IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 26, 2019, accepted April 4, 2019, date of publication April 15, 2019, date of current version April 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2911452

Assessment of Features and Classifiers for Bluetooth RF Fingerprinting

AYSHA M. ALI, EMRE UZUNDURUKAN, AND ALI KARA, (Senior Member, IEEE)
Department of Electrical and Electronics Engineering, Atilim University, 06830 Ankara, Turkey
Corresponding author: Ali Kara (ali.kara@atilim.edu.tr)

ABSTRACT Recently, network security has become a major challenge in communication networks. Most wireless networks are exposed to some penetrative attacks such as signal interception, spoofing, and stray. Radio frequency (RF) fingerprinting is considered to be a promising solution for network security problems and has been applied with various improvements. In this paper, extensive data from Bluetooth (BT) devices are utilized in RF fingerprinting implementation. Hilbert-Huang transform (HHT) has been used, for the first time, for RF fingerprinting of Bluetooth (BT) device identification. In this way, time-frequency-energy distributions (TFED) are utilized. By means of the signals' energy envelopes, the transient signals are detected with some improvements. Thirteen features are extracted from the signals' transients along with their TFEDs. The extracted features are pre-processed to evaluate their usability. The implementation of three different classifiers to the extracted features is provided for the first time in this paper. A comparative analysis based on the receiver operating characteristics (ROC) curves, the associated areas under curves (AUC), and confusion matrix are obtained to visualize the performance of the applied classifiers. In doing this, different levels of signal to noise ratio (SNR) levels are used to evaluate the robustness of the extracted features and the classifier performances. The classification performance demonstrates the feasibility of the method. The results of this paper may help readers assess the usability of RF fingerprinting for BT signals at the physical layer security of wireless networks.

INDEX TERMS Bluetooth, classification, Hilbert-Huang transform, network security, radio frequency fingerprinting, wireless networks.

I. INTRODUCTION

Network security has become a fundamental issue in the wireless networks. Threats like signal interception, spoofing and strays are considered to be major forms of penetrations. Extensive studies have been conducted to enhance the security of wireless networks at physical layer. Mobile devices, such as cell phones and tablets, are considered as the most widely used wireless devices. These mobile devices need to be supported by multiple security approaches [1].

Radio Frequency (RF) fingerprinting is introduced as a solution to enhance the security at physical layer. RF fingerprinting is a technique by which wireless devices can be identified and classified [2]–[4]. In [2] spectral power density fingerprints are used to provide hardware specific identification. In [3], RF fingerprinting is proposed to mitigate primary user emulation attacks. This mitigation is accomplished via low-end software-defined cognitive radio networks. A recent work [5] confirms that RF fingerprinting is an enhancing security technique for wireless networks. The works presented in [6], [7] used RF fingerprinting for identifying individual transmitters. A study [6] proposed Specific Emitter Identification (SEI) method based on Hilbert-Huang transform (HHT). HHT is applied to obtain time-frequency-energy distributions (TFED) of transient signals [8]–[10]. HHT is a self-adaptive signal analysis method; therefore, no preceding information is needed in signal analysis [6]. However, a study [7] proposed SEI method based on signal's energy trajectory gained by the high order cumulants. The works in [6], [7] applied SEI to the transients of the collected GSM signals.

In RF fingerprinting, the first stage is detection of the signal transient or transient signals of continuously transmitting devices. In the literature, many techniques are employed to detect signal transients [5], [11], [13], [14], [16]–[21] although there is still room for improvements. The next

The associate editor coordinating the review of this manuscript and approving it for publication was Ghufran Ahmed.

is extracting unique features from the detected transients. Finally, the extracted features are used in classification of transmitting devices. It has been seen that HHT based RF finger printing along with multiple classifiers for Bluetooth (BT) signals have not been investigated. Especially, testing and performance of this method with extensive BT data has not been studied yet.

In this study, HHT based RF fingerprinting is implemented for BT signals with relatively large number of devices and transients. BT signals are collected in the laboratory from 20 cell phone devices (5 manufacturers with various model and serial number). 150 transients are recorded from each device. BT signals are captured directly by high sampling rate oscilloscope and are not down converted as in many published works. An improved transient detection technique based on energy envelope is developed. The HHT are applied for producing TFED from which unique features are extracted. The extracted features are further smoothed by filtering in order to increase the classification performance. Three different classifiers are employed and their performances are compared. As to the knowledge of the authors, this work is the first in terms of the size of the data, the practical techniques applied and comparative analysis of the classifiers with such big data set. To investigate the applied techniques to noisy signals, we added different levels of noise captured in the data colletion stage to the transients. The considered classifiers are applied with the features extracted from the noisy transients. A comparison between the performance of the classifiers based on four different levels of noise are provided. The classifiers' performances are evaluated based on criteria as in the literature, namely, confusion matrix, the Receiver Operating Characteristics (ROC) and the area under the ROC curve (AUC).

The paper is organized as follows; section 2 describes the data acquisition system and high level signal processing functionalities. In section 3, an improved algorihm for transient detection is presented. Section 4 presents the HHT/TFED method used in transient signal decomposition. In section 5, the extraction of the features in time and frequency domain (TFED), and their assessments are discussed. Section 6 presents the classifiers, and their performance comparison with the date set. Section 7 demonstrates the effectiveness of the applied techniques with noisy signal along with peformance analysis. Finally, concluions are drawn in section 8.

II. DATA ACQUISITION AND SIGNAL PROCESSING

For precise analysis and quantifying of classification and identification performances, BT signals were captured in laboratory environment. The laboratory is isolated in an underground floor (−2) where no other devices or equipments are switched on. The ambient temperature and humidity are very stable in the laboratory as the change in environmental conditions may affect the performance of classifications [12]. Fig. 1 shows the block diagram of the signal capturing/acquisition and off-line processing.

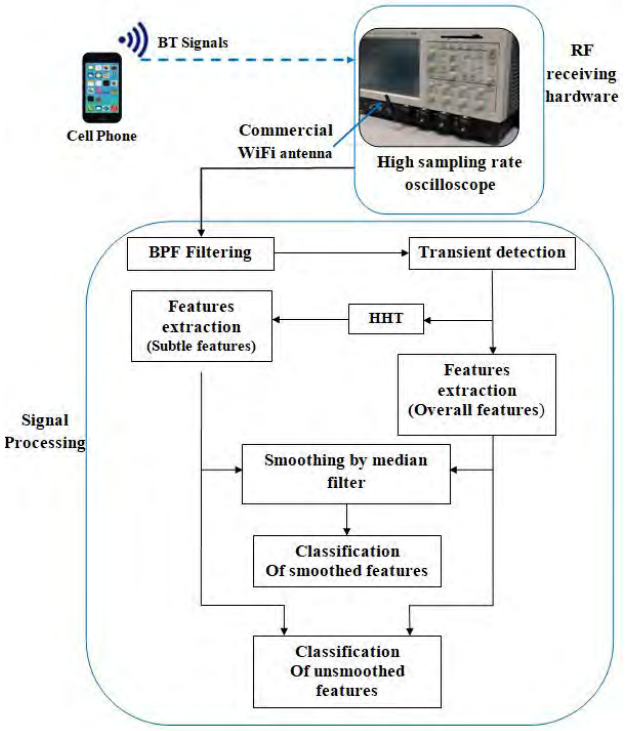

FIGURE 1. Signal acquisition setup and flow chart of signal processing.

BT signals are collected from different cell phone brands, models and serial numbers (identical brand and model but different serial numbers of a model). Bluetooth (BT) system has an operational band between 2400 MHz and 2483.5 MHz (Industrial, Scientific and Medical band or ISM). BT signals are captured directly by high sampling rate oscilloscope (hihg-end device) without down conversion in order to minimize impairments on distinctive transient characteristics. At first, the cell phones are set in flight mode to isolate any unwanted signals generated from the cell phones. Then, BT is switched on manually. By this way, the cell phones only emit BT signals. As the receiver characteristics play a critical role, high end oscilloscope(TDS7404 DSO 4 GHz/20 GSPS) connected to a commerical WiFi modem antenna (operating at 2.4 GHz ISM band) is used to capture BT signals [3]. The performance of high-end devices (expensive equipment such as high sampling, stable, oscilloscopes and/or spectrum analysers) in RF fingerprinting has been studied experimentally in [3]. Typical length of captured signals was 261800, for example for a group of devices, corresponding to 13 μs at 20 GSPS. The distance between the cell phones and the receiving antenna is kept 30 cm for all the devices. According to Nyquist theorem, for a transient at 2.835 GHz signal(the upper bound of BT band), the minimum sampling rate is required to be 4.97 GHz. However, BT signals were sampled at the highest possible frequency (20 GHz) in order to preserve the phase characterstatististics of the transients. The objective is to maximize the use of distinctive characters of the features. Filtering is needed for removing noise and

components introduced by the sampling oscilloscope [13]. A Band Pass Filter (BPF)(FIR filter) was used to eliminate the unwanted frequency components. It was a conventional digital filter to pass only ISM2400 band. In this way, the resultant signals contain only noisy BT emissions. Then, the signals were normalized before the detection of the transients, and extracting of the features. Typical size of the detected transient signals was 30850 samples which corresponds to 1.5 μs duration (it may slightly vary from device to device). Next, HHT is applied to the detected transients and subtle features along with the overall features are extracted in parallel. The subtle and overall features are smoothed by means of median filter (a 4th-order, one-dimensional median filter is implemented) to improve the classification performance. Consequently, we evaluated the classifiers' performances based on smoothed and unsmoothed features to investigate the effects of the smoothing process.

Ten different models among five brands were selected for the experiments. For each model, two different (series or serial number) cell phones were acquired. By this way, a database of 20 cell phones has been created with 150 transients for each device. The total number of the transients was adequate for the analysis of classification performance and the robustness of the features. Table 1 lists the devices, brands and models. Here, A and B indicate the series of the same brand and model devices. A typical transient captured in the laboratory is also shown in Fig. 2. As expected, the detection of the transients is the key point for the analysis of such huge data-set, and then its usability for further work. Then, the first step in the development deals with the transient detection algorithms and possible improvements (detection of start and end points of transient signals).

TABLE 1. The classes.

Class No.	Brand	Model
1	Huawei	G5 A*
2	Huawei	G5 B*
3	I-Phone	5S A
4	I-Phone	5S B
5	I-Phone	6S A
6	I-Phone	6S B
7	I-Phone	6S Plus A
8	I-Phone	6S Plus B
9	I-Phone	7 A
10	I-Phone	7 B
11	LG	G4 A
12	LG	G4 B
13	SamSung	Note3 A
14	SamSung	Note3 B
15	SamSung	S3 A
16	SamSung	S3 B
17	SamSung	S4 A
18	SamSung	S4 B
19	Sony	C4 A
20	Sony	C4 B

A & B refer to two different serial numbers of a given model

III. TRANSIENT SIGNALS DETECTION

The detection of signal transient is an essential step to identify and discriminate transmitting devices via transient based RF fingerprinting techniques. This can be achieved by means of signal characteristics such as amplitude, frequency, phase, and energy. Before the transient detection, the captured signals are pre-filtered to remove undesired components and noise. This is done by applying a band-pass filter (BPF) with cut off frequencies selected for the band of operation [13]. The authors have already studied [22] the performances of various phase and amplitude/energy based techniques. Use of energy envelope of signals is very common technique for the detection of transients, more specifically, the start and end point of transients. However, it is necessary to improve the proposed techniques for both more practical implementations and lower computational loads. This requires somehow intuitive approaches to be implemented with extensive data-set. The algorithm developed for the transient detection follows this approach. The pseudo code that shown in Fig. 3 represents some parts of the algorithm.

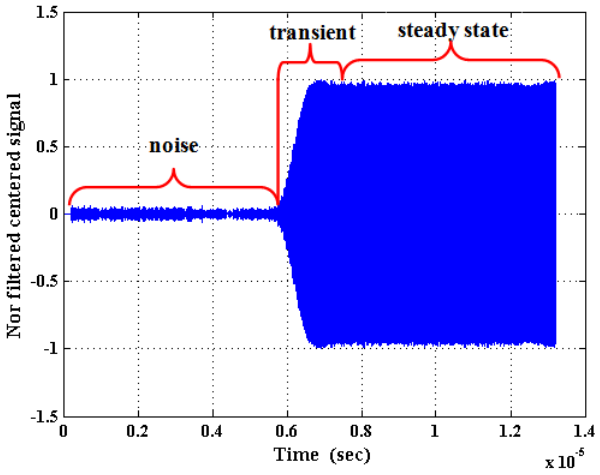

FIGURE 2. A Typical BT signal captured in the laboratory (high SNR).

A sample outcome of the transient detection algorithm is shown in Fig. 4 where the start and the end points of the transients are indicated automatically. The algorithm is based on the use of the local energy envelope maxima. The multiple local energy maximums leads to interference in the transient durations of different classes. Then, the transient duration cannot be used as a robust feature. This is resolved by exploiting the transient features to re-capturing the transient end.

Therefore, the energy based transient detection algorithm has been improved by introducing several more stages. In this technique, the features to be extracted in the next stage is also considered earlier in the detection of the transients. While it requires more signals from multiple devices, overall, this makes the process more accurate, and keep the features more robust. Firstly, the processes estimate the end of the transient using conventional methods. Then, it introduces further stages for precise detection of the transient start and end points. The following illustrates the major steps of the algorithm:

Input: Normalized denoised BT signal.
Output: The transient portion & its features
Initialization: Thresholds. T1 through T6 (*Assigned to fit input signals*)
Steps
Compute the energy envelope (EE) of the input signal
Starting detection by using energy envelope
for *i=1: length (EE)* do
if *T1*mean (EE (I: i+T2)) < EE(i+T3)*
then *Start point=i*
end if
end for
End detection by using energy envelope
for *j= length (EE) :-1: Start point* do
if *mean (EE(length(EE):-1:j-T4))-EE(j)>T5*
assign point j.
end if
end for
Max_point= max(EE(j:T6))
Windowing EE(Max_point:length(EE))
for k=1:# windows do
if *mean(diff(window $_{k-1}$)) <0 && If mean(diff(window $_k$)) >0*
then *the end point=index of window $_k$(1)*
end if
end for

FIGURE 3. Pseudo code of the transient detection algorithm.

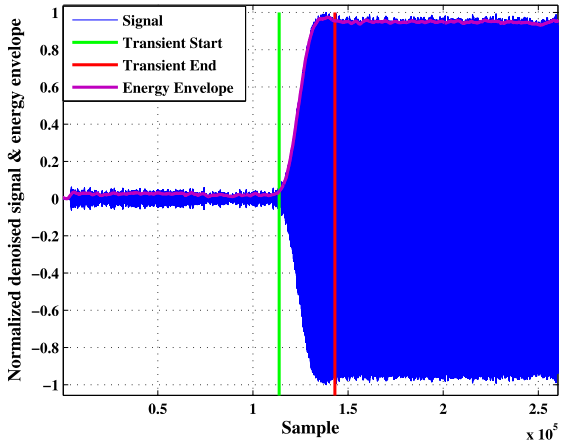

FIGURE 4. Sample BT signal: its energy envelope and components (high SNR).

1) Acquire the transient durations of a couple records (typically 2/3 of the records).
2) Calculate the medians of transient durations.
3) Identify some critical features of each record, and calculate the extreme values (maximum/minimum) of each feature.
4) By means of the parameters in (3), calculate the end of the transient, for all records, by

$$E_{i,j} = S_{i,j} + \widehat{\Delta}_i \pm R_{i,j}, \tag{1}$$

where i represents the index number of class, j represents the index number of record, S represents the start point of the signal transient, E represents the end point of the signal transient, $\widehat{\Delta}$ represents the median of the transients duration of some number of records, R is a random integer.

Note that some of the features are extracted from the transient signal itself before applying HHT, that is, to generate TFED. These are described in details in section 5.

IV. TFED GENERATION

The HHT is applied to the detected transients in order to obtain the Hilbert spectrum of the time-frequency-energy plane, and extract the subtle signal features along the time axis and frequency axis of the TFED. The HHT has been introduced as a data analysis tool for various forms of signals [6], [25]–[27]. Empirical Mode Decomposition (EMD) and Hilbert Transform (HT) are used to accomplish the HHT. The first step of HHT is to obtain intrinsic mode functions (IMFs) of the transient signals. This can be done by applying the EMD to the transient signal.

A. EMPIRICAL MODE DECOMPOSITION (EMD)

The EMD is a systematic way, involves approximation with splines, to extract IMFs from a signal [28], [29] based on a sifting procedure [30]. The generated IMFs which are a series of stationary and linear sub-signals must satisfy the two mandatory conditions, for more information refer to [6]. The EMD applied to the detected transient signal of each BT signal in order to generate its IMFs. The Hilbert transform is needed to be applied to each extracted IMFs to generate their instantaneous characteristics. These characteristics are expressed as the instantaneous amplitude (IA), instantaneous phase (iph), and Instantaneous frequency (IF) [5], [28]. Having obtained the characteristics of each IMF, the Hilbert spectrum (HS) can be generated.

B. HILBERT SPECTRUM(HS)

The Hilbert Spectrum represents a three dimensional plot that is illustrating the energy distribution as a function of time and frequency [31], [32]. In order to help readers implement HS when necessary, some major steps of the implementation are shared in this work. The major steps in generating the HS can be described as follows:

1) Introduce the generated instantaneous characteristics of the IMFs of the transient
2) Calculate the scaled instantaneous frequency S_{IF} from

$$S_{IF} = \frac{ULFS}{max(IFs) - min(IFs)} * IFs, \tag{2}$$

where ULFS is the upper limit of frequency scale, and IFs are the instantaneous frequencies of all IMFs.
3) Initialize the number of frequency bins (FB) and the number of time segments (TS), and define the bin space as $\frac{ULFS}{FB}$.
4) For each IMF, obtain the HS as a weighted sum of IA at the m_{th} frequency bin along time segments as

$$HS_b(m, t) = IA_b(t)\omega^m(t), \tag{3}$$

where b represents the index of imf, $IA_b(t)$ is the corresponding instantaneous amplitude to the imf_b. The

weight factor $\omega^m(t) = 1$, if the $S_{IF}(t)$ lies in the m_{th} frequency bin, else $\omega^m(t) = 0$.

5) Construct the overall HS of the signal transient by using

$$HS = \sum_{b=1}^{B} (HS_b(m, t)). \quad (4)$$

where B is the number of the IMFs of the transient.

A sample Hilbert spectrum obtained from captured transients is shown in Figure 5. Here, the TFED is expressed as normalized amplitude distribution over the time and frequency plane.

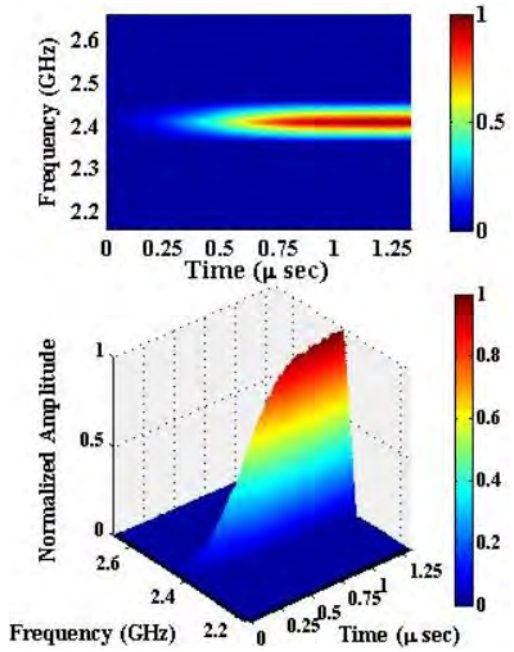

FIGURE 5. Typical Hilbert spectrum sample of BT signals.

V. EXTRACTION OF FEATURES

In RF fingerprinting, the transient signal or steady signal signal can be used in feature extraction [33]. Each might have show different performance depending on the signal types and feature sets. This has been reported extensively in the literature for different signals and systems. While the features are extracted from the steady state signal in [33]–[36], transient signals are used in [6], [7], [12], [14], [15]. In this work, the features are extracted from the transient signals as the literature reports higher classification performance [33]. That is because the transients of devices operating at ISM band are long enough to extract distinguishing features without the steady state signals. More importantly, the transient signals contains some unique hardware characteristics that cannot be forged easily [15]. Therefore, the transmitted BT signals contain uniquely identifying features (fingerprints) of the devices classes) that can be used in RF fingerprinting development. It has been known that unique features could be extracted via the transient signal transformations or decompositions in time or frequency domains. However, some of the features can be directly extracted from the transient signal, for example, the duration of the transient signal or the slope of the transient. Some other features can be extracted from the transients' instantaneous characteristics such as the statistical moments of instantaneous amplitude, phase and frequency distribution [6].

Normalization of the extracted features is necessary in case of comparing two or more devices according to two or more features, shown in Fig. 9 and Fig. 10, as the difference of features ranges may vary. On the other hand, normalization of extracted features is unecessary while comparing two or more devices according to one feature, shown in Fig. 6, Fig. 7 and Fig. 8.

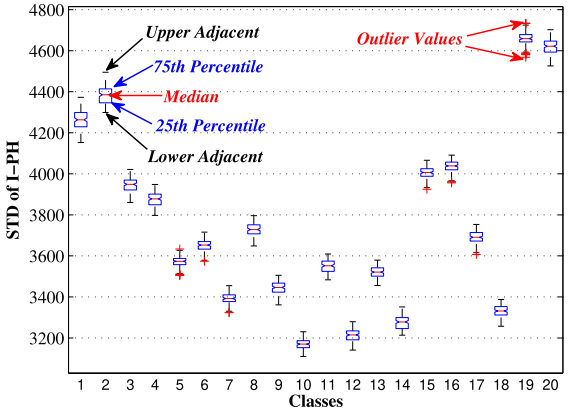

FIGURE 6. std of the 20 cell phones' transients iph.

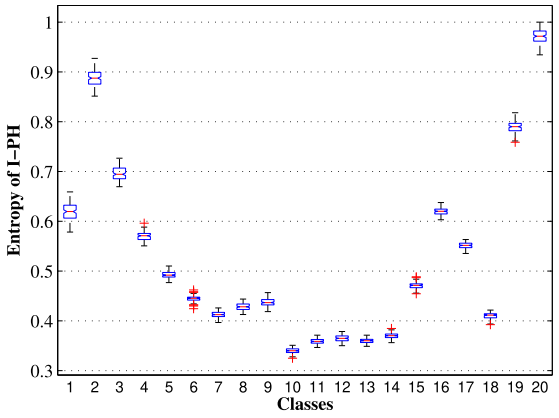

FIGURE 7. The entropy of the 20 cell phones' transients iph.

Features that are considered in this study are listed in Table 2. The features in the table are divided into three groups. Those in the first group are extracted directly from the transient signal (overall features). The other two groups of features are extracted from the TFED of BT signals (subtle features).

In Table 2, the Polyfit coefficient of sum of transient energy distribution along time axis (f_{11}) is referring to a Matlab's polyfit function, where the extracted feature is the coefficient of the highest order term of the polynomial.

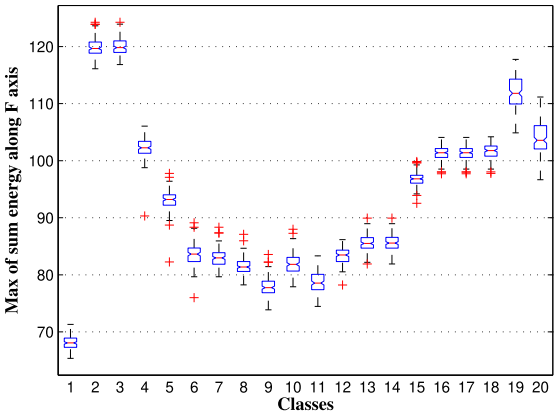

FIGURE 8. Maximum of summation energy along frequency axis of TFED.

TABLE 2. Feature sets in RF fingerprints.

Feature group	Feature name	Feature label
Transient signal and energy envelope	Duration of transient	f_1
	Total energy of transient energy	f_2
	Total energy of transient energy envelope	f_3
	Variance of transient energy envelope	f_4
	Std of inst. phase of transient signal	f_5
	Entropy of inst. phase of transient signal	f_6
TFED of the transient signal along time axis	Length of transient energy distribution	f_7
	Slope of transient energy distribution	f_8
	Variance of sum of transient energy distribution	f_9
	Maximum of sum of transient energy distribution	f_{10}
	Third order polynomial fitting coefficient of sum of transient energy distribution	f_{11}
TFED of the transient signal along frequency axis	Maximum of sum of transient energy distribution	f_{12}
	Variance of sum of transient energy distribution	f_{13}

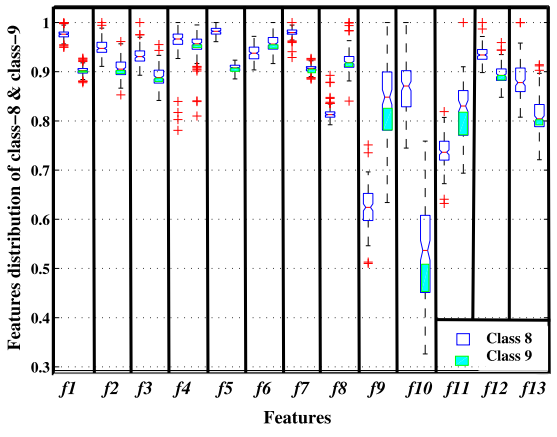

FIGURE 9. Normalized features distribution of two same brand different model cell phones.

Total of 13 features with their labels are listed in the table. By means of the extracted features, the feature vector set is formed. One of the aims of this paper is to analyse the effects of filtering of the features on the classification performance. In order to do so, each feature set is smoothed by using median filter before classification (we used the 4th-order, one-dimensional median filter). Then, the classification are performed with the unsmoothed and smoothed features in order to explore the effects of smoothing process. However, firstly, the robustness of the features can be examined by using Box plot technique. This technique is used to illustrate, to some extent, the discrimination capability of the extracted features. Typical plots are constructed by means of the following five criteria (shown in Fig. 6): The lower adjacent, the 25th percentile, the median, the 75th percentile, and the upper adjacent. The 25th percentile is a value of the considered data group at which 25% of the data values are less than it. A 50% of the data group or the inter-quartile range (IQR) are located between the 25th percentile and the 75th percentile. This means that 25th of the data values are greater than the 75th percentile value. The median criterion, which is illustrated by a red line in the middle of the box, represents the 50% percentile. The lower and upper adjacent represent the whisker boundaries. These boundaries can be determined by multiplying a value (usually 1.5) times the IQR, and subtracting or adding it to the 25th percentile or the 75th percentile, respectively. The values of the data group that are out of the whisker boundaries are outliers [37].

The standard deviation (std) of the transients' instantaneous phases (*iph*) is considered (see Fig. 7) as one of the most important discriminative features. The feature is completely separable, if the box values, including the whiskers and outlier values, of a certain class are separated from the boxes' values of other considered classes. For instant, the std of iph box values of class 1 is completely separated from the boxes' values of the classes 3 through 20. However, the feature is considered to be a nearly separable feature, if only the whiskers or outliers are inseparable. As an example, (see Fig. 7), the std of iph box values of class 1 is nearly separated from the box values of the class 2. If the feature is not adequate to completely separate the considered classes, some other features can be included. In order to illustrate this, two more features are considered, namely the entropy of transients' iphs and the maximum of summation energy along frequency axis of TFED (see Fig. 8, and Fig. 9), respectively. With the new features, the classes 1 and 2 are completely separated. Two different models of the cell phones from the same brand, namely I-Phone-6s-Plus-B(Class-8) and I-Phone-7-A (class-9) are subjects to different features (see Fig. 10). These two classes have been identified to be the most challenging among the considered classes in our study. However, it should be noted that some features can distinguish the class-8 and the class-9 perfectly. It seems that the features $f7$, $f8$, & $f10$; and the features $f1$, & $f5$ would work well in this case. Moreover, as the next stage, some features are quite discriminative for the cell phones of the same model (the same brand and the same model but different serial number) as shown in (see Fig. 10). In this figure, Samsung Note-3-A and Samsung Note-3-B, which are defined by class-13 and the class-14, respectively, can be perfectly distinguished. This

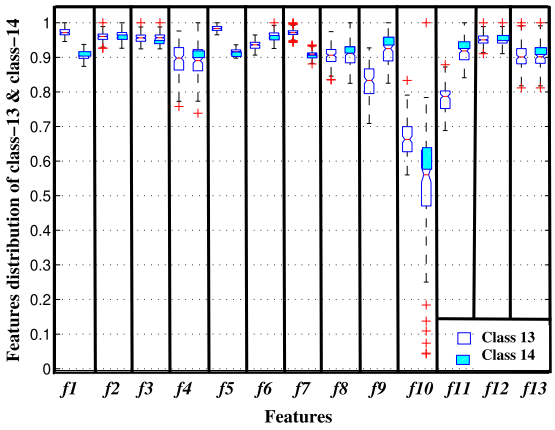

FIGURE 10. Normalized features distribution of two same brand same model different serial number cell phones.

is due to the completely separable features $f1$, & $f5$; and the nearly separable $f7$, $f9$, & $f11$. Each of the rest of the twenty studied classes is classified and identified based on its extracted unique features. Based on the BT signals' records extracted features of the studied cell phones, the classification of these devices can be accomplished.

VI. CLASSIFICATION

Classifier performance has not been studied comprehensively in RF fingerprinting of BT signals. This section describes how classifiers are implemented on the features discussed in the previous section. In reference with Table 1 and 2, thirteen extracted features represent the RF fingerprints of twenty classes each with one hundred fifty records. Each extracted feature data of each class is divided into two data groups namely, training set and testing set, with 40% (60 records) and 60% (90 records) of the total data (150 records), respectively. For each set, several classifiers, namely, complex decision tree, Linear Support Vector Machine (L-SVM), and Linear Discriminant Analysis (LDA) are applied.

The complex decision tree is a supervised learning algorithm that is very common in classification problems. The tree is composed of three types of nodes; root node which contains all the data, splits nodes which generate internal nodes of the tree, and leaves which are defined as terminal nodes. The splits nodes are estimated from a statistical procedure applied to the training data. The splits nodes are used as a mapper from the root node to the leaf decision nodes. The main idea of this procedure is splitting the data into subsets, each subset belongs to unique class. The procedure of the decision tree classifiers depends on recursively partitions of the input data set into smaller subdivisions. The learning algorithms of the decision tree require high-quality training data to generate the model. Therefore, a set of training data that represents the main data must be available to generate an accurate decision tree classifier model [38].

L-SVM, as the second classifier studied in this paper, is a supervised machine learning algorithm that is utilized for classification and regression process. The idea behind the L-SVM procedure is mapping the input data x into high dimensional space of data by using a mapping function $\phi(.)$ and a linear function $f(x) = w\phi(x) + b$, where w, and b are optimized coefficients. The linear function is used to separate the data in the space and generate the hyperplane [39], [40]. The construction of L-SVM is to minimize the upper bound error by maximizing the margin between the separating hyper planes. By using the margin between the planes, separation of the data can be achieved. The advantage of L-SVM technique is the property of condensing information in the training set by using small number of data points.

The optimum hyperplane of the L-SVM classifier is located between the positive and negative points with maximum margin (see Fig. 11). The tips on the hyperplanes H_1 and H_2 are the support vectors, while H_0 represent the median in between H_1 and H_2. $d+$ and $d-$ represent the shortest distances to the closest positive and negative points, respectively. The margin of a separating hyperplane is given by $(d+)$ + $(d-)$.

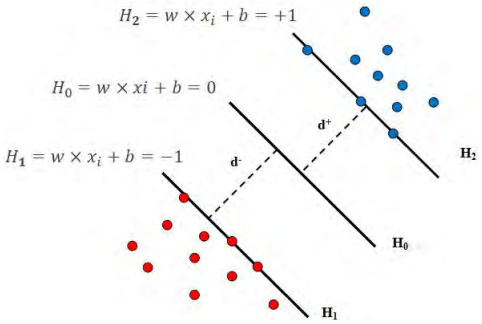

FIGURE 11. Separating hyperplane with maximum margin.

A linear discriminant function in Eq. 5 defines a decision boundary in the feature space.

$$\hat{y}(t) = \omega^T \varnothing(x) + b, \tag{5}$$

where x represents a patron and the feature vector function $\varnothing(x)$ represents the mapping of the patron. ω and b are parameters to be determined by a learning process of a training set $(x1, y1), \ldots (xi, yi), \ldots (xn, yn)$. In order to determine the hyperplane that maximizes the margin between classes, the solutions of the following optimization problem need to be achieved:

$$min\rho(\omega, b) = \frac{1}{2}\omega^2, \quad subject\ to\ \forall_i\ \ y_i[\omega^T \varnothing(x) + b]. \tag{6}$$

By means of the Lagrangian duality theory, the optimization problem in Eq. 6 can be simplified via solving the following problem:

$$\begin{aligned} max\ D(\alpha) &= \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} y_i \alpha_i y_j \alpha_j \varnothing(x_i)^T \varnothing(x_j) \\ subject\ to\ \forall_i\ \ \alpha_i &\geqslant 0 \\ \sum_i y_i \alpha_i &= 0, \end{aligned} \tag{7}$$

where α_i^* is the solution of the optimization problem that determines the optimal parameters of the optimal hyperplane, so that the direction parameter ω^* can be expressed as

$$\omega^* = \sum_i \alpha_i^* y_i \varnothing(x_i). \tag{8}$$

Then,the linear discriminant function can then be written as

$$\hat{y}(t) = \sum_{i=1}^{n} y_i \alpha_i^* \varnothing(x_i)^T \varnothing(x) + b^*. \tag{9}$$

By introducing a kernel $K_{i,j}$ and the duality, the discriminant function can be rewritten as

$$\hat{y}(t) = \sum_{i=1}^{n} y_i \alpha_i^* K_{x_i,x} + b^*. \tag{10}$$

Consequently, the bias parameter $b*$ can be obtained [46].

As the last classifier used in this work, the Linear Discriminant Analysis (LDA) is a dimensionality reduction technique. LDA is closely related to the Principal Component Analysis (PCA) technique, where both of them are used for reduction of feature dimensions. However, LDA technique is used not only for data dimension reduction but also as a classifier to model the difference between the classes of data. The aim of this technique is to project some higher dimensional feature vectors into lower dimensional vectors through a linear transformation procedure. The generated vectors are mostly the bests in separating the different classes [41], [42].

The considered classifiers are trained by using the training set and then, the trained classifier performance is analysed by using the testing set. This is usually not clear in most published works. Moreover, the effect of the smoothing of the features are also studied. When the classifiers are trained with unsmoothed features, the training accuracy were 98.93%, 97.3% and 92% for complex decision tree, L-SVM, and LDA types of classifiers, respectively.

After the training process, the considered classifiers are applied to the test set which represents 60% of the total data(150 records), corresponding to 90 transients for each class or device. The testing demonstrates different classifiers' performances for both smoothed and unsmoothed features as illustrated in Table 3. It can be seen from the table that the classification accuracy of both smoothed and unsmoothed features is more than 90%. As expected, the classifiers with smoothed features have higher accuracy than the unsmoothed ones. These results can be expected because filtering of the features concentrates the features in the feature space. Consequently, the distance between classes in the feature space becomes greater than the case of unfiltered features so that the interference between features is reduced dramatically. Because of this, the classification performance of smoothed or filtered features is higher than the case of unfiltered features. Finally, about five records out of 1800 records on average are misclassified in the testing stage when the considered classifiers are applied to the smoothed features. This corresponds to about 0.3% misclassification which is insignificant.

TABLE 3. Classifiers' performances at high SNR (30-35dB).

Classifier	Smoothed	Un-smoothed
L-SVM	99.8%	97.2%
Complex Tree	99.6%	97.59%
LDA	99.7%	91.3%

Performance of classifiers is generally evaluated by using confusion matrix. However, the receiver operating characteristics (ROC) curve and the area under the ROC curve (AUC) can additionally be used in evaluations. These performance metrics can visualize the performance of the classifiers [43], [44].

Because of the effectiveness of the L-SVM classifier, confusion matrix for the L-SVM at high SNR (30- 35 dB) is created as shown in Fig. 12. The confusion matrix presents the number of classified records, the number of misclassified records, and their classification rate (in percentage). According to the confusion matrix in Figure 12, class 10 has the lowest classification rate (92.2%), while the highest rate (100%) is achieved in Class 2, 14, 19 and 20. The average classification rate is 97.2%. On the other hand, referring to Table 3, only three records were misclassified when smoothed features are used (99.8%). This demonstrates that the use of the smoothed features increases the classification performance substantially.

VII. NOISE PERFORMANCE

In the previous section, we present the results at high SNR (30- 35 dB), that is, on the originally captured data. To investigate the noise peformance of the proposed method, noisy transients with different SNRs levels were created. In doing this, the SNR of the transients were reduced by adding the captured channel noise to the original transients at various levels. The following SNR levels are considered in creating noisy data sets: (18-23 dB), (12-15 dB), and (8-10 dB). Here the distribution of the number of transients at specific SNR level follows Gaussian distribution (that is, the most of the transients have around 9 dB SNR). The SNR of the noisy transients is calculated by using the average energy of the noisy transient signal (s) and the average energy of the noise signal (n) as follows

$$SNR = 10\ log_{10}((s/n) - 1). \tag{11}$$

The considered classifiers were applied to the overall feature, extracted from the noisy transients, and the subtle features, extracted from the TFEDs. Table 4 presents the performance of the classifiers for the three SNR levels.

Considering Table 3 and Table 4, it should be noted that the average of the classification accuracy decreases with lower SNR, as expected. For instance, performance of the L-SVM classifier decreases to 79.3% when the SNR level is reduced to (8-10 dB) level. However, the L-SVM is still the most robust classifier against noise. The confusion matrix of the L-SVM classifier at (8-10 dB) SNR is shown in Fig. 13.

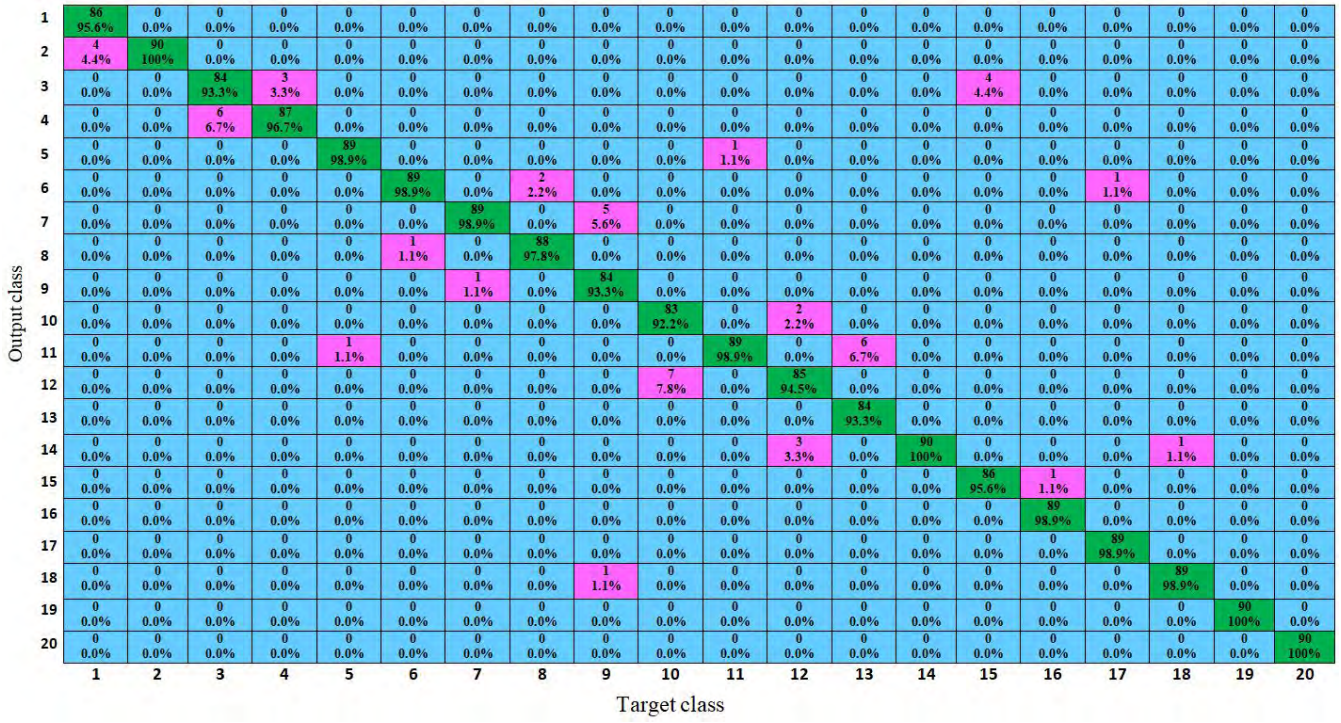

FIGURE 12. Confusion matrix for the L-SVM Classifier at high SNR.

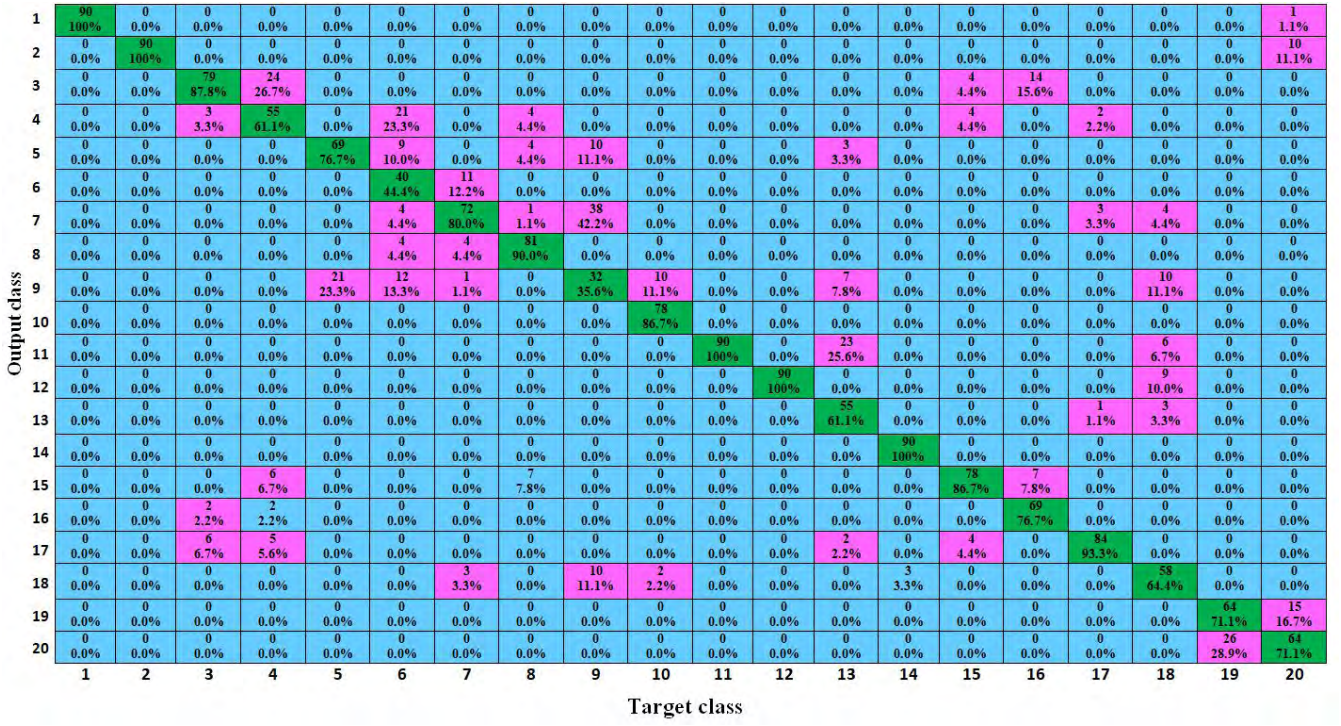

FIGURE 13. Confusion matrix for L-SVM, SNR level of 8-10 (smoothed features).

When the confusion matrices in Figure 12 and Figure 13 are compared along with the tables given, the accuracy decreases susbtantially. At lowest SNR level (8-10 dB) of Fig. 13, some of the devices seem to be very difficult to classify, such as Class 4,6 and 9, and even Class 9 is misclassified (35.6%) as Class 7 (42.2%). However, the overall

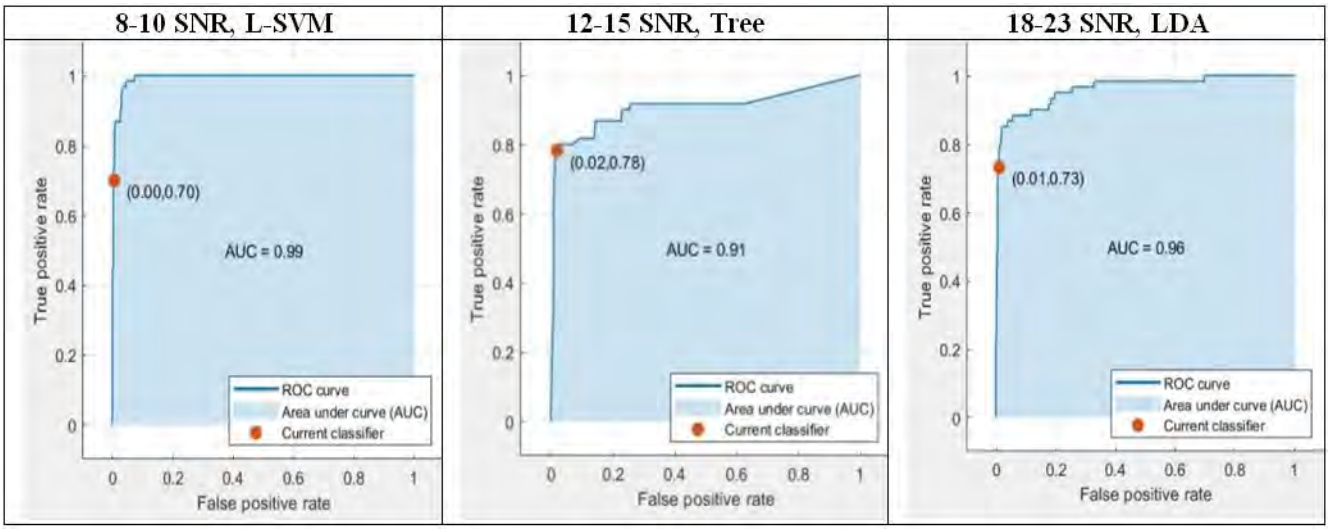

FIGURE 14. ROC generated for the classifiers at different SNR levels (unsmoothed features).

TABLE 4. Classifiers, test preformances with different levels of SNR (smoothed features).

Classifier	SNR		
	(8-10)	(12-15)	(18-23)
L-SVM	79.3%	82.1%	90.53%
Complex Tree	66.8%	68.8%	85.4%
LDA	76.6%	77.8%	83.6%

TABLE 5. Areas under ROC curves of the classifiers at different SNR levels (unsmoothed features).

Classifier	SNR			
	(8-10)	(12-15)	(18-23)	High SNR
L-SVM	0.9895	0.991	0.9955	1.00
Complex Tree	0.8975	0.9105	0.9425	0.999
LDA	0.9480	0.9485	0.9530	0.9995

performance is still high compared with the case when unsmoothed features are used in the same classifier. Here, it should be noted that the classification involves 20 devices all together. One-to one (binary) classification may still work well for all classes except Class 9.

The ROC is another metric to measure the performance of the classifiers. The ROC of the classifiers implemented in this work is shown in Fig. 14 at different SNR levels. Moreover, the AUCs of the relevant classfiers are also shown in Table 5 along with Fig. 14. It should be noted that the performance of the L-SVM in discriminating the classes(devices) is clearly visible in these metrics. The L-SVM classifier yields a high classification performance, by achieving very high AUC (0.9895) at low SNR levels. Here, only unsmoothed features are used as the objective is to compare the perfromance of the classifiers without smoothing of the features. It is expected that the performance would increase further when the smoothing is applied to the feature set.

VIII. CONCLUSIONS

Bluetooth (BT) signals are collected from different cell phone brands and models. Two devices of ten different cell phones, each with one hundred and fifty records, are considered in this study. The collected BT transmission signals are de-noised by filtering. The transient signals of the de-noised transmission signals are detected by using signal energy envelope. By using the energy envelope, the assigned thresholds in the detection process are adaptive to all 3000 considered records, which indicates that the technique is efficient.

Based on the EMD method, the HHT is applied to the detected transient signals to obtain their TFED. Thirteen features are extracted from three feature groups namely, the transient signal and its energy envelope, summation of TFED of the transient signal along time axis, and summation of TFED of the transient signal along frequency axis. Among thirteen features, the std of transient instantaneous phase, the entropy of the transient instantaneous phase, and the duration of the transient are considered as the most robust features. These features are dominant in constructing the signal RF fingerprint for each record.

Before the classification, the robustness of the extracted features are investigated by using the box plot demonstration. In order to identify the twenty classes (devices), three most common classifiers are applied to the extracted features. The data set is split into training and test data in order to see realistic performance of the classifiers. The test data are utilized to train the classes via the generated trained classifier models. The classifiers' performances are evaluated with different levels of SNR. The classification results demonstrate that the method is quite effective in BT device discrimination, and it achieved quite high performance at high SNR levels which implies that the feature set are quite unique for use in RF fingerprinting. On the other hand, as expected, the performance decreases at low SNR levels when the smoothing

is not applied to the features. Some specific devices seem to be very difficult to classify at lower SNR levels, such as Class 4, 6 and 9 (some IPhone models). This suggests that some new decomposition techniques along with robust feature sets should be studied extensively in discriminating of devices. Overall, this study, with very big but diverse data set, proves that the L-SVM achieves the highest accuracy for different SNR levels. Based on these results, BT based RF fingerprinting might be considered, to some extent, in security of wireless networks, or might help researcher investigate the proposed features in RF fingerprinting.

The future work will concentrate on various techniques toward decomposition of the signal and use only robust features. There will also be research toward how the same set of features can be effectively used when a downconversion and lower sampling rate is used. It is also possible to utilize some deep learning apparoaches when there is a big data set [34].

REFERENCES

[1] B. W. Ramsey, T. D. Stubbs, B. E. Mullins, M. A. Temple, and M. A. Buckner, "Wireless infrastructure protection using low-cost radio frequency fingerprinting receivers," *Int. J. Crit. InfraStruct. Protection*, vol. 8, pp. 27–39, Jan. 2015.

[2] W. C. Suski, M. A. Temple, M. J. Mendenhall, and R. F. Mills, "Using spectral fingerprints to improve wireless network security," in *Proc. Global Telecommun. Conf.* New Orleans, LO, USA, Nov. 2008, pp. 1–5.

[3] S. U. Rehman, K. W. Sowerby, and C. Coghill, "Radio-frequency fingerprinting for mitigating primary user emulation attack in low-end cognitive radios," *IET Commun.*, vol. 8, no. 8, pp. 1274–1284, May 2014.

[4] J. L. Padilla, P. Padilla, J. F. Valenzuela-Valdés, and J. Ramírez, "RF fingerprint measurements for the identification of devices in wireless communication networks based on feature reduction and subspace transformation," *Measurement*, vol. 58, pp. 468–475, Dec. 2014.

[5] O. Ureten and N. Serinken, "Wireless security through RF fingerprinting," *Can. J. Elect. Comput. Eng.*, vol. 32, no. 1, pp. 27–33, 2007.

[6] Y. Yuan, Z. Huang, H. Wu, and X. Wang, "Specific emitter identification based on Hilbert-Huang transform-based time-frequency-energy distribution features," *IET Commun.*, vol. 8, no. 13, pp. 2404–2412, Sep. 2014.

[7] Y. J. Yuan, Z. Huang, and Z.-C. Sha, "Specific emitter identification based on transient energy trajectory," *Prog. Electromagn. Res. C*, vol. 44, pp. 67–82, Dec. 2013.

[8] J. Zhang, Y. Li, and J. Yin, "Modulation classification method for frequency modulation signals based on the time–frequency distribution and CNN," *IET Radar, Sonar Nav.*, vol. 12, no. 2, pp. 244–249, Oct. 2017.

[9] I. Paraskevas, K. Prekas, S. M. Potirakis, and M. Rangoussi, "On the use of time-frequency distributions for the power quality problem of harmonics," in *Proc. 7th Medit. Conf. Exhib. Power Gener. Transmiss, Distrib. Energy Convers.*, Agia Napa, Cyprus, Nov. 2010, pp. 1–5.

[10] I. Shafi, J. Ahmad, S. I. Shah, and F. M. Kashif, "Evolutionary time-frequency distributions using Bayesian regularised neural network model," *IET Signal Process.*, vol. 1, no. 2, pp. 97–106, Jun. 2017.

[11] M. J. Hinich, "Detecting a transient signal by bispectral analysis," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. 38, no. 7, pp. 1277–1283, Jul. 1990.

[12] O. H. Tekbas, N. Serinken, and O. Ureten, "An experimental performance evaluation of a novel radio-transmitter identification system under diverse environmental conditions," *Can. J. Electr. Comput. Eng.*, vol. 29, no. 3, pp. 203–209, Jul. 2004.

[13] E. Uzundurukan, A. M. Ali, and A. Kara, "Design of low-cost modular RF front end for RF fingerprinting of Bluetooth signals," in *Proc. 25th Signal Process. Commun. Appl. Conf. (SIU),* Antalya, Turkey, May 2017, pp. 1–4.

[14] O. Ureten and N. Serinken, "Detection of radio transmitter turn-on transients," *Electron. Lett.*, vol. 35, no. 23, pp. 1996–1997, Nov. 1999.

[15] M. Barbeau, J. Hall, and E. Kranakis, "Detection of rogue devices in Bluetooth networks using radio frequency fingerprinting," in *Proc. 3rd IASTED Int. Conf. Commun. Comput. Netw.*, Oct. 2006, pp. 4–6.

[16] M. Frisch and M. Messer, "Transient signal detection using prior information in the likelihood ratio test," *IEEE Trans. Signal Process.*, vol. 41, no. 6, pp. 2177–2192, Jun. 1993.

[17] L. Daudet, "A review on techniques for the extraction of transients in musical signals," in *Proc. Int. Symp. Comput. Music Model. Retr.* Berlin, Germany: Springer, Sep. 2005, pp. 219–232.

[18] Y. L. Gong, S. X. Wang, H. Z. Wang, and Y. S. Dai, "Detecting a transient signal by integrated polyspectrum," *Electron Lett.*, vol. 33, no. 12, pp. 1011–1012, Jun. 1997.

[19] D. O. Walsh and P. A. Delaney, "Detection of transient signals in multipath environments," *IEEE J. Ocean. Eng.*, vol. 20, no. 2, pp. 131–138, Apr. 1995.

[20] S. Colonnese and G. Scarano, "Transient signal detection using higher order moments," *IEEE Trans. Signal Process.*, vol. 47, no. 2, pp. 515–520, Feb. 1999.

[21] N. K. Mishra and G. V. Anand, "Transient signal detection using wigner-ville distribution and wavelet denoising," in *Proc. IEEE Region 10th Conf.*, Hong Kong, China, Nov. 2006, pp. 1–4.

[22] A. M. Ali, E. Uzundurukan, and A. Kara, "Improvements on transient signal detection for RF fingerprinting," in *Proc. IEEE Signal Process. Commun. Appl. Conf. (SIU),*, Antalya, Turkey, May 2017, pp. 1–4.

[23] S. U. Rehman, K. W. Sowerby, and C. Coghill, "RF fingerprint extraction from the energy envelope of an instantaneous transient signal," in *Proc. Aust. Commun. Theory Workshop*, Feb. 2012, pp. 90–95.

[24] J. Hall, M. Barbeau, and E. Kranakis, "Enhancing intrusion detection in wireless networks using radio frequency fingerprinting," in *Proc. Commun. Internet Inf. Technol.*, Saint Thomas, Virgin Islands, Nov. 2004, pp. 201–206.

[25] C.-P. Lai, R. M. Narayanan, and Q. Ruan, "Hilbert-Huang transform (HHT) analysis of human activities using through-wall noise radar," in *Proc. Int. Symp. Signals, Syst. Electron.*, Jul. /Aug. 2007, pp. 115–118,

[26] A. Gururani, S. R. Mohanty, and J. C. Mohanta, "Microgrid protection using Hilbert–Huang transform based-differential scheme," *IET Gener. Transmiss. Distrib.*, vol. 10, no. 15, pp. 3707–3716, Nov. 2016.

[27] R. A. Thuraisingham, "Estimation of Teager energy using the Hilbert–Huang transform," *IET Signal Process.*, vol. 9, no. 1, pp. 82–87, Feb. 2015.

[28] D. Naccache, "Instantaneous frequency analysis," in *Proc. Int. Conf. Opt. Commun. Syst.* Reykjavik, Iceland, Jul. 2013, p. 11.

[29] J. Li, J. Wang, X. Zhang, and W. Tang, "Empirical mode decomposition based on instantaneous frequency boundary," *Electron. Lett.*, vol. 53, no. 12, pp. 781–783, May 2017.

[30] D. Kim and H.-S. Oh, "EMD: A package for empirical mode decomposition and Hilbert spectrum," *R J.*, vol. 1, no. 1, pp. 40–46, May 2009. [Online]. Available: http://cran.r-project.org/web/packages/EMD/index.html

[31] M. K. I. Molla and K. Hirose, "Hilbert spectrum in time-frequency representation of audio signals considering disjoint orthogonality," *Adv. Adapt. Data Anal.*, vol. 2, no. 3, pp. 313–336, Jul. 2010.

[32] M. I. A. Khan, M. E. Hamid, and T. Nakai, "Systolic Phase Detection from Pulsed Doppler Ultrasound Signal using EMD-DHT based approach," *Int. J. Signal Process. Image Process. Pattern Recognit.*, vol. 7, no. 5, pp. 207–216, 2014.

[33] P. Scanlon, I. O. Kennedy, and Y. Liu, "Feature extraction approaches to RF fingerprinting for device identification in femtocells," *Bell Labs Tech. J.*, vol. 15, no. 3, pp. 141–151, Dec. 2010.

[34] K. Merchant, S. Revay, G. Stantchev, and B. Nousain, "Deep learning for RF device fingerprinting in cognitive communication networks," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 160–167, Feb. 2018.

[35] I. O. Kennedy, P. Scanlon, F. J. Mullany, M. M. Buddhikot, K. E. Nolan, and T. W. Rondeau, "Radio transmitter fingerprinting: A steady state frequency domain approach," in *Proc. IEEE 68th Veh. Technol. Conf.*, Sep. 2008, pp. 1–5.

[36] I. O. Kennedy, P. Scanlon, and M. Buddhikot, "Passive steady state RF fingerprinting: A cognitive technique for scalable deployment of Co-channel femto Cell underlays," in *Proc. 3rd IEEE Symp. New Frontiers Dyn. Spectr. Access Netw.*, Oct. 2008, pp. 1–12.

[37] J. L. Hintze, *Box Plots Number Cruncher Statistical Systems*. Kaysville, UT, USA: NCSS User's Guide I, 2007, ch. 152.

[38] M. A. Friedl and C. E. Brodley, "Decision tree classification of land cover from remotely sensed data," *Remote Sens. Environ.*, vol. 61., no. 3, pp. 399–409, Sep. 1997.

[39] S. Wang, B. Meng, and H. Tian, "A modeling method based on wavelet support vector machine," in *Proc. Chin. Control Decision Conf.*, Xuzhou, China, May 2010, pp. 3113–3116.

[40] S. Amari and S. Wu, "Improving support vector machine classifiers by modifying kernel functions," *Neural Netw.*, vol. 12, no. 6, pp. 783–789, Jul. 1999.
[41] L. H. Chiang, M. E. Kotanchek, and A. K. Kordon, "Fault diagnosis based on Fisher discriminant analysis and support vector machines," *Comput. Chem. Eng.*, vol. 28, no. 8, pp. 1389–1401, Jul. 2004.
[42] D. AbuZeina and F. S. Al-Anzi, "Employing fisher discriminant analysis for Arabic text classification," *Comput. Electr. Eng.*, vol. 66, pp. 474–486, Feb. 2018.
[43] C. Cortes and M. Mohri, "AUC optimization vs. error rate minimization," in *Proc. 16th Int. Conf. Neural Inf. Process. Syst.*, Dec. 2003, pp. 313–320.
[44] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, 2006.
[45] S. Yu, J. Kim, and S. Yoon, "R Ensemble learning can significantly improve human microRNA target prediction," *Methods*, vol. 69., no. 3, pp. 220–229, 2014.
[46] L. Bottou, and C. J. Lin, "Support vector machine solvers," in *Large Scale Kernel Machines*, L. Bottou, O. Chapelle, D. DeCoste, and J. Weston, Eds. Cambridge, MA, USA: MIT Press, 2007.

AYSHA M. ALI was born in Al-bayda, Libya, in 1980. She received the B.Sc. degree in telecommunications engineering from Omar Al-Mukhtar University, Al-bayda, in 2002, and the M.Sc. degree in telecommunications engineering from Cankaya University, Turkey, in 2006. In 2007, she joined the Electrical and Electronics Engineering Department, Omar Al-Mukhtar University. She is currently pursuing the Ph.D. degree with Atilim University. Her current research interest includes radio emitter identification.

EMRE UZUNDURUKAN was born in Ankara, Turkey, in 1988. He received the electrical electronics engineering degree and master's degree in electrical electronics engineering from Atilim University, Ankara, in 2016 and 2018, respectively, where he is currently pursuing the Ph.D. degree. He is currently a Research Assistant with the School of Civil Aviation, Atilim University. His Ph.D. study is focused on device classification using RF fingerprinting. His research interests include RF fingerprinting applications, electronic warfare systems, and signal processing.

ALI KARA (SM'93–M'03–SM'12) was born in Amasya, Turkey, in 1972. He received electronics engineering degree from Erciyes University, Kayseri, in 1992, the M.Sc. degree from Cukurova University, Adana, in 1996, and the Ph.D. degree from Hacettepe University, Ankara, in 2002. He was with Polytechnic University (ECE), Brooklyn, NY, USA, from 1999 to 2000, where he conducted theoretical and experimental research as a Research Assistant in a project sponsored by Symbol Technologies (Motorola). He joined the Department of Electrical and Electronics Engineering, Atilim University, in 2000, and devoted himself to founding the Department. He has published in refereed journals/conferences, and has led several projects in the areas of Radio Propagation, Virtual and Remote Laboratories (VRL) and Radar&Electronic Warfare Systems. In one of the projects, he led a group with Atilim University to develop a VRL platform on RF and Communications (unique platform funded by European Commission). On the other hand, he was with TUBITAK-BILGEM as a Chief Researcher and a Consultant on defense and security related projects, from 2006 to 2012. He is currently with Atilim University, where he has been a Professor, since 2015, and also the Director of the Graduate School of Natural and Applied Sciences. He is actively researching in the areas of RF fingerprinting, locating and identification of radio emitters including radars and wireless devices, and radio aspects of wireless communications including channel modeling and antennas.