Wireless Networks (2024) 30:4761–4772
https://doi.org/10.1007/s11276-022-03055-0

Support vector machine approach of malicious user identification in cognitive radio networks

Kaleem Arshid[2] · Zhang Jianbiao[2] · Iftikhar Hussain[3] · Gebrehiwet Gebrekrstos Lema[1] · Muhammad Yaqub[2] · Rizwan Munir[4]

Accepted: 9 June 2022 / Published online: 17 September 2022

Abstract

In cognitive radio network (CRN), effective spectrum management provides better quality of service. The spectrum is limited but the significance of the spectrum is increasing at each network generation. Due to the ineffective spectrum allocation policies, several researches have indicated that a vast segment of the licensed radio is not viably used. A CRN is an intelligent spectrum utilization innovation that provides better spectrum interface. Spectrum sensing detects unused spectrum in the manner that protects interferences to the authorized users. In principle, the secondary user (SU) receives the primary user (PU) signal and reports it to the fusion center for decision or spectrum allocation. The SU cooperates in the detection of the presence or absence of the PU. This type of spectrum sensing is called cooperative spectrum sensing. However, the significance of this type of spectrum sensing is blurred by the security problems. Malicious users can deliberately report misleading information regarding the presence of the PU. Hence, in this paper, a support vector machine learning algorithm is proposed to statistically learn the behavior of the malicious users and it classifies the legitimate SU and malicious users. A particle swarm optimization algorithm is also integrated to learn the smallest possible distinguishable malicious users' energy report deviation from the legitimate SUs. The probability of detection and energy of detection have been applied to evaluate the contribution of the proposed method. Finally, the simulation results have confirmed that better spectrum management can be derived from the proposed statistical approach.

Keywords CRN · Spectrum management · Network security

Abbreviations

AN	Always no
AWGN	Additive White Gaussian Noise
AY	Always yes
CRN	Cognitive radio network
ERFC	Complementary error function
FC	Fusion center
FCC	Federal Communication Commission
GA	Genetic algorithm
ML	Machine learning
P_d	Probability detection
P_f	Probability of false alarm
PSO	Particle swarm optimization
PU	Primary user
SNR	Signal to noise ratio
SU	Secondary user
SVM	Support vector machine

✉ Gebrehiwet Gebrekrstos Lema
g.jcool.com@gmail.com

[1] School of Electrical and Computer Engineering, Mekelle University, Mekelle, Ethiopia

[2] Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

[3] Collage of Science and Engineering, Hamad Bin Khalifa University, Ar-Rayyan, Qatar

[4] Beijing University of Postdoc and Telecommunication, Beijing 10086, China

1 Introduction

The recent trends of wireless communication have shown attractive features. The mobility support of wireless communication was the core among the other critical solutions. The wireless network relies on the spectrum to communicate for different purposes. However, the spectrum is crowded by increased number of users and diverse applications. Due to the spread of the wireless networks and researches have shown that the spectrum is inefficiently utilized such that they often apply fixed frequency [1].

The unlicensed bands are rarely available and the licensed bands are either underutilized or unoccupied. This problem happens due to the static spectrum assignment and roughly speaking; the dynamic spectrum management simplifies the problem. Hence, the idea of CRN gains promising importance because it permits licensed users to access the spectrum dynamically and the SU can access the licensed and unlicensed spectrum systematically. CRN-based spectrum management is an efficient spectrum utilization, which learns the environment and adjusts its transmission parameters. The key concerns of CRN are, its sensing and cognitivism capabilities. The sensing of a SU can be degraded by fading and hidden terminal problems. This challenge can be solved by cooperative spectrum sensing which involves reporting of sensing results of multiple SUs.

The federal communication commission (FCC) has indicated that only 15 to 85% of the licensed spectrum is utilized [2]. To effectively utilize the radio, an adaptive intelligence has shown fascinating importance to solve the network congestion [3]. With the concept of CRN-based spectrum management, the PUs can transmit at any time with no restrictions, but the SUs has to obtain the benefit of spectrum access when the PUs do not use the corresponding spectrum [4]. A light overhead physical layer network security has been proposed using machine learning [5], but neither the spectrum sensing nor the malicious user detection has been covered. A PU traffic pattern-based opportunistic spectrum handoff has been evaluated to reactively or proactively release the channel depending on the arrival rate of the PU [6], however, the malicious user attack was not considered in the study.

To solve the existence of MUs, heuristic approaches including a genetic algorithm (GA) are applied [7–9]. Besides, machine learning (ML) techniques are other good candidates which learns the environment and derives the pattern of the different users. As representative ML-based classification and regression algorithms, there are k-nearest neighbor, decision tree, naïve Bayes, and logistic regression has shown promising performance, [10–12].

The performance of cooperative sensing is analyzed with p-norm detector and improved energy detector and the performance is evaluated on a generalized κ-ν fading channel, [13]. Throughput enhancement were significant using sequential technique [14], however, the classification of the malicious users was not considered. The machine learning method has been applied to distinguish the PUs and malicious users [15]. However, it was not easy to reliably conclude the type of user as the MU itself can act as legitimate SUs. The machine learning-based sparse coding technique was proposed to detect the PU emulation and jamming attacks, in which the signal obtained from a channel dependent is used to differentiate between a spectrum hole, a legitimate PU, and MU [16]. However, the work fails to distinguish the type of secondary users. The multi-user multi-input multi-output system is evaluated using weighted eigenvalue detection. The research reveals interesting output including improved throughput, reduces energy consumption, and reduces error probability [17]. However, the MUs detection problem has continued a CRN challenge. On the other hand, for defending against a primary user emulation attack, a dogfight approach is used to choose the safest channel [18]. This defense technique uses simple channel sensing and evading the primary user emulation attack. An energy efficient spectrum sensing has also been proposed [19], however, the malicious user detection challenge was not solved in the CRN spectrum management.

Similar work has conducted that incorporates two hidden Markov models [20] to distinguish between malicious and legitimate users. However, the work relies on predefined static pattern. The support vector machine has also shown promising solution to classify the malicious and legitimate user [21, 22] but most of the researches didn't exploit how the fusion center can operate to effectively manage the malicious users and an unrealistic assumption blurs the classification.

Recently, a network security against the primary user emulsion issue of CRN is modeled [23], however, it lacks to exploit the environmental dynamics of the channel. On the other hand, the type of CRN user classification is studied using support vector machine learning. It uses the signal to noise ratio (SNR) to decide the user classification [24]. However, it didn't consider the deliberate misleading case of the secondary user because the secondary user can report high or low energy deliberately even when the actual SNR is different.

Though cooperative spectrum sensing provides a promising performance, the existence of malicious users (MUs) deteriorates the performance of the CRN by sending falsified results to a fusion center (FC). Hence, the proposed work focuses on how to combat the MUs to enhance spectrum management. Besides, it is not easy to differentiate the MUs from the legitimate SUs if the MUs' energy level is very close to the legitimate SUs. For this reason, particle swarm optimization is applied to study the smallest possible distinguishable MUs' energy report difference from the legitimate SUs.

In general, due to its mobility support, the demand of wireless network is increasing continuously. Besides, when the number of network users increase, the scarcity of the spectrum becomes obvious. However, the spectrum is a limited natural resource and hence a special spectrum management becomes critical. The CRN is one of the intelligent spectrum utilization mechanisms that detects unused spectrum and assigns to the needy without interfering to the legitimate user.

The concept of CRN highly relies on sensing and cognitivism. While the cognitivism focusses on the intelligence in decision making, the sensing is the capability to identify the presence of the PU. It has also indicated that if we try to sense the channel by the central fusion center, there would be high probability that the sensing result could be degraded by fading and hidden terminal problems. For this reason, the significance of cooperative spectrum sensing sounds ideal because this covers spatial distribution. The cooperative spectrum sensing can sense better than a single fusion center. But the significance of cooperative spectrum sensing didn't come for free. The MUs can deliberately send misleading information about the presence of the PU. Hence, this work proposes support vector machine algorithm to study the MUs' misleading information. The support vector machine learns the behavior of the MUs and develops a pattern to classify MUs and legitimate users.

Recent research has indicated that simplified assumptions have used to classify the legitimate and malicious users. These includes majority vote, linear searching technique, centralized security (more overhead), and a predefined fixed MU features. On the other hand, researchers assume that the MUs report sensing energy far below or above operating energy report of legitimate SUs, for example [21]. Hence, in this paper, the predefining hypothesis is almost fully avoided and the SVM formulates a pattern by statistically learning to the behavior of the MUs. The optimization tool is also introduced to fully automate the assumptions and classifications.

2 Literature review

It has been several years since the idea of cooperative spectrum sensing becomes attractive. The security challenges, reliable malicious user detection, have made it impractical to apply. Several approaches of malicious user detection have proposed [25, 26], however, either they treat the legitimate or malicious users are well known.

Without complicating the sensing, a compromised sensor node was able to discover using a group voting scheme [27]. On time basis, each node has given a weight and to poll the vote the quality of data transmission and weights are combined. Similarly, a compressed sensing technique is applied [28] in which the MU is removed by processing the signal at the fusion center. A simple majority vote based malicious user detection has applied [29] and the decision is considered anomaly if more than half of the votes suggests malicious user, legitimate user otherwise. This technique provides limited significance because it tends to decide using simple signal strength and it sounds not ideal in tiny sensor networks. Similarly, a simple voting mechanism has proposed for determining the presence of malicious users in the SUs region [30]. In this case, an outlier detection identifies the malicious user using a decentralized approach that uses the geographical correlation of received signal strength in near proximity to the SU. Another simple onion peeling method is introduced as a defensive mechanism against compromising users [31]. This onion peeling method sets-up a threshold to determine the status of the user (MU or LU) and if the current weight exceeds the threshold, the user's reports are discarded. This cannot classify if the MUs report below threshold weightage.

On a distributed network, a consensus suggestion mechanism is proposed [32]. With many trials, each legitimate SU are made to choose a neighbor with whom to share sensory data. By comparing the received reports to the local mean value, a trustworthy neighbor was identified and the users with different value from the local reference mean are thrown out.

An optimization problem has evaluated with the goal of maximizing CRN energy efficiency while keeping outage probability as a constraint [33]. To attain the CRN energy efficiency, a low-complexity linear searching technique is used, however, it is hard to achieve the goal under multiple MUs scenario. Besides, machine learning technique has applied in CRN in which the frequency band, modulation type, distance, power, and interference influence the system [34, 35]. However, the machine learning power to adapt the real-time radio resources is not explored.

On the other hand, the concept of long short-term memory added with the concept of extreme learning machines [36] is applied for 5G CRN. This avoids the need for large computational cost and it improves the sensing performance because it considers the energy and distance environmental statistics. However, its long time training and centralized security overhead have made it not ideal for 5G CRN networks. The SVM-based spectrum mobility and handover issues have characterized well under dynamic user features [37, 38], however, the impact of malicious users is not studied. Likewise, in a cooperative spectrum sensing, without prior-knowledge on the primary user, the detection performance has improved with the conventional neural network method [39], but the major security drawback of cooperative spectrum sensing is ignored. The support vector machine is also used to play with transmission power and interference to maximize the network throughput [40] but it is hard to classify the different CRN user without predefined fixed features. Finally, the combined method of support vector machine and genetic algorithm (GA) have shown sounding potential to estimate the number of primary users in a CRN [41]. However, the combination of machine learning and optimization didn’t apply to exploit the significance of cooperative sensing and CRN security solutions.

3 Methods

In this paper, a cooperative spectrum sensing method is used to distinguish a malicious user in a cognitive radio network. This cooperative spectrum sensing relies on energy detection to determine the presence or absence of the MUs. In principle, the SU receives the signal of the legitimate user and sends to the fusion center for decision. Malicious users can also send a false information that claims the presence or absence of the legitimate user deliberately to mislead the decision. Hence, the proposed support vector machine learning analyzes the report of the energy and categorizes the energy as SU’s report or MU’s report. The legitimate user and malicious user are categorized using a probability of detection function. The probability of detection and probability of false alarm functions have used to evaluate the user mobility and signal to noise ratio effects. The learning capability of the proposed algorithms statistically studies the behavior of the users based on their energy report.

More specifically, in a wireless network, there are certain number of legitimate SUs and MUs. Both the legitimate SUs and MUs can report their sensing result to the fusion center. The legitimate SUs do this because they are trying to use the channel when the PU are not currently using the channel. The MUs report an energy detection without actually sensing. On applying the classification method, the energy report of all legitimate SUs and MUs is collected. The collected information is used for training. Using the training data, a model is developed. Then once the model is developed, for any energy report from any of the users (legitimate SUs and MUs), we predict the category of the user. Finally, the boundary of the legitimate SUs and MUs are simulated. In the legitimate SUs and MUs classification, the MUs report that the presence or absence of the PUs by sending higher or lower energy levels. In the analysis, the effect of signal to noise ratio on the probability of detection is analyzed. The impact of the MUs on the probability of detection is also studied.

Finally, the malicious users and legitimate users are distinguished by a hyper line. The matlab tool is used to evaluate the learning and to simulate the results for visualization. The numerical evaluations are conducted on 6 MHz bandwidth and 20 microsecond sensing period. More specific experiment parameters and settings are given in Table 1.

Table 1 Simulation parameters

Simulation parameter	Value or considered
Number of legitimate SUs	12
Number of MUs	4
Probability of PU	0.5
Number of iterations	10,000
Channel model	AWGN
Attack type	AY and AN
Sensing type	Energy detection

4 System model

In this study, a CRN which characterizes primary users (PUs), secondary users, and a fusion center. The secondary users can be legitimate cognitive radio users (SUs) or malicious users (MUs). Both the malicious and legitimate SUs can perform spectrum sensing and transmission-reception capabilities.

They can also transmit sensing results to the fusion center (FC), as shown in Fig. 1. The SUs, in general, are able to cooperate to determine the presence of the PU in the channel. Each SU can perform local sensing by applying energy detection method of spectrum sensing. In Fig. 1, there are S number of SUs and M number of MUs. There is also a single fusion center, usually a base station, the performs spectrum assignment decisions. The absence and presence of the PU are given by the H_0 and H_1 terms, respectively.

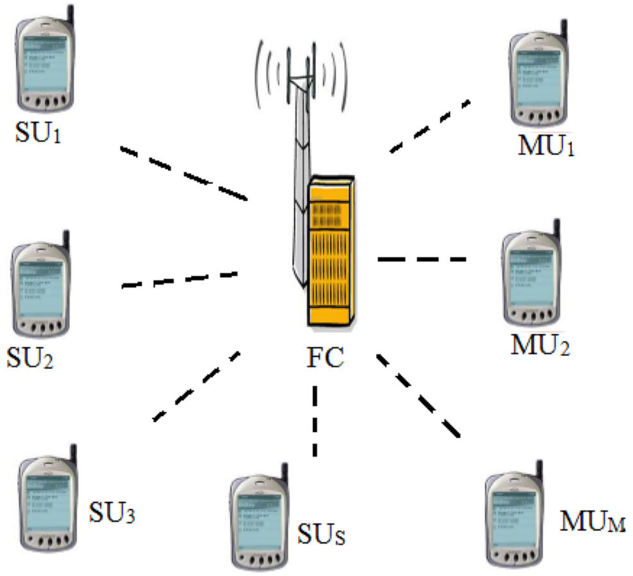


Fig. 1 Proposed system model

$$y = \begin{cases} \text{n; H}_0, \\ \text{s} + \text{n; H}_1, \end{cases} \tag{1}$$

where n is the additive white Gaussian noise (AWGN) of the channel, s is the transmitted signal by the PU. To determine the existence of the primary user, the energy detection technique is considered. This can be described by the sum of the energies of the users, given as:

$$E_i = \frac{1}{N}\sum_{i=1}^{N} |y_i|^2, \tag{2}$$

where $N = 2TW$, with T is the sensing duration and W is the bandwidth, and y_i denotes the i-th sample of the received signal. Using the Nyquist sampling theorem, the minimum sampling rate (f_s) should be atleast twice the operating bandwidth, $\text{f}_\text{s} = 2$ W. Considering the central limit theorem, if N is large enough ($N > 200$), the combined signal can be well approximated as a Gaussian random variable under hypotheses H_0 and H_1, with means μ_0 and μ_1 and variances $\sigma^2{}_0$ and $\sigma^2{}_1$, which are given by

$$\begin{aligned} \text{H}_{0:}\mu_0 &= \text{N}, \mu_0^2 = \text{N(y + 1)} \\ \text{H}_{1:}\mu_1 &= N,\ \mu_1^2 = \text{2N(2Y + 1)} \end{aligned} \tag{3}$$

To describe the probability of detection and false alarm, the Gaussian distribution can be used as an approximation to the real-valued random variables with the mean μ and the variance σ^2. The mean describes the location of the peak and the variance describes the width of the Gaussian distribution.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{4}$$

Considering the maximum posteriori, the probability of false alarm (P_f) and probability detection (P_d) can be expressed as (5):

$$\begin{aligned} P_d(\lambda) &= P_r[y > \lambda | H_1] \\ P_f(\lambda) &= P_r[y > \lambda | H_0], \end{aligned} \tag{5}$$

where λ is the decision threshold of maximum posteriori detection. If we have a normal variable $X\ with\ N(\mu;\ \sigma^2)$, the probability that $y > \lambda$ is

$$P_r[y > \lambda] = Q\left(\frac{\lambda - \mu}{\sigma}\right) \tag{6}$$

The Q-function $Q\ (x)$, given in (7), is also used to express the probability of Gaussian random variables. For all real-valued x, $Q\ (x)$ is defined as the probability of standard normal random variable (zero mean, unit variance) exceeds λ:

$$Q(\lambda) = \frac{1}{\sqrt{2\pi}}\int_{\lambda}^{\infty} e^{-\frac{t^2}{2}} dt = \frac{1}{\lambda\sqrt{2\pi}} e^{-\frac{\lambda^2}{2}} \tag{7}$$

The Q-function is also directly related to complementary error function (erfc) as

$$Q(\lambda) = \frac{1}{2} erfc\left(\frac{\lambda}{\sqrt{2}}\right) \text{ where } erfc(\lambda) = \frac{2}{\sqrt{\pi}}\int_{\lambda}^{\infty} e^{-t^2} dt \tag{8}$$

Considering the central limit theorem for large number of samples, the Chi-square distribution is approximate as Gaussian distribution and the probabilities of false alarm and detection are, respectively, simplified as [42]:

$$P_f = \frac{1}{2} erfc\left[\frac{1}{\sqrt{2}}\left(\frac{\lambda - N\sigma_u^2}{\sigma_u^2\sqrt{2N}}\right)\right] \tag{9}$$

$$P_d = \frac{1}{2} erfc\left[\frac{1}{\sqrt{2}}\left(\frac{\lambda - N(y_i + 1)\sigma_u^2}{\sigma_u^2\sqrt{2N(2y_i + 1)}}\right)\right], \tag{10}$$

where $\sigma^2{}_\text{u}$ is the noise variance and y_i is the Signal to Noise Ratio of the PU at the ith SU.

5 Proposed machine learning solution

In this study, S number of SUs and M number of MUs are considered, as shown in Fig. 1. It is assumed that the number of legitimate secondary users are more than the number of malicious users, M < S. The SUs listens to the channel for the presence of the PU and they cooperate to sense the activity of the PU channel to the fusion center (FC). The MUs can send false information which damages the significance of the CRN system. The MUs may send deliberate misleading information to the FC. This information can falsely indicate the presence of the PU or the absence of the PU. Hence, the MUs send either higher energy detection to indicate the presence of the PU on the channel, called always yes (AY), or it sends lower energy detection to indicate the absence of the PU, called always no (AN). If sufficient learning data is available, machine learning are fascinating solutions without explicitly programmed.

In this work, machine learning is used to statistically classify the legitimate SUs and MUs. The support vector machine learning-based statistical classification [43] is employed. Statistical machine learning relies on averaging the energy reporting of the users (SUs and MUs) on continuous several observations. The legitimate SUs reports the right sensing results in which the energy of detection can be high or low depending on the received signal level from the PU. However, the MUs reports either higher energy detection (AY) or lower energy detection (AN) irrespective of the PU status. Hence, the proposed statistical machine learning averages the energy reporting and classify the type of the user.

The training data set (D) is described as:

$$D = \{(E_i, C_i) | E_i \in \Re, C_i \in \{1, -1\}\} \tag{11}$$

Here the C_i is the class of MUs or SUs indicator vector, and the class "1" and "− 1" represent legitimate SUs and MUs, respectively. The E_i is the energy vector determined by the maximum posteriori energy detection. After the sensing duration is over, the energy reports of all users (MUs and SUs) is applied to the machine learning for building a model, called exploration phase. From the machine learning trained model, a hyper line is formed to classify the MUs and SUs. The FC, then, uses this line to distinguish the class. This line, also called hyperplane, can be formulated by:

$$\begin{aligned} h.x + p &= a \\ h.x + p &= -a, \end{aligned} \tag{12}$$

where p is a threshold, a is any constant, and h is the weight vector obtained in training phase. In the training, the two different classes are categorized as the MUs and SUs. The pseudocode of the proposed machine learning is summarized as:

1. **Initialize the learning**
   a. Initialize the number of SUs and MUs
   b. Generate sensing reports of SUs and MUs
   c. Export sensing energy reports
2. **Exploration**
   a. Load the training data set
   b. Develop a model
   c. Capture model parameters
3. **Exploitation**
   a. For any similar data & user
   b. Predict SUs & MUs boundary
   c. Scatter the class of each user
   d. Display visualization
4. **End**

Many researches assume that the MUs report either quite higher energy than the legitimate SUs or much lower energy than the legitimate SUs. However, the MUs may learn the behavior of the legitimate SUs and they can report an energy detection that closely resembles the SUs. In that case, it is difficult to distinguish the MUs from the SUs. Hence, in this paper, an optimization tool is integrated to learn the smallest possible distinguishable MUs' energy report deviation from the SUs. To optimize the smallest possible malicious user energy deviation, a PSO is used. This energy deviation is the optimal distinguishable energy that the malicious user can apply to falsely report the presence or absence of the PU. The PSO is characterized as an iterative optimal searching algorithm. PSO is initialized with a group of random particles and searches for the optimal solutions by updating generations. In every iteration, each particle is updated by determining the nearest best values.

The optimization algorithm particles represent the MUs and legitimate SUs. The velocity and position indicates the energy levels of the MUs and legitimate SUs, respectively. The Global best, G(best), and Private best, P(best), represents the current smallest detectable energy level between the MUs and SUs among all of the users. The summary of the PSO algorithm is given in Fig. 2.

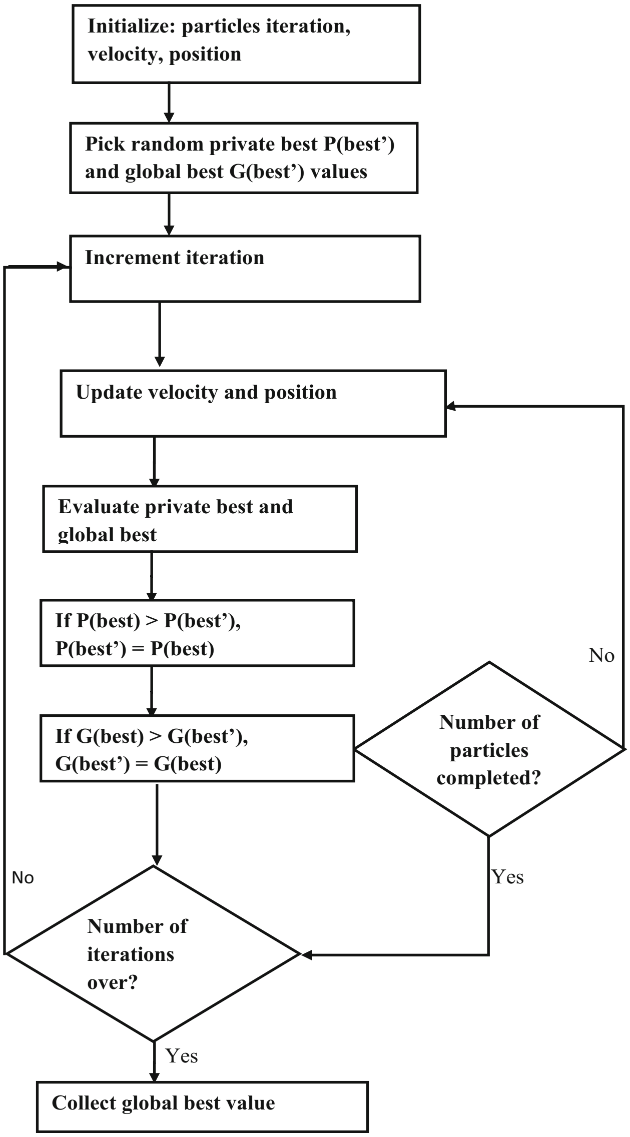

Initialize: particles iteration, velocity, position
Pick random private best P(best') and global best G(best') values
Increment iteration
Update velocity and position
Evaluate private best and global best
If P(best) > P(best'), P(best') = P(best)
If G(best) > G(best'), G(best') = G(best)
Number of particles completed?
No
Yes
Number of iterations over?
No
Yes
Collect global best value

Fig. 2 PSO working skeleton

6 Results and discussion

To evaluate the performance of the proposed optimization and CRN user classification, MATLAB 2018a is used. The rest of the simulation parameters are given in Table 1.

Before the malicious CRN users detection and FC decision mechanism, the major factors that determine the probability of detection is shown in Fig. 3. The CRN is primarily proposed to alleviate the spectrum scarcity when the number of users are increasing in number and functionality. The ultimate significance of wireless communication is its mobility capability and the CRN assumes mobile users share the licensed spectrum.

When the users are mobile, they logically receive different signal strength depending on the legitimate SUs location and PU position. Hence, the legitimate SUs can be largely affected by the SNR, as shown in Fig. 3.

On the other hand, mobile users experience different probabilities of detection. The probability of false alarm and the probability of detection is almost directly proportional, as shown in Fig. 4. At the same time, Fig. 4 indicates that different probability of detection can be reported to the FC from different legitimate SUs.

The probability of detection (Pd) is also largely affected by the presence of malicious users. Depending on the attack type (AY or AN), the probability of detection deviates from the right detection. Considering the energy of detection, covered in Figs. 7 and 8, the AY and AN type of attack report higher energy and lower energy, respectively, without sensing the actual presence of the PU. In the presence of the MUs, the average probability of detection violates the right probability of detection, shown in Fig. 5.

The energy detection reports, Fig. 6, indicates that different CRN users (including the legitimate SUs and MUs) reports different energy levels. Simply collecting the energy level reports at the FC, it is not easy to distinguish the MUs from the legitimate SUs. This diminishes the significance of the CRN system in several years.

By applying statistical machine learning, the behaviour of the MUs and legitimate SUs is studied. The AY MUs reports higher energy to deliberately mislead the decision making center, Fig. 7. The proposed machine learning statistically learns which user is energy detection reporting from sensing and which user is reporting without the knowledge of the presence of the PU. The learning averages the energy report of the individual users over 10,000 sample space and it classifies the MUs from the legitimate SUs. Similarly, the AN MUs reports lower energy to deliberately mislead the decision making center, Fig. 8.

Hence, in this work, an optimization tool is used to determine the minimum energy report of the MUs different from the legitimate SUs' energy report. Several researches assumed that there is a large energy gap between the MUs and legitimate SUs, however, the MUs may learn and apply an energy report quite close to the legitimate SUs. Though the learning itself determines the pattern of the MUs, the optimization exploits the behaviour of the MUs in millions of iterations. The optimization enables the researchers to consider a more realistic scenarios. The machine learning closely studies the MUs and legtimate SUs energy detection pattern and formulates a new model. The new model is derived from exploring the training data set and the model is used to exploit any future predictions.

The proposed machine learning also provides a decision line, Figs. 7 and 8, and the FC easily outlines that the users above the decision line (in Fig. 7) are MUs and the users below the decision line (in Fig. 8) are MUs. This provides an easier decision framework to the fusion center.

The support vector machine has proposed to classify normal SUs and MUs with the range of sensing energies at the legitimate SUs lies on the range of 90–108 [21]. The AN and AY energies are assumed far below and above the

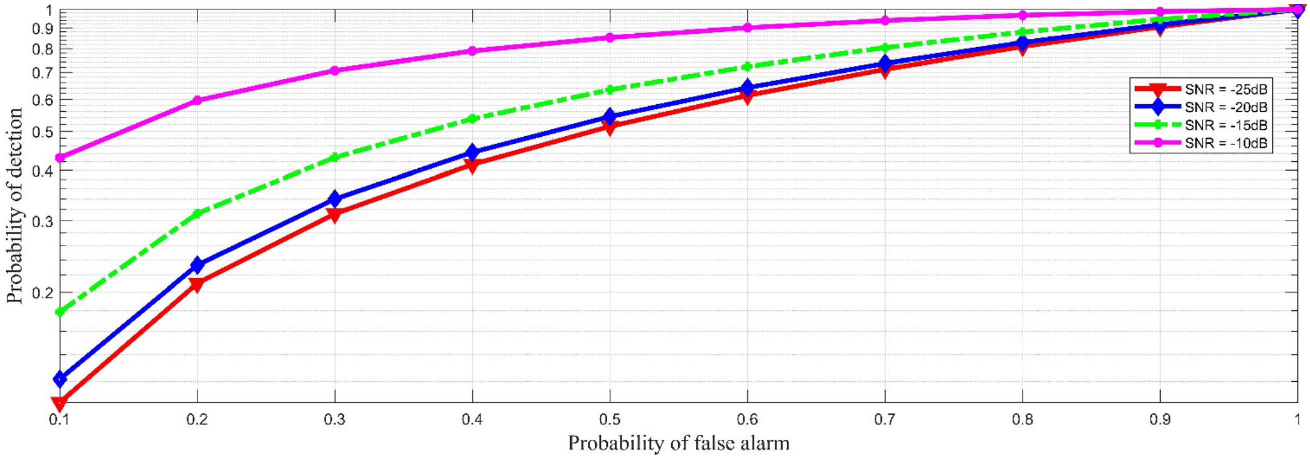


Fig. 3 SNR effects on the probability of detection

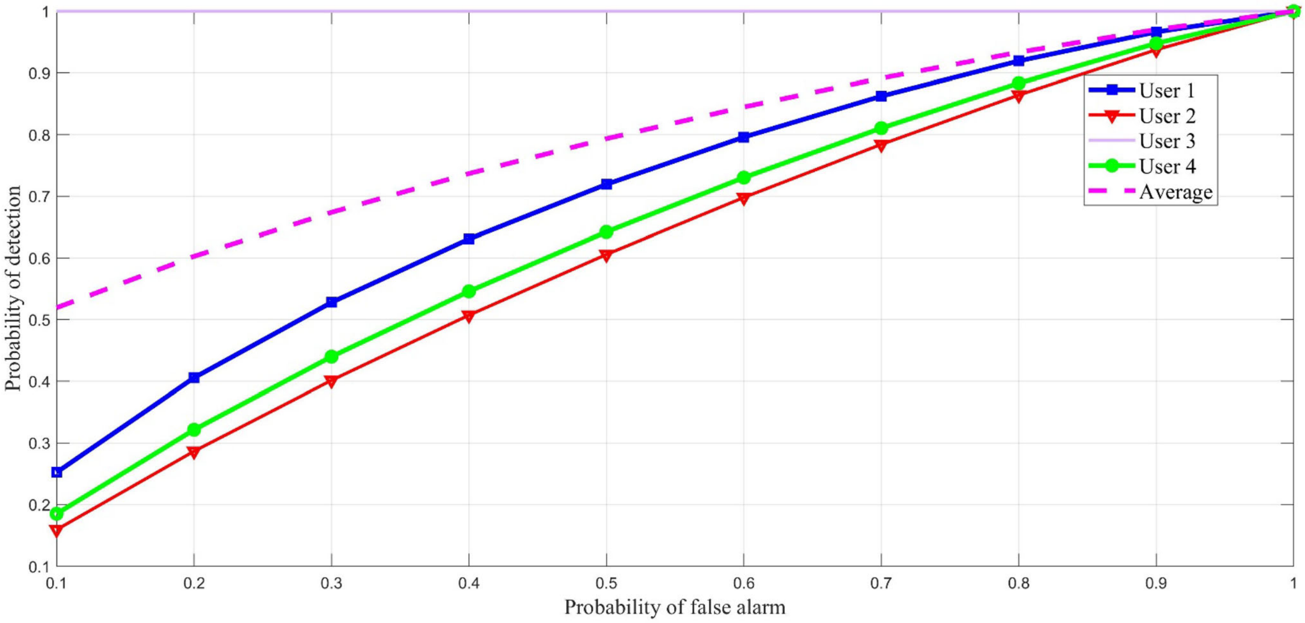


Fig. 4 Probability of detection of random SUs

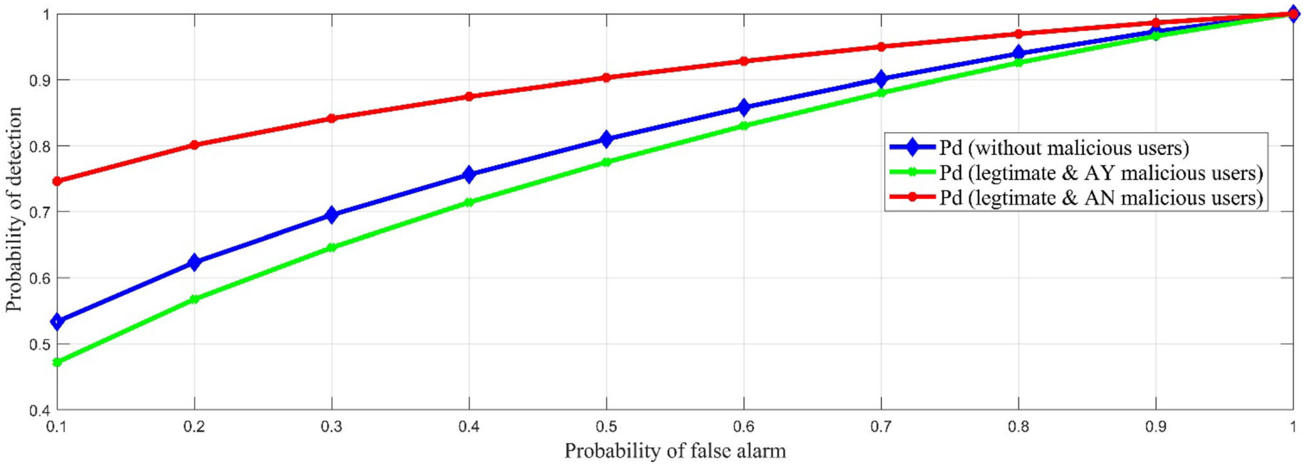


Fig. 5 The impact of the MUs on the probability of detection

legitimate energy ranges. However, the more it is assumed that the MUs do not operate closer to the range of the legitimate user the easier it will be to classify but a little far from the reality. Figures 7 and 9 of [21] indicate that the assumption of the sensing energies of the MUs are around 9. In this paper, we have introduced an optimization algorithm that determines the minimum sensing energy gap between the legitimate and malicious users. Consequently, the proposed solution learns and classifies the legitimate and malicious users even when their sensing energies are close to one as it can be shown in Figs. 7 and 8 of this work.

To sum up, the cooperative spectrum sensing enables ideal solution to sense the existence of the PU. Instead of using another device or employing the unreliable centralized sensing, the cooperative sensing offloads the FC task. Hence, the spectrum management is simplified by applying the cooperative spectrum sensing. However, the cooperative spectrum sensing is highly exposed to malicious user attacks. For this reason, the major contribution of this paper focuses on the MUs detection mechanisms. The machine learning is used to study the characteristics of the MUs and then to determine the pattern of the malicious users. The learning averages the energy detections of the individual users. The legitimate users apply valid spectrum sensing while the MUs send any energy level to the FC without actually sensing the presence or absence of the PU. The proposed machine learning watches the energy report of

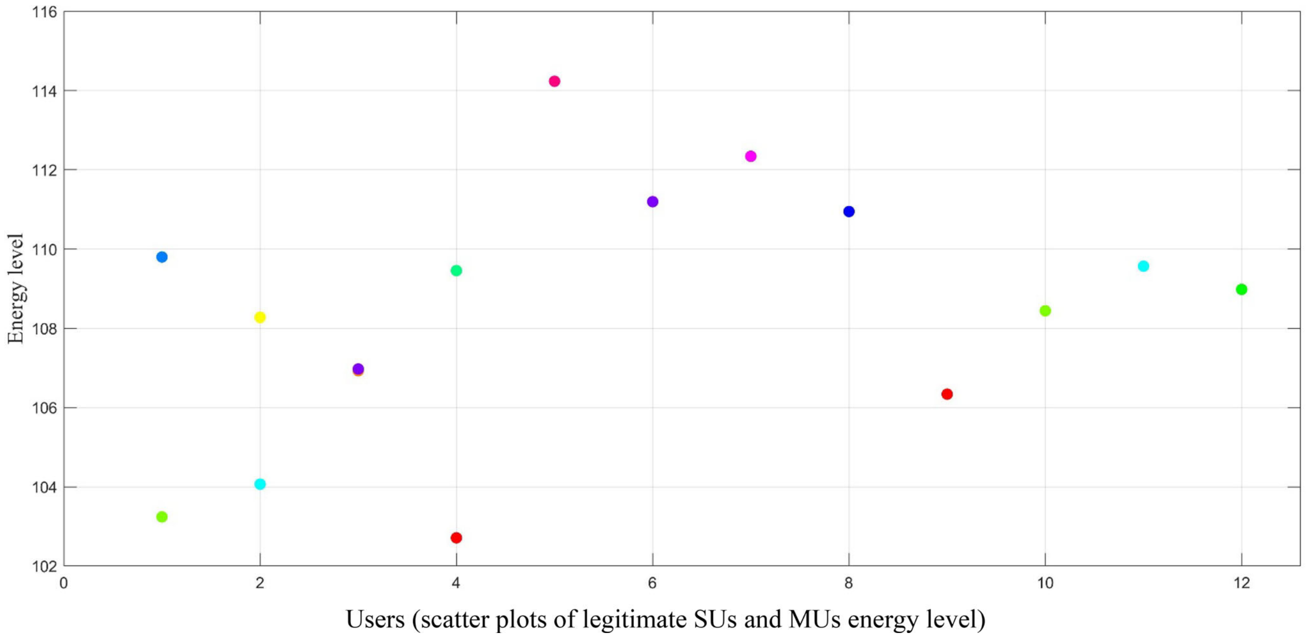


Fig. 6 Unclassified CRN users (legitimate SUs and MUs)

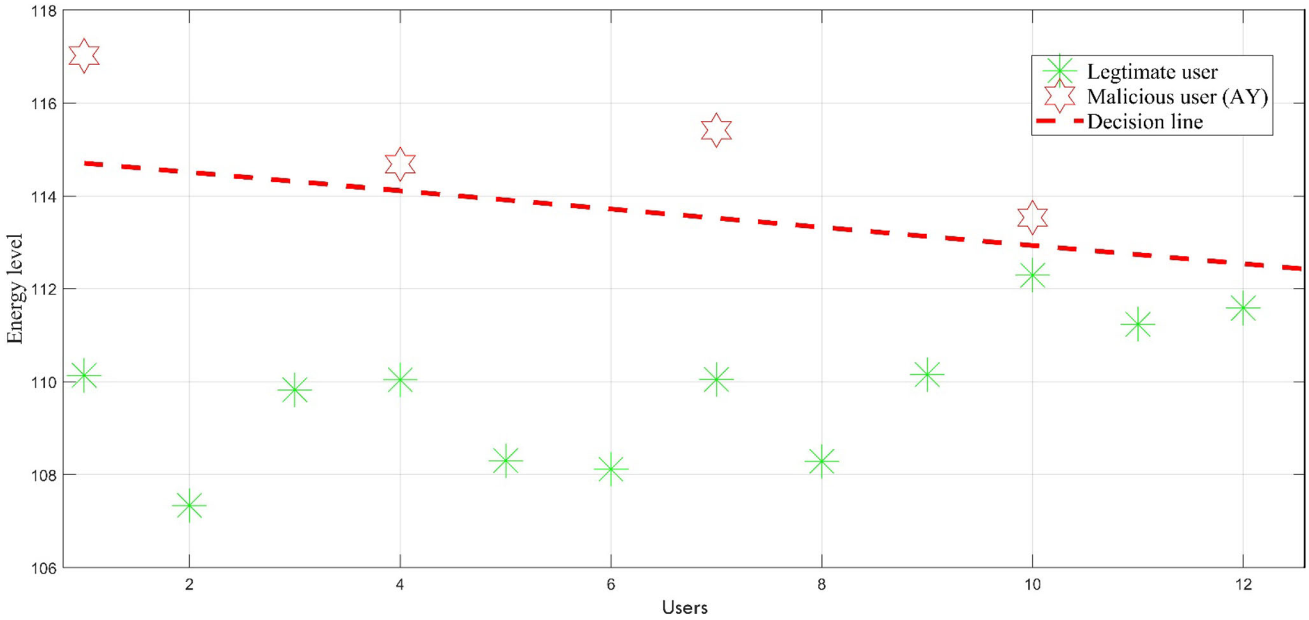


Fig. 7 CRN users classification (legitimate SUs and AY type of MUs)

every user and determines a pattern that characterizes the type of user. The optimization algorithm is also applied to realize the energy reporting by the MUs.

7 Conclusion

To solve the CRN spectrum management challenge, a cooperative spectrum sensing is proposed. The legitimate SUs-based of PU spectrum sensing simplifies the spectrum management of the CRN system. However, this method of spectrum sensing is highly exposed to malicious user attacks. Machine learning is proposed to statistically study the experience of malicious users. Furthermore, the learning classifies the MUs and legitimate SUs with a decision line. Depending on the type of attack (AY or AN), the MUs are classified as above or below the decision line and that region is blacklisted from further accommodation in the spectrum management. The MUs are assumed that they can act quite similar behavior to the legitimate SUs. The

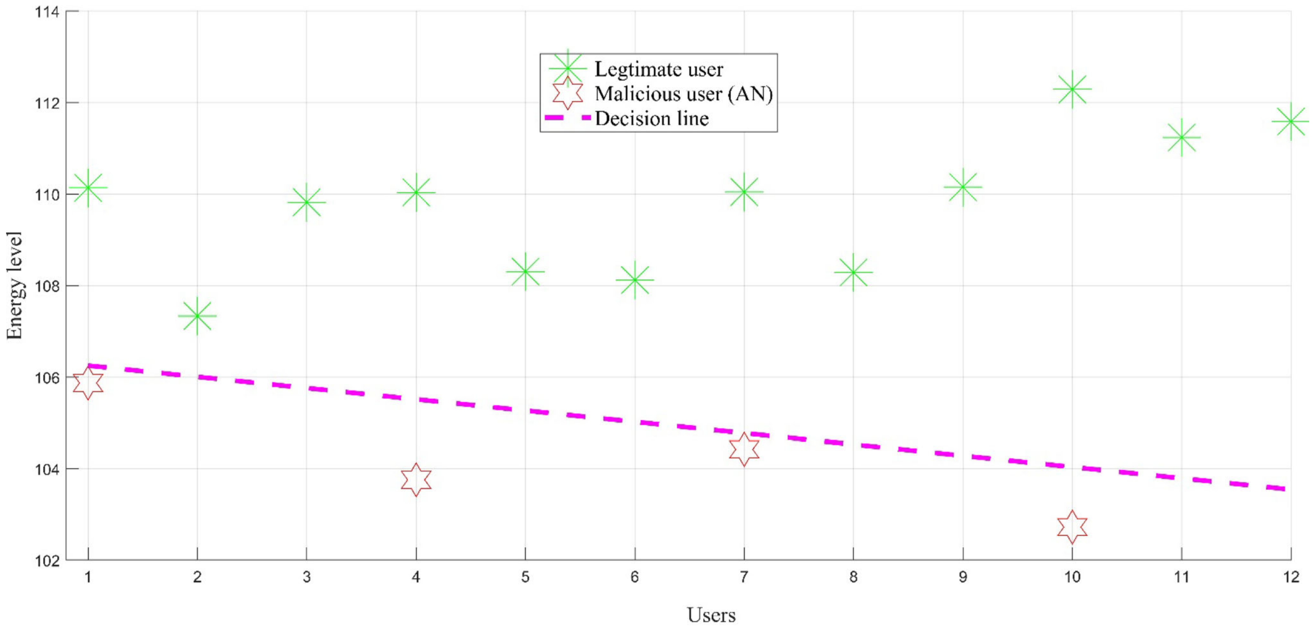


Fig. 8 CRN users classification (legitimate SUs and AN type of MUs)

simulation results have shown that the proposed method distinguishes the MUs from the legitimate SUs well. The FC can easily decide the spectrum management by considering the decision line to separate the MUs and legitimate SUs.

Author contributions GGL has coordinated, summarized, carried out in the sequence alignment and KA drafted the manuscript. ZJ conceived of the study, designed the study and performed the statistical analysis. IH directed the writing of the paper. MY and RM has drafted the writing orientation.

Funding The research was not funded.

Data availability We declare that the matlab code used for the simulation will not be shared, and we assure that we will send it on demand.

Declarations

Conflict of interest The authors declare that there is no conflict of interest regarding the publication of this paper.

Consent for publication We agree the publication of the paper.

References

1. Sansoy, M., Buttar, A. S., Singh, K.. (2015). Cognitive radio: Issues and challenges. *Journal of Network Communications and Emerging Technologies (JNCET* 2), *4*.
2. Baker, A., Brogan, P., Carare, O., Copeland, N., DeGraba, P., Kauffman, S., Lafontaine, P., Matraves, C., Prince, J., Sullivan, S., & Sun, P. (2020). Economics at the FCC 2019–2020: Spectrum policy, universal service, inmate calling services, and telehealth. *Review of Industrial Organization, 57*(4), 827–858.
3. Haykin, S. (2005). Cognitive radio: Brain-empowered wireless communications. *IEEE Journal on Selected Areas in Communications, 23*(2), 201–220.
4. Zhai, L., Wang, H., & Gao, C. (2016). A spectrum access based on quality of service (QoS) in cognitive radio networks. *PLoS ONE, 11*(5), e0155074.
5. Lema, G. G., Weldemichael, K. S., & Weldemariam, L. E. (2021). Performance evaluation of cooperative mobile communication security using reinforcement learning. *Heliyon, 7*(5), e07108.
6. Arshid, K., Hussain, I., Bashir, M. K., Naseem, S., Ditta, A., Mian, N. A., Zahid, M., & Khan, I. A. (2020). Primary user traffic pattern based opportunistic spectrum handoff in cognitive radio networks. *Applied Sciences, 10*(5), 1674.
7. Mehboob, U., Qadir, J., Ali, S., & Vasilakos, A. (2016). Genetic algorithms in wireless networking: Techniques, applications, and issues. *Soft Computing, 20*(6), 2467–2501.
8. Khan, M. S., Gul, N., Kim, J., Qureshi, I. M., & Kim, S. M. (2020). A genetic algorithm-based soft decision fusion scheme in cognitive iot networks with malicious users. *Wireless Communications and Mobile Computing, 2020*, 1–10.
9. Azmat, F., Chen, Y., & Stocks, N. (2015). Analysis of spectrum occupancy using machine learning algorithms. *IEEE Transactions on Vehicular Technology, 65*(9), 6853–6860.
10. Wang, F., Zhen, Z., Wang, Bo., & Mi, Z. (2018). Comparative study on KNN and SVM based weather classification models for day ahead short term solar PV power forecasting. *Applied Sciences, 8*(1), 28.
11. Elangovan, K., Tamilselvam, Y. K., Mohan, R. E., Iwase, M., Takuma, N., & Wood, K. L. (2017). Fault diagnosis of a reconfigurable crawling–rolling robot based on support vector machines. *Applied Sciences, 7*(10), 1025.
12. Jan, S. U., Lee, Y.-D., Shin, J., & Koo, I. (2017). Sensor fault classification based on support vector machine and statistical time-domain features. *IEEE Access, 5*, 8682–8690.
13. Jain, M., Kumary, V., Gangopadhyay, R., & Debnath, S. (2015). Improved p-norm energy detector in Generalized κ-μ fading

channel for spectrum sensing in cognitive radio. In: *2015 International Conference on Communications, Signal Processing, and their Applications (ICCSPA'15)*, pp. 1–4. IEEE.

14. Miah, M. S., Schukat, M., & Barrett, E. (2018). An enhanced sum rate in the cluster based cognitive radio relay network using the sequential approach for the future Internet of Things. *Human-Centric Computing and Information Sciences, 8*(1), 1–27.
15. Albehadili, A., Ali, A., Jahan, F., Javaid, A. Y., Oluochy, J., & Devabhaktuniz, V. (2019). Machine learning-based primary user emulation attack detection in cognitive radio networks using pattern described link-signature (PDLS). *2019 Wireless Telecommunications Symposium (WTS)* (pp. 1–7). IEEE.
16. Furqan, H. M., Aygül, M. A., Nazzal, M., & Arslan, H. (2020). Primary user emulation and jamming attack detection in cognitive radio via sparse coding. *EURASIP Journal on Wireless Communications and Networking, 2020*(1), 1–19.

17 Miah, M. S., Hossain, M. A., Ahmed, K. M., Rahman, M., Calhan, A., & Cicioglu, M. (2021). Machine learning-based malicious user detection in energy harvested cognitive radio-internet of things. *TechRxiv*. https://doi.org/10.36227/techrxiv.13681672.v1

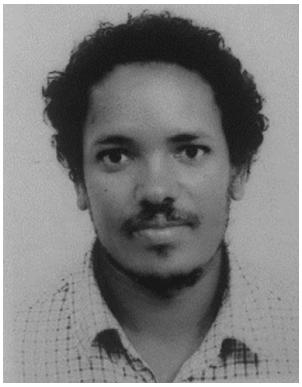
18. Li, H., & Han, Z. (2010). Dogfight in spectrum: Combating primary user emulation attacks in cognitive radio systems, part i: Known channel statistics. *IEEE Transactions on Wireless Communications, 9*(11), 3566–3577.
19. Arshid, K., Jianbiao, Z., Hanif, I., Munir, R., Yaqub, M., & Tariq, U. (2020). Energy detection based spectrum sensing strategy for CRN. *2020 IEEE international conference on artificial intelligence and information systems (ICAIIS)* (pp. 107–112). IEEE.
20. He, X., Dai, H., & Ning, P. (2013). Hmm-based malicious user detection for robust collaborative spectrum sensing. *IEEE Journal on Selected Areas in Communications, 31*(11), 2196–2208.
21. Khan, M. S., Khan, L., Gul, N., Amir, M., Kim, J., & Kim, S. M. (2020). Support vector machine-based classification of malicious users in cognitive radio networks. *Wireless Communications and Mobile Computing*. https://doi.org/10.1155/2020/8846948
22. Hossain, M. S., & Miah, M. S. (2021). Machine learning-based malicious user detection for reliable cooperative radio spectrum sensing in Cognitive Radio-Internet of Things. *Machine Learning with Applications, 5*, 100052.
23. Chhetry, B., & Marchang, N. (2021). detection of primary user emulation attack (PUEA) in cognitive radio networks using one-class classification. *arXiv preprint* arXiv:2106.10964
24. Cadena Muñoz, E., Martínez, L. F. P., & Triviño, J. E. O. (2020). Detection of malicious primary user emulation based on a support vector machine for a mobile cognitive radio network using software-defined radio. *Electronics, 9*(8), 1282.
25. Kaligineedi, P., Khabbazian, M., & Bhargava, V. K. (2010). Malicious user detection in a cognitive radio cooperative sensing system. *IEEE Transactions on Wireless Communications, 9*(8), 2488–2497.
26. Yadav, S., & Nene, M. J. (2013). Rss based detection and expulsion of malicious users from cooperative sensing in cognitive radios. *2013 3rd IEEE international advance computing conference (IACC)* (pp. 181–184). IEEE.
27. Li, T., Song, M., & Alam, M. (2008). Compromised sensor nodes detection: A quantitative approach. *2008 the 28th International conference on distributed computing systems workshops* (pp. 352–357). IEEE.
28. Qin, Z., Gao, Y., Plumbley, M. D., Parini, C. G., & Cuthbert, L. G. (2013). Low-rank matrix completion based malicious user detection in cooperative spectrum sensing. *2013 IEEE global conference on signal and information processing* (pp. 1186–1189). IEEE.
29. Liu, F., Cheng, X., & Chen, D. (2007). Insider attacker detection in wireless sensor networks. *IEEE INFOCOM 2007–26th IEEE international conference on computer communications* (pp. 1937–1945). IEEE.
30. Chen, C., Song, M., Xin, C., & Alam, M. (2012). A robust malicious user detection scheme in cooperative spectrum sensing. *2012 IEEE global communications conference (GLOBECOM)* (pp. 4856–4861). IEEE.
31. Wang, W., Li, H., Sun, Y., & Han, Z. (2009). Catchit: Detect malicious nodes in collaborative spectrum sensing. *GLOBECOM 2009–2009 IEEE Global Telecommunications Conference* (pp. 1–6). IEEE.
32. Yu, F. R., Tang, H., Huang, M., Li, Z., & Mason, P. C. (2009). Defense against spectrum sensing data falsification attacks in mobile ad hoc networks with cognitive radios. *MILCOM 2009–2009 IEEE military communications conference* (pp. 1–7). IEEE.
33. Lin, X., Li, J., Wu, J., Liang, H., & Yang, W. (2019). Making knowledge tradable in edge-AI enabled IoT: A consortium blockchain-based efficient and incentive approach. *IEEE Transactions on Industrial Informatics, 15*(12), 6367–6378.
34. Zhou, X., Sun, M., Li, G. Y., & Juang, B.-H.F. (2018). Intelligent wireless communications enabled by cognitive radio and machine learning. *China Communications, 15*, 16–48.
35. Gupta, A., Majumder, S. (2019). Cognitive radio spectrum classification using FLA-SVM. In: *Proceedings of the 2019 9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference (IEMECON), Jaipur, India*. pp. 144–149.
36. Mohanakurup, V., Baghela, V. S., Kumar, S., Srivastava, P. K., Doohan, N. V., Soni, M., & Awal, H. (2022). 5G Cognitive radio networks using reliable hybrid deep learning based on spectrum sensing. *Wireless Communications and Mobile Computing*. https://doi.org/10.1155/2022/1830497
37. Wang, Y., Zhang, Z., Ma, L., & Chen, J. (2014). SVM-based spectrum mobility prediction scheme in mobile cognitive radio networks. *The Scientific World Journal*. https://doi.org/10.1155/2014/395212
38. Lema, G. G. (2021). Handover performance evaluation under dynamic user characteristics. *Wireless Personal Communications, 120*(4), 2991–3004.
39. Xu, M., Yin, Z., Zhao, Y., & Wu, Z. (2022). Cooperative spectrum sensing based on multi-features combination network in cognitive radio network. *Entropy, 24*(1), 129.
40. Jan, S. U., Vu, V.-H., & Koo, I. (2018). Throughput maximization using an SVM for multi-class hypothesis-based spectrum sensing in cognitive radio. *Applied Sciences, 8*(3), 421.
41. Han, R., Du, L., Liu, T., & Chen, Y. (2017). SVM-GA based method for estimation of a large number of primary users in mobile cognitive radio networks. In: *Proceedings of the 3rd international conference on communication and information processing*.
42. Khan, M. S., Jibran, M., Koo, I., Kim, S. M., & Kim, J. (2019). A double adaptive approach to tackle malicious users in cognitive radio networks. *Wireless Communications and Mobile Computing, 2019*, 1–9.
43. Zhang, D., Zhai, X. (2011). SVM-based spectrum in cognitive radio. In: *2011 7th International Conference on Wireless Communications, Networking and Mobile Computing, Wuhan, China*

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Kaleem Arshid has received his B.Sc. in Information Technology from The University Of Azad Jammu and Kashmir Muzaffarabad in 2013, Pakistan and his MPhil in Computer Science from Preston University Kohat in 2017, Pakistan. From 2017 to 2019, he was working as a Lecturer in Department of Computer Science, University of Sargodha, Pakistan. Currently, he is a Doctoral student at Beijing University of Technology, Beijing, China. His research interest cognitive radio network security, machine learning, block chain, and 5G wireless networks.

Zhang Jianbiao has received his B.Sc., M.Sc., and Ph.D. degrees in computer science from Northwestern Polytechnic University, Xi'an, China, in 1992, 1995, and 1999, respectively. From 1999 to 2001, he was a Postdoctoral Fellow with Beijing University, Beijing, China. He is currently a Professor and a Ph.D. Supervisor with the Faculty of Information Technology, Beijing University of Technology. His research interests include network and information security and trusted computing. He has published over 80 journal/conference papers.

Iftikhar Hussain has graduated in computer science, specialized in software engineering in 2009. In 2017, he obtained his Ph.D. from Hasselt University, Belgium. He has worked on renowned research projects and taught undergraduate and graduate courses. He worked his research in the domain of agent-based modelling and developed multiple simulation systems for several supply-chain, transportation and waste management applications using dedicated modelling and simulation methodology (CRIO meta-model). Currently, he is working at Heriot-Watt University, Edinburgh Campus, UK. His research interests focus in the fields of system modelling and simulations, organizational-based and agent-based modelling, multi-agents in transportation, modelling interactions and negotiations, modelling and prediction of human behaviour, machine learning, and matching algorithms.



Gebrehiwet Gebrekrstos Lema has received his B.Sc in Electronics and Communication Engineering from Mekelle Institute of Technology in 2010 and his M.Sc in Communication Engineering from Ethiopian Institute of Technology-Mekelle in 2015. He was attending his Ph.D in TU of Ilmenau, Germany since April 2017 and currently he is working as an assistant professor in Mekelle University. His research interest focuses in cognitive radio network, QoS enhancement, antenna design, self-organized networks, cellular future networks, optimization technique, beam forming, radar systems, mobile and wireless communication, signal processing, data and computer networking.

Muhammad Yaqub received his B.Sc. and M.Sc. degree from COMSATS University Islamabad, Pakistan, in 2016 and 2019, respectively. From 2016 to 2018, he was a Research Assistant with the COMSATS University Islamabad, Pakistan. From Feb 2019, he has been a Lecturer with Department of Computer Science at The University of Lahore, Chenab Campus. He is currently pursuing his Ph.D. degree at Beijing University of Technology, Beijing, China. His research interest includes the Medical Image processing, Modern Artificial Intelligence Techniques, Computer vision and Machine learning. Mr. Author's awards and honors include the Higher Education Commission Fellowship (HEC Pakistan), the U.S.A MeritScholarship in Pakistan, and the ICT funded FYP Award.

Rizwan Munir has received his B.Sc. degree in Information Technology and the M.Sc. degree in Computer Science with AJK COMSATS University, Pakistan, in 2012 and 2016, respectively. He is currently pursuing his Ph.D. degree with Beijing University of Posts and Telecommunication, China. His research interests are involving wireless networks, software defined networks, and quality of service, particularly with applications of machine intelligence and deep learning.