Received: May 23, 2023. Revised: September 9, 2023.

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Optimizing Spectrum Sensing in Cognitive Radio Networks Using Bayesian-Optimized Random Forest Classifier

Raghavendra L R[1]* **Manjunatha R C[1]**

[1]*Global Academy of Technology, Visvesveraya Technological University, Belagavi, 590018, India*
* Corresponding author's Email: raghavendra.lr52@gmail.com

Abstract: Achieving high spectral efficiency is a fundamental objective in cognitive radio (CR) networks, where fast and precise spectrum sensing plays a vital role. The cognitive radios rely on spectrum sensing to detect available frequencies, identify idle channels, and utilize them for transmitting data. A new approach to spectrum sensing (SS) is presented in this study, using a machine learning-based model. The model leverages different techniques for feature extraction and separates redundant and non-redundant features to enhance detection performance and reduce training overhead. A hybrid approach is introduced, combining energy, MME, and cyclostationary features extracted from sample data with NCA. The selected features are then trained with a Bayesian-optimized random forest classifier (RF) as a machine learning model. The proposed algorithms are evaluated in terms of average training time, classification speed, and accuracy. The experimental results demonstrate that the proposed bayesian-optimized random forest classifier (RF) achieved a probability of detection of 0.94.

Keywords: Bayesian optimization, Cognitive radio, Energy detection, MME, Neighborhood component analysis, Random forest.

1. Introduction

The international telecommunications union (ITU) is responsible for managing technology, information, and telecommunication affairs on a global level, including coordinating the use of the electromagnetic spectrum. Despite this, each country has its own regulatory bodies, such as the federal communications commission (FCC), that govern the use of this valuable resource and exercise sovereign authority over it.

A cognitive radio user is a technology that observes and determines whether or not the main user is present. If the main user (PU) is not present, the open spectrum must be used by the second user [1]. Due to key user identification, this is not as dependable for the ordinary user as many dimming alternatives [2]. The source of the problem is secondary user access to the primary user licencing range as a result of this issue. This question was posed in order to increase detection accuracy in order to overcome shared spectrum perception issues. The fundamental idea behind collaborative spectrum sensor performance is to allow secondary users to increase performance [3, 4].

Outside of voting rule N, the cognitive recognition spectrum is specified. At the merger centre, when at least secondary users employ an external SU to identify primary users N [5]. Within seconds, consumers are raising their power usage to reveal the fusion center's (FC) spectral sensitivity and sensitivity. Plans for energy storage have been developed to increase energy efficiency [6, 7]. Spectrum allocation can minimise time and power usage if the SNR is too high or the main user (PU) is too high. If this is not the case, the spectrum detection sensitivity will be employed again to increase performance. A paradigm for optimising power usage is presented [8, 9], with time and transmission time recorded. As a result, PU improves energy efficiency by decreasing interference. In order to enhance energy efficiency, many well-known channels for efficient input recognition [10] have been proposed.

Each secondary user knows the channel in some

modes, yet the same second user FC and other secondary users can transmit the same message [11]. The authors of [12] presents a dependable high energy threshold circuit for improving spectral energy detection capability. The restriction in [13] was presented as adaptive, based on the effect of the SU transmit power. The detection threshold is solely determined by the statistical features of the received signal, as explained in [14]. The authors get the appropriate threshold value in [15] by using the Lagrange multiplier approach. The authors of [13] presents a two-threshold approach that varies from the standard one-threshold design and greatly enhances detection performance. More time is necessary to detect the spectrum before receiving results in [16], but in [17], the authors suggest a technique for generating an adaptive threshold value for an unlicensed 2.4 GHz WLAN channel so that this approach may be used in practise, particularly with cellular sensors. The author specifies the maximum number of items in [18]. If the length of time spent detecting a spectrum exceeds the upper limit, the SU will switch to detecting a new spectrum. The authors of reported energy detection-based spectrum sensing based on the adaptive threshold spectrum energy detection approach in [19].

The contribution of the proposed approach for spectrum sensing with energy detection, min-max eigenvalue, and cyclostationary features using feature selection through neighborhood component analysis (NCA) with Bayesian optimized random forests in cognitive radio can be summarized as follows:

Improved accuracy: The proposed approach utilizes a combination of energy detection, min-max eigenvalue, and cyclostationary features with feature selection through NCA and Bayesian optimized random forests to achieve improved classification accuracy. The use of multiple features and feature selection ensures that only the most informative features are used for classification, thereby improving the accuracy of the spectrum sensing process.

Robustness: By incorporating multiple features and utilizing the random forest algorithm, the suggested method is capable of handling variations in the received signal, such as noise and interference. Due to its resistance to overfitting and ability to manage noisy and incomplete data, the random forest algorithm is particularly suitable for the cognitive radio environment.

Efficiency: The proposed approach is computationally efficient due to the use of feature selection through NCA and Bayesian optimization, which allows for the selection of the most informative features while reducing the dimensionality of the input data. This results in faster training and testing times, making the proposed approach suitable for real-time spectrum sensing applications.

Flexibility: The proposed approach can be adapted to different cognitive radio environments by adjusting the input features and hyperparameters of the random forest algorithm. This allows for the optimization of the algorithm for different frequency bands and environments, making it more versatile and adaptable than traditional spectrum sensing (SS) techniques.

Overall, the proposed approach for spectrum sensing with energy detection, min-max eigenvalue, and cyclostationary features using feature selection through NCA with Bayesian optimized random forests in cognitive radio offers improved accuracy, robustness, efficiency, and flexibility compared to traditional spectrum sensing techniques. These benefits make it a promising approach for future cognitive radio applications.

The study begins by providing a comprehensive literature review in section 2, highlighting the relevant research in the field. The materials and methods used in the research paper are presented in section 3. In section 4, the proposed methods are explained in detail. The results of the MATLAB-based simulation are presented and analyzed in section 5. Finally, the paper concludes with a summary of the findings and conclusions in section 6.

2. Literature review

The widespread utilization of machine learning (ML) approaches within cognitive radio (CR) systems is evident. A variety of methods are utilized in cognitive radio systems, serving a range of purposes, which include signal classification, feature detection, power allocation, rate adaptation, reconfiguration of system parameters, and CSS (cooperative spectrum sensing). To cater to the specific prerequisites of different CR tasks, machine learning techniques such as SVM (support vector machines), ANN (artificial neural networks), the dirichlet process mixture model, and reinforcement learning (RL) are employed.

A comprehensive overview of ML techniques used for various CR tasks is presented in [11]. Depending on the unique needs, the authors propose that different machine learning algorithms may be used to various cognitive radio (CR) tasks. For Markov decision processes and game theory-based learning, reinforcement learning (RL) is proposed as the optimal approach. The dirichlet process mixture model, a non-parametric learning approach, is

deemed ideal for signal classification and feature recognition since it does not require previous knowledge about the number of mixture components. Artificial neural networks (ANNs), on the other hand, do not require previous information about the distribution of the observed process. The SVM approach outperforms ANNs for short training samples, but it requires previous knowledge about the observed process's distribution and data labelling. Theory-based learning algorithms are employed for power allocation and rate adaptation, whereas threshold learning is used for system parameter reconfiguration. Threshold learning is very useful in managing certain parameters under uncertain settings with limited training data.

The literature on cognitive radio and machine learning approaches is extensive. One research [12], for example, suggests an ML-based CSS method in which the energy vector components represent anticipated energy levels at each CR device. The results show that ML-integrated CSS approaches may learn implicitly from their surroundings. Another work [13] describes an ML-based multiband spectrum sensing strategy that use the greedy technique to track primary user (PU) occupancy statistics and estimate secondary user (SU) detection performance. This strategy picks sub-bands for the secondary network that provide high-throughput spectrum opportunities. In [14], an artificial neural network is used to detect the dominant radio signal in a cognitive environment. The authors of [15] implements a NN-based channel state predictor for multi-SU in a CR. Finally, in [16], authors employed the dirichlet process as a framework for non-parametric Bayesian learning in CRs.

ML techniques have shown great potential in improving the performance of CR systems and enabling more efficient spectrum utilization.

A cooperative spectrum sensing (CSS) approach based on the K-nearest neighbours algorithm is presented in reference [17]. The scheme consists of a training phase followed by a classification phase in which each user makes a judgement, with the results being processed at a fusion centre. The simulation outcomes indicate that the proposed approach has superior performance in detecting primary users (PUs) compared to conventional techniques.

Multiple supervised and unsupervised ML approaches were investigated in a recent work [18] to explore spectrum occupancy. According to the findings, the SVM method is the most successful classification method. This is consistent with the preceding explanation of the benefits of SVM for few training samples and prior knowledge of the observed process distribution. Furthermore, according to reference [19], the SVM method's success in allocating resources such as power and channel in CR networks is proved, emphasising once again the potential of ML techniques in improving the effectiveness and efficiency of CR systems. These additional contributions further demonstrate the wide range of applications and benefits of ML techniques in CR systems, including spectrum sensing and resource allocation. The authors of [20] provided a comprehensive review of recent advancements in SS for CR. The discussion begins by elucidating the half-duplex and full-duplex paradigms, with a particular focus on the operational modes within full-duplex. A thorough analysis of the collision and throughput aspects of full-duplex operation modes is presented [20]. Additionally, the authors of [21] investigated the application of cognitive radio to address bandwidth inefficiency in orthogonal frequency division multiplexing (OFDM) with a cyclic prefix (CP). They proposed integrating cyclostationary spectrum sensing and explores an alternative approach without the CP. One potential disadvantage of the approach presented in [21] is that the integration of CSS and the removal of the cyclic prefix (CP) may introduce additional complexity and computational overhead to the system. The authors of [22] explored the use of Bayesian changepoint detection method and incorporates knowledge about the environment and user mobility parameters to improve spectrum occupancy detection in a mobile cognitive radio scenario. The drawback of this method is that it exhibits reduced performance at higher SNR, indicating a potential limitation in scenarios with stronger signals or lower noise levels. In [23], the authors introduced the segment-based-CR-VANET architecture, aiming to address spectrum scarcity in vehicle Ad hoc networks (VANETs) and improve road safety and congestion management. By employing a hybrid machine learning. One limitation of the segmentation-based approach is that it faces challenges in accurately determining the optimal sub-segmentation based on probability values, potentially affecting the overall performance and accuracy of the spectrum sensing process. The authors of [24] proposed a CSS framework and explored different feature vector combinations with supervised machine learning methods. The disadvantage of the approach is that it does not provide a comprehensive analysis of the drawbacks or limitations of the proposed CSS framework, such as potential challenges in real-world implementation, scalability, or robustness to varying network conditions. The authors of [25] introduced an energy-detector-based feature vector for machine learning training and proposes a Bayesian optimized

support vector machine (SVM) approach to improve the detection rate. The drawback of this research is that the approach achieves a relatively low probability of detection rate of 0.84. While the probability of false alarm is set at 0.1, the detection rate could be improved to provide more accurate and reliable spectrum sensing results.

Drawbacks of previous approaches:

- One drawback of using SVM as a classifier for spectrum sensing in cognitive radio networks is the requirement for a large amount of labeled training data. SVMs perform well in scenarios with a sufficient amount of labeled data, but in practice, obtaining such data can be challenging and time-consuming. Moreover, SVMs tend to be computationally intensive, especially when dealing with high-dimensional feature spaces, which can limit their real-time performance in spectrum sensing applications.
- An inherent drawback of using artificial neural networks as classifiers in cognitive radio spectrum sensing is their susceptibility to overfitting. ANN models can easily become overly complex and capture noise or irrelevant features in the training data, leading to poor generalization performance on unseen data. Additionally, training ANNs requires substantial computational resources, and the training process can be time-consuming, hindering their deployment in real-time spectrum sensing systems.
- A limitation of the k-NN classifier for spectrum sensing in cognitive radio networks is its sensitivity to irrelevant and noisy features in the dataset. Since k-NN relies on the distances between feature vectors, noisy or irrelevant features can significantly impact the classification accuracy. Furthermore, k-NN suffers from the curse of dimensionality, where the effectiveness of the classifier decreases as the number of dimensions (features) increases. This can be problematic in spectrum sensing scenarios with a large number of features.

Advantages of using Bayesian optimized random forest classifier in this paper:

- Robustness to noise and outliers: Random forest is known for its robustness to noise and outliers in the dataset. The ensemble nature of random forest, where multiple decision trees are combined, helps reduce the impact of individual noisy or outlier samples on the final classification result. This robustness is particularly advantageous in cognitive radio spectrum sensing, where the presence of noise and outliers is common due to the nature of wireless communication.
- Scalability and efficiency: Random forest can handle large-scale datasets and perform well with high-dimensional feature spaces. The training and testing of random forest can be parallelized, enabling efficient computation on modern hardware architectures. This scalability is advantageous in cognitive radio networks, where large amounts of spectrum data need to be processed in real-time.
- Probabilistic outputs: Bayesian optimization can be applied to random forest to provide probabilistic outputs, estimating the uncertainty associated with each classification decision. This is valuable in cognitive radio spectrum sensing, as it allows for quantifying the confidence of the classifier in its predictions. The probabilistic outputs can be used to set appropriate thresholds and optimize the probability of detection, enhancing the overall performance of the spectrum sensing system.

3. Materials and methods

3.1 Eigenvalue based detection technique

Blind spectrum sensing, also known as eigenvalue-based detection, is a proposed method for improving cognitive radio's spectrum sensing performance without requiring any information about the licensed user's signal. The technique employs random matrix theory to formulate the detection process, which is mathematically intricate but dependable. The goal of signal identification is to determine whether or not the primary user's signal is present in the receiver. The detection method is comprised of two hypotheses: H_0, which suggests that the primary user's signal is not there, and H_1, which shows that the primary user's signal is accessible. These hypotheses can be expressed quantitatively as follows:

$$H_0: x(n) = \eta(n) \tag{1}$$

$$H_1: x(n) = \bar{s}(n) + \eta(n) \tag{2}$$

In above equation, $\bar{s}(n)$ represents the received

signal samples that are impacted by path loss, multipath fading, and temporal dispersion, whereas $\eta(n)$ represents the received white noise. The white noise signal is assumed to have the same distribution, with a mean of zero and a variance of σ_η^2.

The receiver's signal can be characterised as:

$$x(n) = \sum_{j=1}^{P} \sum_{k=0}^{N_{ij}} h_j(k) s_j(n-k) + \eta(n) \quad (3)$$

The channel response is denoted by $h_j(k)$, where j represents the channel index, and k denotes the sample index. The number of transmitters, or source signals, is represented by P, and the channel order is denoted by N_{ij}.

The detection probability (P_d), which is the chance of correctly detecting a primary user signal when it is absent (i.e., in the H_1 scenario), and the false alarm probability (P_f), which is the chance of erroneously detecting a primary user signal when it is absent (i.e., in the H_0 scenario), can be used to evaluate the performance of detection methods.

3.1.1. Maximum-minimum eigenvalue (MME) detection

The MME approach is used for signal detection when there is no prior knowledge of the signal's or noise's strength. The detection procedure includes comparing a threshold to the greatest eigenvalue to the minimum eigenvalue ratio. MME detection involves the following steps:

Step 1: Determine the received signal's sample covariance matrix:

$$R_x(N_s) = \frac{1}{N_s} \sum_{n=L-1}^{L-2+N_s} \hat{x}(n) \hat{x}^{\dagger}(n) \quad (4)$$

Number of samples collected is denoted by N_s.

Step 2: Determine the maximum and minimum eigenvalues of matrix $R_x(N_s)$ as λ_{max} and λ_{min}, respectively.

Step 3: The decision is made by comparing the ratio of λ_{max} to λ_{min} with a threshold γ_1. If the ratio is greater than γ_1, then it is decided that the signal exists ("yes decision"); otherwise, it is concluded that the signal does not exist ("no decision").

3.2 Energy detection

Considering, a signal $x(t)$ received by the detector to measure its energy, the signal $x(t)$, passes through a filter of bandwidth W. The squared output signal of a filter is integrated over a time interval T. The resulting integrator output V is then compared to a detection threshold, Γ_{th}, to determine the presence or absence of the primary user signal.

First, we will analyze the energy detector considering that the noise level, N_0, is known. This detector, as previously stated, calculates the energy of the incoming signal, $x(t)$, and contrasts it with a predetermined threshold, η_{th}. As a result, the decision characteristics for an energy detector are as follows:

$$T = \sum_{n=1}^{N} \big(y(n)\big)^2 > \eta_{th} \quad (5)$$

Where, N is the number of samples of the received signal. We consider the detection of the signal transmitted within the interval of duration T in a bandwidth W. For simplicity, the noise is considered to be additive white Gaussian with a power spectral density N_0 and the product TW large enough to allow the use of the Gaussian approximation in the test statistics.

The mean μ and the variance σ^2 of the energy detector formulated under the hypotheses H_0 and H_1, are obtained with the following equations:

$$H_0: \ \mu_{H_0} = N_0 TW, \quad \sigma_{H_0}^2 = N_0^2 TW \quad (6)$$

$$H_1: \mu_{H_1} = N_0 TW(SNR + 1), \ \sigma_{H_1}^2 = N_0^2 TW(2SNR + 1) \quad (7)$$

Assuming the hypotheses presented in Eqs. (6) and (7) and a sufficiently large number of primary user signal samples, the central limit theorem can be employed to estimate the test statistic $T(x)$ as a Gaussian distribution, as demonstrated by the following equations:

$$H_0: \ T(x) \sim \mathcal{N}(N_0 TW, N_0^2 TW) \quad (8)$$

$$H_1: T(x) \sim \mathcal{N}\left(N_0 TW(SNR + 1), N_0^2 TW(2SNR + 1)\right) \quad (9)$$

Once, we have the variance and mean of the signal $x(t)$ for each of the hypotheses, H_0 and H_1. The probability of detection (P_d) and the probability of false alarm (P_{fa}) can be approximated using Eqs. (10) and (11) respectively:

$$P_d = P(T(x) \geq \eta_{th}; H_1) = Q\left(\frac{\eta_{th} - \mu_{H_1}}{\sqrt{\sigma_{H_1}^2}}\right) \quad (10)$$

$$P_{fa} = P(T(x) \geq \eta_{th}; H_0) = Q\left(\frac{\eta_{th} - \mu_{H_0}}{\sqrt{\sigma_{H_0}^2}}\right) \quad (11)$$

Where, $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\left(\frac{x(n)^2}{2}\right)} dx$ is the Q function.

Using Eqs. (7) and (10), we obtain the detection probability:

$$P_d = Q\left(\frac{Q(P_{fa})^{-1} - \sqrt{TW}SNR}{\sqrt{1+2SNR}}\right) \quad (12)$$

We now wish to determine the minimum SNR required for the energy detector to identify the signal from a PU. When we solve for the variable SNR in Eq. (12), we get:

$$SNR = \frac{Q^{-1}(P_{fa})}{\sqrt{TW}} + \frac{Q^{-1}(P_d)}{TW}[Q^{-1}(P_d)] - \sqrt{\left(Q^{-1}(P_d)\right)^2 + TW + 2\sqrt{TW}Q^{-1}(P_{fa})} \quad (13)$$

3.3 Neighborhood component analysis (NCA)

It is a feature selection technique that is commonly used for supervised machine learning problems. It is based on the idea of maximizing the class separation in the feature space, while keeping the data points close to their original positions in the original feature space. NCA performs feature selection by finding the linear transformation of the original feature space that maximizes the likelihood of the data given the class labels. It achieves this by optimizing a cost function that balances the trade-off between class separation and data preservation.

Assuming a dataset of $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, where $x_i \in R^d$ represents the i^{th} data point, and $y_i \in R$ represents its corresponding label. Let's explore the use of the leave-one-out cross-validation technique for K-nearest neighbors (KNN) algorithm. Suppose we want to predict the label of the i^{th} sample, then the following can be done:

1. Determine the Euclidean distance between a given point i and all other points in the dataset, denoted as, $d_{ij} = \|x_i - x_j\|_2$
2. Select the k-nearest neighbors with the smallest distances, represented as, $d_{ij_1}, d_{ij_2}, \ldots, d_{ij_k}$
3. Employ the majority voting approach to predict the outcome by considering the labels of these k neighbors, i.e., $Vote(y_{j_1}, y_{j_2}, \ldots, y_{j_k})$

The preceding method presents a typical KNN approach. Now, let's introduce the Stochastic 1-NN Enhancement technique:

1. Compute the distribution of the nearest neighbor for a given data point i:

$$p_{ij} = \frac{\exp\left(-\|x_i - x_j\|_2^2\right)}{\sum_{k \neq i} \exp(-\|x_i - x_k\|_2^2)}, \; p_{ii} = 0 \quad (14)$$

2. By using the probability distribution p_{ij}, j where $j \neq i$ and $j \in [1, n]$, we can randomly sample a point k. Subsequently, we can forecast the label of the i^{th} data point as y_k.

From the above discussion, it is evident that the genuine label of the i^{th} data point is y_i. If the predicted label, y_k, equals the true label, y_i, the prediction is deemed accurate. Suppose $C_i = \{j | y_j = y_i\}$ represents the index set of the same category as the i^{th} data point. Then, the probability of correctly predicting the label of the i^{th} data point using the Stochastic 1-NN approach is given by:

$$p_i = \sum_{j \in C_i} p_{ij} \quad (15)$$

Consequently, the optimization objective for all data points is:

$$f = \sum_{i=1}^{n} p_i = \sum_{i=1}^{n} \sum_{j \in C_i} p_{ij} \quad (16)$$

Mahalanobis distance: Let $X = [x_1; x_2; \ldots; x_n]^T$ be the data sample matrix, which can be represented as $X = [f_1; f_2; \ldots; f_d]$ based on feature angles. Here, $S \in R^{d \times d}$ denotes the covariance matrix between data samples, and the expression is as follows:

$$S_{ij} = Cov(i, j) = \frac{1}{n}\left(f_i - mean(f_i)\right)^T \left(f_i - mean(f_i)\right) \quad (17)$$

The Mahalanobis distance is defined as:

$$d(x_i, x_j) = \sqrt{(x_i - x_j)^T S^{-1} (x_i - x_j)} \quad (18)$$

3.4 Cyclostationary-based sensing

Cyclostationary feature detection is a method used to identify primary user transmissions by exploiting the cyclostationary features of received

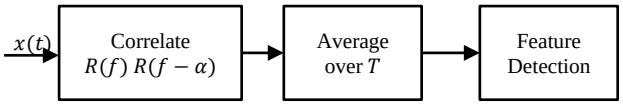

Figure. 1 Diagram illustrating the cyclostationary feature detector

signals. These characteristics result from the periodicity of the signal or its statistics, including the mean and autocorrelation, or they may be deliberately induced to facilitate spectrum sensing. Rather than relying on the power spectral density (PSD), the cyclic correlation function is utilized to identify signals present in a specific spectrum. The detection algorithms based on cyclostationary features can differentiate between noise and primary user signals, as noise is wide-sense stationary (WSS) with no correlation, while modulated signals are cyclostationary with spectral correlation due to the redundancy of signal periodicities. Moreover, cyclostationary features can be employed to distinguish between various types of transmissions and primary users.

Fig. 1 illustrates the implementation of the spectrum correlation function for cyclostationary feature detection, which includes detected features such as the number of signals, their modulation types, symbol rates, and interference presence. If the correlation factor surpasses a certain threshold, it indicates the existence of a primary user in the radio environment. Although it outperforms the energy detector by distinguishing between signal power and noise power, its computational complexity and significant processing time often degrade the cognitive radio's performance.

4. Proposed method

Fig. 2 depicts the proposed method of spectrum sensing with energy detection, min-max eigenvalue, and cyclostationary features. The extracted features are filtered, selecting the most informative subset of features for classification using neighbourhood component analysis (NCA). Further selected features are trained with Bayesian optimized random forests in cognitive radio spectrum sensing.

Let x be a matrix of N energy measurements taken in a frequency band over a certain time period, and let c be a matrix of M cyclostationary features extracted from the received signal. Let R be the covariance matrix of the received signal in the frequency band, and let λ_{min} and λ_{max} be the minimum and maximum eigenvalues of R, respectively.

The matrices x, c and λ can be expressed as:

$$x = [x_1, x_2, x_3, \dots, x_n], c = [c_1, c_2, c_3, \dots, c_m], \\ R = [\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_m] \quad (19)$$

Where, n is the number of observations and m is the number of features.

Let y be a binary label that indicates whether the frequency band is occupied ($y = 1$) or unoccupied ($y = 0$).

Let $F(x, c)$ be a feature selection function that chooses the most informative subset of features from the input matrix (x, c) based on the NCA algorithm. NCA is a distance-based feature selection method that seeks to maximize the classification accuracy by preserving the local neighborhood structure of the data.

4.1 Feature selection function

$$f(F(x,c), \lambda_{min}, \lambda_{max}) \\ = mode\{t_1(F(x,c), \lambda_{min}, \lambda_{max}), \\ t_2(F(x,c), \lambda_{min}, \lambda_{max}), \dots, t_k(F(x,c), \lambda_{min}, \lambda_{max})\} \quad (20)$$

Where, $f(F(x,c), \lambda_{min}, \lambda_{max})$ is the predicted label for input vector $(F(x,c), \lambda_{min}, \lambda_{max})$, $t_i(F(x,c), \lambda_{min}, \lambda_{max})$ is the predicted label for input vector $(F(x,c), \lambda_{min}, \lambda_{max})$ from the i^{th} decision tree in the forest, and $mode\{\}$ is the function which is designed to identify the predicted label that appears most frequently among the decision trees..

$F(x,c)$ can be expressed as:

$$F(x,c) = [f_1, f_2, f_3, \dots, f_k] \quad (21)$$

Where, k is the number of selected features and $f_1 = [x_1, c_1, \lambda_1]$.

The random forest algorithm is used to classify the selected features based on the vectors. The random forest algorithm comprises a collection of decision trees that are generated using a random subset of training data and input features. For each input vector $(F(x,c), \lambda_{min}, \lambda_{max})$, the random forest algorithm provides a set of predicted labels. The predicted label for a given vector $(F(x,c), \lambda_{min}, \lambda_{max})$ is obtained by taking a majority vote of the predicted labels from each decision tree in the forest.

The effectiveness of the random forest method with feature selection may be assessed using performance measures like as the receiver operating characteristic (ROC) curve and the area under the curve (AUC). These metrics allow for the evaluation of the balance between the probability of false alarm and the probability of detection, as well as the optimisation of the algorithm's performance for a

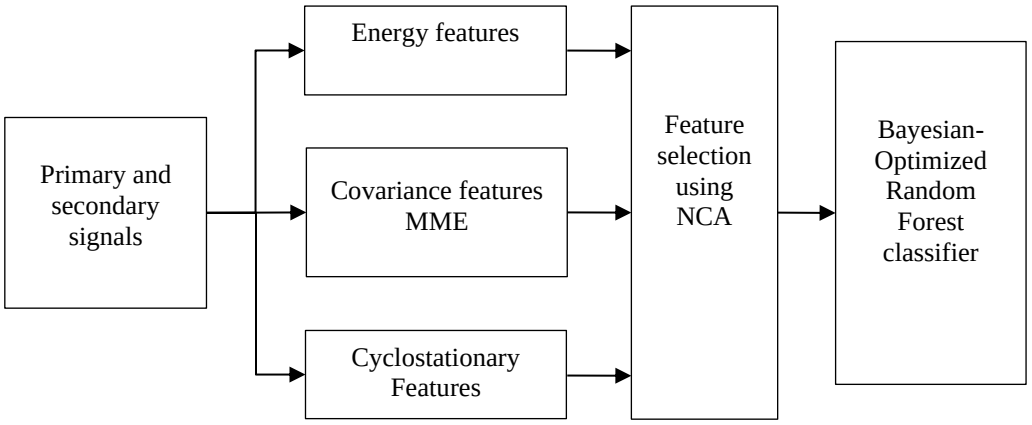

Figure. 2 Proposed spectrum sensing with BO-RF

specific application.

4.2 Classifiers

4.2.1. Bayesian optimized random forest (RF) classifier

Random forest (RF) is a technique for solving classification and regression problems using ensemble learning. It is an enhanced version of the decision tree algorithm.

Compared to a single decision tree, random forest classifiers are more accurate and reliable due to their ability to reduce overfitting by combining the outcomes of multiple decision trees. Moreover, they can efficiently handle high-dimensional and large datasets. It's worth mentioning that the random forest classifier includes a decision tree as one of its base classifiers:

$$\{h(x, \Theta_k),\ k = 1, \dots L\} \tag{22}$$

Random forests are made up of a group of binary decision trees that have been infused with randomness.

Suppose we have q random variables, denoted as Θ_1 through Θ_q, which have no correlation with $\mathcal{L}_n$. Let us consider $\left(\hat{h}(\Theta_1), \dots, \hat{h}(\Theta_q)\right)$ as a collection of tree predictors. By employing the method explained below to merge these arbitrary trees, we can obtain the predictor of random forests, which is denoted as $\hat{h}_{RF}$:

- Average of individual tree predictions in regression:

$$\hat{h}_{RF}(x) = \frac{1}{q}\sum_{l=1}^{q} \hat{h}(x, \Theta_l) \tag{23}$$

- Majority vote among individual predictions trees in classification:

$$\hat{h}_{RF}(x) = \arg\max_{1\le k\le K} \sum_{l=1}^{q} 1_{\hat{h}(x,\Theta_l)=k} \tag{24}$$

The term "random forest" describes a method in which each tree relies on a distinct random variable (in addition to $\mathcal{L}_n$), and each separate predictor is explicitly forecasted for each tree.

To create a Bayesian optimized random forest classifier, one combines the random forest algorithm and Bayesian optimization into a mathematical formula.

Random forest algorithm: Let D be the training dataset, where each observation x_i is a d-dimensional feature vector, and y_i is the corresponding label.

Draw bootstrap samples of $D, D_1, D_2, \dots, D_B$, each of size n, from D.

For each bootstrap sample D_b, fit a decision tree T_b on D_b. At per capita node of the tree, arbitrarily choose m features from the d features, and choose the best feature/split-point among them.

To form the random forest model $f(x)$, one must merge the B decision trees. This model forecasts the label y for a new observation x by selecting the most frequent outcome among the B trees.

Bayesian optimization: Assume that θ represents the hyperparameters of the random forest classifier. These hyperparameters include B, the number of trees, the maximum depth of each tree, the number of features (m) to examine at each split, and so on. Let $L(\theta|D)$ be the log-likelihood of the hyperparameters θ given the training dataset D.

Define a prior distribution $P(\theta)$ over the hyperparameters θ.

For each iteration t:

a. Sample θ_t from the prior distribution $P(\theta)$.
b. Fit the random forest model $f_t(x)$ using the hyperparameters θ_t and the training dataset D.
c. Evaluate the log-likelihood $L(\theta_t|D)$ of the

hyperparameters θ_t on the validation dataset D_{val}.

d. Update the posterior distribution $P(\theta|D, D_{val})$ using Bayes' rule:

$$P(\theta|D, D_{val}) \propto P(\theta) * P(D_{val}|\theta) * P(D|\theta) \quad (25)$$

Where, $P(D_{val}|\theta)$ is the likelihood of the validation dataset given the hyperparameters θ, and $P(D|\theta)$ is the likelihood of the training dataset given the hyperparameters θ.

Return the hyperparameters θ^* that maximize the posterior distribution $P(\theta|D, D_{val})$, and fit the final random forest model $f^*(x)$ using θ^* and the full training dataset D.

The formula for the Bayesian optimized random forest classifier is therefore:

$$f^*(x) = \arg\max_{\theta}\left(P(\theta|D, D_{val}) * f_\theta(x)\right) \quad (26)$$

Where, $f_\theta(x)$ is the random forest model with hyperparameters θ, and $P(\theta|D, D_{val})$ is the posterior distribution over θ obtained by Bayesian optimization.

The random forest classifier's Bayesian optimization involves tuning various hyperparameters, such as the maximum tree depth, the minimum number of samples required to split a node, and the number of trees in the forest.

5 Simulation and results

5.1 Simulation parameters

This study focuses on examining a cooperative spectrum sensing (CSS) scenario wherein secondary users (SUs) are dispersed in a 5-by-5 grid topology spanning across a 4000m × 4000m area. To provide a visual representation of this topology, kindly refer to Fig. 3. Various simulation parameters were established for this study, including a bandwidth of 5 MHz, a sensing duration of 100 μs, a noise spectral density of -153 dBm, and a path-loss exponent of 4. Both shadow fading and multi-path fading components are assumed to remain constant with a value of 1. Furthermore, the transmit power of each primary user (PU) was established at 200 mW. This study examines two PUs situated at fixed locations of (500 m, 500 m) and (-1500 m, 0 m), respectively. The probability of a PU being active is 0.5, and the states of the two PUs are considered to be statistically independent.

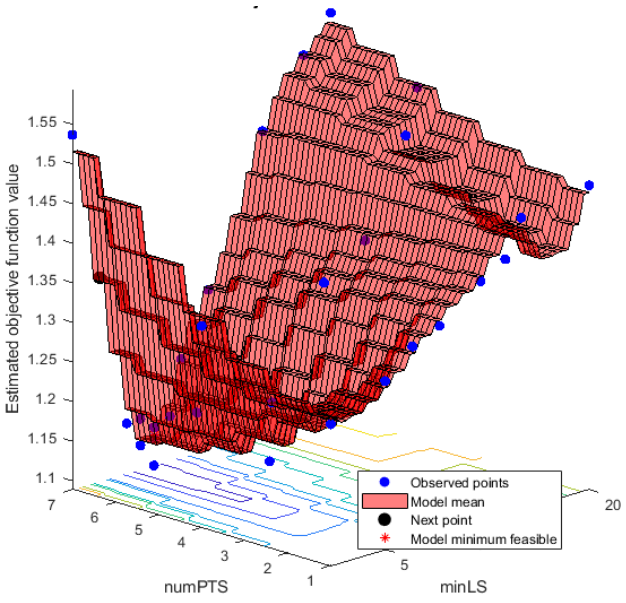

Figure. 3 Objective function model

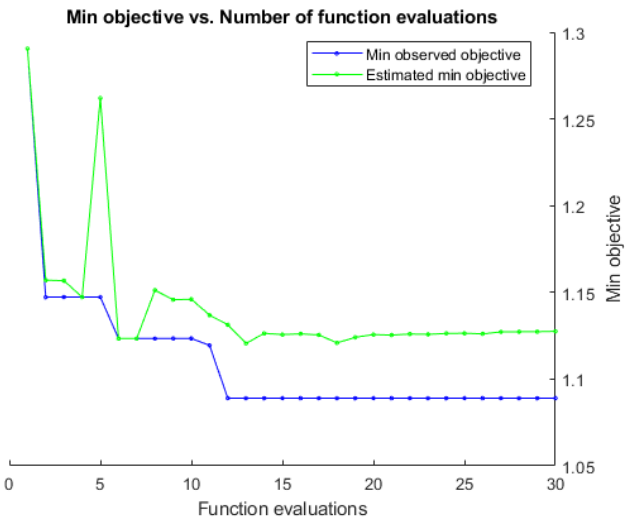

Figure. 4 Objective evaluation for quantile error function

Table 1. Simulation parameters for machine learning

PU Cartesian Position in Meters	[0 0]× e^3;
SU Cartesian position in meters	[00.5; 00.75; 01]× e^3;
PU transmission probability	0.5
PU transmission power in W	0.2
MCS realization	$5e^4$;
SU spectrum sensing period in seconds	$5e^{-6}$;
SU spectrum sensing bandwidth in hertz	$5e^6$;
Noise PSD in dBm/Hz	-153;
No of primary and secondary user	1,3
Training size	5000 samples

5.2 Probability of detection

The probability of detecting primary user (PU)

activity accurately is a crucial aspect of any spectrum sensing technique. This metric, called the probability of detection, determines the effectiveness of the sensing technique in identifying and detecting PU signals with precision. In contrast, the probability of false alarm measures the likelihood of a secondary user (SU) detecting the presence of a PU signal (H_1) when, in reality, no PU is active in the channel (H_0). On the other hand, the probability of missed detection represents the opposite scenario where the SU fails to detect a PU signal (H_1) when one is present, leading to channel underutilization. Both the missed detection and probability of false alarm are essential in evaluating the performance and reliability of a sensing technique. Lower probability of false alarm indicates fewer instances of unnecessary disruptions or interference caused by incorrectly sensing the presence of a PU, while a lower probability of missed detection ensures accurate sensing and detection of PU signals, preventing channel underutilization.

Therefore, achieving a balance between these two probabilities is necessary to ensure optimal spectrum utilization and minimize interference. By optimizing the sensing technique and utilizing appropriate algorithms, a high probability of detection can be achieved while keeping the missed detection and probabilities of false alarm low, thereby enhancing the overall reliability and performance of spectrum sensing.

A monte carlo simulation was executed with a 5×10^4 implementation to assess the effectiveness of the previous model in determining the channel state. The simulation scenario included one primary user (PU) and three secondary users (SUs) operating in an additive rayleigh fading channel, as illustrated in Fig. 3. The simulation parameters were based on the values presented in Table 1.

The Bayesian optimized random forest method was compared against traditional AND, OR, and maximum ratio combining (MRC) analysis methods to assess the performance in Figs. 5 and 6. Both linear and Gaussian kernel functions were considered for the Bayesian optimized SVM.

Upon inspection, it was observed that the MRC method showed the highest performance, trailed by the Bayesian optimized SVM with a linear kernel. In accumulation to these methods, an ROC (receiver operating characteristic) curve was generated by incorporating the singular energies in every single SU. The SUs' resulting average signal-to-noise ratios were 2 dB, 9 dB, and 14 dB, which were affected by their respective distances from the PU. These differences contributed to distinct channel recognition characteristics exhibited by each SU.

By evaluating various threshold settings, the

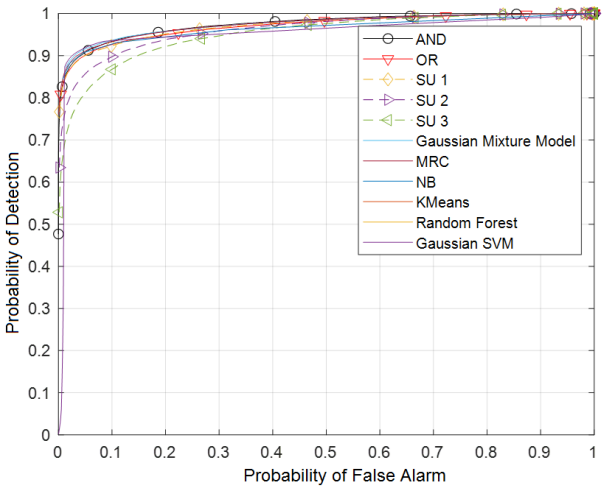

Figure. 5 Comparative evaluation of different methods for probability detection for a primary user signal

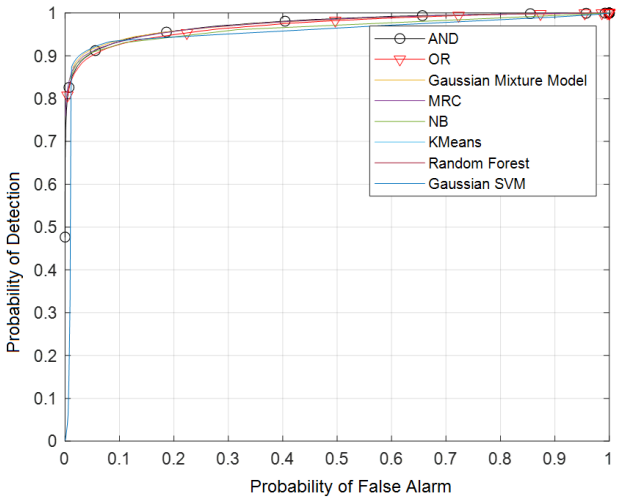

Figure. 6 Comparative evaluation of different methods for probability detection for a primary user signal

ROC curve has offered significant insights regarding the trade-off between the probability of detection and the probability of false alarm. By analyzing the ROC curve, it became possible to understand the system's performance under different operating conditions and make informed decisions regarding the selection of suitable detection thresholds.

5.3 Training duration for different classifiers

Various factors, such as the size of the training dataset, the classifier's complexity, the hyperparameters utilized, and the available computational resources, can affect the training duration of different classifiers.

It is important to note that these results are specific to the dataset and the experimental setup used in this study.

According to Fig. 7 the Naïve bayes classifier has a relatively high training duration among the

Table 2. Comparison of different machine learning algorithms in terms of training duration and classification performance

Method	Training duration in NCA based features in sec	Training duration in normal features in sec	Classification performance in NCA accuracy %	Classification performance in without NCA accuracy %
NB	24.9	36.89	90.34%	89.6%
KNN	15.8	26.75	87.64%	82.23%
BO-RF	21.65	29.87	94.56%	93.78%
GMM	25.98	39.78	90.23%	90.45%
SVM	20.67	35.76	92.34%	91.42%

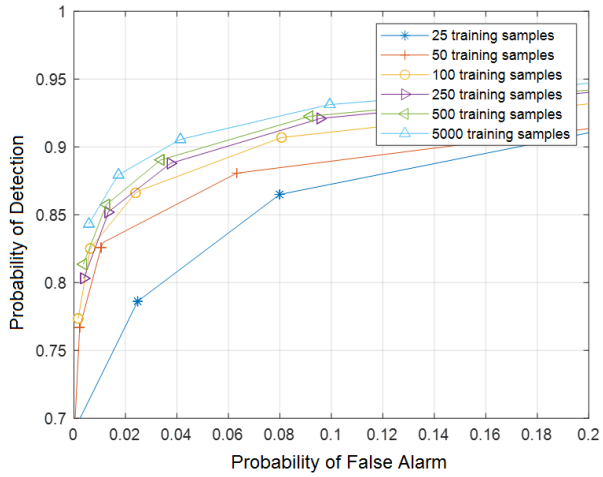

Figure. 7 Different training sample for detection of probability and false alarm for Naïve bayes classifier

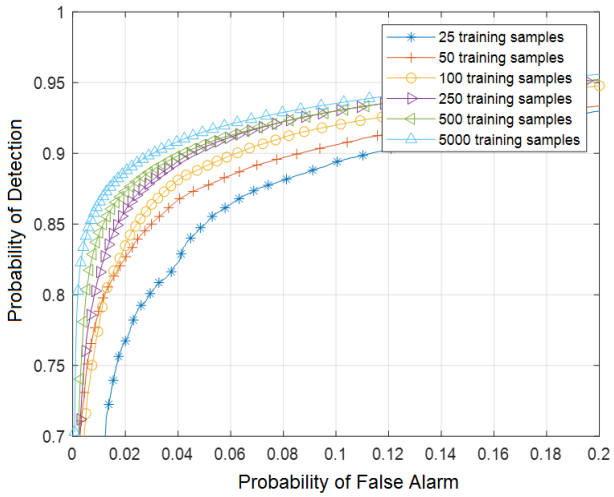

Figure. 8 Different training sample for detection of probability and false alarm for random forest classifier

unsupervised classifiers, taking 1.12796 seconds for 1000 samples. Optimal outcome achieved after 50000 training sample as compared to the other lower training sample. It can be seen that better probability of detection is achieved for lower false alarm as training sample will go higher side.

As depicted in Fig. 8, the random forest classifier has the lengthiest training time among all the classifiers. The analysis of the performance based on the number of training samples indicates that achieving an optimal outcome requires a higher number of training samples, specifically 50,000 in this case. Comparing the results with lower training sample sizes, it is observed that as the training sample size increases, the probability of detection improves while maintaining a lower false alarm rate.

This observation suggests that a larger training sample size allows the model to learn more effectively and capture the underlying patterns and characteristics of the data. With more training samples, the model can generalize better and make more accurate predictions, resulting in a higher probability of detection. At the same time, the false alarm rate remains low, indicating a reliable and robust performance of the detection system.

It ensures that the model can effectively learn and adapt to different scenarios, leading to improved detection accuracy and reliability. Table 2 provides a comparison of different machine learning algorithms in terms of training duration and classification performance, considering the use of neighborhood component analysis (NCA) based features and normal features.

Here is the analysis of the results:

Naive bayes (NB): The training duration for NB is 24.9 seconds with NCA based features and 36.89 seconds with normal features. The classification performance shows an accuracy of 90.34% with NCA and 89.6% without NCA.

K-nearest neighbor (KNN): KNN has a shorter training duration compared to NB, with 15.8 seconds for NCA based features and 26.75 seconds for normal features. However, the classification performance is slightly lower, with an accuracy of 87.64% with NCA and 82.23% without NCA.

Bayesian optimized random forest (BO-RF): BO-RF requires a training duration of 21.65 seconds with NCA based features and 29.87 seconds with normal features. It shows a higher classification

Table 3. Comparative analysis with different work

Methods	Probability of detection	Probability of false alarm
Proposed Bayesian optimized Random Forest	**0.94**	**0.1**
[23]	0.7	0.1
[24]	0.5	0.1
[25]	0.8	0.1

performance compared to NB and KNN, with an accuracy of 94.56% with NCA and 93.78% without NCA.

Gaussian mixture model (GMM): GMM has a longer training duration, with 25.98 seconds for NCA based features and 39.78 seconds for normal features. The classification performance is similar to NB, with an accuracy of 90.23% with NCA and 90.45% without NCA.

Support vector machine (SVM): SVM shows a moderate training duration of 20.67 seconds with NCA based features and 35.76 seconds with normal features. The classification performance is relatively high, with an accuracy of 92.34% with NCA and 91.42% without NCA.

The results indicate that BO-RF achieves the highest classification performance among the analyzed algorithms, followed by SVM. However, NB and GMM also exhibit competitive performance.

The current study aims to explore the potential of utilizing machine learning systems in cooperative spectrum sensing (CSS) and focuses on improving classification speed, training time, and classification performance. Table 3 provides a comparative analysis of different methods for spectrum sensing in cognitive radio networks. The methods are evaluated based on two key metrics: probability of detection and probability of false alarm. In previous research, [23] introduced a two-dimensional distance vector that was transformed into an m-dimensional energy vector for classification purposes using the K-nearest neighbor (KNN) algorithm. However, this method may not be effective in poor signal conditions and fewer user scenarios compared to energy detector-based methods. Another research paper, [24], proposed a cognitive radio performance fuzzy logic and Naïve bayes classifier for detection in the vehicle Ad-hoc network (VANET) platform, achieving a detection rate of 0.5 and a probability of false alarm set at 0.1.

In contrast, [25] utilized centralized cooperative spectrum sensing (CCSS) techniques for cognitive radio networks utilizing an energy detector approach. The proposed Bayesian-optimized Random Forest classifier achieved a probability of detection of 0.94, indicating its high accuracy in correctly detecting available frequencies. The probability of false alarm is maintained at a constant value of 0.1, implying a low rate of false detections.

Overall, the comparative analysis of various methods in cooperative spectrum sensing reveals that incorporating machine learning techniques and Bayesian optimization could improve the accuracy of detection. The proposed method in this study has the potential to outperform existing approaches and improve the performance of cooperative spectrum sensing systems.

6 Conclusion

The article explores various approaches for cooperative spectrum sensing (CSS) in cognitive radio networks, with a focus on utilizing supervised and unsupervised learning techniques. The main objective is to determine the availability of channels for secondary users while ensuring the presence of primary users using machine learning algorithms.

In this study, the received energy level at the secondary users (SUs) is considered a crucial feature for identifying the availability of channels. The researchers propose a bayesian-optimized random forest classifier, which is compared with other classifiers such as Naive bayes, support vector machines (SVM), Gaussian mixture models (GMM), and maximum ratio combining (MRC) in terms of performance.

The evaluation of the classifiers takes into account various factors including training time, classification delay, and receiver operating characteristic (ROC) curves. The results indicate that the bayesian-optimized random forest classifier outperforms the other cooperative spectrum sensing algorithms, achieving a high probability of detection of 0.94. This finding suggests that the classifier is effective in accurately identifying the presence or absence of primary users in cognitive radio networks.

Based on the findings, the article concludes that bayesian-optimized random forest classifiers hold promise as methodologies for cooperative spectrum sensing in cognitive radio networks. However, the study emphasizes the importance of precise training of energy vectors for making accurate decisions. It

suggests that there is room for improvement in the proposed CSS techniques by gradually training the classifiers with individual energy vectors. This incremental training approach enables the classifiers to adapt to changing environments without the need for complete retraining. Consequently, this method has the potential to enhance the performance and adaptability of the CSS techniques in cognitive radio networks.

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

This paper conceptualization, software simulation, verification of results and original draft preparation has been done by Raghavendra L R. The supervision and final approval have been done by Dr. Manjunatha R C.

References

[1] R. Sarikhani and F. Keynia, "Cooperative Spectrum Sensing Meets Machine Learning: Deep Reinforcement Learning Approach", *IEEE Communications Letters*, Vol. 24, No. 7, pp. 1459-1462, 2020.

[2] Z. Shi, W. Gao, S. Zhang, J. Liu, and N. Kato, "AI-Enhanced Cooperative Spectrum Sensing For Non-Orthogonal Multiple Access", *IEEE Wireless Communications*, Vol. 27, No. 2, pp. 173-179, 2019.

[3] W. Ning, X. Huang, K. Yang, F. Wu, and S. Leng, "Reinforcement Learning Enabled Cooperative Spectrum Sensing In Cognitive Radio Networks", *Journal of Communications and Networks*, Vol. 22, No. 1, pp. 12-22, 2020.

[4] A. Haldorai, U. Kandaswamy, A. Haldorai, and U. Kandaswamy, "Cooperative Spectrum Handovers In Cognitive Radio Networks", *Intelligent Spectrum Handovers In Cognitive Radio Networks*, pp. 1-18, 2019.

[5] A. Coluccia, A. Fascista, and G. Ricci, "Spectrum Sensing by Higher-Order SVM-based Detection", In: *Proc. of International Conf. of 27th European Signal Processing Conference (EUSIPCO)*, pp. 1-5, 2019.

[6] A. A. Olawole, F. Takawira, and O. O. Oyerinde, "Fusion Rule And Cluster Head Selection Scheme In Cooperative Spectrum Sensing", *IET Communications*, Vol. 13, No. 6, pp. 758-765, 2019.

[7] J. Chen, S. Huang, H. Li, X. Lv, and Y. Cai, "PSO-based Agent Cooperative Spectrum Sensing In Cognitive Radio Networks", *IEEE Access*, Vol. 7, pp. 142963-142973, 2019.

[8] R. Aswatha, V. Seethalakshmi, K. Murugan, N. Sathishkumar, A. Reethika, and S. Gunanandhini, "Implementation of Cooperative Spectrum Sensing Using Cognitive Radio Testbed", *Indian Journal of Science and Technology*, Vol. 13, No. 13, pp. 1355-1366, 2020.

[9] I. Kakalou, D. Papadopoulou, T. Xifilidis, K. E. Psannis, K. Siakavara, and Y. Ishibashi, "A Survey On Spectrum Sensing Algorithms For Cognitive Radio Networks", In: *Proc. of International Conf. of 7th International Conference on Modern Circuits and Systems Technologies (MOCAST)*, pp. 1-4, 2018.

[10] A. Rai, A. Sehgal, T. L. Singal, and R. Agrawal, "Spectrum Sensing and Allocation Schemes for Cognitive Radio", *Machine Learning and Cognitive Computing for Mobile Communications and Wireless Networks*, pp. 91-129, 2020.

[11] P. Anker, "From Spectrum Management To Spectrum Governance", *Telecommunications Policy*, Vol. 41, No. 5-6, pp. 486-497, 2017.

[12] L. Hu, R. Shi, M. Mao, Z. Chen, H. Zhou, and W. Li, "Optimal Energy-Efficient Transmission For Hybrid Spectrum Sharing In Cooperative Cognitive Radio Networks", *China Communications*, Vol. 16, No. 6, pp. 150-161, 2019.

[13] G. Eappen and T. J. P. C. Shankar, "Hybrid PSO-GSA For Energy Efficient Spectrum Sensing In Cognitive Radio Network", *Physical Communication*, Vo. 40, p. 101091, 2020.

[14] G. P. Aswathy and K. Gopakumar, "Sub-Nyquist Wideband Spectrum Sensing Techniques For Cognitive Radio: A Review And Proposed Techniques", *AEU-International Journal of Electronics and Communications*, Vol. 104, pp. 44-57, 2019.

[15] I. Develi, "Spectrum Sensing In Cognitive Radio Networks: Threshold Optimization And Analysis", *EURASIP Journal on Wireless Communications and Networking*, Vol. 1, pp. 1-19, 2020.

[16] G. Mahendru, A. Shukla, and P. Banerjee, "A Novel Mathematical Model For Energy Detection Based Spectrum Sensing In Cognitive Radio Networks", *Wireless Personal Communications*, Vol. 110, pp. 1237-1249, 2020.

[17] R. B. Chaurasiya and R. Shrestha, "Hardware-Efficient And Fast Sensing-Time Maximum-Minimum-Eigenvalue-Based Spectrum Sensor

For Cognitive Radio Network", *IEEE Transactions on Circuits and Systems I: Regular Papers*, Vol. 66, No. 11, pp. 4448-4461, 2019.

[18] S. H. Kamel, M. B. A. E. Malek, and S. E. E. Khamy, "Compressive Spectrum Sensing Using Chaotic Matrices For Cognitive Radio Networks", *International Journal of Communication Systems*, Vol. 32, No. 6, p. e3899, 2019.

[19] M. Ramchandran and E. N. Ganesh, "Energy Efficient and Interference-aware Spectrum Sensing Technique for Improving the Throughput in Cognitive Radio Networks", In: *Proc. of International Conf. of IOP Conference Series: Materials Science and Engineering*, Vol. 993, No. 1, p. 012092, 2020.

[20] A. Nasser, H. A. Haj Hassan, J. A. Chaaya, A. Mansour, and K. C. Yao, "Spectrum Sensing For Cognitive Radio: Recent Advances And Future Challenge", *Sensors*, Vol. 21, No. 7, p. 2408, 2021.

[21] A. Kumar and P. N. Kumar, "OFDM System With Cyclostationary Feature Detection Spectrum Sensing", *ICT Express*, Vol. 5, No. 1, pp. 21-25, 2019.

[22] A. S. V. Chede, F. Rottenberg, M. Dossou, and J. Louveaux, "Use of Bayesian Changepoint Detection For Spectrum Sensing In Mobile Cognitive Radio", In: *Proc. of International Conf. of 93rd Vehicular Technology Conference (VTC2021-Spring)*, pp. 1-5, 2021.

[23] M. A. Hossain, R. M. Noor, K. L. A. Yau, S. R. Azzuhri, M. R. Z'aba, I. Ahmedy, and M. R. Jabbarpour, "Machine Learning-Based Cooperative Spectrum Sensing In Dynamic Segmentation Enabled Cognitive Radio Vehicular Network", *Energies*, Vol. 14, No. 4, p. 1169, 2021.

[24] W. Wu, Z. Li, S. Ma, and J. Shi, "Performance Improvement For Machine Learning-Based Cooperative Spectrum Sensing By Feature Vector Selection", *IET Communications*, Vol. 14, No. 7, pp. 1081-1089, 2020.

[25] A. D. Vishwakarma and G. A. Kulkarni, "Cooperative Spectrum Sensing using Rule based Hard Decision and Soft Decision with Bayesian Optimized Support Vector Machine", *International Journal of Intelligent Engineering and Systems*, Vol. 15, No. 1, pp. 405-413, 2022.