Ensemble Deep Learning Based Cooperative Spectrum Sensing with Stacking Fusion Center

Hang Liu*, Xu Zhu† and Takeo Fujii‡
*‡Advanced Wireless and Communication Research Center, The University of Electro-Communications
†Corporate Research and Development Center, Toshiba Corporation

***Abstract*—In this paper, an ensemble learning (EL) framework is adopted for cooperative spectrum sensing (CSS) in an orthogonal frequency division multiplexing (OFDM) signal based cognitive radio system. Each secondary user (SU) is accordingly considered as a base learner, where the local spectrum sensing is for investigating the probability of PU being inactive or active. The convolution neural networks with simple architecture are applied given its strength in image recognition as well as the limited computation ability of each SU, meanwhile, the cyclic spectral correlation feature is introduced as the input data. Here, as for the supervised learning, the bagging strategy is helped to establish the training database. For the global decision, the fusion center employs the stacked generalization for further combination learning the SU output of classification pre-prediction of the PU status. Our method shows significant advantages over conventional CSS methods in term of the detection probability or false alarm probability performance.**

I. Introduction

On account of the explosively increasing spectrum resource demands in wireless communication, cognitive radio (CR) as an intelligent system, is proposed to tackle the challenge of surmounting the bandwidth limitations, by adapting its parameters to recycle allotted channels. In CR system, spectrum sensing can preserve the spectrum entitlement for primary user (PU), while assists secondary user (SU) to opportunistically access the idle bands. However, an non-cooperative spectrum sensing may suffer unreliable detection caused by PU experiencing shadowing [1] and multi-path fading [2].

In contrast, cooperative spectrum sensing (CSS) yields a better solution to overcome the hidden terminal problem by locating multiple spatial distributed SUs [3]. In CSS, some of the non-CSS efficient approaches can be directly adopted, such as energy detection [4] and cyclostationary feature detection [5], then the local detection results will be submitted to a fusion center (FC) for final decision of PU's activity status. However, two main global decision schemes usually suffer disappointment: a hard fusion, where individual SU makes one-bit decision of 0/1 (inactive/active PU) from locally recorded data, often results in inferior performance due to the uncomprehensive combination rule [6]; and a soft fusion, which performs centralized processing of estimated parameters collected from SUs, loses competitiveness due to some tough requirement of the prior knowledge [7].

Although machine learning has been applied in CSS, the unsatisfactory capability of learner or system structure [6][8] cannot make the best use of combination strategy of CSS to provide accomplished results. In this paper, an ensemble learning (EL), as an application methodology in machine learning, is adapted into CSS to compensate the mentioned defectiveness. In EL, multiple weak learners are integrated by certain ensemble strategy to form one strong learner [9]. Obviously, the SUs are therefore considered as weak learners when their sensing is to predict the probability of PU activity status: the convolutional neural network (CNN) with fine image processing ability, is employed locally but only executed in simple structure for computation limitation and low requirement thanks to EL; such CNN is trained by databases formed by bagging strategy [10]; the cyclostationary feature plane induced by pilots tones [11] aids the learning as the input data. The FC can be realized by particular ensemble strategy: another deep neural network (DNN) learner is adopted for global decision as the stacking generalization strategy from EL [12]; the predictions of PU status are reported in as the input for the FC learner.

The contributions and novelties are as follows: (1) aiming at orthogonal frequency division multiplexing (OFDM) signal based CR system, a novel ensemble deep learning based CSS scheme is proposed, which shows a performance advantage over conventional methods; (2) fine strategies including bagging and stacking from EL are adapted and localized in CSS for better detection ability; (3) the investigations of how SU learner-structure and number affect the performance are implemented and provided for future reference.

The paper is organized as follows: the system model and framework is given in Section II; the detailed scheme of ensemble deep learning based CSS is described in Section III; the result evaluations and conclusions are discussed in Sections IV and V, respectively.

II. System Model and Framework

A. System Fundamental Model of CSS

An OFDM signal based CSS is investigated in this research. The considered cooperative system shares an allocated spectrum channel with one PU which is assumed having a generalized model that alternate between active and inactive states in the transmission environment. The centralized CSS network is comprised of M SUs with the index of $i = 1, ..., M$ and located in the two-dimensional space. For each individual SU, previous s-tudies suggest an effective binary hypothesis to illustrate the local spectrum sensing, which is considered as

$$\begin{aligned} H_0 : x_i[n] &= w_i[n], \quad &\text{Inactive PU} \\ H_1 : x_i[n] &= s[n] + w_i[n], \quad &\text{Active PU} \end{aligned} \tag{1}$$

where n is the sampling index, s, x and w is the transmitted signal from PU, the received signal at one SU end, and the additive white Gaussian noise (AWGN), respectively. Each SU reports its local sensing result independently to FC where the global decision of PU activity status is made by according to certain combination rule. The system model is shown in Fig. 1.

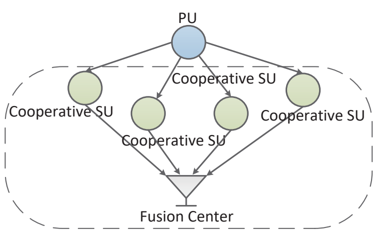


Fig. 1. Cooperative spectrum sensing system model.

B. Framework of Ensemble Deep Learning Based CSS

1) Ensemble Learning Overview: The concept of EL is to improve a prediction by creating a strong high-accuracy learner which combines the outputs of a collection of base/weak learners. The EL can be established from a set of learners of different algorithms or independent learners of the same algorithm. Either way, the multiple base learners learn the original datasets from their overlapping or independent databases in the training phase respectively and then draw the learning conclusion respectively. However, each single learner tends to give a biased prediction which can be fixed by the "ensemble" strategy, which means the poor results will be integrated and optimized by certain combination scheme, and a more accurate decision will be obtained.

2) Establishment of Ensemble Deep Learning Based CSS: As we mentioned before, the brilliance of CSS based CR system is that all the distributed cooperative SUs are united as a collective organization where a fine global decision is given by a competent FC after weighting all local sensing results by certain rule. Such principle is also the key point to the EL where all the base learners combine by certain combination strategy to form a stronger learner and provide the final decision. The commonality of characteristics allow us to perform the EL into the CSS; therefore the advantages and fine unique strategies of the former can affect and help improve the detection performance. The established EL based CSS is shown in Fig. 2.

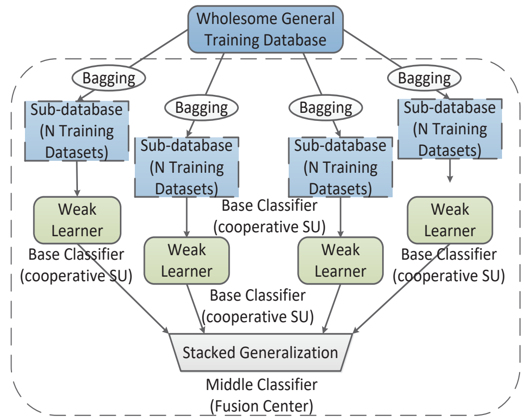


Fig. 2. Cooperative spectrum sensing system model.

In the EL based CSS, each SU strength is further developed and consolidated, while a slight of deficiency is compensated. However, in term of the weak-learner selection, a comparative fine one which can provide a valuable reference is still favored. The DNNs that is proved of its competence for CR, can assist us here as the base learner to achieve higher accuracy. Furthermore, since the base learner is required to study and differentiate inactive/active PU, the base learner is also analogous to a weak/base classifier. As is shown in Fig. 2, inspired by the bagging strategy in EL, which improves the accuracy by training each weak learner using sub-databases independently and randomly drawn/generated from a wholesome general database. In this research, such general training database is established beforehand, then a fixed predetermined number of datasets are randomly picked out to form sets of sub-databases for the training of base classifiers. Besides, the stacked generalization or stacking strategy in EL is adopted for FC to integrate the base learning results. The stacking strategy is a combining learner which is trained to make a final prediction using all the predictions of the other learners as additional inputs. In this research, according to such stacked generalization scheme, a DNN is applied again and required to make the final decision as the middle learner/classifier.

III. Concrete Scheme of Ensemble Deep Learning Based CSS

A. Local Sensing - Classification Converted Spectrum Sensing

In the EL based CSS, each SU learner is required to achieve effective local sensing for offering valuable contributions to the global decision, where the required sensing information is the perditions of PU's statuses.

Firstly, for better understanding and modeling, the local detection is converted to signal classification. Inspired by Eq. (1), H_0 and H_1 are treated as two categories to state the inactive and active PU [13], as:

- C_0: PU is inactive - only AWGN existing in the transmission environment;
- C_1: PU is active - PU signal is currently transmitting in the AWGN transmission environment.

Then, a hard decision of C_0 or C_1 from the base classifier, will be collected and reported to the FC.

1) Data Collection: The built-in signal structures of OFDM signals including a preamble, framed data packages and inserted pilot tones help it achieve robustness against multi-path interference. Therefore, for the considered supervised, feature-extraction-based learning, an especially valuable feature of the PU signal can be extracted as the input data for base classification.

Normally, when random user data is transmitted on the individual sub-carriers, the spectral correlation is expected to be zero. The built-in pilot tones are the only exceptions, where the correlated data is transmitted in parallel on dedicated sub-carriers according to various distribution strategies. This can result in the spectral feature contributions to the cyclic signature which helps implement signal detection and identification [14].

One can decompose a signal which owns a series of weighted spectral components into sinusoidal waveforms, to realize statistical spectrum analysis. The second-order spectral expression deeply engages with the definition of cyclic spectral analysis. Assuming that a signal $x(t)$ is exhibiting second order periodical, an important expression to measure its spectral correlation is the cyclic correlation spectral periodogram, as

$$S_{X_T}^{\alpha}(f) = \left[X_T\left(t, f+\frac{\alpha}{2}\right)\cdot X_T^{*}\left(t, f-\frac{\alpha}{2}\right)\right]/T, \quad (2)$$

with frequency components $f+\alpha/2$, $f-\alpha/2$, where f and α are coordinates spectral location (shift center) and spectral separation (shift amount), respectively. Then, the spectral coherence density (SCD) is defined as :

$$S_{X_T}^{\alpha}(f)_{\Delta t}=\frac{1}{\Delta t}\int_{-\Delta t/2}^{\Delta t/2}\frac{1}{T}X_T\left(t,f+\frac{\alpha}{2}\right)^{*}X_T^{*}\left(t, f-\frac{\alpha}{2}\right)dt, \quad (3)$$

where X_T is the local spectral representation as:

$$X_T(u_1, u_2) = \int_{u_1-T/2}^{u_1+T/2} x(t)^{*}e^{-i2u_2ft}dt. \quad (4)$$

Plotting the SCD over the bi-frequency plane (with f and α) allows us to visualize second-order cyclic spectral correlation feature, while provide the extracted input feature data for classification. The PU detection of OFDM signal, is, therefore transformed into the classification by its SCD plane, i.e. image processing.

2) Base learner of Convolutional Neural Networks: To further exploit and learn the SCD plane of OFDM signal, a qualified image processing is necessary. A CNN belongs to deep, feed-forward artificial neural networks, which is applied and specialized in visual imagery analysis. To require minimal preprocessing, multilayer perceptrons were redesigned as a variation version for CNNs. Inspired by biological processes, individual units respond to only a restricted region known as the receptive field, where different units partially overlap so that the entire input field is covered.

Note that, the supervised learning requires a training input database for training phase to determine the parameters within CNN structure, as well as a testing input database for testing phase to perform final performance measurement. The input data is a matrix of SCD feature plane in this research, which can be expressed as:

$$\mathbf{x}_b = \left[x_{i,j}, ..., x_{F,A}\right], i\in[i,F], j\in[i,A], b\in[i,B], \quad (5)$$

where b is the dataset index in the training/testing database, F and A express the total pixel resolution of SCD feature plane, respectively. The core of CNN is the feature extraction realized by convolutional and pooling layers, the output of which, accordingly, is as:

$$\mathbf{o}_l = pool\left(\sigma\left(\mathbf{w}_l^v\cdot\mathbf{o}_{l-1}+\mathbf{b}_l^v\right)\right), v\in[1,c_l], l\in[1,L-1] \quad (6)$$

where assuming $\mathbf{o}_0 = \mathbf{x}_b$, L denotes the layer number of the CNN structure, c, w and b denotes the number of convolutional filters, the weight and the bias parameters in each layer, σ denotes the activation function which, as a non-linear function, can transform the output to a manageable and measurable data range for network training. After the feature extracted by convolutional and pooling layers, it is concatenated into a dense vector in full connection layer for final, high-level extraction. Then, the final output is expressed as:

$$\mathbf{y}_b=\mathbf{w}_f\left(f\left(pool\left(\sigma\left(\mathbf{w}_l^v\cdot\mathbf{o}_{l-1}+\mathbf{b}_l^v\right)\right)\right)\right)+\mathbf{w_f}, b\in[i,B], \quad (7)$$

where f is a linear function for the concatenation process containing $\mathbf{w}_f$ and $\mathbf{w_f}$ denoting the weight and bias parameters of the full connection layers. Finally, the class-predicting layer employs the softmax regression which considers the posterior probability to quantity classification confidence, as $\Psi_\theta(\mathbf{y}_b) = e^{\theta_i\mathbf{y}^i}/\sum_K e^{\theta_i\mathbf{y}^k}$, where y_i denotes the label predictions for each class $i=1,2,...,K$, ($K=2$ in this study) and the optimal parameters θ is obtained by minimizing the cost function in the gradient descent.

The CNN structure is competent to exploit the SCD plane and provide promising sensing results. As a base classifier, the most powerful classifier is not necessarily required, while the limited computation ability should also be taken into account. Therefore, CNNs with simple architecture are employed for this research. The detailed parameter of CNN structure will be given later.

B. Global Decision-Stacking Converted Semi-soft Fusion Center

Since the classification results are the predictions of C_0 or C_1, such outputs of the SU local sensing are reported to the FC and leading to a hard global decision. As mentioned before, a learning tool is required to realize the stacking generalization strategy and integrate the collected local outputs. In order to fully accomplish this task, a DNN is again applied as the middle classifier and give the final PU detection result. Unlike the complex task of the SCD feature analysis, each SU only reports one-bit decision of 0/1 to the FC. Therefore, a full connection deep network is able to achieve the pursuing performance, which will also mitigate computation pressure for the FC.

IV. Result Evaluations

A. Simulation and Training Setup

1) Experimental Settings: Throughout the whole research, a single PU which transmits OFDM signal of 802.11g protocol with 16QAM, and multiple location-fixed SUs which sense the PU activity statuses in AWGN environments while report the sensing results to FC through a perfect error-free channel, are investigated for the CSS. The number of the SUs varies to evaluate the performance generally.

GPU acceleration is applied, with the hardware status of 3.20 GHz Intel Core i7-6900K CPU and Nvidia GeForce GTX 1080 GPU, running MATLAB R2016b on a 64-bit Windows operating system. For each base classifier functioning at SU end where the local sensing is converted by classification, the structure beginning with several sets of consecutive convolutional layers with Relu activation function and average pooling layers, following by several dense full connection layers with Relu activation function, and ending with the softmax layer for providing C_0 or C_1 predictions, is applied. The batch size and learning rate are 100 and 0.01, respectively. Several different structures of SU classifiers are investigated for general evaluation, whose detailed information will be given later.

2) Training Process: The input SCD-plane data for training/testing the SU classifier has the size of 64×64, to reserve its useful characteristics while save the computation and storage load. Then two wholesome general training databases containing such input data and sharing the same structure are needed to be prepared, where the 1st one is prepared for base training the SU classifier and the 2nd one is prepared for middle training the FC classifier. In both wholesome general training databases, the C_0 and C_1 classes both have 100,000 mentioned input datasets for their own learning. However, to balance the performance with the hardware computation ability and weigh each cooperative equally, all the sub-databases for training base classifier, that are constituted by randomly picking out database from the general training database according to the bagging strategy, have the same size of: 10,000 for C_0 and 10,000 for C_1. Notice that, the training database for C_1 contains some slightly polluted PU signal data (SNR = -10 dB) to improve the detection performance for local sensing further.

In the training phase, the realization flow as follows:

1) Each SU classifier is trained by its own sub-databases from the 1st general training database to determine the parameters of CNN structure;
2) The determined base classifiers test the sub-databases from the 2nd general training database, and for every pre-labeled dataset they provide the class predictions;
3) The class predictions combining with the label of every dataset are reported for training and determining of structural parameters of the FC classifier;
4) The entire architecture of ensemble deep learning based CSS is fully established and ready for final detection performance testing.

Notice, each SU is asked to report the local sensing result after every observation period which equals to the duration of OFDM one symbol. Two times of the local reports are accumulated for middle classification in FC. Then, since CSS includes M SUs, one input dataset is in the size of $M \times 2$, and a total number of $M \times 10,000 \times 2$ datasets is used for FC training.

In the final performance testing phase, 5,000 datasets are tested for each SNR value of environment situation.

B. Results Analysis

Fig. 3 shows performance variation curves of p_d (probability of detection) and p_f (probability of false alarm). To better and fairly investigate the performance of the proposed method, the comparisons with geostationary feature detection and energy detection with hard OR-rule FC and $p_f = 0.001$ are exhibited.

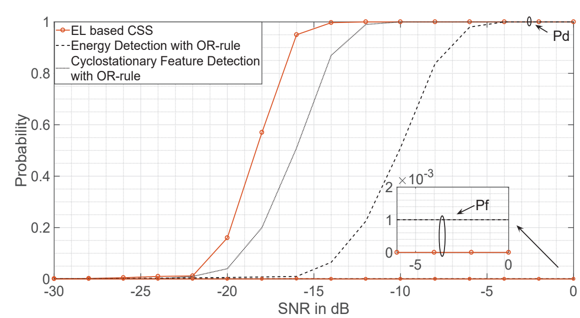


Fig. 3. General performance comparison of p_d and p_f variation curves. The SUs number M is 4.

In term of p_d curves, the energy detection cannot provide a good detection performance, in contrast, the cyclostationary feature detection can give a comparatively better detection capability owing to the efficient cyclic spectral feature which as mentioned before is also applied in the proposed method. However, our proposed ensemble deep learning based CSS shows its superior to both of the conventional detection methods thanks to the competent ensemble deep learning architecture and superior strategies. In term of p_f curves, even though that of the conventional methods are set to be low, the proposed ensemble deep learning based CSS can provide even lower values of 0.00005 from the middle training results. The satisfactory p_fs are supported by the stacking strategy in the FC, where a small dimensionality of input data and a big training database guarantee a nearly perfect fitting training for such high accuracy.

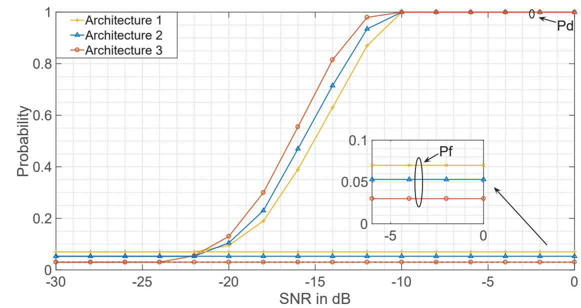
TABLE I
CONDITIONS OF THREE LOCAL BASE CLASSIFIERS

Architecture	Time Complexity	Structure
1	1.27×10^6	1 convolution layer (3*3) 1 pooling layer 1 full connection layer
2	1.74×10^6	1 convolution layer (5*5) 1 pooling layer 1 full connection layer
3	4.29×10^6	2 convolution layer 2 pooling layer 1 full connection layer

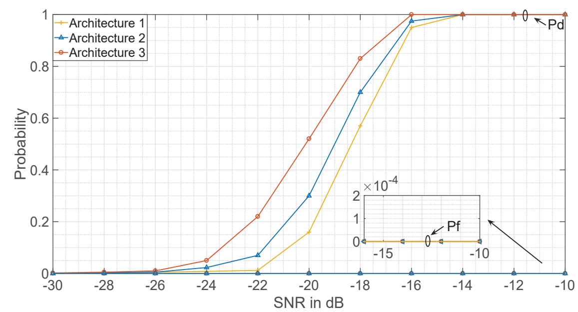
Three CNN architectures for the base SU classifier are discussed to see how local sensing effects the global decision. The detailed conditions are shown in Table I.

As is shown in Fig. 4(b), along the complexity of CNN structure grows, a more fitting training can be reached for each base classifier, which leads to the rise of the local p_d and the drop of the local p_f. In the same trend, the global sensing combining base CNN classifiers of the most advanced structure has the highest p_d. However, when comes to determine the whole architecture, one has to balance the local computation ability with the sensing performance. Moreover, in the global sensing three p_f curves drop to low values regardless of how they behave in the local sensing, which indicates that the stacking strategy can help compensate some deficiencies brought by the base classifier to a certain degree.

Finally, the performance comparison of the proposed method with different SU number is exhibited in Fig. 5. As is shown, when the number of the cooperative SUs increases from 4 to 64, the p_d has an accompanying rise. Such behavior is supported by the increasing dimensionality of the input data for the middle classifier, which provides more built-in data structure feature and allows a better learning. However, the p_d values remain almost still after 64 SUs appear, since the increasing dimensionality could not provide any additional helpful



(a) Local sening performance comparison



(b) Global sensing performance comparison

Fig. 4. Performance comparison of p_d and p_f variation curve for different base classifier architectures. The SUs number M is 4.

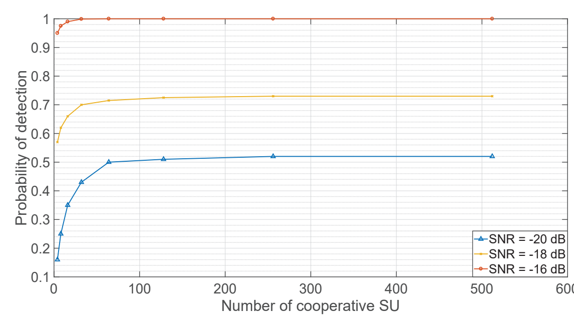


Fig. 5. Performance comparison of p_d and p_f variation curve at SNR -16, -18 and -20 dB for different numbers of SUs where Ms equal 4, 8, 16, 32, 64, 128, 256 and 512 respectively.

data structure feature anymore. Therefore, when a large group of SUs are performed as CSS, a cluster approach that divides the SUs into small groups may help improve the proposed method.

V. CONCLUSIONS

This research proposed a novel ensemble deep learning based CSS method. The EL model is borrowed, where each cooperative SU and the FC are driven by base and combination classifier, then the final sensing result is given by the upgraded strong integrant. The local sensing is converted from signal classification with CNN as deep learning classifier and cyclic spectral correlation feature as efficient input data while the bagging strategy from EL being adopted for the establishment of the training database. The global decision utilizes the stacking strategy which integrates all the reported class predictions through another deep learning classifier. The performance comparison shows the great superiority of the proposed method over conventional methods in both p_d and p_f. In addition, the increasing number and complexity level of SUs help further improve the performance of the proposed method.

Future work will include the investigations for mobile CSS with more complex modal and in harsher transmission environment. Also, other better weak learner type, combination strategy or establishment scheme for training database should also be discussed later to pursue a possibly better detection performance.

REFERENCES

[1] D. Cabric, S. M. Mishra, and R. W. Brodersen, "Implementation issues in spectrum sensing for cognitive radios," in *Signals, systems and computers, 2004. Conference record of the thirty-eighth Asilomar conference*, vol. 1. IEEE, 2004, pp. 772–776.

[2] A. Sahai, R. Tandra, S. M. Mishra, and N. Hoven, "Fundamental design tradeoffs in cognitive radio systems," in *Proceedings of the first international workshop on Technology and policy for accessing spectrum*. ACM, 2006: 2.

[3] I. F. Akyildiz, B. F. Lo, and R. Balakrishnan, "Cooperative spectrum sensing in cognitive radio networks: A survey," *Physical communication*, vol. 4, no. 1, pp. 40–62, 2011.

[4] F. F. Digham, M.-S. Alouini, and M. K. Simon, "On the energy detection of unknown signals over fading channels," *IEEE transactions on communications*, vol. 55, no. 1, pp. 21–24, 2007.

[5] A. Fehske, J. Gaeddert, and J. H. Reed, "A new approach to signal classification using spectral correlation and neural networks," in *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 2005.(DySPAN 2005).* IEEE, 2005, pp. 144–150.

[6] K. M. Thilina, K. W. Choi, N. Saquib, and E. Hossain, "Machine learning techniques for cooperative spectrum sensing in cognitive radio networks," *IEEE Journal on selected areas in communications*, vol. 31, no. 11, pp. 2209–2221, 2013.

[7] D. Hamza, S. Aïssa, and G. Aniba, "Equal gain combining for cooperative spectrum sensing in cognitive radio networks," *IEEE Transactions on wireless communications*, vol. 13, no. 8, pp. 4334–4345, 2014.

[8] W. Lee, M. Kim, D.-H. Cho, and R. Schober, "Deep sensing: Cooperative spectrum sensing based on convolutional neural networks," *arXiv preprint arXiv:1705.08164*, 2017.

[9] X.-L. Zhang and D. Wang, "A deep ensemble learning method for monaural speech separation," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 24, no. 5, pp. 967–977, 2016.

[10] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 4, pp. 463–484, 2012.

[11] H.-S. Chen, W. Gao, and D. G. Daut, "Spectrum sensing for ofdm systems employing pilot tones," *IEEE Transactions on Wireless Communications*, vol. 8, no. 12, 2009.

[12] G. Wang, J. Hao, J. Ma, and H. Jiang, "A comparative assessment of ensemble learning for credit scoring," *Expert systems with applications*, vol. 38, no. 1, pp. 223–230, 2011.

[13] H. Liu, X. Zhu, and T. Fujii, "Primary user detection in cognitive radio using spectral-correlation features and stacked denoising autoencoder," in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC).* IEEE, 2017, pp. 1–5.

[14] M. Adrat, J. Leduc, S. Couturier, M. Antweiler, and H. Elders-Boll, "2nd order cyclostationarity of ofdm signals: Impact of pilot tones and cyclic prefix," in *IEEE International Conference on Communications, 2009 (ICC'09).* IEEE, 2009, pp. 1–5.