Convolutional LSTM-based Long-Term Spectrum Prediction for Dynamic Spectrum Access

Bethelhem S. Shawel
SECE
AAiT, AAU
Addis Ababa, Ethiopia
bethelhem.seifu@aait.edu.et

Dereje H. Woldegebreal
SECE
AAiT, AAU
Addis Ababa, Ethiopia
dereje.hailemariam@aait.edu.et

Sofie Pollin
Department ESAT
KU Leuven University
Leuven, Belgium
sofie.pollin@esat.kuleuven.be

***Abstract*—The concept of Dynamic Spectrum Access (DSA) with Cognitive Radio (CR) as a key enabler is considered as a promising solution to alleviate the inefficient use of the radio spectrum. Relying on the presumed knowledge of the spectrum occupancy from sensing, geo-location databases or prediction, DSA allows opportunistic users to share spectrum bands in a non-interfering manner when the bands are not in use by their respective incumbent owners. Several literatures have presented prediction algorithms in order to get meaningful data about future spectrum usage; however, most of them only exploit the spectrum data in time, space and/or frequency dimension(s) to provide a short term, i.e., single next step, prediction. In this work, we propose a novel approach with Convolutional Long Short-Term Memory (ConvLSTM) Deep Learning Neural Network for a long-term temporal prediction that is trained to learn joint spatial-spectral-temporal dependencies observed in spectrum usage. Real environment measurement data from Electrosense are used to evaluate the prediction accuracy of the proposed network for increasing future time steps and different spectrum channels. Prediction result for the next 180 minutes for UHF bands of 450-520 MHz is presented for a 4 km^2 area in Spain indicating the prominent and stable prediction performance of ConvLSTM network.**

***Index Terms*—ConvLSTM Neural Network, Deep Learning Network, Long term Prediction, Spectrum Prediction, Dynamic Spectrum Access.**

I. Introduction

The recent technological advancements of wireless communication and rapid growth of bandwidth-hungry innovative services are demanding availability of broader spectral band and hence, changing the conventional course of spectrum access. In order to accommodate the growing demand and optimize the use of spectrum, the concept of dynamic spectrum access (DSA) has been adopted as one spectrum sharing scheme. DSA is an optimal solution to improve spectrum utilization in time, frequency and location by allowing opportunistic users to share spectrum bands in a non-interfering manner when the bands are not in use by their respective incumbent owners. This spectrum sharing scheme requires prior knowledge of spectrum usage pattern, that can be obtained through geo-location databases or spectrum sensing, and can further be optimized by spectrum prediction techniques.

Spectrum prediction is a method of sensing the spectrum bands to infer relevant information about future spectrum usage. It effectively exploits hidden usage patterns available in historical spectrum data. Generally, it focuses on attaining future parameters such as (i) channel status, i.e., prediction of the spectrum status as *"idle"* or *"busy"*, (ii) duty cycle, i.e., prediction of average fraction of time the spectrum channel is occupied, (iii) signal power level, i.e., prediction of the power level on a specific channel to define the quality. Space, frequency and time dimensions of spectrum usage pattern are random in nature depending on factors such as the radio propagating environment, transmitter power, users mobility and others. Predictions based on measured data from different spectrum sensors (SS) can be used to understand and exploit such multi-dimensional dependencies of usage pattern. Even though in most cases the prediction is done to provide single future step prediction, DSA requires the availability of accurate long-term (multiple future steps) prediction. It is undeniable that long-term prediction suffers from various additional complications, such as accumulation of errors and reduced accuracy [1], [2]. However, it is possible to explore different prediction techniques in order to provide multiple future predictions with a tolerable error.

Extensive research has been carried out on various prediction techniques and applications. Several of those explore models based on Linear Regression Analysis, Bayesian Inference or Markov Analysis. However, these models suffer from limitations to capture nonlinearities due to multidimensional nature of spectrum data, increased complexity in discrete state representations and scalability issues which will hinder the accuracy of channel quality or states prediction [3], [4]. In recent years, as Deep Learning (DL) is gaining popularity for its improved likelihood of estimations with self-learning ability on large data, more and more spectrum learning algorithms are proposed to provide more accurate information. The learning algorithms are applicable for both classification and regression problems. In dealing with classification problems where feature categorization is the aim, the most popular DL network is Convolutional Neural Network (CNN) [5]. Whereas in the case of regressional problems, where learning the sequential dependency is the goal, recurrent neural network (RNN) based models such as Long Short-Term Memory (LSTM) is the most practical to use. LSTM network has achieved great success in many sequence prediction applications with its ability to cap-

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

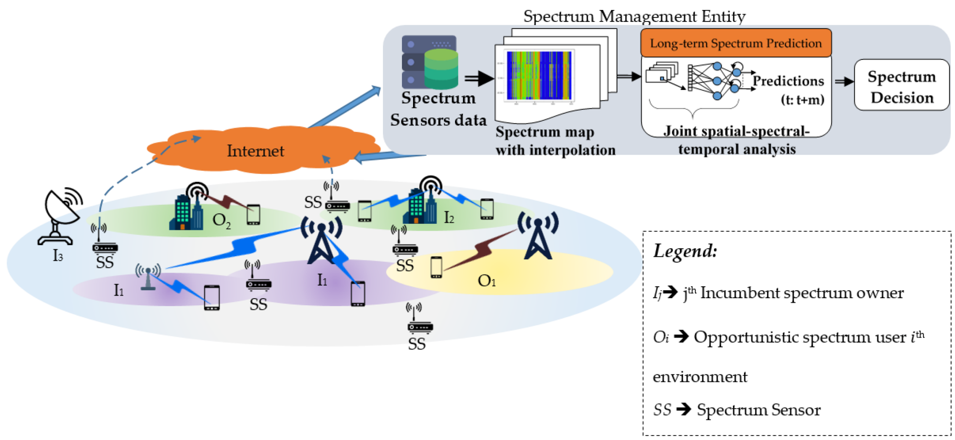

Fig. 1. System model for prediction enabled Spectrum Management Entity.

ture long-range temporal dependences of sequences. In [6] and [7], the use of LSTM for spectrum prediction was presented to provide the channel occupancy prediction. In our previous work in [8], we explored the accuracy of long-term signal level predictions on different spectrum bands by analyzing the spectral-temporal correlations with LSTM based DL networks. Similarly in [9], optimized LSTM model with Taguchi Method is presented to predict signal level in different channels and evaluate the impact of hyper-parameter and network depth size on prediction accuracy. An approach to provide a long-term prediction was proposed in [10] from an image inference perspective. By converting spectrum prediction into 3-order Tensor completion problem, they were able to predict one day ahead spectrum state with reasonable error margin. Even with such broad work in spectrum prediction, finding those that dealt with long-term spectrum prediction based on joint spatial-spectral-temporal dependencies is limited.

Considering the aforementioned discussions in spectrum prediction, this work will address the long-term prediction problem by learning the joint spatial-spectral-temporal dependencies observed in spectrum data. Taking spectrum measurement data in the form of received signal power in dBm from sparsely distributed SS, spatially interpolated spectrum map using Inverse Distance Weighting (IDW) method is created similar to [11] for a particular region in Spain, within 6 Km radius from the center of Madrid. As a result, the spectrum data becomes a function of frequency, time and space. In order to provide the long-term predictions of spectrum data, we proposed to apply the Convolutional LSTM (ConvLSTM) based DL network with a sequence to sequence architecture. In summary, the main contributions of this work are:

- Formulate a new long-term spectrum prediction scheme with ConvLSTM-based DL network to capture the joint spatial-spectral features and temporal dependencies of the spectrum data.
- Evaluate the proposed network with real environment spectrum data and assess the long-term prediction accuracy in different spectrum bands.

The remainder of this paper is organized as follows. After the introductory Section I detailing spectrum prediction, Section II is dedicated to preliminary discussions on system model and constructing interpolated spectrum map. After that, details of long-term spectrum prediction scheme is presented in Section III. Section IV provides dataset description and experimental results, followed by concluding remarks in Section V.

II. System Model

The intent of the long-term spectrum prediction problem defined in this work is to use the previously observed spectrum measurement data from multiple locations over wide spectrum band to forecast fixed length and multiple-slot future spectrum data, in terms of received power level, in a particular region. In a practical scenario, as illustrated in Fig. 1, for a heterogeneous network environment with I multiple incumbent spectrum owners and O opportunistic network users, a centrally controlled database entity could be responsible for identifying available spectrum channels based on long term predictions. Suppose, K sparsely distributed SS are continously monitoring the power level from multiple frequency bands, F, and sending the received signal power level (P, that varies over time) to the management entity. The spatial region of interest is partitioned into X grids in one (or rows) side and Y grids on the other side (or columns). Then the interpolated spectrum map is created for a grid indicated as (x, y), where $x \in 1, 2, ...X$ and $y \in 1, 2, ...Y$ by fusing/combining measured data from the five SS. There are several techniques such as non-geostatistical and geostatistical (univariate/multivariate) interpolating methods that can be considered depending on sample size and the nature of data sampling [12]. In this work, due to the unsystematic nature of SS placement, we have chosen to use IDW method to linearly combine the measured values from SS with a weighting coefficient ω_k that is an inverse function of the distance from a SS the grid of interest. For a given grid in space (x, y), the interpolated spectrum measurement value $p^{(x,y)}_{(f,t)}$ at a particular time, t and channel, f

is evaluated based on measurement data from those K nearby SS as [12]:

$$p_{(f,t)}^{(x,y)} = \sum_{k=1}^{K} \omega_k p_{(f,t)}^{k} \quad (1)$$

where $p_{(f,t)}^{k}$ is measured power in dBm by the k^{th} SS and ω_k refers to the weighting coefficient, assigned to each SS, expressed as:

$$\omega_k = \frac{d_k^{-n}}{\sum_{j=1}^{K} \frac{1}{d_j^n}} \quad (2)$$

where n is a path-loss exponent (between 2 to 6) and d_k are the Euclidean distances from the grid of interest to the k^{th} SS.

The interpolated spectrum map, thus, can be represented by a tensor χ with $X \times Y \times F \times t$ representing the 4 dimensionalities. Furthermore, if we consider a grid referencing approach, where a particular location l representing the grid in space (x, y) is defined for square unit area, then it is possible to simplify the above tensor representation with only $l \times F \times t$ dimensions. For sequentially obtained spectrum data $\chi_1, \chi_2, \chi_3, \chi_4, ...$, the long-term spectrum prediction problem evaluated with a DL network is defined as

$$\langle \chi_{t-n}, ...\chi_{t-2}, \chi_{t-1} \rangle \longmapsto \langle \chi_t, \chi_{t+1}, ..., \chi_{t+m} \rangle$$

where n and m represents the historical observations and the future instants in time, respectively.

III. Long-term Spectrum Prediction Network

A. Convolutional LSTM Networks

The impressive success of DL prediction problems that involve processing time-sequence information have been achieved with RNN architectures that are *"deep in time"*. The conventional RNN has the capability to model the dynamic temporal behavior of sequential inputs by having a recurrent hidden state whose activation at each time step depends on that of the previous time. However, it faces a significant limitation with the ability to back-propagate an error through a long-range temporal interval, also known as *vanishing gradient problem*. In order to solve such difficulties and effectively capture the long-range temporal dependencies, LSTM is found to be more efficient [9]. Incorporating memory units that explicitly allow the network to learn when to *"forget"* previous hidden states, and when to update hidden states given new information have improved gradient training for LSTM Networks.

The joint multi-dimensional dependencies in space, frequency and time that are observed in spectrum data can be considered as a complex higher-order learning problem for LSTM networks. Although they are already proven at handling spectral-temporal dependencies for spectrum prediction, maintaining structural locality and solving such extended problems will be challenging [13]. Thus, in order to acquire a joint spatial-spectral feature and also effectively analyzes temporal dependency for spectrum data, ***ConvLSTM*** that is discussed in [14], is considered in this work.

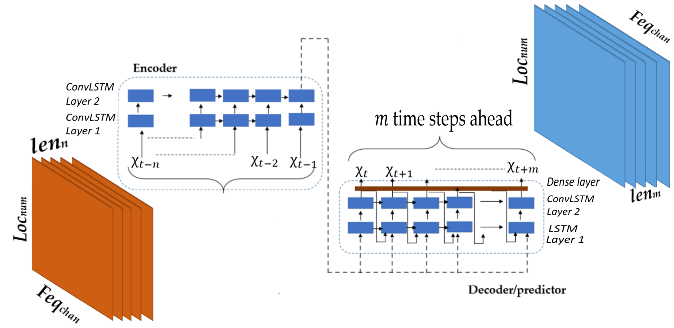

Fig. 2. The learning and predicting ConvLSTM-based DL network with sequence-to-sequence architecture proposed for Long-term spectrum prediction.

ConvLSTM combines CNN and LSTM taking the best of the two worlds to learn the multi-dimensional dependencies in sequential data [15]. CNN is widely known feed-forward network with its capability to extract features from a multi-dimensional input data through a convolution operation of the input with a filter (kernel). Whether it 2D layout in the case of image or 3D structure in video frames, its ability to automatically discover relevant contextual and spatial features with reduced number of parameters makes it more relevant [15]. In ConvLSTM architecture, the inputs to the network are priorly transformed or convolved with feature extraction parameters (*weights*) to produce a fixed-length matrix representation. Key equations that define ConvLSTM for a given input tensor χ_t are given in (3):

$$\begin{aligned}
i_t &= \sigma(W_{\chi_i} * \chi_t + W_{Hi} * H_{t-1} + W_{Ci} \odot C_{t-1} + b_i) \\
f_t &= \sigma(W_{\chi_f} * \chi_t + W_{Hf} * H_{t-1} + W_{Cf} \odot C_{t-1} + b_f) \\
o_t &= \sigma(W_{\chi_o} * \chi_t + W_{Ho} * H_{t-1} + W_{Co} \odot C_{t-1} + b_o) \\
C_t &= f_t \odot C_{t-1} + i_t \odot g(W_{pc} * \chi_t + W_{Hc} * H_{t-1} + b_c) \\
H_t &= o_t \odot \hbar(C_t)
\end{aligned} \quad (3)$$

where $\odot$ and $*$ denotes the Hadamard product and the convolution operator, respectively. i_t, f_t, o_t, C_t and H_{t-1} are matrices all representing the gating units (input, output, forget), cell unit and the hidden state, respectively. The weights, W_i, W_f, W_c, W_o, corresponds to feature extracting convolutional filter matrices, which are multidimensional arrays in nature, used to transform in the state-to-state and input-to-state convolutional transitions.

The selection of input and output activation functions, ($\hbar$ and g), depends on the learning problem at hand. The prediction of spectrum data from previous observations is real-valued problem, so rather than letting those functions be $tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$, *hyperbolic tangent* to squash the values within the range of [-1,1], it is more advisable to use rectifier - *ReLU*, *R (z) = max(0,z)*, as activation function. Whereas the gate activation functions usually takes the $\sigma(z) = (1+e^{-x})^{-1}$, sigmoid non-linearity to squash the values within the range of [0,1]. w_p and b are Weights and biases, respectively, learned by the network by minimizing the loss between the ConvLSTM outputs and the actual samples during network training.

Extending the discussion to multi-step prediction problem for spectrum data instantiating a one class of **sequential learning task**: *multi input-multi output*, that can be implemented with an *encoder-decoder* network arrangement. As illustrated in Fig. 2, the *encoder* maps input sequences to a fixed-length matrix and the *decoder (predictor)* unfolds the matrix and generate a sequence of arbitrary length data. It comprises two 2D-ConvLSTM hidden layers at the encoder, and LSTM hidden layer that is used to capture memory and hidden states from the encoder output, a 2D-ConvLSTM hidden layer and fully connected (dense) layer at the decoder. The bottom of ConvLSTM layer at the encoder receives a len_w sequence of loc_{num} X $ferq_{chan}$ dataset where len_w represents the number historical observations, and loc_{num} and $ferq_{chan}$ represent the number of locations considered and the multiple spectrum channels in a band, respectively, to be filtered by a kernel with a stride of 1 sample. Similarly at the decoder side, the dimensionality of predicted output is given as a len_m sequence of loc_{num} X $ferq_{chan}$ dataset where len_m represents the long-term prediction length. With a simplified representation of space grids to location mapping, we were able to make use of 2D kernels with a very small size of 3×3 at the bottom layer of encoder, and 1×1 at the top layer of both encoder and decoder.

B. Network Training

The network training is implemented in R with the help of TensorFlow framework. We selected the Nesterov accelerated adaptive moment estimation (NADAM) algorithm as the optimizer with tuned hyper-parameters of $\beta_1 = 0 : 9$, $\beta_2 = 0 : 999$, $\epsilon = 10^{-8}$, $\lambda = 0.004$ and $\ell_r = 0.0002$ and training loss measured in Mean Square Error (MSE). Additional hyper-parameters to consider are the number of convolutional filters where the minimum selection criteria of filter size is based on the multiple spectral-spatial features to be extracted.

The dataset is divided in 80/10/10 ratio for network training, validating and testing, respectively. In order to reduce the effect of over-fitting, we considered unit dropouts at both encoder and decoder with $p = 0.3$ probability during the pre-training phase and batch normalization as regularization techniques. Furthermore, in order to account the faulty measurements that might have an impact on the interpolated spectrum data, the network is trained with an added Gaussian noise with variance 0.2.

IV. Experimental Evaluation

A. Data Description

The performance of the proposed network is evaluated on aggregated spectrum measurement data from Electrosense[1] open API for UHF bands of 450-520 MHz. Measurement data from five SS located in Spain, with in 6 Km radius from the center of Madrid is considered. The resolution bandwidth of each individual spectrum channel is 200 kHz with 3 minutes resolution time, which corresponds to 840,000 measurement data per sensor obtained for the duration of 5 days.

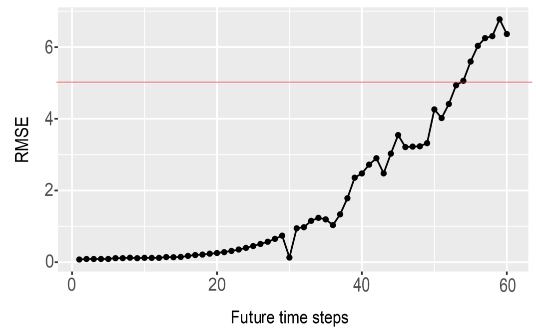

Fig. 3. Predicting RMSE for Spectrum data for varying predicting time step

B. Experimental Results and Discussions

The spectrum prediction performance is evaluated using Root Mean Square Error:

$$RMSE = \sqrt{\frac{1}{T}\sum_{j=1}^{T}(y_j - y\prime_j)^2} \tag{4}$$

where T is the number of samples considered, and $y\prime_j$ and y_j are the predicted and actual values, respectively.

RMSE values w.r.t. future time steps and different spectrum channel are presented to evaluate prediction accuracy of the network. The spectrum data prediction in terms of signal power level is done for the next 150 minutes for 70 MHz bandwidth over a selected 4 km^2 area[2]. The predicted output is based on 120-time steps (or 6 hrs. of past observations) to predict an output with dimensions of 50 time steps, 1600 location points and 350 spectrum channel. As shown in Fig. 3, an increase in future time steps increases the accumulated error that is propagated from the previous predictions resulting in RMSE performance to consistently increase. However, even with that, we were able to manage values below 5.012, which is the combined sample standard deviation we set as a benchmark. The spectrum prediction for a particular location $l = 268$ (UTM coordinates of $X = 592875.8$ m and $Y = 638032.6$ m) is illustrated for 450-520 MHz in Fig. 4. It shows that the prediction is more accurate in terms of capturing spectral dependencies and long term temporal spectrum usage patterns than changes that occur for a short time interval.

RMSE values averaged for all location points within the 4 km^2 area is presented for different spectrum channels in Fig. 5. As shown, 95% of RMSE values are deviating within ± 0.07 from the mean indicating the stable prediction performance of the proposed network.

[1] available at https://electrosense.org/

[2] $(2km$ X $2km)$ square with starting point at [40.33839, -3.774233]

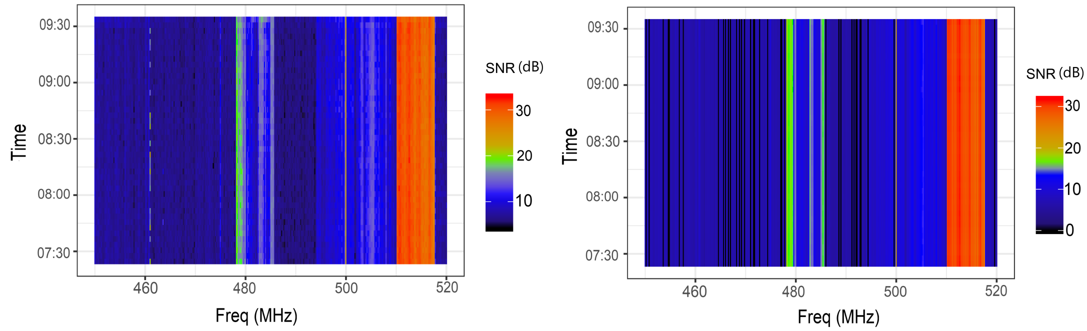

Fig. 4. Comparison between the measured spectrum data (left) and the corresponding predicted data (right) for time slot of 3 hrs. duration.

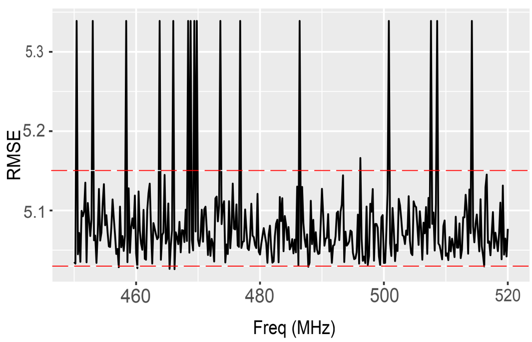

Fig. 5. RMSE performance in different spectrum channels averaged for a given $4km^2$ area

V. Conclusions and Future Works

In this work, we propose ConvLSTM-based DL neural network with sequence-to-sequence architecture for long-term spectrum prediction problem by capture the joint spatial-spectral-temporal dependencies observed in spectrum data. To evaluate the prediction accuracy of the proposed network in terms of RMSE, measurement data from five SS located in Spain, with in 6 Km radius from the center of Madrid is used. The long-term prediction as SNR in dB is done for the next 180 minutes for UHF bands of 450-520 MHz over the selected 4 km^2 area. As future time steps increases, the accumulated error propagated from the previous predictions increase resulting in RMSE value to consistently rise. Even with that, RMSE values obtained are less than 5.012, which is the combined sample standard deviation we set as benchmark. Similarly,the 95% average RMSE value for different spectrum channels deviated within ± 0.07 from the mean indicating the stable prediction performance of the proposed network. As for future work, the degree of predictability with respect to the increased length of future prediction and incomplete measured data is an area worth investigating.

References

[1] X. Wu, Y. Wang, J. Mao, Z. Du, and C. Li, “Multi-step prediction of time series with random missing data,” *Applied Mathematical Modelling*, vol. 38, no. 14, pp. 3512–3522, doi: 10.1016/j.apm.2013.11.029.

[2] J. Zhang, G. Ding, Y. Xu, and F. Song, “On the usefulness of spectrum prediction for dynamic spectrum access,” in *International Conference on Wireless Communications Signal Processing*, Oct 2016, pp. 1–4, doi: 10.1109/WCSP.2016.7752555.

[3] H. Eltom, S. Kandeepan, R. J. Evans, Y. C. Liang, and B. Ristic, “Statistical spectrum occupancy prediction for dynamic spectrum access: a classification,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2018, no. 1, p. 29, doi: 10.1186/s13638-017-1019-8.

[4] J. Wu and Y. Li, “A survey of spectrum prediction methods in cognitive radio networks,” vol. 1834, 04 2017, p. 020018, doi: 10.1063/1.4981557.

[5] M. Kulin, T. Kazaz, I. Moerman, and E. D. Poorter, “End-to-end learning from spectrum data: A deep learning approach for wireless signal identification in spectrum monitoring applications,” *IEEE Access*, vol. 6, pp. 18 484–18 501, 2018, doi: 10.1109/ACCESS.2018.2818794.

[6] L. Yu, J. Chen, and G. Ding, “Spectrum prediction via long short term memory,” in *IEEE International Conference on Computer and Communications*, Dec 2017, pp. 643–647, doi: 10.1109/Comp-Comm.2017.8322623.

[7] L. Yu, Q. Wang, Y. Guo, and P. Li, “Spectrum availability prediction in cognitive aerospace communications: A deep learning perspective,” in *2017 Cognitive Communications for Aerospace Applications Workshop (CCAA)*, June 2017, pp. 1–4, doi:10.1109/CCAAW.2017.8001877.

[8] B. S. Shawel, D. H. Woldegebreal, and S. Pollin, “Deep-learning based cooperative spectrum prediction for cognitive networks,” *International Conference on Information and Communication Technology Convergence*, pp. 133–137, 2018, doi: 10.1109/ICTC.2018.8539570.

[9] L. Yu, J. Chen, G. Ding, Y. Tu, J. Yang, and J. Sun, “Spectrum prediction based on taguchi method in deep learning with long short-term memory,” *IEEE Access*, vol. 6, pp. 45 923–45 933, 2018.

[10] J. Sun, J. Wang, G. Ding, L. Shen, J. Yang, Q. Wu, and L. Yu, “Long-term spectrum state prediction: An image inference perspective,” *IEEE Access*, vol. 6, pp. 43 489–43 498, 2018, doi: 10.1109/ACCESS.2018.2861798.

[11] A. Agarwal and R. Gangopadhyay, “Predictive spectrum occupancy probability-based spatio-temporal dynamic channel allocation map for future cognitive wireless networks,” *Trans. Emerging Telecommunications Technologies*, vol. 29, no. 8, 2018, doi: 10.1002/ett.3442.

[12] J. Li and A. D. Heap, “A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors,” *Ecological Informatics*, vol. 6, no. 3, pp. 228–241, doi: 10.1016/j.ecoinf.2010.12.003.

[13] Y. Zhang, W. Chan, and N. Jaitly, “Very deep convolutional networks for end-to-end speech recognition,” in *IEEE International Conference on Acoustics, Speech and Signal Processing*, March 2017, pp. 4845–4849, doi: 10.1109/ICASSP.2017.7953077.

[14] L. Mou, L. Bruzzone, and X. Zhu, “Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. PP, 03 2018, doi: 10.1109/TGRS.2018.2863224.

[15] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-k. Wong, and W.-c. Woo, “Convolutional lstm network: A machine learning approach for precipitation nowcasting,” in *International Conference on Neural Information Processing Systems*, ser. NIPS’15. Cambridge, MA, USA: MIT Press, 2015, pp. 802–810.