Received 6 June 2020; revised 4 July 2020; accepted 5 July 2020. Date of publication 9 July 2020; date of current version 28 July 2020.

Digital Object Identifier 10.1109/OJAP.2020.3008130

Antenna Classification Using Gaussian Mixture Models (GMM) and Machine Learning

YIHAN MA AND YANG HAO (Fellow, IEEE)

Electronic Engineering Department, Queen Mary University of London, London E1 4NS, U.K.

CORRESPONDING AUTHOR: Y. HAO (e-mail: y.hao@qmul.ac.uk)

This work was supported by IET AF Harvey Research Prize.

ABSTRACT Radio frequency fingerprinting (RFF) is the concept arising from classification of wireless emitters due to their unique radio frequency features. RFF has been further extended to applications including both RF devices classification and wireless signal identification. In this paper, we adopt Gaussian Mixture Models (GMM) technique as feature extraction approach and firstly apply it to extract RFF of antennas. 9 classical antennas with 3 different load conditions (open, short, match) were studied in our experiment. Moreover, we also made a theoretical analysis about the reason scattered signal has the unique features. Specifically, we adopt the Random Noise Radar (RNR) technique to obtain reflected RF signals of antenna under test (AUT) and apply the GMM technique to fit RF signals and then extract the RFF of AUT. A support vector machine (SVM) is proposed to recognize the RFF at different signal-to-noise ratio (SNR) environment. Compared with the conventional feature extraction approaches, for example, from variance, skewness and kurtosis (VSK) values, our method demonstrates better performance on large datasets with classification accuracy above 88% using a SVM classifier. Moreover, the accuracy remains higher than 75% even when the Signal to Noise Ratio (SNR) is equal to 0dB, indicating that the proposed approach has the strong capability of noise immunity.

INDEX TERMS Antenna, Gaussian mixture models, radar signal processing, classification algorithm, machine learning.

I. INTRODUCTION

RADIO frequency fingerprinting (RFF) technique has widely been applied to enhance the security of wireless communications for applications in Internet of Things (IoT) [1]. The underlying principle is that each wireless signal possesses its own unique RF characteristics, which can be used as multi-factor authentication. Over past a few years, several approaches based on amplitude, frequency, and phase parameters have been proposed and studied in time-domain and frequency-domain [2], [3]. Normally, devices under test (DUT) are deployed in the test scenario, and a set of receivers is applied to collect scattered signals. By setting certain data processing workflows, we may obtain unique features so as to achieve device and signal classification.

With the countless consumption of human and financial resources in counterfeits detection, the RFF technique has great market prospects [4]. Moreover, the radio frequency fingerprinting also has its potential to address physical layer security concerns by exploiting unique physical mechanisms of radio propagation [5]–[7], and, therefore, the RFF technique can effectively prevent intrusion by unauthorized devices without upper layer identification [8], [9]. Signal generated from RF devices normally comes from modulated signals, which can be described from its instantaneous characteristics. In [7], researchers capture ramp-up waveforms from several different Wi-Fi radios and investigate the instantaneous attributes of the signals. From the experiment results, the authors find the instantaneous amplitudes among different devices present distinctive features, which can be used as the RFF to recognize radios. After reducing the dimensionality of raw instantaneous amplitude sequence with principal component analysis (PCA) [10], all the radios can be classified with 2% classification error. Besides the instantaneous amplitude, signal's instantaneous phase and frequency can also be applied as the RFF [8]. For the time-domain sequence of samples, the first step is to calculate the instantaneous characteristics (amplitude, phase, and frequency) based on the Hilbert transform. Then each instantaneous sequence

is divided into N_R regions, and the statistical fingerprint features, standard deviation (σ), variance (σ^2), skewness (γ), and kurtosis () are generated from all subsequences and the total sequence ($N_R + 1$ regions). Therefore, each subregion fingerprint can be expressed as a 1×4 vector $\boldsymbol{F}_{R_i} = [\sigma_{R_i}\ \sigma^2_{R_i}\ \gamma_{R_i}\ _{R_i}]_{1\times4}$, and the total regional statistical fingerprint can be concatenated from these subregion fingerprint as $\boldsymbol{F}^r = [\boldsymbol{F}_{R_1} \vdots \boldsymbol{F}_{R_2} \cdots \boldsymbol{F}_{R_{N_R+1}}]_{1\times4(N_R+1)}$. Finally, the time-domain fingerprint with instantaneous amplitude (a), phase (φ) and frequency (f) is formed as $F = [F^a \vdots F^\varphi \vdots F^f]_{1\times4(N_R+1)\times3}$. This statistical feature extraction approach can successfully recognize WiMax devices with above 95% classification accuracy. Similar approaches based on statistical RFF have also been investigated in the frequency-domain and wavelet-domain [2].

Associated with RFF is object classification, its development has been accelerated with the machine learning technique, and found many applications in areas, such as medical diagnosis, sentiment analysis, video surveillance, and authentication. The object classification based on reflected RF signals has been recently studied [11]. Unlike capture the transmit signals to recognize these RF devices, the scattering signal of object is used to detect the shape classes. In this paper, four different shape families (cone, cylinder, plate and spheroid) with unknown roll rates, tumble rates, and unknown initial orientations can be classified based on the noisy monostatic Radar Cross Section (RCS). Objects can absorb and reflect incident electromagnetic waves, which can be captured and analyzed by arbitrary data acquisition systems. Due to the diversity of physical constructions and material configurations among different objects, reflected signals always contain unique characteristic of various objects under the test.

Antenna is a special object, because its scattered signals consist of two parts: the reflected signals from its physical construction and the reradiated signals based on its antenna characteristics [12]. Based on the successful application of RFF in the RF devices authentication, the authors in [13] use similar approach in [8] to extracted the statistical parameters of subregions, i.e., VSK (variance, skewness and kurtosis) values from scattered signals of the antenna as its subregion fingerprinting, and concatenate them together as the regional RFF. Due to the regional RFF themselves are strongly correlated with each other, Multiple Discriminant Analysis (MDA) algorithm is applied to achieve dimensionality reduction. After reducing the dimensionality of RFF, it can achieve antenna classification with a Maximum Likelihood (ML) classifier. In this experiment, the authors successfully classify two UWB antennas (Log-periodic and Vivaldi antennas) with different load conditions (open, short and match) based on the RFF of scattering signals. Moreover, the authors also extended this technique to identifying faulty conditions of devices after the receive antenna [14]. Although the fingerprinting processing based on VSK values can achieve desirable classification accuracies in a small database, it is difficult to maintain the level of accuracy when processing big data.

In this paper, we investigate the application of probabilistic model Gaussian Mixture Models (GMM) to extract RFF from its scattered signals to identify different antennas with arbitrary terminations. White Gaussian Noise (WGN) sequence with stochastic property is generated as the transmitter signals from a Random Noise Radar (RNR) system to actively interrogate targets. The Gaussian distribution has excellent effect to describe the signal which is irrelevant to temporal ordering. Compared with previous research, we first proposed using the GMM technique to extract RFF of scattered signal and apply it to antenna classification. In addition, we also discuss the reasons scattered signal of AUT can be used as the distinctive features. In comparison with the statistical feature extraction approach, the model-based method provides visible accuracy improvement in antenna classification. The layout of this paper is arranged as follows: Section II presents the RFF extraction methodology including data acquisition, theoretical analysis and fingerprinting extraction; Section III compares the classification results of VSK approach with GMMs feature extraction method and provides the classification accuracy vary with different Signal-to-Noise Ratios (SNRs); Finally, the paper is concluded in Section IV.

II. METHODOLOGY

For the object classification purposes, the process of classification can be summarized as follows: data collection, pre-processing, feature extraction and classification [15]. In this section, we mainly introduce the process of feature extraction, including 1) data acquisition; 2) theoretical analysis; 3) RF fingerprinting (RFF) feature extraction.

A. DATA ACQUISITION

For collecting RF signals from different antennas, a radar system is set up in our numerical study. Specifically, two wideband Log-Periodic Antennas (LPAs) (covering the frequency band from 250MHz to 800MHz) are applied, one is used for transmitting, and another is used for receiving. The antennas under test (AUT) are placed far away from the Tx and Rx antennas, as depicted in Fig. 1. A long white noise sequence is used as the incidence to the AUT loaded with different terminations. Scattered signals from the AUT will be collected by the Rx antenna. Preprocessing is the first step after obtaining raw scattered signals. In this experiment, data filtering with passband between 300MHz and 700MHz is used to remove the unwanted features from raw signals, and data normalization is applied to get rid of the influence of amplitude error and regulate the data. In order to consider the application in the real environment, the additive white Gaussian noise model is implemented after data normalization. Fig. 2 illustrates the whole demonstration process for antenna classification.

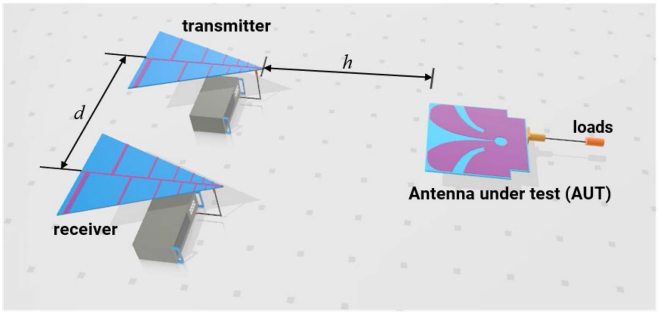

FIGURE 1. The experimental scenario for antenna classification. The distance between transmit antenna and receive antenna d is 0.5m, and the distance between transmitter and AUT h is 1.25m, with the AUT bore-sighted on the transmit and receive antennas.

TABLE 1. Antenna types applied in analysis.

Antenna	Size (mm^2)	Gain (dB)	Bandwidth (MHz)
	480 MHz		
Dipole	272×10	2.09	480-520
Bowtie	290×170	2.30	480-750
Planar Elliptical Dipole	240×200	2.25	480-1100
Printed Folded Dipole	248×168	3.61	480-580
Planar Inverted F	220×110	2.16	480-540
Microstrip	222×222	6.23	480-489
Yagi	550×200	8.37	480-575
Log-Periodic	328×250	6.20	480-1250
Vivaldi	290×223	6.46	480-1500

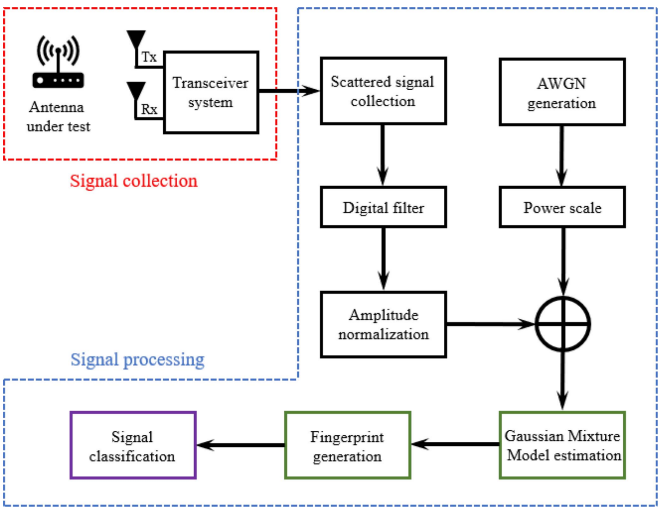

FIGURE 2. Overall process for signal collection, SNR addition, Model fitting, fingerprint generation, and classification.

Overall, we have studied and investigated 9 different types of antennas, namely, Dipole, Bowtie, Planar Elliptical Dipole, Printed Folded Dipole, Planar Inverted F Antenna (PIFA), Microstrip, Yagi, Log-Periodic, and Vivaldi Antennas, respectively with 3 different load conditions (open, short and match) to establish a dataset for all AUTs of 27 samples from numerical simulations. A detail description of these antennas is presented in Table 1. In this experiment, we expect to consider the influence of structure and properties to the classification accuracy. To guarantee the collected signals are comparable, all antennas are designed to operate at the lowest frequency of 480MHz.

B. THEORETICAL ANALYSIS

The underlying principle that scattering signals can be classified for different antennas with different load conditions can be inferred from [12]. Antennas can be referred to as a special electromagnetic wave scatter whose scattering properties are directly linked to their loads and structures. When the antenna is matched, the scattering signals are only related to its structural mode, otherwise, consists of structural mode scattering and antenna mode scattering. The structural mode scattering is related to the physical structure of antennas as those of other passive components, and the antenna mode scattering is influenced by its electromagnetic characteristics. The relationship between the two modes is given by [16]:

$$\sigma^{total} = \left|\sqrt{\sigma^{st}} + \sqrt{\sigma^{an}}e^{j\varphi}\right|^2 \tag{1}$$

where σ^{total} is the total radar cross-section (RCS), σ^{st} is the structural mode RCS, and σ^{an} is the antenna mode RCS. φ is the phase difference between the two modes.

The simplified structural scattering field can be expressed as follow [17]:

$$\overrightarrow{\mathrm{H}}_{\mathrm{st}}^{\mathrm{s}} = \frac{\mathrm{jk}}{4\pi \mathrm{r}}\mathrm{e}^{-\mathrm{jkr}} \cdot \iint \left[\left(2\hat{\mathrm{n}} \times \overrightarrow{\mathrm{H}}^{\mathrm{i}}\right) \times \hat{\mathrm{r}}\right] \cdot \mathrm{e}^{-\mathrm{jk}\hat{\mathrm{r}}\cdot\overrightarrow{\mathrm{r}'}}\,\mathrm{ds}' \tag{2}$$

where k is the wavelength, $\overrightarrow{H}^{\mathrm{i}}$ is the incident magnetic field, $\hat{n}$ is the unit vector normal to structure, $\hat{r}$ is the direction of incident wave, $\overrightarrow{r'}$ is the integration variable vector from origin to any surface point.

In this study, the incident magnetic field $\overrightarrow{H}^{i}$ is dependent on the direction of arrival (DoR) of the incident wave, that is, $\overrightarrow{H}^{i} = H_i(\omega, r, \theta, \varphi)$. Moreover, the scattering electric field of structural mode $\overrightarrow{E}^{s}_{st} = E^{s}_{st}(\omega, r, \phi, \varphi, s')$ is dependent on the bandwidth and direction of incident wave and the physical structure of antenna.

The scattering electric field of antenna mode can be expressed as follow:

$$\overrightarrow{E}^{s}_{an} = \overrightarrow{E}^{i}\overrightarrow{G}(\omega', \theta', \varphi')\frac{\lambda'\mu\Gamma}{4\pi r} \tag{3}$$

where $G(\omega', \theta', \varphi')$ is the radiation pattern of the scattering antenna, Γ represents the voltage reflection coefficient when antenna is mismatched. The scattering electric field of antenna mode $\overrightarrow{E}^{s}_{an} = E^{s}_{an}(\omega, r, \theta, \varphi, \Gamma, G)$ is dependent on the bandwidth and direction of incident wave and the load condition and radiation pattern of AUT.

In this study, we can classify scattered signals into the following three parts:

- Leakage signals from the Tx antenna;
- Structural mode scattering;
- Antenna mode scattering.

Based on (1) - (3), the received electric field can be expressed as (neglecting the polarization losses):

$$\overrightarrow{E_r} = E_{le}(\omega, r, \theta'', \varphi'') + \Big[E^s_{st}(\omega, r, \theta, \varphi, s') + E^s_{st}(\omega, r, \theta, \varphi, \Gamma, G) \cdot e^{j\emptyset_1}\Big] \cdot e^{j\emptyset_2} \tag{4}$$

where $E_{le}(\omega, r, \theta'', \varphi'')$ is the leakage electric field from transmit port, $\emptyset_1$ is the phase difference between scattering electric field of structural mode and scattering electric field of antenna mode, $\emptyset_2$ is the phase difference between scattering electric field and leakage electric field.

For the same antenna in different load conditions: the leakage signals $E_{le}(\omega, r, \theta'', \varphi'')$ and structure mode signals $\overrightarrow{E}^s_{st} = E^s_{st}(\omega, r, \phi, \varphi, s')$ parts among different load conditions are the same. However, because the reflection coefficient Γ of AUT is related to the terminations, the generation of antenna mode signals $\overrightarrow{E}^s_{an} = E^s_{an}(\omega, r, \theta, \varphi, \Gamma, G)$ vary with loads. Therefore, antennas with different terminations can be classified according to the terminations, the generation of antenna mode signals $\overrightarrow{E}^s_{an} = E^s_{an}(\omega, r, \theta, \varphi, \Gamma, G)$ vary with loads. Therefore, antennas with different terminations can be classified accordingly.

For the different antennas with same load conditions: the leakage signals $E_{le}(\omega, r, \theta'', \varphi'')$ are identical. While the antenna structure s' in structure mode $E^s_{st}(\omega, r, \phi, \varphi, s')$ and the radiation pattern G in $E^s_{an}(\omega, r, \theta, \varphi, \Gamma, G)$ are different, which result in the various scattering signals from the AUT.

C. RF FINGERPRINTING FEATURE EXTRACTION

Skewness and kurtosis are the statistical parameters which have been widely implemented to describe the histogram of signals. Specifically, the skewness can be regarded as the degree of distortion from the symmetrical bell curve in terms of the histogram of a sequence. Positive skewness means the tail on the curve's right-side is larger than the tail on the left-side, and negative skewness means the tail towards on the opposite side. The kurtosis is all about the tails of the distribution. High kurtosis in a histogram indicates data has heavy tails or outliers, and low kurtosis presents light tails. For the signals with only one "peak" in the distribution of data, they can be fitted well by a single value of skewness and kurtosis. While for the signals with more than one "peak", they will generally generate a poor fit [18]. In our experiment, although the quasi-white noise signals in the transmit port can be regarded as the signal with one "peak" Gaussian distribution, the scattering signals always present multiple "peaks" properties in the data. To fit scattering signals appropriately, we introduce the Gaussian mixture models (GMMs) algorithm in this section.

GMMs is a model-based fitting algorithm to partition data sequences to pre-defined normal distribution models [19]. As a probabilistic method for obtain statistical features of clusters, GMMs is quite suitable to fit data with characteristics of probability distribution. A Gaussian mixture is an equation that is consist of several normal distributions, which is defined as (5):

$$p(x) = \sum_{m=0}^{M} \omega_m p_m(x) = \sum_{m=0}^{M} \omega_m \mathcal{N}(x; u_m, \Sigma_m) \tag{5}$$

where M is the number of mixtures, $\mathcal{N}(x; u_m, \Sigma_m)$ is a normal distribution with mean value u_m and covariance matrix Σ_m, and ω_m is the mixture weight with the constraint that $\sum \omega_m = 1$. Therefore, the raw scattered signals can be expressed with solely u_m, Σ_m, ω_m and M parameters in the GMM.

The number of mixtures M can be selected from the Bayesian information criterion (BIC), BIC is a criterion for the model selection among a finite set of models by introducing a penalty term for the number of parameters in the model, and it can be defined as:

$$\mathrm{BIC}(\theta) = -2\log p(x \mid \theta) + k\log(n) \tag{6}$$

where θ is the model, $p(x|\theta)$ is the maximized value of the likelihood function of model θ, k is the number of parameters in the model θ, and n is the number of observations. This criterion can evaluate and select suitable number of mixtures with minimum BIC values from the generated model.

As for other parameters u_m, Σ_m, and ω_m are estimated using the Expectation-Maximization (EM) algorithm [20]. This iterative method yields a Maximum Likelihood (ML) estimate, via the estimation formulas:

$$\mu'_m = \frac{\sum_{t=1}^{T} p_m(x_t) x_t}{\sum_{t=1}^{T} p_m(x_t)} \tag{7}$$

$$\sum_m^{'} = \frac{\sum_{t=1}^{T} p_m(x_t)(x_t - \mu_m)^T (x_t - \mu_m)}{\sum_{t=1}^{T} p_m(x_t)} \tag{8}$$

$$w'_m = \frac{\sum_{t=1}^{T} p_m(x_t)}{\sum_{t=1}^{T} \sum_{m=1}^{M} p_m(x_t)} \tag{9}$$

Based on the EM iteration method, signals can be optimized and fitted as several Gaussian models with a certain mean value u_m, covariance matrix Σ_m, and mixture weight ω_m. In our experiment, the number of mixtures M is optimized as 5, and the type of covariance matrix is selected as diagonal matrix to guarantee only the diagonal parameters of covariance matrix are estimated. Therefore, The RFFs with 15 parameters $[\omega_1, \mu_1, \sum_1, \omega_2, \mu_2, \sum_2, \ldots, \omega_5, \mu_5, \sum_5,]$ are extracted from GMMs.

III. RESULT AND DISCUSSION

Previous research [2], [7], [12], and [14] mostly apply Multiple Discriminant Analysis (MDA) algorithm to reduce features dimension and then use Maximum Likelihood (ML) approach as the classifier for classification. In their research, the length of fingerprint sequence corresponding with the number of sub-regions in the processed signals. Normally, the length of features extracted in these researches is

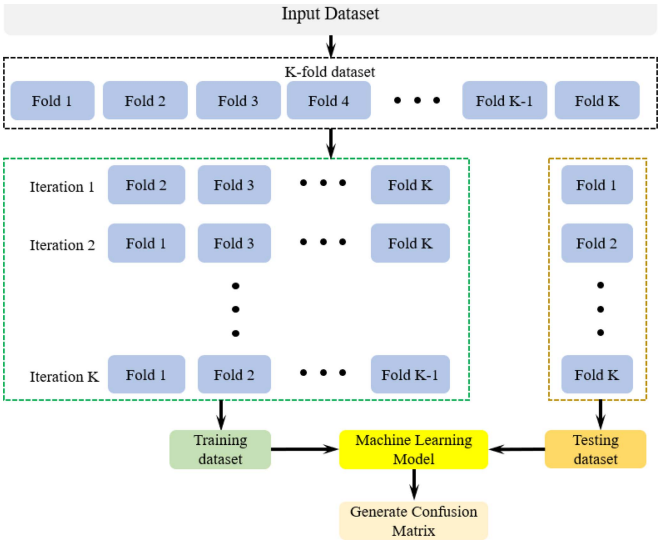

FIGURE 3. Machine Learning classification process with K-fold cross validation.

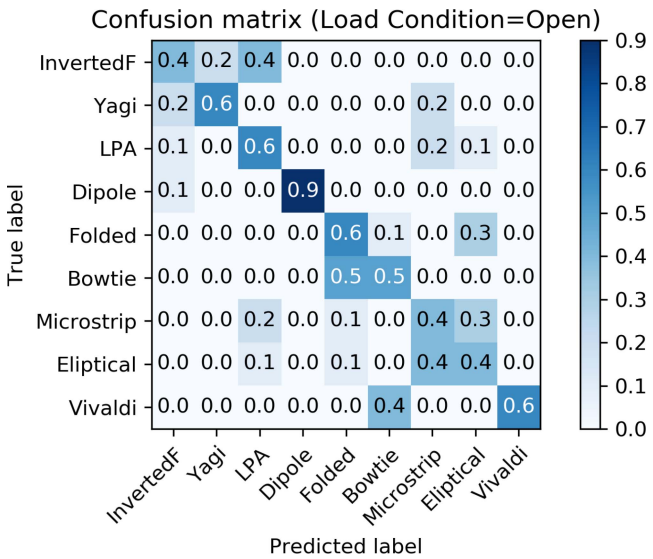

FIGURE 4. Normalized confusion matrix for VSK features with open termination.

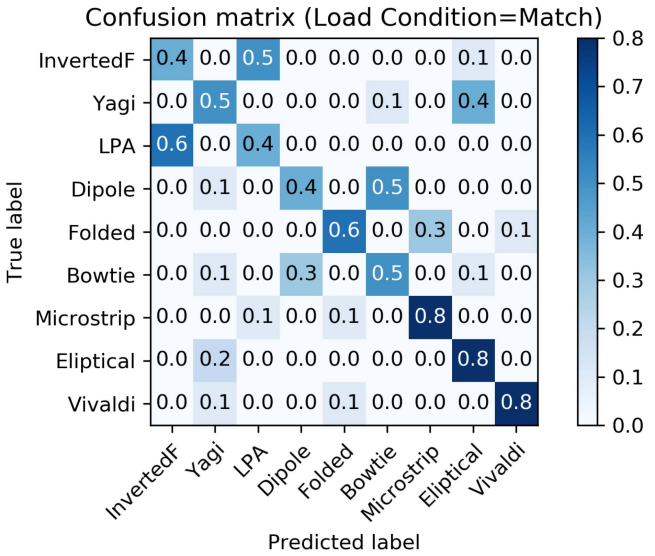

FIGURE 5. Normalized confusion matrix for VSK features with match termination.

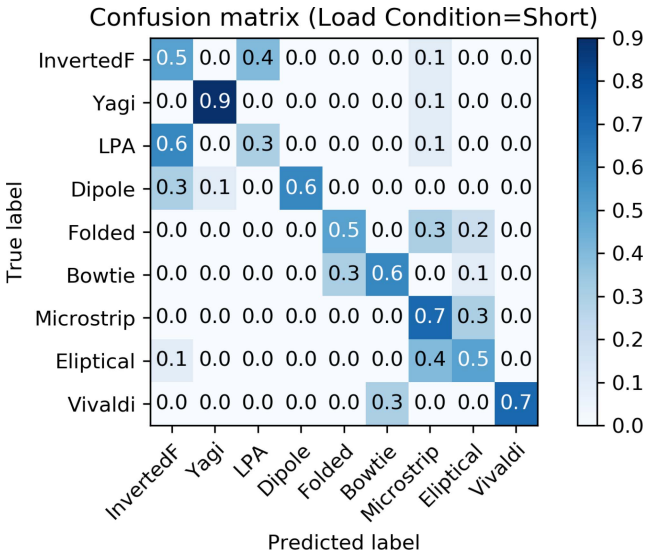

FIGURE 6. Normalized confusion matrix for VSK features with short termination.

more than several hundreds. With the MDA algorithm, a n-dimensional space can be projected to a (d-1) dimensional space, where d is the number of categories. While it is unnecessary to apply dimensionality reduction of a dataset with only 15 features. In the result section, three pattern classification methods have been implemented: support vector machine (SVM), random forest (RF), and k-nearest neighbour (KNN) [21], [22].

The traditional SVM classifier is a non-probabilistic binary linear classifier. With the one-against-one strategy, multi-class SVM can be achieved [21]. As one of the most popular strategy for supervised machine learning and classification, the SVM technique is suitable to analyse small dataset with good generalization capabilities and stability. In this experiment, we adopt one-against-one approach with linear kernel function which would extend SVM to $k(k-1)/2$ sub-classifiers, where k is the classes. RF is an ensemble learning approach for classification or regression, which consists of a large number of individual decision trees [22]. Based on the bagging algorithm with ensemble learning technique, RF robust to outliers and non-linear data. A RF classifier with 200 decision trees is applied to our experiment. KNN algorithm is one of the simplest algorithms and it is also one of widely used approach in classification and quantification problems. 1-NN classifier is selected to find the label of single nearest neighbour as the object's label. In order to make full use of dataset and reduce variability, K-fold cross validation technique is applied in our experiment as well. The process of ML with cross validation is shown in Fig. 3. The input features are partitioned into K equal sub-regions, then collect the K-1 folds as the training dataset, and the rest

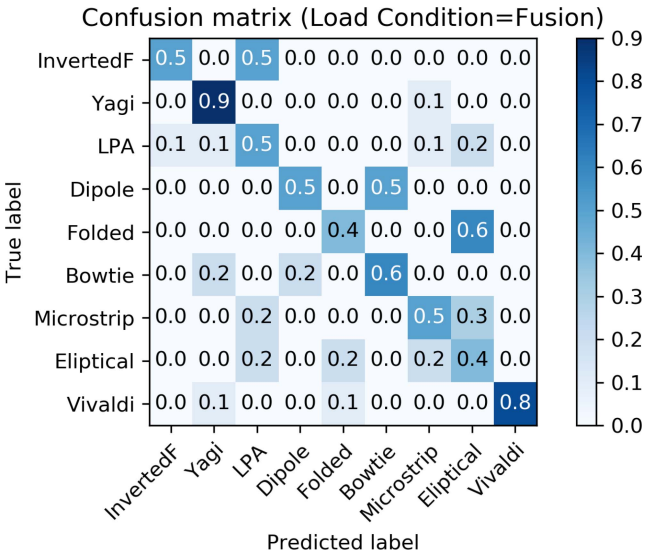

FIGURE 7. Normalized confusion matrix for VSK features with fusion termination.

TABLE 2. Classification result in different classifier at different SNR.

	Accuracy (%)		
VSK statistical fingerprint			
SNR\Classifier	SVM	RF	KNN
0 dB	43.8	42.9	34.8
10 dB	51.9	52.2	39.2
20 dB	56.5	55.9	41.3
30 dB	56.7	56.3	41.3
GMMs fingerprint			
SNR\Classifier	SVM	RF	KNN
0 dB	78.6	80.2	64.2
10 dB	81.4	81.1	65.5
20 dB	89.8	88.7	71.1
30 dB	91.1	90.2	73.0

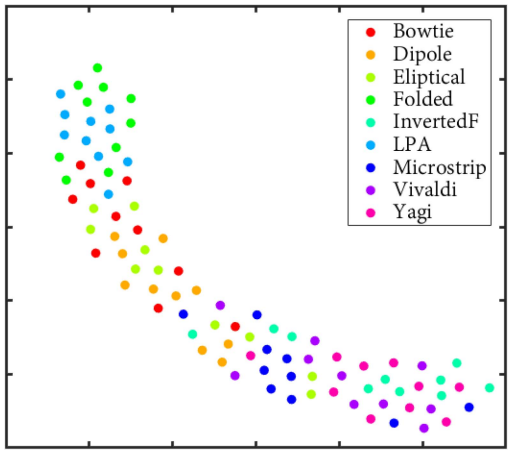

FIGURE 8. Visualized results from VSK statistical features.

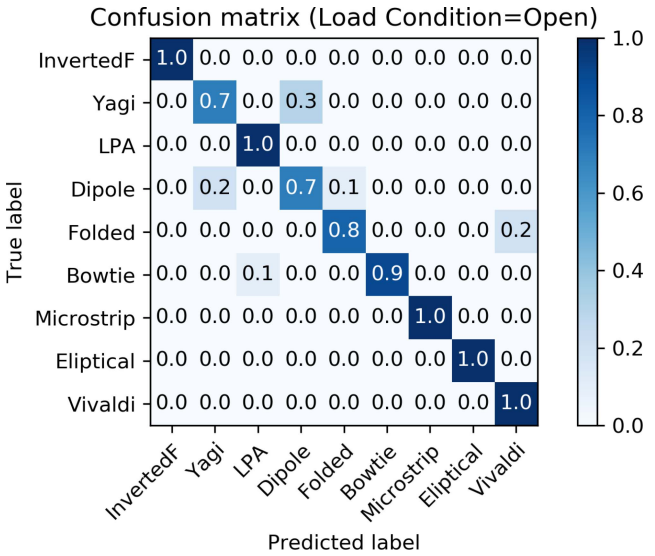

FIGURE 9. Normalized confusion matrix for GMM feature with open termination.

one fold as the testing dataset. A machine learning model can be generated from the K-1 folds training dataset, and a confusion matrix from the rest one fold testing dataset can be obtained with this machine learning model. Repeat this process K iterations until the confusion matrix for each fold is obtained. Finally, accumulating all the classification results together to finish the ML process.

Table 2 shows the classification accuracies in three different classifiers at various Signal-to-noise Ratio (SNR). All the results are shown under arbitrary terminations. It reflects the great differences that exist between performance of VSK statistical fingerprint and of GMMs fingerprint. The feature extraction approach from the VSK statistical fingerprint is widely used in pervious researchers [2], [7], [13], and [14]. Both SVM and RF classifiers superior to the KNN classifier in terms of the GMMs features and VSK features. Moreover, the performance in SVM classifier is slightly better than in RF classifier. It can be seen from the table that the accuracy of SVM and RF classifiers is about 56% with VSK features at SNR=30dB environment. When SNR decrease to 0dB, the accuracy steady decrease to about 43%. While the GMMs fingerprint still presents about 80% accuracy at SNR = 0dB.

Fig. 4-7 presents the confusion matrix with 3 different loads (open, short and match) separately and random selected terminations based on the approach of statistical feature extractions, approach a K=10 K-fold cross validation with SVM classifier. The average classification over 10 replications of SVM ranged from 55.6% (open) to 62.2% (short). Short load conditions generate slightly better results, but the difference is not significant. T-distributed Stochastic Neighbour Embedding (t-SNE) algorithm is a nonlinear dimensionality reduction algorithm used for visualizing high dimensional data in low-dimensional 2D or 3D space [23]. Fig. 8 shows the results of our VSK statistical features with t-SNE plot at SNR = 30dB. The large overlaps among different antenna features also indicates the classification result is undesirable.

The average classification accuracy based on GMMs approach with SVM classifier is shown in Fig. 9-12. As expected, the overall accuracy of prediction compared to the previous approach is largely improved. The

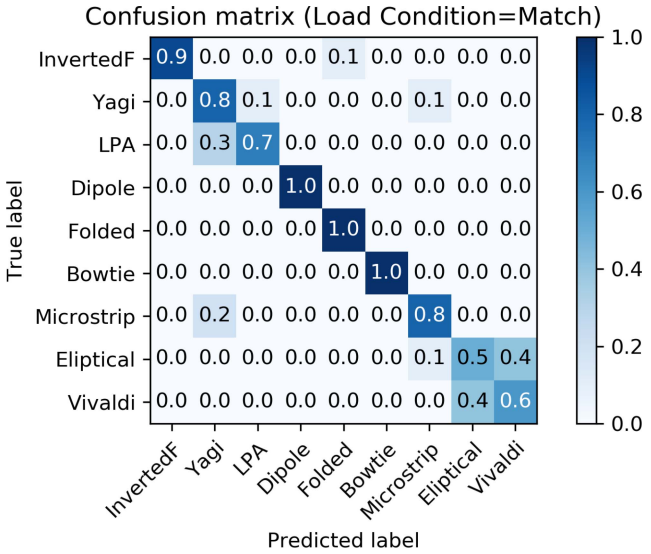

FIGURE 10. Normalized confusion matrix for GMM feature with match termination.

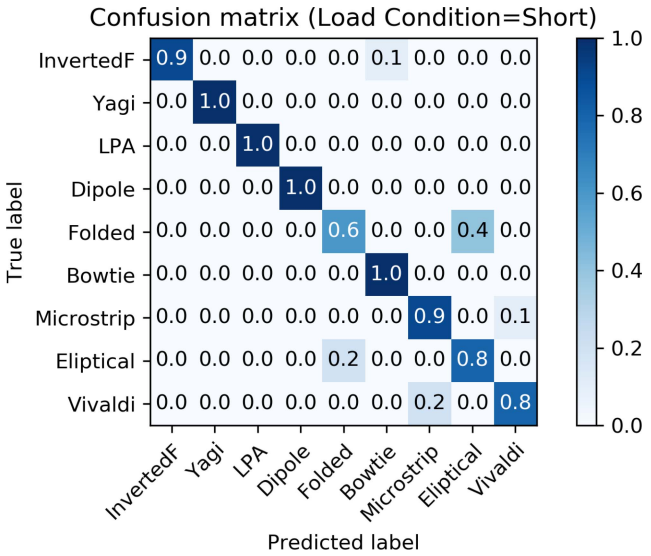

FIGURE 11. Normalized confusion matrix for GMM feature with short termination.

average accuracy between 81.1% (match) and 91.1% (fusion), which indicates the GMMs based feature extraction approach is fairly promising. Fig. 13 presents the t-SNE plot from GMMs features with fusion termination at SNR = 30dB. From the visual inspection, it is easy to identify that GMMs features generate excellent classification accuracy.

From the results we can infer that the accuracy is slightly influenced by load conditions in terms of antennas classification. In the previous approach based on statistical parameters feature extraction, the average accuracy of random select terminations is 56.7%, which is similar to other load conditions. In our approach, although the average accuracy of mixed loads is the highest among four situations, it is only 1.1% and 2.2% higher than open loads and short respectively. Moreover, almost 20% accuracy is improved from

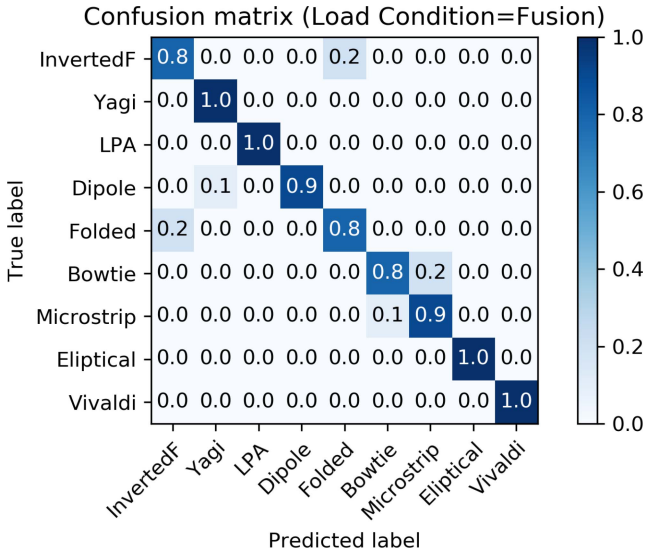

FIGURE 12. Normalized confusion matrix for GMM feature with fusion termination.

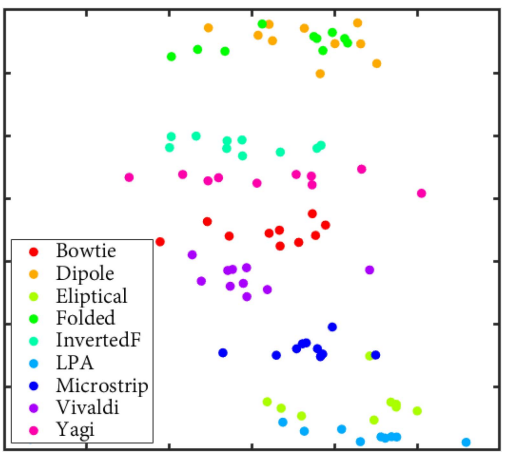

FIGURE 13. Visualized results from GMMs features.

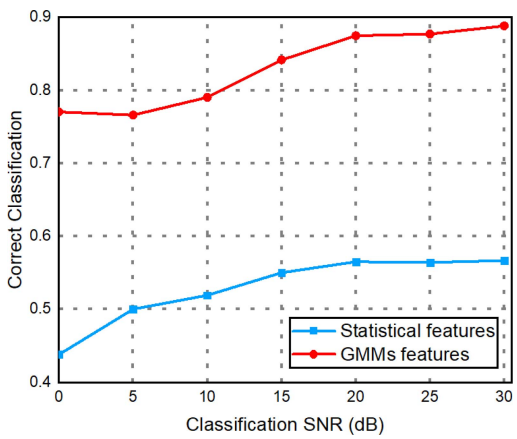

FIGURE 14. Average classification accuracy under different RF fingerprints.

our approach, which also indicates the GMMs outperforms the solely VSK statistical parameters extraction in terms of feature extraction process.

The classification accuracy under different Signal-to-Noise Ratios (SNR) with SVM classifier are also investigated in our experiment. Fig. 14 illustrates the results under different SNRs for antenna classification with arbitrary terminations. When SNR is below 5dB, the accuracy based on statistical features is less than 50% and begins to decrease rapidly. Above 20dB, the accuracy remained constant at 56.5%. Whereas for the GMM approach, over the SNR from 20dB to 30dB, the accuracy remained level, and it exhibits decreased performance at SNR$<$20dB and achieves about 78% accuracy at SNR $=$ 10dB.

IV. CONCLUSION

In this paper, we have investigated the application of Gaussian Mixture Models (GMM) approach to extract RF fingerprinting (RFF) of scattered signals from Random Noise Radar (RNR) system. The stochastic scattered signals are robust to noisy environment and are suitable for fitting with GMM. Mean values, weights and covariances of mixed models extracted from GMM are then applied to classify antennas with various terminations. It is demonstrated that the Random Noise Radar (RNR) transceiver system can be used to classify antennas with arbitrary terminations based on reflected signals. Our experiments on the antenna datasets confirm the effectiveness of our framework and its superiority over the traditional statistical parameter extraction approach. The fingerprint generated from the GMMs yields robust results at SNR below 20dB, with about 78% accuracy obtained at 10dB.

REFERENCES

[1] J. Zhang, S. Rajendran, Z. Sun, R. Woods, and L. Hanzo, "Physical layer security for the Internet of Things: Authentication and key generation," *IEEE Wireless Commun.*, vol. 26, no. 5, pp. 92–98, Oct. 2019.

[2] R. W. Klein, M. A. Temple, and M. J. Mendenhall, "Application of wavelet-based RF fingerprinting to enhance wireless network security," *J. Commun. Netw.*, vol. 11, no. 6, pp. 544–555, Dec. 2009.

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

[4] G. DeJean and D. Kirovski, "RF-DNA: Radio-frequency certificates of authenticity," in *Proc. Int. Workshop Cryptographic Hardw. Embedded Syst.*, 2007, pp. 346–363.

[5] B. Danev, D. Zanetti, and S. Capkun, "On physical-layer identification of wireless devices," *ACM Comput. Surveys*, vol. 45, no. 1, pp. 1–29, 2012.

[6] W. Wang, Z. Sun, S. Piao, B. Zhu, and K. Ren, "Wireless physical-layer identification: Modeling and validation," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 9, pp. 2091–2106, Sep. 2016.

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

[8] M. D. Williams, S. A. Munns, M. A. Temple, and M. J. Mendenhall, "RF-DNA fingerprinting for airport WiMax communications security," in *Proc. 4th Int. Conf. Netw. Syst. Security*, Melbourne, VIC, Australia, 2010, pp. 32–39.

[9] X. Wang, Y. Zhang, H. Zhang, X. Wei, and G. Wang, "Identification and authentication for wireless transmission security based on RF-DNA fingerprint," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, no. 1, p. 230, 2019. [Online]. Available: https://doi.org/10.1186/s13638-019-1544-8

[10] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *J. Educ. Psychol.*, vol. 24, no. 6, pp. 417–441, 1993.

[11] E. Wengrowski, M. Purri, K. Dana, and A. Huston, "Deep CNNs as a method to classify rotating objects based on monostatic RCS," *IET Radar Sonar Navig.*, vol. 13, no. 7, pp. 1092–1100, Jul. 2019.

[12] Y. Liu, S.-X. Gong, and D.-M. Fu, "A novel model for analyzing the RCS of microstrip antenna," in *IEEE Antennas Propag. Soc. Int. Symp. Dig. USNC/CNC/URSI North Amer. Radio Sci. Meeting (Cat. No.03CH37450)*, vol. 4. Columbus, OH, USA, 2003, pp. 835–838.

[13] M. W. Lukacs, A. J. Zeqolari, P. J. Collins, and M. A. Temple, "'RF-DNA' fingerprinting for antenna classification," *IEEE Antennas Wireless Propag. Lett.*, vol. 14, pp. 1455–1458, 2015.

[14] M. Lukacs, P. Collins, and M. Temple, "Device identification using active noise interrogation and RF-DNA 'fingerprinting' for non-destructive amplifier acceptance testing," in *Proc. IEEE 17th Annu. Wireless Microw. Technol. Conf. (WAMICON)*, Clearwater, FL, USA, 2016, pp. 1–6.

[15] G. Baldini, G. Steri, R. Giuliani, and C. Gentile, "Imaging time series for Internet of Things radio frequency fingerprinting," in *Proc. Int. Carnahan Conf. Security Technol. (ICCST)*, Madrid, Spain, 2017, pp. 1–6.

[16] W. Wang, Y. Liu, S. Gong, Y. Zhang, and X. Wang, "Calculation of antenna mode scattering based on method of moments," *Progr. Electromagn. Res. Lett.*, vol. 15, pp. 117–126, 2010, doi: 10.2528/PIERL10051704.

[17] W. Stutzman and G. Thiele, *Antenna Theory and Design*. Hoboken, NJ, USA: Wiley, 2013.

[18] P. Westfall, "Kurtosis as peakedness, 1905–2014.R.I.P." *Amer. Stat.*, vol. 68, no. 3, pp. 191–195, 2014.

[19] A. Samé, F. Chamroukhi, G. Govaert, and P. Aknin, "Model-based clustering and segmentation of time series with changes in regime," *Adv. Data Anal. Classif.*, vol. 5, no. 4, pp. 301–321, 2011.

[20] T. Moon, "The expectation-maximization algorithm," *IEEE Signal Process. Mag.*, vol. 13, no. 6, pp. 47–60, Nov. 1996.

[21] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 415–425, Mar. 2002.

[22] T. K. Ho, "Random decision forests," in *Proc. 3rd Int. Conf. Doc. Anal. Recogn.*, vol. 1. Montreal, QC, Canada, 1995, pp. 278–282.

[23] L. van der Maaten and G. Hinton. "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

YIHAN MA received the bachelor's and master's degrees from the School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China, in 2014 and 2017, respectively. He is currently pursuing the Ph.D. degree with the Queen Mary University of London, U.K. His research interests include the antenna design and machine learning.

YANG HAO (Fellow, IEEE) received the Ph.D. degree in computational electromagnetics from the Centre for Communications Research, University of Bristol, Bristol, U.K., in 1998.

He was a Postdoctoral Research Fellow with the School of Electronic, Electrical and Computer Engineering, University of Birmingham, Birmingham, U.K. He is currently a Professor of antennas and electromagnetics with the Antenna Engineering Group, Queen Mary University of London, London, U.K. Over the past seven years, he has led several major research projects with the total funding exceeding £25M as PI. This includes his Directorship of £4.6M EPSRC QUEST Programme Grant and Co-I and Management Board Membership of £23M Cambridge Graphene Centre. His work has been recognised both nationally and internationally through his books "Antennas and Radio Propagation for Body-Centric Wireless Communications" and "FDTD Modeling of Metamaterials: Theory and Applications," (Artech House, USA) and highly cited papers published in leading journals, including *Nature Group Publications*, *Physical Review Letters*, *Applied Physics Letters*, IEEE proceedings, and transactions. These research results have been taken up by industry in the U.K., and overseas. His work has raised the international status of U.K. antenna research, and attracted more than 14 000 citations leading to several patents with industries. He has supervised more than 66 Ph.D. students and postdoctoral, employed by the industry and universities as professors. Four of them set up respective spinouts. His research on transformation optics and metamaterials have led to many tangible benefits for a range of industrial products. One example is lens antenna designs for satellite communications. He has been instrumental in establishing a complete supply chain, entailing the electromagnetic modeling, materials design and fabrication, lens structure manufacture and antenna system development. This technology has been fully scoped and is currently commercialised under a $45M startup of Isotropic System Limited.

Prof. Hao won many accolades, including the prestigious AF Harvey Prize in 2015, the BAE Chairman's Silver Award in 2014, and the Royal Society Wolfson Research Merit Award in 2013. He was a Strategic Advisory Board Member for EPSRC, where he is committed to championing RF/microwave engineering for reshaping the future of U.K. manufacturing and electronics. He made some important contributions to the roadmap of machine-learning for the Royal Society in 2016, and white papers on future U.K. Industry Strategic for the IET in 2017. Along with many invitations to international conferences, he was asked to chair the Theo Murphy international scientific meeting by the Royal Society. Internationally, he has achieved his reputation for his leadership in IEEE, being an Editor-in-Chief for IEEE ANTENNAS AND WIRELESS PROPAGATION LETTERS from 2013 to 2017 and founding a new open access journal *EPJ Applied Metamaterials* dedicated to metamaterials research in 2015. He currently serves as the AdCom Member and the Chair of Publication Committee for IEEE Antennas and Propagation Society.