IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 25, 2019, accepted July 12, 2019, date of publication July 16, 2019, date of current version August 5, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2929316

Improved Blind Spectrum Sensing by Covariance Matrix Cholesky Decomposition and RBF-SVM Decision Classification at Low SNRs

JIANRONG BAO[1,2], (Senior Member, IEEE), JIANYUAN NIE[1], CHAO LIU[1], BIN JIANG[1], FANG ZHU[1], AND JIANHAI HE[3]

[1]School of Communication Engineering, Hangzhou Dianzi University, Hangzhou 310018, China
[2]Zhejiang Provincial Key Laboratory of Information Processing, Communication and Networking, Hangzhou 315800, China
[3]School of Electronic Information Engineering, Ningbo Polytechnic, Ningbo 315800, China

Corresponding author: Chao Liu (liuchao@hdu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant U1809201, in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LY17F010019, in part by the Zhejiang Provincial Science and Technology Plan Project under Grant LGG18F010011 and Grant LGG19F010004, and in part by the Open Project of the Zhejiang Provincial Key Laboratory of Information Processing, Communication and Networking, Zhejiang, China, under Grant IPCAN-1802.

ABSTRACT An improved blind spectrum sensing scheme is established by the covariance matrix Cholesky decomposition and radial basis function (RBF)-support vector machine (SVM) decision classification at low signal-to-noise ratios (SNRs). Under strong background noises, the proposed scheme improves the recognition rate of primary users (PUs) than that of the current blind spectrum sensing. First, the ratio of the maximum-to-minimum eigenvalue of a covariance matrix obtained by the Cholesky decomposition is used to construct the statistics. Second, the statistics are labeled with "+1" or "−1," namely, the energy characteristics of the training samples are extracted and marked with "+1" for PUs and "−1" for noises. Finally, an RBF-SVM classification model, with an intelligent RBF as the SVM kernel function, is obtained by training the above-mentioned statistics and the labels. Thus, the received signals are classified as PUs or not be trained in the SVM model. The threshold possesses self-learning ability, and it distinguishes PU signals from noises effectively. The classification among PU signals and noises is implemented by the optimal SVM decision boundary, derived from maximizing the margin of the decision boundary of trained samples for efficient detection. In addition, the complexity of the statistic construction is lower than that of the conventional maximum minimum eigenvalue (MME). The simulation results show that the RBF in our scheme has 77.5% accuracy at −10 dB, and it outperforms linear kernel function significantly by about 27.5% in accuracy at −10 dB. In addition, the average error probability of the proposed scheme is reduced by about 26% when compared with those of original SVM schemes at −20 dB. The proposed scheme also outperforms the current MME detection in detection probability over 10% at −20 dB. Therefore, the proposed blind spectrum sensing scheme can be efficiently used to detect the PUs by the covariance matrix Cholesky decomposition and the RBF-SVM decision classification in the fifth-generation (5G) communications, especially at low SNRs.

INDEX TERMS Blind spectrum sensing, Cholesky decomposition, RBF-SVM, covariance matrix, decision classification.

I. INTRODUCTION

Spectrum sensing is one of significant techniques in cognitive radio (CR), which helps alleviate spectrum shortage in wireless communications. Traditionally, spectrum sensing mainly included energy detection (ED), matched filter detection, and cyclostationary feature detection methods [1], [2]. The ED method without prior information of a primary user (PU) usually had low computational complexity. However, it was sensitive to the uncertainty of noise range and it could not distinguish a PU signal from noises at low signal-to-noise ratios (SNRs) [3]. Despite of the short detection duration and

The associate editor coordinating the review of this manuscript and approving it for publication was Wen-Long Chin.

satisfactory precision in matched filter detection, it required prior information of PUs in advanced and it had high computational complexity [4]. Cyclostationary feature detection had high recognition rate at low SNRs. But its complexity was high, which also led to successive large detection latency [5].

Generally, spectrum sensing methods needed prior information about signals or noises, which confines their practical use. Thus, some blind methods occurred [6], [7], where the eigenvalue, and also the covariance based sensing methods were widely investigated. They just consider the correlation among signals without any prior information. The eigenvalue-based spectrum sensing methods performed better than the ED ones, when PU signals were highly correlated [8]. In [9], a maximum eigenvalue (ME) method about the covariance matrix was proposed. However, it was restricted in practice due to the difficulty of obtaining the detection threshold derived from the impractical assumptions that the number of received signals and the dimension of the covariance matrix were infinite. Consequently, it did not permit an accurate decision threshold, which needed to be operated with a finite number of signal samples [10]–[12]. The spectrum sensing methods by Cholesky decomposition overcame the difficulty of decision thresholds for improved performance, when compared with the eigenvalue-based methods. The test statistic in [13] was the ratio of the sum of all squares of the elements from the signal covariance matrix obtained by Cholesky decomposition, to the sum of the squares of the diagonal elements from that matrix. But in [14], it was the ratio of the first diagonal element to the last diagonal element of the covariance matrix by Cholesky decomposition. Thus, the performance of spectrum sensing in [13], [14] mainly depended on the selected elements of the covariance matrix obtained by Cholesky decomposition. This effect resulted in non-robust detection performances of these methods. In [15], a blind detection was proposed for good performance, featured with Cholesky decomposition of a covariance matrix as the criterion to determine the vacant radio frequency (RF) band. However, its decision threshold was fixed with its derivation rather than its self-learning capability.

In CR systems, cognitive devices were capable of learning and reasoning, because they needed to identify the activities of PUs in their RF environments [16]. Also machine learning (ML) had been utilized for spectrum sensing in CR [17], [18]. And supervised and non-parametric k-nearest neighbor (KNN) had been proposed for the same purpose. It extracted the linear combination of features in high dimensional signals and thus reduced classification complexity. However, it may result in a majority of large capacity samples of k neighbors, when they were unbalanced. For instance, the sample size of a class was large, while those of other classes were quite small [19]. In [20], a novel spectrum sensing framework is proposed by the Bayesian machine learning approach. And it introduced Bayesian inference into group sensing data with common spectrum states without prior knowledge of the state number. In [21], [22], a deep learning-based convolutional neural network (CNN) was suggested to merge sensing results in cooperative spectrum sensing scenarios. In these studies, classifier based learning methods exhibited excellent signal detection capability. Simultaneously, a support vector machine (SVM) method was proposed in spectrum detection [23]. Usually, an SVM-based classifier performed better in practice compared with other techniques due its kernel function tricks [24], [25]. In [26], the authors used an SVM and a weighted k-nearest neighbor based technique for spectrum sensing in a cooperative scenario. But the input of the SVM and the w-KNN classifier is based on the original sensing signals rather than the features of sensing signals. And it would increase the amount of calculation. Then a hardware implementation of k-means clustering for spectrum sensing in a cognitive radio system was proposed. It is easy to be implemented in hardware equipment. However, the selection of the initial cluster center had large impact on the clustering results [27]. In this method, unclassifiable problems in an eigenspace were transformed to a high-dimensional space, where classification was feasible by using a linear hyper-plane.

Confronted with the aforementioned spectrum sensing methods on current studies, especially the poor recognition rate under low SNRs and large complexity, we propose an efficient blind spectrum sensing method based on Cholesky decomposition and the radial basis function-support vector machine (RBF-SVM) decision classification at low SNRs. The main contributions of the proposed scheme are summarized as follows.

- **The actual decision threshold possesses self-learning ability by the SVM technique.**
  The proposed spectrum detection method is suited for various signal detections without prior knowledge of signals, channels and noise power. The actual decision threshold possesses self-learning ability based on the SVM, which distinguishes signals from noises effectively.
- **The RBF kernel function maps a nonlinear space to a linear space by non-linear transform and it efficiently works under mixed situations of many PU signals and large background noises.**
  At low SNRs, the proposed scheme can achieve satisfactory detection, because the optimal decision boundary established by the SVM can maximize the margin between the separating hyper-plane and received samples by minimizing the upper bound of Vapnik-Chervonenkis dimension [28]. Different from a linear kernel function, a radial basis function (RBF) kernel is applied to map a nonlinear space to a linear space by non-linear transform. Thus, it can efficiently work under mixed situations with many PUs and large background noises.
- **Collaborative spectrum sensing with multiple antennas of the SUs.**
  A new statistic vector is proposed with the spectrum sensing information of all SUs and it is constructed by the ratio of the maximum and the minimum

eigenvalues of Cholesky decomposition of the sample covariance matrix. But the method of labeling the above vector is just related to the spectrum sensing situation of most SUs. Therefore, it can effectively achieve collaborative spectrum sensing with the antennas of these SUs.

- **Low complexity by the training and the testing procedures in the proposed scheme.**
  The proposed scheme mainly includes the statistic construction, the training and the testing procedures. An original SVM scheme only includes the latter two ones. But the complexity of the former one is negligible. The size of the training and testing sample sets in the proposed scheme remarkably decreases by the statistic construction , when compared with that in an original SVM scheme. Therefore, much lower complexity is obtained in the training and testing procedures of the proposed scheme.

The remainder of this paper is organized as follows. Section II introduces a system model of spectrum sensing in a CR system. Section III presents a new spectrum sensing scheme based on the covariance matrix Cholesky decomposition and the RBF-SVM decision classification on the basis of the aforementioned model to improve detection performance and reduce complexity. Section IV compares the complexity of the proposed scheme with those of other existing schemes. Subsequently, Section V presents the numerical simulations and result analyses to verify good performance of the proposed scheme, especially with an RBF kernel function, by conducting comparisons and analyses with other existing detection algorithms. Section VI provides the summary in this study.

II. SYSTEM MODEL OF SPECTRUM SENSING IN CR SYSTEMS

A. SYSTEM MODEL OF SPECTRUM SENSING

Suppose that a pair of a primary user transmitter (PU-T), a PU receiver (PU-R) and some secondary users (SUs) belong to different communication terminals, and there is not direct information between any two of them. So the SUs cannot obtain channel status information from the PU-T. Then, a typical cognitive radio network (CRN) system model is shown in Fig. 1. A cognitive base station (CBS) first detects PU signals in the detection channel at all divided frequency bands. Then, the CBS transmits the status of a PU-R and determines the idle spectrum. When the PU-R is in the detection area, real-time detection continues until it leaves the detection area and releases its previously occupied spectrum. Thus, the SU can reuse this free spectrum band. It can avoid interference to the PU-R without transmission when it senses at least one PU-R in detection areas. The SU must release and transfer the spectrum to a buffer and enable access for the PU-R. Cognitive devices simultaneously detect other free spectrum for SU utilization.

Suppose that an SU has a multi-antenna device with M antennas in a CRN, and a PU is located a little far away from the SU, but still within the detection area of constant channel gain. Then, the modeling of spectrum sensing as a binary hypothesis testing problem under Neyman-Pearson and Bayes tests is expressed as

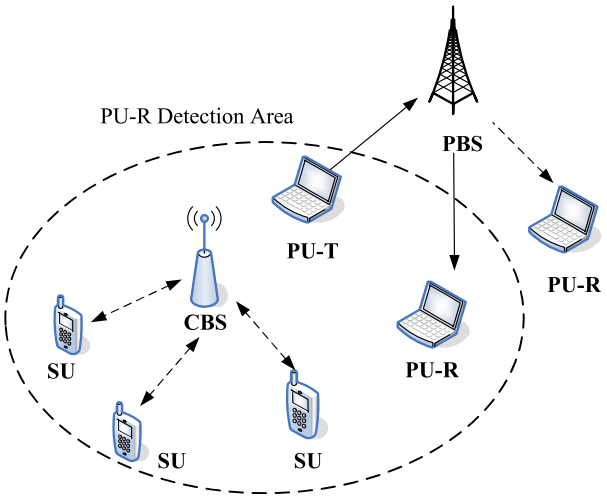


FIGURE 1. Typical cognitive radio network architecture.

$$\begin{cases} H_0 : x_m(n) = \eta_m(n) \\ H_1 : x_m(n) = h_m(n) * s_m(k) + \eta_m(n), \end{cases} \tag{1}$$

where $s_m(n)$ and $\eta_m(n)$ $(n = 1, 2, \cdots, N, m = 1, 2, \cdots, M)$ represent the received PU signals and the additive white Gaussian noises (AWGNs) with zero mean and variance σ_η^2 at the $k-$th sample and the $m-$th antenna in an SU, respectively. N is the total number of sensing samples in one observation obtained in a sensing slot. $x_m(n)$ indicates the signals received by an SU. $h_m(n)$ is the channel gain. Operator "*" denotes the convolutional calculation. H_1 and H_0 denote the hypothesis of the presence of a PU or not, respectively. By the collected samples from M antennas of an SU, a sample matrix is listed as

$$X = \begin{bmatrix} x_1(1) & x_1(2) & \cdots & x_1(N) \\ x_2(1) & x_2(2) & \cdots & x_2(N) \\ \vdots & \vdots & \ddots & \vdots \\ x_M(1) & x_M(2) & \cdots & x_M(N) \end{bmatrix}. \tag{2}$$

The performance of spectrum sensing is mainly evaluated by three following items, namely, the detection probability(P_d), the false alarm probability (P_f), and the missed alarm probability (P_m). They are expressed as

$$\begin{cases} P_d = P\ (D_1 | H_1) \\ P_f = P\ (D_1 | H_0) \\ P_m = P\ (D_0 | H_1), \end{cases} \tag{3}$$

where D_1 and D_0 represent the existence of a PU or not determined by SU detectors.

The definitions of the aforementioned three items are listed as follows:

- P_d: when a PU is present, the SU detector determines the probability of its presence.
- P_f: when a PU is not present, the SU detector determines the probability of its presence.
- P_m: when a PU is present, the SU detector determines the probability of its none existence.

Suppose that $P(H_1)$ and $P(H_0)$ are the *prior* probability of the existence of a PU or not, respectively. And they are all initially set as 0.5. The average error rate (P_e) is calculated by P_f and P_m, and it is multiplied by its average *prior* probability, respectively. Then, P_e is expressed as

$$\begin{aligned} P_e &= P(H_0)P_f + P(H_1)P_m \\ &= P(H_0)P(D_1/H_0) + P(H_1)P(D_0/H_1). \end{aligned} \tag{4}$$

B. PRIOR WORKS

- *MME Method*

The statistical covariance matrices, *i.e.*, R_x, of received signals under the above two hypotheses when $N \to \infty$ are different and they are jointly expressed as

$$R_x = E(XX^T) = \begin{cases} \sigma_\eta^2 I_M, & H_0 \\ R_s + \sigma_\eta^2 I_M, & H_1, \end{cases} \tag{5}$$

where $E(\cdot)$ denotes the expectation operator and $(\cdot)^T$ represents the transpose operator. I_M denotes the identity matrix of order M and R_s represents the covariance matrix of PU signals.

Let $\tilde{R}_x$ be the normalized covariance matrix, defined as $\tilde{R}_x = (N/\sigma_\eta^2)R_x$. $\tilde{R}_x$ is in proportional to R_x and $\tilde{R}_x$ is a Wishart matrix under the hypothesis H_0. λ_1 and λ_M denote the maximum and minimum eigenvalue of $\tilde{R}_x$. The maximum-minimum eigenvalue (MME) algorithm is related to the ratio of the maximum to the minimum eigenvalue under alternate hypothesis. The test statistic of the MME detection is defined as $T_{MME} = \frac{\lambda_1}{\lambda_M}$. According to the analysis of MME detection, under the assumption $\lim_{N\to\infty} \frac{M}{N} = a$ (a is a constant with range of $0 < a < 1$), the approximate expressions for P_f and the decision threshold γ_{MME} are given in [11] as

$$\gamma_{MME} = \frac{(\sqrt{N}+\sqrt{M})^2}{(\sqrt{N}-\sqrt{M})^2}\left[1+\frac{(\sqrt{N}+\sqrt{M})^{-2/3}f^{-1}(1-P_f)}{(NM)^{-1/6}}\right], \tag{6}$$

where F_1 is the inverse function of F_1^{-1}, *i.e.*, the cumulative distribution function (CDF) of the Tracy-Widom distribution with order 1 [12]. u and v are given as follows.

$$\begin{cases} u = (\sqrt{N-1}+\sqrt{M})^2 \\ v = (\sqrt{N-1}+\sqrt{M})\left(\dfrac{1}{\sqrt{N-1}}+\dfrac{1}{\sqrt{M}}\right)^{1/3}. \end{cases} \tag{7}$$

- *EME Method*

The test statistic of energy with minimum eigenvalue (EME) algorithm is the ratio of the average energy of received signals to the minimum eigenvalue of the covariance matrix from received signal samples and it is expressed as

$$T_{EME} = \frac{1/MN(\sum_{i=1}^{M}\sum_{k=1}^{N}|x_i^2(k)|)}{\lambda_M}. \tag{8}$$

The corresponding threshold is given in [13] as

$$\gamma_{EME} = (\sqrt{\frac{2}{MN}}Q^{-1}(P_f)+1)\frac{N}{(\sqrt{N}-\sqrt{M})^2}, \tag{9}$$

where $Q^{-1}(\cdot)$ is the inverse function of $Q(\cdot)$ function and it is expressed as

$$Q(x) = \int_x^{+\infty} \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}t^2)dt. \tag{10}$$

- *CAV Method*

The covariance absolute value (CAV) algorithm is constructed according to the difference of the received signal covariance matrix given PU presence or not. The CAV detection statistic is expressed as

$$T_{CAV} = T_1/T_2, \tag{11}$$

where T_1 and T_2 are the sums of the absolute values of all diagonal elements of $\tilde{R}_x$, respectively. They are expressed as

$$T_1 = \frac{1}{M}\sum_{i=1}^{M}\sum_{j=1}^{M}|R_{i,j}|, \tag{12}$$

$$T_2 = \frac{1}{M}\sum_{i=1}^{M}|R_{i,i}|, \tag{13}$$

where $R_{i,j}$ and $R_{i,i}$ are the elements of $\tilde{R}_x$ with coordinate (i, j) and (i, i), respectively.

And the CAV detection threshold is given in [13] as

$$\gamma_{CAV} = \frac{1+(M-1)\sqrt{2/N\pi}}{1+Q^{-1}(P_f)\sqrt{2/N}}. \tag{14}$$

.

III. BLIND SPECTRUM SENSING BY CHOLESKY DECOMPOSITION AND RBF-SVM

In this section, an efficient CR spectrum sensing scheme with statistic construction by covariance matrix Cholesky decomposition and RBF-SVM is established. It includes the blind detection with the covariance Cholesky factorization (CCF) and the SVM. Then the flow diagram of the scheme is proposed and analyzed as follows.

A. BLIND DETECTION BY COVARIANCE CHOLESKY FACTORIZATION

Since $\tilde{R}_x$ is a non-negative defined symmetric matrix, there exists a unique Cholesky decomposition of the matrix. Let U denote the matrix obtained by Cholesky decomposition

of $\tilde{R}_x$. Then, based on Cholesky decomposition theorem, U is expressed as

$$\tilde{R}_x = U \cdot U^T, \tag{15}$$

According to the above theorem, the matrix U by the CCF under the above two hypotheses are also different. The matrix U obtained by Cholesky decomposition under the hypothesis H_0 and $N \to \infty$ is a diagonal matrix with the same diagonal elements. The eigenvalues of a diagonal matrix are the diagonal elements. But U obtained by Cholesky decomposition under the hypothesis H_1 and $N \to \infty$ is a lower triangular matrix with different diagonal elements, which is given by

$$U = \begin{bmatrix} u_{11} & 0 & \cdots & 0 \\ u_{21} & u_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ u_{M1} & u_{M2} & \cdots & u_{MM} \end{bmatrix}, \tag{16}$$

where $u_{i,j} \geq 0$ $(i \geq j)$ is calculated by

$$u_{i,i} = \sqrt{(\tilde{R}_{i,j} - \sum_{k=1}^{i-1} u_{i,k}^2)}, \quad if\ i = j, \tag{17}$$

$$u_{i,j} = \frac{\tilde{R}_{i,j} - \sum_{k=1}^{j-1} u_{i,k} \cdot u_{j,k}}{u_{j,j}} \quad if\ i > j, \tag{18}$$

where $\tilde{R}_{i,j}$ is the (i, j)-th element of $\tilde{R}_x$.

Let U_k be the k-th matrix U in the SUs, where $k = 1, 2, \cdots, K$ and K is the number of the SU. λ_i^k represents the descending ordered eigenvalues of U_k, where $i = 1, 2, \cdots, M$, namely, $\lambda_1^k \geq \lambda_2^k \geq \cdots \geq \lambda_M^k$. The diagonal elements of U_k are the eigenvalues of U_k and they are not ordered. Similar to the maximum minimum eigenvalue (MME) scheme, in our scheme, the statistic T_k is established by using the ratio of the maximum to the minimum eigenvalue of the matrix U_k obtained by the covariance matrix Cholesky decomposition. Thus, the test statistic and detection threshold under the two hypotheses are expressed as

$$T_k = \frac{\lambda_1^k}{\lambda_M^k} \underset{H_0}{\overset{H_1}{\gtrless}} 1, \quad k = 1, 2, \cdots, K, \tag{19}$$

where T_k is greater or equal to 1 and the statistic vector is expressed as $T = [T_1, T_2, \cdots, T_K]$.

In practice, the statistical covariance matrix $\tilde{R}_x$ can only be estimated by sample covariance matrix R_N that is defined as $R_N = 1/N(XX^T)$ and the decision threshold obtained by the following SVM model, which will be discussed in the next subsection.

B. CLASSIFICATION OF PUS AND NOISES BY THE SVM

The proposed scheme mainly relies on the statistic vector T to solve the optimal decision boundary. Decision function $f(T)$ is generated from T by mapping the kernel function and then solving the optimal equations. Decisions and classifications are rapidly executed by the established decision functions, as the testing signals are received by statistic construction.

As shown in Fig. 2, the circles filled with white and black color represent two groups of training samples. The classification interval determined by the hyper-plane is $2/\|w\|$. The samples on two broken lines are support vectors. The interval hyper-plane changes with the change of samples.

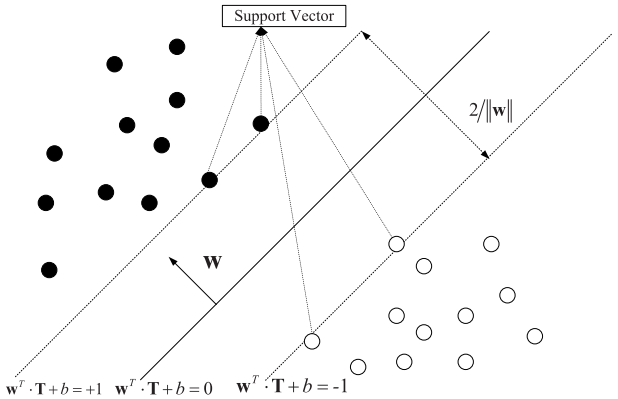


FIGURE 2. Maximum interval of a typical hyper-plane.

The training samples are represented as $G = \{(T_i, f_i)|i = 1, 2, \cdots, L\}$. T_i represents an input data vector and f_i denotes that T_i belong to a certain class of two labels as "+1" or "−1". The decision result T_i is f_i. The statement $f_i \in \{-1, +1\}$ is the binary class label with hypothesis H_0 or H_1, respectively. L is the number of training samples. The nonlinear SVM achieves classification with the optimal separation hyper-plane. Then the optimal hyperplane and the classification decision function are represented respectively as

$$w \cdot \phi(T) + b = 0, \tag{20}$$

$$f(T) = sign(w \cdot \phi(T) + b), \tag{21}$$

where w and b are respectively the weighting vector and the bias. $\phi(T)$ is mapping function and it maps T into a high dimensional space. Symbolic function sign(x) is defined as $sign(x) = +1$ when $x > 0$, $sign(x) = -1$ when $x < 0$ and $sign(x) = 0$ when $x = 0$.

To alleviate the over fitting in high-dimension, soft margin has been introduced in [29]. A new slack variable is increased for margin adjustment. Then, the optimal hyperplane is present as

$$\min_w \frac{1}{2}\|w\|^2 + C\left(\sum_{i=1}^{L} \xi_i\right), \tag{22}$$

$$st: f_i[(w \cdot \phi(T_i))) + b] \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad \forall i \geq 0, \quad i = 1, 2, \cdots L, \tag{23}$$

where ξ_i is the slack variable and C is the penalty factor.

The optimization of (22) and (23) can be solved by a standard Lagrangian multiplier method. New parameters of α_i and β_i are Lagrangian multipliers, which satisfy the relationship of $\alpha_i > 0$ and $\beta_i > 0$. So the objective function turns

into

$$L_a(w, b, \alpha, \beta) = \|w\|^2 \Big/ 2 + C\left(\sum_{i=1}^{L} \xi_i\right) - \sum_{i=1}^{L} \alpha_i\{f_i[(w \cdot \phi(T_i)) + b] - 1 + \xi_i\} - \sum_{i=1}^{L} \beta_i \xi_i. \quad (24)$$

Take partial derivative of $L_a(w, b, \alpha, \beta)$ with independent variable w and b, and then set the results as 0, respectively, there are

$$\begin{cases} \dfrac{\partial L_a(w, b, \alpha, \beta)}{\partial(w)} = 0 \quad \Rightarrow w = \sum_{i=1}^{L} \alpha_i f_i \phi(T_i) \\ \dfrac{\partial L_a(w, b, \alpha, \beta)}{\partial(b)} = 0 \quad \Rightarrow \sum_{i=1}^{L} \alpha_i f_i = 0. \end{cases} \quad (25)$$

After deriving the above Lagrangian function and then bringing it into (22) and (23), the original optimization problem is converted into its dual problems [26]. Then the optimal hyperplane can be modeled as

$$\max_{\alpha} \sum_{i=1}^{L} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{L} \alpha_i \alpha_j f_i f_j K(T_i, T_j), \quad (26)$$

$$s.t.: \sum_{i=1}^{L} f_i \alpha_i = 0, \quad 0 \le \alpha_i \le C, \ i = 1, 2, \cdots L, \quad (27)$$

where $K(T_i, T_j) = \phi(T_i) \cdot \phi(T_j)$ is the kernel function which represents a legitimate inner product in Eigen space.

After solving (26) and (27), the final classification function is given as

$$f(T) = sign[\sum_{i=1}^{L} \alpha_i f_i K(T, T_i) + b], \quad (28)$$

where $f(T)$ determines the existence of PUs based on detected samples. $f(T) = +1$ represents H_1 with the existence of a PU. Otherwise, the PU does not exist.

In classification, it is difficult for a linear hyperplane to distinguish between the two classes. Then, a kernel function $K(T_i, T_j)$ is used to transform the data from input space to a higher dimensional space. So it is possible for a linear hyperplane to differentiate the data elements of the two classes. Then three kernel functions are utilized and given as

$$K(T_i, T_j) = T_i \cdot T_j^T. \quad (29)$$

$$K(T_i, T_j) = \left[(T_i, T_j) + p\right]^q. \quad (30)$$

$$K(T_i, T_j) = \mathrm{e}^{-\|T_i - T_j\|_2^2 / 2\sigma^2}. \quad (31)$$

Let $\theta = -1/2\sigma^2$ and then the theoretical analysis about the preference of the RBF kernel function can be expresses as

$$K(T_i, T_j) = \mathrm{e}^{\theta \|T_i - T_j\|_2^2} = \mathrm{e}^{\theta(\|T_i\|_2^2 + \|T_j\|_2^2)} \mathrm{e}^{2\theta T_i T_j}, \quad (32)$$

where $\mathrm{e}^{2\theta T_i T_j} = 2\theta(\sum_{i=0}^{n} \frac{(T_i T_j)^n}{n!} + o((T_i T_j)^n))$ according to Taylor Expansion. It can ignore the remainder $o((T_i T_j)^n)$ when $n \to +\infty$. Thus, (32) can be represented as

$$K(T_i, T_j) = \sum_{i=0}^{n} \sqrt{2\theta / n!}\, T_i^n e^{\theta \|T_i\|_2^2} \sum_{j=0}^{n} \sqrt{2\theta / n!}\, e^{\theta \|T_j\|_2^2}. \quad (33)$$

Kernel function can be converted into an inner product form of two vectors, namely, $K(T_i, T_j) = \phi(T_i) \cdot \phi(T_j)$. Thus, the mapping function $\phi(T_i)$ is present as

$$\phi(T_i) = e^{\theta \|T_i\|_2^2} \left[\sqrt{2\theta / 0!}\, T_i^0, \sqrt{2\theta / 1!}\, T_i^1, \cdots, \sqrt{2\theta / n!}\, T_i^n\right]^T. \quad (34)$$

Thus, the RBF kernel function maps non-linear spaces to high-dimensional linearly separable spaces through nonlinear mapping function (34). $\phi(T_i)$ satisfies the following relationship as

$$\|\phi(T_i)\|_2^2 = < \phi(T_i), \phi(T_i) > = K(T_i, T_i) = 1. \quad (35)$$

It indicates that $\phi(T_i)$ implements the normalization of the kernel value after T_i is mapped to a high-dimensional space.

C. FLOW DIAGRAM OF THE PROPOSED SPECTRUM DETECTION

The proposed scheme includes training and testing procedures. In the training procedure, it includes the construction of training vector T and the corresponding label. The construction of the training set with bind spectrum sensing is shown in Fig. 3 and it is further depicted in Tab. 1.

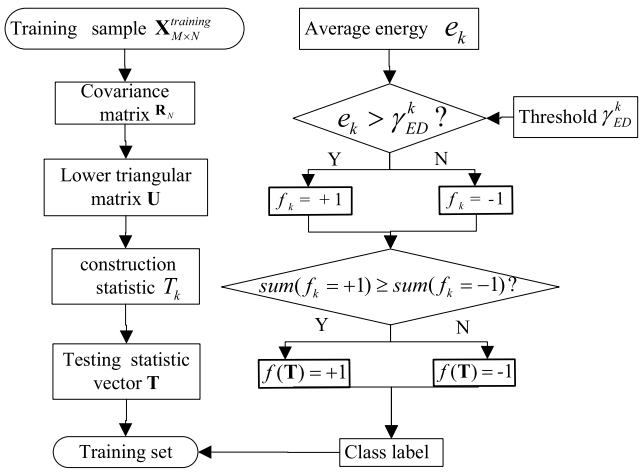


FIGURE 3. Flow diagram of the construction of a training set.

The statistic construction of the SVM training sets is a key factor in spectrum sensing. It is the ratio of the maximum to the minimum eigenvalue of matrix U obtained by covariance matrix Cholesky decomposition. The statistic vectors T and the corresponding labels of "+1" and "−1" are trained to generate an optimized SVM classification model, which determines the presences of the PUs in the testing process.

TABLE 1. The procedures of the proposed scheme.

Procedures of the training set construction with bind spectrum sensing.
Step 1). First, training sensing signal matrix $X_{M\times N}$ is consisted with sensing signal $x_m(n)$. Second, the sample covariance matrix R_N is constructed with $X_{M\times N}$ and the matrix U is obtained by the Cholesky decomposition of R_N. The maximum and the minimum eigenvalues of U_k are then constructed as the statistic $T_k = \lambda_1^k/\lambda_M^k$ in (19). Then, goto step 2).
Step 2). First, the sample label is set as $f_k = +1$, where the average energy of the sensing signal is greater than the detection threshold. Otherwise, it is set as $f_k = -1$, where $k = 1, 2, \cdots, K$ and K is the number of the SUs. The average energy of the sensing signals and the energy detection threshold are calculated as $e = \frac{1}{MN}(\sum_{m=1}^{M}\sum_{n=1}^{N}|x_m(n)|^2)$ and $\gamma_{ED} = \sigma_n^2(2 + \sqrt{2/MN}Q^{-1}(P_f))$, respectively [3]. Then, goto step 3).
Step 3). The total number of $f_k = +1$ and $f_k = -1$ are calculated, respectively. The statistic vector T is set to $f(T) = +1$, where the number of $f_k = +1$ is greater than that of $f_k = -1$. Otherwise, it is set as $f(T) = +1$. Then, goto step 4).
Step 4). The training set consists of the statistic vector T and the labels is trained to generate a classifier model.

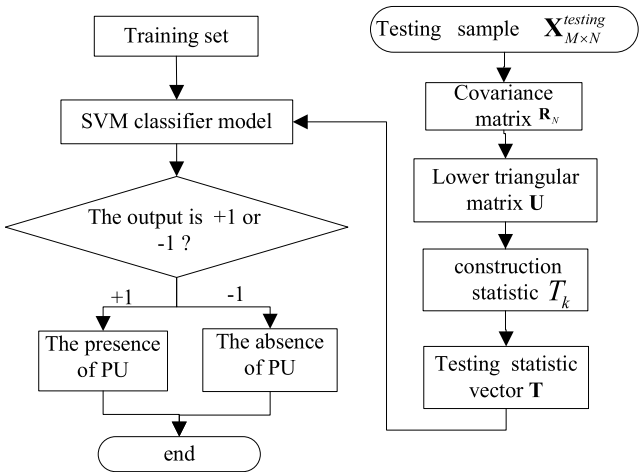


FIGURE 4. Flow diagram of the proposed scheme.

Namely, the output is "+1" or "−1" corresponding to the spectrum occupied by PUs or not. Finally, the flow diagram of the proposed scheme is shown in Fig. 4.

IV. COMPUTATIONAL COMPLEXITY ANALYSES OF THE PROPOSED SPECTRUM SENSING SCHEME

In this section, the computational complexity of the proposed spectrum sensing scheme is analyzed in the following two aspects and it is represented as follows.

First, in the construction of the test statistic, the complexity of the proposed scheme mainly lies in the following two aspects. One is the computation of the covariance matrix, where $M(M+1)N/2$ multiplications and $M(M+1)(N-1)/2$ additions are required. Here, M, N are the number of antennas in an SU and the total number of sensing samples in one observation obtained in a sensing slot as described in Subsection II(A). Another is the Cholesky decomposition of the covariance matrix. Since the MME, the EME and the CAV schemes are also required to compute the covariance matrix, the computations of the covariance matrix are all the same among them. So the complexity difference mainly lies in the latter, *i.e.*, the eigenvalue decomposition of the covariance matrix. In this work, the Gaxpy algorithm [30] is used to perform the Cholesky decomposition. The complexity of the Gaxpy algorithm is $o(M^3/3)$, while the complexity of the eigenvalue decomposition of the covariance matrix by the QR decomposition in the MME, and the EME schemes are $o(M^3)$ [30]. Then the complexity of the proposed scheme is much lower than that of the MME one. The CAV scheme needs M^2 multiplications and $(M+2)(M-1)$ additions. The ED and EME need MN multiplications and $M(N-1)$ additions. In conclusion, the computational complexity of the proposed statistic construction is compared with those of the MME and the ED statistic construction in Tab. 2.

TABLE 2. Complexity comparison among the detection schemes.

Scheme Name	Total Computational Complexity
ED Scheme	$M(2N-1)$
MME Scheme	$M(M+1)(2N-1)/2 + o(M^3)$
EME Scheme	$M(M+1)(2N-1)/2 + M(2N-1) + o(M^3)$
CAV Scheme	$M(M+1)(2N-1)/2 + M^2 + (M+2)(M-1)$
Proposed Scheme	$M(M+1)(2N-1)/2 + o(M^3/3)$

Second, the input of the classifier is given in form of test statistics in (19), which is obtained by Cholesky decomposition over sample covariance matrix of sensing signals rather than the input of original sensing signals. In addition, the value of the test statistic is approximately 1, which reduces the calculations in the training procedure of the SVM classification model. Suppose that L and L_1 are the number of training and testing samples, respectively. The size of training sample set is $L \cdot K \cdot M \cdot N$, and the size of testing sample set is $L_1 \cdot K \cdot M \cdot N$. The size of training and testing sample sets are then reduced to $L \cdot K$ and $L_1 \cdot K$, respectively, by both the Cholesky decomposition of the covariance matrix and the construction of the test statistics. Therefore, the size of training and testing sample sets in proposed scheme are remarkably decreased, when compared with that of the traditional SVM spectrum sensing scheme in [23].

V. NUMERICAL SIMULATIONS AND RESULT ANALYSES

In numerical simulations, ideal historical training samples are required in the implementation of the proposed spectrum sensing scheme. The generation of a good SVM classifier with ideal historical data can guarantee the detection performance. In this study, statistic vector T and corresponding labels are used as training samples. They can be obtained on the basis of simulations of an ideal wireless environment by using a reliable perception algorithm. In this experiment, wireless signals are simulated with PUs in BPSK modulation. Then, the cosine carriers are multiplied by the baseband signals of the PUs to generate the binary phase shift keying (BPSK) modulated transmission signals. Given that the

simulations are performed in an ideal AWGN channel with zero mean $\mu = 0$ and variance $\sigma_\eta^2 = 1$, the channel gains for all PU signals are assumed as $h = 1$. The *prior* probability $P(H_0)$ and $P(H_1)$ that represent the presence of at least a PU or not are all set as 0.5. The P_f in the next simulation is set as 0.01. The bandwidth is set as $W = 1.5 \times 10^4$, and the sampling frequency is $F_s = 2W = 3 \times 10^4$. The sampling time is 0.01s (*i.e.*, $t = 0.01s$). The number of antennas of an SU is 10 (*i.e.*, $M = 10$). And the number of SUs is 10 (*i.e.*, $K = 10$). The simulations in each SNR are performed at least for 1 thousand times to fulfill the requirement of Monte Carluo experiments.

A. KERNEL FUNCTION PERFORMANCE AND ANALYSES

In the SVM, two groups of samples are easily separated with a hyperplane. However, the aliased PUs and noises are difficult to be classified in practice. Local classification charts are simulated and listed with linear, polynomial, and RBF kernel functions. The detection probability of the energy detection is nearly up to 100% at 10 dB. Thus, the method of labels is simulated at 10 dB. The parameter of the RBF kernel function is set as $\sigma = 2$ in (31). The parameter of polynomial kernel function is set as $p = 0, q = 2$ in (30). Penalty factor C is set as "$+\infty$" in (22). As shown in Figs. 5, 6, and 7, PUs and noises are classified based on training and testing conditions. They are obtained by inputting the statistic vector T into the SVM classification model. The curve in the figure refers to the hyper-plane, which can distinguish PUs from noises.

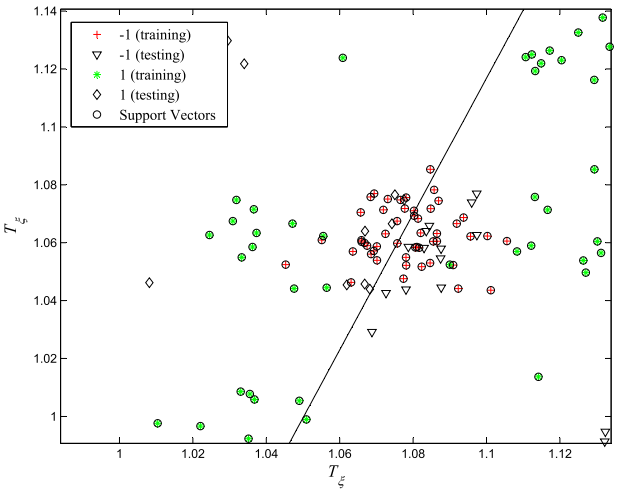


FIGURE 5. Plot of test data classification by using a linear kernel function.

Fig. 5 shows the local classification chart of a sample set by using a linear kernel function. A linear hyper-plane is observed and support vectors that related to the SVM classifier model are labeled with all training samples. But the support vectors have approximately more than half of the errors on training samples. The reasons can be explained as follows. The wireless channel includes PU signals, noises and environmental interferences which belong to the nonlinear space. So the linear kernel function is only suitable for

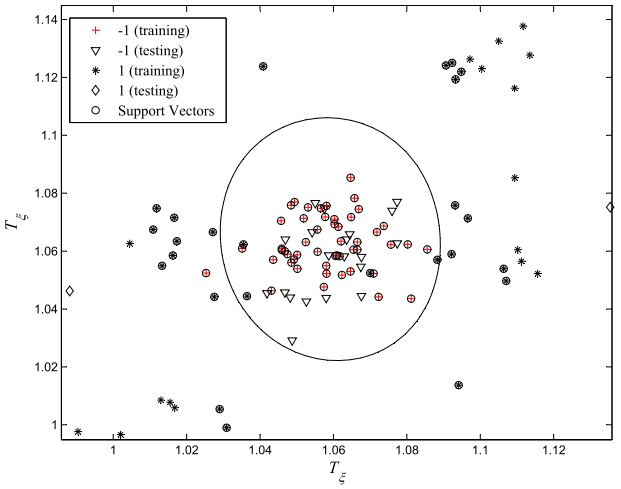


FIGURE 6. Plot of test data classification by using the polynomial kernel function.

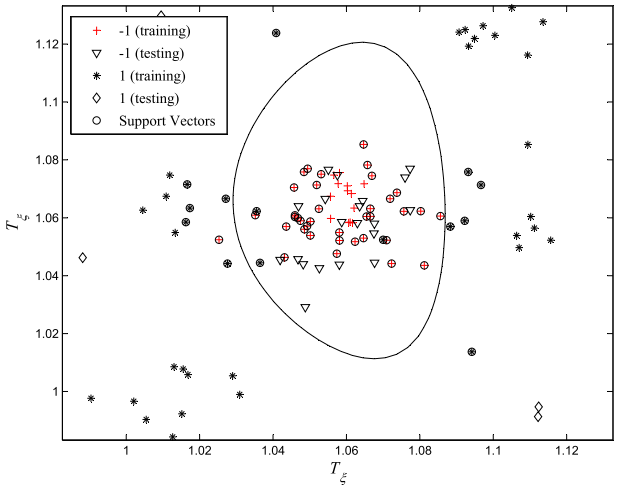


FIGURE 7. Plot of the test data classification by using the RBF kernel function.

linear spaces. Thus, the separation of PUs and noises cannot be solved on such occasions.

Fig. 6 shows the local classification of a sample set by using the polynomial kernel function. Compared with the linear hyper-plane of the linear kernel function, the hyper-plane can easily separate PUs from noises. The polynomial kernel functions are less than those of the linear kernel functions in term of the number of support vectors with five errors on the training samples.

As shown in Fig. 7, the RBF kernel function needs less support vectors to successfully classify the testing samples compared with those of contrast polynomial and linear kernel-based schemes. In this case, support vectors have four errors on the training samples. Compared with the polynomial kernel function, the RBF successfully classifies more testing samples.

The performance of the classifier is evaluated by the accuracy, *i.e.*, the ratio of correct classifications to the number

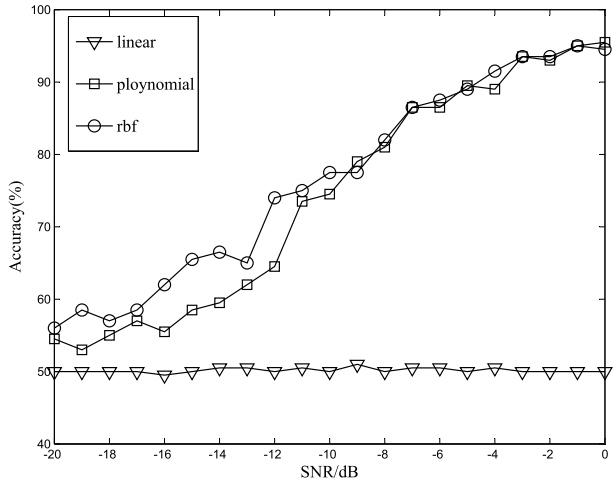


FIGURE 8. Comparison of the SVM-based classifiers trained with linear, polynomial and RBF kernel functions in the proposed scheme.

of total test observations. Fig. 8 shows the accuracy of the three aforementioned kernel functions based on different SNRs. The number of training and testing samples are all set as 1000. The classifier trained by either the RBF or the polynomial kernel function always outperforms the linear kernel function, and it attains accuracy about 50%, irrespective of SNRs. The former two use one of the two classifiers of either the RBF or the polynomial kernel function of the same increased accuracy along with the growth in SNRs. However, the RBF kernel value avoids large amount of computations, because it is approximately 1 or 0, when $\|T_i - T_j\|_2 \approx 0$ or $\|T_i - T_j\|_2 >> 0$ in (31), respectively. Thus, the RBF can be efficiently used as a kernel function in the proposed scheme.

According to the above simulations, the number of training samples should be equilibrium, since they remarkably affects the training efficiency. Excessive samples lead to long learning time and deteriorate the training efficiency. Few training samples result in insufficient learning. Thus, the studies on the number in sample training are conducted as follows. As shown in Fig. 9, the number of training sample is set from 300 to 2700. So the training sample number in testing is set as 1000.

The simulation results of average error rate are obtained by using the above parameters combined with the RBF kernel function in the SVM, as shown in Fig. 9. In Fig. 9, P_e decreases with the increased SNRs and the increased number of training samples. P_e is less than 0.1 at 0 dB. With 1500 training samples, the values of P_e are 0.52, 0.47, 0.35, 0.09, and 0.02, when the SNRs are set as -20dB, -15dB, -10dB, -5dB and 0dB, respectively. The increase in training samples leads to much larger complexity. However, a few training samples lead to an increase in P_e. When the number of training samples is increased to 1500, P_e is approximately stable. Therefore, the number of training samples is optimally adopted as 1500 in successive simulations.

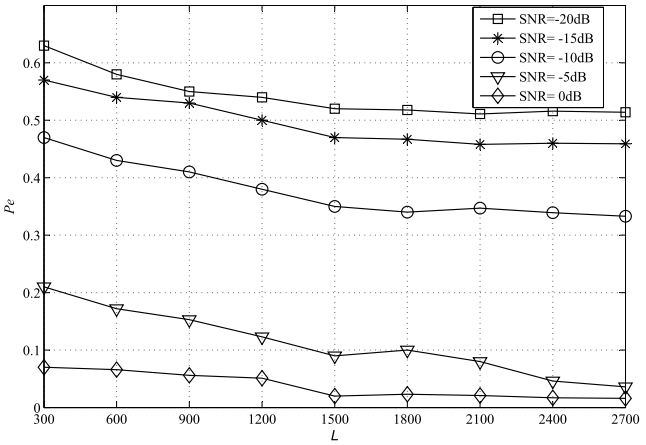


FIGURE 9. Relationships of $P_e \sim L$ with SVM training data at different SNRs.

B. PERFORMANCE COMPARISONS OF THE SPECTRUM SENSING SCHEMES

Next, in the ED [3], the MME [13], the EME [6], the CAV [7] and the SVM [23] schemes, the spectrum sensing schemes are compared in term of P_e under different SNRs with different N. For the SVM and the proposed scheme, the number of testing sample are all selected as 1500. The SNR ranges from -20 dB to 0 dB. P_e is obtained as in Fig. 10 through numerical simulations.

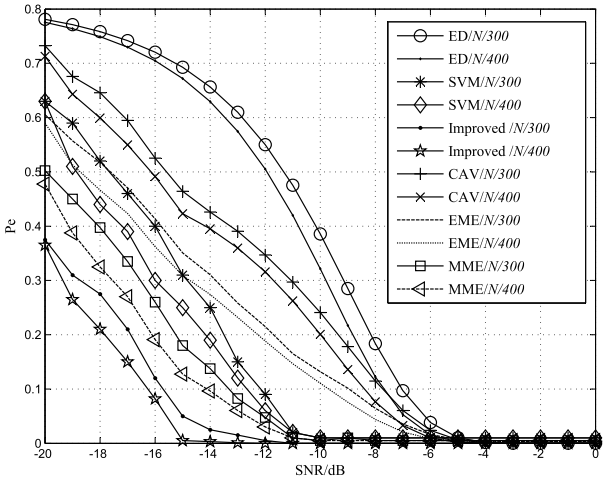


FIGURE 10. Comparisons among various spectrum sensing schemes in terms of average error rate P_e at different SNRs.

As shown in Fig. 10, P_e decreases with increased SNRs in various spectrum sensing schemes. The proposed spectrum sensing obtains apparently lower P_e than the energy detection, the SVM detection and other blind spectrum sensing schemes with the same N. With the increase of N from 300 to 400, P_e of all schemes are reduced. The MME scheme has lower P_e compared with other blind spectrum sensing schemes, including the CAV and EME schemes. Compared with blind spectrum sensing, the proposed scheme declines rapidly in term of P_e. And it achieves a perfect performance

when $N = 400$. Even at low SNR of -15dB, P_e reaches 0.365. This phenomenon can be explained by the following reasons. First, at low SNRs, the capability of the energy detection is suppressed, and PU signals are submerged into large background noises. At this situation, the SVM scheme has much better performance, because the optimal decision boundary established by the SVM maximizes the margin between the separated hyper-plane and the received data. Second, compared with the SVM and other blind spectrum sensing schemes, the superiority of the proposed scheme lies in features extracted from Cholesky decomposition covering all key information and the decision threshold is based on the SVM self-learning, which can distinguish the signals from noises effectively.

The receiver operating characteristic (ROC) curve is another effective tool to evaluate the performance of a detection scheme. Fig. 11 exhibits the ROC situations between the energy detection [3] and the machine learning schemes, including k-means [17], the KNN [19] and the SVM schemes.

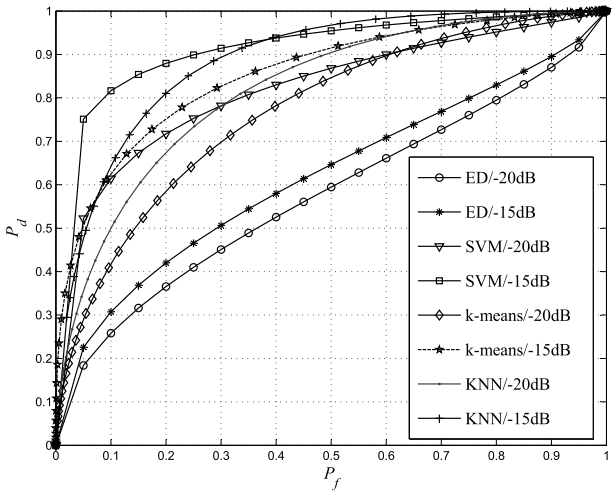


FIGURE 11. ROC curves of machine learning schemes and an energy detection one.

At each SNR, the relationship of $P_d \sim P_f$ is obtained and shown in Fig. 11, where the detection probability P_d increases with the increased false alarm probability P_f. Given the same false alarm probability, large SNR leads to better detection probability. It shows that the SNR becomes larger, which makes the signal energy larger and thus it improves the detection performance. The three machine learning methods perform better than the energy detection in term of ROC at -15 dB and -20 dB. Specifically, the SVM classifier performs better than k-means classifier in the false alarm probability region of 0.1-0.6 at -20 dB and it also performs better than KNN classifier in the false alarm probability region of 0.1-0.3 at -20 dB. Further, SVM classifier also performs comparatively better than the k-means classifier in all false alarm probability at -15 dB and it also performs better than the KNN classifier in the false alarm probability region of 0.1-0.4 at -15 dB, approximately. In summary, the KNN method achieves higher performance in comparison to the k-means and the energy detection due to the exploitation of the localized information, whereas the SVM classifier achieves a higher detection probability by mapping feature space to a higher dimension with the help of kernel function. Thus, blind spectrum sensing by the SVM method obtains better detection performance compared with the energy detection and k-means scheme.

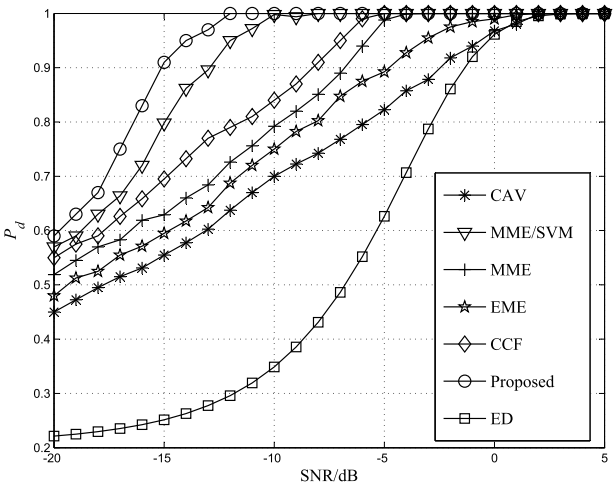


FIGURE 12. Comparison of detection probabilities among the proposed and existing blind detection schemes, and also the energy detection one.

Fig. 12 illustrates the detection probability of the proposed scheme with other contrast blind spectrum sensing schemes. In this figure, the P_d of the proposed scheme and the MME based SVM ones all increase much more rapidly than those of conventional bind spectrum sensing and also the energy detection schemes. This phenomenon can be explained as follows. First, the blind spectrum sensing schemes, including the MME, the EME and the CAV schemes, outperforms the traditional energy detection. Because the thresholds in them are independent of noises. They are only dependent on the parameters, such as N, M and P_f in (6), (9), (14). Thus, they had stable and excellent performance at low SNRs. Second, compared with the asymptotic threshold in (6) of the MME, EME, and CAV detection, the decision threshold in (19) for the CCF detection can be calculated by using an exact non-asymptotic expression without asymptotic assumptions that $\lim_{M,N\to\infty} \frac{M}{N}$ is a constant. Third, the energy detection and the blind detection schemes by the covariance matrix of received signals mainly rely on detection thresholds. Once the detection thresholds are inaccurate, the detection performance is seriously deteriorate. In addition, a classifier in the proposed scheme is constructed to achieve spectrum sensing by history learning data and it has self-learning capability rather than the fixed decision threshold. Therefore, the proposed scheme outperforms the traditional ED, other blind spectrum sensing schemes and the MME based SVM schemes in detection accuracy and efficiency.

Subsequently, to simplify the computational complexity analyses, the normalized sensing time is defined as $\hat{t} = (t_i - t_{min})/(t_{max} - t_{min})$, where $\hat{t}$ is the normalized sensing time, t_{min} and t_{max} represent the minimum and the maximum sensing time, respectively. The sensing time of the proposed scheme includes the time of statistic construction, the training and testing processes. The testing number of the SVM and the proposed schemes are selected as 1500. To compare sensing time, the number of the simulations with energy detection and other blind spectrum sensing are all selected as 1500. At the range of N from 100 to 600, we obtain the changes of the normalized sensing time at -10 dB shown in Fig. 13. The ED [3], MME [13], the EME [6], the CAV [7], and the SVM [23] schemes are compared with our proposed one in term of normalized sensing time in Fig. 13.

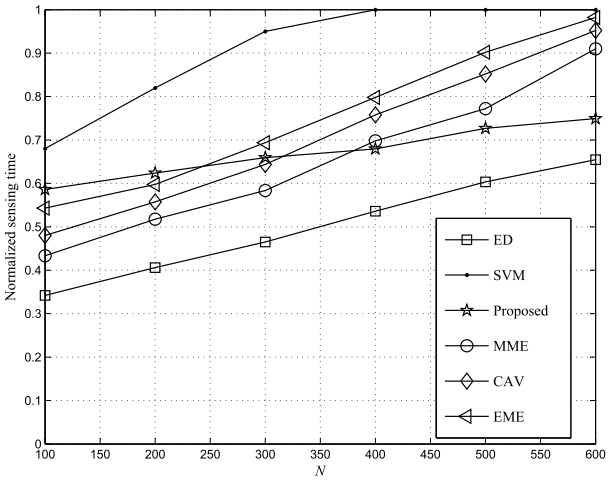


FIGURE 13. Normalized sensing time based on different spectrum sensing schemes.

Fig. 13 shows the normalized sensing time of various spectrum sensing schemes with different dimension N. The normalized sensing time of all schemes increase along with the increase of N. But in our scheme, it has no significant change with the increase of N. Compared with the original SVM spectrum sensing scheme, our scheme has much lower normalized sensing time. The reasons can be explained as follows. The statistics of all schemes are proportional to the signal dimension N shown in Tab. 2. Thus, the normalized sensing time of the statistic construction is related to N. However, compared with the training sample sets of the original SVM scheme in [23], the size of the training sample sets in proposed scheme are remarkably decreased from $L \cdot K \cdot M \cdot N$ to $L \cdot K$, namely, the training sample sets of the proposed scheme are only related to the size of training samples(L) and the number of the SUs (K). And the complexity of the proposed scheme, the SVM, the ED and other MME similar blind spectrum sensing schemes are $o(M(M+1)(2N-1)/2 + o(M^3/3)) + o(L^2 \cdot K^2) + o(L_1^2 \cdot K)$, $o(L^2 \cdot (K \cdot M \cdot N)^2) + o(L_1^2 \cdot (K \cdot M \cdot N))$, $o(L \cdot K \cdot M \cdot (2N-1))$ and $o(L \cdot K \cdot (M(M+1)(2N-1)/2 + M(2N-1) + o(M^3)))$, respectively, where $o(L^2 \cdot K^2)$ and $o(L_1^2 \cdot K)$ are the time complexity of training and testing processes of the proposed scheme, respectively. Thus, compared with the original SVM scheme, the proposed scheme reduces the waiting time for the SUs to find a new channel significantly and thus it decreases communication interruption.

VI. CONCLUSION

In this study, we present an efficient blind spectrum sensing by covariance matrix Cholesky decomposition and RBF-SVM decision classification at low SNRs. The decision is rapidly made by using an established SVM classification model. It has several advantages as follows. First, the proposed scheme can be used for various signal detection occasions without prior knowledge of signals, channels and noise power. The actual decision threshold possesses self-learning ability based on the SVM, which distinguishes signals from noises effectively. Second, at low SNRs, the proposed scheme can achieve satisfactory detection, because the optimal decision boundary established by the SVM maximizes the margin between the separated hyper-plane and received samples. Finally, the proposed scheme outperforms the conventional MME scheme in computational complexity. Simulation results demonstrate that the proposed scheme has better performance than conventional detection ones, especially at low SNRs, which verifies its suitability in 5G communications.

REFERENCES

[1] L. Yu, Q. Wu, and J. Wang, "Optimal adaptive multiband spectrum sensing in cognitive radio networks," *KSII Trans. Internet Inf. Syst.*, vol. 8, no. 3, pp. 984–996, Mar. 2014.

[2] O. P. Awe, A. Deligiannis, and S. Lambotharan, "Spatio-temporal spectrum sensing in cognitive radio networks using beamformer-aided SVM algorithms," *IEEE Access*, vol. 6, pp. 25377–25388, Apr. 2018.

[3] J. Luo, G. Zhang, H. Xu, Y. Zhou, and W. Yang, "An energy detection based signal detection method with multiple antennas in cognitive radio," in *Proc. IEEE Adv. Inf. Technol., Electron. Autom. Control Conf. (IAEAC)*, Chongqing, China, Dec. 2015, pp. 187–191.

[4] A. Surampudi and K. Kalimuthu, "An adaptive decision threshold scheme for the matched filter method of spectrum sensing in cognitive radio using artificial neural networks," in *Proc. 1st India Int. Conf. Inf. Process. (IICIP)*, Delhi, India, Aug. 2016, pp. 1–5.

[5] J. Du, H. Huang, X. J. Jing, and X. Chen, "Cyclostationary feature based spectrum sensing via low-rank and sparse decomposition in cognitive radio networks," in *Proc. 16th Int. Symp. Commun. Inf. Technol. (ISCIT)*, Qingdao, China, Sep. 2016, pp. 615–619.

[6] Y. Zeng and Y.-C. Liang, "Eigenvalue-based spectrum sensing algorithms for cognitive radio," *IEEE Trans. Commun.*, vol. 57, no. 6, pp. 1784–1793, Jun. 2009.

[7] Y. Zeng and Y.-C. Liang, "Spectrum-sensing algorithms for cognitive radio based on statistical covariances," *IEEE Trans. Veh. Technol.*, vol. 58, no. 4, pp. 1804–1815, May 2009.

[8] F. Zhou, N. C. Beaulieu, Z. Li, and J. Si, "Feasibility of maximum eigenvalue cooperative spectrum sensing based on Cholesky factorisation," *IET Commun.*, vol. 10, no. 2, pp. 199–206, 2016.

[9] N. Pillay and H.-J. Xu, "Eigenvalue-based spectrum sensing using the exact distribution of the maximum eigenvalue of a Wishart matrix," *IET Signal Process.*, vol. 7, no. 9, pp. 833–842, Dec. 2013.

[10] A. Maali, H. Semlali, N. Boumaaz, and A. Soulmani, "Energy detection versus maximum eigenvalue based detection: A comparative study," in *Proc. 14th Int. Multi-Conf. Syst. Signals Devices (SSD)*, Marrakech, Morocco, Mar. 2017, pp. 1–4.

[11] C. B. A. Wael, N. Armi, and B. P. A. Rohman, "Spectrum sensing for low SNR environment using maximum-minimum eigenvalue (MME) detection," in *Proc. Int. Seminar Intell. Technol. Appl. (ISITIA)*, Lombok, Indonesia, Jul. 2016, pp. 435–438.
[12] S. M. Putri and I. Sugihartono, "Energy efficiency in cognitive radio with cooperative MME (maximum to minimum eigenvalue) spectrum sensing method," in *Proc. Int. Seminar Intell. Technol. Appl. (ISITIA)*, Surabaya, Indonesia, May 2015, pp. 379–384.
[13] X. Yang, K. Lei, S. Peng, and X. Cao, "Blind detection for primary user based on the sample covariance matrix in cognitive radio," *IEEE Commun. Lett.*, vol. 15, no. 1, pp. 40–42, Jan. 2011.
[14] F. Zhou, Z. Li, J. Si, and L. Guan, "An efficient spectrum sensing algorithm for cognitive radio based on finite random matrix," in *Proc. IEEE 25th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Washington, DC, USA, Sep. 2014, pp. 1223–1227.
[15] Z. Li, F. Zhou, J. Si, P. Qi, and L. Guan, "Feasibly efficient cooperative spectrum sensing scheme based on Cholesky decomposition of the correlation matrix," *IET Commun.*, vol. 10, no. 9, pp. 1003–1011, 2016.
[16] P. Anal and S. P. Maity, "On outage minimization in cognitive radio networks through routing and power control," *Wireless Pers. Commun.*, vol. 98, no. 1, pp. 251–269, Jan. 2018.
[17] M. Bkassiny, Y. Li, and S. K. Jayaweera, "A survey on machine-learning techniques in cognitive radios," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1136–1159, 3rd Quart., 2013.
[18] Y. Lu, P. Zhu, D. Wang, and M. Fattouche, "Machine learning techniques with probability vector for cooperative spectrum sensing in cognitive radio networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCMC)*, Doha, Qatar, Apr. 2016, pp. 1–6.
[19] C. An, D. Zhang, C. Yuan, L. Li, X. Zhao, S. Geng, W. Zheng, and W. Shao, "Spectrum sensing based on kNN algorithm for 230 MHz power private networks," in *Proc. 12th Int. Symp. Antennas, Propag. EM Theory (ISAPE)*, Hangzhou, China, Dec. 2018, pp. 1–4.
[20] Y. Xu, P. Cheng, Z. Chen, Y. Li, and B. Vucetic, "Mobile collaborative spectrum sensing for heterogeneous networks: A Bayesian machine learning approach," *IEEE Trans. Signal Process.*, vol. 66, no. 21, pp. 5634–5647, Nov. 2018.
[21] G. J. Mendis, J. Wei, and A. Madanayake, "Deep learning-based automated modulation classification for cognitive radio," in *Proc. IEEE Int. Conf. Commun. Syst. (ICS)*, Shenzhen, China, Dec. 2016, pp. 1–6.
[22] S. Riyaz, K. Sankhe, S. Ioannidis, and K. Chowdhury, "Deep learning convolutional neural networks for radio identification," *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 146–152, Sep. 2018.
[23] D. Zhang and X. Zhai, "SVM-based spectrum sensing in cognitive radio," in *Proc. 7th Int. Conf. Wireless Commun., Netw. Mobile Comput. (WCNMC)*, Wuhan, China, Sep. 2011, pp. 1–4.
[24] S. Hou and R. C. Qiu, "Kernel feature template matching for spectrum sensing," *IEEE Trans. Veh. Technol.*, vol. 63, no. 5, pp. 2258–2271, Jun. 2014.
[25] A. Patle and D. S. Chouhan, "SVM kernel functions for classification," in *Proc. Int. Conf. Adv. Technol. Eng. (ICATE)*, Mumbai, India, Jan. 2013, pp. 1–9.
[26] K. M. Thilina, K. W. Choi, N. Saquib, and E. Hossain, "Pattern classification techniques for cooperative spectrum sensing in cognitive radio networks: SVM and W-KNN approaches," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Anaheim, CA, USA, Dec. 2012, pp. 1260–1265.
[27] A. Agarwal, H. Jain, R. Gangopadhyay, and S. Debnath, "Hardware implementation of k-means clustering based spectrum sensing using usrp in a cognitive radio system," in *Proc. Int. Conf. Adv. Comput., Commun. Inf. (ICACCI)*, Udupi, India, Sep. 2017, pp. 1772–1777.
[28] Y. Tan and J. Wang, "A support vector machine with a hybrid kernel and minimal Vapnik-Chervonenkis dimension," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 4, pp. 385–395, Apr. 2004.
[29] Y. Hamasuna, Y. Endo, and S. Miyamoto, "Support vector machine for data with tolerance based on hard-margin and soft-margin," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Hong Kong, Jun. 2008, pp. 750–755.
[30] G. H. Golub and C. F. van Loan, *Matrix Computations*, 4th ed. London, U.K.: The Johns Hopkins Univ. Press, 2013.

JIANRONG BAO (S'06–M'11–SM'19) received the B.S. degree in polymeric materials and engineering and the M.S.E.E. degree from the Zhejiang University of Technology, Hangzhou, China, in 2000 and 2004, respectively, and the Ph.D.E.E. degree from the Department of Electronic Engineering, Tsinghua University, Beijing, China, in 2009.

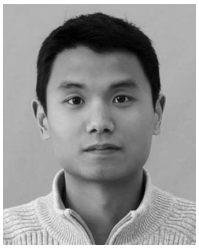
He was a Postdoctoral Researcher with Zhejiang University, from 2011 to 2013, and Southeast University, from 2014 to 2017, and a Visiting Scholar with Columbia University, New York, NY, USA, in 2015. He is currently a Professor with the School of Communication Engineering, Hangzhou Dianzi University, Hangzhou, China. He is also a Visiting Scholar with the Zhejiang Provincial Key Laboratory of Information Processing, Communication and Networking, Zhejiang, China. His main research interests include modern wireless communications, cognitive radio, information theory and coding, communication signal processing, and wireless sensor networks.

JIANYUAN NIE received the B.S.E.E from the School of Electronic Information Engineering, Taiyuan University of Science and Technology, Taiyuan, China, in 2017. She is currently pursuing the master's degree with the School of Communication Engineering, Hangzhou Dianzi University, Hangzhou, China.

Her research interests include cognitive radio, compressed sensing, and cooperative communications.

CHAO LIU received the B.S.E.E. and Ph.D.E.E. degrees from the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China, in 2000 and 2005, respectively.

He is currently an Associate Professor with the School of Information Engineering, Hangzhou Dianzi University, Hangzhou, China. His research interests include modern wireless communication and coding and MIMO multi-user detection.



BIN JIANG received the B.S.E.E. and M.S.E.E. degrees from the School of Communication Engineering, Hangzhou Dianzi University, Hangzhou, China, in 2002 and 2007, respectively.

He is currently a Senior Engineer with the School of Communication Engineering, Hangzhou Dianzi University. His main research interests include wireless communications, signal processing, information theory, and channel coding.

FANG ZHU received the B.S.E.E. degree from the Department of Electrical Engineering and Computer Science, Hubei University of Technology, Wuhan, China, in 1997, and the M.S.E.E. degree from the Department of Radio Engineering, Southeast University, Nanjing, China, in 2004.

He is currently a Lecturer with the School of Communication Engineering, Hangzhou Dianzi University, Hangzhou, China. His main research interests include wireless digital communications, iterative signal processing, the IoT, and embedded systems.

JIANHAI HE received the B.S.E.E. degree from Hangzhou Dianzi University, Hangzhou, China, in 2002, and the M.S.E.E. degree from Zhejiang University, Zhejiang, China, in 2011.

He is currently an Associate Professor with the School of Electronic Information Engineering, Ningbo Polytechnic. His current research interests include wireless communications, wireless sensor networks, and the embedded IoT systems.