Geo-Location Information Aided Spectrum Sensing in Cellular Cognitive Radio Networks [†]

Siji Chen [1] , **Bin Shen** [1,*] , **Xin Wang** [1] **and Sang-Jo Yoo** [2]

[1]   School of Communication and Information Engineering (SCIE), Chongqing University of Posts and Telecommunications (CQUPT), Chongqing 400065, China; cqchensj@outlook.com (S.C.); s170101160@stu.cqupt.edu.cn (X.W.)

[2]   Department of Information and Communication Engineering, Inha University, Incheon 402751, Korea; sjyoo@inha.ac.kr

[*]   Correspondence: shenbin@cqupt.edu.cn; Tel.: +86-23-6246-0222

[†]   This paper is an extended version of our paper published in Chen, S.; Shen, B.; Wang, X.; Wu, H. SVM and Decision Stumps Based Hybrid AdaBoost Classification Algorithm for Cognitive Radios. In Proceedings of the International Conference on Advanced Communications Technology (ICACT), PyeongChang, Korea, 17–20 February 2019.

check for
updates

Abstract: Apart from the received signal energy, auxiliary information plays an important role in remarkably ameliorating conventional spectrum sensing. In this paper, a novel spectrum sensing scheme aided by geolocation information is proposed. In the cellular cognitive radio network (CCRN), secondary user equipments (SUEs) first acquire their wireless fingerprints via either received signal strength (RSS) or time of arrival (TOA) estimation over the reference signals received from their surrounding base-stations (BSs) and then pinpoint their geographical locations through a wireless fingerprint (WFP) matching process in the wireless fingerprint database (WFPD). Driven by the WFPD, the SUEs can easily ascertain for themselves the white licensed frequency band (LFB) for opportunistic access. In view of the fact that the locations of the primary user (PU) transmitters in the CCRN are either readily known or practically unavailable, the SUEs can either search the WFPD directly or rely on the support vector machine (SVM) algorithm to determine the availability of the LFB. Additionally, in order to alleviate the deficiency of single SUE-based sensing, a joint prediction mechanism is proposed on the basis of cooperation of multiple SUEs that are geographically nearby. Simulations verify that the proposed scheme achieves higher detection probability and demands less energy consumption than the conventional spectrum sensing algorithms.

Keywords: spectrum sensing; geolocation; wireless fingerprint database; support vector machine; dynamic spectrum access; cognitive radio

1. Introduction

With the aim of identifying the status of the licensed spectrum and enabling secondary access, spectrum sensing techniques have been extensively investigated for cognitive radio (CR) in recent years [1–5]. In order to guarantee that, as little interference is generated to the primary user (PU) as possible, secondary users (SUs) in the cognitive radio network (CRN) can only gain opportunistic access to the licensed frequency band (LFB) when they detect no PU activity over it. However, the task of ascertaining PU behavior over the licensed spectrum is practically challenging if conventional spectrum sensing methods, e.g., energy detection (ED), are concerned, since the signals transmitted from the PUs are usually subject to complicated radio propagation effects before reaching the SUs, especially when the PU signals encounter severe power attenuation and fast fluctuations. In addition, if it is assumed

that the SUs in the CRN are cellular secondary user equipments (SUEs) (Correspondingly, the CRN is referred to as the cellular cognitive radio network (CCRN) hereinafter.) periodically collecting spectrum observations over the LFB and reporting them to their serving base-stations (BSs) for global decisions on the status of the LFB, it would probably arouse another problematic issue of power deficiency for battery driven SUE devices. In particular, the battery time that the SUEs could sustain may be significantly shortened due to the dual-band power consumption.

In the existing literature, the spectrum sensing techniques can be roughly categorized into four types. Firstly, threshold-test-based signal detection, such as ED [6–8], cyclostationary feature detection [9–11], and matched filtering detection [12–14] serve as common methods for the SUEs to gain awareness of the LFB's PU occupancy status. Secondly, for achieving higher accuracy and reliability, cooperative spectrum sensing (CSS) algorithms have been extensively investigated [15–18], where space diversity in the CRN is exploited and different hard decision fusion (HDF) schemes and soft data fusion algorithms are proposed under different criteria [15]. Thirdly, different from the single/multi-user-based signal detection schemes, some hybrid spectrum sensing strategies that combine signal processing techniques and geolocation databases have been proposed [19–21]. The geolocation database stores in advance some information about the PUs, such as the PU transmitter's (PUT's) power, geographical position, statistical behavior, and so forth. Supported by this additional a priori information, the database-based sensing approaches substantially relieve the complexity requirement and power demand on the SUEs' devices [22,23]. Fourthly, as one kind of promising solution, machine learning-based spectrum sensing (MLSS) schemes have also been researched in recent years [24–28], where it has been testified that some typical supervised and unsupervised machine learning methods work well in discerning the spectrum observations collected from different situations and even outperform some conventional CSS methods, for instance, the AND, OR, and Majority criteria-based HDF schemes. It is of special importance to note that the conventional threshold-test methods, CSS schemes, and MLSS algorithms usually operate on the spectrum observations only and aim to detect the weakest PU signal possible. It certainly helps significantly in relieving the hidden terminal effect, but the spectrum opportunity is unfortunately ignored when the SUEs are sufficiently far away from the PUs. On the other hand, the conventional geolocation database supported sensing method operates on the statistical knowledge that is drawn over a relatively long period. It, therefore, lacks the flexibility of being able to adapt to the transient behavior of the PUs and it may even fail to achieve the sensing agility requirement if a plug-and-play sensing method is demanded.

In order to strengthen the SUEs' capability of squeezing the opportunistic spectrum chances and enhance their energy efficiency in sensing, we propose a geolocation information aided spectrum sensing scheme on the basis of the wireless fingerprint database (WFPD) and machine learning (ML) algorithms. The philosophy of resorting to WFPD and ML algorithms in the proposed sensing scheme is twofold. Firstly, the WFPD helps the SUEs a lot in easily identifying their own geographical locations and hence they are able to search the database and easily obtain the LFB availability information corresponding to their current positions in the CCRN. This mechanism is of great benefit for the SUEs' energy efficiency in sensing because the conventional periodic acquisition of spectrum observations over the LFB is circumvented in a proactive manner. In this way, a large portion of energy consumed in scanning the LFB can be saved. Secondly, assisted by the information and data stored in the WFPD, it is easier for the SUEs to take spectrum sensing decisions when ML methods are adopted for processing the spectrum observations with high efficiency and precision. Furthermore, and to be more specific, aided by the WFPD, geolocation related spectrum availability for a specific SUE can be achieved according to the distance between the SUE and the PUTs. This geolocation information enables the SUEs to suppress the nuisance interference generated by themselves to the PU receivers in their vicinity. In practice, however, locations of the PUTs can be either readily known or completely unknown [29]. For the two different situations, wireless fingerprint (WFP) searching-based sensing scheme and the MLSS algorithm-based HDF scheme are proposed. It is verified in simulations that the proposed

scheme outperforms the conventional HDF-based CSS algorithms in terms of sensing performance. Meanwhile, the proposed scheme also achieves higher energy efficiency than the conventional periodic spectrum sensing schemes.

The rest of this paper is organized as follows. In Section 2, we describe the CCRN and present the signal model. Section 3 briefly reviews the non-geolocation information aided spectrum sensing methods. In Section 4, we propose the geolocation information aided spectrum sensing schemes for different scenarios. In Section 5, performance evaluation results for the proposed schemes are presented. Finally, Section 6 concludes the paper.

2. System Model

2.1. Cellular Cognitive Radio Network

We consider a CCRN, where there are multiple BSs serving a number of SUEs over the CCRN's own frequency band denoted by F_1. By means of any possible strategy or mechanism, the SUEs attempt to seize possible opportunities of accessing the LFB, denoted as F_2, to offload the traffic that could not be undertaken over their own frequency band F_1.

The target geographic area of the CCRN is divided into Q grids with the same area (The area of the grids reflects the spatial resolution requirement in status identification of the LFB.), where any specific SUE in the CCRN can be located in terms of the grid it is currently camping in. If one database is established in advance, with its data tables storing information about the availability of the LFB within each specific grid in the CCRN, then the problem of spectrum sensing could be easily solved by first positioning the SUEs and then searching the database for the pre-stored spectrum availability information. This kind of database-driven spectrum sensing mechanism is of particular benefit in practice. Normally as battery supported devices communicating with the BSs over F_1 and simultaneously monitoring F_2, the SUEs in the CCRN inevitably encounter a problem of power deficiency, which is a critical issue to be tackled. In this sense, with the support of the database, the SUEs are provided an option of determining the status of F_2 via simple database operations.

2.2. Signal Model

It is assumed that K BSs continuously operate over F_1 in the CCRN and simultaneously P PUTs serve their PU receivers over F_2 within the same geographic area. When communicating with its serving BS (the k-th BS) over F_1, the SUE in the q-th grid (Without loss of generality, we refer to the SUE in the q-th grid as the q-th ($q = 1, 2, \cdots, Q$) SUE in this paper, even though in practice there may be multiple SUEs camping in the q-th grid simultaneously.) receives the signal as

$$\bar{y}_{qk}(i) = \sqrt{\bar{E}_k} \bar{h}_{kq} \bar{s}_k(i) + \sum_{j=1, j \neq k}^{K} \sqrt{\bar{E}_j} \bar{h}_{jq} \bar{s}_j(i) + \bar{n}_q(i), \tag{1}$$

where i is the time index for signal samples, $\bar{E}_k$ is the transmit power of the k-th BS, $\bar{h}_{kq}$ is the channel coefficient from the k-th BS to the q-th SUE, $\bar{s}_k$ is the downlink signal transmitted by the k-th BS, and $\bar{n}_q(i) \sim \mathcal{CN}(0, \bar{\sigma}_0^2)$ is the complex additive white Gaussian noise (AWGN) corresponding to the q-th SUE in the F_1 band.

In addition to communicating over F_1, the q-th SUE keeps repeatedly observing the LFB F_2 with the sensing period T_s in conventional sensing methods. It obtains the signal received from the PUTs in the m-th sensing operation as

$$y_{qm}(i) = \begin{cases} \sum_{p=1}^{P} \sqrt{I^+(E_p^{(m)})} h_{pq}^{(m)} x_p^{(m)}(i) + n_q^{(m)}(i), & \mathcal{H}_1 \\ n_q^{(m)}(i), & \mathcal{H}_0 \end{cases} \tag{2}$$

where i stands for the time index within the m-th spectrum observation, $I^+(E_p^{(m)}) = \max(E_p^{(m)}, 0)$ with $E_p^{(m)} \in \{E_p, 0\}$ referring to the transmit power level (For the purpose of simplicity, we assume in this paper that the PUTs have only two possible transmit power levels, although, in practice, there are usually multiple transmit power levels that need to be identified.) of the p-th PUT, $h_{pq}^{(m)}$ denotes the channel coefficient from the p-th PUT to the q-th SUE, $x_p^{(m)}(i)$ is the transmit symbol of the p-th PUT, $n_q^{(m)}(i) \sim \mathcal{CN}(0, \sigma_0^2)$ is the complex AWGN of the q-th SUE in the F_2 band, and the hypotheses $\mathcal{H}_1$ and $\mathcal{H}_0$, respectively, represent the case of at least one PUT being busy and the case that all the PUTs are idle.

The channel coefficient $h_{pq}^{(m)}$ can be modeled as

$$h_{pq}^{(m)} = \sqrt{\mathrm{PL}(\|\mathbf{C}_p^{\mathrm{PUT}} - \mathbf{C}_q^{\mathrm{SUE}}\|)} \cdot \zeta_p \cdot v_p, \tag{3}$$

where $\mathbf{C}_p^{\mathrm{PUT}} = [C_{p,x}^{\mathrm{PUT}}\ C_{p,y}^{\mathrm{PUT}}]^T$ stands for the position of the p-th PUT, $\mathbf{C}_q^{\mathrm{SUE}} = [C_{q,x}^{\mathrm{SUE}}\ C_{q,y}^{\mathrm{SUE}}]^T$ is the position of the q-th SUE, $\| . \|$ is the Euclidean distance, $\mathrm{PL}(d) = d^{-a}$ is the path-loss component for the relative distance d with the path-loss exponent a, ζ_p is the shadowing component, and v_p is the multipath fading component in accordance with Rayleigh distribution.

Within the m-th sensing operation, the q-th SUE acquires the spectrum observation vector as $\mathbf{y}_{qm} = [y_{qm}(1), y_{qm}(2), \cdots, y_{qm}(2W\tau)]^T$, where W is the PU signal bandwidth in F_2 and τ is the sensing interval. Over M consecutive sensing interval τ within each T_s, the aggregate observations can be organized as $\mathbf{Y}_q = [\mathbf{y}_{q1}, \mathbf{y}_{q2}, \cdots, \mathbf{y}_{qM}]$, where M is the total number of spectrum observation segments. Based on the raw data $\mathbf{Y}_q$, the spectrum occupancy status of F_2 with respect to the q-th grid can be decided, through either conventional signal detection methods or MLSS algorithms. It is noteworthy that depending on the grid granularity in the CCRN, some neighboring grids may obtain the same spectrum decisions. Once all the grids requesting for spectrum opportunities are aware of their own spectrum status of F_2, the spectrum sensing task is fully fulfilled in the CCRN.

3. Non-Geo-Location Information Aided Spectrum Sensing

3.1. Conventional Threshold-Test-Based Spectrum Sensing Methods

3.1.1. Single User-Based Energy Detection

According to the well-known ED, the q-th SUE measures the received signal energy in $\mathbf{y}_{qm}$ as

$$Y_{qm} = \|\mathbf{y}_{qm}\|^2 = \sum_{i=1}^{2W\tau} |y_{qm}(i)|^2, \tag{4}$$

and obtains the test statistic Z_q and the spectrum decision D_q via the threshold-test

$$Z_q = tr(\mathbf{Y}_q \mathbf{Y}_q^H) = \sum_{m=1}^{M} \|\mathbf{y}_{qm}\|^2 \gtrless \lambda_q \begin{matrix} \nearrow D_q = \hat{\mathcal{H}}_1, \\ \searrow D_q = \hat{\mathcal{H}}_0, \end{matrix} \tag{5}$$

where λ_q is the pre-calibrated threshold, depending on the desired false alarm probability under the criterion of constant false alarm probability (CFAP), $\| . \|$ is the vector norm, $tr(.)$ is the matrix trace operator, and the test decision D_q is either $\hat{\mathcal{H}}_0 = 0$ for the case of all PUTs being idle or $\hat{\mathcal{H}}_1 = 1$ for the case of any PUT being active in the CCRN, respectively.

3.1.2. Multi-User-Based HDF Sensing

On obtaining the spectrum decision D_q, the q-th SUE may cooperate with multiple SUEs in its neighbourhood to strengthen the sensing reliability through the HDF criteria, e.g., AND, OR, and Vote, as

$$\Lambda_q = \sum_{t \in \mathbb{Q}_q, t \neq q} D_t + D_q \gtrless \eta \begin{array}{l} \nearrow \tilde{D}_q = \hat{\mathcal{H}}_1, \\ \searrow \tilde{D}_q = \hat{\mathcal{H}}_0, \end{array} \tag{6}$$

where D_t is the spectrum decision of the t-th SUE, $\mathbb{Q}_q$ is the index set of the q-th SUE's neighboring SUEs with the set cardinality denoted as $|\mathbb{Q}_q|$, and η is an integer threshold for the HDF schemes. It can be observed that the HDF scheme boils down to the AND scheme for $\eta = |\mathbb{Q}_q| + 1$, the OR scheme for $\eta = 1$, and the η_0-out-of-$(|\mathbb{Q}_q| + 1)$ (a.k.a Vote) scheme for $\eta = \eta_0$, respectively.

The sensing performance is usually quantified in terms of receiver operating characteristics (ROC), which gives the detection probability $P_D = Pr(\hat{\mathcal{H}}_1 | \mathcal{H}_1)$ as a function of the probability of false alarm $P_{FA} = Pr(\hat{\mathcal{H}}_1 | \mathcal{H}_0)$. By varying the detection threshold, the operating point of a detector can be chosen anywhere along its ROC curve.

3.2. Machine Learning-Based Spectrum Sensing

In recent years, machine learning-based classification algorithms have been proposed to identify the LFB status for CR systems, since the ultimate task of sensing is in some sense equivalent to classifying the spectrum observations as the data containing PU signal components or the data composed of noise only. Typically, ML algorithm such as support vector machine (SVM) and clustering algorithms can be utilized for spectrum sensing. For the purpose of simplicity, we assume in this paper that the transmit powers of the PUTs are fixed within each sensing period T_s, i.e., $E_p^{(m)} = E_{p,h}$ for all $m = 1, 2, \cdots, M$, and the p-th PUT has only two possible power levels $E_{p,h} \in \{E_p, 0\}$. Practically in the m-th sensing operation, the P PUTs may operate in mutually independent modes and hence there are in total 2^P operation states of the PUTs in the CCRN, where for the h-th state, the corresponding power vector of the PUTs can be expressed as

$$\mathbf{P}_h = \left[I^+(E_{1,h}), I^+(E_{2,h}), \cdots, I^+(E_{p,h}), \cdots, I^+(E_{P,h}) \right]^T, \ h \in \{0, 1, \cdots, 2^P - 1\}. \tag{7}$$

For the SVM-based sensing data classifier, we can get the q-th energy vector from Equation (4) within each sensing period T_s as

$$\boldsymbol{\theta}_q = \left[Y_{q1}, Y_{q2}, \cdots, Y_{qM} \right]^T, \tag{8}$$

where Y_{qm} is the received signal energy captured for the q-th SUE.

Prior to feeding $\boldsymbol{\theta}_q$ into the SVM, we need to obtain enough number of energy vectors to form the training data set. Given a sufficiently large number of energy vectors that may be accumulated in a relatively long time, we utilize the training data set and the training label set as

$$\begin{aligned} \boldsymbol{\Theta}_{q,L} &= \left\{ \boldsymbol{\theta}_q^{(1)}, \boldsymbol{\theta}_q^{(2)}, \cdots, \boldsymbol{\theta}_q^{(l)}, \cdots, \boldsymbol{\theta}_q^{(L)} \right\}, \\ \mathbf{C}_{q,L} &= \left\{ C_q^{(1)}, C_q^{(2)}, \cdots, C_q^{(l)}, \cdots, C_q^{(L)} \right\}, \end{aligned} \tag{9}$$

where L is the number of energy vectors adopted for training and the corresponding label $C_q^{(l)}$ is either $\hat{\mathcal{H}}_0$ or $\hat{\mathcal{H}}_1$. After the SVM classifier is successfully trained, it can be used to classify the spectrum observations. In addition, it is worth noting that for conventional ED and MLSS schemes, only when the PUTs are operating in the $h = 0$ state, i.e., $\mathbf{P}_0 = [0, 0, \cdots, 0]^T$, the label $C_q^{(l)}$ can be set as $\hat{\mathcal{H}}_0$. It is apparent that in fact the state $\mathbf{P}_0 = [0, 0, \cdots, 0]^T$ means that no active PUT is found in the CCRN at all and it is the only state under which the SUEs are allowed to access the LFB F_2. Therefore, it is

actually an over-strict condition for the SUEs to gain spectrum opportunities over F_2 only in case that the PUTs' state is identified as $\mathbf{P}_0$, whereas the spectrum opportunities possibly hidden in the states $\mathbf{P}_h, h \in \{1, 2, \cdots, 2^P - 1\}$, are ignored.

In the MLSS schemes, when the test data $\boldsymbol{\theta}_q$ is fed into the readily trained classifier and the label allocated to it is $\hat{C}_q$, the spectrum sensing accuracy can be evaluated as

$$P_A = \sum_{p=1}^{2} \mathbb{E}_q \left[Pr(\hat{C}_q = C_q | \mathcal{H}_p) Pr(\mathcal{H}_p) \right] \approx \sum_{p=1}^{2} \frac{1}{Q} \sum_{q=1}^{Q} I(\hat{C}_q = C_q) Pr(\mathcal{H}_p), \tag{10}$$

where $\mathcal{H}_p$ has only two possible hypothesis as $\mathcal{H}_1$ and $\mathcal{H}_2$, C_q is the true label of $\boldsymbol{\theta}_q$, $\mathbb{E}[.]$ is the mathematical expectation, $Pr(A)$ is the probability of an event A, and $I(.)$ is the indicator function.

4. Geo-Location Information Aided Spectrum Sensing

In the previous section, the conventional signal detection-based sensing schemes and the MLSS algorithms obtain the spectrum decisions from the spectrum observations only. Without additional information that may be helpful in sensing, the aforementioned schemes can merely acquire the spectrum decisions under the very stringent constraint of ascertaining whether PU signal components exist in the spectrum data under test. Drawbacks of these methods are apparent. Firstly, a relatively large portion of the potential spectrum opportunities in the CCRN may be wasted. Even when the SUE is sufficiently far away from the active PUTs, it can not be granted permission to access F_2. Interestingly, this problem may occur when cooperation of some SUEs is evoked to enhance the reliability in sniffing the weak PU signals (Bearing in mind that hidden terminal effect may be harmful, we usually find that the conventional spectrum sensing schemes are designed to prevent the SUEs from accessing the LFB when weak PU signals are detected in the spectrum observations.). Secondly, without offline history information of the spectrum status, the SUEs need to repeatedly monitor the LFB F_2, even when its status is relatively stable or only varying slowly over time. This case actually imposes a heavy burden on the power consumption of the SUEs.

Aiming to tackle these problems encountered in conventional spectrum sensing schemes, we propose to exploit geolocation information to assist the SUEs in sensing operations. Tailored for geolocation information-based sensing scheme, the WFPD can be viewed as an indispensable infrastructure consisting of a large variety of different data. Specifically, aided by the WFPD, the SUEs are easy to locate themselves in the CCRN, with short time and high precision. When the positions of the PUTs are possibly available to the SUEs and the WFPD stores records of the PU behaviors, the SUEs are capable of quickly identifying the LFB availability with respect to their own locations.

4.1. Geographical Region-Based Spectrum Status Identification

According to the robustness requirement of spectrum sensing stipulated in the IEEE 802.22 standard [30], a region division strategy over the active PUTs is considered in the CCRN. The target geographical area is divided into black, grey, and white regions, as shown in Figure 1 for example. Centered with an active PUT, the black region with radius D_t is closely surrounded by the grey region with inner radius D_t and outer radius D_p. Meanwhile, the grey region is surrounded by the white region. Inside the black region, the PUTs have exclusive right to use the LFB F_2 and the SUEs are absolutely forbidden to operate over F_2, thus tremendously eliminating any possible interference to the PU receivers inside it. The temporal spectrum opportunities may be found in the grey region, where the SUEs can opportunistically gain access, and in the white region, where the SUEs are of sufficiently long distances from the PU receivers and therefore are allowed to transmit with their maximum power at any time, without causing severe interference to the PU receivers.
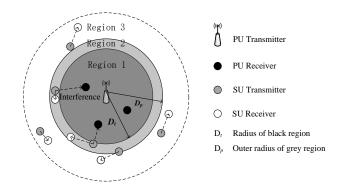
Figure 1. Geo-regions-based spectrum status identification.

Specifically, radius of the black region surrounding the *p*-th active PUT can be defined as [31]

$$D_t \leq \beta_t \left(\frac{E_p}{\sigma_0^2 (2^{\xi_0} - 1)} \right)^{1/a} \tag{11}$$

where ξ_0 is a threshold constant and β_t is an adjusting coefficient. The outer radius of the gray region is

$$D_p = \left[\frac{2 E_p W M \tau}{\sigma_0^2 \left(q_f \sqrt{2 W M \tau} + q_m^2 + q_m \sqrt{2 W M \tau + 2 q_f \sqrt{2 W M \tau} + q_m^2} \right)} \right]^{1/a} \tag{12}$$

where $q_f = Q^{-1}(\varepsilon_f)$, $q_m = Q^{-1}(\varepsilon_m)$, the parameter ε_f and ε_m are constraint constants of P_{FA} and P_{M} (probability of miss detection), respectively, with $P_{\text{FA}} \leq \varepsilon_f$, $P_{\text{M}} \leq \varepsilon_m$ required, and $Q^{-1}(.)$ is the inverse *Q*-function. For the purpose of ensuring the quality of PU communication and guaranteeing the spectrum opportunities for the SUEs, the region partitioning can be flexibly adjusted, as shown in Figure 2.
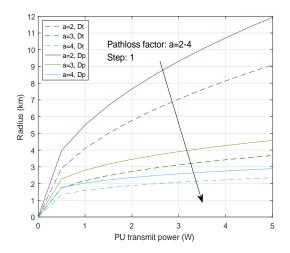

Figure 2. Relationship between the region radius and primary user transmitter (PUT) power.

The spectrum availability of F_2 in different regions can be denoted by the flags given in Table 1, where the PUT_Flag is set identical with the Region_Flag. The PUT_Flag with value 1 means no spectrum opportunity over F_2 is found for the SUEs, −1 means the SUEs are safe to transmit over F_2, and 0 implies potential spectrum opportunities might be discovered in the grey region. As an enabling label for the SUEs to access F_2, the next step of spectrum sensing is to identify the region that the SUEs are camping in.

Table 1. Region oriented spectrum availability.

Region Type	Region_Flag	PUT_Flag	Availability of F_2 for the SUEs
Black Region	1	1	strictly non-accessible
Grey region	0	0	uncertain
White region	−1	−1	freely accessible

4.2. WFPD Aided SUE Positioning

After defining the regions in the CCRN, the next step of spectrum sensing is to find the locations of the SUEs. When the SUEs are able to know their own locations, they can make a quick spectrum decision based on the spectrum availability information stored in the WFPD. Since it is well known that wireless positioning technology based on received signal strength (RSS) and time of arrival (TOA) are common methods that may be utilized practically, they are both employed in this paper. The TOA positioning technique refers to the method of estimating the time that the BS downlink reference signal takes to arrive at the SUE. Observing the surrounding BSs' downlink reference signals in routine operations such as synchronization tracking and reference signal received power (RSRP) measurement, the SUEs can measure the downlink reference signals' RSS and TOA of neighboring BSs to locate themselves in the CCRN [32,33].

Based on the received signal in Equation (1), the q-th SUE utilizes the locally generated reference signal over F_1 to seek the peak of the correlation output as the TOA. The signal correlation is first obtained as

$$R_{qk}(l) = |\sum_{l_1=1}^{L_1} \bar{\mathbf{y}}_{qk}(l+l_1) \cdot \mathbf{r}_{q,\text{local}}(l_1)|^2, \tag{13}$$

where $\bar{\mathbf{y}}_{qk} = \left[\bar{y}_{q1}, \bar{y}_{q2}, \cdots, \bar{y}_{qL_2}\right]^T$ is the k-th BS's downlink reference signal received by the q-th SUE, $\bar{\mathbf{y}}_{qk}(l+l_1)$ denotes the $(l+l_1)$-th signal element in $\bar{\mathbf{y}}_{qk}$, $\mathbf{r}_{q,\text{local}} = \left[\bar{r}_{q1}, \bar{r}_{q2}, \cdots, \bar{r}_{qL_1}\right]^T$ is the local signal generated by the q-th SUE, $\mathbf{r}_{q,\text{local}}(l_1)$ is the l_1-th signal sample, and the parameters L_1 and L_2 are, respectively, the length of the local signal and the length of the received signal.

Therefore, the estimate of TOA of the k-th BS at the q-th SUE is obtained as

$$T_{qk} = \frac{1}{F_s} \arg\max_l (R_{qk}(l)), \tag{14}$$

where F_s is the sampling frequency of the BS's downlink signal.

Before the SUEs can truly perform sensing operations, it is assumed that one WFPD has already been established in advance, as shown in Table 2, and it is fully accessible for all the SUEs in the CCRN. Since there are Q grids in the CCRN, the number of WFPs in the WFPD is set as Q, too. Each WFP consists of the TOA estimations (For the case of RSS information-based WFPD, RSS data is used instead.) with respect to the K BSs, the corresponding K BS-IDs, PUT_Position_Flag, Region_Flag, the LFB F_2's availability flag (PUT_Flag), the LFB F_2's status update timer (Update_Timer), received signal $\{\mathbf{y}_{qm}\}_{m=1}^M$ and signal energy $\{Y_{qm}\}_{m=1}^M$. In addition, there is a public flag PUT_State_Flag_h among the Q WFPs, where h indicates that all the data in Table 2 are actually obtained from the PUTs' power state $\mathbf{P}_h$.

If the q-th SUE is triggered to sense the LFB F_2, it first sends its TOA estimations, i.e., the TOA fingerprint, to the WFPD. The TOA fingerprint is a combination of the estimated TOAs and the corresponding BS-IDs, e.g., the q^*-th WFP $\{\text{TOA}_{kq^*}\}_{k=1}^K$. The q-th SUE is located in the q^*-th grid based on the WFP search when its WFP best matches the q^*-th WFP in the WFPD as

$$q^* = \arg\min_{q' \in \{1,2,\cdots,Q\}} \beta_{q'} = \arg\min_{q' \in \{1,2,\cdots,Q\}} \sqrt{\sum_{k=1}^K \left[T_{qk} - \text{FP}_{q'}(k)\right]^2}, \tag{15}$$

where β'_q is the root square of the TOA estimation error, q^* is the index of the WFP that best matches the fingerprint reported by the q-th SUE, and $\text{FP}_{q'}(k)$ represents the k-th fingerprint feature, $\text{TOA}_{kq'}$ in Table 2, i.e., the TOA estimation of the reference signal from the k-th BS to the q'-th grid in the CCRN. In addition, Equation (15) requires that $\beta_{q^*} < \psi$ and $\beta_{q'} > \psi$, $q' \in \{1, 2, \cdots, Q\}/q^*$, where ψ is the threshold used to guarantee the maximum tolerable error in localizing the q-th SUE. If $\beta_{q^*} > \psi$, the q-th SUE fails to find the grid it is camping in and has to search the WFPD with its newly measured fingerprint again. It is worth mentioning that depending on the grid area and the error threshold ψ, WFP duplications [29] may be encountered by the SUEs. This situation usually results from the small grid area and a high ψ setting in practice. In this case, the WFP duplications would not give rise to wrong spectrum decisions because the grids that have the same WFP will have the same spectrum label as well.

Table 2. Data table of the time of arrival (TOA) wireless fingerprint database.

Data Type	WFP 1	WFP 2	$\cdots$	WFP q	$\cdots$	WFP Q
Geo-Location	$\mathbf{C}_1^{(\text{SUE})}$	$\mathbf{C}_2^{(\text{SUE})}$	$\cdots$	$\mathbf{C}_q^{(\text{SUE})}$	$\cdots$	$\mathbf{C}_Q^{(\text{SUE})}$
WFP Feature	TOA_{11} TOA_{21}	TOA_{12} TOA_{22}	$\cdots$ $\cdots$	TOA_{1q} TOA_{2q}	$\cdots$ $\cdots$	TOA_{1Q} TOA_{2Q}
(TOA)	$\vdots$ TOA_{K1}	$\vdots$ TOA_{K2}	$\vdots$ $\cdots$	$\vdots$ TOA_{Kq}	$\vdots$ $\cdots$	$\cdots$ TOA_{KQ}
Spectrum Label	Region_Flag_1 PUT_Flag_1	Region_Flag_2 PUT_Flag_2	$\cdots$ $\cdots$	Region_Flag_q PUT_Flag_q	$\cdots$ $\cdots$	Region_Flag_Q PUT_Flag_Q
Timer	Update_Timer_1	Update_Timer_2	$\cdots$	Update_Timer_q	$\cdots$	Update_Timer_Q
Most Recent Observations	$\mathbf{y}_{11}^h, \mathbf{y}_{12}^h, \cdots, \mathbf{y}_{1M}^h$ $Y_{11}^h, Y_{12}^h, \cdots, Y_{1M}^h$	$\mathbf{y}_{21}^h, y_{22}^h, \cdots, \mathbf{y}_{2M}^h$ $Y_{21}^h, Y_{22}^h, \cdots, Y_{2M}^h$	$\cdots$ $\cdots$	$\mathbf{y}_{q1}^h, \mathbf{y}_{q2}^h, \cdots, \mathbf{y}_{qM}^h$ $Y_{q1}^h, Y_{q2}^h, \cdots, Y_{qM}^h$	$\cdots$ $\cdots$	$\mathbf{y}_{Q1}^h, \mathbf{y}_{Q2}^h, \cdots, \mathbf{y}_{QM}^h$ $Y_{Q1}^h, Y_{Q2}^h, \cdots, Y_{QM}^h$
Public Labels	PUT_State_Flag_h, $h \in \{0, 1, \cdots, 2^P - 1\}$; PUT_Position_Flag					

4.3. Grid Oriented Spectrum Decision Making

In the previous subsection, the q-th SUE finds its location in the q^*-th grid via WFP matching in the WFPD. In order to determine which region the q^*-th grid belongs to, the distances between the q^*-th grid and the active PUTs need to be evaluated. However, calculating the distances demands geolocation information of the q^*-th grid and the active PUTs, whereas in practice $\mathbf{C}_p^{\text{PUT}}$ may be either readily available or completely unknown for the SUEs. It actually imposes two different situations to be dealt with by different approaches.

4.3.1. PUT's Geo-Location Information Aided Spectrum Decision Making

When the locations of the PUTs in the CCRN are readily known to the SUEs, it is set in the WFPD that PUT_Position_Flag $= 1$. Once the location of the q-th SUE is determined as in the q^*-th grid, the flag Region_Flag_q^* can be immediately determined, depending on its distance to the PUTs and the predefined region radius D_t and D_p, as

$$\text{Region\_Flag\_}q^* = \begin{cases} 1 & \sum_{p=1}^{P} \|\mathbf{C}_p^{\text{PUT}} - \mathbf{C}_{q^*}^{\text{SUE}}\| \cdot I(\hat{E}_{p,h} = E_p) < \|\hat{\mathbf{P}}_h\|_0 \cdot D_t, \\ 0 & \|\hat{\mathbf{P}}_h\|_0 \cdot D_t \leq \sum_{p=1}^{P} \|\mathbf{C}_p^{\text{PUT}} - \mathbf{C}_{q^*}^{\text{SUE}}\| \cdot I(\hat{E}_{p,h} = E_p) \leq \|\hat{\mathbf{P}}_h\|_0 \cdot D_p, \\ -1 & \sum_{p=1}^{P} \|\mathbf{C}_p^{\text{PUT}} - \mathbf{C}_{q^*}^{\text{SUE}}\| \cdot I(\hat{E}_{p,h} = E_p) > \|\hat{\mathbf{P}}_h\|_0 \cdot D_p, \end{cases} \quad (16)$$

where $\hat{\mathbf{P}}_h$ is the power vector corresponding to the PUTs' current operating state classified by the machine learning algorithms as $\hat{\mathcal{C}}_{q^*} = h$, $\hat{E}_{p,h}$ is the power level accredited for the p-th PUT according to the h-th state of the P PUTs, and $\|\cdot\|_0$ is the l_0-norm. As for the label Region_Flag_q^*,

Region_Flag_q^* = 1 means that the SUE in the q^*-th grid is strictly prohibited from accessing the LFB F_2, Region_Flag_q^* = 0 implies that the q^*-th grid is inside in the grey area and the SUEs inside it might cause interference to the potential PU receivers in its vicinity when they transmit over F_2, and Region_Flag_q^* = -1 claims that the q^*-th grid is sufficiently far away from any active PUT in the CCRN and hence the SUEs in the q^*-th grid can freely gain access to the LFB without yielding any interference to the PU receivers in the proximity of themselves.

In this way, the unavailable/available/uncertain status of the LFB F_2 can be determined as PUT_Flag_q^* = $1/-1/0$, respectively, depending on the region that the q^*-th grid belongs to. In the proposed sensing scheme, PUT_Flag_q^* serves as the final sensing decision for the SUE located in the q^*-th grid in the CCRN. The Update_Timer_q^* in Table 2 counts the time since the last update operation of PUT_Flag_q^*. When it reaches an upper limit φ, it automatically returns to zero and sets the PUT_Flag_q^* as 0, meaning that the status of the PUT needs to be reconfirmed. If no data can be obtained for updating the PUT_Flag_q^*, it is better to keep it as 0 to prevent possible interference to the PU receivers. It is of particular importance to note that when PUT_Position_Flag_q^* = 0, the PUT_Flag_q^* needs to be set as 0, meaning that Equation (16) is not applicable due to the lack of $\mathbf{C}_p^{PUT}$. It is therefore necessary to determine whether the LFB F_2 is accessible by the q-th SUE, by means of either conventional sensing methods or MLSS algorithms. In other words, without knowledge of the PUTs' locations, the SUEs are no longer able to decide the region that their locations belong to, but have to merely rely on the received signal energy to identify the status of the F_2. Figure 3 gives a depiction of the grid oriented spectrum sensing scenario, where the q-th SUE first performs TOA estimation based on the received reference signals from the three BSs in the CCRN, and it is then able to locate itself through WFP matching operation in the WFPD. Subsequently, the q-SUE can identify its spectrum label from the WFP it matches in the WFPD or through different grid oriented spectrum decision making methods described within this subsection.
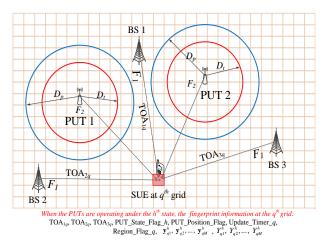

When the PUTs are operating under the hth state, the fingerprint information at the qth grid:
TOA$_{1q}$, TOA$_{2q}$, TOA$_{3q}$, PUT_State_Flag_h, PUT_Position_Flag, Update_Timer_q,
Region_Flag_q, $\mathbf{y}_{q1}^h$, $\mathbf{y}_{q2}^h$, ..., $\mathbf{y}_{qM}^h$, Y_{q1}^h, Y_{q2}^h, ..., Y_{qM}^h

Figure 3. Grid-based sensing scenario and data collection.

4.3.2. Machine Learning Aided Spectrum Decision Making

Intrinsically as classifiers or clustering algorithms, typical ML methods, e.g., SVM [34], K-means [35], and K-nearest neighbors (KNN) [36], can be employed for identifying the PUT_Flag_q^* for the q-th SUE. In the sequel, we take the SVM classification algorithm as an example (Compared to the SVM algorithm adopted for spectrum decision making, the proposed scheme can be implemented with the K-means and KNN algorithms in a similar workflow. The difference between employing SVM and K-means lies in that the SVM algorithm requires a readily available label set in training whereas the K-means algorithm only demands the number of clusters in training, significantly relieving the implementation requirement.) and describe how the spectrum sensing decision is made for the SUEs in the q-th grid.

Similar to the conventional MLSS methods, we utilize the training data set and the corresponding label set as

$$
\begin{aligned}
\boldsymbol{\Theta}_{q,L} &= \left\{ \boldsymbol{\theta}_q^{(1)}, \boldsymbol{\theta}_q^{(2)}, \cdots, \boldsymbol{\theta}_q^{(l)}, \cdots, \boldsymbol{\theta}_q^{(L)} \right\}, \\
\mathbb{C}_{q,L} &= \left\{ \mathcal{C}_q^{(1)}, \mathcal{C}_q^{(2)}, \cdots, \mathcal{C}_q^{(l)}, \cdots, \mathcal{C}_q^{(L)} \right\},
\end{aligned}
\tag{17}
$$

where the training label set $\mathbb{C}_{q,L}$ is different form the training label set $\mathbf{C}_{q,L}$ in Equation (9) in that $\mathcal{C}_q^{(l)} \in \{0, 1, \cdots, 2^P - 1\}$.

The SVM originally provides a binary model in machine learning which strives to find a linearly separable hyperplane with the help of support vectors that lie closest to the decision surface by maximizing the margin of the classifier while minimizing the sum of classification errors [24], as shown in Figure 4, where $\mathbf{x}^{(n)}$ is the n-th training sample, $\mathbf{x}^{*(s)}$ is the s-th test sample, $y^{(n)}$ and $y^{*(s)}$ are, respectively, their corresponding labels, and the number of training samples and test samples is N and S, respectively.
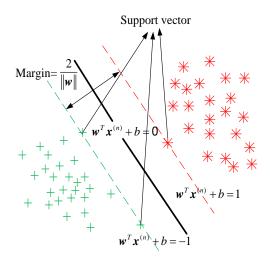

Figure 4. Support vector machine (SVM) model © 2019 IEEE [28].

The learning strategy of SVM is to maximize the margin shown in Figure 4 and its learning goal is to find a hyperplane in the multi-dimensional samples space. The hyperplane equation can be expressed as

$$
\boldsymbol{\omega}^T \mathbf{x}^{(n)} + b = 0,
\tag{18}
$$

where $\boldsymbol{\omega}$ is the weighting vector and b is the bias.

During training, the SVM should satisfy the following condition for all $n = 1, 2, \cdots, N$ as

$$
y^{(n)} = \begin{cases} 1, & \text{if } \boldsymbol{\omega}^T \mathbf{x}^{(n)} + b \geq 1, \\ -1, & \text{if } \boldsymbol{\omega}^T \mathbf{x}^{(n)} + b \leq -1. \end{cases}
\tag{19}
$$

We need to minimize the vector norm of $\boldsymbol{\omega}$ so as to maximize the margin, and hence the objective function is

$$
\begin{array}{ll}
\text{Min} & \frac{1}{2} \|\boldsymbol{\omega}\|^2 \\
\text{s.t.} & y^{(n)} (\boldsymbol{\omega}^T \mathbf{x}^{(n)} + b) \geq 1.
\end{array}
\tag{20}
$$

where $\boldsymbol{\omega}$ and b in the optimal hyperplane can be obtained by solving the objective function.

In practice, when a test sample $\mathbf{x}^{*(s)}$ is fed into the SVM model, the SVM can determine which class it belongs to through the following rules

$$
\hat{y}^{*(s)} = \begin{cases} 1, & \text{if } \boldsymbol{\omega}^T \mathbf{x}^{*(s)} + b \geq 1, \\ -1, & \text{if } \boldsymbol{\omega}^T \mathbf{x}^{*(s)} + b < -1, \end{cases}
\tag{21}
$$

where $\hat{y}^{*(s)}$ is the predicted label of the *s*-th binary test sample.

However, practically the samples are not always linearly separable. For this case, the hyperplane satisfying such conditions does not exist at all. Then, we need to find a fixed nonlinear mapping function $\phi(.)$ to map the non-linear samples into a new feature space and use a linear SVM in the feature space [34]. Hence, the nonlinear SVM should satisfy the following condition for all $n = 1, 2, \cdots, N$ as

$$y^{(n)} = \begin{cases} 1, & \text{if} \quad \boldsymbol{\omega}^T \phi(\mathbf{x}^{(n)}) + b \geq 1, \\ -1, & \text{if} \quad \boldsymbol{\omega}^T \phi(\mathbf{x}^{(n)}) + b \leq -1. \end{cases} \tag{22}$$

The decision rule for nonlinearly SVM is given as

$$\hat{y}^{*(s)} = \begin{cases} 1, & \text{if} \quad \boldsymbol{\omega}^T \phi(\mathbf{x}^{*(s)}) + b \geq 1, \\ -1, & \text{if} \quad \boldsymbol{\omega}^T \phi(\mathbf{x}^{*(s)}) + b < -1. \end{cases} \tag{23}$$

While the training energy vectors have been mapped into a higher dimensional feature space, practically we cannot achieve a perfect linearly separable hyperplane that satisfies the condition in Equation (20) for each $\mathbf{x}^{(n)}$. Therefore, we rewrite the optimization problem as a convex optimization problem as follows

$$\begin{aligned} \text{Min} \quad & \tfrac{1}{2}\|\boldsymbol{\omega}\|^2 + \Lambda \sum_{n=1}^{N} \xi^{(n)} \\ \text{s.t.} \quad & y^{(n)}(\boldsymbol{\omega}^T\mathbf{x}^{(n)} + b) \geq 1 - \xi^{(n)}, n = 1, 2, \cdots, N \\ & \xi^{(n)} \geq 0, n = 1, 2, \cdots, N, \end{aligned} \tag{24}$$

where Λ is a soft margin constant, for which a larger Λ means the assignment of a higher penalty to errors, and $\xi^{(n)}$ is a slack variable. As it is well known that the Radial basis function (rbf) kernel is a popular kernel function used in various kernelized learning algorithms to map a feature space to a higher dimension, it is adopted for the SVM classifier in this paper.

In order to obtain the final spectrum sensing decision for the SUEs located in the q^*-th grid, PUT_Flag_q^*, there are two types of SVM classifiers that need to be utilized in two consecutive stages. The first type SVM classifier, denoted as T1-SVM, is first evoked to classify the current operating state of the PUTs as $\hat{\mathcal{C}}_{q^*} \in \{0, 1, \cdots, 2^P - 1\}$ and the second type SVM classifier, denoted as T2-SVM, is subsequently triggered to identify the availability of the LFB F_2 as PUT_Flag_$q^* \in \{1, 0, -1\}$. For the training of the T1-SVM classifiers, it is assumed that there are enough number of SUEs pre-traversing all the grids in the CCRN and experiencing all the 2^P operating states of the P PUTs over a relatively long time, for example a duration of hundreds or even thousands of spectrum sensing periods T_s. In this way, $\boldsymbol{\Theta}_{q,L}$ and $\mathbb{C}_{q,L}^{(1)}$ are collected with a sufficiently large L and used as the training data set and label set for the T1-SVM classifier, respectively, with $C_{q,L}^{(1)} \in \{0, 1, \cdots, 2^P - 1\}$. For the T2-SVM classifier training, the classification models are trained under different operating states of the PUTs. Given the label $C_{q,L}^{(1)} = h$ from the T1-SVM, the training data set $\breve{\boldsymbol{\Theta}}_{q,h} = \{\breve{\boldsymbol{\theta}}_q^{(1)}, \breve{\boldsymbol{\theta}}_q^{(2)}, \cdots, \breve{\boldsymbol{\theta}}_q^{(L_h)}\} \subset \boldsymbol{\Theta}_{q,L}$ is taken from $\boldsymbol{\Theta}_{q,L}$, with $C_{q,L}^{(1)}$ identifying the PUTs' operating state that the data set $\breve{\boldsymbol{\Theta}}_{q,h}$ is obtained under. The label set for T2-SVM training is $\mathbb{C}_{q,h}^{(2)}$ with $C_{q,h}^{(2)} \in \{1, 0, -1\}$. Specifically, for the training process, we have $\sum_{h=0}^{2^P-1} L_h = L$ and $\bigcup_{h=0}^{2^P-1} \breve{\boldsymbol{\Theta}}_{q,h} = \boldsymbol{\Theta}_{q,L}$. After training, the current operating state of the PUTs can be predicted by the T1-SVM classifier as a label $\hat{\mathcal{C}}_q^{(1)} = h$, indicating that the current state is the *h*-th state. Subsequently, under the $\hat{\mathcal{C}}_q^{(1)}$-th PUTs' state, the T2-SVM operates to determine the final spectrum sensing decision PUT_Flag_q^* as $\hat{\mathcal{C}}_q^{(2)}$.

Unlike the binary SVM classifier presented in Equation (23), T1-SVM and T2-SVM are required to work as non-binary classifiers and hence they need to be modified for the purpose of being able to make multiple classifications. We adopt a one-versus-all (OVA) [37,38] scheme to fulfill the multi-classification task. Without loss of generality, assuming that a four-class problem for the T1-SVM

classification model is to be solved, i.e., $P = 2$, we need a series of four SVM classifiers, denoted as c_1, c_2, c_3, and c_4, respectively. During the learning phase, the i-th binary classifier c_i outputs a label '+1' when it is determined that the training data belongs to the positive class, whereas it produces a label '−1' when the training data is classified as being in the negative class. With four binary SVMs operating in the OVA scheme, the task of multi-classification can be completed. As shown in Table 3, the binary SVM classifier c_1 in the first row is trained by assigning the positive label to it while the remaining c_2, c_3, and c_4 binary SVM classifiers are assigned the negative label. The classification label l_1 is actually a combination of the individual labels of the four classifiers in the first row and the label l_1 is represented by the label set of the four SVM classifiers as one final label. Similarly, for the second classifier c_2 in row 2 of Table 3, the positive label is assigned to it and the negative label is assigned to all the other classifiers. In general, for the c_i classifier in the i-th row of Table 3, we assign the positive label to it and the negative label to the remaining classifiers in the l_i row. In this way, the multiple classification task to obtain $\hat{C}_q^{(1)}$ and $\hat{C}_q^{(2)}$ can be fulfilled with different number of different binary SVM classifiers, respectively.

Table 3. Classification coding matrix of the one-versus-all (OVA) scheme.

	c_1	c_2	c_3	c_4
l_1	+1	−1	−1	−1
l_2	−1	+1	−1	−1
l_3	−1	−1	+1	−1
l_4	−1	−1	−1	+1

Due to the limited spectrum observation of the q-th SUE, $\boldsymbol{\theta}_q$, it is straightforward to exploit spatial diversity to ameliorate the spectrum sensing accuracy through SUE cooperation. Specifically, we assume the q'-th SUE ($q' \in \mathbb{Q}_q$) stays in one of the $|\mathbb{Q}_q|$ grids surrounding the q-th grid and they may have the same true PU_Flag_q', $q' \in \mathbb{Q}_q$ with that of the q-th grid. Accordingly, the spectrum sensing scheme only depending on $\boldsymbol{\theta}_q$ is defined in this paper as a single-SUE-based spectrum prediction scheme and the one depending on $\boldsymbol{\theta}_q$ and the spectrum observations $\boldsymbol{\theta}_{q'}$, $q' \in \mathbb{Q}_q$ from the SUEs in $\mathbb{Q}_q$ as a joint-SUE spectrum prediction scheme.

For the single-SUE spectrum prediction, the spectrum sensing decision for the SUE located in the q-th grid is predicted as PU_Flag_q. For the joint-SUE spectrum prediction, a sum of the labels from the q-th grid and its neighboring grids in $\mathbb{Q}_q$ is firstly obtained as

$$\bar{\Lambda}_q = \text{PU\_Flag\_}q + \sum_{t \in \mathbb{Q}_q} \text{PU\_Flag\_}t, \tag{25}$$

and the final decision PU_Flag_q is made as

$$\text{PU\_Flag\_}q = \begin{cases} -1 & \bar{\Lambda}_q \leq \bar{\eta}_1 \\ 0 & \bar{\eta}_1 < \bar{\Lambda}_q < \bar{\eta}_2 \\ 1 & \bar{\Lambda}_q \geq \bar{\eta}_2 \end{cases} \tag{26}$$

where $\bar{\eta}_1$ and $\bar{\eta}_2$ are the lower and upper integer threshold designed for the joint-SUE prediction, respectively.

The proposed geolocation information aided spectrum sensing scheme is described in detail in Algorithm 1.

Algorithm 1 Geo-location information aided spectrum sensing.

Input: $\Theta_{q,L}$, $\mathbb{C}_{q,L}$, T_{qk}, WFPD.　　% for the q-th SUE, $q = 1, 2, \cdots, Q$

Output: PUT_Flag_q.

1: train the q-th T1-SVM classifier via $\{\Theta_{q,L}, \mathbb{C}^{(1)}_{q,L}\}$

2: train the q-th T2-SVM classifier via $\{\breve{\Theta}_{q,h}, \mathbb{C}^{(2)}_{q,h}\}^{2^P-1}_{h=0}$

3: the q-th SUE estimates K TOAs (fingerprint) from the K BSs over F_1

4: **if** $\beta_{q^*} < \psi$ and it satisfies (15) **then**

5:　　the estimated fingerprint best matches the q^*-th WFP in the WFPD

6:　　the q-th SUE is located in the q^*-th grid

7: **else** return to step 2

8: **end if**

9: the q-th SUE reads the labels in the q^*-th WFP and obtain

10:　　$\{$PUT_Position_Flag_q^*, Update_Timer_q^*, PUT_Flag_$q^*\}$

11: **if** PUT_Position_Flag_$q^* = 1$ **then**

12:　　**if** Update_Timer_$q^* < \varphi$ **then**

13:　　　　**if** PUT_Flag_$q^* \neq 0$ **then**

14:　　　　　　output PUT_Flag_q = PUT_Flag_q^*

15:　　　　　　Exit to step 33

16:　　　　**end if**

17:　　**end if**

18: **end if**

19: **if** Update_Timer_$q^* = \varphi$ or PUT_Flag_$q^* = 0$ **then** the q-th SUE

20:　　% single-SUE prediction

21:　　obtains the sensing observation $\boldsymbol{\theta}_q$ over the LFB F_2

22:　　predicts the state of the PUTs as $\hat{\mathcal{C}}^{(1)}_{q^*}$ via $\boldsymbol{\theta}_q$ and the trained T1-SVM classifier

23:　　predicts PUT_Flag_q^* as $\hat{\mathcal{C}}^{(2)}_{q^*}$ via $\boldsymbol{\theta}_q$, $\hat{\mathcal{C}}^{(1)}_{q^*}$ and the trained T2-SVM classifier

24:　　**if** Joint-SUE prediction is required **then**

25:　　　　reads received signals of the $|\mathbb{Q}_q|$ grids that neighbor the q^*-th grid

26:　　　　predicts PUT_Flags_q', $q' \in \{\mathbb{Q}_q\}$ independently, via single-SUE prediction

27:　　　　obtains PUT_Flag_q^* by HDF voting based on PUT_Flags_q', $q' \in \{\mathbb{Q}_q \cup \{q^*\}\}$

28:　　**end if**

29:　　update PUT_Flag_q^* in the q^*-th WFP

30:　　clear and restart Updata_Timer_q^* in the q^*-th WFP

31: **end if**

32: output PUT_Flag_q = PUT_Flag_q^*

33: **if** A new sensing request occurs **then** return to step 2

34: **end if**

35: Exit

Similar to Equation (10), the spectrum sensing accuracy of the proposed scheme is

$$
\begin{aligned}
P_A &= \sum_{h=0}^{2^P-1} \mathbb{E}_q \left[Pr(\mathcal{H}_h) Pr(\hat{\mathcal{C}}_q^{(1)} = \mathcal{C}_q^{(1)} | \mathcal{H}_h) Pr\left(\hat{\mathcal{C}}_q^{(2)} = \mathcal{C}_q^{(2)} | \hat{\mathcal{C}}_q^{(1)} = \mathcal{C}_q^{(1)}, \mathcal{H}_h \right) \right] \\
&\approx \frac{1}{Q} \sum_{q=1}^{Q} \sum_{h=0}^{2^P-1} Pr(\mathcal{H}_h) I\left(\hat{\mathcal{C}}_q^{(2)} = \mathcal{C}_q^{(2)}, \hat{\mathcal{C}}_q^{(1)} = \mathcal{C}_q^{(1)} | \mathcal{H}_h \right),
\end{aligned}
\tag{27}
$$

where $\mathcal{H}_h$ is the h-th hypothesis indicating that the PUTs operate under the h-th state.

In order to compare the average energy consumption of the q-th SUE in the conventional non-geolocation aided spectrum sensing scheme and the proposed geolocation-based spectrum sensing scheme, we adopt the following energy consumption calculation as

$$
E_{q,\mathrm{avg}} = \lim_{N=(N_1+N_2)\to\infty} \frac{1}{N_1+N_2} \left[\underbrace{(N_1+N_2)E_{q,F_1} + (N_1+N_2)E_{\mathrm{WFP},F_1}}_{E_1} + \underbrace{v_{q,F_2} N_2 \cdot (E_{\mathrm{scan},F_2} M + E_{\mathrm{report},F_1})}_{E_2} \right],
\tag{28}
$$

where E_{q,F_1} is the power consumption of the q-th SUE in routine operations such as synchronization and signal measurement over the F_1 frequency band, E_{WFP,F_1} is the power consumed in WFP matching operations in the WFPD, E_{scan,F_2} is the power consumed in capturing energy observations of spectrum samples over the LFB F_2, E_{report,F_1} is the power consumption of reporting the spectrum observations to the q-th SUE's serving BS for classification operations, $v_{q,F_2} \in \{0,1\}$ indicates whether the q-th SUE identifies the status of the LFB F_2 through the WFP matching procedure ($v_{q,F_2} = 0$) or it is required to make a decision for the LFB F_2 spectrum ($v_{q,F_2} = 1$) based on the spectrum observations, N_1 and N_2 are, respectively, the number of operations performed over a relatively long period of time, corresponding to the E_1 and E_2 power consumption. As it is easy to observe that when $v_{q,F_2} = 0$, the power consumption of E_2 is saved because the q-th SUE is capable of discerning the availability of F_2 based on the WFPD only. With a well established WFPD, periodical spectrum observation acquisition from the LFB F_2, which is intrinsically demanded by the non-geolocation-based spectrum sensing scheme, is thus fully circumvented and the power consumption is significantly reduced because instead of operating on dual frequency bands F_1 and F_2 simultaneously, the q-th SUE only operates on the single frequency band F_1.

5. Simulation and Analysis

In this section, the performance of the proposed geolocation information aided spectrum sensing scheme is evaluated via Matlab 2016b and compared with the conventional sensing algorithms. The conventional scheme refers to the energy detection-based sensing scheme [6,7,15], including the AND and OR criteria-based hard decision fusion schemes. These schemes are claimed as conventional because they do not rely on the WFP matching mechanism but keep periodically scanning the licensed spectrum. We consider a 6 km× 6 km CCRN area consisting of 5625 grids, with the size for each grid as 80 m × 80 m. In simulations, the main parameters are given in Table 4, where their values are chosen under practical concern. For example, the minimum bandwidth of F_2 is set as 5 MHz, which is a good choice for the 4 G LTE technology-based SUEs. The time-frequency product $W\tau$ is 500, which is sufficiently large for obtaining the energy sample.

Table 4. Simulation parameters.

Parameter	Value
CCRN Area	6 km $\times$ 6 km
Grid Area	80 m $\times$ 80 m
Number of BSs, K	3
Number of Grids, Q	5625
Number of PUTs, P	2
Bandwidth of F_2, W	5 MHz
Sensing Interval, τ	100 μs
Sensing Period, T_s	100 ms
Maximum Time in Update_Timer_q, φ	1000 T_s
Number of sensing segments, M	10
Pathloss component, α	4
Shadowing component, ζ_p	2 dB
Fading component, ν_p	5 dB
Length of data for T1-SVM training, L	22,500
Length of data for T2-SVM training, L_h	5625

Figure 5 depicts the spectrum sensing scenario, where there are three BSs and two PUTs in the CCRN. When the two PUTs are both active, Figure 6a presents the ideally identified three regions for accessing the LFB F_2 and the regions predicted by the proposed algorithm are shown in Figure 6b. It is worth noting that Figure 6 is drawn from the grid oriented spectrum decisions, where the q-th grid in the figure is displayed as white, black, or grey according to PUT_flag_$q \in \{-1, 1, 0\}, q \in \{1, 2, \cdots, Q\}$.

For the WFPD aided SUE localization, we compare the positioning accuracy performance of the RSS and TOA schemes, as shown in Figure 7. It is easy to see that the differences between the actual positions and the estimated positions of the SUEs are apparent for the RSS scheme, whereas the differences are trivial for the TOA estimate-based scheme. It means that the accuracy of TOA-based positioning is much better than that of the RSS-based scheme, because the latter one is inherently more sensitive to the strength variations of the received reference signal over F_1.
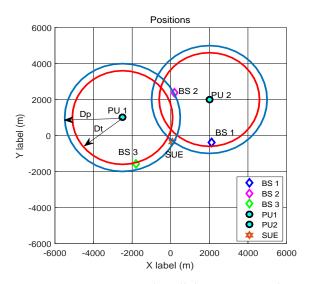

Figure 5. Spectrum sensing scenario in the cellular cognitive radio network (CCRN).
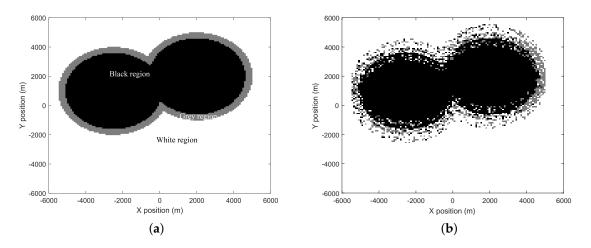
Figure 6. Region division when there are two active PUTs in the CCRN. (**a**) Regions ideally identified when there are two active PUTs in the CCRN. (**b**) Regions predicted when there are two active PUTs in the CCRN (E_P = 5 W).
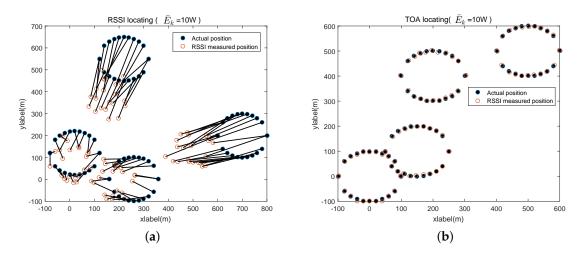

Figure 7. Accuracy comparison of TOA and RSS-based SUE positioning. (**a**) Accuracy of received signal strength (RSS)-based secondary user equipment (SUE) positioning. (**b**) Accuracy of time of arrival (TOA)-based SUE positioning.

For the power efficiency, the conventional spectrum sensing scheme and the proposed WFP-based spectrum sensing scheme are compared in Figure 8. It is assumed that both E_{scan,F_2} and E_{WFP,F_1} are approximately equal to E_{q,F_1}, and E_{report,F_1} is only a portion, 30%, of E_{WFP,F_1}. When the PUTs' locations are already known to the SUEs or the spectrum availability information stored in the WFPD is within the newly updated period, the SUEs are able to determine the spectrum availability by searching the WFPD only. In other words, it is not necessary for the the SUEs to operate in the dual-band mode, and, therefore, the power consumptions of E_{scan,F_2} and E_{report,F_1} are totally saved. It is shown that the proposed scheme saves much more energy than the conventional non-WFPD aided sensing scheme. In the conventional sensing scheme, it requires the SUEs to keep sensing the LFB F_2 periodically whenever there is a request to access it, whereas for the proposed scheme, the SUEs only need to sense the spectrum on condition that the F_2 status in the WFPD is outdated or there is a request to obtain the spectrum observations over F_2.
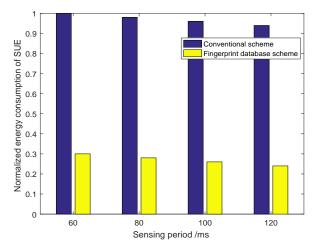
Figure 8. Power consumption of conventional sensing scheme and the wireless fingerprint database (WFPD) aided scheme with different sensing period T_s.

For the joint-SUE prediction-based sensing gain, we compare the spectrum prediction accuracies for the joint-SUE prediction with $|\mathbb{Q}_q| = 8$ and the single-SUE-based prediction, using SVM, KNN, and K-means algorithms. Figure 9 gives the prediction accuracy according to the distance between the SUE and one active PUT, where the PUT is located at the 0 km point and the grey region is about from 2~3.5 km. It is shown that the prediction accuracy is not satisfactory within the grey region, because it suffers an ambiguity in discerning the data collected from the boundaries of the three regions. As shown, the prediction accuracy achieves 100% in the black and white regions and decreases when the SUE moves around the grey area. Due to the fact that the energy observations collected from the region borders are statistically indiscernible and it is hard for ML algorithms to classify them, the SUEs need to be conservative to gain spectrum opportunities in this region. We also compare KNN, SVM, K-means, and conventional sensing schemes in Figure 10, where the ROC curves for different schemes are depicted. The proposed SVM aided sensing scheme outperforms the other ML algorithms-based scheme and the conventional ED-based sensing schemes.
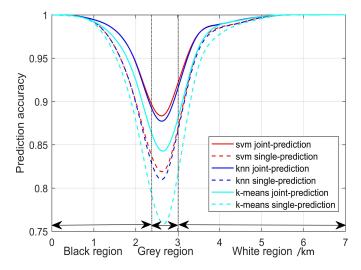

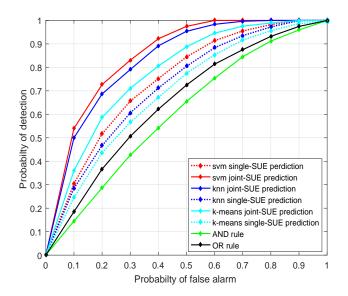**Figure 9.** Prediction accuracy of different machine learning (ML) algorithms.

Figure 10. Receiver operating characteristics (ROC) of different sensing algorithms.

6. Conclusions

In this paper, a geolocation information-based spectrum sensing mechanism is proposed for the SUEs in the CCRN. By formulating the first task of sensing as identifying the positions of the SUEs through the WFP matching operation in the WFPD, we tackle the second task as ascertaining the grid oriented spectrum availability through either spectrum labels in the WFPD or ML algorithm aided spectrum observation classification. On the condition that the PUTs' locations are readily known, the SUEs just need to check the LFB occupancy status in the WFPD or the distance between the SUE and the active PUTs, whereas when the PUTs' locations are unknown, the SUEs have to gather data from their neighboring grids to obtain the final spectrum decision, with the help of MLSS algorithms. Simulation results verified that the TOA estimation-based WFP scheme is superior to the RSS-based scheme for the first task. As for the grid oriented spectrum decision making mechanism, the SVM algorithm is verified to achieve higher spectrum prediction accuracy than the KNN and K-means algorithms. Meanwhile, the proposed scheme exhibits the best performance in terms of detection probability, compared with the ED-based HDF methods. Since the problem of only two transmit power levels of the PUTs are investigated in this paper, the methods and analysis for the case of multiple power levels are to be addressed in future work. Due to its salient power-saving capability in sensing operations, the proposed geolocation information aided spectrum sensing scheme can be used as one practical candidate solution in the CCRN.

Author Contributions: Conceptualization, B.S.; Data curation, S.C.; Formal analysis, S.C.; Investigation, B.S. and S.-J.Y.; Software, S.C. and X.W.; Supervision, B.S.

References

1. Mitola, J., III; Maguire, G.Q. Cognitive radio: Making software defined radio more personal. *IEEE Pers. Commun.* **1999**, *6*, 13–18. [CrossRef]
2. Axell, E.; Leus, G.; Larsson, E.G. Overview of spectrum sensing for cognitive radio. In Proceedings of the 2nd International Workshop on Cognitive Information Processing, Elba Island, Italy, 14–16 June 2010; pp. 322–327.
3. Meena, M.; Bhagari, F.; Rajendran, V. Spectrum sensing using cognitive radio technology. In Proceedings of the International Conference on Communication and Signal Processing (ICCSP), Melmaruvathur, India, 6–8 April 2017; pp. 1654–1657.

4.   Sun, D.; Song, T.; Gu, B.; Li, X.; Hu, J.; Liu, M. Spectrum sensing and the utilization of spectrum opportunity tradeoff in cognitive radio network. *IEEE Commun. Lett.* **2016**, *20*, 2442–2445. [CrossRef]

5.   Awin, F.; Abdel-Raheem, E.; Tepe, K. Blind spectrum sensing approaches for interweaved cognitive radio system: A tutorial and short course. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 238–259. [CrossRef]

6.   Urkowitz, H. Energy detection of unknown deterministic signals. *Proc. IEEE* **1967**, *55*, 523–531. [CrossRef]

7.   Sobron, I.; Diniz, P.S.R.; Martins, W.A.; Velez, M. Energy detection technique for adaptive spectrum sensing. *IEEE Trans. Commun.* **2015**, *63*, 617–627. [CrossRef]

8.   Makarfi, A.U.; Hamdi, K.A. Interference analysis of energy detection for spectrum sensing. *IEEE Trans. Veh. Technol.* **2013**, *62*, 2570–2578. [CrossRef]

9.   Gardner, W.A.; Napolitano, A.; Paura, L. Cyclostationarity: Half a century of research. *Signal Process.* **2006**, *86*, 639–697. [CrossRef]

10. Jang, W.M. Blind cyclostationary spectrum sensing in cognitive radios. *IEEE Commun. Lett.* **2014**, *18*, 393–396. [CrossRef]

11. Lunden, J.; Koivunen, V.; Huttunen, A.; Poor, H.V. Collaborative cyclostationary spectrum sensing for cognitive radio systems. *IEEE Trans. Signal Process.* **2009**, *57*, 4182–4195. [CrossRef]

12. Van, T.H.L. *Detection, Estimation, and Modulation Theory: Part I*; John Wiley and Sons, Inc.: Hoboken, NJ, USA, 1968.

13. Conte, E.; Lops, M.; Ricci, G. Adaptive matched filter detection in spherically invariant noise. *IEEE Signal Process. Lett.* **1996**, *3*, 248–250. [CrossRef]

14. Lin, W.; Liu, K.; Wang, C. Differential matched filter architecture for spread spectrum communication systems. *Electron. Lett.* **1996**, *32*, 1539–1540. [CrossRef]

15. Axell, E.; Leus, G.; Larsson, E.G.; Poor, H.V. Spectrum Sensing for cognitive radio: State-of-the-art and recent advances. *IEEE Signal Process. Mag.* **2012**, *29*, 101–116. [CrossRef]

16. Guo, H.; Jiang, W.; Luo, W. Linear soft combination for cooperative spectrum sensing in cognitive radio networks. *IEEE Commun. Lett.* **2017**, *21*, 1573–1576. [CrossRef]

17. Kim, J.; Choi, J.P. Sensing coverage-based cooperative spectrum detection in cognitive radio networks. *IEEE Sens. J.* **2019**, *19*, 5325–5332. [CrossRef]

18. Jia, M.; Liu, X.; Gu X.; Guo, Q. Joint cooperative spectrum sensing and channel selection optimization for satellite communication systems based on cognitive radio. *Int. J. Satell. Commun. Netw.* **2015**, *35*, 139–150. [CrossRef]

19. Barrie, M.; Delaere, S.; Sukareviciene, G.; Gesquiere, J.; Moerman, I. Geolocation database beyond TV white spaces? Matching applications with database requirements. In Proceedings of the IEEE International Symposium on Dynamic Spectrum Access Networks, Bellevue, WA, USA, 16–19 October 2012; pp. 467–478.

20. Fanan, A.; Riley, N.G.; Mehdawi, M.; Alfahad, O. Implementation of combined geolocation database and infrastructure sensing in TV bands using different spectrum devices. In Proceedings of the 25th Telecommunication Forum (TELFOR), Belgrade, Serbia, 21–22 November, 2017; pp. 1–4.

21. Awin, F.A.; Alginahi, Y.M.; Abdel-Raheem, E.; Tepe, K. Technical issues on cognitive radio-based internet of things systems: A Survey. *IEEE Access* **2019**, *7*, 97887–97908. [CrossRef]

22. Mauwa, H.; Bagula, A.; Zennaro, M.; Pietrosemoli, E.; Lysko, A.; Brown, T.X. Systematic analysis of geolocation and spectrum sensing as access methods to TV white space. In Proceedings of the ITU Kaleidoscope: ICTs for a Sustainable World (ITU WT), Bangkok, Thailand, 14–16 November 2016; pp. 1–8.

23. Gao, B.; Bhattarai, S.; Park, J.M.J. Incentivizing spectrum sensing in database driven dynamic spectrum sharing. In Proceedings of the 35th Annual IEEE International Conference on Computer Communications, San Francisco, CA, USA, 10–14 April 2016; pp. 1–9.

24. Thilina, K.M.; Choi, K.W.; Saquib, N.; Hossain, E. Machine learning techniques for cooperative spectrum sensing in cognitive radio networks. *IEEE J. Sel. Areas Commun.* **2013**, *31*, 2209–2221. [CrossRef]

25. Li, Z.; Wu, W.; Liu, X.; Qi, P. Improved cooperative spectrum sensing model based on machine learning for cognitive radio networks. *IET Commun.* **2018**, *12*, 2485–2492. [CrossRef]

26. Tian, J.F.; Cheng, P.; Chen, Z.; Li, M.Q.; Hu, H.L.; Li, Y.H.; Vucetic, B. A machine learning-enabled spectrum sensing method for OFDM systems. *IEEE Trans. Veh. Technol.* **2019**, *68*, 11374–11378. [CrossRef]

27. Bkassiny, M.; Li, Y.; Jayaweera, S.K. A survey on machine-learning techniques in cognitive radios. *IEEE Commun. Surv. Tutorials* **2013**, *15*, 1136–1159. [CrossRef]

28. Chen, S.; Shen, B.; Wang, X.; Wu, H. SVM and Decision Stumps Based Hybrid AdaBoost Classification Algorithm for Cognitive Radios. In Proceedings of the 2019 21st International Conference on Advanced Communication Technology (ICACT), PyeongChang, Korea, 17–20 February 2019; pp. 492–497.
29. Qureshi, U.M.; Umair, Z.; Hancke, G.P. Problem of wireless fingerprint duplication in fingerprinting-based indoor localization. In Proceedings of the 2018 IEEE 27th International Symposium on Industrial Electronics (ISIE), Cairns, QLD, Australia, 13–15 June 2018; pp. 1270–1275.
30. Cordeiro, C.; Challapali, K.; Birru, D.; Shankar, S. IEEE 802.22: The first worldwide wireless standard based on cognitive radios. In Proceedings of the First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, Baltimore, MD, USA, 8–11 November 2005; pp. 328–337.
31. Wei, Z.; Feng, Z.; Zhang, Q.; Li, W. Three regions for space-time spectrum sensing and sccess in cognitive radio networks. *IEEE Trans. Veh. Technol.* **2015**, *64*, 2448–2462. [CrossRef]
32. Sadowski, S.; Spachos, P. RSSI-based indoor localization with the internet of things. *IEEE Access* **2018**, *6*, 30149–30161. [CrossRef]
33. Guvenc, I.; Chong, C.C. A survey on TOA-based wireless localization and NLOS mitigation techniques. *IEEE Commun. Surv. Tutor.* **2009**, *11*, 107–124. [CrossRef]
34. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
35. Peng, X.; Zhou, C.; Hepburn, D.M.; Judd, M.D.; Siew, W.H. Application of K-Means method to pattern recognition in on-line cable partial discharge monitoring. *IEEE Trans. Dielectr. Electr. Insul.* **2013**, *20*, 754–761. [CrossRef]
36. Zhang, S.; Li, X.; Zong, M.; Zhu, X.; Wang, R. Efficient kNN classification with different numbers of nearest neighbors. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 1774–1785. [CrossRef]
37. Shu, W.; Cai, K. A SVM multi-class image classification method based on DE and KNN in smart city management. *IEEE Access* **2019**, *7*, 132775–132785. [CrossRef]
38. Wang, S.; Guo, X.; Tie, Y.; Lee, I.; Qi, L.; Guan, L. Graph-based safe support vector machine for multiple classes. *IEEE Access* **2018**, *6*, 28097–28107. [CrossRef]