	97.09	98.00	96.00	**99.00**	98.00	96.00	97.00	98.00	98.00	98.00	98.00	98.00
Count 20 chips NO case ×30 sorted	63.60	66.95	71.88	72.35	62.49	73.78	71.87	73.94	73.62	71.23	73.94	**77.60**	76.17
Count 20 chips WITH case ×30 sorted	70.32	77.14	78.89	79.68	79.52	81.27	80.00	80.63	80.63	79.05	**83.33**	80.95	80.79
Count 20 papers ×10	75.71	75.24	**91.43**	84.29	87.14	89.52	84.76	86.19	90.00	87.62	88.10	83.33	83.81
Distance 3 mugs 10 distances	47.22	64.44	80.56	**88.00**	47.56	75.11	72.22	77.33	80.56	50.78	62.56	64.78	63.67
Distance 3 mugs grouped by material	84.33	**100.00**	93.44	**100.00**	80.89	90.33	**100.00**	**100.00**	92.44	80.89	94.67	99.00	**100.00**
Distance 7 slotting	**100.00**	**100.00**	**100.00**	98.33	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
Flip 10 creditcards NO case	83.64	86.36	84.09	88.18	82.73	85.91	**89.55**	88.64	85.45	88.18	85.00	87.73	86.82
Flip 10 creditcards WITH case	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**

Table 3. *Cont.*

	RF :OB	RF :Ot	SVM-L :R1000	SVM-L :t	SVM-L :OB	SVM-L :OR1000	SVM-L :Ot	SVM-L :All	SVM-G :R1000	SVM-G :OB	SVM-G :OR100	Stack :All	Stack+ :All
Flip 52 cards	96.18	98.18	98.18	**99.09**	96.36	98.18	**99.09**	98.18	98.18	98.18	97.18	98.18	98.18
Identify 10 creditcards NO case	87.27	**91.82**	81.82	90.00	83.64	84.55	90.00	86.36	83.64	84.55	88.18	88.18	88.18
Identify 10 creditcards WITH case	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
Identify 12 printed designs	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	99.23	**100.00**	**100.00**	**100.00**
Identify 12 touch on numpad	98.33	97.82	97.95	98.59	98.97	98.72	98.59	98.59	98.21	**99.23**	98.85	98.97	99.10
Identify 5 colors × 20 chips	75.42	83.33	85.83	**96.67**	87.50	88.33	**96.67**	**96.67**	85.00	88.33	87.08	94.58	96.25
Identify 6 tagged plastic cards	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
Identify 6 users by palm	82.86	90.95	90.00	95.24	90.95	91.43	94.29	**95.24**	89.52	91.90	92.38	94.29	**95.24**
Identify 6 users by touch behavior sorted	82.56	96.03	85.64	98.72	85.38	88.72	98.72	98.59	89.10	91.41	91.92	98.46	**98.85**
Identify 7 dominoes	98.75	98.75	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
Identify 9 touch on half-sphere	99.33	98.67	99.50	99.00	98.67	**99.67**	99.00	99.33	99.50	98.50	99.00	98.83	99.00
Order 3 coasters NO case	79.58	83.13	80.00	85.21	81.04	83.12	85.21	**86.46**	80.63	85.00	83.54	82.08	83.54
Order 3 creditcards NO case sorted	82.39	85.99	84.12	89.01	86.58	87.87	**90.26**	89.63	84.15	89.63	87.79	88.38	89.01
Order 4 creditcards NO case sorted	55.51	56.10	58.64	65.91	61.16	60.72	**67.40**	65.47	58.04	63.40	63.09	64.43	65.47
Order 4 creditcards WITH case sorted	**99.49**	**99.49**	99.23	**99.49**	**99.49**	99.23	**99.49**	**99.49**	99.23	99.23	99.23	**99.49**	**99.49**
Rotation interval half numeric	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	99.17	**100.00**	**100.00**
Rotation interval one numeric	99.17	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	99.17	**100.00**	**100.00**	**100.00**	**100.00**
Slide inside-out on desk surface	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	99.09	**100.00**	**100.00**	**100.00**	**100.00**
Slide outside-in ruler	**100.00**	**100.00**	**100.00**	99.09	**100.00**	**100.00**	99.09	99.09	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
session1 × 10	99.81	99.81	**100.00**	99.81	99.81	99.81	99.81	99.81	**100.00**	99.81	99.81	99.81	99.81
session2 × 10	**99.62**	99.43	99.25	99.06	**99.62**	99.25	99.43	99.43	99.06	99.25	**99.62**	99.43	99.43
session3 × 10	99.62	**99.81**	**99.81**	99.43	**99.81**	**99.81**	99.62	**99.81**	**99.81**	**99.81**	**99.81**	**99.81**	**99.81**
session4 × 10	**100.00**	**100.00**	**100.00**	99.43	**100.00**	**100.00**	99.43	99.43	**100.00**	**100.00**	**100.00**	99.81	99.81
session5 × 10	97.92	98.49	**99.25**	**99.25**	98.49	**99.25**	**99.25**	**99.25**	98.87	98.30	99.06	98.49	98.87
session6 × 10	**99.25**	**99.25**	99.06	**99.25**	**99.25**	99.06	**99.25**	**99.25**	99.06	**99.25**	**99.25**	**99.25**	**99.25**
Mean	89.82	92.48	92.84	**94.68**	91.35	93.28	94.38	94.52	92.97	92.38	93.25	93.94	94.07

Table 4 presents the average accuracy of each one of the pairs (Classifier:Feature Set) assessed throughout the 34 datasets. It also shows the average ranks computed using all the methods in the experimental setup. In terms of average accuracies, the best results obtained by Shyong Yeo et al. [18] appear in the lower third of the table; SVM-L:OB (Linear SVM trained using the concatenation of Raw and Basic features) achieves an average accuracy of 91.36%. The same classifier, when trained using all features or TSFRESH with feature selection, obtains an accuracy higher than 94.5%. In terms of average ranks, the two pairs with top ranks are SVM-L:All and Stacking+:All, with average ranks below 12.7. The average rank for SVM-L:OB is 18.13.

Table 4. Average accuracies and ranks from all the datasets.

Method	Accuracy	Method	Rank
SVM-L:t	94.678	SVM-L:All	12.632
SVM-L:All	94.524	Stacking+:All	12.676
SVM-G:t	94.466	SVM-L:Ot	13.456
SVM-L:Ot	94.382	Stacking:All	14.265
SVM-G:All	94.188	SVM-L:t	14.853
Stacking+:All	94.075	SVM-L:OR1000	15.368
SVM-G:Ot	94.025	SVM-G:OR100	15.441
Stacking:All	93.938	SVM-G:O	16.824
SVM-G:T	93.359	SVM-L:OR100	16.971
SVM-G:OT	93.338	SVM-G:OB	17.176
SVM-L:OR1000	93.283	SVM-G:All	17.309
SVM-L:OT	93.283	SVM-L:R1000	17.338

Table 4. *Cont.*

Method	Accuracy	Method	Rank
SVM-G:OR100	93.252	SVM-G:OR1000	17.588
SVM-L:T	93.205	SVM-G:Ot	17.662
SVM-G:OR1000	93.070	RF:t	17.794
SVM-G:R1000	92.969	SVM-G:t	17.853
RF:t	92.926	SVM-L:OB	18.132
SVM-L:R1000	92.840	RF:All	18.176
RF:All	92.720	SVM-G:R1000	18.265
RF:Ot	92.479	RF:Ot	18.353
SVM-G:OB	92.375	RF:T	19.279
SVM-L:OR100	92.141	SVM-L:O	19.603
RF:OT	92.100	SVM-L:OT	20.176
RF:T	92.028	RF:OT	20.426
SVM-G:O	91.987	SVM-L:R100	21.044
SVM-L:OB	91.355	SVM-G:OT	21.103
SVM-G:R100	90.877	SVM-G:T	21.382
SVM-L:R100	90.712	SVM-L:T	21.544
RF:OR1000	90.706	RF:OR1000	21.779
RF:OR100	90.110	SVM-G:R100	22.765
SVM-G:B	90.100	RF:OR100	22.897
SVM-L:O	89.989	RF:OB	23.176
RF:OB	89.823	RF:R1000	23.338
RF:R1000	89.761	RF:O	24.706
RF:O	89.333	RF:R100	25.971
RF:R100	88.747	SVM-G:B	26.603
RF:B	87.508	SVM-L:B	28.088
SVM-L:B	87.068	RF:B	28.985

The best five pairs according to the average accuracy and rank (in Table 4) use the feature sets (t), (All), and (Ot). Table 5 summarizes the results in Table 4 averaging for each feature set the corresponding values of RF, SVM-L and SVM-G. According to both the average accuracies and ranks, the three best feature sets are (t), (All), and (Ot).

Table 5. Average accuracies and ranks for each feature set, from all the datasets. For each feature set, the values in these tables are the averages for RF, SVM-L, and SVM-G in Table 4.

Feature Set	Accuracy	Feature Set	Rank
t	94.024	All	16.039
All	93.811	Ot	16.490
Ot	93.628	t	16.833
OT	92.907	OR1000	18.245
T	92.864	OR100	18.436
OR1000	92.353	OB	19.495
R1000	91.857	R1000	19.647
OR100	91.834	O	20.377
OB	91.184	OT	20.569
O	90.437	T	20.735
R100	90.112	R100	23.260
B	88.226	B	27.892

Figure 4 shows, for each one of the three classification methods, the differences in terms of accuracy between each feature set and the feature set (OB) used by Yeo et al. [18]. Each boxplot is from the corresponding differences from the 34 datasets. The average differences are clearly favorable for several of the alternative feature sets. Medians of the differences are close to 0 or negative. As shown in Tables A1–A4, there are several datasets with 100% accuracy for all or many of the feature sets. For several feature sets, the boxplot are mostly in the positive region, positive differences are greater than negative differences.

For SVM-G the boxplots are less favorable to the alternative feature sets, but, as shown in Table 4, SVM-L has better results than SVM-G.

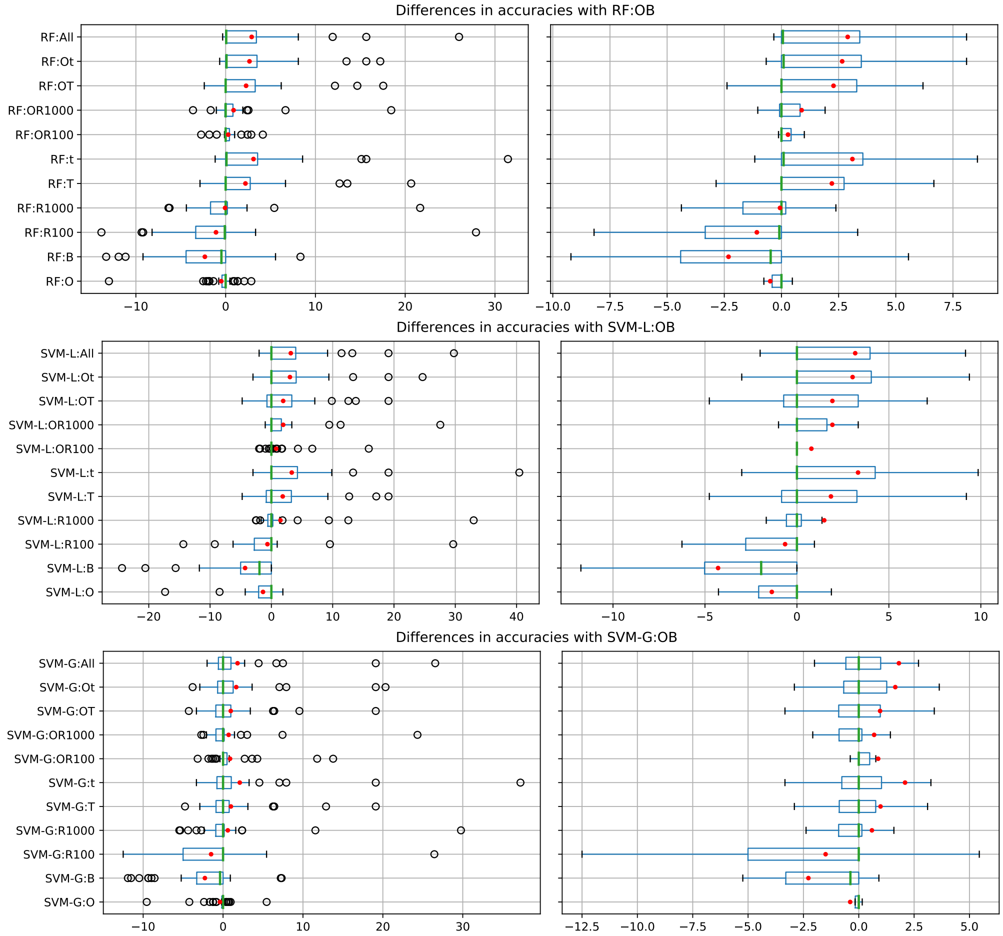


Figure 4. Boxplots of the differences in accuracy between the different feature sets and the feature set (OB), for the three considered classifiers. The boxplots on the right do not include the outliers. The average differences are marked with a red dot (•).

Given the large number of (Classifier:Feature Set) tested methods, it is preferable to obtain the average ranks by considering smaller pair groups. Therefore, methods were divided depending on the type of classifier used: RF, SVM-L, and SVM-G. These average ranks are shown in Table 6.

Table 6. Average ranks for each classification method.

RF		SVM-L		SVM-G	
Method	**Rank**	**Method**	**Rank**	**Method**	**Rank**
RF:t	4.676	SVM-L:All	4.765	SVM-G:OR100	5.397
RF:All	4.691	SVM-L:Ot	5.059	SVM-G:Ot	5.647
RF:Ot	4.941	SVM-L:OR1000	5.515	SVM-G:All	5.647
RF:T	5.500	SVM-L:t	5.529	SVM-G:t	5.868
RF:OT	5.750	SVM-L:OR100	6.000	SVM-G:O	5.941
RF:OR1000	6.456	SVM-L:R1000	6.294	SVM-G:OB	5.941
RF:OR100	6.500	SVM-L:OB	6.412	SVM-G:OR1000	6.132
RF:OB	7.015	SVM-L:O	6.956	SVM-G:R1000	6.471
RF:R1000	7.162	SVM-L:OT	6.956	SVM-G:OT	6.985
RF:O	7.515	SVM-L:R100	7.426	SVM-G:T	7.265
RF:R100	8.412	SVM-L:T	7.662	SVM-G:R100	7.735
RF:B	9.382	SVM-L:B	9.426	SVM-G:B	8.971

The use of the Nemenyi test [44] was also proposed by Demšar [41] to compare methods in a pairwise way. For a certain level of confidence (α), the test determines a critical difference (CD) value. If the difference between the average rankings of two methods is greater than CD, the null hypothesis, H_0, that both methods have equal performance, is rejected. Figure 5 shows the Nemenyi's CD diagrams. In these diagrams, thick horizontal lines are used to connect methods whose difference in average ranks is smaller than CD.

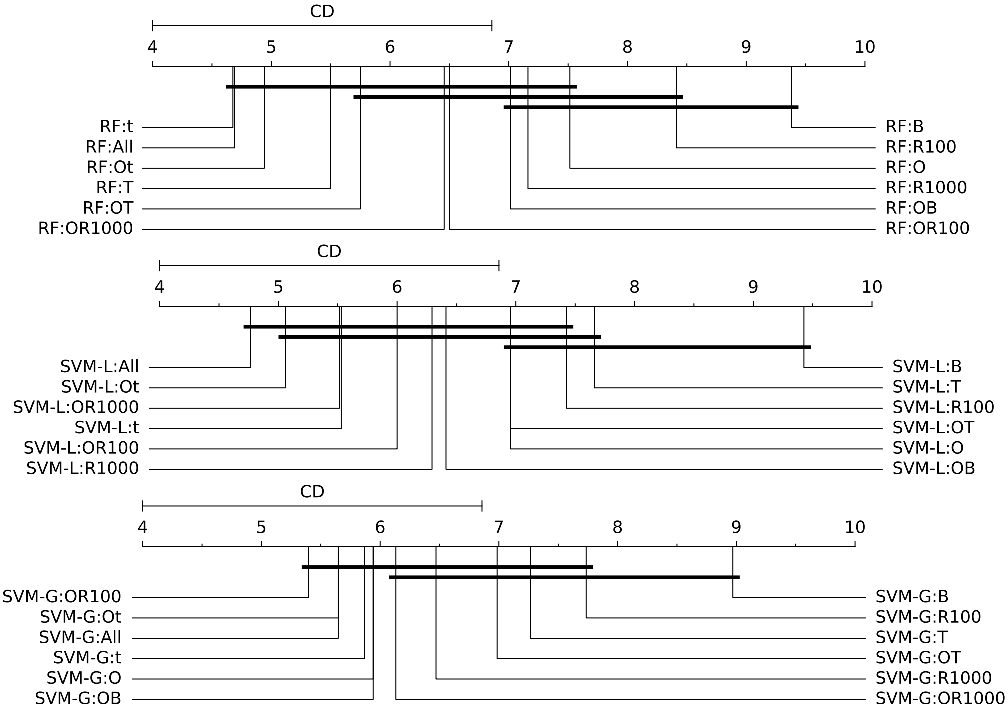


Figure 5. Critical difference diagrams for the Nemenyi test ($\alpha = 0.05$).

The methods were also compared using the Bayesian Signed-Rank Test [43], the Bayesian framework equivalent version of the Wilcoxon Signed-Rank Test. In this test, the value of the Region of Practical Equivalence (ROPE) was set to 1% for accuracy. Two methods were considered equivalent when the difference in their performance was smaller than this ROPE value. The test determines three probability values, corresponding to the following cases: (1) one method is better than the other; (2) vice versa; (3) they are in the ROPE.

Figures 6–8 show the Bayesian Signed-Rank Tests posteriors, for the RF, SVM-L, and SVM-G classifiers, respectively. In these figures, the (OB) feature set is compared against each one of the other feature sets, for the corresponding classifier. For each feature set, there is a triangle. In these triangles [43], the bottom-left and bottom-right regions correspond to the case where one method is better than the other or vice versa. The top region represents the case where the ROPE is more probable. The corner triangles show the probability of each region. The left region in the triangle is for OB and the right region for the other feature set.

Figure 6 shows that, for RF, the feature set with more favorable results when compared to (OB) is (All), with a probability of 0.767, while it is 0.000 for (OB). Figure 7 shows that, for SVM-L, the best feature set is (Ot): its probability is 0.821, while it is 0.000 for (OB). In Figure 8, for SVM-G, the results are less favorable for the alternative cases to (OB). The best feature set is (Ot), with a probability of 0.244, being 0.001 the probability for (OB). The classification results for the three different classifiers therefore show an improvement that can be considered as significant, when using the different types of proposed feature sets, versus the features proposed in [18].

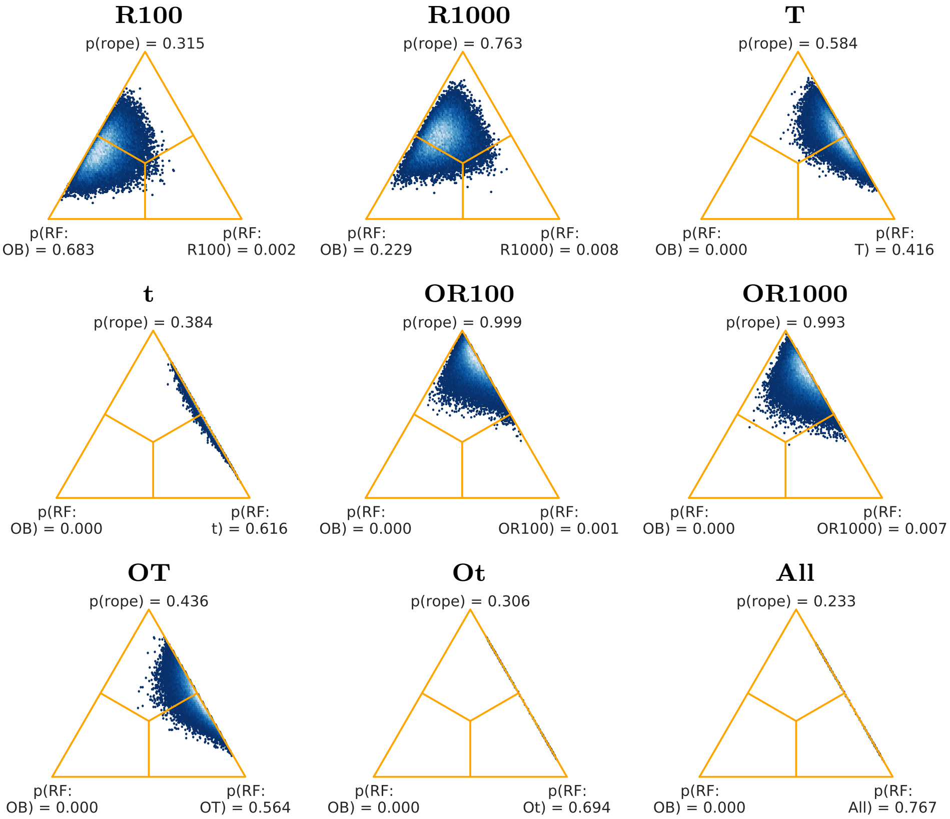


Figure 6. Posteriors for the Bayesian sign-rank tests for RF, from all the datasets.

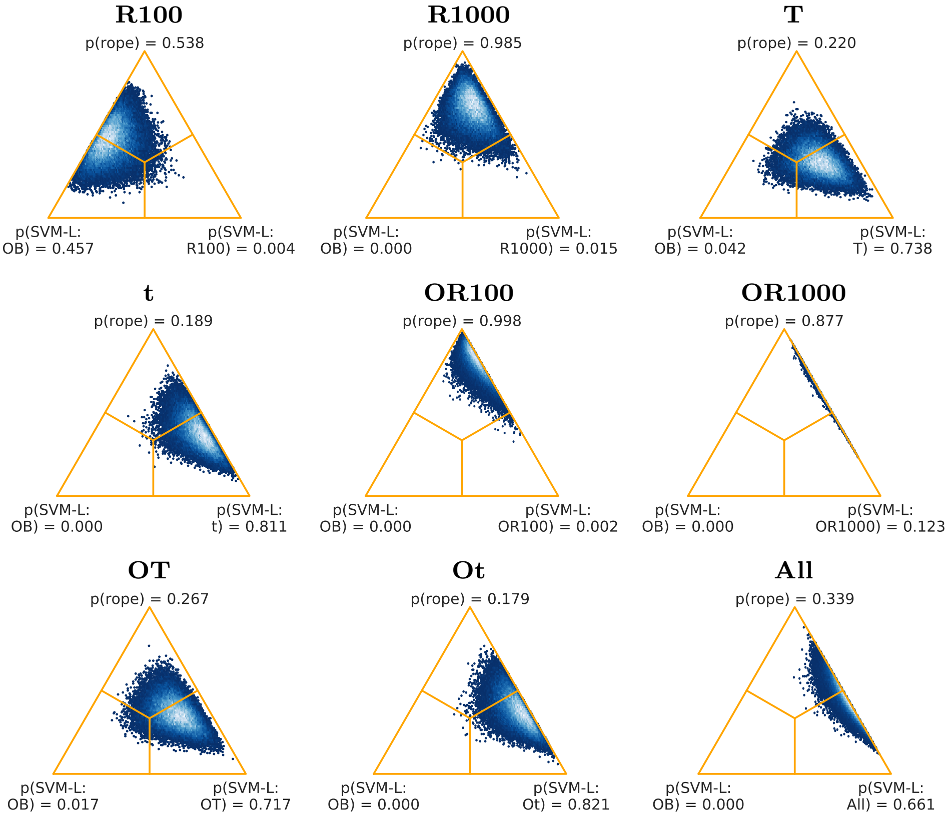


Figure 7. Posteriors for the Bayesian sign-rank tests for SVM-L, from all the datasets.

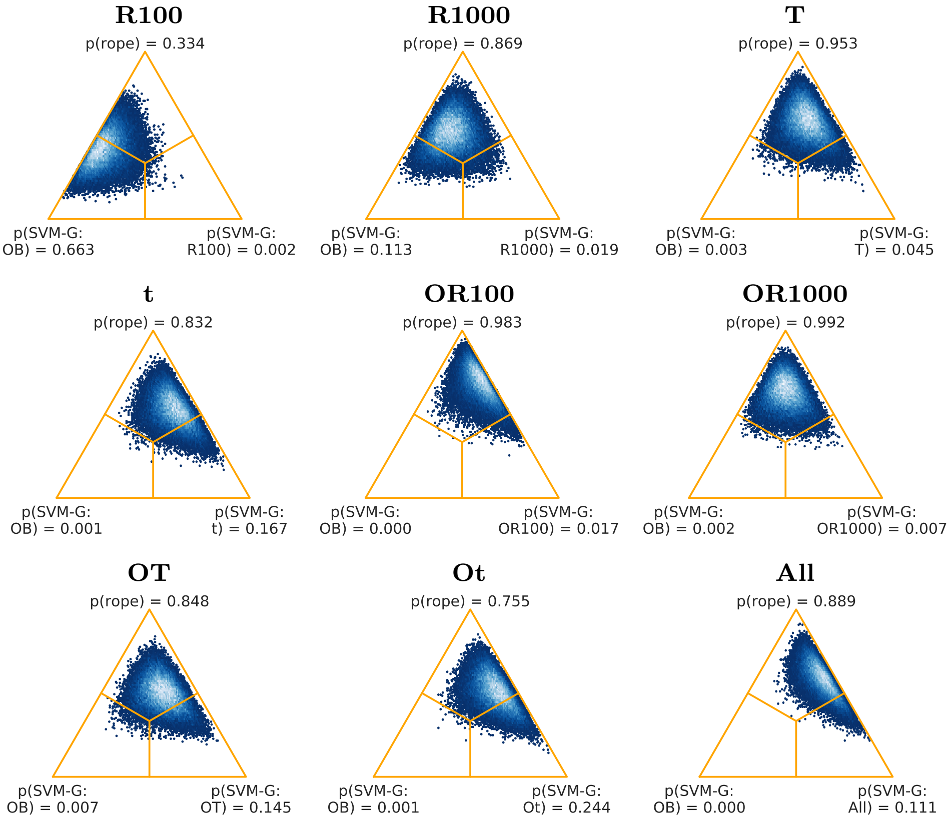


Figure 8. Posteriors for the Bayesian sign-rank tests for SVM-G, from all the datasets.

3.2. Results for the So-Called Difficult Datasets

An important part of the datasets reached a classification accuracy near or equal 100% (Tables A1–A4, in Appendix A) , while others did not even reach 70%. This was the reason we considered splitting up the dataset into two groups, one being formed only by what we may call the difficult datasets, for which the classification accuracy using SVM-L (best classifier in the previous work) was ≤90% using the original set of raw features. The list of difficult datasets (in Table A2) is the following: (1) Count 20 chips NO case x30 sorted; (2) Count 20 chips WITH case x30 sorted; (3) Count 20 papers x10; (4) Distance 3 mugs 10 distances; (5) Distance 3 mugs grouped by material; (6) Flip 10 creditcards NO case; (7) Identify 10 creditcards NO case; (8) Identify 5 colors x 20 chips; (9) Identify 6 users by palm; (10) Identify 6 users by touch behavior sorted; (11) Order 3 coasters NO case; (12) Order 3 creditcards NO case sorted; (13) Order 4 creditcards NO case sorted.

For the other datasets, simple methods may have good accuracy results, with little room for improvement. The entire experimental framework was repeated considering only this (difficult) subgroup of datasets, and the results are as follows.

Table 7 shows the average accuracies and average ranks, obtained using only the difficult datasets. The average accuracy for SVM-L:OB is 78.199 and for SVM-L:t is 87.173. The average rank of SVM-L:OB is 23.115 and 6.654 for SVM-L:All. Table 8 summarizes the results in Table 7 for each feature set, averaging the results of RF, SVM-L and SVM-G. The best feature set is (t), with an average accuracy of 85.58% and an average rank of 10.987. The average accuracy is 77.922 and the rank is 23.090 for (OB).

Table 7. Average accuracies and ranks from the difficult datasets .

Method	Accuracy	Method	Rank
SVM-L:t	87.173	SVM-L:All	6.654
SVM-G:t	87.092	SVM-L:Ot	7.154
SVM-L:All	86.550	SVM-L:t	7.538
SVM-G:All	86.278	SVM-G:t	7.731
SVM-L:Ot	86.226	SVM-G:Ot	7.846
SVM-G:Ot	85.924	SVM-G:All	8.308
Stacking+:All	85.215	Stacking+:All	9.577
Stacking:All	84.908	Stacking:All	11.654
SVM-L:T	84.458	SVM-G:OT	11.808
SVM-G:OT	84.407	SVM-L:OT	12.654
SVM-L:OT	84.319	SVM-L:T	13.077
SVM-G:T	84.297	SVM-G:T	13.500
SVM-G:OR100	83.199	SVM-G:OR100	14.462
SVM-L:OR1000	83.128	SVM-L:OR1000	15.115
SVM-G:OR1000	82.891	SVM-G:OR1000	15.654
SVM-G:R1000	82.522	SVM-G:OB	16.423
RF:t	82.475	SVM-G:O	17.462
SVM-L:R1000	82.026	RF:t	17.692
RF:All	81.967	SVM-G:R1000	18.077
RF:Ot	81.345	RF:All	19.000
SVM-G:OB	80.921	RF:Ot	19.346
RF:OT	80.458	SVM-L:R1000	19.885
RF:T	80.400	SVM-L:OR100	20.808
SVM-L:OR100	80.365	RF:T	20.808
SVM-G:O	79.875	RF:OT	22.115
SVM-L:OB	78.199	SVM-L:OB	23.115
SVM-G:R100	77.307	SVM-L:O	27.115
SVM-L:R100	76.674	SVM-L:R100	27.885
RF:OR1000	76.643	SVM-G:R100	28.077
SVM-G:B	75.809	RF:OR100	28.846
RF:OR100	75.442	SVM-G:B	28.923
SVM-L:O	74.660	RF:OR1000	29.000
RF:OB	74.647	RF:OB	29.731
RF:R1000	74.360	RF:O	30.154
RF:O	73.736	RF:R1000	31.731
RF:R100	72.264	RF:R100	33.538
RF:B	70.128	RF:B	34.231
SVM-L:B	68.541	SVM-L:B	34.308

Table 8. Average accuracies and ranks for each feature set, from the difficult datasets. For each feature set, the values in these tables are the averages for RF, SVM-L, and SVM-G in Table 7.

Feature Set	Accuracy	Feature Set	Rank
t	85.580	t	10.987
All	84.932	All	11.321
Ot	84.498	Ot	11.449
OT	83.062	OT	15.526
T	83.051	T	15.795
OR1000	80.887	OR1000	19.923
OR100	79.669	OR100	21.372
R1000	79.636	OB	23.090
OB	77.922	R1000	23.231
O	76.090	O	24.910
R100	75.415	R100	29.833
B	71.492	B	32.487

Given again the large number of methods (Classifier:Feature Set) tested, they were divided depending on the type of classifier used: RF, SVM-L, and SVM-G. These average ranks (for the difficult datasets) are shown in Table 9.

Table 9. Average ranks for each classification method, for difficult datasets.

RF		SVM-L		SVM-G	
Method	**Rank**	**Method**	**Rank**	**Method**	**Rank**
RF:t	2.808	SVM-L:All	2.923	SVM-G:Ot	3.308
RF:All	2.923	SVM-L:Ot	3.192	SVM-G:t	3.577
RF:Ot	3.538	SVM-L:t	3.346	SVM-G:All	3.808
RF:OT	4.692	SVM-L:OT	4.885	SVM-G:OT	5.000
RF:T	4.731	SVM-L:OR1000	5.231	SVM-G:T	6.000
RF:OR100	6.962	SVM-L:T	5.423	SVM-G:OR100	6.192
RF:OR1000	7.346	SVM-L:R1000	7.154	SVM-G:OB	6.962
RF:O	7.692	SVM-L:OR100	7.231	SVM-G:OR1000	7.000
RF:OB	8.000	SVM-L:OB	7.923	SVM-G:O	7.269
RF:R1000	8.846	SVM-L:O	9.385	SVM-G:R1000	7.692
RF:R100	10.038	SVM-L:R100	9.692	SVM-G:R100	10.308
RF:B	10.423	SVM-L:B	11.615	SVM-G:B	10.885

Figure 9 shows the critical difference diagrams for the Nemenyi test, for the three classifiers. The difference of the average ranks between RF:OB and the best feature sets with RF is greater than the critical difference. The distance of SVM-L:OB to the best feature sets with SVM-L is also greater. Nevertheless, the differences for SVM-G:OB and other feature sets with SVM-G are smaller than the critical difference.

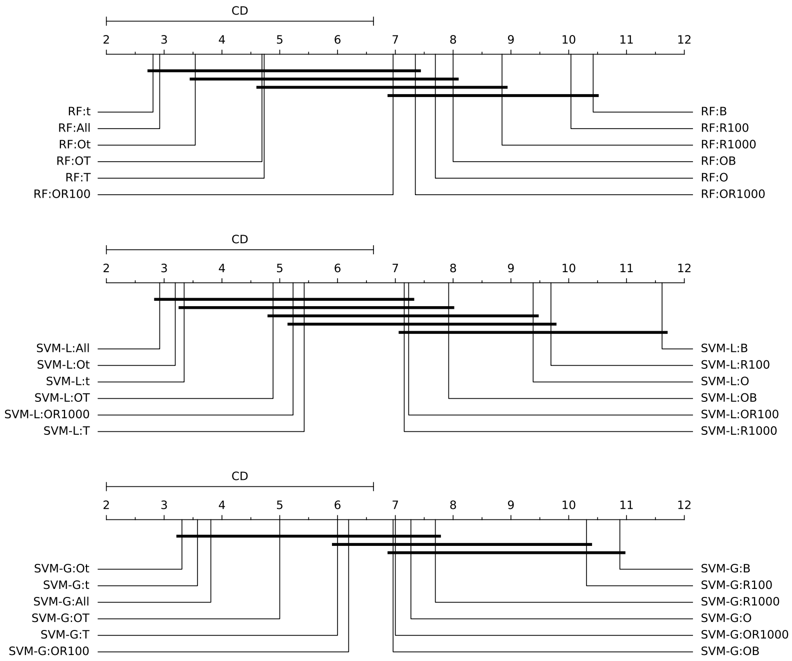


Figure 9. Critical difference diagrams for the Nemenyi test, for the difficult datasets ($\alpha = 0.05$).

Figures 10–12 show the Bayesian Signed-Rank Tests posteriors, for the RF, SVM-L, and SVM-G classifiers, respectively, for the difficult datasets. In Figure 10, for RF, the feature sets with more favorable results when compared to (OB) are (t), (OT), (Ot), and (All), with a probability of 1.000 for the corresponding feature set, and 0.000 for (OB). For SVM-L (Figure 11), the best feature sets are, again, (t), (OT), (Ot), and (All), with a probability of 1.000 for the corresponding feature set, and 0.000 for (OB). In Figure 12, for SVM-G, the results are best for (Ot), with a probability of 0.997, and a probability of 0.001 for (OB), followed by (All), with a probability of 0.983, and a probability of 0.000 for (OB), and by (t), with a probability of 0.989, and a probability of 0.002 for (OB).

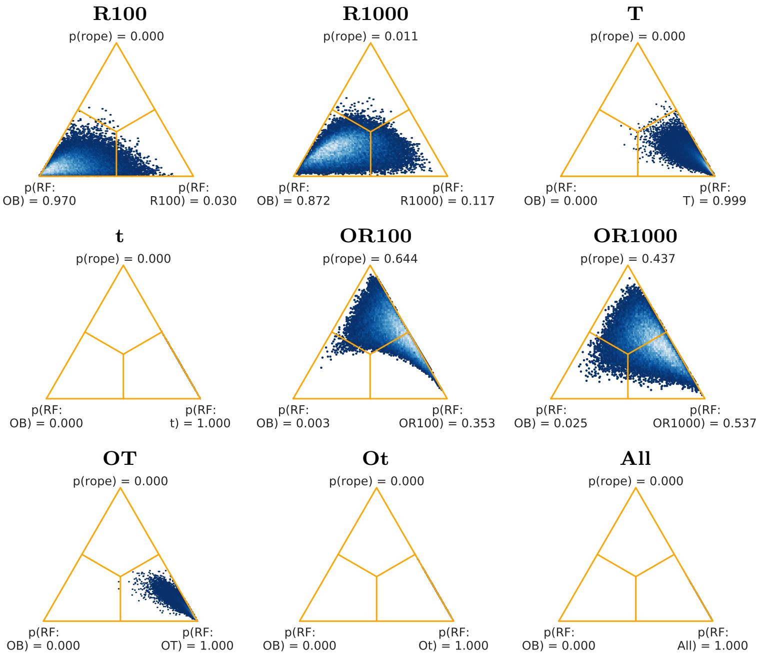


Figure 10. Posteriors for the Bayesian sign-rank tests for RF, from the difficult datasets.

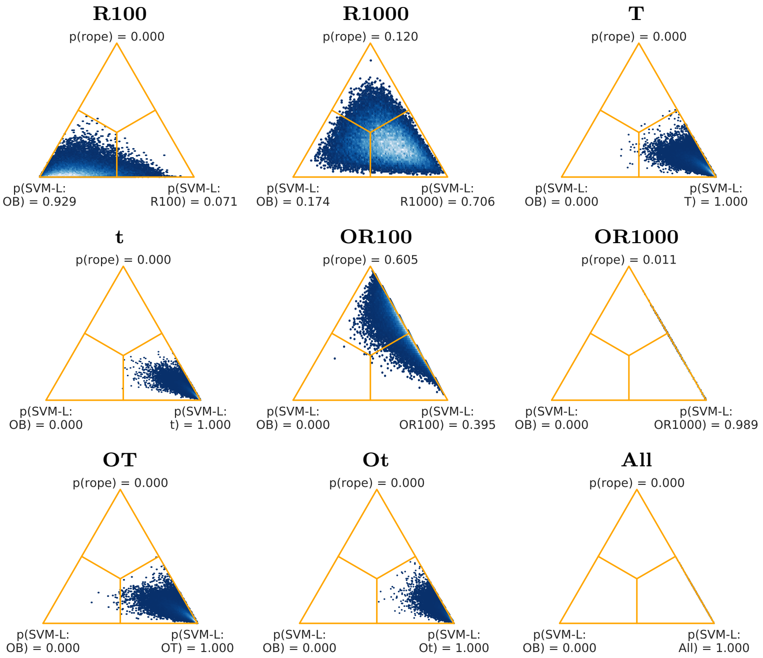


Figure 11. Posteriors for the Bayesian sign-rank tests for SVM-L, from the difficult datasets.

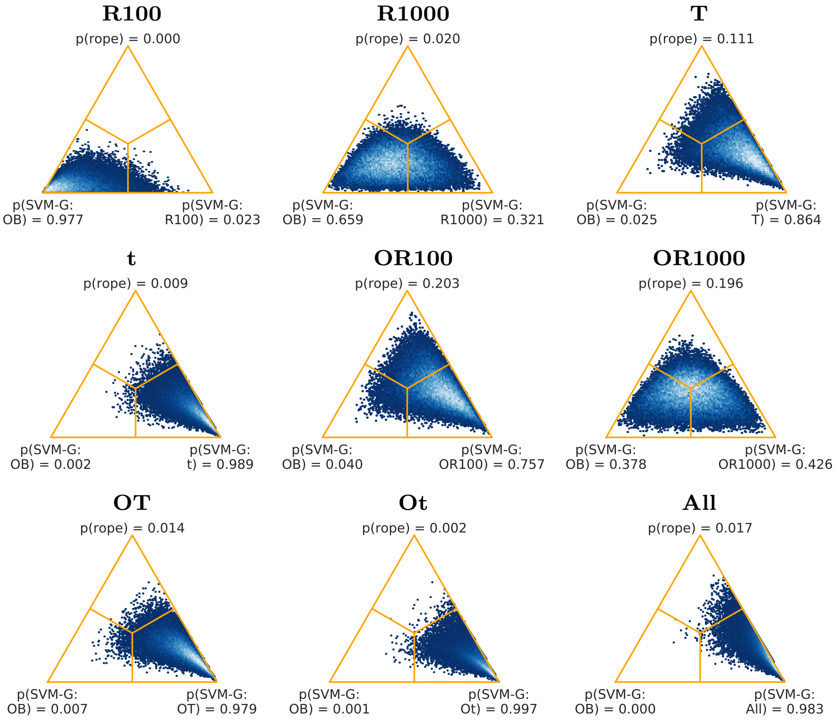


Figure 12. Posteriors for the Bayesian sign-rank tests for SVM-G, from the difficult datasets.

4. Conclusions

This paper presents a comparative analysis of different types of classification methodologies, applied on a series of datasets of raw signals acquired by a portable radar sensor, for different types of materials. In particular, twelve different types of feature vectors obtained from the original raw dataset were obtained, applying different types of feature extraction strategies. These feature vectors were subsequently combined with two classification methods (Random forests and SVM with linear and radial kernel types). A stacked generalization (Stacking) approach was also considered which involved base classifiers created using Random Forest and SVM trained using a subset of the different sets of features. The classification results shown outperformed the corresponding ones obtained by Shyong Yeo et al. [18], when considering the complete collection of datasets, as well as in a wider margin when using the partial so-called difficult datasets. In particular, the difference between the use of the TSFRESH with feature selection (t) features and the original and basic (OB) features (used in [18]), for the complete group of datasets, is almost 4% in accuracy. Moreover, this difference increases to almost 7.7%, for (t) vs. (OB) as well, for the so-called difficult subgroup.

From a classifier performance point-of-view, SVM with linear kernel (with default options) has the best global results (the methods with best average accuracy and rank in Tables 4 and 7 use SVM-L), being much less costly than SVM with Gaussian kernel (with parameter adjustment) and Stacking. This suggest that it is not necessary to use expensive methods when using adequate feature extraction methods.

Potential future lines of research include the creation of our own datasets to explore the use of the radar sensor in problems that may have an industrial interest, for instance, in non-destructive testing or in the signal analysis of trash composites, to discern or classify them.

Author Contributions: J.F.D.-P., P.L.-C. and J.J.R. made the design of experiments, and carried them out. All the authors analysed the results and wrote the paper up. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Spanish Ministry of Science and Innovation under project PID2020-119894GB-I00, Junta de Castilla y León under project BU055P20 (JCyL/FEDER, UE) co-financed through European Union FEDER funds. José Luis Garrido-Labrador was supported by the predoctoral grant (BDNS 510149) awarded by the Universidad de Burgos, Spain.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data and codes will be available upon demand.

Acknowledgments: We would like to thank the authors of [18] for making the datasets available (https://github.com/tcboy88/solinteractiondata (accessed on 18 January 2021). This work was supported by the Spanish Ministry of Science and Innovation under project PID2020-119894GB-I00, Junta de Castilla y León under project BU055P20 (JCyL/FEDER, UE) co-financed through European Union FEDER funds. José Luis Garrido-Labrador was supported by the predoctoral grant (BDNS 510149) awarded by the Universidad de Burgos, Spain.

Conflicts of Interest: Authors declare that they do not have any conflict of interests.

Appendix A. Tables with Classification Results for All Datasets

This section includes four tables corresponding to the complete classification results, for the 34 datasets, when applying RF, SVM-L, SVM-G, and Stacking classification strategies.

Table A1. Results for classifier RF with different feature sets.

	O	B	OB	T	OT	R100	R1000	t	Ot	OR100	OR1000	All
Count + Order Lego	97.09	90.45	96.09	94.27	97.09	**98.00**	**98.00**	**98.00**	97.09	96.18	**98.00**	96.09
Count 20 chips NO case ×30 sorted	65.68	57.57	63.60	65.51	66.95	49.76	57.24	**68.70**	66.95	67.74	65.18	68.68
Count 20 chips WITH case ×30 sorted	70.79	61.11	70.32	73.33	75.56	66.98	69.84	75.24	**77.14**	72.06	72.86	75.71
Count 20 papers ×10	**78.57**	75.71	75.71	72.86	73.33	74.76	76.67	75.24	75.24	76.19	77.14	78.57
Distance 3 mugs 10 distances	48.56	52.78	47.22	67.89	64.78	75.11	68.89	**78.67**	64.44	49.67	65.67	73.22
Distance 3 mugs grouped by material	71.33	92.67	84.33	97.89	99.00	80.78	80.00	**100.00**	**100.00**	85.33	84.11	**100.00**
Distance 7 slotting	**100.00**	98.33	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
Flip 10 creditcards NO case	82.27	80.00	83.64	86.82	85.00	79.55	79.55	**87.27**	86.36	80.91	83.64	83.64
Flip 10 creditcards WITH case	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
Flip 52 cards	94.36	96.27	96.18	97.18	97.27	95.36	97.27	**98.18**	**98.18**	96.27	97.18	97.27
Identify 10 creditcards NO case	85.45	83.64	87.27	**91.82**	90.91	80.91	85.45	90.91	**91.82**	85.45	83.64	90.00
Identify 10 creditcards WITH case	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
Identify 12 printed designs	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
Identify 12 touch on numpad	97.56	91.67	98.33	97.82	98.08	96.15	97.05	97.95	97.82	98.21	**98.59**	98.46
Identify 5 colors × 20 chips	75.00	73.33	75.42	82.08	81.25	78.75	80.83	**83.75**	83.33	76.25	82.08	82.50
Identify 6 tagged plastic cards	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
Identify 6 users by palm	82.86	79.52	82.86	89.52	89.05	82.86	85.24	**91.43**	90.95	85.71	85.24	90.95
Identify 6 users by touch behavior sorted	80.38	71.41	82.56	95.26	94.74	73.21	80.38	**97.69**	96.03	82.95	80.90	94.49
Identify 7 dominoes	96.25	**100.00**	98.75	**100.00**	98.75	97.50	**100.00**	98.75	98.75	98.75	**100.00**	**100.00**
Identify 9 touch on half-sphere	99.00	94.67	**99.33**	98.00	99.00	96.00	97.33	98.17	98.67	98.33	99.17	99.00
Order 3 coasters NO case	80.42	71.25	79.58	**85.42**	82.71	76.25	75.21	82.92	83.13	80.21	79.17	83.33
Order 3 creditcards NO case sorted	80.40	69.08	82.39	80.40	**85.99**	73.20	78.12	84.26	**85.99**	82.32	82.28	85.40
Order 4 creditcards NO case sorted	56.85	43.59	55.51	56.40	56.69	47.31	49.25	56.10	56.10	55.94	54.47	**59.07**
Order 4 creditcards WITH case sorted	**99.49**	98.46	**99.49**	**99.49**	**99.49**	**99.49**	**99.49**	**99.49**	**99.49**	**99.49**	**99.49**	**99.49**
Rotation interval half numeric	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
Rotation interval one numeric	99.17	**100.00**	99.17	**100.00**	99.17	**100.00**	**100.00**	**100.00**	**100.00**	99.17	99.17	**100.00**
Slide inside-out on desk surface	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
Slide outside-in ruler	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
session1 × 10	99.81	99.25	99.81	99.81	99.81	99.81	**100.00**	99.81	99.81	99.81	99.81	99.62
session2 × 10	99.25	98.30	**99.62**	99.43	99.43	99.25	99.43	99.43	99.43	**99.62**	99.43	**99.62**
session3 × 10	**99.81**	99.62	99.62	**99.81**	99.62	**99.81**	**99.81**	**99.81**	**99.81**	**99.81**	**99.81**	99.62
session4 × 10	99.81	99.81	**100.00**	**100.00**	99.81	**100.00**	**100.00**	99.81	**100.00**	**100.00**	99.81	**100.00**
session5 × 10	97.92	97.55	97.92	**98.68**	98.68	97.55	97.74	98.68	98.49	98.11	97.92	98.49
session6 × 10	**99.25**	**99.25**	**99.25**	**99.25**	**99.25**	99.06	99.06	**99.25**	**99.25**	**99.25**	**99.25**	**99.25**
Mean	89.33	87.51	89.82	92.03	92.10	88.75	89.76	**92.93**	92.48	90.11	90.71	92.72
Average rank	7.51	9.38	7.01	5.50	5.75	8.41	7.16	**4.68**	4.94	6.50	6.46	4.69

Table A2. Results for classifier SVM-L with different feature sets.

	O	B	OB	T	OT	R100	R1000	t	Ot	OR100	OR1000	All
Count + Order Lego	**99.00**	96.09	**99.00**	96.00	96.00	98.00	98.00	96.00	96.00	**99.00**	98.00	97.00
Count 20 chips NO case ×30 sorted	60.11	60.42	62.49	71.71	72.35	60.90	71.88	72.35	71.87	69.17	73.78	**73.94**
Count 20 chips WITH case ×30 sorted	79.52	67.78	79.52	78.73	79.37	73.33	78.89	79.68	80.00	81.27	**81.27**	80.63
Count 20 papers ×10	88.57	79.52	87.14	82.38	82.38	88.10	**91.43**	84.29	84.76	88.10	89.52	86.19
Distance 3 mugs 10 distances	39.11	27.00	47.56	64.67	61.33	77.22	80.56	**88.00**	72.22	63.44	75.11	77.33
Distance 3 mugs grouped by material	63.56	77.00	80.89	**100.00**	**100.00**	90.44	93.44	**100.00**	**100.00**	85.22	90.33	**100.00**
Distance 7 slotting	**100.00**	**100.00**	**100.00**	98.33	**100.00**	**100.00**	**100.00**	98.33	**100.00**	**100.00**	**100.00**	**100.00**
Flip 10 creditcards NO case	79.55	75.45	82.73	86.82	87.27	79.55	84.09	88.18	**89.55**	82.73	85.91	88.64
Flip 10 creditcards WITH case	**100.00**	96.82	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
Flip 52 cards	96.36	92.45	96.36	98.18	98.18	97.09	98.18	**99.09**	**99.09**	94.36	98.18	98.18
Identify 10 creditcards NO case	80.00	81.82	83.64	88.18	87.27	79.09	81.82	**90.00**	**90.00**	83.64	84.55	86.36
Identify 10 creditcards WITH case	**100.00**	**100.00**	**100.00**	98.18	98.18	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
Identify 12 printed designs	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
Identify 12 touch on numpad	**99.36**	94.62	98.97	98.59	98.85	97.82	97.95	98.59	98.59	98.97	98.72	98.59
Identify 5 colors × 20 chips	85.00	76.67	87.50	94.58	94.58	81.25	85.83	**96.67**	**96.67**	87.08	88.33	**96.67**
Identify 6 tagged plastic cards	**100.00**	**100.00**	**100.00**	97.14	97.14	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
Identify 6 users by palm	87.62	85.71	90.95	92.86	92.86	86.19	90.00	95.24	94.29	90.00	91.43	**95.24**
Identify 6 users by touch behavior sorted	82.18	61.03	85.38	98.08	97.95	71.03	85.64	**98.72**	98.72	85.64	88.72	98.59
Identify 7 dominoes	**100.00**	**100.00**	**100.00**	96.25	97.50	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
Identify 9 touch on half-sphere	97.50	95.17	98.67	98.83	98.83	97.67	99.50	99.00	99.00	98.50	**99.67**	99.33
Order 3 coasters NO case	82.92	77.50	81.04	84.58	85.42	76.67	80.00	85.21	85.21	82.71	83.12	**86.46**
Order 3 creditcards NO case sorted	82.32	75.59	86.58	89.01	89.01	81.07	84.12	89.01	**90.26**	84.74	87.87	89.63
Order 4 creditcards NO case sorted	60.13	45.54	61.16	66.36	66.37	51.93	58.64	65.91	**67.40**	61.01	60.72	65.47
Order 4 creditcards WITH case sorted	**99.49**	99.23	**99.49**	**99.49**	**99.49**	**99.49**	99.23	**99.49**	**99.49**	**99.49**	99.23	**99.49**
Rotation interval half numeric	**100.00**	**100.00**	**100.00**	99.17	99.17	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
Rotation interval one numeric	**100.00**	**100.00**	**100.00**	99.17	99.17	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
Slide inside-out on desk surface	**100.00**	**100.00**	**100.00**	97.27	97.27	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
Slide outside-in ruler	**100.00**	**100.00**	**100.00**	98.18	99.09	**100.00**	**100.00**	99.09	99.09	**100.00**	**100.00**	99.09
session1 × 10	99.81	99.62	99.81	99.62	99.62	**100.00**	**100.00**	99.81	99.81	99.81	99.81	99.81
session2 × 10	**99.62**	99.06	**99.62**	98.87	99.25	**99.62**	99.25	99.06	99.43	**99.62**	99.25	99.43
session3 × 10	**99.81**	99.43	**99.81**	99.62	99.62	**99.81**	**99.81**	99.43	99.62	**99.81**	**99.81**	**99.81**
session4 × 10	**100.00**	**100.00**	**100.00**	99.62	99.62	**100.00**	**100.00**	99.43	99.43	**100.00**	**100.00**	99.43
session5 × 10	98.87	97.55	98.49	**99.25**	**99.25**	98.87	**99.25**	**99.25**	**99.25**	**99.25**	**99.25**	**99.25**
session6 × 10	**99.25**	**99.25**	**99.25**	**99.25**	**99.25**	99.06	99.06	**99.25**	**99.25**	**99.25**	99.06	**99.25**
Mean	89.99	87.07	91.35	93.20	93.28	90.71	92.84	**94.68**	94.38	92.14	93.28	94.52
Average rank	6.96	9.43	6.41	7.66	6.96	7.43	6.29	5.53	5.06	6.00	5.51	**4.76**

Table A3. Results for classifier SVM-G with different feature sets.

	O	B	OB	T	OT	R100	R1000	t	Ot	OR100	OR1000	All
Count + Order Lego	**98.00**	97.00	**98.00**	95.09	95.09	97.00	**98.00**	95.09	95.09	**98.00**	**98.00**	96.00
Count 20 chips NO case ×30 sorted	71.39	60.74	71.23	72.50	73.93	63.60	73.62	72.35	74.25	73.94	73.46	**75.68**
Count 20 chips WITH case ×30 sorted	79.68	69.68	79.05	79.52	81.43	75.40	80.63	80.95	81.90	**83.33**	82.06	81.75
Count 20 papers ×10	88.57	79.05	87.62	82.86	83.33	87.62	**90.00**	84.29	83.81	88.10	89.05	85.71
Distance 3 mugs 10 distances	46.56	58.00	50.78	63.67	60.33	77.22	80.56	**88.00**	71.11	62.56	75.11	77.33
Distance 3 mugs grouped by material	71.33	88.22	80.89	**100.00**	**100.00**	86.33	92.44	**100.00**	**100.00**	94.67	88.33	**100.00**
Distance 7 slotting	**100.00**	**100.00**	**100.00**	98.33	96.67	**100.00**	**100.00**	98.33	98.33	**100.00**	**100.00**	98.33
Flip 10 creditcards NO case	87.27	85.00	**88.18**	85.45	86.36	81.82	85.45	88.18	87.73	85.00	85.45	87.73
Flip 10 creditcards WITH case	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
Flip 52 cards	**98.18**	96.27	**98.18**	**98.18**	**98.18**	96.18	**98.18**	**98.18**	**98.18**	97.18	96.18	**98.18**
Identify 10 creditcards NO case	**90.00**	85.45	84.55	85.45	86.36	79.09	83.64	89.09	88.18	88.18	83.64	84.55
Identify 10 creditcards WITH case	**100.00**	**100.00**	**100.00**	99.09	99.09	**100.00**	**100.00**	99.09	99.09	**100.00**	**100.00**	99.09
Identify 12 printed designs	**100.00**	99.23	99.23	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	99.23
Identify 12 touch on numpad	99.10	96.15	**99.23**	98.33	98.33	97.82	98.21	98.46	98.46	98.85	98.46	98.46
Identify 5 colors × 20 chips	86.67	85.00	88.33	94.58	94.58	75.83	85.00	**96.25**	**96.25**	87.08	87.50	95.83
Identify 6 tagged plastic cards	**100.00**	**100.00**	**100.00**	98.57	98.57	**100.00**	**100.00**	98.57	98.57	**100.00**	**100.00**	98.57
Identify 6 users by palm	89.52	86.67	91.90	92.86	91.90	85.71	89.52	**94.76**	93.33	92.38	90.48	93.81
Identify 6 users by touch behavior sorted	91.41	82.05	91.41	97.82	97.82	82.95	89.10	**98.46**	**98.46**	91.92	90.26	98.08
Identify 7 dominoes	**100.00**	98.75	**100.00**	98.75	98.75	**100.00**	**100.00**	98.75	98.75	**100.00**	**100.00**	98.75
Identify 9 touch on half-sphere	98.33	97.00	98.50	98.83	98.83	98.33	**99.50**	99.00	99.00	99.00	**99.50**	99.33
Order 3 coasters NO case	83.75	76.04	85.00	85.00	**86.04**	76.25	80.63	85.42	85.21	83.54	83.75	86.04
Order 3 creditcards NO case sorted	88.38	78.13	89.63	89.63	88.38	79.89	84.15	87.79	**90.26**	87.79	87.17	89.04
Order 4 creditcards NO case sorted	63.84	51.48	63.40	66.50	**66.81**	53.28	58.04	66.65	66.51	63.09	61.32	66.06
Order 4 creditcards WITH case sorted	99.23	99.23	99.23	99.23	99.23	99.23	99.23	99.23	99.23	99.23	99.23	**99.49**

Table A3. *Cont.*

	O	B	OB	T	OT	R100	R1000	t	Ot	OR100	OR1000	All
Rotation interval half numeric	**100.00**	**100.00**	**100.00**	**100.00**	99.17	**100.00**	**100.00**	**100.00**	**100.00**	99.17	99.17	**100.00**
Rotation interval one numeric	99.17	99.17	**100.00**	99.17	99.17	**100.00**	99.17	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
Slide inside-out on desk surface	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	99.09	99.09	**100.00**	**100.00**	**100.00**	99.09	**100.00**
Slide outside-in ruler	**100.00**	**100.00**	**100.00**	99.09	99.09	**100.00**	**100.00**	99.09	99.09	**100.00**	**100.00**	99.09
session1 × 10	99.81	99.62	99.81	99.62	99.62	**100.00**	**100.00**	99.81	99.81	99.81	99.81	99.81
session2 × 10	**99.62**	98.87	99.25	98.87	99.06	**99.62**	99.06	99.06	99.25	**99.62**	99.43	99.25
session3 × 10	**99.81**	99.43	**99.81**	99.43	99.43	**99.81**	**99.81**	99.43	99.43	**99.81**	**99.81**	99.62
session4 × 10	**100.00**	**100.00**	**100.00**	99.43	99.62	**100.00**	**100.00**	99.25	99.25	**100.00**	99.81	99.43
session5 × 10	98.68	97.92	98.30	99.06	99.06	98.87	98.87	99.06	99.06	99.06	**99.25**	98.87
session6 × 10	**99.25**	**99.25**	**99.25**	**99.25**	**99.25**	98.87	99.06	**99.25**	**99.25**	**99.25**	99.06	**99.25**
Mean	91.99	90.10	92.38	93.36	93.34	90.88	92.97	**94.47**	94.02	93.25	93.07	94.19
Average rank	5.94	8.97	5.94	7.26	6.99	7.74	6.47	5.87	5.65	**5.40**	6.13	5.65

Table A4. Results for Stacking.

	Stacking	Stacking+
Count + Order Lego	**98.00**	**98.00**
Count 20 chips NO case ×30 sorted	**77.60**	76.17
Count 20 chips WITH case ×30 sorted	**80.95**	80.79
Count 20 papers ×10	83.33	**83.81**
Distance 3 mugs 10 distances	**64.78**	63.67
Distance 3 mugs grouped by material	99.00	**100.00**
Distance 7 slotting	**100.00**	**100.00**
Flip 10 creditcards NO case	**87.73**	86.82
Flip 10 creditcards WITH case	**100.00**	**100.00**
Flip 52 cards	**98.18**	**98.18**
Identify 10 creditcards NO case	**88.18**	**88.18**
Identify 10 creditcards WITH case	**100.00**	**100.00**
Identify 12 printed designs	**100.00**	**100.00**
Identify 12 touch on numpad	98.97	**99.10**
Identify 5 colors × 20 chips	94.58	**96.25**
Identify 6 tagged plastic cards	**100.00**	**100.00**
Identify 6 users by palm	94.29	**95.24**
Identify 6 users by touch behavior sorted	98.46	**98.85**
Identify 7 dominoes	**100.00**	**100.00**
Identify 9 touch on half-sphere	98.83	**99.00**
Order 3 coasters NO case	82.08	**83.54**
Order 3 creditcards NO case sorted	88.38	**89.01**
Order 4 creditcards NO case sorted	64.43	**65.47**
Order 4 creditcards WITH case sorted	**99.49**	**99.49**
Rotation interval half numeric	**100.00**	**100.00**
Rotation interval one numeric	**100.00**	**100.00**
Slide inside-out on desk surface	**100.00**	**100.00**
Slide outside-in ruler	**100.00**	**100.00**
session1 × 10	**99.81**	**99.81**
session2 × 10	**99.43**	**99.43**
session3 × 10	**99.81**	**99.81**
session4 × 10	**99.81**	**99.81**
session5 × 10	98.49	**98.87**
session6 × 10	**99.25**	**99.25**
Mean	93.94	**94.07**
Average rank	1.60	**1.40**

References

1. Stergiopoulos, S. *Advanced Signal Processing Handbook: Theory and Implementation for Radar, Sonar, and Medical Imaging Real Time Systems*; CRC Press: Boca Raton, FL, USA, 2000.
2. Gini, F.; Rangaswamy, M. *Knowledge-Based Radar Detection, Tracking, and Classification*; John Wiley and Sons: Hoboken, NJ, USA, 2007.
3. Ptak, P.; Hartikka, J.; Ritola, M.; Kauranne, T. Aircraft classification based on radar cross section of long-range trajectories. *IEEE Trans. Aerosp. Electron. Syst.* **2015**, *51*, 3099–3106; [CrossRef]
4. Watts, S. *Airborne Maritime Surveillance Radar, Volume 1*; Morgan and Claypool Publishers: San Rafael, CA, USA, 2018; pp. 2053–2571. [CrossRef]

5. Ajadi, O.A.; Barr, J.; Liang, S.Z.; Ferreira, R.; Kumpatla, S.P.; Patel, R.; Swatantran, A. Large-scale crop type and crop area mapping across Brazil using synthetic aperture radar and optical imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *97*, 102294. [CrossRef]
6. Spagnuolo, O.S.; Jarvey, J.C.; Battaglia, M.J.; Laubach, Z.M.; Miller, M.E.; Holekamp, K.E.; Bourgeau-Chavez, L.L. Mapping Kenyan Grassland Heights Across Large Spatial Scales with Combined Optical and Radar Satellite Imagery. *Remote Sens.* **2020**, *12*, 1086. [CrossRef]
7. Lei, W.; Jiang, X.; Xu, L.; Luo, J.; Xu, M.; Hou, F. Continuous Gesture Recognition Based on Time Sequence Fusion Using MIMO Radar Sensor and Deep Learning. *Electronics* **2020**, *9*, 869. [CrossRef]
8. Kang, S.W.; Jang, M.H.; Lee, S. Identification of Human Motion Using Radar Sensor in an Indoor Environment. *Sensors* **2021**, *21*, 2305. [CrossRef]
9. Klavestad, S.; Assres, G.; Fagernes, S.; Grønli, T.M. Monitoring Activities of Daily Living Using UWB Radar Technology: A Contactless Approach. *IoT* **2020**, *1*, 320–336. [CrossRef]
10. Park, D.; Lee, S.; Park, S.; Kwak, N. Radar-Spectrogram-Based UAV Classification Using Convolutional Neural Networks. *Sensors* **2021**, *21*, 210. [CrossRef] [PubMed]
11. Kim, W.; Cho, H.; Kim, J.; Kim, B.; Lee, S. YOLO-Based Simultaneous Target Detection and Classification in Automotive FMCW Radar Systems. *Sensors* **2020**, *20*, 2897. [CrossRef]
12. Senigagliesi, L.; Ciattaglia, G.; De Santis, A.; Gambi, E. People Walking Classification Using Automotive Radar. *Electronics* **2020**, *9*, 588. [CrossRef]
13. Yeo, H.S.; Quigley, A. Radar sensing in human-computer interaction. *Interactions* **2017**, *25*, 70–73. [CrossRef]
14. Wu, C.; Zhang, F.; Wang, B.; Liu, K.R. MSENSE: Towards mobile material sensing with a single millimeter-wave radio. *Proc. ACM Interactive Mob. Wearable Ubiquitous Technol.* **2020**, *4*, 1–20.
15. Choi, J.W.; Ryu, S.J.; Kim, J.H. Short-Range Radar Based Real-Time Hand Gesture Recognition Using LSTM Encoder. *IEEE Access* **2019**, *7*, 33610–33618. [CrossRef]
16. Hof, E.; Sanderovich, A.; Salama, M.; Hemo, E. Face Verification Using 802.11 waveforms. In Proceedings of the 2020 IEEE International Conference on Human-Machine Systems (ICHMS), Rome, Italy, 6–8 April 2020; pp. 1–4.
17. Omer, A.E.; Safavi-Naeini, S.; Hughson, R.; Shaker, G. Blood glucose level monitoring using an FMCW millimeter-wave radar sensor. *Remote Sens.* **2020**, *12*, 385. [CrossRef]
18. Yeo, H.S.; Minami, R.; Rodriguez, K.; Shaker, G.; Quigley, A. Exploring tangible interactions with radar sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2018**, *2*, 1–25. [CrossRef]
19. Ishii, H. Tangible bits: Beyond pixels. In Proceedings of the 2nd International Conference on Tangible and Embedded Interaction, Bonn, Germany, 18–20 February 2008.
20. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
21. Boser, B.E.; Guyon, I.M.; Vapnik, V.N. A Training Algorithm for Optimal Margin Classifiers. In Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory, Pittsburgh, PA, USA, 27–29 July 1992; pp. 144–152.
22. Wolpert, D.H. Stacked generalization. *Neural Netw.* **1992**, *5*, 241–259. [CrossRef]
23. Dempster, A.; Petitjean, F.; Webb, G.I. ROCKET: Exceptionally fast and accurate time classification using random convolutional kernels. *Data Min. Knowl. Discov.* **2020**. [CrossRef]
24. Christ, M.; Braun, N.; Neuffer, J.; Kempa-Liehr, A.W. Time series feature extraction on basis of scalable hypothesis tests (tsfresh–a python package). *Neurocomputing* **2018**, *307*, 72–77. [CrossRef]
25. Dau, H.A.; Bagnall, A.; Kamgar, K.; Yeh, C.C.M.; Zhu, Y.; Gharghabi, S.; Ratanamahatana, C.A.; Keogh, E. The UCR time series archive. *IEEE/CAA J. Autom. Sin.* **2019**, *6*, 1293–1305. [CrossRef]
26. Bagnall, A.; Dau, H.A.; Lines, J.; Flynn, M.; Large, J.; Bostrom, A.; Southam, P.; Keogh, E. The UEA multivariate time series classification archive. *arXiv* **2018**, arXiv:1811.00075.
27. Bagnall, A.; Davis, L.; Hills, J.; Lines, J. Transformation based ensembles for time series classification. In Proceedings of the 2012 SIAM International Conference on Data Mining, Anaheim, CA, USA, 26–28 April 2012; pp. 307–318.
28. Large, J.; Kemsley, E.K.; Wellner, N.; Goodall, I.; Bagnall, A. Detecting forged alcohol non-invasively through vibrational spectroscopy and machine learning. In Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, Melbourne, Australia, 3–6 June 2018; pp. 298–309.
29. Ye, L.; Keogh, E. Time series shapelets: A new primitive for data mining. In Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, 28 June–1 July 2009; pp. 947–956.
30. Gandhi, A. *Content-Based Image Retrieval: Plant Species Identification*; Oregon State University: Corvallis, OR, USA, 2002.
31. Lee, D.J.; Archibald, J.K.; Schoenberger, R.B.; Dennis, A.W.; Shiozawa, D.K. Contour matching for fish species recognition and migration monitoring. In *Applications of Computational Intelligence in Biology*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 183–207.
32. Christ, M.; Kempa-Liehr, A.W.; Feindt, M. Distributed and parallel time series feature extraction for industrial big data applications. *arXiv* **2016**, arXiv:1610.07717.
33. Benjamini, Y.; Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **2001**, *29*, 1165–1188. [CrossRef]

34. Hsu, C.W.; Chang, C.C.; Lin, C.J. A Practical Guide to Support Vector Classification. 2003. Available online: http://www.datascienceassn.org/sites/default/files/Practical%20Guide%20to%20Support%20Vector%20Classification.pdf (accessed on 18 January 2021).
35. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [CrossRef]
36. Ho, T.K. C4.5 Decision Forests. In Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, Canada, 14–16 August 1995; pp. 278–282.
37. Garrido-Labrador, J.L.; Puente-Gabarri, D.; Ramírez-Sanz, J.M.; Ayala-Dulanto, D.; Maudes, J. Using Ensembles for Accurate Modelling of Manufacturing Processes in an IoT Data-Acquisition Solution. *Appl. Sci.* **2020**, *10*, 4606. [CrossRef]
38. Prieto, O.J.; Alonso-González, C.J.; Rodríguez, J.J. Stacking for multivariate time series classification. *Pattern Anal. Appl.* **2015**, *18*, 297–312. [CrossRef]
39. Garcia-Ceja, E.; Galván-Tejada, C.E.; Brena, R. Multi-view stacking for activity recognition with sound and accelerometer data. *Inf. Fusion* **2018**, *40*, 45–56. [CrossRef]
40. Ouyang, Z.; Sun, X.; Chen, J.; Yue, D.; Zhang, T. Multi-view stacking ensemble for power consumption anomaly detection in the context of industrial internet of things. *IEEE Access* **2018**, *6*, 9623–9631. [CrossRef]
41. Demšar, J. Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.* **2006**, *7*, 1–30.
42. Benavoli, A.; Corani, G.; Mangili, F. Should we really use post-hoc tests based on mean-ranks? *J. Mach. Learn. Res.* **2016**, *17*, 152–161.
43. Benavoli, A.; Corani, G.; Demšar, J.; Zaffalon, M. Time for a Change: A Tutorial for Comparing Multiple Classifiers Through Bayesian Analysis. *J. Mach. Learn. Res.* **2017**, *18*, 1–36.
44. Nemenyi, P. Distribution-Free Mulitple Comparisons. Ph.D. Thesis, Princeton University, Princeton, NJ, USA, 1963.