LSTM-Based Spectrum Occupancy Prediction Performance with Thin Plate Spline Model

Dinushika Chathurangani Alahakoon*, Sithamparanathan Kandeepan*, Xinghuo Yu*, and Gianmarco Baldini†
*School of Engineering, RMIT University, Melbourne, VIC 3001, Australia
†European Commission, Joint Research Centre, 21027 Ispra, Italy
Email: s3959584@student.rmit.edu.au, kandeepan@ieee.org, xinghuo.yu@rmit.edu.au, Gianmarco.baldini@ec.europa.eu

***Abstract*—Prediction-based Dynamic Spectrum Access (DSA) in Cognitive Radio (CR) is a promising approach for efficient frequency spectrum utilisation for fast and reliable next generation of wireless communication. This led us to meticulously examine the relation of spectrum occupancy characteristics on prediction performance to accurately predict spectral opportunities in a dynamic radio environment. Our research involves utilising a Long-Short-Term Memory (LSTM)-based deep learning model to predict binary spectrum occupancy, which is characterised by the discrete-time Markov process. The LSTM prediction error is analysed and modelled using a Thin-Plate Spline (TPS) model for arbitrary binary Markovian model parameters. The proposed TPS model for prediction error is presented here along with the model verification results, which show a clear correlation between prediction performance and channel occupancy patterns and gives superior accuracy for conveniently computing the prediction error under the proposed settings.**

***Index Terms*—Error modeling, Markovian spectrum occupancy, prediction error, spectrum occupancy prediction, TPS**

I. Introduction

Pursuing larger capacity and higher reliability have always been critical objectives as we progress through the evolution of wireless systems. As wireless networks approach the sixth generation (6G) through the fifth generation (5G), the world of communication is becoming more concerned with artificial intelligence (AI) solutions, and is about to undergo significant improvements with AI applications [1]. The dynamic wireless environment is poised to see crucial involvement from several important ideas, such as Cognitive Radio (CR), Dynamic Spectrum Access (DSA), and spectrum utilization. CR and DSA, in particular, represents a paradigm shift in wireless communication systems, with its primary goal being to enhance the utilization of the radio frequency spectrum [2].

Prediction-based DSA has being deployed to mitigate the shortfalls of fundamental cognitive radio spectrum management and improve spectrum efficiency, reduce energy consumption, and reduce the latency of dynamic spectrum access [2],[3],[4]. Spectrum prediction is employed to predict the future status of the channel in advance and allows secondary users to have seamless handovers in cognitive radio networks. There are different models such as auto-regressive, Markov, Bayesian network, and Deep Learning (DL) that can be used for spectrum prediction in cognitive radio [2]. These models can capture the temporal, spectral and spatial dynamics of spectrum occupancy and provide probabilistic or deterministic predictions of future spectrum availability. DL is a popular AI tool due to the capacity to learn the complex and nonlinear features of the spectrum environment from the observed data with the architecture based on Artificial Neural Networks (ANN) [2]. Researchers have shown keen interest in the computational prowess and accuracy of DL models for spectrum prediction [3] and their ability to effectively adapt to complex environments [5].

Deep Neural Network [5], Convolutional Neural Network (CNN) [6], Long-Short-Term Memory (LSTM)[5] and Recurrent Neural Network (RNN) [7], are popular DL approaches used in spectrum prediction. Neural networks consist of an input layer, an output layer, and intermediate hidden layers comprised of artificial neurons. RNN is well known for addressing time series with dependencies thanks to its rapid convergence and ability to manage complex nonlinear data using the feedback architecture of recurrent cell [8]. LSTM is an optimized version of the RNN network with the ability to handle time series with dependencies due to its RNN features and improved capabilities, including ease of training, memory cells, and long-term dependencies [9].

The accuracy of spectrum occupancy prediction results relies, in part, on the predictive model used. A model can be selected based on the performance compared to other prediction models [1],[5], or the structure and hyper-parameter selections of the prediction models for specific application [10],[11],[12]. The accuracy of predictions is significantly influenced by the input spectrum data, making it crucial to assess how well the prediction model aligns with the intended Cognitive Radio (CR) environment [5],[1]. Predicting spectrum occupancy has been investigated recently for various occupancy models using different prediction techniques [9],[12],[13]. It is desirable to theoretically quantify the prediction performance for predicting the spectrum occupancy using machine learning methods for various parametric settings, and this will allow the wireless system engineers to schedule wireless transmissions better in a spectrally crowded environment using dynamic spectrum scheduling. Especially in a dynamic environment with time-varying parameters, it is important to have a theoretical/computational model to estimate the performance of spectrum prediction.

In this letter, we model the error performance for predicting the binary channel states of spectrum occupancy using LSTM-based classification, considering a discrete two-state Markov based spectrum occupancy model for arbitrary state

transition probabilities. Our previous work on prediction error modelling was limited to the particular case where the transition probabilities were the same for the two states [14], and in this paper, we present it for any arbitrary values of state-transition probabilities. Non-parametric regression is preferred over parametric methods for error modeling due to the complex relationship between transition probabilities and error performance [15],[16]. Additionally, it avoids forcing the data into predefined base functions or models. After assessing parametric and non-parametric models, Thin-Plate Spline (TPS) was found to be the best fit. The TPS regression is a well established non-parametric regression method that provides a straightforward and effective regression [15]. Consequently, the performance model presented here is based on a TPS interpolation that offers a closed-form series expression and model coefficients. The model was further tested with additional data for validation. To the authors' knowledge, such a prediction error performance model has not been presented in the literature to date.

In the remainder of the paper, we present the system model for prediction in Section-II, the Thin-Plate Spline modeling in Section-III together with the verification of the proposed model in Section-IV. Finally, the Section V concludes the paper.

II. SYSTEM MODEL

The LSTM-based prediction model is presented in this section. The spectrum occupancy is modelled using a discrete two-state Markov process and the binary spectrum occupancy prediction is performed using an LSTM-based classification model that predicts the channel status with a single step ahead.

A. Spectrum Occupancy Model

We consider a single discrete time channel with binary occupancy level $X(n)$, where the binary states of $X(n)$ are modelled by a stationary discrete time Markov process [17], as defined below,

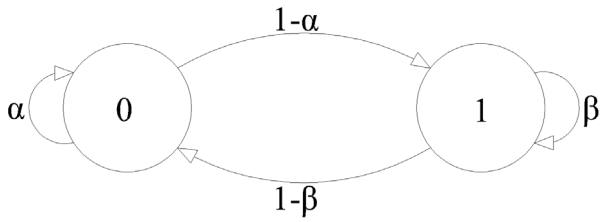

Fig. 1. State transition diagram.

$$X(n) = \begin{cases} 0 & ; \text{ Idle} \\ 1 & ; \text{ Busy} \end{cases} \tag{1}$$

The initial state is Idle, and the transition probability matrix P_T is defined as,

$$P_T = \begin{bmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{bmatrix} \tag{2}$$

where,

$$\alpha = Pr[X(n+1) = 0 | X(n) = 0] \tag{3}$$

$$\beta = Pr[X(n+1) = 1 | X(n) = 1] \tag{4}$$

The steady state probabilities are,

$$P(X(n) = 0) = \frac{1-\beta}{2-(\alpha+\beta)} \tag{5}$$

$$P(X(n) = 1) = \frac{1-\alpha}{2-(\alpha+\beta)} \tag{6}$$

B. LSTM Model for Temporal Spectrum Prediction

The LSTM-based deep learning model [14] considered for classification and prediction in this work consists of five layers, including the sequence input layer, LSTM layer, fully connected layer, softmax activation layer and classification layer, as depicted in Fig.2. The data series for channel occupancy, which was acquired through the model described in Section II-A, was utilized to train the proposed LSTM model.

The Markov input data sequence, $X(n)$ with N time steps is reshaped into a $(N-1) \times 2$ matrix using a sliding window approach to prepare data for the sequence input layer with an input size of two. Every row of the reshaped matrix is an input containing the current and previous spectrum occupancy states, $[X(n-1),\ X(n)]$. The sequence input layer converts the input sequence into a compatible vector for the next layer.

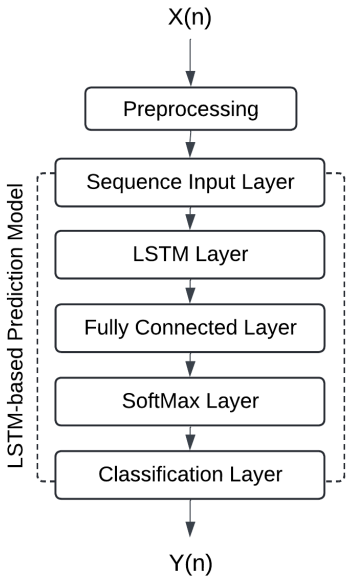

Fig. 2. LSTM-based prediction model.

The LSTM layer, composed of gated units of three sigmoid gates and a tanh layer, performs the spectrum prediction's central computational part. The number of hidden units or the gated units selection determines the capacity of the LSTM layer to extract and process more features of the input data. Similarly, more hidden units may cause overfitting and impact the computational speed. One hidden unit has been configured

in this particular model. The output of the LSTM layer is forwarded to the fully connected layer.

The result from the fully connected layer is then passed through the softmax layer, which converts raw scores into meaningful probabilities, denoted as q and $1-q$, for the labels "idle" and "busy," respectively. The array of softmax output is,

$$V(n) = [q, 1-q]; 0 \leq q, 1-q \leq 1 \tag{7}$$

During training and validation, the classification layer calculates the cross-entropy loss between the actual label and predicted probabilities resulting from the softmax layer. During the prediction, the prediction output $Y(n)$, which is the predicted channel state for $X(n+1)$, is classified by the classification layer according to the (8),

$$Y(n) = \begin{cases} 0 & ; q > 1-q \\ 1 & ; q < 1-q \end{cases} \tag{8}$$

C. Probability of Prediction Error

The prediction error occurs when the predicted state does not match with the corresponding actual state of the channel, and the probability of prediction error P_e is given by $Pr[Y(n) \neq X(n+1)]$, which can be written as,

$$\begin{aligned} P_e = {} & Pr[Y(n)=0|X(n+1)=1]Pr[X(n+1)=1] \\ & + Pr[Y(n)=1|X(n+1)=0]Pr[X(n+1)=0] \end{aligned} \tag{9}$$

Based on the above expression, P_e becomes a surface curve when observed with respect to the transition probabilities α and β. In this work we model P_e as a function of α and β using a TPS based surface fitting model, as described in the next section.

III. Modeling the Probability of Error with TPS Regression Method

The Thin Plate Spline regression is a known mathematical technique to interpolate and smooth grid data to fit a surface [18], This non-parametric regression method allows to obtain a straight forward mathematical representation of the dataset and assigns weights to each data point [19]. The LSTM prediction error probability P_e becomes a surface curve when observed as a function of the two transition probabilities α and β. Upon obtaining sample data points for P_e from computer simulations across arbitrary set of α and β values, Multivariate TPS regression can be utilized for surface fitting.

The basic concept of TPS is as follows. Residual sum of squared (RSS) (10) represent the goodness of fit of TPS function.

$$RSS = \sum_{i=1}^{M} \left(P_e(c_i)) - \overline{P}_e(c_i)\right)^2 \tag{10}$$

M is the number of modeling data points and i=1,2,..M. c_i denotes the (α, β) coordinates of i^{th} data point. $\overline{P}_e(c_i)$ is the model expression representing prediction error obtained from the regression analysis and $P_e(c_i)$ is the actual prediction error at the i^{th} data point.

Smoothness of the curve fitting is calculated by the Integrated Squared Second Derivative (ISSD) of the smoothing spline (11).

$$\begin{aligned} ISSD = \lambda \iint_{\mathbb{R}^2} & \left(\frac{\partial^2 \overline{P}_e(c)}{\partial \alpha^2}\right)^2 \\ & + 2\left(\frac{\partial^2 \overline{P}_e(c)}{\partial \alpha \partial \beta}\right)^2 + \left(\frac{\partial^2 \overline{P}_e(c)}{\partial \beta^2}\right)^2 d\alpha d\beta. \end{aligned} \tag{11}$$

λ is the Smoothing parameter and, $\lambda > 0$. Cross-validation score quantifies smoothing parameter in TPS model[18]

TPS estimates a smoothing spline function $\overline{P}_e(c)$ that efficiently minimizes the Penalized Sum of Square(PSS) in (12) by not precisely interpolating all the input data points but rather by using a smoothing approach [20],[19].

$$PSS = RSS + ISSD \tag{12}$$

Then, $\overline{P}_e(c)$ is defined to be in the form of (13) [18],

$$\overline{P}_e(c) = \sum_{k=1}^{K} b_k Q_k(c) + \sum_{i=1}^{M} w_i U(r_i) \tag{13}$$

Satisfying,

$$N^T w = 0 \tag{14}$$

Where, K is the number of linear terms in the equation and $K = 3$. The function Q_k represents the independent variables of the surface curve (P_e) such that $Q_0(c) = 1$, $Q_1(c)$=α and $Q_2(c)$=β. b_k and w_i are model coefficients. w is a M×1 matrix of w_i coefficients. N is a M×K matrix where each row of N is, N_i = [1 α_i β_i]. The terms r_{ij} and $U(r)$ are respectively given by,

$$r_i = ||c - c_i|| = ||(\alpha, \beta) - (\alpha, \beta)_i|| \tag{15}$$

$$U(r_i) = r_i^2 \ln(r_i^2) \tag{16}$$

The coefficients b_k and w_i in (13) are known as the model parameters obtained by solving matrix form of (13) and (14) for the given data set. The key objective of the TPS modeling is to obtain the model parameters that would fully define the mathematical representation of P_e allowing regression for any arbitrary values of α and β. This regression model generates a total of $M + K$ coefficients to mathematically describe P_e, with M coefficients for the nonlinear terms and K coefficients for linear polynomial terms.

A. Accuracy of Model

The accuracy of the model can be defined by calculating the residual ξ between the true (P_e) and model ($\overline{P}_e$) values during the validation process using computer simulations, the residual at a particular input data point (α_i, β_j) is given by,

$$\xi_{ij} = \overline{P}_e(\alpha_i, \beta_j) - P_e(\alpha_i, \beta_j) \tag{17}$$

Where, P_e is computationally obtained from the simulations and $\overline{P}_e$ is obtained from (13). Once the residual error is obtained from (17), we compute the root mean squared error, σ_ξ probability density function, f_ξ and the cumulative distribution function, F_ξ to present the accuracy of the developed model.

IV. Model Development and Simulation Results

The study involves transition probabilities of Idle and busy states ranging from 0.1 to 0.9 in increments of 0.1 creating different channel occupancy scenarios. For each α and β pair, The same LSTM model was trained using the modeled spectrum usage of 10,000 time steps with initial learner rate at 0.05 and piece-wise learning rate schedule. Number of maximum epochs was set as 50. The trained LSTM model forecasts spectrum occupancy across 10,000 time steps for each input pattern. This prediction is repeated 1,000 times for robust and reliable results, and the average result is computed.

Graphs in Fig.3 illustrate the variation in prediction error for each α and β pair. The line graphs depict the P_e for a specific β, showing an increase in P_e even when a single state reaches random transitions. The highest prediction error is noted when both α and β equal 0.5, indicating a state of absolute randomness and minimal predictability. Those plots prove that the prediction model does not perform at the same level of prediction results for each channel occupancy condition.

Further, these prediction results show a sensitivity to the transition probabilities despite the long-term channel occupancy described by steady state probabilities. The plot in Fig.4, where $\alpha = \beta$, illustrates P_e's variation across different state transition probabilities, assuming equal steady-state probabilities for channel states. This confirms that α and β impact P_e regardless of steady-state probabilities, affirming P_e's dependency on α and β, and enabling its formulation as a function of α and β.

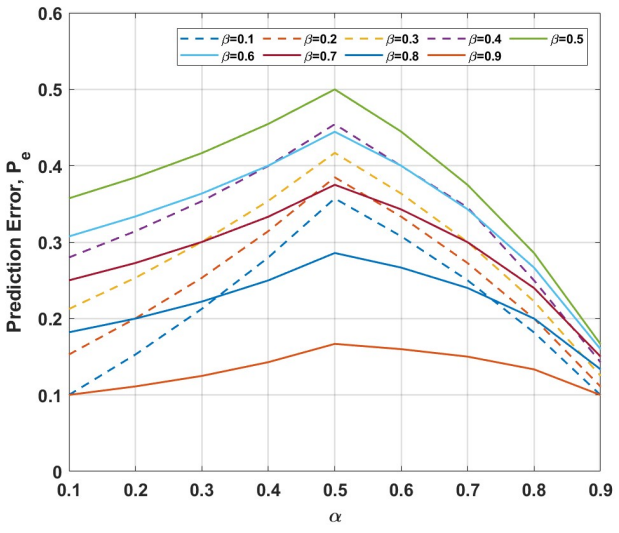

Fig. 3. LSTM prediction error vs transition probabilities.

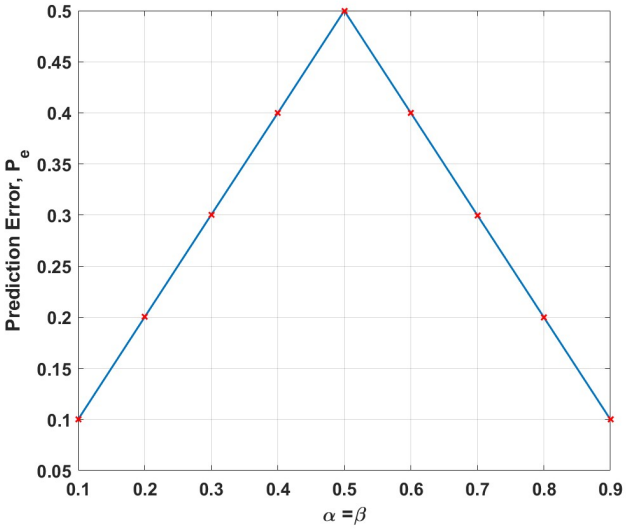

Fig. 4. LSTM prediction error vs equal transition probabilities.

TABLE I
Interpolation Models Comparison

Interpolation Method	Number of coefficients in CFMR	Validation σ_ξ
Nearest Neighbour	No CFMR	0.0423
Linear	No CFMR	0.0151
Cubic spline	1024	0.0134
Support vector regression	82	0.0487
Gaussian process regression	No CFMR	0.0161
TPS	84	0.0136

A. TPS Model Development

Table I presents a comparative analysis of TPS interpolation and other prevalent models from the literature [15],[16],[19] and practice. While cubic spline interpolation exhibits marginally superior accuracy compared to TPS, the Closed Form Mathematical Representation (CFMR) of TPS with 84 coefficients is notably less complex than the Cubic spline interpolation with 1024 coefficients. Consequently, TPS is preferred over the cubic spline approach and other methods.

It is imperative to note that an increased number of data points can lead to the model being over-fitted to the data and increased model complexity while less number of data points may not catch the real variation of the data set. Hence, the TPS model (13) was simulated using three sets of prediction error data points in equally distributed α and β coordinates. The first set includes 22 data points, the second set includes 41 data points, and the third set includes 81 data points. Subsequently, each TPS model developed with these data sets is referred to as TPS_{22}, TPS_{41} and TPS_{81}, respectively. Model testing and validation were performed to identify the best model for LSTM occupancy prediction error modelling.

B. Testing And Validation of Proposed TPS Model

According to the basic concept of smoothing splines, the models are not supposed to interpolate all the input data points

TABLE II
TPS_M MODEL ERRORS

Model - TPS_M	Modeling σ_ξ	Validation σ_ξ
TPS_{81}	3.1709e-16	0.0136
TPS_{41}	2.1555e-16	0.0139
TPS_{22}	1.5150e-16	0.0183

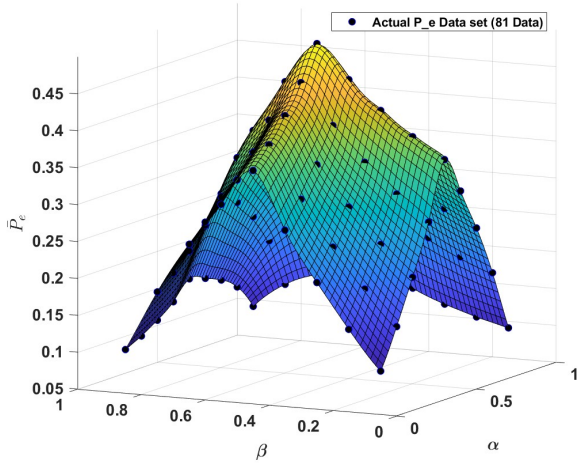

Fig. 5. TPS_{81} Model fit.

precisely, but smoothly capturing the pattern within the data points. Hence, the proposed TPS error model was validated using a dataset of 64 P_e data.

The RMSE results presented in Table II make it clear that a model's modeling error is inversely proportional to the number of data points used for modeling. As the number of data points increases, resulting in a more intricate model, the modeling error also escalates. However, models using more data points tend to exhibit improved accuracy in predicting data during validation. Table II provides the validation σ_ξ, which is critical as it reflects the model's capacity to predict the P_e for diverse spectrum occupancy patterns.

Based on the tabulated σ_ξ results, it can be concluded that the TPS_{81} model exhibits superior accuracy compared to the other two models while f_ξ distribution of both TPS_{81} and TPS_{41} on par with each other. Another approach for selecting the optimal TPS model rather than using the maximum accuracy strategy is to consider the minimum acceptable error strategy, which is not applied in this case. The visual representation of TPS_{81} model fit can be found in fig.5. Fig.7 illustrates the cumulative distribution (F_ξ) of validation residuals of TPS_{81}, which lies within -0.055 and 0.017 and notably, the magnitude of 90.6% of these residuals is less than 0.017. The validation results with higher accuracy confirms that prediction performance of DL models such as LSTM can be formulated as a function of α and β.

The model parameters of TPS_{81} can be found in Table III and Table IV. To facilitate comprehension, (13) has been adjusted as follows (18). coefficients in Table III and Table IV follows the format of (18).

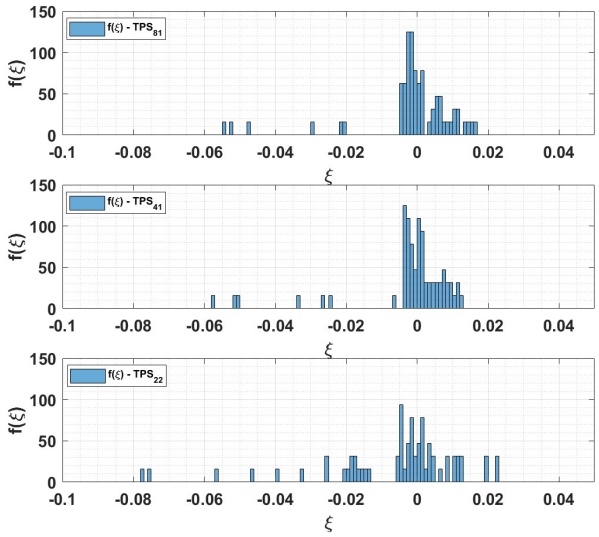

Fig. 6. f_ξ of Validation residuals.

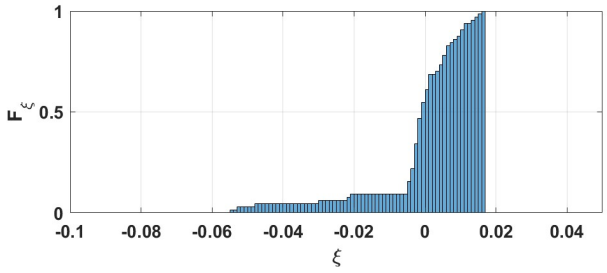

Fig. 7. F_ξ of TPS_{81} Validation residual.

$$\overline{P}_e(\alpha,\beta) = \sum_{k=1}^{K} b_k Q_k(c) + \sum_{i=1}^{R}\sum_{j=1}^{S} w_{ij} U(r_{ij}) \tag{18}$$

Where,

$$r_{ij} = ||c - c_{ij}|| = ||(\alpha,\beta) - (\alpha_i,\beta_j)|| \tag{19}$$

$M = R \times S$ with $i = 1,2,\ldots,R$, $j = 1,2,\ldots,S$. The R and S values are both 9. R is number of α values and S is number of β values used for LSTM prediction, creating a data set of 81 P_es, ranging from 0.1 to 0.9 with a 0.1 increment. Hence, the total number of data points, M, is 81.

V. Conclusion

This paper proposes an error modeling for Long Short-Term Memory (LSTM) channel occupancy prediction. We clearly define a two-state Markovian spectrum occupancy model, an LSTM prediction model, and a Thin Plate Spline (TPS) model for error modeling, along with the performance evaluation matrices. We created an error model using TPS for 81 LSTM prediction error data points. Testing and validation confirmed that a model based on 81 equally distributed data points

TABLE III
COEFFICIENTS OF NONLINEAR ELEMENTS OF TPS_{81}

	β_1 = 0.1	β_2 = 0.2	β_3 = 0.3	β_4 = 0.4	β_5 = 0.5	β_6 = 0.6	β_7 = 0.7	β_8 = 0.8	β_9 = 0.9
α_1 = 0.1	W_{11}=-0.3208	W_{12} =0.0127	W_{13} =0.0821	W_{14} =-0.7578	W_{15} =1.6331	W_{16} =-0.6288	W_{17}=0.1054	W_{18}=0.1230	W_{19}=-0.1229
α_2 = 0.2	W_{21}=0.0188	W_{22}=0.0498	W_{23}=0.1163	W_{24}=-0.7225	W_{25}=0.7925	W_{26}=-0.6765	W_{27}=0.1052	W_{28}=0.0694	W_{29}=-0.1116
α_3 = 0.3	W_{31}=0.0713	W_{32}=0.1250	W_{33}=0.3096	W_{34}=-0.4702	W_{35}=1.1361	W_{36}=-0.4135	W_{37}=-0.0784	W_{38}=0.2226	W_{39}=-0.0533
α_4= 0.4	W_{41}=-0.7579	W_{42}=-0.7178	W_{43}=-0.4777	W_{44}=-1.1955	W_{45}=0.4058	W_{46}=-1.3004	W_{47}=0.4447	W_{48}=-0.4475	W_{49}=-0.3013
α_5 = 0.5	W_{51}=1.6438	W_{52}=0.7923	W_{53}=1.1041	W_{54}=0.358	W_{55}=1.6332	W_{56}=0.2942	W_{57}=0.359	W_{58}=0.7093	W_{59}=0.0725
α_6 = 0.6	W_{61}=-0.6406	W_{62}=-0.6783	W_{63}=-0.4334	W_{64}=-0.9926	W_{65}=0.2492	W_{66}=-0.8117	W_{67}=-0.2344	W_{68}=-0.0261	W_{69}=-0.375
α_7 = 0.7	W_{71}=0.1141	W_{72}=0.0443	W_{73}=0.2138	W_{74}=-0.3528	W_{75}=0.6696	W_{76}=-0.316	W_{77}=0.0259	W_{78}=0.2004	W_{79}=-0.1666
α_8 = 0.8	W_{81}=0.1352	W_{82}=0.043	W_{83}=0.198	W_{84}=-0.1568	W_{85}=0.6833	W_{86}=-0.0424	W_{87}=0.216	W_{88}=0.2678	W_{89}=-0.1028
α_9 = 0.9	W_{91}=-0.1294	W_{92}=-0.0924	W_{93}=-0.072	W_{94}=-0.3561	W_{95}=0.0746	W_{96}=-0.3757	W_{97}=-0.1699	W_{98}=-0.1023	W_{99}=0.2567

TABLE IV
COEFFICIENTS OF LINEAR ELEMENTS OF TPS_{81}

b_0	b_1	b_2
0.3453	-0.0828	-0.0822

provides greater accuracy for the scenario under consideration. Our study reveals a clear correlation between prediction performance and channel occupancy patterns. This insight can be instrumental in selecting an appropriate prediction model for specific input spectrum data or determining the extent of accuracy achievable through incremental training of a prediction model in advance.

REFERENCES

[1] O. Ozyegen, S. Mohammadjafari, E. Kavurmacioglu, J. Maidens, and A. B. Bener, "Experimental results on the impact of memory in neural networks for spectrum prediction in land mobile radio bands," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 2, pp. 771–782, 2020.

[2] X. Xing, T. Jing, W. Cheng, Y. Huo, and X. Cheng, "Spectrum prediction in cognitive radio networks," *IEEE Wireless Communications*, vol. 20, no. 2, pp. 90–96, 2013.

[3] G. Ding, Y. Jiao, J. Wang, Y. Zou, Q. Wu, Y.-D. Yao, and L. Hanzo, "Spectrum inference in cognitive radio networks: Algorithms and applications," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 1, pp. 150–182, 2017.

[4] P. Chauhan, S. K. Deka, B. C. Chatterjee, and N. Sarma, "Cooperative spectrum prediction-driven sensing for energy constrained cognitive radio networks," *IEEE Access*, vol. 9, pp. 26107–26118, 2021.

[5] O. Omotere, J. Fuller, L. Qian, and Z. Han, "Spectrum occupancy prediction in coexisting wireless systems using deep learning," in *2018 IEEE 88th Vehicular Technology Conference*, pp. 1–7, IEEE, 2018.

[6] M. Jia, X. Zhang, J. Sun, X. Gu, and Q. Guo, "Intelligent resource management for satellite and terrestrial spectrum shared networking toward b5g," *IEEE Wireless Communications*, vol. 27, no. 1, pp. 54–61, 2020.

[7] S. S. Fernandes, M. R. Makiuchi, M. V. Lamar, and J. L. Bordim, "An adaptive recurrent neural network model dedicated to opportunistic communication in wireless networks," in *2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 01–08, 2018.

[8] A. M. Schäfer and H.-G. Zimmermann, "Recurrent neural networks are universal approximators," *International journal of neural systems*, vol. 17, no. 04, pp. 253–263, 2007.

[9] X. Wang, T. Peng, P. Zuo, and X. Wang, "Spectrum prediction method for ism bands based on lstm," in *2020 5th International Conference on Computer and Communication Systems (ICCCS)*, pp. 580–584, IEEE, 2020.

[10] N. Radhakrishnan and S. Kandeepan, "An improved initialization method for fast learning in long short-term memory-based markovian spectrum prediction," *IEEE Transactions on Cognitive Communications and Networking*, vol. 7, no. 3, pp. 729–738, 2020.

[11] N. Radhakrishnan, S. Kandeepan, X. Yu, and G. Baldini, "Soft fusion based cooperative spectrum prediction using lstm," in *International Conf. on Signal Processing and Communication Systems*, pp. 1–7, 2021.

[12] 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)*, pp. 1–4, IEEE, 2017.

[13] L. Guo, J. Lu, J. An, and K. Yang, "Dsil: An effective spectrum prediction framework against spectrum concept drift," *IEEE Transactions on Cognitive Communications and Networking*, 2024.

[14] N. Radhakrishnan, S. Kandeepan, X. Yu, and G. Baldini, "Performance analysis of long short-term memory-based markovian spectrum prediction," *IEEE Access*, vol. 9, pp. 149582–149595, 2021.

[15] E. Benini and R. Ponza, "Nonparametric fitting of aerodynamic data using smoothing thin-plate splines," *AIAA journal*, vol. 48, no. 7, pp. 1403–1419, 2010.

[16] S. Shokrzadeh, M. J. Jozani, and E. Bibeau, "Wind turbine power curve modeling using advanced parametric and nonparametric methods," *IEEE Transactions on Sustainable Energy*, vol. 5, no. 4, pp. 1262–1269, 2014.

[17] H. Eltom, S. Kandeepan, B. Moran, and R. J. Evans, "Spectrum occupancy prediction using a hidden markov model," in *International Conf. on Signal Processing and Communication Systems*, pp. 1–8, 2015.

[18] P. J. Green and B. W. Silverman, *Nonparametric regression and generalized linear models: a roughness penalty approach*. Crc Press, 1993.

[19] W. Keller and A. Borkowski, "Thin plate spline interpolation," *Journal of Geodesy*, vol. 93, pp. 1251–1269, 2019.

[20] F. Bookstein, "Principal warps: thin-plate splines and the decomposition of deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 6, pp. 567–585, 1989.