IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 30, 2018, accepted March 23, 2018, date of publication April 11, 2018, date of current version May 24, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2825603

Spatio-Temporal Spectrum Sensing in Cognitive Radio Networks Using Beamformer-Aided SVM Algorithms

OLUSEGUN PETER AWE[1], (Member, IEEE), ANASTASIOS DELIGIANNIS[2], (Member, IEEE), AND SANGARAPILLAI LAMBOTHARAN[2], (Senior Member, IEEE)

[1]Department of Electronic and Electrical Engineering, Obafemi Awolowo University, Ife 220282, Nigeria
[2]Wolfson School, Loughborough University, Loughborough LE11 3TU, U.K.

Corresponding author: Anastasios Deligiannis (a.deligiannis@lboro.ac.uk)

This work was supported in part by the Petroleum Technology Development Fund of Nigeria and in part by the Engineering and Physical Science Research Council of U.K. under Grant EP/M015475/1.

ABSTRACT This paper addresses the problem of spectrum sensing in multi-antenna cognitive radio system using the support vector machine (SVM) algorithms. First, we formulated the spectrum sensing problem under multiple primary users scenarios as a multiple state signal detection problem. Next, we propose a novel beamformer-aided feature realization strategy for enhancing the capability of the SVM for signal classification under both single and multiple primary users conditions. Then, we investigate the error correcting output codes-based multi-class SVM algorithms and provide a multiple independent model alternative for solving the multiple state spectrum sensing problem. The performance of the proposed detectors is quantified in terms of probability of detection, probability of false alarm, receiver operating characteristics (ROC), area under ROC curves, and overall classification accuracy. Simulation results show that the proposed detectors are robust to both temporal and joint spatio-temporal detection of spectrum holes in cognitive radio networks.

INDEX TERMS Cognitive radio, spectrum sensing, support vector machine, beamforming, multiple primary users.

I. INTRODUCTION

Cognitive radio (CR) has been widely accepted as a veritable solution to the problem of scarcity of the radio spectrum. However, the successful implementation of CR depends on the use of reliable spectrum sensing schemes [2], [3]. In general, spectrum sensing is a process by which unlicensed users also known as secondary users (SUs) acquire information about the status of the radio spectrum allocated to a licensed user or primary user (PU) for the purpose of accessing unused licensed bands on an opportunistic basis without causing intolerable interference to the transmissions of the licensed user [4].

It has been established that for cognitive devices to be really cognizant of the PUs' activities in their radio frequency (RF) environment, it is imperative that they be equipped with both learning and reasoning functionalities [2], [5]. Consequently, in fairly recent times, machine learning (ML) algorithms have been studied for spectrum sensing in CR. For example, Thilina *et al.* [6] proposed the supervised, non-parametric K-nearest neighbor (kNN) and support vector machine (SVM) classifiers for temporal spectrum sensing. The semi-supervised K-means clustering and expectation-maximization (EM) algorithms were also proposed in [6] while the unsupervised parametric, soft-assignment variational Bayesian inference technique was studied in [7]. A tracking technique for adapting parametric classifiers to changing channel conditions was also presented in [8]. Furthermore, in [9]–[11] the adaptation of deep convolutional neural networks (CNN) for detecting modulation patterns was considered. Moreover, Rajendran *et al.* [12] investigated a long short term memory based deep learning classifier utilizing the temporal dependencies between amplitude and phase characteristics to differentiate various modulation schemes. In [13]–[15], a deep learning based CNN was proposed for combining sensing results in cooperative spectrum sensing. In these works, classifier based learning methods have been well demonstrated to exhibit excellent signal detection capability. Nevertheless, it is important to note that the classifiers considered are primarily designed for binary classification problems. Hence, their applications have been limited to

temporal spectrum sensing. However, the main reason for spectrum sensing in CR is to reliably detect unused bands at a specific *time* and *location* [16]. Thus, detection of spatial spectrum holes is an important research problem.

A. MOTIVATION AND RELATED WORKS

In the future deployment of CRs, the PU-SU system might be a cellular network with the possibility of frequency re-use among nearby cells, thereby guaranteeing the existence of multiple PUs in the vicinity of the SUs [17]. For example, the PUs might be high power macro cell base stations while the SU might be a low power micro cell base station (SBS) located at the cell edge of multiple macrocells. In scenarios like this, it is apparent that when one or more PUs are not transmitting in the direction of the SU, spectrum holes are created in the geographical locations of such PUs (spatial spectrum hole) which if detected, can be utilized by the SUs operating within such locations. By using transmission strategies such as beamforming and user-location based power allocation, more efficient utilization of spectrum resources can be achieved. It is noteworthy that the classifiers investigated in the works above are not capable of handling the spectrum sensing problem ensuing under this condition.

There are some reported research works on spatial-temporal spectrum sensing. For example, [18] proposed addressing the sensing problem using kridged Kalman filter where channel gain maps are being tracked in real time. In [19], the problem of joint detection and localization of PUs was approached using distributed antenna system while [20] investigated spatio-temporal opportunity detection for spectrum in heterogeneous CR networks using a two dimensional sensing framework which exploits correlation in time and space simultaneously. Although the spectrum sensing problem has been well researched, joint spatio-temporal sensing within the framework of ML has not been considered despite the excellent detection capability of ML based schemes as evident in [5]– [8].

B. CONTRIBUTIONS

In this study, extending our work in [1] we consider multi-antenna SUs, propose and investigate a novel and efficient strategy for adopting the SVM algorithms in both temporal and joint spatio-temporal spectrum sensing. We consider scenarios with multiple PUs and re-formulated the sensing problem as a multiple state signal detection problem which facilitates the application of multi-class [1] SVM learning algorithms. Our goal is to determine the actual state of the PU-SU network at any given time as a function of PUs' activities. This means that we desire to know the actual number of active PUs and their locations in the network during the sensing interval. Furthermore, we introduce a beamforming based pre-processing technique for realizing our energy features in order to enhance the performance of the SVM algorithm. We choose the SVM owing to its superior performance in comparison with other traditional pattern classifiers. Specifically, the main contributions of this paper are as follows.

- Reformulate the spectrum sensing problem under multiple users scenarios as a multiple state signal detection problem and demonstrate the applicability of multi-class learning algorithms for addressing it.
- Introduce and develop an efficient beamformer aided feature realization strategy for enhancing the detection capability of classifier based learning algorithms.
- Propose and investigate the ECOC based multi-class algorithms within the context of spectrum sensing and provide a multiple independent model (MIM) alternative for solving the multi-state spatio-temporal spectrum sensing problem.

We quantify our results in terms of probabilities of detection and false alarm, receiver operating characteristics (ROC), area under ROC curve (AuC) and overall classification accuracy. The proposed schemes are blind and the exact knowledge in terms of the PU signal, noise or the channel gain is not required. To the best of our knowledge, this is the first application of beamforming aided multi-class learning algorithms in the context of CR spectrum sensing and we hereby demonstrate the advantages and evaluate the performance of the proposed schemes.

The remainder of the paper is organized as follows. In section II, we describe the system model and assumptions. Section III presents the beamformer design technique. In section IV, we describe our feature vectors realization procedure while the proposed SVM based algorithms are presented in Section V. Numerical results are discussed in section VI followed by the conclusion in section VII.

II. SYSTEM MODEL AND ASSUMPTIONS

We consider a PU-SU network where the SUs and PUs are fixed, multi-antenna devices with beamforming capabilities. It is assumed that the SUs are equipped with M antennas and operating in the coverage areas of P PU transmitters. The PUs are assumed to be geographically separated but operating within the same frequency band. We further assumed that the PUs activities are such that when all PUs are inactive, spectrum holes are available both temporally and spatially at the respective PU's location. However, when only p PUs, where $p < P$ are active, spatial spectrum holes are available at some $\bar{p} = P - p$ PUs' geographical locations (the coverage areas of the $\bar{p}$ inactive PUs) which can be utilized by the SUs during the $\bar{p}$ PUs' idle period. Such spatially available bands could serve the purpose of base-to-mobile communications as well as device-to-device communications that is being proposed as an integral part of the next-generation cellular networks [21].

We first define a class in our classification problem as the number of active PUs in the network at any given point in time. Hence, the set of all possible classes can be defined as

$$\mathcal{P} = \{C_1, C_2, \ldots, C_P\} = \{C_i\}_{i=1}^{P}. \quad (1)$$

[1] †In this context, the term multi-class denotes more than two classes.

Within each class, there is a set of possible states where each state indicates the different combinations of active PUs. For example, for class 3, i.e. for C_3, three PUs are active. Hence, out of P possible PUs, there are $\binom{P}{3}$ combinations called states where $\binom{P}{i} = \frac{P!}{(P-i)!\,i!}$. In general, for the i^{th} class, we will have $Q(i) = \binom{P}{i}$ possible states written as

$$C_i = \{\mathcal{S}_1^i, \mathcal{S}_2^i, \ldots, \mathcal{S}_{Q(i)}^i\} = \{\mathcal{S}_q^i\}_{q=1}^{Q(i)} \tag{2}$$

where $\mathcal{S}_q^i$ is a particular selection of PUs in class C_i. Our spectrum sensing problem is therefore formulated as determining not only the availability of spectrum hole but also the state of the network, i.e. to determine which primary user(s) are active. Hence, we write the received signal model under this scenario as a multiple hypotheses testing of the form

$$H_0 : \mathbf{y}(n) = \boldsymbol{\eta}(n) \tag{3}$$

$$H_{i,q} : \mathbf{y}(n) = \sum_{p \in \mathcal{S}_q^i} \bar{\boldsymbol{\phi}}_p s_p(n) + \boldsymbol{\eta}(n), \quad \forall \mathcal{S}_q^i \in C_i,\ \forall C_i \in \mathcal{P} \tag{4}$$

where H_0 implies that *all* PUs are inactive and $H_{i,q}$, means that i number of PUs corresponding to the q^{th} state are active. Therefore, the alternative hypothesis for H_0 is

$$H_1 = \bigcup_{\substack{i=\{1,\ldots,P\} \\ q=\{1,\ldots,Q(i)\}}} H_{i,q}. \tag{5}$$

Furthermore, $\mathbf{y}(n) = [y_1(n), y_2(n), \ldots, y_M(n)]^T$ is the vector of instantaneous signal received at the SU over bandwidth ω of interest within which the PUs operate, $\bar{\boldsymbol{\phi}}_p = [\bar{\phi}_{1,p}, \bar{\phi}_{2,p}, \ldots, \bar{\phi}_{M,p}]^T$ is the vector of channel coefficients between the p^{th} PU and the SU. The remaining parameters in (4) are $s_p(n)$ which is the instantaneous PU signal, assumed to be an independent and identically distributed (*i.i.d*) process with mean zero and variance, $\mathbb{E}|s_p(n)|^2 = \sigma_{s_p}^2$, and $\boldsymbol{\eta}(n) = [\eta_1(n), \eta_2(n), \ldots, \eta_M(n)]^T$, which is the vector of noise, $\eta_m(n)$, assumed to be an *i.i.d* circularly symmetric complex zero-mean Gaussian with variance, $\mathbb{E}|\eta_m(n)|^2 = \sigma_\eta^2$.

Under H_0, all PUs are inactive and it corresponds to the null hypothesis. On the other hand, H_1 corresponds to *composite* alternative hypothesis where at any given time, *at least one* PU is active during the sensing interval. It is apparent that this composite hypothesis intuitively embeds P classes of alternative hypotheses each of which may comprise of one or more possible network states. Our goal is to learn the peculiar attributes that uniquely characterize each state under H_1, and to use this knowledge to discriminate them. In the context of our multi-class problems, one or more states may be classified more accurately than others within the same class, or in another class. Therefore, if i and q denote the class and state index respectively, we define the overall classification accuracy as

$$CA_{ovr} = \frac{1}{(Y_P + 1)} \left\{ \sum_{i=1}^{P} \sum_{q=1}^{Q_i} p(H_i|H_i)\, p(H_q|H_q) + p(H_0|H_0) \right\} \tag{6}$$

where P is the total number of classes, Q_i is the number of states in the i^{th} class, $Y_P = card(\bigcup_{i=1}^{P} Q_i)$ is the total number of states present in all classes being considered under H_1, and $card(G)$ implies cardinality. We wish to emphasize here that if $Q_i = card(Q_i),\ \forall\, i, j \in P, \exists j : Q_i \neq Q_j$.

III. NOVEL BEAMFORMER DESIGN TECHNIQUE FOR FEATURES REALIZATION

In this section, we describe a beamforming technique for enhancing the receive signal-to-noise-ratio (*SNR*) and hence, the quality of received PU(s) signals used for realizing the feature vectors of our spectrum sensing scheme. We assume that the SU's antennas are identical and equally spaced so as to form a uniform linear array. Let $s_m(n)$ denote the discrete time PU's signal arriving at the m^{th} antenna of the array at an angle of arrival (AOA), θ, assumed to be uniformly distributed within the interval $[\theta_{min}, \theta_{max}]$. The total azimuth coverage of the array is restricted to 180° so that the array scans the entire range, $\Theta \in [-90°, 90°]$ for θ [22]. Multiple arrays may be deployed if 360° coverage is required. In our beamformer design, Θ is partitioned into K sectors denoted as $\{\bar{\theta}_k\}_{k=1}^{K}$, where each sector has a width $\bar{\theta}_k$ and it is assumed that the AOA of the PU signals, θ, lies within $\bar{\theta}_k$. For example, if we let $K = 9$ so that $\bar{\theta}_k$ equals 20° $\forall\ k \in K$, $\bar{\theta}_1 \in [-90, -70)$, $\bar{\theta}_2 \in [-70, -50)$, and so on. The goal here is to design a unique beamformer for each sector $\bar{\theta}_k \in \Theta$ which maximizes the array gain within $\bar{\theta}_k$ and minimizes it elsewhere, i.e. in all $\bar{\theta}_j \in \Theta \setminus \bar{\theta}_k$.

Let the sector of interest $\bar{\theta}_k$ be further represented as a set of $\bar{K}$, fine angular sub-partitions described by $\{\tilde{\theta}_{\bar{k}}\}_{\bar{k}=1}^{\bar{K}}$. Additionally, let the *desired* beampattern for the sector be represented by $\phi(\tilde{\theta}_{\bar{k}})$ and the array response vector associated with θ be written as $\mathbf{a}(\theta) = [1\ \mathrm{e}^{-j\theta}\ \cdots\ \mathrm{e}^{-j(M-1)\theta}]^T$. If the beamformer required to obtain $\phi(\tilde{\theta}_{\bar{k}})$ is denoted by $\mathbf{w}_{\tilde{\theta}_{\bar{k}}}$, we wish to determine a rank one matrix, $\mathbf{R} = \mathbf{w}_{\tilde{\theta}_{\bar{k}}} \mathbf{w}_{\tilde{\theta}_{\bar{k}}}^H$ that minimizes the difference between the desired beampattern, $\phi(\tilde{\theta}_{\bar{k}})$ and the *actual* receive beampattern, $\mathbf{a}^*(\tilde{\theta}_{\bar{k}})\mathbf{R}\mathbf{a}(\tilde{\theta}_{\bar{k}})$ in the least squares sense [23]. The operation $(\cdot)^*$ denotes the conjugate transpose. Hence, our beamformer design task becomes a beampattern matching problem which can be formulated mathematically as an optimization problem of the form

$$\begin{aligned} &\underset{\tilde{\alpha}, \mathbf{R}}{\text{minimize}}\ t \\ &\text{subject to} \sum_{\bar{k}=1}^{\bar{K}} \left[\tilde{\alpha}\phi(\tilde{\theta}_{\bar{k}}) - \mathbf{a}^*(\tilde{\theta}_{\bar{k}})\mathbf{R}\mathbf{a}(\tilde{\theta}_{\bar{k}}) \right]^2 < t, \\ &\qquad \mathbf{R} \succeq 0,\ rank(\mathbf{R}) = 1 \end{aligned} \tag{7}$$

where $\tilde{\alpha}$ is a scaling factor whose optimal value can be obtained jointly as part of the solution of the optimization problem. However, due to the matrix rank constraint in (7) the problem is rendered non-convex. So, we relax the rank restriction imposed on $\mathbf{R}$ and recast (7) into a semi-definite optimization problem which can be solved to obtain optimal $\mathbf{R}$ ($\mathbf{R}_{\tilde{\theta}_k}^{opt}$). The desired beamformer weights, $\mathbf{w}_{\tilde{\theta}_k}$ can be

obtained as the principal eigenvector of $\mathbf{R}_{\bar{\theta}_k}^{opt}$ multiplied by the square root of the principal eigenvalue [24].

IV. BEAMFORMER AIDED ENERGY FEATURE VECTORS REALIZATION STRATEGY

In many practical spectrum sensing scenarios, the presence of heavily built structures between the PU and the SU may result in the PU(s) signals arriving the SU via multiple, strong paths. Where there are multiple PUs, it is possible that reflections from these multiple sources arrive at the SU receiver at azimuth angles that fall within the same sector, $\bar{\theta}_k$. This is treated as a case of overlapping reflections. On the other hand, the various reflections may be received by the SU at widely separated AOAs coinciding with different sectors, thus, considered as a non-overlapping case.

Without loss of generality, we demonstrate the usefulness of the beamformers presented in section III under the above scenarios and describe the algorithms for deriving our beamformer based energy features. To perform spectrum sensing, we propose to implement the entire learning process in two stages described as the qualification and training stages. In both single PU and multiple PUs scenarios, the goal of the qualification stage is to identify the sole beamformer or set of beamformers (as the case may be) whose output will yield the desired high quality energy features. We also seek to find a way of associating the energy sample obtained from each beamformer with the PUs. Thus, the qualification stage is accomplished in two phases namely, the identification phase and the association phase.

1) IDENTIFICATION PHASE

Here, the SU employs the beamformers to *learn* the direction of arrival (DOA) of the PU signals and the corresponding azimuth angles. The strategy for implementing the identification phase is described as follows. Let the discrete time signal received at the M-element array SU be represented as

$$\mathbf{y}(n) = \begin{cases} \mathbf{a}(\theta)s(n) + \boldsymbol{\eta}(n), & \text{if } P = 1 \\ \sum_{i=1}^{\mathrm{P}} \mathbf{a}(\theta_i)s_i(n) + \boldsymbol{\eta}(n), & \text{if } P > 1 \end{cases} \tag{8}$$

where $\mathbf{y}(n) \in \mathbb{C}^M$. If the beamformer designed for the k^{th} sector is $\mathbf{w}_{\bar{\theta}_k}$, the beamformer output can be expressed as

$$x_k(n) = \mathbf{w}_{\bar{\theta}_k}^H \mathbf{y}(n). \tag{9}$$

Suppose we collect N samples of $x_k(n)$, the qualifying energy feature is derived as

$$\vartheta_k = \frac{1}{N}\sum_{n=1}^{N} |x_k(n)|^2. \tag{10}$$

Further, if the vector of energy samples obtained for all beamformers is denoted as $\boldsymbol{\vartheta} \triangleq [\vartheta_1, \vartheta_2, \ldots, \vartheta_K]$, where $\boldsymbol{\vartheta} \in R^K$, we seek to determine the set of qualified beamformer(s) whose output will yield the *true* feature vector. To achieve this, we apply the decision threshold, ζ_1 defined for a target false alarm probability, $\bar{P}_{fa}$ as [25]

$$\zeta_1 = \sigma_{s_p}^2 \gamma^{-1}\left(1 + \sqrt{\frac{2}{N}} Q^{-1}(\bar{P}_{fa})\right) \tag{11}$$

where $\gamma = \frac{\sigma_{sp}^2}{\sigma_\eta^2}$ is the receive SNR of the PU(s) signal measured at the SU under hypothesis $H_{i,q}$ that corresponds to the state, $\mathcal{S}_{Q(P)}^P$ (i.e. when all PUs are active) and Q^{-1} denotes the inverse Q-function, $Q(x) = \frac{1}{2\pi}\int_x^{+\infty} \exp(-t^2/2)dt$. The true dimension of the feature vector, $\bar{S} \ll K$ given by $card(\boldsymbol{\vartheta} : \vartheta_k > \zeta_1, \forall k \in K$ can thus be determined. In addition, since $\bar{S} \ll M$ it follows that through the qualification phase, the dimension of the feature vector in the input space is significantly reduced. This also leads to reduction in the computational overhead of the sensing algorithm in comparison with the non-beamformer based alternatives. Furthermore, in multiple PUs scenarios, if the AOA's of the individual PU(s) signals and the attendant energy sample(s), $\vartheta_k, \forall k \in \bar{S}$ can be identified, we opine that it is very feasible to use independent binary classifiers to monitor the activities of individual PU without recourse to multi-class algorithms.

2) ASSOCIATION PHASE

Here, the process for associating the energy samples with their respective sources is described. Under multiple PUs scenarios, in addition to determining the qualified beamformers set via the identification phase, we need to know the particular beamformer(s) with which respective PU signal can be associated, thus enabling the SU to know the source(s) responsible for the signals derived at every beamformer. This will facilitate the use of multiple, independent SVM models (MIMSVM) to simultaneously monitor the activities of all PUs as a viable alternative to multi-class SVM (MSVM) algorithms [26].

In general, let the set of qualified beamformers be represented by $\mathcal{B} = \{b_1, b_2, \ldots, b_{\bar{S}}\} \in \{\mathbf{w}_{\bar{\theta}_1}, \mathbf{w}_{\bar{\theta}_2}, \ldots, \mathbf{w}_{\bar{\theta}_K}\}$. Further, let the corresponding sequences of the PU signal samples derived at the beamformers output be described as $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{\bar{S}}]$ where $\mathbf{x}_{\bar{s}} \triangleq [x_{\bar{s}}(1), x_{\bar{s}}(2), \ldots, x_{\bar{s}}(N)]^T$, N is the number of samples, $\forall \bar{s} \in \bar{S}$. To solve the beamformer association problem, we assume that at least one of the multipaths is from the LOS indicating the known direction of the PUs. The beamformer corresponding to the LOS signal of the i^{th} PU (PU_i) is denoted as b_{ref}^i, where $b_{ref}^i \in \mathcal{B}$. However, we need to associate the other multipaths with each of the PU. We do this by performing the cross correlation between the known beamformer's output and every other beamformer's output and compare the result to a threshold, ζ_2. The estimate of the cross correlation between the sequences derived at the output of any two beamformers, $\mathbf{x}_i$ and $\mathbf{x}_j$ is computed for various delays as

$$R_{\mathbf{x}_i\mathbf{x}_j}(\tau) = \frac{1}{N}\sum_{n=1}^{N}[\mathbf{x}_i(n)\mathbf{x}_j^*(n+\tau)], n = 1, \ldots, N. \tag{12}$$

The test statistic for comparison is therefore derived as

$$\mho_d = \sum_{\tau=-\tau_d}^{\tau_d} |R_{\mathbf{x}_{ref}\mathbf{x}_{\bar{s}}}(\tau)|^2, \tag{13}$$

where $R_{\mathbf{x}_{ref}\mathbf{x}_{\bar{s}}}(\tau)$ denotes the τ-lag cross correlation between $\mathbf{x}_{ref}$ and $\mathbf{x}_{\bar{s}}$ and $\mho_d$ is the sum of square of the magnitude of cross correlation returns over the search interval, $[-\tau_d, \tau_d]$. The search interval must be carefully chosen to capture the likely delay, τ', between $\mathbf{x}_{ref}$ and the reflected version which may be present in $\mathbf{x}_{\bar{s}}$. It should be noted here, that the exact amount of the delay, τ', may *not* be known a priori, so τ_d should be sufficiently large. To determine the presence of $\mathbf{x}_{ref}$ in $\mathbf{x}_{\bar{s}}$, we compare $\mho_d$ to ζ_2 defined by

$$\zeta_2 = \tilde{\varrho}|R_{\mathbf{x}_{ref}\mathbf{x}_{ref}}(0)|^2 \tag{14}$$

where $\tilde{\varrho}$ is an appropriate scalar that should be chosen carefully as a trade-off between probability of detection and false alarm rate. If $\mho_d \geq \zeta_2$, we conclude that $\mathbf{x}_{ref}$ is present in $\mathbf{x}_{\bar{s}}$, that is, the same PU source signal is arriving from two different directions. If $\mho_d < \zeta_2$, then signals arriving from the two different directions are independent. Meanwhile, having identified the DOA of the PU's signal via the qualified beamformers set, during the *training* stage, the SU derives the required training energy features from the qualified beamformers set *only*. The remaining beamformers' output are simply ignored.

V. BEAMFORMER AIDED SVM LEARNING ALGORITHMS

For completeness, we first consider the application of the proposed beamformer aided SVM algorithm to solve the spectrum sensing problem under single primary user scenarios.

A. SPECTRUM SENSING USING BEAMFORMER-DERIVED FEATURES AND BINARY SVM CLASSIFIER UNDER SINGLE PRIMARY USER CONDITION

Under single PU scenario, our beamforming based spectrum sensing problem can be formulated as a binary hypothesis testing problem of the form

$$x_k(n) = \mathbf{w}_{\bar{\theta}_k}^H \mathbf{y}(n) \begin{cases} \mathbf{y}(n) = \boldsymbol{\eta}(n) & \mathrm{H}_0 \\ \mathbf{y}(n) = \mathbf{a}(\theta_k)s(n) + \boldsymbol{\eta}(n) & \mathrm{H}_1, \end{cases} \quad \forall k \in \mathcal{B} \tag{15}$$

where $x_k(n)$ is the instantaneous signal at the output of the k^{th} beamformer. Without loss of generality, we assume non-overlapping multipath scenario and collect D independent energy vectors, each comprising energy samples realized according to (10) for training purpose. Let $\mathcal{S} = \{(\boldsymbol{\vartheta}_1, \bar{l}_1), (\boldsymbol{\vartheta}_2, \bar{l}_2), \ldots, (\boldsymbol{\vartheta}_D, \bar{l}_D)\}$ represent the training data set where $\boldsymbol{\vartheta}_i \in \mathcal{R}^{\bar{S}}$ is an $\bar{S}$-dimensional feature vector and $\bar{l}_i \in \{-1, 1\}$ is the corresponding class label. The classification task using the soft margin SVM can be formulated as an optimization problem of the form [27], [28]

$$\begin{aligned} \underset{\vec{z},b,\xi}{\text{minimize}} \quad & \langle \vec{z}.\vec{z} \rangle + \Gamma \sum_{i=1}^{D} \xi_i \\ \text{subject } to \quad & \bar{l}_i(\langle \vec{z}.\beta(\boldsymbol{\vartheta}_i) \rangle + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0, \quad i = 1, 2, \ldots, D. \end{aligned} \tag{16}$$

where $\vec{z}$ is the weight vector which is normal to the SVM's separating hyperplane, $\langle \vec{z}.\vec{z} \rangle$ denotes inner product, b is the bias describing the perpendicular distance between the origin and the hyperplane, ξ is a slack variable, Γ is a soft margin parameter known as the box constraint and $\beta(\boldsymbol{\vartheta})$ is a suitable kernel function [28], [29]. The convex optimization problem can be solved using the Langrangian method of multipliers and applying the Karush-Kuhn-Tucker conditions [30]. We solved the dual form of (16) using quadratic programming algorithm [31] and thus, the kernel based SVM classifier for predicting the status of the PU is obtained. The PU's status, H_0 or H_1 is then determined in terms of the class of a new observed data vector, $\boldsymbol{\vartheta}^{new}$, as

$$\bar{l}(\boldsymbol{\vartheta}^{new}) = sgn\left(\sum_{i=1}^{\mathrm{N}_s} \bar{l}_i \alpha_i \beta(\boldsymbol{\vartheta}^{new}, \boldsymbol{\vartheta}_i) + b \right) \tag{17}$$

where N_s is the number of support vectors. In Algorithm 1, a summary of the proposed beamformer-aided spectrum sensing technique for single PU scenarios is presented.

Algorithm 1 Beamformer Aided SVM Algorithm for Spectrum Sensing Under Single PU Scenarios

Learning process:

Qualification stage

1.) Load $\mathbf{w}_{\bar{\theta}_k}$, $\forall k \in K$ and scan Θ .

2.) Compute $\boldsymbol{\vartheta} \in R^K$, under H_1 using (10).

3.) Apply ζ_1 in (11) to $\boldsymbol{\vartheta}$ in (*ii*) to find $\mathcal{B}$.

Training stage

4.) Using $\mathcal{B}$ in (*iii*) obtain set $\{D\}$ of $\boldsymbol{\vartheta} \in R^{\bar{S}}$, under H_0 and H_1 using (10).

5.) Generate prediction model $\bar{l}(\boldsymbol{\vartheta}^{new})$ in (17) from $\{D\}$ in (*iv*).

Prediction process:

do repeat

6.) Obtain test sample, $\boldsymbol{\vartheta}^{new} \in R^{\bar{S}}$.

7.) Classify $\boldsymbol{\vartheta}^{new}$ in (*vi*) to decide H_0 or H_1.

B. ECOC BASED BEAMFORMER AIDED MULTICLASS SVM FOR SPECTRUM SENSING UNDER MULTIPLE PRIMARY USER SCENARIOS

One significant limitation of the conventional SVM (CSVM) algorithm adopted in section V-A is that it is mainly designed for binary classification tasks and as such, it fits well for addressing temporal spectrum sensing. In this subsection, we introduce error correcting output codes (ECOC) based

MSVM technique as a more suitable tool for solving spatio-temporal spectrum holes detection problem ensuing under multiple PU scenarios.

In reality, due to mutual interference, the number of active transmitters that can simultaneously transmit in the same spectral band within a given geographical area is limited [18]. It is reasonable therefore to assume that the actual number of PUs to be considered operating in the vicinity of the SU, P is small. With this in mind and without loss of generality, we describe the proposed ECOC MSVM algorithm given a scenario with two PUs whose signals are received via LOS. Let us consider that the signals from PU_1 and PU_2 arrive the SU at AOAs corresponding to $\theta_1 \in \bar{\theta}_3$ and $\theta_2 \in \bar{\theta}_6$ respectively. In this case, the multiple hypotheses problem defined in (3) and (4) translates to four hypotheses testing problem. If we let the indexes, i and q in $H_{i,q}$ indicate the class and state respectively, these hypotheses can be written as

$$H_0 : \mathbf{x}(n) = [\mathbf{w}_{\bar{\theta}_3}^H \boldsymbol{\eta}(n)\ \mathbf{w}_{\bar{\theta}_6}^H \boldsymbol{\eta}(n)]^T \tag{18a}$$

$$H_{1,1} : \mathbf{x}(n) = [\mathbf{w}_{\bar{\theta}_3}^H \mathbf{y}_1(n)\ \mathbf{w}_{\bar{\theta}_6}^H \boldsymbol{\eta}(n)]^T,$$
$$\mathbf{y}_1(n) = \mathbf{a}(\theta_1)s_1(n) + \boldsymbol{\eta}(n) \tag{18b}$$

$$H_{1,2} : \mathbf{x}(n) = [\mathbf{w}_{\bar{\theta}_3}^H \boldsymbol{\eta}(n)\ \mathbf{w}_{\bar{\theta}_6}^H \mathbf{y}_2(n)]^T,$$
$$\mathbf{y}_2(n) = \mathbf{a}(\theta_2)s_2(n) + \boldsymbol{\eta}(n) \tag{18c}$$

$$H_{2,2} : \mathbf{x}(n) = [\mathbf{w}_{\bar{\theta}_3}^H \mathbf{y}_1(n)\ \mathbf{w}_{\bar{\theta}_6}^H \mathbf{y}_2(n)]^T,$$
$$\mathbf{y}_1(n) = \mathbf{a}(\theta_1)s_1(n) + \boldsymbol{\eta}(n),$$
$$\mathbf{y}_2(n) = \mathbf{a}(\theta_2)s_2(n) + \boldsymbol{\eta}(n). \tag{18d}$$

where $\mathbf{x}(n)$ is the instantaneous received signal vector derived at the output of $\mathcal{B}$. Under this operating condition, it is assumed that only one of the four states in (18) can exist during any sensing duration and our goal is to declare spatial spectrum hole in the operating environment of any inactive PU(s). To address this multi-state signal detection problem, we seek to learn the peculiar attribute(s) of each state using the MSVM algorithms.

In general, the implementation of the MSVM classification technique can be approached in two ways. One way is the direct approach where multi-class problem is formulated as a single, large, all-in-one optimization problem that considers all support vectors at once [32]. However, the number of parameters to be estimated will increase as the number of classes to be discriminated increases. This method is also less stable which affects the classifier's performance [33]. The alternative approach which we adopted in this study is to treat the multi-state problem as multiple binary classification tasks. It requires the construction of multiple binary SVM models from the training data by using one-versus-all (OVA) or one-versus-one (OVO) methods [32], [34], [35]. In this application, the data set that were collected under each of the four hypotheses represents each system state and can also be viewed as the unique classes that we seek to be able to distinguish. In the OVA method, we construct $J = 2^P$ binary SVM models for J classes of interest where the j^{th} model, $\forall j \in J$ is trained on two classes of training set. The pair of classes are realized by assigning positive labels to all data points in the j^{th} class and negative labels to all remaining training data points. So, suppose we have D training data points in our training set, $\mathcal{S}$, where in this application $\boldsymbol{\vartheta}_i \in R^2$, $i = 1, \ldots, D$ and $\bar{l}_i \in \{1, \ldots, J\}$ is the class to which $\boldsymbol{\vartheta}_i$ belongs, the j^{th} SVM solves the optimization problem defined in the primal form as

$$\begin{aligned}
\underset{\vec{z}^j, b^j, \xi^j}{\text{minimize}} \quad & \langle \vec{z}^j . \vec{z}^j \rangle + \Gamma^+ \sum_{i|\bar{l}_i=j}^{D} \xi_i^j + \Gamma^- \sum_{i|\bar{l}_i\neq j}^{D} \xi_i^j \\
\text{subject to} \quad & (\langle \vec{z}^j . \beta(\boldsymbol{\vartheta}_i) \rangle + b^j) \geq 1 - \xi_i^j, \ \ \textit{if } \bar{l}_i = j \\
& (\langle \vec{z}^j . \beta(\boldsymbol{\vartheta}_i) \rangle + b^j) \leq -1 + \xi_i^j, \ \ \textit{if } \bar{l}_i \neq j \\
& \xi_i^j \geq 0, \quad i = 1, \ldots, D.
\end{aligned} \tag{19}$$

where as in (16), the training set are mapped into high dimensional feature plane via appropriately selected kernel function $\beta(\boldsymbol{\vartheta})$, Γ^+ and Γ^- are penalty parameters and $\sum_i \xi_i^j$ is an upper bound on the number of training errors. It should be noted that the original binary SVM soft margin objective function in (16) assigns equal cost, Γ to both positive and negative misclassifications in the penalty component. We have modified this in (19) in order to accommodate the imbalance in the number of training examples in the two classes arising under the OVA scheme. This is a cost sensitive learning approach for addressing the sensitiveness of the SVM algorithm to class imbalance training data [36]. Appropriate values for the penalty parameters can be obtained by setting the ratio $\frac{\Gamma^+}{\Gamma^-}$ to $\frac{card(C_{maj})}{card(C_{min})}$ where C_{maj} and C_{min} refer to the majority and minority class respectively. By solving the dual form of (19), we obtain the required J decision functions

$$\sum_{i=1}^{N_s^j} \bar{l}_i^j \alpha_i^j \beta(\boldsymbol{\vartheta}^{new}, \boldsymbol{\vartheta}_i^j) + b^j, \ \forall j \in J \tag{20}$$

for classifying any new data point, $\boldsymbol{\vartheta}^{new}$ and determining the corresponding PU network state.

The second learning strategy is the OVO method whereby each binary learner trains only on a pair of classes, $j_j, j_q \in J$. To apply this technique, all training examples in class j_j are taken as the positive class while those in the j_q class are treated as the negative class. All remaining examples, $J \setminus \tilde{j}$ where $\tilde{j} = \{j_j \cup j_q\}$ are simply ignored. Therefore, similar to (19) the formulation of the optimization problem in the primal form under the OVO scheme can be expressed as [32]

$$\begin{aligned}
\underset{\vec{z}^{jq}, b^{jq}, \xi^{jq}}{\text{minimize}} \quad & \langle \vec{z}^{jq} . \vec{z}^{jq} \rangle + \Gamma \sum_{i=1}^{\bar{D}} \xi_i^{jq} \\
\text{subject to} \quad & (\langle \vec{z}^{jq} . \beta(\boldsymbol{\vartheta}_i) \rangle + b^{jq}) \geq 1 - \xi_i^{jq}, \ \ \textit{if } \bar{l}_i = j \\
& (\langle \vec{z}^{j} . \beta(\boldsymbol{\vartheta}_i) \rangle + b^{jq}) \leq -1 + \xi_i^{jq}, \ \ \textit{if } \bar{l}_i = q \\
& \xi_i^{jq} \geq 0, \quad i = 1, \ldots, \bar{D}.
\end{aligned} \tag{21}$$

where $\bar{D} = card(\tilde{j})$. We solve the dual form of (21) for every possible $j-q$ pair and obtain $\binom{J}{2}$ SVM decision functions

$$\sum_{i=1}^{N_s^{jq}} y_i^{jq}\alpha_i^{jq}\beta(\boldsymbol{\vartheta}^{new}, \boldsymbol{\vartheta}_i^{jq}) + b^{jq}, \quad \forall j\ q\ pair \tag{22}$$

for classifying new data point, $\boldsymbol{\vartheta}^{new}$.

C. PREDICTING PRIMARY USERS' STATUS VIA ECOC BASED CLASSIFIER'S DECODING

Network state prediction via MSVM classifiers can be achieved using the decision acyclic graph (DAG) or ECOC approach [32]. While the DAG is suitable for decoding in OVO scheme, the ECOC can be used for both OVA and OVO schemes. The ECOC provides a framework that enables us to design the selection of the pairs of classes during training. It also allows us to take advantage of the dependencies among different labels and the predictions made by the individual binary classifier model towards minimizing overall classification error [35], [37]. For the OVA model in (19), we implement the ECOC by constructing a coding matrix, $\mathcal{M} \in \{+1, -1\}^{J\times L}$ where J is the number of classes in the complete training data set and L is the number of binary learners required to solve our multi-state problem. It should be noted that for the OVA method, $J = L$, and one way of constructing the coding matrix $\mathcal{M}$ is to choose a square, symmetric matrix with $+1s$ on the leading diagonal *only* so that $Tr(\mathcal{M}) = J$. The coding matrix for implementing the OVA scheme where $J = L = 4$ is shown in Table I. To solve our four-state problem, $J = 4$ decision models are required, thus we apply the OVA matrix in Table 1. During the learning stage, for the n^{th} learner, we assign positive label to all examples in the n^{th} row and negative label to all training examples in the remaining rows. For example, the first binary learner l_1 in the first column is trained by assigning the positive label to all the training examples in the j_1 class while all training examples in the j_2 through j_4 classes are assigned the negative label as shown.

TABLE 1. One-versus-all coding matrix.

	l_1	l_2	l_3	l_4
j_1	+1	-1	-1	-1
j_2	-1	+1	-1	-1
j_3	-1	-1	+1	-1
j_4	-1	-1	-1	+1 .

Similarly, in the case of the OVO scheme, we design the coding matrix, $\mathcal{M} \in \{+1, -1, 0\}^{J\times L}$ such that the n^{th} learner trains on two classes only. Assuming that this learner is used to train on the j^{th} and q^{th} classes, the rows of $\mathcal{M}$ that corresponds to the j^{th} and q^{th} classes of interest are labeled $+1$ and -1 respectively while all remaining rows are ignored by assigning them 0s. This procedure is repeated until all required $\binom{J}{2}$ classifier functions are realized. Table 2 shows the OVO coding matrix $\mathcal{M}$ for solving our four-class problem.

TABLE 2. One-versus-one coding matrix.

	l_1	l_2	l_3	l_4	l_5	l_6
j_1	+1	+1	+1	0	0	0
j_2	-1	0	0	+1	+1	0
j_3	0	-1	0	-1	0	+1
j_4	0	0	-1	0	-1	-1 .

It could be observed that in the two ECOC strategies described above, each class in the training set is associated with a row of $\mathcal{M}$ thereby constituting a unique codeword for each class. To decode the class for a new (test) data point, we adopt the loss-weighted decoding strategy [34] which takes into account the confidence level that each binary classifier attaches to its prediction as calculated from the actual score obtained from the decision function in (17). In loss-weighted decoding, if the set of predicted scores for a new observation, $\boldsymbol{\vartheta}^{new}$ jointly returned by all MSVM decision models is denoted by

$$\bar{\Omega}(\boldsymbol{\vartheta}^{new}) = \left\{\bar{\omega}_{l_1}(\boldsymbol{\vartheta}^{new}), \bar{\omega}_{l_2}(\boldsymbol{\vartheta}^{new}), \cdots, \bar{\omega}_{l_L}(\boldsymbol{\vartheta}^{new})\right\} \tag{23}$$

for a set of L learners, where $\bar{\omega}_{l_l}(\boldsymbol{\vartheta}^{new})$ is the actual predicted score produced by the decision model of the l^{th} learner, then we classify a test data point as belonging to the $\hat{j} \in \{1, \ldots, J\}$ class that offers the minimum sum of binary losses over all learners. This is obtained by using [37]

$$\hat{j} = \underset{j}{\operatorname{argmin}} \frac{\sum_{l=1}^{L} |m_{jl}| g(m_{jl,\theta_{l_l}})}{\sum_{l=1}^{L} |m_{jl}|} \tag{24}$$

where $\hat{j}$ is the predicted class for the test data point, m_{jl} refers to the element jl of $\mathcal{M}$, (i.e. the label for class j of learner l) and $g(.,.)$ is an appropriate binary loss function specifically chosen for the SVM classifier. A good choice for the binary loss is the hinged function whose score domain lies in $(-\infty, \infty)$, and is defined by

$$g(\bar{l}_l, \theta_l) = \frac{max(0, 1 - \bar{l}_l\theta_{l_l})}{2} \tag{25}$$

where $\bar{l}_l$ is the class label for the l^{th} binary learner of the class being considered. In this study, we used (24) to obtain the class of test data points which corresponds to the true state of the PU activities that we wish to predict during the sensing interval. In Algorithm 2, the procedure for solving our multiple PUs sensing problem using the proposed beamformer derived feature based ECOC multi-class SVM is presented. An alternative approach based on the proposed MIMSVM method is also presented in Algorithm 3.

VI. NUMERICAL RESULTS AND DISCUSSION

We evaluate the performance of the proposed beamformer-aided SVM algorithms for single and multiple PUs' scenarios. The CSVM algorithm was applied under the single PU considerations while the MSVM and MIMSVM algorithms were implemented for the multiple PUs scenarios. For the beamformer design, we set $\Theta \in [-90°, 90°]$, K equals 9, and sector width, $\bar{\theta}_k$ equals 20° $\forall\ k \in K$. Furthermore,

Algorithm 2 Beamformer Aided ECOC MSVM Algorithm for Spectrum Sensing Under Multiple PU Scenarios

Learning process:
Qualification stage
1.) Load $\mathbf{w}_{\bar{\theta}_k}, \forall k \in K$ and scan Θ,.
2.) Compute $\boldsymbol{\vartheta} \in R^K$, under $\mathcal{S}^P_{Q(P)}$ using (10).
3.) Apply ζ_1 in (11) to $\boldsymbol{\vartheta}$ in (*ii*) to find $\mathcal{B}$.
Training stage
4.) Obtain set $\{D\}$ of $\boldsymbol{\vartheta} \in R^{\bar{S}}$ from $\mathcal{B}$ in (*iii*) under H_0, $H_{i,q}$, i.e. $\forall \mathcal{S}^i_q \in \mathcal{C}_i, \forall \mathcal{C}_i \in \mathcal{P}$ using (10).
5.) Generate J decision models in (20) or (22) from $\{D\}$ in (*iv*).
Prediction process:
do repeat
6.) Obtain test sample $\boldsymbol{\vartheta}^{new} \in R^{\bar{S}}$ using (10).
7.) Classify $\boldsymbol{\vartheta}^{new}$ in (*vi*) using (*v*) to decide system state, H_0 or $H_{i,q}$.

Algorithm 3 Beamformer Aided MIMSVM Algorithm for Spectrum Sensing Under Multiple PU Scenarios

Learning process:
Qualification stage
1.) Load $\mathbf{w}_{\bar{\theta}_k}, \forall k \in K$ and scan Θ.
2.) Compute $\boldsymbol{\vartheta} \in R^K$, under $\mathcal{S}^P_{Q(P)}$ using (10).
3.) Apply ζ_1 in (11) to $\boldsymbol{\vartheta}$ in (*ii*) to find $\mathcal{B}$.
4.) **for** $i = 1$ to P, **do**
5.) Identification by computing $\mho_d$ in (13) $\forall$ pair $\{b^i_{ref}, b_{\bar{s}}\}, \forall b_{\bar{s}} \in \mathcal{B}, b^i_{ref} \neq b_{\bar{s}}$.
6.) Associate $\{b^i_{ref}, b_{\bar{s}}\} \subset \mathcal{B}$ with PU_i if $\mho_d(b^i_{ref}, b_{\bar{s}}) \geq \zeta_2$.
7.) **end for**
Training stage
8.) **for** $i = 1$ to P, **do**
9.) Obtain set $\{D\}$ for PU_i in (*vi*) under H_0 and H_1 using (10).
10.) Generate independent models in (17) from $\{D\}$ in (*ix*).
11.) **end for**
Prediction process:
12.) $\forall i \in P$, **do repeat**
13.) Obtain test sample $\boldsymbol{\vartheta}^{new} \in R^{\bar{S}}$ using (10).
14.) Classify $\boldsymbol{\vartheta}^{new}$ (*xiii*) using corresponding decision model in (*x*) and decide the PUs' state, H_0 or H_1.

to apply the decision threshold ζ_1, target false alarm probability, $\bar{P}_{fa}$ was set to 0.01. The results are quantified in terms of probability of detection, probability of false alarm, receiver operating characteristics (*ROC*), area under ROC curve (*AuC*) and overall classification accuracy (CA_{ovr}).

A. SINGLE PRIMARY USER SCENARIO

Under this scenario, our aim is simply to detect the presence or absence of the PU. For the purpose of simulation, under H_1 we assumed that the PU signal is *BPSK* modulated. We further consider that during the sensing interval, the transmission is multipath propagated and received at the SU via two strong components at AOAs, $\theta \in [-45°, -35°] \in \bar{\theta}_3$ and $\theta \in [15°, 20°] \in \bar{\theta}_6$. The delay between the arrival of the two multipath components is assumed to be 5 symbols and the total received power is normalized to unity. It is further assumed that the noise is circularly symmetric complex additive white Gaussian with power, σ^2_η. The PU's signal and the noise are assumed to be uncorrelated. To investigate the performance of the resulting beamformer derived feature vector, we set $\gamma = 0$ *dB* and applied the CSVM using linear kernel with box constraint, $\Gamma = 0.9$. We generated 2000 sets of energy vectors through random realizations of the channels, out of which 400 were used for training and the rest for testing purpose. The number of antennas at the SU is assumed to be $M = 8$ with spacing $d = 0.5\lambda$. Fig. 1 shows the performance of the proposed beamformer based SVM binary classifier (BFSVM) in terms of the ROC curves for fixed number of received signal samples, $N = 500$ when the $SNR = -15$ *dB*, -18 *dB* and -20 *dB* in comparison with the alternative in which the use of beamformers is not considered, that is, the non-beamformer based scheme (NBFSVM). As expected, the BFSVM scheme takes advantage of the beamforming array gain of $10\log_{10} M$ *dB* and thus exhibits significant performance improvement compared to the NBFSVM scheme. Specifically, at $Pfa = 0.1$, the Pd achieved by the BFSVM scheme is about 0.99 whereas the NBFSVM achieves about 0.74 when SNR is -15 *dB*. In terms of *AuC*, the BFSVM scheme yields 0.9933 while NBFSVM offers 0.9149 at SNR of -15 *dB*. Similar trend can be observed in all cases of SNR considered. The proposed BFSVM scheme consistently outperforms the NBFSVM scheme which demonstrates the potential of the beamformer derived features to enhance the capability of the SVM binary classifier. It is strikingly interesting to note that the dimension of the feature vector of the BFSVM scheme in the input space is far less than that of the NBFSVM scheme which indicates that the proposed scheme offers significant reduction in implementation complexity. In Fig. 2, we show the effect of varying the number of PU signal samples, N on the performance of the proposed scheme where the SNR is kept at -20 *dB*. As seen, when N is increased from 500 to 2000 and Pfa is 0.1, about 40% improvement in performance is observed for the BFSVM scheme, where Pd is increased from 0.45 to 0.85. On the other hand, the NBFSVM method yields only about 24% improvement, i.e. Pd is increased from 0.25 to 0.49. Furthermore, given the same Pfa and $N = 2000$, the proposed scheme attains the Pd of 0.85 against 0.49 yielded by the NBFSVM alternative. Similarly, from the *AuC* perspective, an increase from 0.7926 to 0.9487 and from 0.6635 to 0.8094 is observed for the BFSVM and NBFSVM respectively for fixed SNR of -20 *dB* as N is increased.

We conclude our investigation on the single PU scenario, in Fig. 3 where we evaluate the impact of receive SNR on

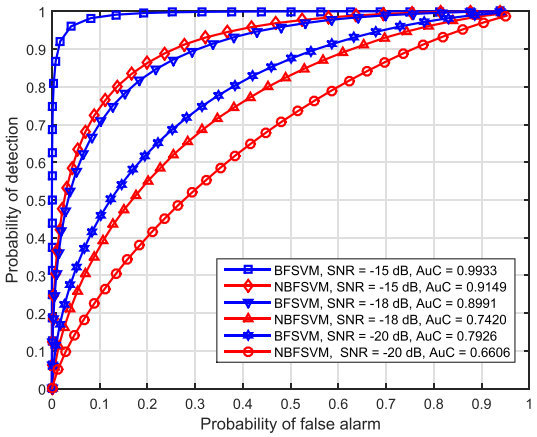

FIGURE 1. ROC performance comparison between beamformer based and non-beamformer based SVM schemes under different *SNR*, number of PU, $P = 1$ and number of samples, $N = 500$.

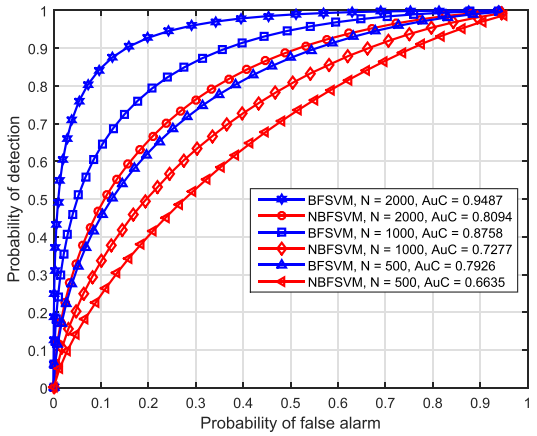

FIGURE 2. ROC performance comparison between beamformer based and non-beamformer based SVM schemes with different number of samples, *N*, and $SNR = -20$ *dB*.

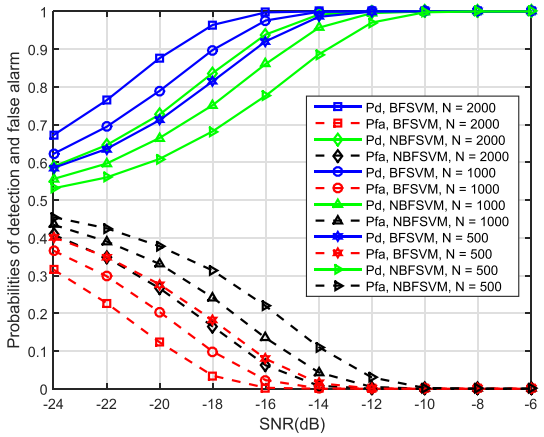

FIGURE 3. Performance comparison between beamformer based and non-beamformer based SVM schemes showing probabilities of detection and false alarm versus *SNR*, with different sample number, *N*.

Pd and *Pfa* and compare both metrics under the BFSVM and NBFSVM schemes. As expected, the performance of both schemes improves as *SNR* is increased. However, the proposed BFSVM scheme outperforms the NBFSVM scheme as seen for example at *SNR* of -20 *dB* where the BFSVM scheme attains *Pd* of about 0.88 when $N = 2000$ and *Pfa* ≈ 0.1. On the other hand, the NBFSVM only attains *Pd* of about 0.72 and *Pfa* of about 0.28. Furthermore, as *N* is increased from 500 to 2000 and at *SNR* of -20 *dB*, *Pd* rises in the case of BFSVM scheme from 0.7 to 0.88 (about 18% gain) while *Pfa* reduces from 0.28 to 0.12 (about 16% drop) whereas, for the NBFSVM, rise in *Pd* is from 0.6 to about 0.72 (12 % gain) and *Pfa* reduces from 0.39 to 0.28 (about 11% drop). The performance of the proposed BFSVM scheme at *N* equals 500 almost matches that of the NBFSVM scheme at *N* equals 2000 indicating some savings in sensing time for the same performance level. From the foregoing, it is evident that the proposed beamformer based scheme exhibits a superior performance in terms of improving the usage of the radio spectrum resources and reduced implementation complexity in comparison with the non beamformer based alternative.

B. MULTIPLE PRIMARY USER SCENARIO

We investigate the performance of the beamformer-aided scheme using the ECOC MSVM and MIMSVM algorithms with energy features and quantify the results in terms of CA_{ovr}. We consider a network comprising two PUs operating in the frequency band of interest and transmitting with a specific power such that the *SNR* at the receiver is 0 *dB* and -2 *dB* respectively. The channel coefficients have also been normalized to one. Further, we considered two scenarios for the angle of arrival signals. In the first scenario, the signal from PU1 is received from two paths, the first one arrives with an AOA, $\theta \in [-45°, -35°] \in \bar{\theta}_3$ and the second path comes with an AOA, $\theta \in [15°, 20°] \in \bar{\theta}_6$. Similarly, the two multipath components of PU2 arrive at angles $\theta \in [-20°, -15°] \in \bar{\theta}_4$ and $\theta \in [40°, 45°] \in \bar{\theta}_7$ respectively. In the second scenario, we consider a situation where the multipath components of the first PU arrive with AOAs, $\theta \in [-45°, -35°] \in \bar{\theta}_3$ and $\theta \in [15°, 20°] \in \bar{\theta}_6$. However, for the second PU, they arrive at $\theta \in [15°, 20°] \in \bar{\theta}_6$ and $\theta \in [40°, 45°] \in \bar{\theta}_7$. It means that the beamformer corresponding to $\bar{\theta}_6$ picks up signals from both PUs. We call this scenario the overlapping multipath case. Hence, the first scenario is non overlapping. For each PU, the multipath components received via the two distinct paths are assumed to arrive the receiver with a delay of 5 symbols.

Furthermore, by cross-validation, the SVM box constraint parameter, Γ is 1 and the Gaussian kernel scaling factor, σ is 10. However, when implementing the OVA scheme, the corresponding values for box constraint parameters, Γ^+ and Γ^- are obtained as the ratio of the pair of classes as discussed in section V-B. We generated 2000 sets of energy vectors through random channel realization, trained with 400 of them and used the rest for testing. In Fig. 4, we inves-

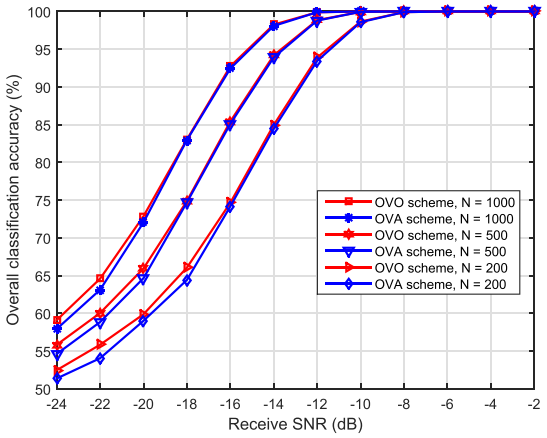

FIGURE 4. Performance comparison between OVO and OVA ECOC MSVM schemes under non-overlapping transmission scenario with different number of samples, N and number of PU, $P = 2$.

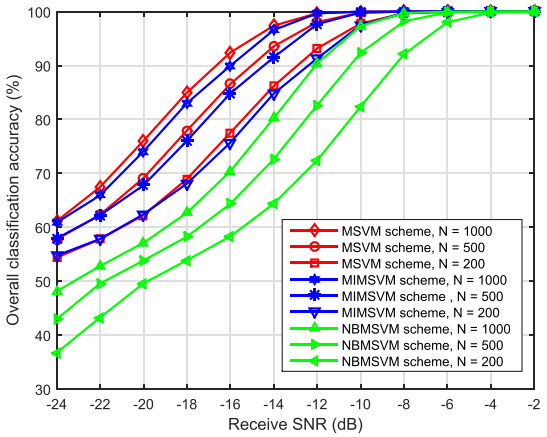

FIGURE 5. Performance comparison of OVO MSVM, MIMSVM and OVO NBMSVM schemes under LOS transmission scenario with different number of samples, N and number of PU, $P = 2$.

tigate the suitability of the OVO and OVA coding techniques by evaluating their performance in terms of CA_{ovr} at different receive SNR using the ECOC MSVM algorithm. The performance evaluation under the non-overlapping transmission scenario indicates that for both schemes, the CA_{ovr} improves as the SNR is increased. For example, when N is 1000, CA_{ovr} increases from about 58% to 100% as SNR is raised from $-24\ dB$ to $-8\ dB$. Similar trend can be observed for various N values. The OVO scheme appeared to slightly outperform the OVA scheme especially in the very low SNR regime. At any rate, in deciding which coding scheme to use, the system's complexity in terms of the number of classifiers required to be constructed by each method, the memory requirement and the training as well as testing time should be considered. In Fig. 5, under the LOS transmission scenario, we investigate the performance of our beamformer aided MIMSVM and ECOC MSVM algorithms over a range of SNR and compare these with the non beamformer based alternative (NBMSVM). As seen, the beamformer aided schemes significantly outperform the NBMSVM. For instance, when $N = 1000$, for the beamformer aided schemes, CA_{ovr} improves from about 60% to 100% when the SNR is increased from $-24\ dB$ to $-12\ dB$, whereas for the NBMSVM scheme, CA_{ovr} only increased from about 48% to 90% for the same SNR increment. In addition, the OVO ECOC MSVM is seen to slightly outperform its MIMSVM counterpart over a considerable portion of the SNR range and for all cases of N.

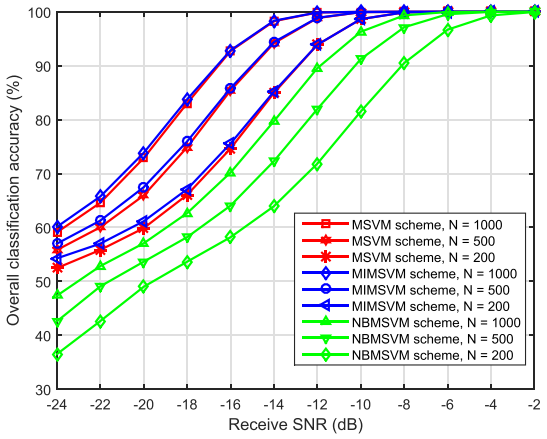

FIGURE 6. Performance comparison of OVO-MSVM, MIMSVM and OVO-NBMSVM schemes under non-overlapping reflection scenario with different number of samples, N and number of PU, $P = 2$.

In Fig. 6 and Fig. 7, we examine the performance of the MIMSVM, ECOC MSVM and NBMSVM schemes under non-overlapping and overlapping transmission scenarios. Both results show that the performance of the three schemes is similar to that seen for the LOS scenario where CA_{ovr} is observed to improve as the receive SNR is increased. It could however be noticed in these two cases, that in addition to offering far less computational complexity, the MIMSVM slightly outperforms the OVO based ECOC MSVM scheme especially in the very low SNR regime. This may largely be due to the fact that the MIMSVM scheme benefits from increase in the dimension of its feature space under these two scenarios.

Furthermore, it can be seen that both the MIMSVM and ECOC MSVM schemes perform equally well and consistently outperform the NBMSVM under the cognitive radio deployment scenarios described in section IV, thereby further lending credence to the robustness of the proposed beamformer based learning approach. Next, in Fig. 8 we show the comparison between the OVO ECOC MSVM and the DAG SVM methods. It is observable here that the ECOC MSVM performs better than the DAGSVM in the low SNR regime. This can be seen for instance at the SNR of $-24\ dB$ where as N is increased from 200 to 1000, we see that the CA_{ovr} rises from about 52% to 60% for the ECOC MSVM whereas in the case of DAGSVM, the rise in CA_{ov} is approximately from 46% to 55%. In Fig. 9, we show the comparison between the SVM and kNN classification techniques where both non-parametic methods are considered under beamformer based multiclass OVO ECOC scheme over a range of SNR and

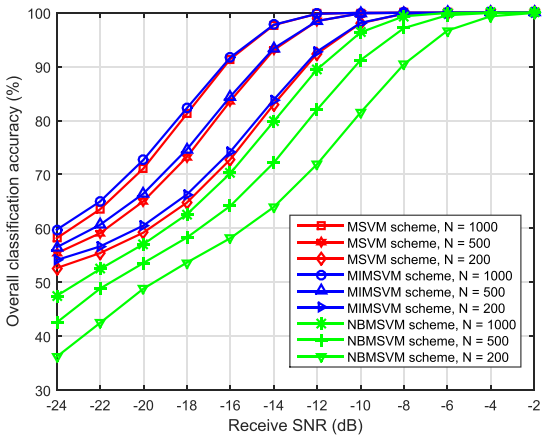

FIGURE 7. Performance comparison of OVO MSVM, MIMSVM and OVO NBMSVM schemes under overlapping reflection scenario with different number of samples, N and number of PU, $P = 2$.

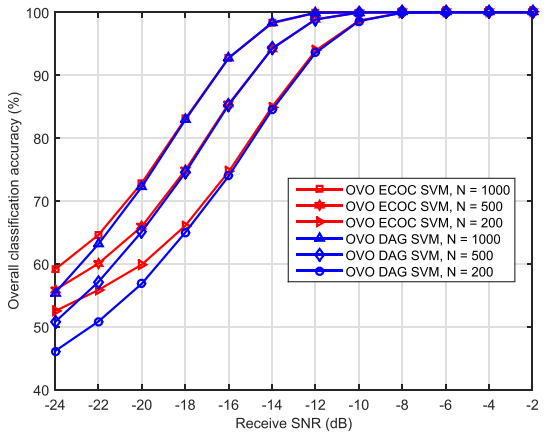

FIGURE 8. Performance comparison between OVO ECOC and DAG based MSVM under non-overlapping reflection scenario with different number of samples, N and number of PU, $P = 2$.

different N. As seen, the SVM consistently outperforms the kNN.

Finally, we conclude our investigation by computing the execution time required to implement the proposed algorithms. In Table III, it can be seen that on the average, it takes about 1.0614s, 2.2492s and 2.2593s to execute the *learning process* section of **Algorithm** 1, 2 and 3 respectively, when $N = 1000$ samples. Similarly, in Table IV, it can be observed that about 0.962ms, 1.200ms and 0.993ms are required to execute the *prediction process* section of **Algorithm** 1, 2 and 3 respectively, with $N = 1000$. The time evaluation was carried out using Matlab 2016a on a core $i5$ 5200U, 2.20 GHz computer with 8.00 GB RAM installed.

In summary, all simulation results indicate that the proposed, beamformer aided scheme offers significant advantage for SVM classifier in solving spectrum sensing problem given both single and multiple primary user scenarios in multi-antenna cognitive radio networks.

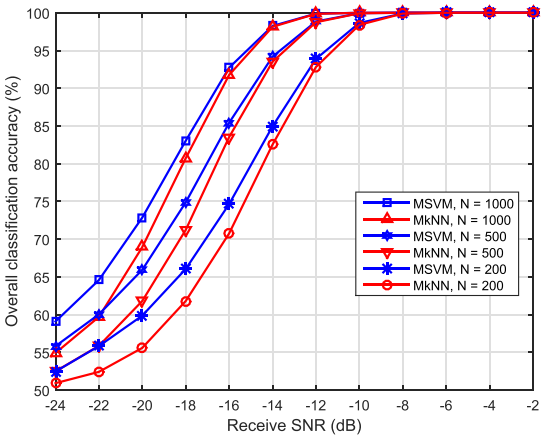

FIGURE 9. Performance comparison of OVO based MSVM and MkNN techniques with different number of samples, N, number of neighbor = 5, and number of PU, $P = 2$.

TABLE 3. Average Learning Duration (Seconds).

No of Samples	100	200	300	400	500	1000
Algorithm 1	0.2752	0.3809	0.5086	0.6012	0.7677	1.0614
Algorithm 2	0.7652	0.9473	1.1430	1.3663	1.5400	2.2492
Algorithm 3	1.0423	1.1683	1.2786	1.4595	1.5682	2.2593

TABLE 4. Average Prediction Duration ($\times 10^{-4}$ Seconds).

No of Samples	100	200	300	400	500	1000
Algorithm 1	1.6415	2.7141	3.8155	4.8259	6.3709	9.6242
Algorithm 2	2.3712	3.4514	4.4584	5.7320	6.4769	12.000
Algorithm 3	1.4264	2.4224	3.3629	4.2431	5.7965	9.9335

VII. CONCLUSION

In this paper, we proposed and investigated the performance of beamformer aided support vector machine algorithms for spectrum sensing in multi-antenna cognitive radio networks. In particular, we have developed algorithms for multiple hypotheses testing facilitating joint spatio-temporal spectrum sensing. Using energy features and the error correcting output codes technique, the key performance metrics of the classifiers were evaluated which demonstrate the superiority of the proposed methods over previously proposed alternatives.

ACKNOWLEDGMENT

This work has been partially supported by the Petroleum Technology Development Fund (PTDF) of Nigeria and the Engineering and Physical Science Research Council (EPSRC) of UK under grant EP/M015475/1. Partial results of this paper were presented at the IEEE ICSPCS, Cairns, Australia, December 2015 [1].

REFERENCES

[1] O. P. Awe and S. Lambotharan, "Cooperative spectrum sensing in cognitive radio networks using multi-class support vector machine algorithms," in *Proc. 9th Int. Conf. Signal Process. Commun. Syst.*, Cairns, QLD, Australia, Dec. 2015, pp. 1–7.

[2] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[3] T. Yucek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Commun. Surveys Tuts.*, vol. 11, no. 1, pp. 116–130, 1st Quart., 2009.

[4] S. Haykin, D. J. Thomson, and J. H. Reed, "Spectrum sensing for cognitive radio," *Proc. IEEE*, vol. 97, no. 5, pp. 849–877, May 2009.

[5] 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.

[6] K. Thilina, K. W. Choi, N. Saquib, and E. Hossain, "Machine learning techniques for cooperative spectrum sensing in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 11, pp. 2209–2221, Nov. 2013.
[7] O. P. Awe, S. M. R. Naqvi, and S. Lambotharan, "Variational Bayesian learning technique for spectrum sensing in cognitive radio networks," in *Proc. 2nd IEEE Global Conf. Signal Inf. Process.*, Atlanta, GA, USA, Dec. 2014, pp. 1353–1357.
[8] O. P. Awe, S. M. Naqvi, and S. Lambotharan, "Kalman filter enhanced parametric classifiers for spectrum sensing under flat fading channels," in *Cognitive Radio Oriented Wireless Networks*, M. Weichold *et al.*, Eds. Springer, 2015, pp. 235–247.
[9] T. O'Shea, J. Corgan, and T. C. Clancy. (Feb. 2016). "Convolutional radio modulation recognition networks." [Online]. Available: https://arxiv.org/abs/1602.04105
[10] E. West and T. O'Shea. (Mar. 2017). "Deep architectures for modulation recognition." [Online]. Available: https://arxiv.org/abs/1703.09197
[11] X. Zhu and T. Fujii, "Modulation classification for cognitive radios using stacked denoising autoencoders," *Int. J. Satellite Commun. Netw.*, vol. 35, no. 5, pp. 517–531, 2017.
[12] S. Rajendran, W. Meert, D. Giustiniano, V. Lenders, and S. Pollin. (Jul. 2017). "Distributed deep learning models for wireless signal classification with low-cost spectrum sensors." [Online]. Available: https://arxiv.org/abs/1707.08908
[13] G. J. Mendis, J. Wei, and A. Madanayake, "Deep learning-based automated modulation classification for cognitive radio," in *Proc. IEEE Int. Conf. Commun. Syst.*, Shenzhen, China, Dec. 2016, pp. 1–6.
[14] K. Karra, S. Kuzdeba, and J. Petersen, "Modulation recognition using hierarchical deep neural networks," in *Proc. IEEE Int. Symp. Dyn. Spectr. Access Netw.*, Piscataway, NJ, USA, Mar. 2017, pp. 1–3.
[15] W. Lee, M. Kim, D. Cho, and R. Schober. (May 2017). "Deep sensing: Cooperative spectrum sensing based on convolutional neural networks." [Online]. Available: https://arxiv.org/abs/1705.08164
[16] C. Zhao, M. Sun, B. Li, L. Zhao, and X. Peng, "Blind spectrum sensing for cognitive radio over time-variant multipath flat-fading channels," *EURASIP J. Wireless Commun. Netw.*, p. 84, May 2014.
[17] L. Wei and O. Tirkkonen, "Spectrum sensing in the presence of multiple primary users," *IEEE Trans. Commun.*, vol. 60, no. 5, pp. 1268–1277, May 2012.
[18] S.-J. Kim, E. Dall'Anese, and G. B. Giannakis, "Cooperative spectrum sensing for cognitive radios using Kriged Kalman filtering," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 1, pp. 24–36, Feb. 2011.
[19] D. Čabrić and M. Erić, "Spatio-temporal spectrum sensing using distributed antenna systems and direct localization methods," in *Proc. IEEE Antennas Propag. Soc. Int. Symp.*, Chicago, IL, USA, Jul. 2012, pp. 1–2.
[20] Q. Wu, G. Ding, J. Wang, and Y. D. Yao, "Spatial-temporal opportunity detection for spectrum-heterogeneous cognitive radio networks: Two-dimensional sensing," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 516–526, Feb. 2013.
[21] X. Lin, J. G. Andrews, and A. Ghosh, "Spectrum sharing for device-to-device communication in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 12, pp. 6727–6740, Dec. 2014.
[22] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Englewood Cliffs, NJ, USA: Prentice Hall, 2005.
[23] A. Deligiannis, J. A. Chambers, and S. Lambotharan, "Transmit beamforming design for two-dimensional phased-MIMO radar with fully-overlapped subarrays," in *Proc. Sensor Signal Process. Defence Conf.*, Edinburgh, U.K., Sep. 2014, pp. 1–6.
[24] A. Deligiannis, S. Lambotharan, and J. A. Chambers, "Game theoretic analysis for MIMO radars with multiple targets," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 6, pp. 2760–2774, Dec. 2016.
[25] Y.-C. Liang, Y. Zeng, E. C. Y. Peh, and A. T. Hoang, "Sensing-throughput tradeoff for cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 4, pp. 1326–1337, Apr. 2008.
[26] O. P. Awe, "Machine learning algorithms for cognitive radio wireless networks," Ph.D. dissertation, Loughborough Univ., Loughborough, U.K., Nov. 2015.
[27] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[28] C. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.
[29] M. A. Davenport, R. G. Baraniuk, and C. D. Scott, "Tuning support vector machines for minimax and neyman-pearson classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 10, pp. 1888–1898, Oct. 2010.
[30] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discovery*, vol. 2, no. 2, pp. 121–167, 1998.
[31] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[32] 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.
[33] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.
[34] E. L. Allwein, R. E. Schapire, and Y. Singer, "Reducing multiclass to binary: A unifying approach for margin classifiers," *J. Mach. Learn. Res.*, vol. 1, pp. 113–141, Sep. 2001.
[35] T. Dietterich and G. Bakiri, "Solving multiclass learning problems via error-correcting output codes," *J. Artif. Intell. Res.*, vol. 2, no. 1, pp. 263–286, 1995.
[36] R. Batuwita and V. Palade, "Class imbalance learning methods for support vector machines," in *Imbalanced Learning: Foundations, Algorithms, and Applications*, H. He and Y. Ma, Eds. Hoboken, NJ, USA: Wiley, 2013, pp. 83–96.
[37] S. Escalera, O. Pujol, and P. Radeva, "Separability of ternary codes for sparse designs of error-correcting output codes," *Pattern Recognit. Lett.*, vol. 30, no. 3, pp. 285–297, 2009.

OLUSEGUN PETER AWE received the Ph.D. degree in electronic, electrical and systems engineering from Loughborough University, U.K., in 2015. He is currently a Lecturer with the Department of Electronic and Electrical Engineering, Obafemi Awolowo University, Ife, Nigeria. His current research interests include machine learning and applications, target tracking, cognitive radio networks, cognitive machine-to-machine communications, and HAP-based communication systems.

ANASTASIOS DELIGIANNIS received the Diploma (bachelor's and master's degrees equivalent) degree from the School of Electrical and Computer Engineering, University of Patras, Greece, in 2012, and the Ph.D. degree in radar signal processing from the Signal Processing and Networks Research Group, Wolfson School of Mechanical, Manufacturing and Electrical Engineering, Loughborough University, U.K., in 2016. Since 2016, he has been a Research Associate in signal processing with Loughborough University. His research focuses on signal processing algorithms, sparse array design, convex optimization, and game theoretic methods, within the radar network framework and wireless communications.

SANGARAPILLAI LAMBOTHARAN received the Ph.D. degree in signal processing from Imperial College London, U.K., in 1997. In 1996, he was a Visiting Scientist with the Engineering and Theory Centre of Cornell University, USA. From 1997 to 1999, he was a Post-Doctoral Research Associate with the Imperial College London. From 1999 to 2002, he was with the Motorola Applied Research Group, U.K., where he investigated various projects, including physical-link layer modelling and performance characterization of GPRS, EGPRS, and UTRAN. From 2002 to 2007, he was a Lecturer and Senior Lecturer with Kings College London and Cardiff University, respectively. He is currently a Professor of digital communications and the Head of the Signal Processing and Networks Research Group, Wolfson School Mechanical, Electrical and Manufacturing Engineering, Loughborough University, U.K. His current research interests include 5G networks, MIMO, radars, smart grids, machine learning, network security and convex optimizations, and game theory. He has published approximately 200 technical journal and conference articles in these areas.