

Convolutional neural networks combined with feature selection for radio-frequency fingerprinting

Gianmarco Baldini¹ | Irene Amerini² | Franc Dimc³ | Fausto Bonavitacola⁴

¹Joint Research Centre, European Commission, Ispra, Italy

²Department of Computer, Control, and Management Engineering, Sapienza University of Rome, Università' la Sapienza, Rome, Italy

³Faculty of Maritime Studies and Transport, Univerza v Ljubljani, Portorož, Slovenia

⁴Department of communications, Fincons, Milan, Italy

Correspondence

Franc Dimc, Faculty of Maritime Studies and Transport, University of Ljubljana, Portorož, Slovenia.

Email: franc.dimc@fpp.uni-lj.si

Abstract

Radio-frequency fingerprinting is a technique for the authentication and identification of wireless devices using their intrinsic physical features and an analysis of the digitized signal collected during transmission. The technique is based on the fact that the unique physical features of the devices generate discriminating features in the transmitted signal, which can then be analyzed using signal-processing and machine-learning algorithms. Deep learning and more specifically convolutional neural networks (CNNs) have been successfully applied to the problem of radio-frequency fingerprinting using a spectral domain representation of the signal. A potential problem is the large size of the data to be processed, because this size impacts on the processing time during the application of the CNN. We propose an approach to addressing this problem, based on dimensionality reduction using feature-selection algorithms before the spectrum domain representation is given as an input to the CNN. The approach is applied to two public data sets of radio-frequency devices using different feature-selection algorithms for different values of the signal-to-noise ratio. The results show that the approach is able to achieve not only a shorter processing time;

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Computational Intelligence* published by Wiley Periodicals LLC.



it also provides a superior classification performance in comparison to the direct application of CNNs.

KEYWORDS

deep learning, feature selection, radio frequency, security, wireless communication

1 | INTRODUCTION

Radio-frequency fingerprinting (RFF) is an identification technique for electronic devices, more specifically wireless devices. RFF does not rely on cryptographic means, but exploits the intrinsic physical features of electronic devices. The concept is that the electronic components of wireless devices have small differences between the different models—and even devices of the same model—that is due to the use of different manufacturing processes or materials in the production phase. These differences do not usually affect the quality of the communications or compliance with the wireless standard, but can be used to distinguish between the devices by analysing the digital output generated by the device (e.g., signal in space).^{1,2} The digital output of the wireless device retains the physical features of the electronic components during transmission, which can be extracted using a radio-frequency sampler or receiver. For example, the non-linearities of an amplifier generate small differences in the signal in space, even if they are not enough to hamper the functioning of the device and the conformity to the standard implemented by the wireless device (e.g., an 802.11g access point in case of a common Wi-Fi device). This general concept has been proven to successfully distinguish between wireless devices with great accuracy (better than 90%) using a hand-crafted set of features (e.g., standard deviation, skewness, kurtosis) and machine-learning algorithms in References 3,4. Deep learning (DL) was also demonstrated to offer superior performance to “shallow machine learning” (e.g., a decision-tree algorithm) at the expense of more computing resources and time.^{5,6} In comparison to approaches based on shallow machine learning and on selected features, the DL algorithm is able to learn the optimal set of features independently.

Two of the most important metrics to evaluate the RFF approach are the identification accuracy and the computing time, which should be minimized. Studies available in the literature¹⁻³ have shown that there is usually a trade-off between the two. These two metrics are also related to the robustness of the approach, where robustness in this context means the ability of the approach to identify the wireless devices when the signal-to-noise ratio (SNR) decreases due to the increasing presence of noise.

In RFF one of the potential problems associated with obtaining good performance according to the metrics identified above is related to the size of the input data, which could be linked to the sampling rate used to collect the signals in space from the wireless devices. High sampling rates are preferable because they provide a high level of granularity, necessary to preserve the discriminating information in the signals.^{1,2,7} On the other hand, high sampling rates generate large samples, which take a lot of time to process using the ML/DL algorithm. It would be preferable to implement a pre-processing step for a dimensionality reduction of the samples, so as to reduce their size before being used as an input for the ML/DL algorithm. The objective is to design such a step in a way that minimizes the computational time (and computational resources needed

for the computation) while maintaining a competitive level of accuracy compared to the baseline using all the input data.

This paper proposes a novel approach (to the best of the authors' knowledge) in the field of RFF that combines three different conceptual elements or steps: (1) a feature-extraction process performed on a segmented spectral domain representation of the original input data, (2) the application of different feature-selection algorithms (FSAs) to identify the most discriminating features and their associated segments in the spectral domain representation, and (3) the application of a convolutional neural network (CNN), which is a kind of DL algorithm, to the reassembled segments identified in the previous step, that is, a reduced size of the input data.

The hypothesis of the approach is that the pre-processing steps (1) and (2) would be able to perform a dimensionality reduction without removing the discriminating features. In addition, this approach should be time efficient. The time needed to extract the features ($Time_{S1}$) added to the time required to apply the FSA ($Time_{S2}$) and the time needed to apply the CNN to the reduced spectral representation ($Time_{S3}$), should not take up more computational time and resources than applying the CNN algorithm directly to the initial input data, which is considered as the baseline case. Then, $Time_{S1} + Time_{S2} + Time_{S3} < Time_{baseline}$. The reason why the spectral domain is used instead of the initial time-domain representation is that studies in the literature^{2,6,8} have shown that the spectral domain representation generally supports a more accurate identification of the wireless devices in the RFF than the time-domain representation. One possible reason for this is that the nonlinearities of electronic components such as filters and amplifiers are more apparent in the spectral domain because their frequency response is different.

In the RFF research literature, one or two of the above elements have been used, but not a combination of all three as proposed in this paper. An approach to dimensionality reduction in RFF was employed in References 3,9 by extracting statistical features from portions of the spectral domain representation of the signal. Subsequently, shallow machine-learning algorithms such as support vector machine (SVM) were used to classify the wireless devices, but no segment-to-feature mapping was used because the SVM was applied directly to the extracted features. The application of a CNN in the RFF domain was shown to be successful in recent literature in comparison to other DL algorithms and shallow machine-learning algorithms, which support its use in this paper.^{5,6} Additional details about the state of art in the literature are provided in Section 3.

Our contribution: We summarize here the key contributions of this paper:

1. We propose a novel RFF approach based on the integration of spectral domain, data segmentation, feature extraction and feature selection in combination with a CNN to support the objective of reducing the input data and consequently the computational time required for source RF identification. The secondary objective is to maintain a competitive identification accuracy compared to the case where no data reduction is performed (which will be referred to as the baseline case in the remainder of this paper).
2. The approach is applied to two different public RFF data sets (described in Section 5).
3. The approach is evaluated for robustness against the presence of noise for decreasing values of the SNR in dB.
4. Different FSAs based on different designs are applied and their performance is compared.

The results show that the proposed approach is not only able to achieve the primary (computational time efficiency) and secondary (identification accuracy) goals, but is also more robust than the baseline case in both data sets under high and medium SNR conditions.



The structure of the paper is as follows. Section 2 introduces the main concepts of data segmentation, feature extraction, feature selection and the CNN, the applications of which to the RFF problem are described in the rest of this paper. Section 3 provides an overview of related studies in the RFF domain or the application of similar approaches to the one proposed in this paper in other domains. Section 4 describes the overall methodology of the proposed approach, including a description of the CNN architecture, the adopted FSAs and the metrics of the evaluation. Section 5 describes the two public data sets used to evaluate the proposed approach. Section 6 provides the results of the evaluation, including a comparison of the different FSAs. Finally, Section 7 draws the conclusions of the paper and describes future developments.

2 | PRELIMINARY KNOWLEDGE

The aim of this section is to provide preliminary information about the main elements used in this study: data segmentation, feature extraction, feature selection and the CNN used for the classification. In addition, Table 1 gives the definition of the acronyms used in this paper.

2.1 | Time-series data mining and segmentation

The approach proposed in this paper is based on the concept of the segmentation of time-series data,¹⁰ where the segmentation is performed in the spectral domain of the initial time series and the goal of the segmentation is to achieve a dimensionality reduction while trying to preserve the most discriminating segments. In general, this is not a trivial problem and many different methods have been proposed in literature. Reference 11 provides a comprehensive overview of the different time-series data mining and segmentation techniques adopted in literature. The techniques can be based on a sliding-window approach with a fixed segment, the identification of specific patterns or change-point detection for adaptive segment-size definition (which is more computationally demanding, as described in Reference 11). If we consider that one of the main objectives of our approach is time efficiency, since the total computation time should be lower than when the CNN is applied directly to the whole input data, we have chosen the sliding-window approach in the spectral domain with a fixed window due to its simplicity and lack of computational complexity. For the same reasons, we chose non-overlapping sliding windows. In a further extension of this work, techniques with adaptive segment-size definition while minimizing the computational time can be investigated.

2.2 | Feature extraction

Another element of the proposed approach is the feature-extraction step, which is applied to the segments identified in the time-series data-segmentation step described in the previous sub-section 2.1. In this case it is applied to the spectral domain representation of the signal.

Feature extraction can be defined as the process of creating derived values (known as features) that are intended to be informative and non-redundant to facilitate subsequent steps such as classification, as in this study.¹² The challenge with the approach is to select features that are both informative (e.g., discriminative for classification) and which have low computational complexity. The features (e.g., standard deviation) were selected based on these criteria and their use

TABLE 1 Definitions.

Acronym	Definition
AWGN	Additive white Gaussian noise
CNN	Convolutional neural networks
CUDA	Compute unified device architecture
DG	Dependence guided
dB	decibel
FFT	Fast fourier transform, $FFT(k)_x = \sum_{j=1}^N x(j)W_N^{(j-1)(k-1)}$, $W_N = e^{(-2\pi i)/N}$, $x = \text{signal}$
FM_i	Feature matrix related to HLF i ($i = 1 \dots 5$)
FN	False negative
FP	False positive
FSA	Feature-selection algorithm
GF	Generalized fisher
GSM	Global system for mobile communications
HLF	High-level feature (e.g., kurtosis)
MI	Mutual information
mRMR	minimum redundancy maximum relevance
PCA	Principal component analysis
QCLP	Quadratically constrained linear programming
RF	Radio frequency
RFF	Radio-frequency fingerprinting
RMS	Root mean square, $RMS_x = \sqrt{\frac{1}{N} \sum_1^N x_n ^2}$, $x = \text{signal}$
RMSProp	Root-mean-square propagation
SDAR	Spectral domain amplitude representation
SDR	Software-defined radio
SF_j	Segment feature ($j+1 \dots 150$)
SGO	Structured graph optimization
SGDM	Stochastic gradient descent with momentum
SNR	Signal-to-noise ratio, $SNR_{dB} = 10 \times \log_{10} \frac{\sqrt{\sum_1^N x_n ^2}}{\sqrt{\sum_1^N y_n ^2}}$, $x = \text{signal}$, $y = \text{noise}$.
SVM	Support vector machine
TN	True negative
TP	True positive
USOBC	Unsupervised simultaneous orthogonal basis clustering
USRP	Universal software radio peripheral

in the RFF literature.³ As described in section 4, the features are only used to indicate the most informative segments, which are related to the features.

2.3 | Feature-selection algorithms

Seven FSAs were used to evaluate the approach proposed in this paper: ReliefF; generalized Fisher (GF) score; structured graph optimization (SGO); dependence guided (DG); minimum redundancy maximum relevance (mRMR); unsupervised simultaneous orthogonal basis clustering (USOBC); and mutual information-based (MI). These algorithms were selected for their computational efficiency, because they are widely used as a baseline for historical reasons (e.g., reliefF), because they are based on different concepts, and because newer concepts (e.g., SGO and DG) have recently been proposed in the research literature. In the following paragraphs we provide a brief description. For additional details the reader can refer to the cited reference in each paragraph.

ReliefF¹³ calculates a feature score (based on the identification of feature-value differences between nearest-neighbor instance pairs) for each feature, which can then be applied to rank and select top scoring features for the feature selection.¹⁴ In this study the reliefF nearest-neighbor parameter (K_R) is used.

The GF score was proposed in Reference 15 and can be used to find a subset of features that maximizes the lower bound of the traditional Fisher score. The resulting feature-selection problem is mixed integer programming, which can be reformulated as quadratically constrained linear programming (QCLP).

SGO was designed to address the problem that conventional embedded, unsupervised methods need to construct the similarity matrix, which makes the selected features highly dependent on the learned structure. The authors of SGO in Reference 13 developed an algorithm that addresses this problem and performs feature selection and local structure learning simultaneously, and where the similarity matrix can thus be determined adaptively. Moreover, the authors in Reference 13 imposed constraints on the similarity matrix to obtain more accurate information about the data structure so that the proposed approach can select more valuable features. SGO adaptively learns the local manifold structure, and thus SGO can select more valuable features than other methods.

DG¹⁶ is a joint learning framework for feature selection and clustering where a projection-free feature-selection model is proposed based on $l_2,0$ -norm equality constraints, and dependence-guided terms are used to enhance the dependence among original data, cluster labels, and selected features.

mRMR is part of the minimal-optimal family of FSAs that seek to identify a small set of features that, when combined, have the maximum-possible predictive power.¹⁷ This method handles each feature separately from the data set and uses the mutual information between them to measure the level of similarity between two features A and B.¹⁸ The motivation for using the mRMR FSA is that it can effectively reduce the redundant features while keeping the relevant features for the model.

USOBC is an unsupervised version of simultaneous orthogonal basis clustering feature selection, which was proposed by the authors in Reference 19. USOBC is a FSA that makes use of the local structure information of the data points in the input data. As described in Reference 19, this FSA does not explicitly adopt pre-computed local structure information, but concentrates on the latent cluster information by conducting orthogonal basis clustering directly on the projected



data points to estimate the latent cluster centers. Since the target matrix is put in a single unified term for the regression of the proposed objective function, the feature selection and clustering are simultaneously performed, with the advantage that the feature selection is computed by the estimated latent cluster centers of the projected data points.

The last FSA used in this study is another (in addition to mRMR) mutual-information-based algorithm proposed by the authors in Reference 20, which takes into account both the class-dependent and class-independent correlations among features. In particular, this FSA includes both relevance and redundancy factors, model redundancy using both class-dependent and class-independent correlation and takes the average redundancy over all the previously selected features.

The MATLAB implementation provided by the authors of References 13,15,16,18-20 and the ReliefF function by Mathworks were used in combination with the matFR toolbox by Reference 21, which was used as a common programming interface.

2.4 | Convolutional neural network

A CNN is a type of DL inspired by human learning because the connectivity pattern between the neurons resembles the organization of an animal's visual cortex. A CNN is a kind of feed-forward neural network that is able to extract features from data with a convolution structure.²² Conventional machine-learning techniques like SVM generally require feature extraction (e.g., the application of a standard deviation to the data) as the prerequisite, and this requires a domain expert or the hand-picking of discriminating features, which can be appropriate for the data set (i.e., the selected features have more discriminating power). DL techniques overcome the problem of feature selection by not requiring pre-selected features, but by extracting the discriminating features from the raw input data automatically for a problem in hand (i.e., in this case the classification of wireless devices using RF fingerprints). As described before, the cost for this higher performance in comparison to other machine-learning algorithms is the need for powerful computing resources. In fact, the recent increased use of DL models is also due to the availability of high-performance computing platforms.²³ DL usually consists of a collection of processing layers that can automatically learn features from data through multiple levels of abstraction. A CNN, often called ConvNet, is a type of deep-learning algorithm. It has a deep feed-forward architecture and has proved to have a high performance for many classification tasks, in particular object recognition and image classification.²⁴ In this case the CNN is used to detect specific patterns related to the RF fingerprints of wireless devices in a one-dimensional space, which is the spectral domain representation (using a fast fourier transform) of the digitized signal in space. The reason for selecting a CNN for this particular problem is because it has proven to have a superior performance in similar RFF problems in References 5,6.

3 | RELATED WORK

RFF is a promising identification and authentication technique for wireless devices that can complement cryptographic techniques. As described before, RFF relies on exploiting the small differences in the electronic components transmitted to the digital signal in space, where they can be collected for analysis. Several studies have described the key elements of RFF and provided an overview of the various RFF concepts.^{1,2,25} These surveys have shown that the first RFF



implementations were based on the extraction of manually created features (e.g., variance, Shannon entropy)^{3,26} or the conversion of the signal in space into a representation in the spectral domain.^{27,28} A new trend is the application of DL to RFF, drawing inspiration from the highly successful application of DL algorithms to image analysis and classification.

CNNs have been successfully used for RFF by either applying the CNN to the original time representation of the signal as I/Q samples in Reference 29 or by implementing a pre-processing step that converts the original time representation into a spectral domain representation that serves as the input to the CNN. In the literature the spectral representation combined with the CNN has generally been shown to be more robust and accurate than the time-domain representation.^{6,8} One possible reason is that the non-linearities of the electronic components like filters and amplifiers are more evident in the spectral domain because their frequency response might be different.² For this reason, the approach proposed in this paper is based on the spectral domain representation given as the input to the CNN. One issue with this approach is that the size of the specific sample that serves as the input to the CNN can be large, since a high level of granularity is required to extract the discriminating features to classify the RF devices. For example, in Reference 30 segments of the samples are randomly chosen with a sliding window to reduce the size of the sample. This processing step is also based on the understanding that the RF fingerprints are not evenly distributed over the entire spectral range, since the nonlinearities often affect certain parts of the spectral domain and not others, or only to a small extent, for example, due to resonance elements in the components. In References 3,9, a dimensionality-reduction approach was used for RFF by extracting statistical features from parts of the spectral domain representation of the signal. Shallow machine-learning algorithms such as SVM were used to classify the wireless devices. One possible approach would be to combine the dimensionality reduction approach presented in Reference 3 with the deep-learning approach that has achieved considerable success in the domain of RFF. One possibility would be to apply the CNN to the feature space created by the application of statistical features from Reference 3 (i.e., standard deviation, skewness and kurtosis), but there is a risk that the feature-extraction step removes the discriminating information from the RF fingerprints. Another option (which was used in this study) would be to use the statistical feature-based space to identify the most discriminating parts of the spectral domain, which can then be prioritized (i.e., using a feature-selection algorithm) and recombined to create a spectral representation that can be fed to the deep-learning algorithm. The hypothesis is that such a pre-processing step would be able to perform dimensionality reduction without removing the discriminating features. In addition, this approach should save time: extracting the features added to apply the feature-selection algorithm and applying deep learning to the reduced spectral representation should not require more computational time and resources than applying the deep-learning algorithm directly.

Along these lines, one recent approach proposed in the literature for another domain (analysis of electrocardiogram signals in biomedical engineering) is presented in Reference 31, where the Shannon entropy measure is applied to the time-frequency representation of the electrocardiogram to produce reduced images using principal component analysis (PCA), which are then fed to a CNN. The study proposed in this paper takes inspiration from Reference 31 but goes a step further by using other features and employing FSAs for the filter type instead of PCA. We used feature-selection algorithms instead of PCA because the approach proposed in this paper is slightly different from Reference 31 as it focuses on selecting the most discriminating segments of the spectral range associated with the features. In addition, we prefer dimensionality-reduction techniques that allow feature ranking rather than creating a new space with reduced dimensions as in PCA.

4 | METHODOLOGY

4.1 | Workflow

Figure 1 presents the workflow. In the initial step all the data samples from both data sets (DATASET1 and DATASET2) are synchronized and normalized by subtracting the mean and by calculating the ratio of the signal with its root mean square (RMS). A detailed description of the data sets is provided in Section 5. The synchronization and normalization steps are introduced to ensure that the proposed approach and the algorithms are based only on the fingerprints of the RF and are not affected by different distances between the transmitter and the receiver used to acquire the RF samples (in DATASET1, for example, these are the distances between the drone controllers and the receiver/sampler used to acquire the signal in space). After synchronization and normalization of the data sets, a fast fourier transform (FFT) is applied to each of the samples of the data set to obtain the spectral domain representation with an amplitude and a phase component. Although both components of the spectral domain can be used, an experimental analysis (not presented here because of space limitations) has shown that the amplitude component alone has greater discriminating power in both data sets. This is referred to and used below as the spectral domain amplitude representation (SDAR). The SDAR is divided into segments $SDARSEG_i$ of equal size $SEG_{S1} = 100$ for DATASET1 and $SEG_{S2} = 50$ for DATASET2 (in an extended study of this paper, SEG_{S1} and SEG_{S2} could be hyper-parameters). Since the SDAR

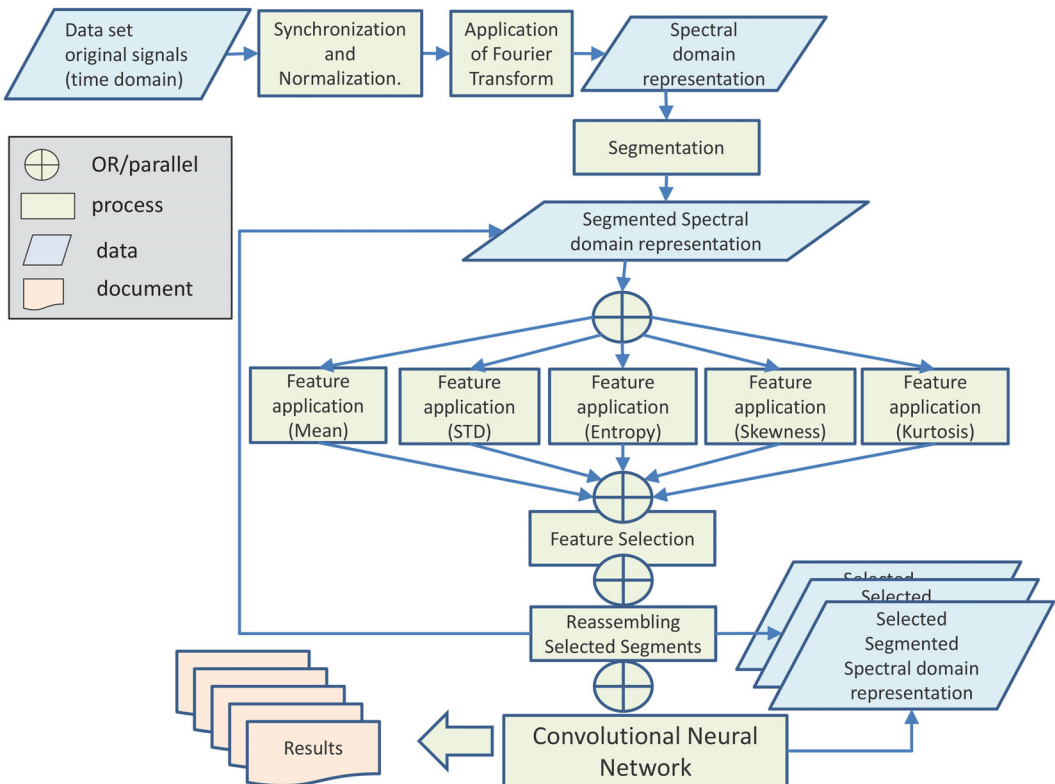


FIGURE 1 Workflow of the proposed approach.

in both data sets is symmetric (because it is the application of the FFT to a real signal), only the first half of the SDAR can be used in the study, which reduces the computational cost. Then, considering that the size of the SDAR is 30,000 in DATASET1 and 2600 in DATASET2 the number of segments is $30000/(2 \cdot 100) = 150$ in DATASET1 and $2600/(2 \cdot 50) = 26$ in DATASET2. Thus, $SDARSEG_i$, $i = 1..150$ in DATASET1 and $SDARSEG_i$, $i = 1..26$ in DATASET2.

For each segment $SDARSEG_i$, five different high-level features are applied. The following features are used in this study: (1) mean, (2) standard deviation, (3) Shannon entropy, (4) skewness, and (5) kurtosis. These features are used because they are adopted in the literature for RFF and because they have a low computing complexity. Then, five feature matrices FM_j ($j=1 \dots 5$) of size $6750 \cdot 150$ are created for DATASET1 and five feature matrices of size $9000 \cdot 26$ for DATASET2 (the application of the FSA is only for the training set portion of 3/4 of the overall data set).

It is important to emphasize that, according to the hierarchy of meanings in this approach, the term “feature” has two different forms. The first type of feature, which we refer to as a high-level feature (HLF), is the feature (e.g., mean, standard deviation) used to convert the SDAR into a FM_j ($j=1 \dots 5$). The second type of feature, which we call the segment feature or SF_{ij} in the rest of this paper, is the value of the j HLF applied to a specific segment $SDARSEG_i$. A more detailed view of these steps is shown in Figure 2.

Then, on each FM_j a feature-selection algorithm (FSA) is applied to reduce the number of features (e.g., from 26 to 10). As described in subsection 2.3, seven different FSAs are used. Then, each algorithm will provide a ranking of features. In this paper the first high-ranking $N_F = 20$ segment features are selected on the basis of criteria explained in Section 6 for DATASET1 and the first high-ranking $N_F = 10$ segment features were selected for DATASET2. In reality each feature corresponds to a particular segment index. Subsequently, the N_F features are used to generate a

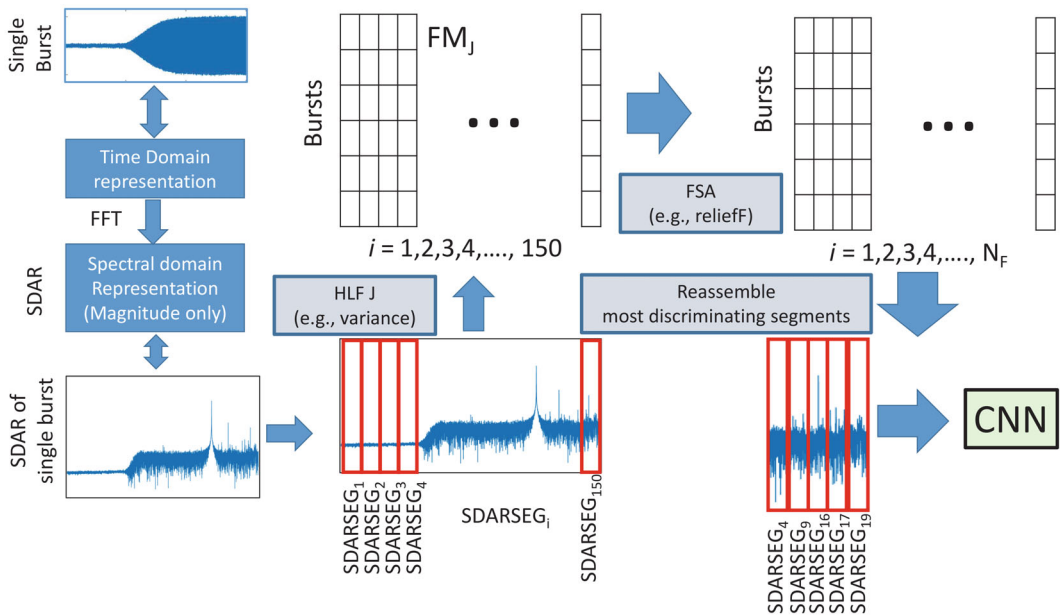


FIGURE 2 Scheme for the application of feature selection in the identification and re-assembling of the most discriminating segments $SDARSEG_i$ of the spectral domain SDAR. An example for the DATASET1 is shown in the figure.

new spectral domain representation of the signals, where N_F segments are continuously recombined. Note that the recombination process is the same for all samples from all instruments in the data set for consistency. The reduction is significant because the new sample size is $20 * 100 = 2000$ instead of the initial 150,000 (a factor reduction of 7.5). The hypothesis is that such a procedure selected only the best discriminating features in the spectral range, and this is based on the understanding that RF fingerprints are mostly related to nonlinearities in certain frequency bands, as they are caused by inadequacies in filters, amplifiers, and so on. In the final step the CNN is applied to the reduced spectral domain representation to perform the classification. For comparison, the CNN is also applied to the original SDAR with all the segments to evaluate the performance of the proposed approach, as is common in the literature⁵ (this is referred to as the baseline in the remainder of this paper).

4.2 | Architecture of the convolutional neural network

A one-dimensional CNN architecture is used because the input is the one-dimensional spectral amplitude SDAR of the RF fingerprint signal, which is reduced with the optimal segments selected by the FSA and then re-assembled. The architecture of the CNN is shown in Figure 3, where three layers are used and the value of the parameters is the result of a grid optimization in the ranges defined below. In Figure 3, FS is the filter size and f is the number of filters in the convolutional layers, P is the pool size, and S is the stride size in the pooling layer. Adam is the CNN solver algorithm with an initial learning rate of 0.005. The optimization ranges were defined as follows: for the filters between 8 and 64, the number of convolutional layers between 2 and 4, the initial learning rate in the value set [0.001, 0.005, 0.01, 0.05, 0.1], the initial convolutional size W_s in the value set [8, 16, 24, 32, 40, 48], and the solvers between Adam, RMSProp and SGDM. The maximum number of epochs was set to 40, as we found that a larger number of epochs was not necessary since the CNN converged before 40 epochs.

Each data set was split into a portion of 3/4 of the total set for the training set, including 1/10 of the training set for validation, and a portion of 1/4 of the total set for testing. The CNN was executed 10 times, randomly shuffling the training and test sets each time, and the results were averaged.

A Windows computing platform with an Intel i9-10885H processor, 2.4 GHz of clock speed, 32 Gbytes of memory, and NVIDIA QUADRO P4000, which is Compute Unified Device Architecture (CUDA enabled) with MATLAB as the programming language was used to compute the results

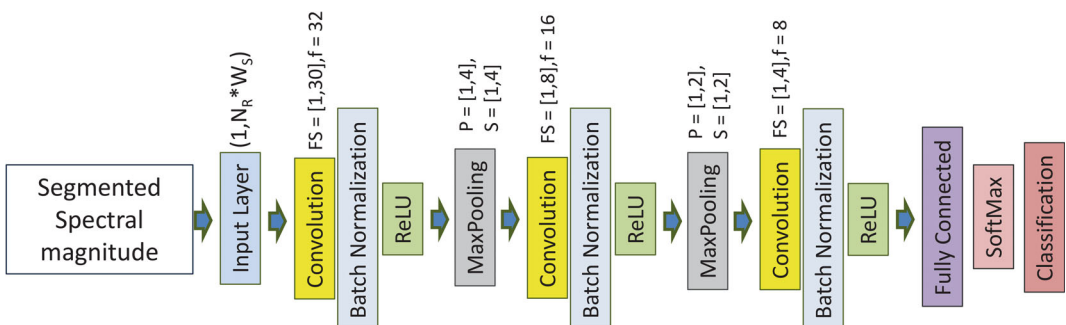


FIGURE 3 1D CNN architecture used in the analysis.



and the execution time. The MATLAB implementations used in this study are listed at the end of Section 2.3. The Deep-Learning MATLAB toolbox (version 2021a) from Mathworks was used to implement the CNN.

5 | DESCRIPTION OF THE DATA SETS

To evaluate our approach we used two different public data sets described in the following subsections.

5.1 | DATASET1: Drone controllers

The first data set was published in Reference 32 and used by the authors in Reference 33. The data set contains several sets of signals collected from radio remote controllers for drones transmitting in the 2.4-GHz band. The sampling frequency used to capture the signals was 20 GSamples per second. Analysis of the data shows that nine controllers from the entire data set have similar signal structures and the same number of samples, and these were selected for analysis. All 1000 data samples from each controller were used, as the total data set consists of 9000 samples. All the samples were normalized and synchronized around the transient, and a time interval of 1.5 microseconds was chosen, corresponding to a vector of length 30,000. Figure 4 shows a sample from each device in the time domain where the transient phase is visible. The transient portion of the signal was chosen because it has been shown in the literature^{2,34,35} that the transient portion of the signal contains the most significant discriminating characteristics of the signal, but the collection of the transient portion requires a relatively high sampling rate, which in turn produces large input files, as in this (where a sampling rate of 20 GSamples per second was used). To

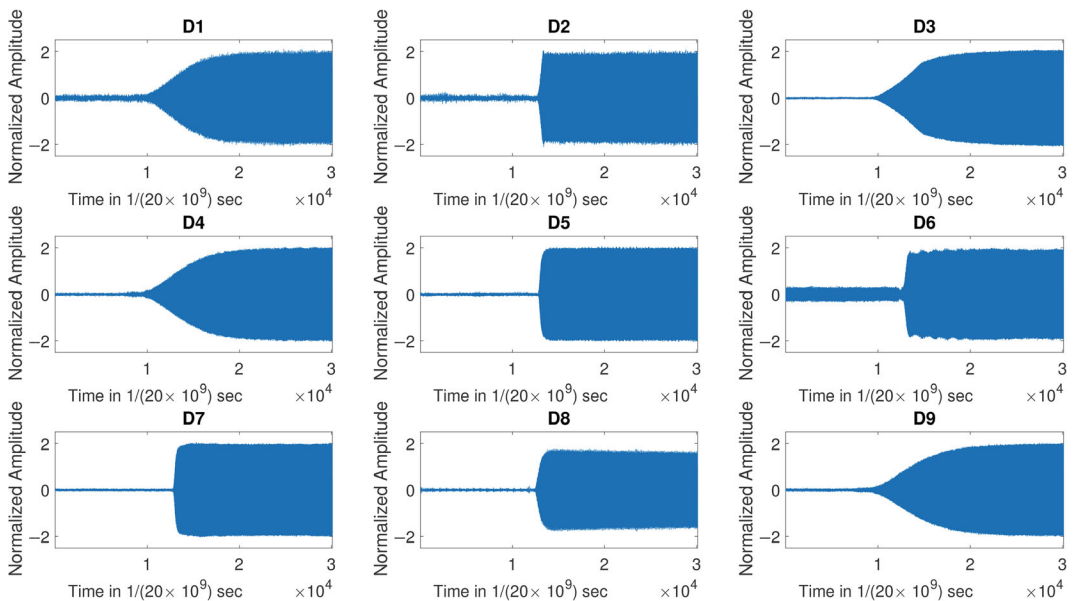


FIGURE 4 Samples from each device used in the first data set (DATASET1).

evaluate the robustness of the approach in the presence of noise, additive white gaussian noise (AWGN) was added to the data set to simulate different SNRs in dB. It should be noted that the RFF problem for this is relatively straightforward since the drone controllers are of different models and the transient portion of the signal is different, as shown in Figure 4. Then, AWGN is added to evaluate the proposed approach under more difficult conditions where the controller signal is obscured by the presence of noise. This is a common approach in the literature where AWGN is used to evaluate the robustness of the proposed algorithm.^{3,5,34}

In the rest of this paper, this data set is identified with the keyword DATASET1.

5.2 | DATASET2: GSM

The second data set is based on the collection of wireless signals based on the global system for mobile communications (GSM) standard from 12 mobile phones of four different models, with a set of three phones for each model. The data set is public and it is available at Reference 36. The data set was created using a controlled test bed in which the RF signals from each of the 12 transmitting wireless devices were collected using a universal software radio peripheral (USRP) N200 receiver of the software-defined radio (SDR) type with a sampling rate of 20 MHz. The SDR was fully disciplined and synchronised with a global navigation satellite system (GNSS) receiver. A total of 1000 bursts were collected from each cell phone to generate a data set of 12,000 samples. For each GSM burst, the data payload was removed, and only the ramp-up, ramp-down, and preamble are maintained, so that the distortion caused by the content (i.e., speech) did not matter to the RFF (the preamble was configured to be the same for all the phones). Subsequently, all the bursts were normalised. As in the first data set, different SNR conditions were generated using the AWGN to evaluate the robustness of the approach to the presence of noise. A graphical representation of one sample for each of the 12 mobile phones is reproduced in Figure 5.

In the rest of this paper, this data set is identified with the keyword DATASET2.

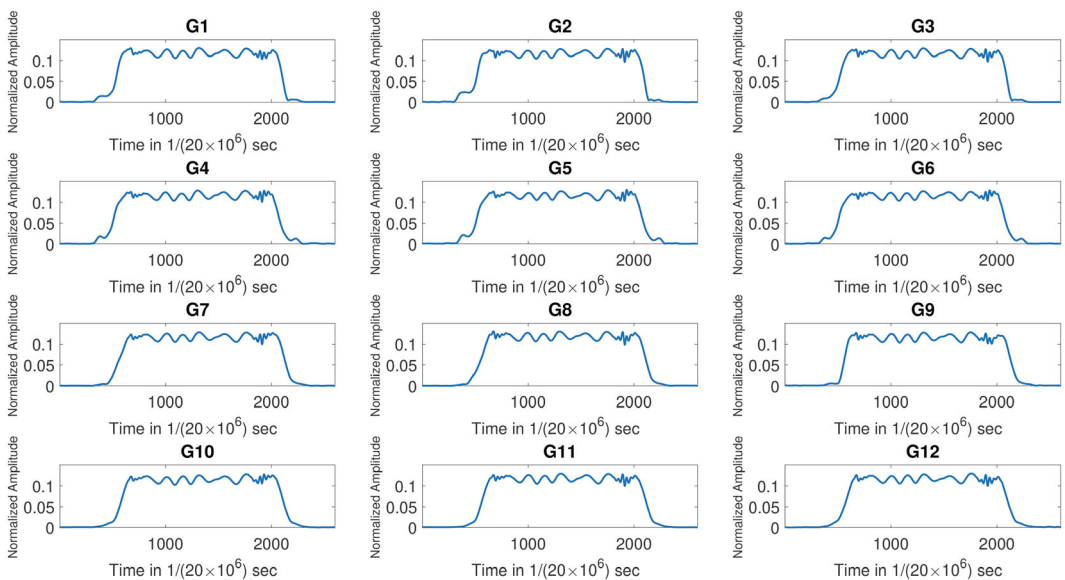


FIGURE 5 Samples from each device used in the second data set (DATASET2).

6 | RESULTS

This section presents the results and related analysis for applying the approach proposed in this paper to DATASET1 and DATASET2. This section is organised as follows. The first subsection 6.1 describes the evaluation metrics. Subsection 6.2 shows the comparison between the different FSAs. Subsection 6.3 shows some examples of the influence of hyperparameters in the approach.

6.1 | Evaluation metrics

The metrics used to evaluate the performance of the proposed approach are accuracy, F-score, and execution time in seconds to measure the computational complexity. Accuracy is the ratio of correct predictions to total predictions. F-Score is the harmonic mean of precision and recall. Precision is the ratio of true positives (TPs) to the sum of Ts and false positives (FPs). Recall is the ratio of TPs to the sum of TPs and false negatives (FNs). The F-score is used because accuracy might not provide a complete understanding on how the FPs and FNPs are distributed in the outcome of the classification results. Since the problem addressed in this paper is a multi-class problem with balanced data sets (DATASET1 and DATASET2), we have implemented the F-score by macro-averaging and taking all the classes as equally important.

In addition we compute the confusion matrices, where the results of the predicted values are compared with the true values. In this paper we use the notation that the predicted values are on the x-axis and the true values are on the y-axis. The execution time in seconds is used as a metric to evaluate the speed of the different approaches. The same computing platform was used to calculate the execution time. The computing platform was described in the previous subsection 4.2.

6.2 | Comparison of the approaches with different FSAs with optimized parameters

As described before, seven different FSAs were used to evaluate the approach on the two data sets across different levels of SNR expressed in dB.

Regarding DATASET1, Figure 6 shows the comparison among the selected approaches once the optimization is performed in relation to the choice of K_R for ReliefF and the specific HLF (e.g., standard deviation) used for the application of the FSA. Figure 6 shows that all FSAs except DG and USOBC provide better accuracy than the baseline in the range of SNR values from -25 to 20 dB, with ReliefF and GF higher than the baseline at -30 dB. It should be noted that both the DG and USOBC FSAs are not robust in the presence of noise. This may be due to the specific characteristics of the data set and the algorithm for that data set. At extreme noise levels ($\text{SNR} \leq -35$ dB), most approaches perform worse than the baseline because the signal is too noisy for the FSA to be able to select the optimal segments of the spectral domain. On the other hand, even the baseline is not able to achieve accuracy when the SNR values are too low. These results at very low SNR values are consistent with the results from the literature on RFF,^{3,5,6} where the accuracy decreases significantly at low SNR values because the presence of noise eventually obscures the RF fingerprints. These results show that the choice of FSA is important to achieve optimal performance, but each algorithm is more-or-less robust for different values of SNR.

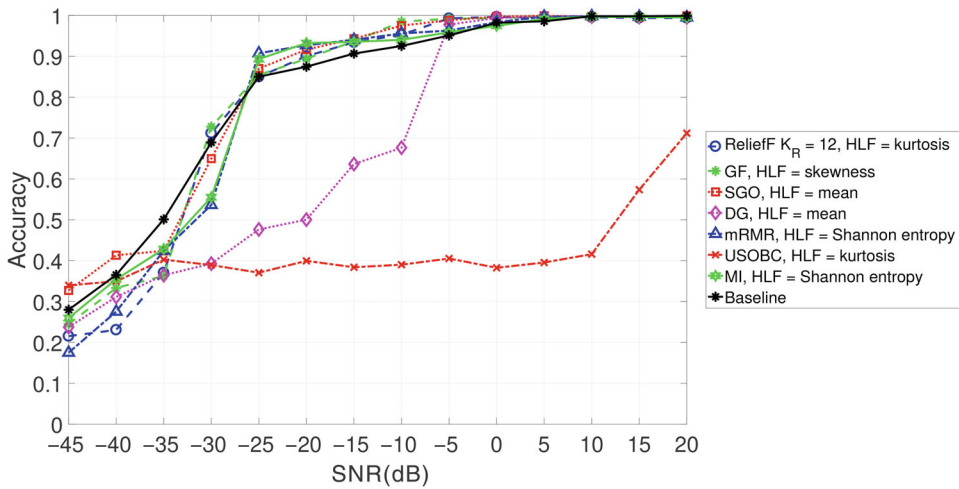


FIGURE 6 DATASET1. Accuracy comparison of the proposed approach based on the application of different FSAs with optimal features and hyperparameters against the baseline for different values of SNR.

The numerical details of the accuracy of results presented in Figure 6 for DATASET1 are shown in Table 2 for specific values of SNR in addition to the corresponding values of the F-score. The best results for each value of SNR are shown in bold. In general, the values of F-score are consistent with the values of the accuracy for the same FSA and the same level of SNR. From the analysis of Table 2 it is clear that ReliefF, SGO, GF, mRMR and MI have better performance than the baseline in terms of accuracy and F-score in presence of low and medium levels of noise (SNR from -30 to -10). At SNR = -10 dB, the difference in accuracy between the best-performing FSA (GF) and the baseline is almost 6%. For values of SNR greater than -10 dB, the accuracy is near ideal for all the algorithms and the differences are minor among the approaches. Both the GF and SGO FSAs have slightly better performance (in terms of accuracy and F-score) than the ReliefF, mRMR, and MI-FSAs in this DATASET1, although performance remains high in absolute terms. It should be noted that the GF and SGO FSAs were introduced more recently than ReliefF, which is much older. Therefore, these two algorithms are more advanced and theoretically better performing than the ReliefF algorithm, which is used for historical reasons. Moreover, an additional advantage of the GF and SGO algorithms is that they do not require the tuning of hyper-parameters like ReliefF.

The results for DATASET2 are shown in Figure 7 and the related detailed values are shown in Table 3. In this second data set the approach partially confirms the results obtained in the first DATASET1, as most FSAs perform better than the baseline for medium and low noise levels (for this data set the range is between 20 and 45 dB). DATASET2 is more challenging than DATASET1 from a classification point of view, as the accuracy decreases significantly at higher SNR values in DATASET1 than in DATASET2. One of the most striking differences is the relatively poor performance of the SGO FSA in the presence of noise, as the accuracy drops significantly below SNR = 30 dB, while all other FSAs perform better than the baseline. One possible reason for the relatively poor performance of the SGO FSA is that the sample space and number of segments in DATASET1 (150) are smaller than DATASET 2 (26). Another difference is that most FSAs (with the exception of SGO) in DATASET2 perform better than the baseline across all the values of SNR, while in DATASET1 this was not in case.



TABLE 2 DATASET1.

SNR value (dB)	Baseline	ReliefF	GF	SGO	DG	mRMR	USOBC	MI
Accuracy								
-30	0.6893	0.7120	0.7280	0.6497	0.3893	0.5364	0.3893	0.551
-20	0.8746	0.9004	0.8942	0.9164	0.508	0.9271	0.3996	0.9324
-10	0.9253	0.9528	0.9848	0.9751	0.655	0.9560	0.3902	0.9404
0	0.9822	0.9964	0.9928	0.9973	0.9937	0.9836	0.3929	0.9747
10	0.9982	0.9955	0.9946	0.9983	0.9973	0.9982	0.4218	0.9973
F-score								
-30	0.6755	0.7091	0.7280	0.6463	0.3912	0.5426	0.3858	0.5575
-20	0.8655	0.8961	0.8877	0.9164	0.501	0.9217	0.3899	0.9300
-10	0.9219	0.9515	0.9848	0.9748	0.648	0.9613	0.3784	0.9427
0	0.9821	0.9964	0.9929	0.9973	0.9937	0.9883	0.3742	0.9694
10	0.9982	0.9955	0.9940	0.9983	0.9983	0.9973	0.4007	0.9982

Note: Identification accuracy and F-score for different values of SNR in dB. The optimal accuracy and F-score for each value of SNR in dB are highlighted in bold.

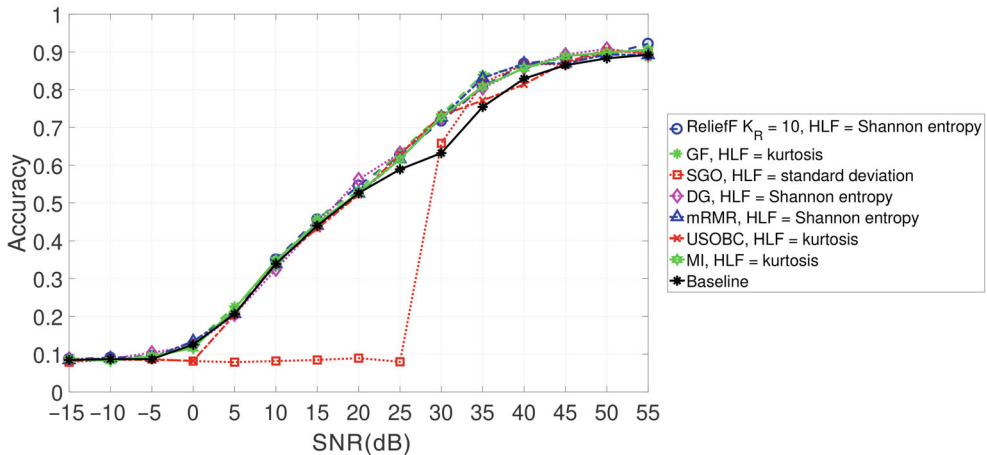


FIGURE 7 DATASET2. Accuracy comparison of the proposed approach based on the application of different feature-selection algorithms with optimal features and hyperparameters against the baseline for different values of SNR.

As shown in Table 3, the relative improvement in accuracy of the approach proposed in this paper for specific FSAs in DATASET2 is even greater than in DATASET1. For example, with an SNR of 35 dB, the approach with GF FSA leads to an 8.6% higher accuracy.

A Wilcoxon rank-sum test was implemented to evaluate the statistical significance test. Examples of the comparison between one FSA and the baseline are shown in the following Table 4 for different values of SNR in dB. A low p -value (less than 0.05) provides evidence to reject the null hypothesis (i.e., that both populations' results are the same) is true.

TABLE 3 DATASET2.

SNR value (dB)	Baseline	ReliefF	GF	SGO	DG	mRMR	USOBC	MI
Accuracy								
0	0.125	0.126	0.127	0.082	0.122	0.134	0.082	0.116
15	0.440	0.456	0.459	0.085	0.443	0.440	0.432	0.447
35	0.754	0.808	0.840	0.817	0.804	0.831	0.772	0.808
55	0.892	0.922	0.885	0.901	0.894	0.892	0.896	0.906
F-score								
0	0.133	0.123	0.120	0.084	0.116	0.133	0.083	0.107
15	0.432	0.455	0.447	0.080	0.438	0.426	0.430	0.437
35	0.727	0.788	0.843	0.825	0.789	0.830	0.795	0.801
55	0.885	0.921	0.882	0.900	0.896	0.897	0.900	0.914

Note: Identification accuracy and F-score for different values of SNR in dB. The optimal accuracy and F-score for each value of SNR in dB are highlighted in bold.

TABLE 4 Wilcoxon rank-sum test between one FSA and the baseline for both data sets.

DATASET1 (ReliefF against baseline)	SNR = -20 dB	SNR = 0 dB
<i>p</i> -value	0.0018	0.000036
DATASET2 (mRMR against baseline)	SNR = 30 dB	SNR = 40 dB
<i>p</i> -value	0.0000364	0.000097

Apart from better identification accuracy, the main achievement of the proposed approach is the computational efficiency, even taking into account that the feature matrix has to be computed and the FSA applied to it (this computation is not present in the baseline approach). As described in the previous sections of this paper, this was the main goal of the approach, which was successfully achieved.

Table 5 shows the computation times using the optimal HLF for each FSA (e.g., mean for SGO) for DATASET1 and DATASET2. For DATASET1 the values are reported at SNR = 20 dB, while for DATASET2 the values are reported at SNR = 45 dB (similar results were obtained for other SNR values, but they are not presented here because of space limitations).

The reported time given in the column 'Time (s) to generate the feature matrix' is specific to the HLF with the best performance, as shown in Table 5. The calculation of the feature matrices FM_j is independent of the application of the FSA. The other Table 6 gives a complete overview of all the computations for all the HLFs and FSAs for both DATASET1 and DATASET2. The gain in computation time is given in parentheses in the last column of Table 5. The gain is calculated as $(T_{Baseline} - T_{FSA})/T_{Baseline}$ and given in percentages (i.e., %).

Table 5 indicates that while most of the computation time is still due to the application of the CNN, the dimensionality reduction manages to achieve a reduction of the computational time on the computing platform used in this study (e.g., 176 s against 567 s for the application of the baseline approach) especially for DATASET1. It is possible to achieve around 70% gain for most of the FSAs. The optimal computing time gain is achieved with MI FSA. For DATASET2, the gain



TABLE 5 Computation times of the different approaches.

DATASET1: Approach (SNR = 20 dB)	Time (s) for feature matrix creation	Time (s) for FSA execution	Time (s) for CNN	Total execution time (s) and (Gain in %)
This approach (GF)	57	1	101	159 (+72)
This approach (ReliefF)	58	16	102	176 (+69)
This approach (SGO)	32	40	104	176 (+69)
This approach (DG)	32	43	103	178 (+68.6)
This approach (mRMR)	46	5	102	153 (+72)
This approach (USOBC)	58	82	103	243 (+57.1)
This approach (MI)	46	4	102	152 (+73.2)
Baseline	0	0	567	567
DATASET2: Approach (SNR = 55 dB)	Time (s) for feature matrix creation	Time (s) for FSA execution	Time (s) for CNN	Total execution time (s) and (Gain in %)
This approach (GF)	11	1	123	134 (+20.2)
This approach (ReliefF)	6	15	123	144 (+14.3)
This approach (SGO)	5	163	123	291 (-73.2)
This approach (DG)	6	132	124	262 (-56.0)
This approach (mRMR)	6	3	122	131 (+22.0)
This approach (USOBC)	11	232	123	366 (-117.9)
This approach (MI)	11	3	123	137 (+18.5)
Baseline	0	0	168	168

TABLE 6 Computation times in seconds for the creation of the feature matrices FM_j .

HLF	DATASET1	DATASET2
Mean	32	4
Standard deviation	40	5
Shannon entropy	46	6
Skewness	57	11
Kurtosis	58	11

in computing time is less than the one reported in DATASET1, but it is still significant, as it is possible to achieve a gain in computing time of around 20%. This is probably because the spectral domain representation SDAR is much larger in DATASET1 (15,000) than DATASET2 (1300). The optimal computing time is achieved with mRMR FSA. In general, both mRMR and MI (which are based on a similar design) are efficient in computing terms. Taking into consideration the reported accuracy and F-score values from Tables 2 and 3, which are generally superior to the baseline, mRMR and MI are to be preferred regarding the objective of computing efficiency. In this data set it should also be noted that some FSAs are slow to converge to the optimal set of features.

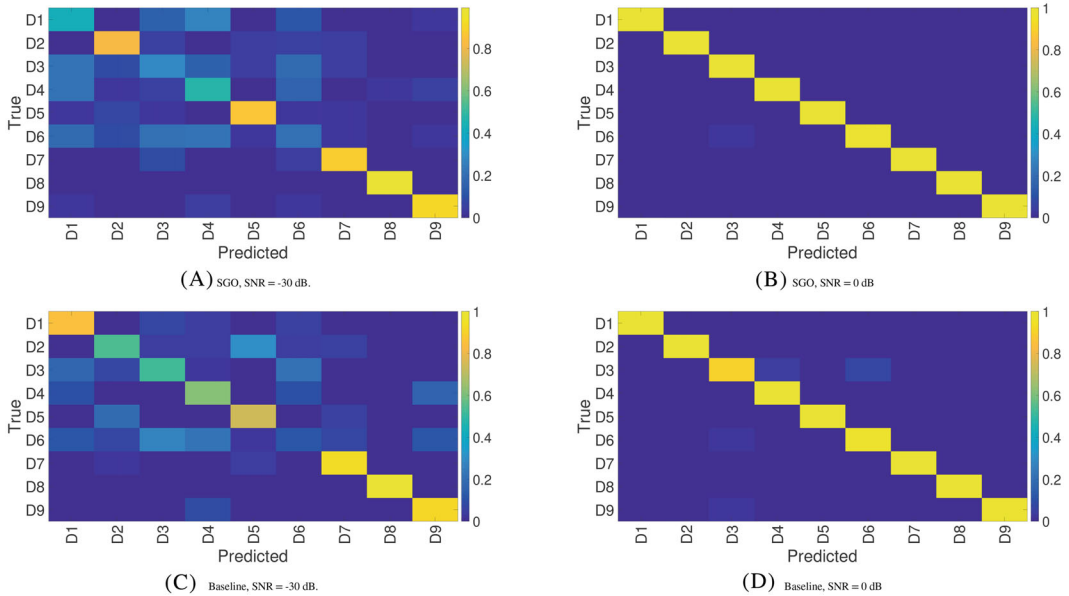


FIGURE 8 DATASET1. Confusion matrices using the SGO approach with HLF = mean (Figure 8A,B) and the baseline (Figure 8C,D) at SNR = -30 dB and SNR = 0 dB. The true values are on the y-axis and the predicted values are on the x-axis.

In particular, we record a very long computing time for the application of SGO, DG and USOBC for this specific data set, which practically excludes them from their application in this data set as the time gain is negative: the sum of the computing times in the pre-processing steps is greater than the baseline case.

To complement the reported values of accuracy and F-score, we also present in the following figures, examples of confusion matrices obtained for both s for the baseline and a specific FSA. Figure 8 shows the confusion matrices obtained with the SGO FSA and the baseline for DATASET1, where the device (i.e., a drone controller) is represented with the term DX ($x = 1 \dots 9$). Figure 9 shows the confusion matrices obtained with the ReliefF FSA and the baseline approach for DATASET2, where the device (i.e., a GSM mobile phone) is represented by the term GX ($x = 1 \dots 12$). In both figures the true values are on the y-axis while the predicted values are on the x-axis.

Figure 8 shows that the confusion matrices at SNR = -30 dB indicate a large number of errors (e.g., FPs and FNs) because the approach has difficulty in discriminating the fingerprints of the drone controllers in the presence of noise. On the other hand, the confusion matrix obtained with the SGO FSA has fewer errors (i.e., values outside the diagonals) than the confusion matrix obtained with the baseline approach, suggesting that the approach proposed in this paper is superior to the baseline approach in terms of prediction accuracy. Similar results (although less clear since the overall accuracy is higher) are shown for the confusion matrices at SNR = 0 dB.

Additionally, it can be seen that RF fingerprints in some data sets are easier to recognize than others: D7, D8, and D9. This is due to the fact that D7, D8, and D9 are different models and the classification between models (i.e., inter-model classification) is easier than the classification within a model (i.e., intra-model classification).^{1,2} The reason is that different models of wireless devices (i.e., inter-model set), which are still conformal to the same wireless standard are designed and

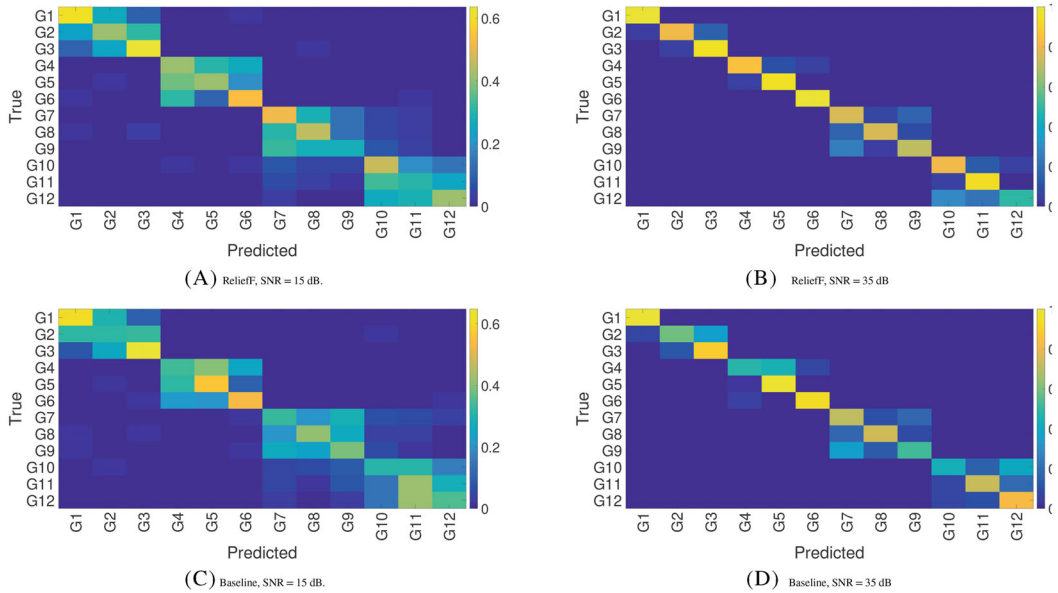


FIGURE 9 DATASET2. Confusion matrices using the ReliefF approach with $K_R = 10$ with HLF = Shannon entropy (Figure 9A,B) and the baseline (Figure 9C,D) at SNR = 15 dB and SNR = 35 dB. The true values are on the y-axis and the predicted values are on the x-axis.

manufactured in different design and production environments, which generates differences in their structure, which translates to different RF fingerprints. On the other hand, wireless devices of the same model (i.e., intra-model set) but different serial numbers have the same design and minor differences in their structure due to the manufacturing process and materials. Because the physical differences among wireless devices of the same model are minor in comparison to wireless devices of different models, the CNN has more difficulty in distinguishing intra-model wireless devices.

In a similar way, we present in Figure 9 the confusion matrices obtained with the DATA SET2 using the ReliefF ($K_R = 10$ and Shannon entropy) obtained at SNR = 15 dB and SNR = 35 dB. The aspect related to the intra-model classification is even more evident in Figure 9, because it is clear that for SNR = 15 dB the approach has some difficulties in identifying the mobile phones belonging to the same model.

6.3 | Evaluation of the impact of parameters

In summary, the proposed approach can be based on a number of parameters: (1) the type of HLF (e.g., mean, standard deviation) used to generate the feature matrix FM_i on which the FSA is applied, (2) the number of highest-ranking segment features N_F obtained from the FSA, (3) the type of FSA algorithm and (4) other hyper-parameters present in the definition of the FSA. In this study, only the ReliefF algorithm has the hyperparameter K_R to be investigated. In this section we evaluate the impact of parameters 1,2, and 4 with the ReliefF FSA and the next subsection will present the results of the evaluation of the different FSAs in comparison to the baseline. The baseline is the approach where the entire spectral domain representation is given as an input to the CNN without the application of the FSA.

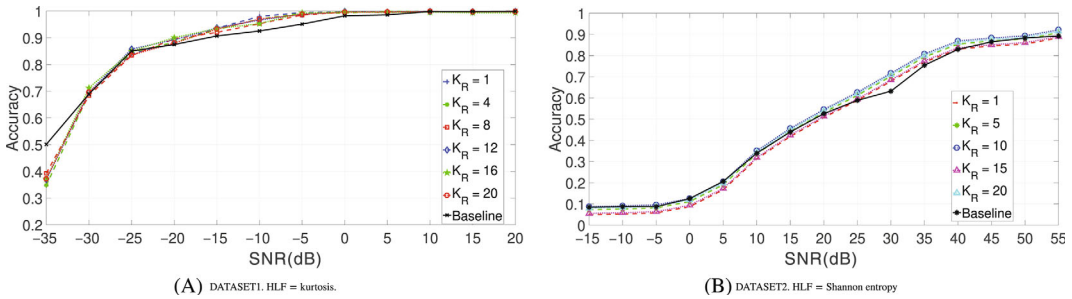


FIGURE 10 Comparison of the proposed approach based on the application of the ReliefF algorithm against the baseline for different values of SNR in dB and values of the ReliefF hyperparameter K_R .

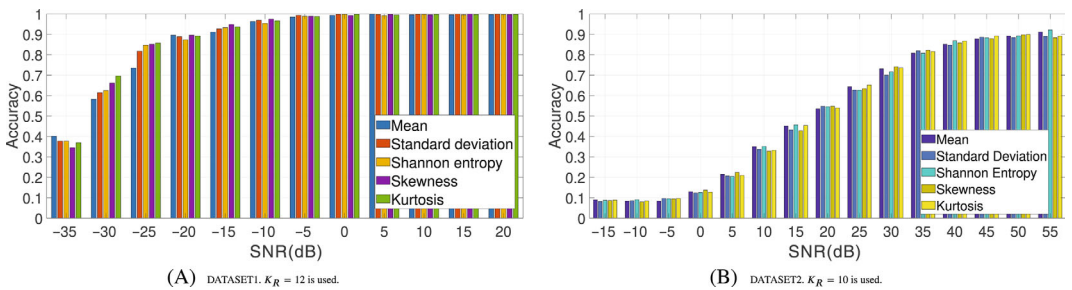


FIGURE 11 Comparison of the proposed approach based on the application of the reliefF algorithm for different HLFs.

In the case of the ReliefF approach, K_R is also a hyperparameter. A grid approach with K_R ranging from 1 to 20 was used to estimate the optimal value of the hyperparameter using accuracy as a metric. Each FSA was tuned for the use of the features and the ReliefF algorithm was optimized for the value of K_R . The results of the optimization of the approach based on the ReliefF algorithms are shown in Figure 10A for DATASET1 (HLF is kurtosis) and in Figure 10B for DATASET2 (HLF is Shannon entropy). The proposed approach is able to outperform the baseline for the majority of the values of SNR in dB apart from SNR = -35 dB when the noise is too high for the ReliefF algorithm to be able to select the appropriate segments for classification. It is clear from Figure 10 that both the values of $K_R = 12$ and $K_R = 16$ achieve the optimal performance at different values of SNR in dB.

Based on the results above, we also evaluated the impact of the choice of the HLF. The Figure 11 of the bar type shows the effects of the choice of HLF in the case of FSA ReliefF with the specific value $K_R = 12$ and for different values of SNR in dB. It can be seen that for the ReliefF FSA, the kurtosis HLF provides the optimal accuracy for most of the values of SNR in dB, with the HLF skewness also achieving a relatively good performance. This is not unusual because the kurtosis feature in the spectral domain has been used in radio-frequency signal processing literature to detect specific patterns, in particular in the transient portions of the signal.³⁷

The number of the highest-ranking segment features N_F obtained from the FSA is another parameter to optimize. This parameter is also important for the proposed approach: if the value of N_F is too large, the overall approach would not be time-efficient because the CNN has samples that are too large as the input. Moreover, segments with low discriminative value could be included

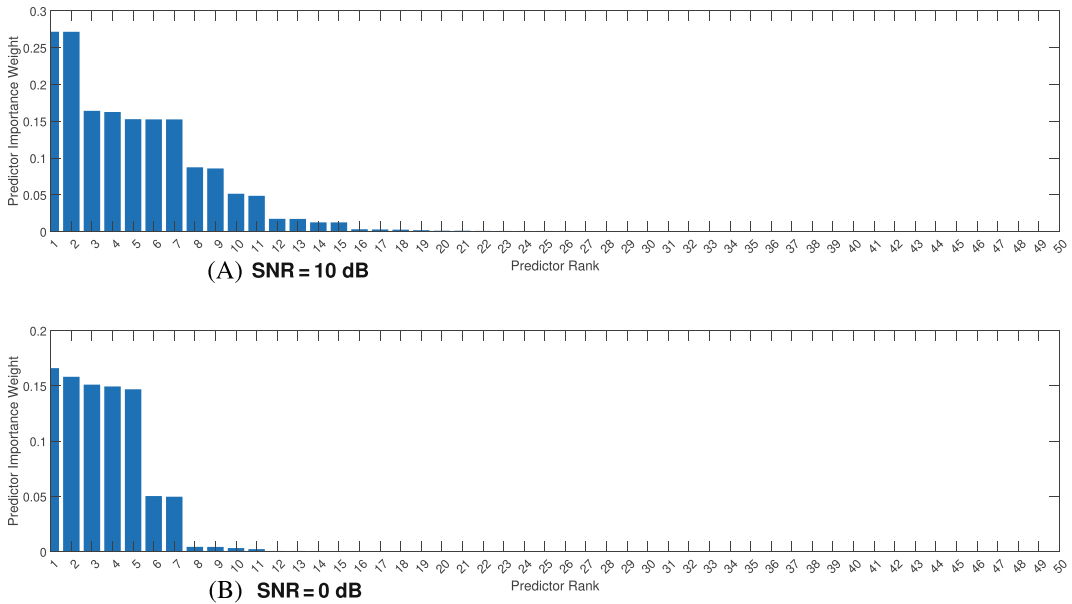


FIGURE 12 Weight ranking of the segment features using ReliefF with $K_R = 12$ for values of SNR = 0 dB and SNR = 10 dB. Only the highest-ranking 50 segment features are shown for readability.

in the model DL. If N_F is too small, the approach might not be able to select the segments with the highest discrimination value, resulting in a worse overall classification. The term 'too small' depends on the data set's properties. There are different potential approaches to determine the optimal value of N_F . One possibility would be to use a similar approach to the other parameters (e.g., the choice of the FSA) where the classification using the CNN is performed for each value of the parameter (e.g., the specific FSA algorithm), but this would be time consuming. In this case a more efficient approach is to evaluate the weights of the ranked segment features provided directly by the application of the FSA, which can provide not only the ranking of the features but also their relative weights. Then, we can use this information to select the optimal value of N_F .

Figure 12 shows the relative weights of the ranking of the segment features for two different values of SNR = 0 dB and SNR = 10 dB calculated with the ReliefF algorithm and $K_R = 12$ for DATASET1. A limit of 50 was set in the figures instead of the full size of the spectral domain for readability (in the current study there are 150 segment features with the maximum possible value of $N_F = 150$). A threshold of $N_F = 20$ includes most of the relevant segment features and that it is not necessary to increase this value. In fact, the number of high-rank features decreases with the lower values of SNR = 0 dB. Similar results are obtained with the other FSAs. A similar approach and result was obtained for DATASET2 with $N_F = 10$ but is not shown here because of space limitations.

7 | CONCLUSIONS

This paper describes a novel approach to improving the computational efficiency and accuracy in RFF based on a pre-processing step prior to applying a CNN to the spectral domain representation of the digital signal emitted by wireless devices. The approach is based on two main hypotheses:

(a) that not all parts of the spectral domain have the same discriminating value because the RF fingerprints are located in certain parts of the spectral domain and (b) that such discriminating parts can be identified with a feature-selection approach that is more time efficient than using a CNN for the same purpose. In practice the approach proposes a combination of feature selection and deep learning (CNN), where feature selection is used to support a dimensionality reduction of the input data for the benefit of the CNN. The approach is evaluated on two different public data sets and achieves a significant improvement over using a CNN directly, both in terms of classification (measured by accuracy and F-score) and computational time. The trade-off is the selection of the appropriate FSA, as differences in the application of FSAs were found.

Future developments of this study could take several directions. One direction would be to study in more detail the parameters associated with data segmentation; another direction would be to study the application of transforms other than the fast Fourier transform. A third direction would be to investigate this unsupervised learning approach with an open data set of wireless devices.

AUTHOR CONTRIBUTIONS

Conceptualization, Gianmarco Baldini; Methodology, Gianmarco Baldini, Irene Amerini; Data Curation, Gianmarco Baldini; writing Gianmarco Baldini, Franc Dimc, Irene Amerini, Fausto Bonavitacola. All authors have read and agreed to the published version of the manuscript.

ACKNOWLEDGMENTS

We acknowledge the authors of Reference 33 for the publication of their data set.

CONFLICT OF INTEREST STATEMENT

The authors declare no potential conflict of interests.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in drone remote controller RF signal dataset at <http://dx.doi.org/10.21227/ss99-8d56>.

REFERENCES

1. Baldini G, Steri G. A survey of techniques for the identification of mobile phones using the physical fingerprints of the built-in components. *IEEE Commun Surv Tutor*. 2017;19(3):1761-1789.
2. Xu Q, Zheng R, Saad W, Han Z. Device fingerprinting in wireless networks: challenges and opportunities. *IEEE Commun Surv Tutor*. 2016;18(1):94-104. doi:10.1109/COMST.2015.2476338
3. Reising DR, Temple MA, Jackson JA. Authorized and rogue device discrimination using dimensionally reduced RF-DNA fingerprints. *IEEE Trans Inf Forens Secur*. 2015;10(6):1180-1192. doi:10.1109/TIFS.2015.2400426
4. Almashaqbeh H, Dalveren Y, Kara A. A study on the performance evaluation of wavelet decomposition in transient-based radio frequency fingerprinting of Bluetooth devices. *Microw Optic Technol Lett*. 2022;64(4):643-649.
5. Merchant K, Revay S, Stantchev G, Nousain B. Deep learning for RF device fingerprinting in cognitive communication networks. *IEEE J Sel Topics Signal Process*. 2018;12(1):160-167.
6. Baldini G, Gentile C, Giuliani R, Steri G. Comparison of techniques for radiometric identification based on deep convolutional neural networks. *Electron Lett*. 2018;55(2):90-92.
7. Patel H, Temple MA, Ramsey BW. Comparison of high-end and low-end receivers for RF-DNA fingerprinting. Paper presented at: 2014 IEEE Military Communications Conference. 2014 24-29.



8. Shen G, Zhang J, Marshall A, Peng L, Wang X. Radio frequency fingerprint identification for LoRa using spectrogram and CNN. Paper presented at: IEEE INFOCOM 2021-IEEE Conference on Computer Communications. 2021 1–10.
9. Reising DR, Temple MA, Oxley ME. Gabor-based RF-DNA fingerprinting for classifying 802.16e WiMAX mobile subscribers. Paper presented at: 2012 International Conference on Computing, Networking and Communications (ICNC). 2012:7-13. doi:10.1109/ICCNC.2012.6167534
10. Ślęzak D, Grzegorowski M, Janusz A, et al. A framework for learning and embedding multi-sensor forecasting models into a decision support system: a case study of methane concentration in coal mines. *Inform Sci*. 2018;451:112-133.
11. Tc F. A review on time series data mining. *Eng Appl Artif Intel*. 2011;24(1):164-181.
12. Guyon I, Gunn S, Nikravesh M, Zadeh LA. *Feature Extraction: Foundations and Applications*. Vol 207. Springer; 2008.
13. Nie F, Zhu W, Li X. Unsupervised feature selection with structured graph optimization. Proceedings of the AAAI Conference on Artificial Intelligence. 2016 30(1).
14. Robnik-Šikonja M, Kononenko I. Theoretical and empirical analysis of ReliefF and RReliefF. *Mach Learn*. 2003;53(1):23-69.
15. Gu Q, Li Z, Han J. Generalized fisher score for feature selection. Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence. 2011 266–273.
16. Guo J, Zhu W. Dependence guided unsupervised feature selection. Proceedings of the AAAI Conference on Artificial Intelligence. 2018 32(1).
17. Peng H, Long F, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Anal Mach Intell*. 2005;27(8):1226-1238.
18. Nguyen XV, Chan J, Romano S, Bailey J. Effective global approaches for mutual information based feature selection. Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2014 512–521.
19. Han D, Kim J. Unsupervised simultaneous orthogonal basis clustering feature selection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015 5016–5023.
20. Herman G, Zhang B, Wang Y, Ye G, Chen F. Mutual information-based method for selecting informative feature sets. *Pattern Recognit*. 2013;46(12):3315-3327.
21. Zhang Z, Liang X, Qin W, Yu S, Xie Y. matFR: a MATLAB toolbox for feature ranking. *Bioinformatics*. 2020;36(19):4968-4969.
22. Li Z, Liu F, Yang W, Peng S, Zhou J. A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE Trans Neural Netw Learn Syst*. 2021;33(12):6999-7019.
23. Chen J, Ran X. Deep learning with edge computing: a review. *Proc IEEE*. 2019;107(8):1655-1674.
24. Nebauer C. Evaluation of convolutional neural networks for visual recognition. *IEEE Trans Neural Netw*. 1998;9(4):685-696.
25. Soltanieh N, Norouzi Y, Yang Y, Karmakar NC. A review of radio frequency fingerprinting techniques. *IEEE J Radio Frequen Identif*. 2020;4(3):222-233.
26. Scanlon P, Kennedy IO, Liu Y. Feature extraction approaches to RF fingerprinting for device identification in femtocells. *Bell Labs Techn J*. 2010;15(3):141-151. doi:10.1002/bltj.20462
27. Zhang J, Wang Q, Guo X, Zheng X, Liu D. Radio frequency fingerprint identification based on logarithmic power cosine spectrum. *IEEE Access*. 2022;10:79165-79179.
28. Kennedy IO, Scanlon P, Mullany FJ, Buddhikot MM, Nolan KE, Rondeau TW. Radio transmitter fingerprinting: a steady state frequency domain approach. Vehicular Technology Conference, 2008. VTC 2008-Fall. IEEE 68th. 2008 1–5.
29. Riyaz S, Sankhe K, Ioannidis S, Chowdhury K. Deep learning convolutional neural networks for radio identification. *IEEE Commun Mag*. 2018;56(9):146-152.
30. Jian T, Gong Y, Zhan Z, et al. Radio frequency fingerprinting on the edge. *IEEE Trans Mob Comput*. 2022;21(11):4078-4093.
31. Asgharzadeh-Bonab A, Amirani MC, Mehri A. Spectral entropy and deep convolutional neural network for ECG beat classification. *Biocybern Biomed Eng*. 2020;40(2):691-700.
32. Ezuma M, Erden F, Anjinappa CK, Ozdemir O, Guvenc I. *Drone Remote Controller RF Signal Dataset*. 2020. doi:10.21227/ss99-8d56

33. Ezuma M, Erden F, Anjinappa CK, Ozdemir O, Guvenc I. Micro-UAV detection and classification from RF fingerprints using machine learning techniques. Paper presented at: 2019 IEEE Aerospace Conference. 2019 1–13.
34. Baldini G, Gentile C. Transient-based internet of things emitter identification using convolutional neural networks and optimized general linear chirplet transform. *IEEE Commun Lett*. 2020;24(7):1482-1486.
35. Köse M, Taşcioğlu S, Telatar Z. RF fingerprinting of IoT devices based on transient energy spectrum. *IEEE Access*. 2019;7:18715-18726.
36. Baldini G. GSM Radio Frequency Fingerprinting data set. 2023. doi:[10.21227/c7sa-wq24](https://doi.org/10.21227/c7sa-wq24)
37. Nita GM. Spectral kurtosis statistics of transient signals. *Month Notice R Astronom Soc*. 2016;458(3):2530-2540.

How to cite this article: Baldini G, Amerini I, Dimc F, Bonavitacola F. Convolutional neural networks combined with feature selection for radio-frequency fingerprinting. *Computational Intelligence*. 2023;39(5):734-758. doi: 10.1111/coin.12592