



Research on RF Fingerprinting Extraction of Power Amplifier Based on Multi-domain RF-DNA Fingerprint

Yihan Xiao^(✉) and Xinyu Li

College of Information and Communication Engineering,
Harbin Engineering University, Harbin, Heilongjiang 150000, China
{xiaoyihan, lixinyu_2019}@hrbeu.edu.cn

Abstract. The uniqueness of the RF signal is caused by the difference in the hardware structure of the transmitter and the differences between the different devices. Among them, RF power amplifier is one of the key components of RF fingerprinting of wireless transmitter. It is an important breakthrough for RF fingerprint generation mechanism and individual identification. This paper proposes a new identification method of power amplifier based on new intelligent feature set, firstly, processing the received signal. The time domain, frequency domain, time-frequency domain, fractal domain transformation and feature extraction are performed. Secondly, the new intelligent feature set of each power amplifier individual can be characterized, and the RF-DNA fingerprint is visualized. Finally, the support vector machine is used to realize the individual recognition by selecting the optimal RBF kernel function. By simulating and verifying the eight power amplifier signals, a new intelligent feature set can be used to uniquely characterize the power amplifier. Under low SNR, the power amplifier individual can be quickly and effectively identified. The recognition rate of more than 80% can be achieved above the -5 dB signal-to-noise ratio.

Keywords: Individual identification · RF-DNA · SVM

1 Introduction

When a wireless device transmits a signal, the subtle features of each device are attached to the radio frequency signal due to hardware differences, and the individual identification of the device can be realized by using these subtle differences, and the power amplifier is a key component of the wireless device transmitter, and has a significant non-Linear characteristics are an important source of RF fingerprint generation and the key to studying the mechanism of RF fingerprint generation. The concept of “radio frequency fingerprint” was first put forward in 2003 [1]. In 2016, Yu Jiabao et al. conducted a comprehensive review of RF fingerprinting technology, pointing out that RF fingerprinting is a method for extracting RF fingerprints of devices by analyzing the communication signals of wireless devices for device identification. It should be universal, unique and short. Time invariance, independence and robustness [2]. In fact, as early as 1995, Choe [3] and Toonstra [4] and others have begun to use

communication signals for individual identification of devices. RF fingerprint technology can be applied to wireless network security authentication, indoor wireless positioning, fault diagnosis, communication countermeasures and radar countermeasures. This technology has been listed by the US military as one of the key technologies for current development, and has high application value in civil and military applications.

In fact, the early technology of radio frequency fingerprint was based on transient signals. Radio frequency fingerprint technology based on steady-state signal was first proposed in 2008 [5]. Because of its good classification performance and higher feasibility of the method, more and more attention has been paid to the technology of RF fingerprint extraction and recognition based on steady-state signals. Radio frequency different native attributes (RF-DNA) is a method to extract features from radio frequency (RF) signals. The signals emitted by each device have unique characteristics and can be used to distinguish the device from other similar devices. This is called fingerprint recognition. Randall W. Klein et al. proposed an RF-DNA fingerprint recognition algorithm based on time domain (TD) features [6]. McKay D et al. proposed RF-DNA fingerprint feature extraction technology based on spectral domain (SD) features [7]. Donald R et al. proposed that the normalized amplitude of the preamble of IEEE802.16e WiMAX signal be transformed into discrete Gabor transform, and the RF-DNA fingerprint signal is generated by Gabor coefficient [8, 9]. Kang used bispectrum theory to extract SURF features [10]. Gok uses variational mode decomposition (VMD) to compute fingerprint features [11]. D'Agostino S uses time domain amplitude characteristics [12]. Ru X uses frequency domain features to accomplish individual recognition of radiation sources [13, 14].

The fingerprint recognition technology based on RF-DNA is one of the hotspots in the field of RF fingerprint identification in recent years. It constructs radio frequency fingerprint of transmitter by extracting statistical features from radio frequency signal, so as to realize the identity authentication of equipment [15]. Based on this research, this paper proposes a method for identifying the power amplifiers by constructing RF-DNA gene profiles: By intercepting the information of the power amplifiers, this paper transforms them into the time domain, the frequency domain, the time-frequency domain and the fractal domain. Genes that can characterize power amplifiers are extracted from various domains to construct RF-DNA gene profiles, which can uniquely characterize power amplifiers and realize the individual identification of power amplifiers (Fig. 1).

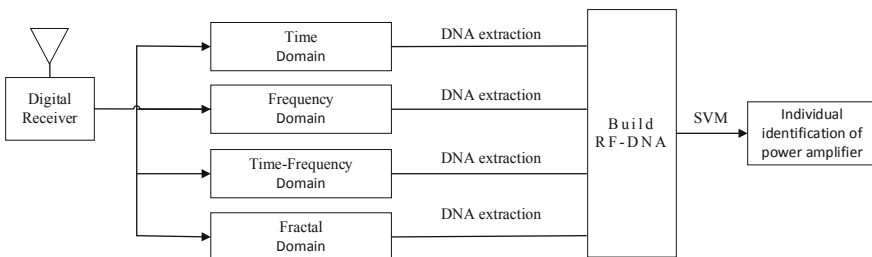


Fig. 1. RF-DNA construction and identification technology block diagram

2 Construction of Power Amplifier RF-DNA

The gene extracted by transforming the intercepted signal in the time domain, the frequency domain, the time-frequency domain, and the fractal domain can reflect the uniqueness of the Power Amplifier, which can be defined as RF-DNA.

2.1 Time Domain RF-DNA RF Fingerprint Extraction Process

For time domain RF-DNA radio frequency fingerprints (F), the basic signal characteristic is instantaneous amplitude. I/Q quadrature sampling is a commonly used method of signal acquisition equipment. The complex time domain signal can be expressed as follows:

$$s_{TD}(n) = I_{TD}(n) + jQ_{TD}(n) \quad (1)$$

Where, $I_{TD}(n)$ and $Q_{TD}(n)$ are the Instantaneous in-phase component and quadrature phase component of $S_{TD}(n)$.

According to the Hilbert transform, the instantaneous amplitude $a(n)$ of the signal can be conveniently calculated according to the Eq. (2).

$$a(n) = \sqrt{I_{TD}^2(n) + Q_{TD}^2(n)} \quad (2)$$

In order to eliminate the influence of the deviation of the acquisition equipment on the subsequent processing, it is necessary to centralize the basic feature. The instantaneous amplitude is centered according to the Eq. (3).

$$a_c(n) = a(n) - \mu_a \quad (3)$$

Where $n = 1, 2, \dots, N_M$, N_M is the total number of samples of the sampled signal, and μ_a is the mean of the instantaneous amplitude of the signal samples.

For the above basic signal characteristic, if directly used as the classification recognition process of the signal, the signal may be limited in data processing due to the high complexity. Applying the inherent statistical characteristics of the basic signal feature as classification features can reduce the feature space dimension used for device classification and reduce the computational burden. Common statistical features commonly used in feature extraction include standard deviation (σ), variance (σ^2), skewness (γ), and kurtosis (κ).

2.2 Frequency Domain RF-DNA RF Fingerprint Extraction Process

In the field of signal processing, most signal processing methods are only for the amplitude information of the signal, and the phase information is also very important for characterizing a signal. Higher-order spectra and higher-order cumulants can describe both the amplitude information of the signal and the phase information of the signal. Therefore, the process of introducing high-order spectra and high-order

cumulants is the simplest, the most in-depth, and the most widely used bispectrum. Some research has been done on the role of bispectrum in the recognition of power amplifier based on RF-DNA.

In this paper, the nonparametric indirect bispectrum estimation results of the power amplifier output signal are analyzed, and the bispectral distribution entropy, bispectral singular spectral entropy, bispectral energy entropy and waveform entropy are extracted, and the four characteristic entropies are combined into four-dimensional features. The vector acts as a fingerprint feature that distinguishes the signals of different power amplifiers. The research will be carried out in the following aspects.

- (a) Receive signals from different power amplifiers;
- (b) Calculate the bispectrum estimate of the signal using the indirect estimation method in the nonparametric method;
- (c) Divide a certain quadrant of the bispectrum into a frequency region of $8 * 8$, and calculate a distribution entropy of the bispectrum estimation. The formula is as follows

$$\|A\| = \sum_{i=1}^N |a_i|, \quad P_i = \frac{|a_i|}{\|A\|} \tag{4}$$

Where $a_i (i = 1, 2, \dots, N)$ represents bispectrum mean of N frequency planes. The bispectral distribution entropy can be defined as

$$E_a = - \sum_{i=1}^N p_i \cdot \lg p_i \tag{5}$$

- (d) Performs the singular value decomposition on the bispectrum estimation result, and gets a series of singular values $\lambda_i (i = 1, 2, \dots, L)$. Then calculates the bispectral singular spectral entropy according to formula (6,7).

$$p_i = \lambda_i / \sum_{i=1}^L \lambda_i \tag{6}$$

$$E_{svd} = - \sum_{i=1}^L p_i \cdot \lg p_i \tag{7}$$

- (e) Energy entropy and waveform entropy are calculated by the surrounding-line integral bispectrum.

$$p_i = b_i / \sum_{i=1}^M |b_i| \tag{8}$$

$$E_b = - \sum_{i=1}^L p_i \cdot \lg p_i \tag{9}$$

$$N_{i,j} = \sum_{i,j} |n(i,j)| \tag{10}$$

$$p_{ij} = |n(i,j)|/N_{i,j} \tag{11}$$

$$E_n = - \sum_{i,j} p_{ij} \cdot \lg p_{ij} \tag{12}$$

In the above formula, E_b and E_n represent waveform entropy and energy entropy. In formula (8), M is the number of lines in the perimeter integral and b_i represents the integral of each circle. In formula (10), $n(i, j)$ represents the result of the third-order cumulant bispectrum estimation.

(f) Form the four-dimensional eigenvector $F = [E_a, E_{svd}, E_b, E_n]$.

2.3 Time-Frequency Domain RF-DNA RF Fingerprint Extraction Process

In a majority of previous related work, RF-DNA fingerprints were predominantly extracted from TD and SD responses, with Wavelet Transform (WT) coefficients being AFIT’s first application of joint 2D features. The use of WT coefficients is consistent with conclusions in [9] indicating that the use of momentary and/or time localized energy as a function of frequency can be effective for describing signals. This motivated the use of Wave Translate (WT) which is calculated as follows:

- Step 1. Perform a n-layer wavelet decomposition on the signal to obtain a series of wavelet coefficients.
- Step 2. The obtained wavelet coefficients include n detail coefficients (corresponding to high frequency components of wavelets) and an approximation coefficient (corresponding to low frequency components of wavelets).
- Step 3. According to the wavelet coefficients, reconstructing the signal at a specified level by using high frequency or low frequency components of the original signal.
- Step 4. Calculate the energy of each reconstructed signal to form a feature vector $F=[e_1, e_2, \dots, e_n]$.

2.4 Fractal Domain RF-DNA RF Fingerprint Extraction Process

Fractal is a general term for a self-similar graphic structure that has no meaning length. This graphic structure’s essence cannot be described by Euclidean measures. Therefore, using dimensions to portray such graphics has been employed widely.

Fractal dimensions can quantitatively describe the complexity of a collection of parts. The definition of fractal dimension is different from the definition of European geometric dimension. It is application-dependent, and its fractal dimension is defined

differently depending on the application. In the field of modulation mode identification, the box dimension and information dimension are commonly used. Among them, the box dimension reflects the collection scale of the classification set, and the information dimension reflects the information of the distribution of the classification set. In this paper, we choose box dimension and information dimension as the basic features in fractal domain.

Box Dimension. The steps of feature extraction based on box dimension are as follows:

- (a) The signal sequence after the pre-processing can be expressed as

$$\{g(i), i = 1, 2, \dots, N\} \quad (13)$$

Where N is the length of the signal sequence

- (b) According to the definition of the box dimension, the signal sequence $\{g(i)\}$ is placed in the unit square, and the minimum interval of the abscissa is $q = 1/N$. $N(q)$ can be expressed as

$$N(q) = N + \frac{\left\{ \sum_{i=1}^{N-1} \max\{g(i), g(i+1)\}q - \sum_{i=1}^{N-1} \min\{g(i), g(i+1)\}q \right\}}{q^2} \quad (14)$$

- (c) The box dimension can be calculated by the following formula

$$D_b = -\frac{\ln N(q)}{\ln q} \quad (15)$$

Information Dimension. The following are the steps for feature extraction using information dimensions.

- (a) Same as box dimension, the preprocessed signal can be expressed as

$$\{f(i), i = 1, 2, \dots, N\} \quad (16)$$

- (b) Signal reconstruction. In the frequency domain, the signal is reconstructed as follows.

$$s(i) = fs(i+1) - fs(i), i = 1, 2 \dots, N-1 \quad (17)$$

The purpose of this is to reduce the impact of some of the in-band noise, and to facilitate the calculation of information dimensions using the following method.

(c) Calculate information dimension using reconstructed signals.

$$L = \sum_{i=1}^{N-1} s_i, P_i = \frac{s_i}{L} \tag{18}$$

$$D_I = \sum_{i=1}^{N-1} P_i \lg(1/P_i) \tag{19}$$

Fractal features are expressed in vector form as $V_F = [D_b, D_I]$

2.5 Conclusion

This chapter introduces the characteristics of the multi-domain which will be used in this paper to identify the amplifiers. The extraction methods of each feature are also be introduced. Table 1 shows the summary of the used features.

Table 1. Feature extraction in different transform domains.

Domain	Feature
Time domain	Standard deviation (σ) Variance (σ^2) Skewness (γ) Kurtosis (κ)
Frequency domain	Waveform entropy Energy entropy Distribution entropy Singular value entropy
Time-frequency domain	The energy of the Reconstructed signal used the 1–4th order high frequency coefficient of wavelet.
Fractal domain	Box dimension Information dimension

3 Power Amplifier Individual Recognition Based on SVM

3.1 Individual Recognition Based on Support Vector Machine

The basic learning principle of Support Vector Machine (SVM) is the principle of SRM (Structure Risk Minimization). The main advantages are fast learning speed, global optimality, fast convergence speed and small mean square error. The support vector machine method is based on the concept of decision plane, which is solved by transforming the classification problem into the quadratic programming problem of the data. Decision planes can distinguish a set of elements belonging to different categories, Assume the following samples exist.

$$(x_1, y_1), \dots, (x_l, y_l), x \in R, y \in \{+1, -1\} \quad (20)$$

l is the number of samples. After the training samples are given, a decision plane is constructed by SVM, also called the optimal classification hyperplane, so that the distance between the different types of training samples to the optimal classification hyperplane is maximized, and then achieves more accurate classification in the training of test samples.

By introducing the concept of kernel function, support vector machine (SVM) can deal with the problems of non-linearity and high-dimensional space. It plays an important role in the theory of SVM. The kernel function is a technique generated as a spatial map. The kernel function is mapped to a higher dimensional space, and the complex inner product operation between two points in the high-dimensional feature space is replaced by a kernel function of the original input space. Thereby avoiding the increase of the SVM algorithm in the case of nonlinear separability. The kernel function selected in this paper is RBF kernel function.

$$K(x, y) = \exp(-\gamma \|x - y\|^2) \quad (21)$$

In formula, $\gamma = 1/2\sigma^2$, γ is the distance between the optimal classification hyperplane and the nearest point. By changing the value of width σ , a larger γ can be selected to reduce the error.

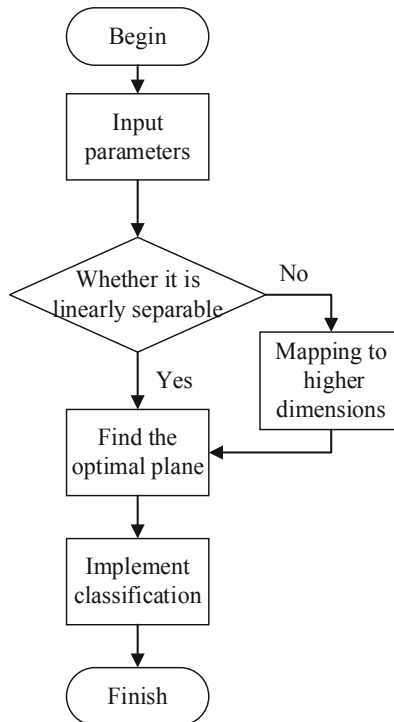


Fig. 2. Flow chart of support vector machine recognition

Classification process: For linear separable categories, support vector machine directly achieves classification by seeking the optimal classification hyperplane; for linear inseparable samples in low-dimensional space, an important idea of support vector machine is to map them to higher-dimensional data space through kernels, so that samples can be classified linearly in high-dimensional space. The flowchart of the support vector machine implementation classification is shown in Fig. 2.

4 Simulation and Verification of Individual Recognition of Power Amplifier

4.1 Experimental Configuration

Eight 433 MHz RF power amplifiers of the same model and the same batch were selected as the research object. The experimental configuration is shown in Fig. 3. The vector signal generator is used to repeatedly transmit an edited 16QAM modulated signal. The carrier frequency of the signal is set to 433 MHz. The sampling rate is 1 MHz and the output power is 0 dB. The 16QAM signal is connected to the power amplifier through a coaxial line. The power amplifier is powered by a 5 V DC voltage. After the amplified signal passes through the 30 dB attenuator, it is directly connected to the RF recorder by a coaxial line. The RF recorder uses quadrature sampling technology with a center frequency set to 433 MHz and a sampling rate set to 4 MHz.

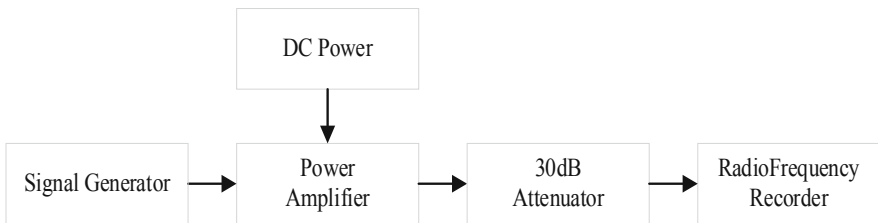


Fig. 3. Experimental configuration

4.2 Data Preprocessing

Since the signal source continuously transmits a QAM signal repeatedly, after using a radio frequency recorder records a piece of data for about 5 s, it is necessary to perform segmentation interception processing on the signal, and intercept each segment of the QAM signal. Specific steps are as follows:

- (a) Instantaneous amplitude calculation. Calculate the instantaneous amplitude characteristic $a(n)$ of the recorded signal.
- (b) Detection of mutation points. The abrupt point detection is performed on $a(n)$, and the index value of the first 40 mutation points of the recorded signal file, that is, the index value of the start time and the end time of each QAM signal is obtained. (According to a complete QAM signal playback time of 90 ms, there are about 55 repeated cycles of QAM signal in 5 s record data.)

- (c) QAM signal interception. For each of the power amplifier's recorded signals, 20 waveforms were intercepted as experimental data, and 8 amplifiers totaled 160 waveforms.

4.3 Analysis of Results

This paper used features of the multi-domain which are introduced in Sect. 2 to construct RF-DNA of the power amplifier. After extracting the multi-domain RF-DNA fingerprint of the signal, a set of high-dimensional eigenvectors is obtained. The main feature extracted from RF-DNA is used as the input of support vector machine, which is divided into test set and training set according to a certain proportion (4:1) for individual classification and recognition. The results are analyzed as follows.

Before the feature extraction, the normalization operation is first performed. The purpose of this step is to avoid the data dominate that SVM model parameters are dominated by data with a larger or smaller distribution range. The comparison chart before and after normalization is as follows:

Figures 4 and 5 represent the signal waveforms before and after normalization. This operation can effectively prevent the subsequent processing from being affected by the wild value.

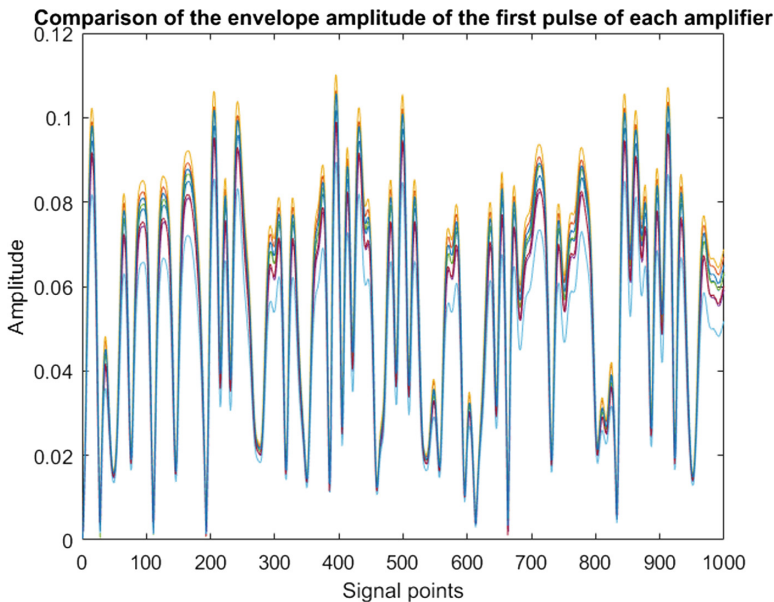


Fig. 4. Comparison of RF signals of eight power amplifiers before normalization

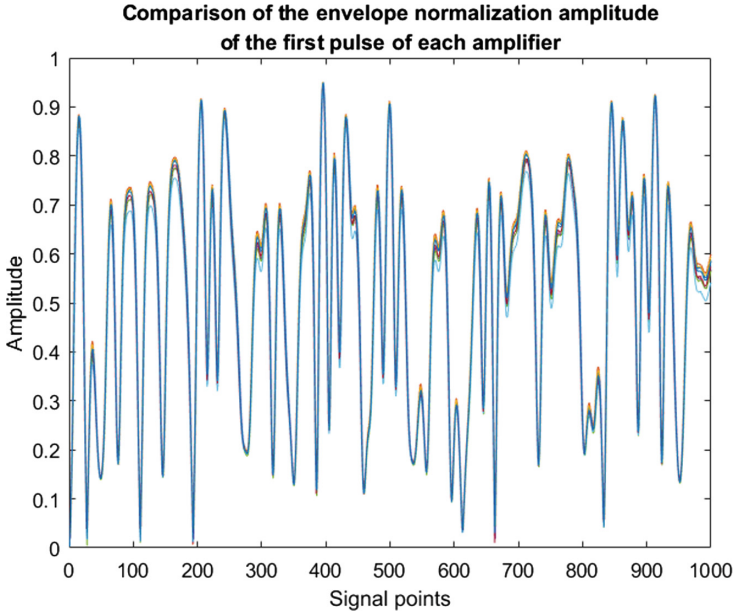


Fig. 5. Normalized comparison of RF signals of eight power amplifiers

Feature Visualization. After feature extraction in different domains, this paper carried out feature fusion to construct multi-domain RF-DNA.

Figure 6 is a graphical representation of the statistical average of the multi-domain fusion features of the eight power amplifiers. The color represents the statistical average of the features extracted from the 20 samples. From this picture, we can see that the characteristics extracted between different power amplifiers are quite different.

This figure can be analyzed from the following aspects:

- (a) From the perspective of a single domain, amplitude features of time domain are more effective than wavelet features of time-frequency domain. This is reflected in the higher recognition rate of time domain features. The main reason for this result is that the 16QAM signal is an amplitude modulation signal, and the amplitude has a greater influence on the signal itself.
- (b) As can be seen from Fig. 7, using multi-domain RF-DNA for individual recognition is better than single domain. The recognition accuracy of eight power amplifiers has reached 95% when the SNR is -3 dB.
- (c) From the rising trend of the three curves, the multi-domain RF-DNA and time-domain amplitude characteristics have great similarities in the rising trend, which indicates to some extent that the time-domain amplitude characteristics play a major role in the fusion characteristics.

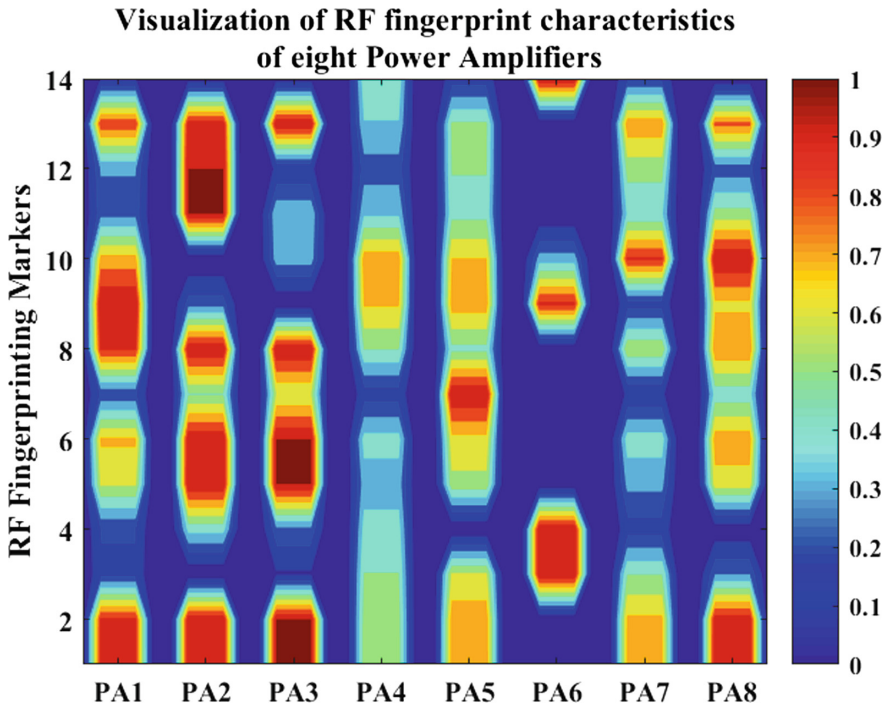


Fig. 6. Visualization of RF fingerprint characteristics

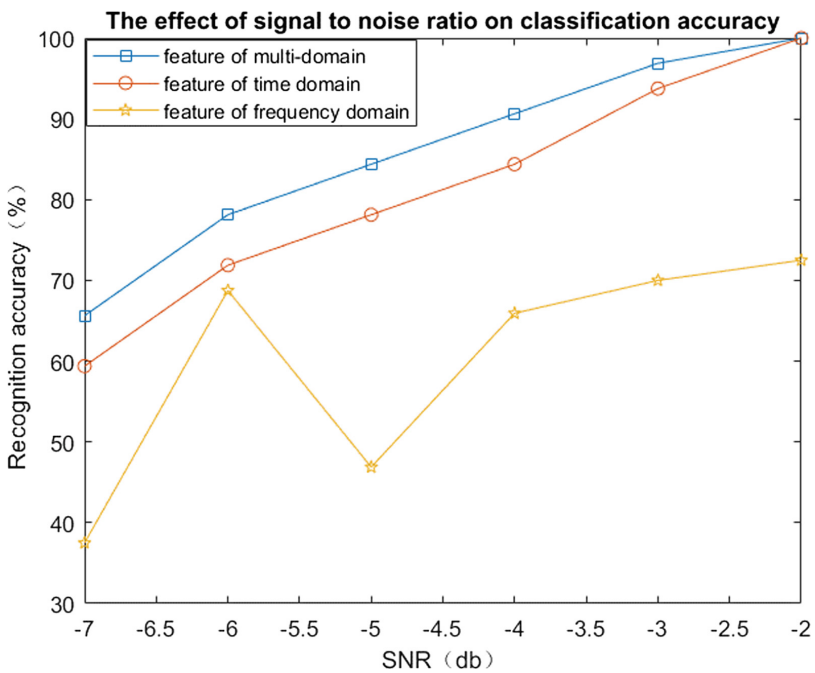


Fig. 7. The effect of signal to noise ratio on classification accuracy

5 Conclusion

In this paper, firstly, the signal is transformed into time-domain, frequency-domain, time-frequency domain and fractal domain to extract 14 features to form RF-DNA. Then we put them together and as the input of the support vector machine, which is determined by selecting the optimal RBF kernel function. Eventually, the recognition rate of the eight power amplifiers is over 95%. This proves that multi-domain RF-DNA is very effective in the identification of individual amplifiers. And even in the case of very low signal-to-noise ratio, the recognition accuracy can meet the required requirements.

Acknowledgment. The authors would like to thank State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System Director Fund (CEMEE2019K0101A).

References

1. Hall, J., Barbeau, M., Kranakis, E.: Detection of transient in radio frequency fingerprinting using signal phase. In: IASTED International Multi-conference on Wireless and Optical Communications (2003)
2. Yu, J.-B., Hu, A.-Q., Zhu, C.-M., Peng, L.-N., Jiang, Y.: RF fingerprinting extraction and identification of wireless communication devices. *J. Cryptologic Res.* **3**(5), 433–446 (2016)
3. Toonstra, J., Nsner, W.: Transient analysis and genetic algorithms for classification. In: Communications, Power, and Computing, pp. 432–437 (1995)
4. Choe, H.C., Poole, C.E., Andrea, M.Y., et al.: Novel identification of intercepted signals from unknown radio transmitters. In: SPIE's 1995 Symposium on OE/Aerospace Sensing and Dual Use Photonics. International Society for Optics and Photonics, pp. 504–517 (1995)
5. Kennedy, I.O., et al.: Radio transmitter fingerprinting: a steady state frequency domain approach. In: Vehicular Technology Conference, VTC 2008-Fall, pp. 1–5. IEEE (2008)
6. Klein, R.W., Temple, M.A., Mendenhall, M.J.: Application of wavelet-based RF fingerprinting to enhance wireless network security. *J. Commun. Netw.* **11**(6), 544–555 (2012)
7. Williams, M.K.D., et al.: RF-DNA fingerprinting for airport WiMax communications security. In: International Conference on Network and System Security, pp. 32–39. IEEE (2010)
8. Reising, D.R., Temple, M.A., Oxley, M.E.: Gabor-based RF-DNA fingerprinting for classifying 802.16e WiMAX mobile subscribers. In: International Conference on Computing, Networking and Communications, pp. 7–13. IEEE (2012)
9. Reising, D.R., Temple, M.A., Jackson, J.A.: Authorized and rogue device discrimination using dimensionally reduced RF-DNA fingerprints. *IEEE Trans. Inf. Forensics Secur.* **10**(6), 1180–1192 (2015)
10. Ru, X., Huang, Z., Liu, Z., et al.: Frequency-domain distribution and band-width of unintentional modulation on pulse. *Electron. Lett.* **52**(22), 1853–1855 (2016)
11. Ru, X.H., Liu, Z., Huang, Z.T., et al.: Evaluation of unintentional modulation for pulse compression signals based on spectrum asymmetry. *IET Radar Sonar Navig.* **11**(4), 656–663 (2017)

12. Dubendorfer, C.K., Ramsey, B.W., Temple, M.A.: An RF-DNA verification process for ZigBee networks. In: Military Communications Conference, Milcom – 2012, pp. 1–6. IEEE (2013)
13. Chen, Y., et al.: Detecting and localizing wireless spoofing attacks. In: Securing Emerging Wireless Systems, pp. 1–18. Springer, Boston (2009). https://doi.org/10.1007/978-0-387-88491-2_8
14. Yuan, H.L.: Research on physical-layer authentication of wireless network based on RF fingerprinting. Southeast University (2011)
15. Bihl, T.J., Bauer, K.W., Temple, M.A.: Feature selection for RF fingerprinting with multiple discriminant analysis and using ZigBee device emissions. *IEEE Trans. Inf. Forensics Secur.* **11**(8), 1862–1874 (2016)