Received 24 December 2023; revised 22 January 2024; accepted 5 February 2024. Date of publication 13 February 2024; date of current version 7 March 2024.

Digital Object Identifier 10.1109/OJCOMS.2024.3365708

Deep Learning Models for Time-Series Forecasting of RF-EMF in Wireless Networks

CHI NGUYEN[1,2], ADNAN AHMAD CHEEMA[1] (Member, IEEE), CETIN KURNAZ[3], ARDAVAN RAHIMIAN[1] (Member, IEEE), CONOR BRENNAN[4] (Senior Member, IEEE), AND TRUNG Q. DUONG[2,5] (Fellow, IEEE)

[1]SenComm Research Lab, School of Engineering, Ulster University, BT15 1AP Belfast, U.K.

[2]School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, BT7 1NN Belfast, U.K.

[3]Department of Electrical and Electronics Engineering, Ondokuz Mayis University, 55270 Samsun, Turkey

[4]School of Electronic Engineering, Dublin City University, Dublin 9, D09Y074 Ireland

[5]Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John's, NL A1C 5S7, Canada

CORRESPONDING AUTHOR: T. Q. DUONG (e-mail: tduong@mun.ca)

This work was supported by the Canada Excellence Research Program.

ABSTRACT Radio-frequency electromagnetic field (RF-EMF) forecasting plays an important role in the evaluation of regulatory compliance, network planning and system optimization. The knowledge of RF-EMF levels is essential to ensure compliance with standards and avoid public health concerns, especially with the arrival of new frequencies and scenarios in fifth-generation (5G) and sixth generation (6G) wireless networks. This work provides a comprehensive study on time series forecasting for RF-EMF measured in frequency from 100 kHz - 3 GHz. The state-of-the-art deep learning model architectures consist of deep neural network (DNN), convolutional neural network (CNN), long-short term memory (LSTM), and transformer are applied for time series forecasting. The prediction performance is evaluated under three different scenarios - namely single-step input single-step output (SISO), multi-step input single-step output (MISO), and multi-step input multi-step output (MIMO). The findings from the simulation demonstrate that SISO forecasting is inadequate in predicting long-term radio-frequency electromagnetic fields (RF-EMF) data as it lacks accuracy while MISO and MIMO forecasting scenarios offer more precise predictions. Specifically, in these two scenarios where the input width and label width are both set to 20 steps, the LSTM and CNN models exhibit superior performance compared to other models. Nonetheless, as the input width and label width in a MIMO scenario increase, the accuracy of both CNN and LSTM models decline considerably, whereas the transformer model consistently maintains good performance. Additionally, the transformer model continues to deliver accurate predictions as the label width and shift length increase, which is not the case for DNN, CNN, and LSTM models.

INDEX TERMS CNN, deep learning, EMF, forecasting, LSTM, RF-EMF, time-series, transformer, 6G.

I. INTRODUCTION

ALTHOUGH the 5G communication technology is still being commercialized across the world, researchers have been paying attention to the 6G to satisfy the future demands by 2030. The motivations for 6G include the explosive growth in mobile traffic, and the emergence of unprecedented applications together with new technologies. In particular, the mobile broadband could reach 17.1 billion subscribers by 2030 and the monthly data consumption by each user may increase from around 5 GB in 2020 to 250 GB in 2030. While 5G has been evolutionary in enabling the Internet-of-Things and Industry 4.0, it may not be able to fully support applications such as smart cities, smart utility services, extended reality, connected autonomous systems, telemedicine and so on. These applications require huge bandwidth on the order of terabit per second along with

image compression (i.e., for holographic-type communications), low latency and high reliability (i.e., for extended reality), and a stringent low latency (i.e., below 1ms for tactile Internet and multi-sense experience). Consequently, 6G has promised to provide more data rate, spectrum and reliability compared to 5G. A remarkable decrease to less than 1ms in delay in 6G (less than 1 ms) is mandatory compared to that in 5G. 6G should increase data rate to a target of 1 Tbps, much higher than 1-20 Gbps in 5G, motivating to a further investigation of frequency higher than 6 GHz. The higher frequency with a spectrum ranging from 73 - 140 GHz and 1 - 10 THz, in small-cell densification could be a key attraction in 6G [1].

Furthermore, artificial intelligence (AI) and data science, including data analytics and data mining, are becoming dominant technologies of 6G development that could have impacts on physical, medium access control, networks and application layers. Particularly, AI has become a key technology required for 6G networks to tackle multiple problems, such as optimization of the radio interface, adaptive beamforming strategies and network management. The crucial role of AI is inevitable, especially when it is hard to solve problems by mathematical equations, or it requires real-time analysis, i.e., for extreme applications such as video monitoring [2].

A. EMF STANDARDS AND LIMITATIONS

Although 6G is expected to create an ecosystem that supports a wide range of applications, there are growing concerns about adverse health effects due to radio-frequency electromagnetic fields (RF-EMF) exposure, urging consideration of RF-EMF exposure based constraints in recent research about 6G. The potential health effects could be a possible heating effect on the exposed tissues, i.e., thermal effect, and whole-body heat stress [3], [4]. Hence, a critical investigation of RF-EMF exposure for multi-band networks is necessary [5]. Due to the problems of RF-EMF exposure in humans, guidelines specified by International Commission on Non-Ionizing Radiation Protection (ICNIRP) (1998) and Information and Communication Technologies Authority (ICTA) (2018) are adopted worldwide. The exposure limits in the research by [6] are more up-to-date and can be adapted by regulators at the national level. The study in [7] provides the exposure level limit for base station operation at 800 MHz, 900 MHz, 1800 MHz, 2100 MHz, and 2600 MHz are 29.1 V/m, 28.8 V/m, 40.72 V/m, and 49.23 V/m respectively, which are recommended by ICTA [8].

Recently, many studies related to 6G wireless networks have considered RF-EMF health concerns by taking into account the constraints of RF-EMF exposure. In particular in the study by [9], the RF-EMF exposure is minimized under the condition of minimum rate requirement for each user in an orthogonal frequency-division multiplexing system. In the study of [10], RF-EMF exposure is minimized subjected to the minimum quality of service (QoS) in a reconfigurable intelligent surfaces (RIS)-aided network. Reference [11] considered the problem of maximization the energy efficiency in a RIS-aided multi-input multiple-output system, subject to maximum power constraints and additional constraints on the maximum exposure of the end-users to electromagnetic radiations. In addition, it is important to investigate the current status of RF-EMF exposure levels in different environments in existing wireless networks to understand current exposure levels and how this information can be used to plan new deployments to minimize exposure and educate the public.

B. RELATED WORK ON RF-EMF

In recent years, RF-EMF related studies have been mainly concentrated on empirical and modeling approaches to ensure safe exposure conditions and to validate that RF-EMF values comply with international and national regulations. In a systematic review of RF-EMF in various micro-environments in the European region by [12], 26 related studies were categorized based on four main methods, namely spot measurement, fixed site measurement, personal measurement with volunteers, and mobile micro-environmental measurement with trained researchers. Most recently, [13] presented an assessment of RF-EMF exposure from the downlink traffic of a commercial 5G mm-Wave base station. The paper concluded that RF-EMF exposure from mm-Wave base station is mostly negligible in a line-of-sight location in the case of no traffic and increases according to the amount of throughput. In contrast to the common belief that densification of 5G networks may increase RF-EMF exposure, [14] provided evidence that the higher number of base stations contributes to reducing RF-EMF exposure from mm-Wave frequencies, beam-forming, and connection of millions of devices.

C. RELATED WORK ON APPLYING ML FOR RF-EMF

Few studies till now have started exploring machine learning (ML) for RF-EMF dataset in wireless networks to solve problems such as clustering, classification and prediction. In [15], an unsupervised ML hierarchical clustering algorithm is applied to determine the patterns of RF-EMF radiation in 205 Greek schools in the region of Thessaly. The analysis is due to the need to identify areas that contain increased RF-EMF values. The RF-EMF strength is determined by the measurements of electric field intensity (in V/m) with the detection range from 27 MHz - 3 GHz and by the frequency selective measurements over interval of 6 minutes at three heights of 110 cm, 150 cm and 170 cm. The RF-EMF dataset contains 6 variables, including area, exists, electric field strength, dense, ratio and limit area. The results show that no specific clustering pattern is found in the observed frequency range. Reference [16], proposed two aritifical neutral network (ANN) models to predict electric field values of random locations which are specific regions for RF-EMF exposure. The RF-EMF values (generated by averaging over 6-minute durations) are measured in the frequency range 27 MHz - 3 GHz at specific locations in the campus of Bursa

Uludag University, Turkey. To predict the average electric field values, two ANN models with different input layers are applied, in which the input of ANN consists of latitude, longitude, and different frequency ranges. Both ANN models obtain good prediction results, with correlation coefficients at 99% for the first ANN and at 92% for the second ANN. Moreover, both the measured and estimated results are in the limit values defined by ICTA.

Regarding time series EMF data, long-short term memory (LSTM) is applied to predict electric field strength levels of high-voltage lines that were measured monthly in Turkey between 2014 and 2018 in the range of 1 Hz - 400 kHz using the Wave control SMP2 device [17]. Based on the 60 monthly time-series data of RF-EMF in the paper, the proposed LSTM is expected to predict the next 12-month electric field values. The RF-EMF dataset containing 60 records that is divided into 42 records for training and 12 records for testing. The LSTM model is compared with other models like autoregressive integrated moving average (ARIMA), extreme learning machine, and multi-layer perceptron etc. It is observed that LSTM model outperforms other models in terms of root mean square error (RMSE). Although being the first study on RF-EMF time-series, the time-series model for in [17] relied on a very limited amount of data for both training and testing, which could lead to an over-fitting model. Furthermore, the dataset was measured in a specific location at the coordinate $(38°N, 41°E)$, thus the time-series problem simply becomes using one input variable and predicting one output variable. In addition, there was no consideration of time series parameters such as the length of input steps (input width), the length of output steps (label width), and the shift of the time-series in [17]. The frequency range of the dataset in [17] is from 1 kHz - 400 kHz, which is very low and hard to cover the current most popular wireless network services such as LTE1800, LTE2600, WLAN, and others.

D. RESEARCH GAPS, MOTIVATIONS AND CONTRIBUTIONS

This section presents the motivations and the contributions of this paper. The motivations of this work are based on the importance of RF-EMF forecasting especially long-term RF-EMF, the significance of RF-EMF predictions based on ad machine learning techniques, and the gaps in previous studies. Also, the differences in comparison with earlier works are highlighted.

It is inevitable that the measurement and analysis of RF-EMF are always further required when it comes to new frequencies, scenarios and applications. However, it is time-consuming and expensive to measure RF-EMF, especially long-term RF-EMF with measurements conducted over days, weeks or even months. In addition, the explosion of new technologies, new frequencies and use cases in 5G/6G networks increases the demands of RF-EMF measurements and modeling remarkably to avoid health hazards, resulting in more time-consuming and costly. To overcome these challenges, better approaches are required for RF-EMF measurements and modeling. In addition, ML techniques should be exploited for RF-EMF modeling, particularly for a better classification and predictions of RF-EMF exposure in similar use cases.

More importantly, RF-EMF forecasting is crucial for assessing regulatory compliance, planning networks, and optimizing systems. The estimation of RF-EMF values helps to monitor the transmit power at base station (BS) to guarantee that radio emissions from these sources remain below the regulatory threshold [18]. In addition, RF-EMF estimation is crucial to optimize the installation of BS locations and other network infrastructure [19]. Recently, the RF-EMF level has been considered as a constraint for the multi-objective optimization problems in wireless networks, such as multiple-input multiple-output (MIMO) system [20], and RIS-aided system [21].

Due to the limited availability of studies on time-series forecasting for RF-EMF data from wireless networks, this paper proposes a comprehensive time-series forecasting for instantaneous long-term RF-EMF data that were measured in a city center in the Altinordu District of Ordu City, Turkey. Although the dataset in this study was conducted in sub-3 GHz, it can still be used for higher bands under new 6G eco-system. This study is different from [17] in that the long-term RF-EMF data were conducted in multiple locations, resulting in a prediction based on multiple variable time-series at the input. Furthermore, the proposed deep learning (DL) model in this paper utilizes instantaneous long-term RF-EMF data to avoid data sparsity, which could lead to over-fitting problems in the training process.

Given the limited research in RF-EMF time-series forecasting and the advancements in machine learning for time-series problems, this study comprehensively explores the application of advanced machine learning techniques for RF-EMF forecasting. Our selection of DL models for instantaneous RF-EMF data is based on recent review studies on DL architectures for time-series forecasting by [22], which compared seven types of DL models and their variants, including multi-layer perceptron, recurrent neutral networks (RNN), echo state networks (ESN), CNN, temporal convolutional networks (TCN), gate recurrent units (GRU) and LSTM. The study trained 3800 models and showed that LSTM and CNN are the best alternatives, with LSTM providing more accurate forecasts.

Motivated by the results of [22], this research employed conventional time-series forecasting models such as CNN and LSTM for the time series forecasting of instantaneous RF-EMF data. In addition, the transformer model which is a novel architecture that leverages the attention mechanism to process the time series sequence [23], has been studied in this paper. Transformer is particularly useful in learning recurrent patterns with long-term dependencies since it can access any part of history regardless of distance [24]. In [24], it was demonstrated that the transformer model provides superior or comparable performance as well as computing

efficiencies compared to RNN and its variants including LSTM and GRU. For these reasons, the transformer model is adopted for RF-EMF time-series forecasting in this paper. Baseline models including linear and DNN have also been considered to compare their performance accuracy. To the best of our knowledge, there is limited research on evaluating DL models' performance for instantaneous RF-EMF data according to time-series forecasting factors, i.e., the single/multi-step ahead, length of input/output steps, and the variation of shift in the time series with multiple input features.

The contribution of this paper includes

- This paper provides a comprehensive time-series forecasting of instantaneous RF-EMF based on long-term RF-EMF measurements in 17 locations in a city center, Ordu, Turkey. This study examines the forecasting of RF-EMF time-series at L1 (location 1) based on multiple time-series inputs from 17 locations.
- Different learning models are used in three different forecasting scenarios, namely SISO, MISO, and MIMO. Linear and DNN models are used in SISO scenario, while DNN, CNN, LSTM and transformer models are evaluated for remaining two scenarios. The models' performance in each scenario is assessed by using various metrics such as root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and coefficient of determination R-squared score (R2 score). The performance of different DL models is evaluated according to the variation of time series parameters, i.e., input width, label width and shifts.
- This study aims to determine the optimized DL model for a RF-EMF forecasting in each scenario to meet the acceptable prediction performance accuracy, with R2 score higher than 0.8 in the case of multi-step output.

The rest of the paper is organized as follows. Section II introduces background on the measured long-term instantaneous RF-EMF and presents data analysis for the dataset. Section III explains window-based time-series forecasting approach, and defines fundamental parameters in a time series window. This Section also presents three time-series forecasting scenarios, including SISO, MISO and MIMO. Section IV provides a detailed presentation of proposed DL models and their corresponding architectures for each scenario. Data pre-processing, model training, and testing process are presented in Section V. The simulation results are discussed in Section VI, and finally Section VII concludes the paper.

II. DATASET

A. DATA MEASUREMENTS

The measurement used in this paper is considered from long-term wideband instantaneous RF-EMF measurements that were conducted in the Altinordu District of Ordu City, Turkey from 100 kHz - 3 GHz, more details on experimental setup can be found in [7]. The purpose of long-term RF-EMF measurement is to capture the daily exposure variations and identify the main sources of RF-EMF. To ensure that rare high peaks were not missed, the measurements were conducted at 15-second intervals over a 24-hour period, thereby enabling the characterization of the time-based RF-EMF features. The measurements were taken at 17 locations in the city center, covering a frequency range of 100 kHz - 3 GHz, starting at 06:00 a.m. and ending at 06:00 a.m. the following day as shown in Fig. 1. Fig. 2 illustrates the long-term RF-EMF data obtained from all locations, which shows that the RF-EMF levels vary during the day with the same pattern/trend, in which the highest values were observed around 12:00 p.m and the lowest during the morning and night hours. This long-term RF-EMF trend motivated the idea of time-series forecasting for this dataset.

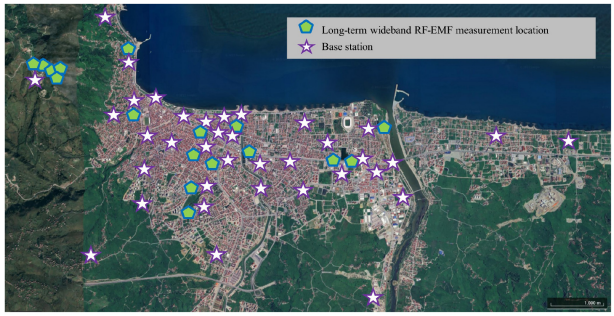


FIGURE 1. Locations of RF-EMF measurements.

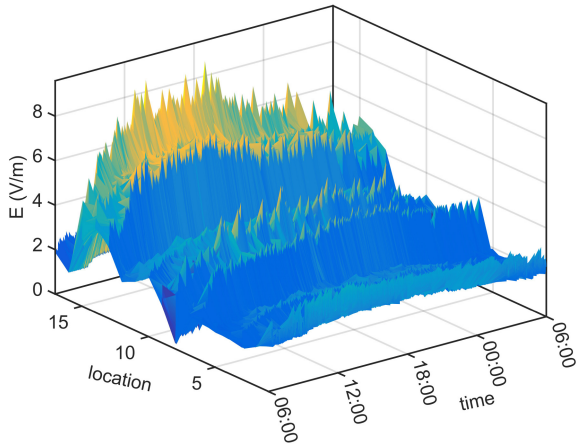


FIGURE 2. RF-EMF long-term data.

The statistics of the dataset are presented in Table 1. The instantaneous long-term RF-EMF values in this paper comprise 5760 data points for each of the 17 locations measured resulting in the dimension of the dataset as (5760, 17). In Table 1, statistical features for each location, including number of samples/data points (count), mean, standard deviation (std), min value, 25%, 50%,

TABLE 1. Statistics of 24-hour instantaneous RF-EMF values in 17 selected measurement locations.

	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15	L16	L17
Count	5760	5760	5760	5760	5760	5760	5760	5760	5760	5760	5760	5760	5760	5760	5760	5760	5760
Mean	2.97	2.02	2.29	2.01	1.92	3.06	2.27	2.45	2.63	3.42	2.66	2.36	4.48	4.95	5.13	2.03	2.60
Std	0.44	0.24	0.44	0.27	0.22	0.96	0.35	1.35	0.50	0.38	0.33	0.19	0.90	0.77	1.34	0.65	0.40
Min	1.85	1.7	1.25	1.52	1.47	1.61	1.76	**0.10**	1.59	2.58	2.17	1.89	3.2	3.63	2.53	1.2	1.83
25%	2.69	1.80	2.03	1.80	1.73	2.16	2.03	1.25	2.27	3.13	2.41	2.27	3.65	4.24	3.99	1.43	2.33
50%	3.09	2.02	2.41	2.05	1.88	3.02	2.26	2.41	2.73	3.5	2.6	2.44	4.49	4.99	5.21	1.92	2.69
75%	3.29	2.15	2.64	2.22	2.12	3.97	2.45	3.69	3.01	3.72	2.94	2.51	5.16	5.61	6.08	2.51	2.82
Max	4.01	2.75	3.19	2.94	2.66	5.16	4.1	6.08	3.84	4.55	3.83	2.8	7.21	6.92	**9.54**	3.91	4.44

TABLE 2. ADF statistics of original series based on 17 locations.

	ADF	p-values	Critical 1%	Critical 5%	Critical 10%
L1	-2.628	0.087	-3.431	-2.862	-2.567
L2	-3.675	**0.004**	-3.431	-2.862	-2.567
L3	-2.552	0.103	-3.431	-2.862	-2.567
L4	-2.685	0.076	-3.431	-2.862	-2.567
L5	-2.646	0.083	-3.431	-2.862	-2.567
L6	-3.744	**0.003**	-3.431	-2.862	-2.567
L7	-2.709	0.072	-3.431	-2.862	-2.567
L8	-4.768	**0.000**	-3.432	-2.862	-2.567
L9	-3.081	**0.028**	-3.432	-2.862	-2.567
L10	-2.819	0.056	-3.432	-2.862	-2.567
L11	-4.284	**0.001**	-3.432	-2.862	-2.567
L12	-1.913	0.326	-3.432	-2.862	-2.567
L13	-2.494	0.117	-3.432	-2.862	-2.567
L14	-2.330	0.163	-3.432	-2.862	-2.567
L15	-3.381	**0.012**	-3.432	-2.862	-2.567
L16	-4.073	**0.001**	-3.432	-2.862	-2.567
L17	-3.689	**0.004**	-3.432	-2.862	-2.567

75%, and the max value are presented in detail. The maximum of long-term RF-EMF value recorded is 9.54 V/m at location L15, while the minimum of long-term RF-EMF value is 0.1 at location L8. Moreover, location L15 recorded the highest average value of 5.314 V/m, while location L8 had the highest standard deviation of 1.358 V/m.

B. STATIONARY TESTING

Deep learning models typically struggle to provide accurate prediction on non-stationary data, and perform better on stationary data [17]. To investigate the statistical characteristics of the RF-EMF dataset, an augmented DickeyŰFuller test (ADF) is conducted. Table 2 shows that the dataset contains stationary and non-stationary features. A feature is considered stationary if its $p-values$ is less than 0.05. Based on this criterion, stationary features are located at L2, L6, L8, L9, L11, L15, L16, and L17. However, the RF-EMF times series are non-stationary, meaning that they need to convert to the stationary dataset before applying machine learning models. The reason behind this transformation is that the non-linear and non-stationary nature of the instantaneous long-term RF-EMF series in this paper makes an accurate forecast difficult [17].

III. TIME SERIES FORECASTING

A. DEFINITION OF WINDOW SIZE IN A TIME SERIES FORECASTING

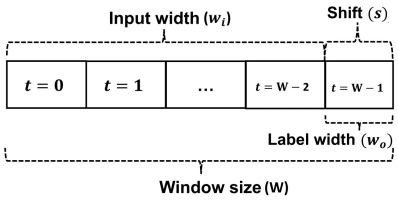
In this paper, the RF-EMF time series data is divided into windows before feeding to DL models. By doing this, the RF-EMF time series prediction becomes time-series forecasting based on windows rather than based on the whole dataset. The window is defined by three main parameters including input width, label width and shift as in Fig. 3. Input width refers to the length of input steps (or input series), label width is the length of output steps (or output series), and shift is the time offset between input series and output series. There are four cases of window size as in Fig. 3. The length of the window size, the input width, and label width are denoted by W, w_i, and w_o, respectively. In Fig. 3(a) and in Fig. 3(b), the label width is illustrated by $w_o = 1$ which corresponds to a window of single-step label. In contrast, Fig. 3(c) and Fig. 3(d) illustrates $w_o > 1$ corresponding to a window of multi-step labels. The symbol s represents the length of shift and defines the future horizon of the time series, regardless of whether the window is a single-step label or multi-step label. The length of the shift (s) together with the input width (w_i) determines the length of the window size that is calculated by $\mathrm{W} = w_i + s$.

B. INPUT AND OUTPUT STEPS OF TIME-SERIES FORECASTING

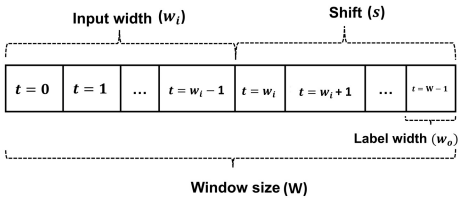
This section presents various use cases/scenarios for time-series forecasting with regard to the number of input and output steps.

1) SINGLE-STEP INPUT, SINGLE-STEP OUTPUT (SISO) PREDICTION

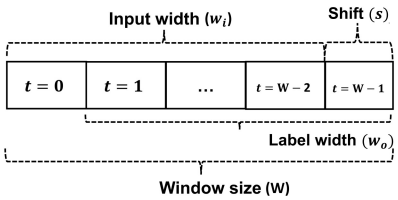
The most simple prediction is that we predict the value of the next time step based on the value of just the previous time step as described in Fig. 4. In particular, the RF-EMF value of time step t can be predicted based on time step $(t-1)$. Fig. 4 provides the visualization of w independent prediction of (input, prediction) by applying a model. Each pair (input, prediction) at time step $((t-1), t)$ is considered as an individual SISO prediction. It is noted that a single prediction can be used for MISO by iteration of SISO forecasting. However, this iteration may cause an accumulation of prediction errors, resulting in a large error in sum.



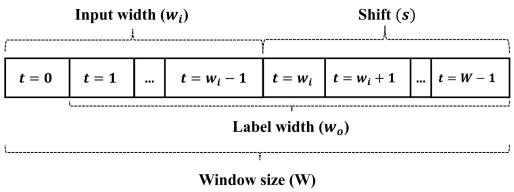
(a) Window 1 for single-step shift and single-step of label width



(b) Window 2 for multi-step shift and single-step label width



(c) Window 3 for single-step shift and multi-step label width



(d) Window 4 for multi-step shift and multi-step label width

FIGURE 3. Combination of different windows based on parameters of time series forecasting.

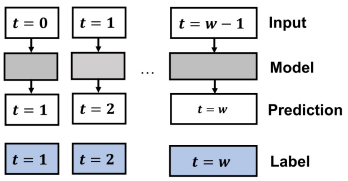


FIGURE 4. Prediction of single-input step and single-output step (SISO).

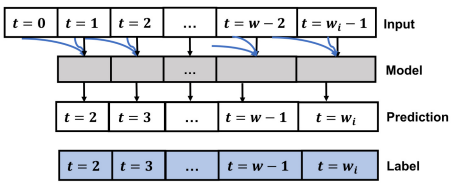


FIGURE 5. Prediction of multi-step input with a small sliding window (example of size of 2 for illustration only).

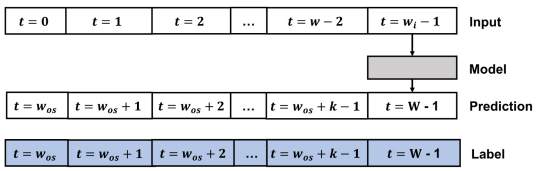


FIGURE 6. Prediction of multi-input prediction without sliding window.

2) MULTI-STEP INPUT (MISO & MIMO) PREDICTION

SISO prediction lacks the context of the current values of its inputs because of no information of how the input time series varies over time. Therefore, a multi-step input is necessary for time-series prediction problem such as in Fig. 5 and Fig. 6. In particular, Fig. 5 is an example of a DL model (e.g., CNN model) using a small sliding window that contains input width $w_i = 2$, output width $w_o = 1$ and shift $s = 1$. The sliding window slides over the whole window size to produce the prediction. Another example of multi-step input prediction (e.g., LSTM model with no return sequence) is described in Fig. 6 in which the prediction of output steps is only based on the last step of the input series. The multi-step input scenario is divided into two scenarios including MISO when label width $w_o = 1$ and MIMO when the length of label width is defined as $w_o > 1$.

IV. DEEP LEARNING MODELS FOR EMF TIME-SERIES FORECASTING

This paper utilizes linear and DNN models for the case of SISO scenario, and DNN, CNN, LSTM, and transformer models for MISO and MIMO scenario. For the SISO case, a linear model is used as a baseline for performance comparison with the other models. On the other hand, DNN model is considered as the baseline in the cases of MISO and MIMO. The goal of this paper is to apply state-of-the-art DL models for RF-EMF time-series forecasting to get high prediction performance, so this section provides a general background of state-of-the-art DL models instead of discussing their architecture and parameters in-depth.

A. PROPOSED DL MODELS

This section presents the DL models used in this paper. The input and output data have three dimensions, with shapes of $(None, w_i, 17)$ and $(None, w_o, 1)$, respectively. Number 17 corresponds to the number of input features (i.e., 17 locations), and number 1 corresponds to the number of output feature (i.e., location L1).

The architecture of DNN models in this paper consists of a *Lambda* layer, one or two hidden layers, and a *Reshape* layer. The *lambda* layer has its own function which is used to modify the input data. Specifically, a *Lambda* layer extracts the information from input data, which is then *Reshape* before being fed to the hidden layer. The *Reshape* layer is

TABLE 3. Model architectures & parameters for SISO case.

Case	No	Model	Layer	Output sequence
SISO	1	Linear	Dense	(None, 1, 1)
	2	Dense	Dense	(None, 1, 1)
			Dense	(None, 1, 1)
			Dense	(None, 1, 1)

used to reshape an output to a desired shape at output. The DNN model employed in MISO case is DNN_1, which uses one hidden dense layer, and while DNN_2 is used in MIMO scenario and contains two hidden dense layers.

This study utilizes Conv1D for the CNN model due to the three-dimensional nature of input data. The *Lambda* layer is adopted in the case of MISO with the same function as in DNN models. Model CNN_1 for MISO is composed of two dense layers, while for MIMO model CNN_1 includes one dense layer. Like DNN models, all CNN models require *Reshape* layer to specify the desired output shape.

Regarding LSTM, this paper proposed two LSTM architectures named LSTM_1 and LSTM_2, in which LSTM_1 contains two LSTM layers and LSTM_2 consists of one LSTM layer. Both LSTM models adopt *Reshape* layer to reshape the output. However, LSTM_1 returns the entire sequence as the output sequence (return_sequence=True), while LSTM_2 only returns the last output in the output sequence (return_sequence=False). As a result, LSTM_2 requires a Dense layer to reshape the output that includes a label width dimension.

The architecture of the transformer model used for RF-EMF time-series data in this paper is inspired by that presented in [23], which is originally for time-series rather than natural language processing and implemented in the Keras library. Transformer architecture consists of encoders and decoders while applying multi-head attention mechanism. To implement the projection layers, a conv1D layer is applied, followed by the average pooling 1D method. Multiple encoders are included in the transformer architecture, and they are connected using *TFOLambda*, namely a TensorFlow operator addition layer. For a more detailed explanation of transformer architecture and its layers, refer to [23].

B. SISO PREDICTION

For SISO case, we select two simple models: linear and Dense. The parameters for these models are described in Table 3. It should be noted that to differentiate the DNN model in SISO from the two remaining scenarios, we referred to the DNN model in this case, SISO, as Dense model. The main difference between dense layer in Linear model and that in Dense model is that the former does not include the activation function, while the latter does. To ensure the activation match to the range of output series of train, valid and test datasets, *Tanh* is used as the activation function for Dense model in this case, as well as the remaining models in this paper.

TABLE 4. Model architectures and parameters for MISO case.

Case	No	Model	Layer	Output sequence
MISO	3	DNN_1	Lambda	(None, 1, 17)
			Dense	(None, 1, 1)
			Reshape	(None, 1, 1)
	4	CNN_1	Conv1D	(None, 1, $filter\_1$)
			Dense	(None, 1, $units\_1$)
			Dense	(None, 1, 1)
			Reshape	(None, 1, 1)
	5	LSTM_1	LSTM	(None, w_i, $cell\_1$)
			LSTM	(None, 1)
			Reshape	(None, 1, 1)
	6	Transformer	Input Layer	(None, w_i, 17)
			Multi head attention	(None, w_i, 17)
			TFOLambda	(None, w_i, 17)
			Conv1D	(None, w_i, 2)
			Conv1D	(None, w_i, 2)
			TFOLambda	(None, w_i, 17)
			Average polling 1D	(None, w_i)
			Dense	(None, 1)
			Reshape	(None, 1, 1)

C. MISO PREDICTION

For MISO prediction, the label width is set to $w_o = 1$. Four models are used to make comparison, as listed in Table 4. The first model is DNN_1, which is a multi-step dense model. In contrast to the dense in Section IV-B that only captures a single input step, the DNN_1 works with multi-step input. The second model, CNN_1 model uses Conv1D layer, where the size of the convolution width is set equal to the length of the input width. Therefore, we have $kernel\_size = (conv\_width,)$, where $conv\_width = w_i$. The number of filters in Conv1D is $filter\_1 = 32$, and the number of dense units of the first Dense layer in CNN_1 is $units\_1 = 32$. In the third model, LSTM_1 model is designed with $return\_sequences = True$, which means that the total sequence is returned in the output. The parameters of the first LSTM layer in LSTM_1 depend on input width w_i, and the number of LSTM cell that is denoted by $cell\_1 = 18$. Although the architecture of the transformer model in [23] is for time-series data, several hyperparameters require an adjustment to work with the RF-EMF dataset. The parameters of the transformer model in this paper are the size of multi-head attention $head\_size = 2$, the output dimension of all sub-layers as well as embedding layers $ff\_dim = 2$, and the number of encoder blocks adopted by $num\_transformer\_blocks = 2$. Again, all models in this section use Tanh activation function.

D. MIMO PREDICTION

For MIMO prediction, five models are considered including DNN_2, CNN_2, LSTM_1, LSTM_2, and Transformer which are compared in Table 5. The difference of MISO models and MIMO models is that the latter are designed

TABLE 5. Model architectures and parameters for MIMO case.

Case	No	Model	Layer	Output sequence
MIMO	7	DNN_2	Lambda	(None, 1, 17)
			Dense	(None, 1, $units\_2$)
			Dense	(None, 1, w_o)
			Reshape	(None, w_o, 1)
	8	CNN_2	Lambda	(None, w_i, 17)
			Conv1D	(None, 1, $filter\_2$)
			Dense	(None, 1, w_o)
			Reshape	(None, w_o, 1)
	9	LSTM_1	LSTM	(None, w_i , $cell\_2$)
			LSTM	(None, w_o)
			Reshape	(None, w_o, 1)
	10	LSTM_2	LSTM	(None, $cell\_2$)
			Dense	(None, w_o)
			Reshape	(None, w_o, 1)
	11	Transformer	Input Layer	(None, w_i, 17)
			Multi head attention	(None, w_i, 17)
			TFOLambda	(None, w_i, 17)
			Conv1D	(None, w_i, 2)
			Conv1D	(None, w_i, 2)
			TFOLambda	(None, w_i, 17)
			Average polling 1D	(None, w_i)
			Dense	(None, w_o)
			Reshape	(None, w_o, 1)

for multi-step output, as reflected in the shape in the last layer. DNN_1 and DNN_2 exhibit similar structures, just as CNN_1 and CNN_2 also possess comparable architecture. But these DNN_2 and CNN_2 differ slightly from DNN_1 and CNN_1, respectively. In particular, number of units in DNN_2 (*unit*_2) is nominated as 32, and the number of filters in Conv1D layer of CNN_2 is set up as *filter*_2 = 32. In MIMO case, LSTM_1 has the same number of cells in a LSTM layer as LSTM_2, indicated by *cell*_2 = 32. Regarding the transformer, the model architecture and hyperparameters are similar to those in the case of MISO, except for the multi-step output w_o in the last layer.

V. PROPOSED DL MODEL TRAINING AND TESTING

A. DATA PRE-PROCESSING

In order to improve the accuracy of time-series forecasting, we apply the differencing technique for the original dataset to make sure the dataset is stationary (which means all features are stationary). The successive differencing is kept applying until the time series becomes stationary series. The auto-correlation function (ACF) is used to check the necessary order of differencing for achieving stationary data. ACF graph decays rapidly around zero value when increasing the order of differencing in the stationary dataset; in contrast, ACF graph firstly decreases and then varies periodically in the non-stationary dataset. This means that if the series still exhibits positive ACF at high lags, it may require

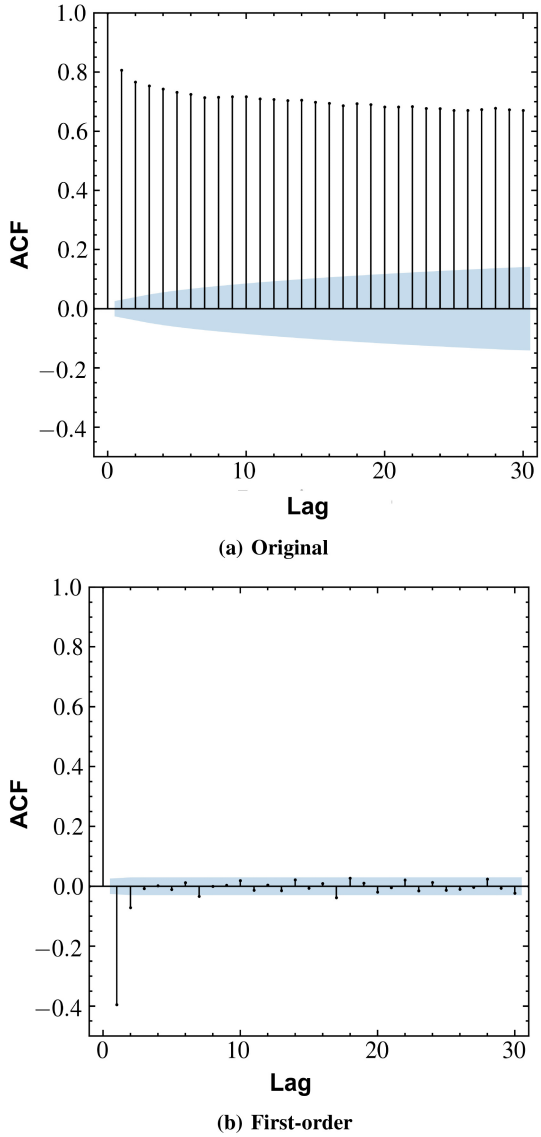


FIGURE 7. ACF of the original data and the first-order difference data at location L1.

further differencing [25], [26]. The ACF figures at all other locations (L2, L3, . . . , L17) produce similar results as that of the first-order differencing at L1. For convenience, only the ACF graphs of location L1 as shown in Fig. 7. These figures illustrate the ACF of the dataset at location L1 before and after implementing first-order differencing. It is observed that the first-order of differencing is sufficient to eliminate the non-stationary characteristics of the RF-EMF series as the ACF values from lag 2 start fluctuating around zero. After obtaining the first-order difference data, we split it into training, validation and testing purpose with the ratio 0.7: 0.2: 0.1. We then apply normalization to the first-order difference on each dataset (train/test/validation) separately, ensuring the range of values falls between [min, max] = [0, 1]. Fig. 8 presents the statistics of training data at 17 locations after applying difference and normalization.

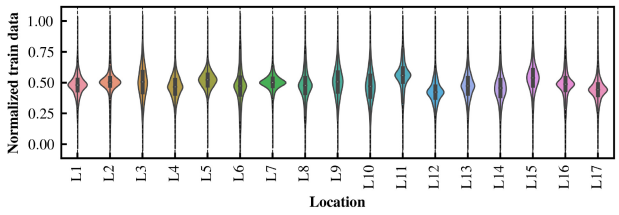


FIGURE 8. Statistics of train data at 17 locations after operating first-order difference and normalization.

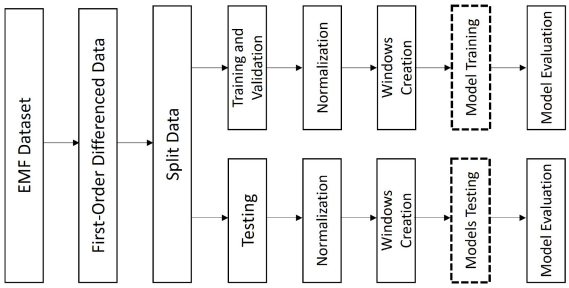


FIGURE 9. Data processing, training and testing process for each DL model.

B. MODEL TRAINING AND TESTING

All processed data is fed to the proposed DL models of each scenario at the same time. In each scenario, all models are trained and their hyperparameters are tuned. The number of epochs is 100, the batch size is kept to 8, and the patience for early stopping is set to 20. The early stopping technique is utilized to monitor the performance of the model on validation dataset. The model training will stop when there is no further improvement in performance on the validation dataset, which maintains the generalization of the DL models on validation (unseen) dataset. The performance of all models is evaluated on both validation and test datasets as seen in Fig. 9.

C. PERFORMANCE METRICS

The performance accuracy of DL models in this paper is evaluated using common performance metrics such as RMSE, mean square error (MSE), MAE, MAPE and R2 score on the normalized test dataset. In particular, RMSE is the square root of an average squared of the difference between the original values and the predicted values, while MSE calculates the average of the squared difference between the label/target and predicted values in the data set. The MSE measures the variance of the residuals, while RMSE estimates the standard deviation of the residuals. RMSE and MSE is expressed as

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}, \tag{1}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2, \tag{2}$$

where y_i is the i_{th} target value and $\hat{y}_i$ is the i_{th} predicted value. The MAE determines the average of residuals by taking the average of the absolute difference between the actual and target values in the dataset and is expressed as

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|. \tag{3}$$

Different from the three above metrics, which calculated the error directly, MAPE is one of the most used metrics that calculate the forecast error by percentage and define as

$$MAPE = \frac{100\%}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right|. \tag{4}$$

Lower values of the RMSE, MSE, and MAPE indicate lower the prediction errors are, and hence a better prediction. Additionally, R2 score is a critical metric in prediction or forecasting which measures the extent to which the predicted values follow the trend of the target values. A higher value of R2 score denotes a better prediction. The R2 score is expressed as

$$R2 = 1 - \frac{SSE}{SST}, \tag{5}$$

where

$$SSE = \frac{1}{N}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \tag{6}$$

$$SST = \frac{1}{N}\sum_{i=1}^{N}(y_i - \mu_y)^2, \tag{7}$$

in which μ_y is the standard deviation of y. The evaluation of the performance metrics is carried out on the normalized test dataset which is divided into windows. As a result, the performance metrics are obtained by taking the average of all windows. This paper presents the first two windows of predicted RF-EMF data w.r.t their target/label, for the purpose of convenient visualization.

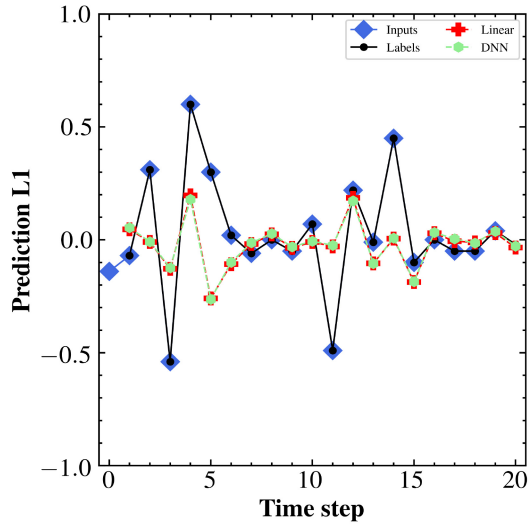
VI. SIMULATION RESULTS

A. SISO PREDICTION

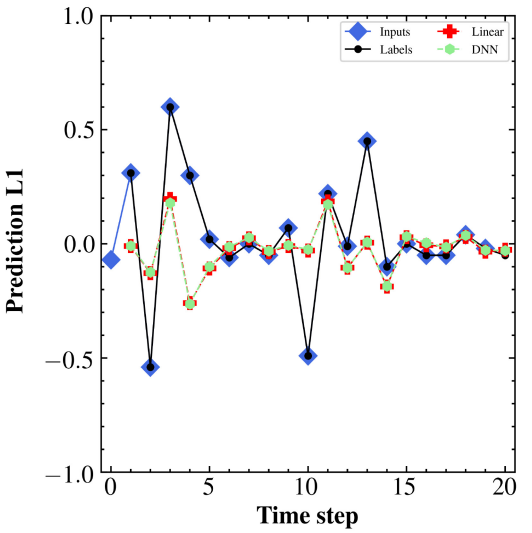
This section focuses on the simplest form of prediction, where a single input step is used to predict one output step. Each (input, output) pair is considered independent, and there is no relationship between the prediction points at different time steps. For the SISO case, $w_i = 1$, $w_o = 1$, and $s = 1$ are set, and 20 (input, output) pairs were analyzed together. The linear and dense models were chosen as the two simple models for this prediction case, as shown in Table 3. The results in Table 6 reveals that the performance accuracy of both models is comparable. The results presented in Fig. 10 indicate that the linear and dense models fail to deliver accurate predictions on the test data in its original scale. Both linear and dense models produce predictions that do not align well with the labeled data. The drawback of utilizing prediction models in SISO case is that the present time step

TABLE 6. Performance accuracy of SISO prediction.

Metrics	Linear	Dense
RMSE	0.0783	**0.0782**
MSE	0.0061	**0.0061**
MAE	0.0587	**0.0586**
MAPE	0.1304	**0.1303**



(a) Window 1



(b) Window 2

FIGURE 10. Prediction of SISO for each 20 continuous (input, label) pairs on original-scaled test data.

lacks any connection with the previous one, prompting the need to explore the prediction of multiple input steps.

B. MISO PREDICTION

Unlike SISO where the output of a single step depends solely on the input at a single time step, the MISO scenario predicts a single output step based on multiple steps from the input time series. In other words, in MISO, the model has the ability to retain and extract more information from multi-step input series to make predictions, which is not the case in SISO. Table 7 indicates that the performance measures of the four DL models in the MISO scenario are better than those of the linear and dense models in the SISO case. LSTM_1 performed better than DNN_1, CNN_1, and Transformer, possibly because the LSTM architecture is specifically designed for long-term memory time-series prediction problems. Fig. 11 illustrates the prediction of the first two windows, which reveals that all four models produce predictions that closely align with the label in the case of a single-step output prediction. In this case, it is evident that LSTM_1 model is the optimal model. However, according to the prediction graph in Fig. 11, DNN_1, CNN_1 and Transformer also provide relatively accurate prediction, only slightly lower than LSTM_1. This is possible as input width is only at $w_i = 20$ and there is single-step output for prediction with a shift length of 1 ($s = 1$) These time-series parameters may not be long enough to create significant differences in terms of the prediction ability among DL models. To further access the performance of these DL models, we will investigate how they perform when the number of prediction steps or the length of the label width, the input width and the length of the shift are increased.

TABLE 7. Performance accuracy of MISO prediction with $w_i = w_o = 20$ and shift $s = 1$.

	DNN_1	CNN_1	LSTM_1	Transformer
RMSE	0.0779	0.0750	**0.0735**	0.0764
MSE	0.0061	0.0056	**0.0054**	0.0058
MAE	0.0584	0.0570	**0.0557**	0.0582
MAPE	0.1301	0.1283	**0.1218**	0.1259

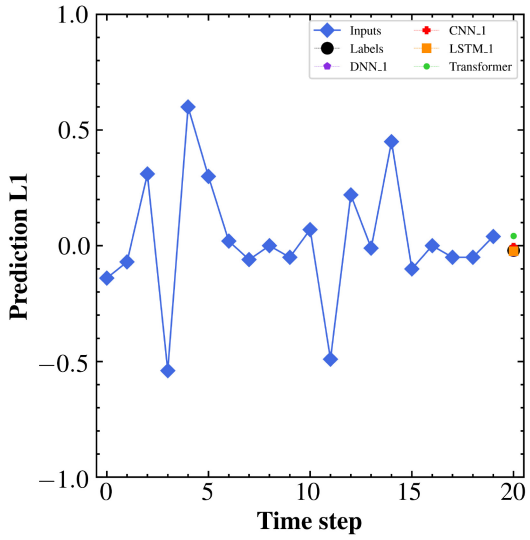
TABLE 8. Performance accuracy of MIMO prediction with $w_i = w_o = 20$.

	DNN_2	CNN_2	LSTM_1	LSTM_2	Transformer
RMSE	0.0821	0.0189	0.0196	0.0201	0.0200
MSE	0.0067	0.0004	0.0004	0.0004	0.0004
MAE	0.0602	0.0103	0.0106	0.0114	0.0096
MAPE	0.1351	0.0223	0.0239	0.0248	0.0215
R2	0.0585	0.9435	0.9460	0.9431	0.9436

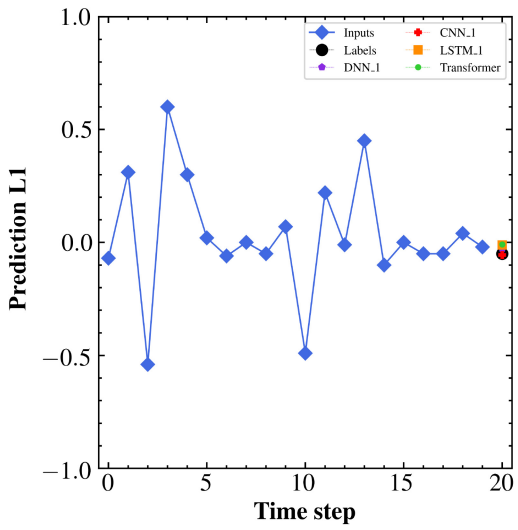
C. MIMO PREDICTION

1) COMPARISON BETWEEN FIVE DL MODELS

To begin with, we examine the performance accuracy when the input width is equal to the label width, that is $w_i = w_o = 20$, and shift $s = 1$. Table 8 illustrates the performance accuracy of MIMO case. When label width is $w_o = 20$, DNN_2 model provides very high RMSE and low R2 score compared to the four remaining models. In contrast, CNN_2 provides the best performance accuracy among four models. LSTM_1, LSTM_2 and transformer show similar performance, whereas DNN is not specifically designed for long-term memory, resulting in poor performance when the number of input width increases. The prediction results of DNN_2 differ significantly from the labels, as shown in Fig. 12, while the remaining models provide predictions

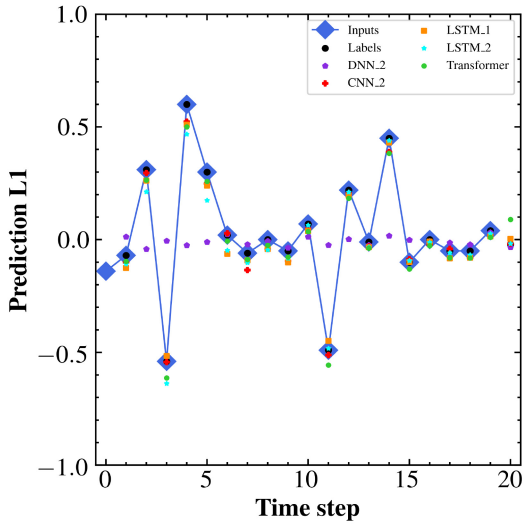


(a) Window 1

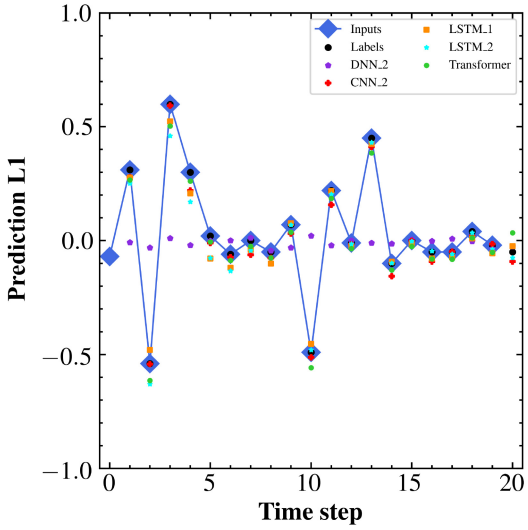


(b) Window 2

FIGURE 11. Prediction of MISO with window size $W = 20$ on original-scaled test set.



(a) Window 1



(b) Window 2

FIGURE 12. Prediction of MIMO in the case of $w_I = w_o = 20$ and $s = 1$ on original-scaled test dataset.

that are relatively closed to the labels and follow the trend well. Table 8 presents various performance metrics for all the observed DL models. Table 8 shows that four models, including CNN_2, LSTM_1, LSTM_2 and Transformer, have achieved the same MSE. In terms of RMSE, CNN_2 obtains a slightly lower error compared to LSTM_1, LSTM_2, and Transformer. As for MAE and MAPE, the transformer has a slightly lower error compared to CNN_2, LSTM_1, and LSTM_2. When it comes to R2 score, LSTM_1 provides a slightly higher accuracy compared to CNN_2, LSTM_1, and LSTM_2.

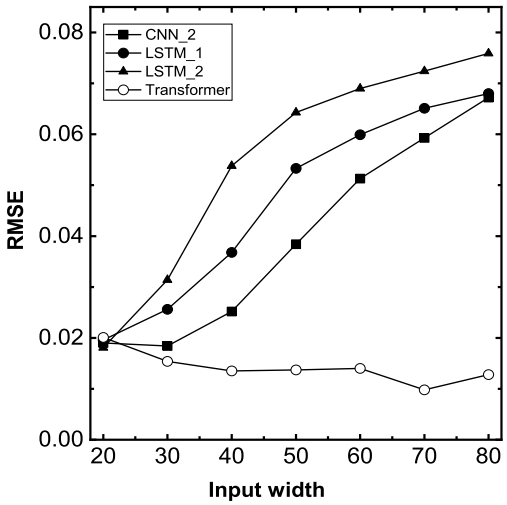
Even though there is a minor variation of performance metrics among CNN_2, LSTM_1, LSTM_2, and transformer in this case; these differences can be disregarded as the four models perform relatively similarly. The slight differences in performance between these four models arise from the variation of the loss during training. In addition, none of the four models demonstrates significant superiority over others, possibly due to input width, which is $w_i = 20$, not long enough to make the distinguish between models' memory and forecasting capabilities. Therefore, the authors are motivated to investigate the performance of these models with different input widths in the following subsection.

2) PERFORMANCE OF FOUR COMPARATIVE MODELS W.R.T INPUT WIDTH

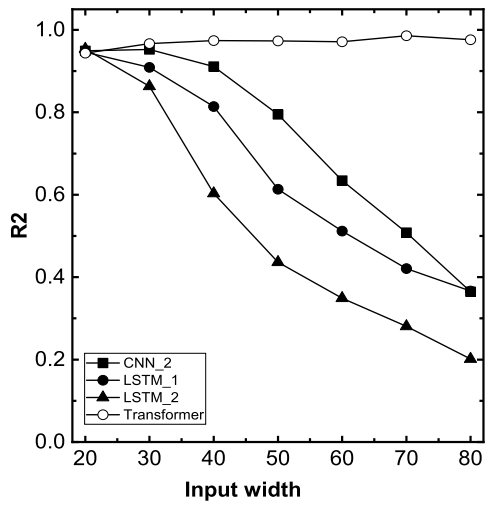
Given that short input width as well as label width (i.e., $w_i = w_o = 20$) and shift $s = 1$, CNN_2 achieves the best performance accuracy. In this section, we investigate the performance accuracy of four models, which are CNN_2,

TABLE 9. Performance accuracy of transformer w.r.t label width given that $w_i = 80$ and $s = 1$.

	80	70	60	50	40	30	20	10
RMSE	**0.0113**	0.0198	0.0262	0.0240	0.0213	0.0160	0.0193	0.0271
MSE	**0.0001**	0.0004	0.0007	0.0006	0.0005	0.0003	0.0004	0.0007
MAE	**0.0056**	0.0159	0.0223	0.0192	0.0164	0.0068	0.0093	0.0159
MAPE	**0.0122**	0.0354	0.0493	0.0402	0.0363	0.0156	0.0209	0.0348
R2	**0.9813**	0.9437	0.9014	0.9163	0.9327	0.9615	0.9436	0.8886



(a) RMSE w.r.t input width



(b) R2 w.r.t t input width

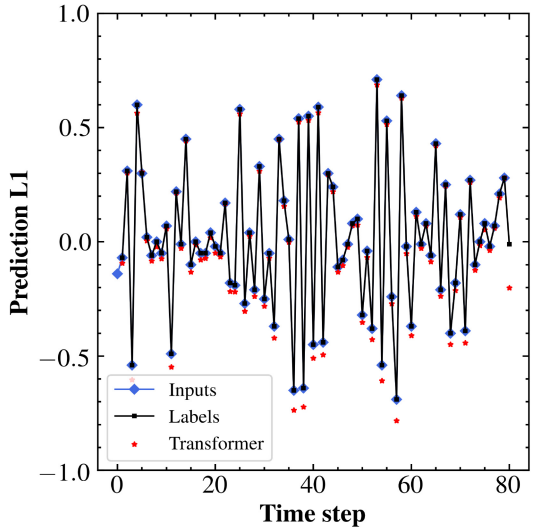
FIGURE 13. Performance accuracy w.r.t input width given that $w_i = w_o$ and $s = 1$.

LSTM_1, LSTM_2, and transformer, as the input width w_i increases while maintaining $w_i = w_o$ and $s = 1$. DNN_2 is excluded from the comparison because of its poor performance in Section VI-C1. As seen in Fig. 13, there is a little difference in prediction performance among the four models when $w_i = 20$. These findings align with those from Section VI-C1. However, as the length of input width w_i increases, the prediction performance gap between four models becomes apparent. Specifically, in Fig. 13 at $w_i = 80$, the transformer model shows good prediction accuracy with low RMSE and high R2 score which is in contrast to a significant decrease in prediction performance in the other three models. The R2 score values of the observed DL models (CNN_2, LSTM_1, and LSTM_2) are at least 0.8 for input widths of 50, 40, and 30, respectively. On the other hand, the transformer model achieved R2 score values higher than 0.9 for all the observed input widths. Therefore, it can be concluded that the transformer model is a suitable option for predicting RF-EMF time-series in a MIMO setting when the input width is relatively high.

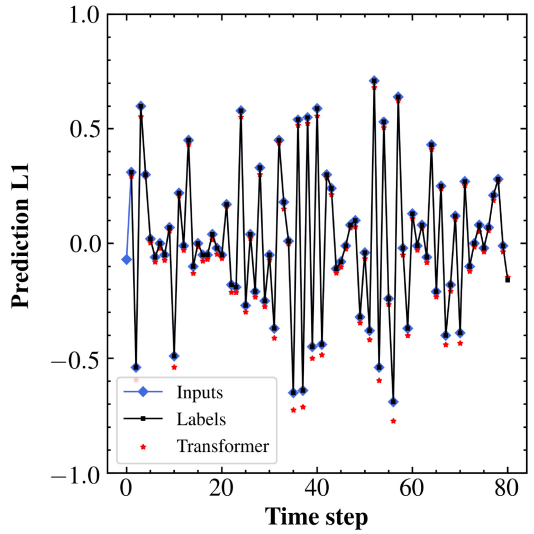
Model LSTM_2 has higher RMSE and lower R2 score compared to LSTM_1. The rationale behind the higher RMSE and lower R2 score of the LSTM_2 model compared to LSTM_1 is that LSTM_1 returns the whole sequence in the output sequence while LSTM_2 only returns the last output. For the input width range of $w_i = 30$ to $w_i = 70$, LSTM_1 shows better performance in terms of prediction accuracy. However, at $w_i = 80$, LSTM_1 and CNN_2 have similar prediction performance, as shown in Fig. 13. It appears that LSTM_1's ability to capture long-term time steps is not effective when the input width is up to $w_i = 80$, which could be improved by using the transformer model.

3) PERFORMANCE ACCURACY OF TRANSFORMER W.R.T LABEL WIDTH

In the previous subsection, we examined how the performance of the four DL models is affected by changes in input width (w_i) while keeping the output width (w_o) constant. In this subsection, we will observe the performance accuracy of the selected optimal model (i.e., transformer) with regard to changes in output width. The reason for choosing the transformer model is due to its good performance even when input width increases. To conduct the experiment, the parameters are set to $w_i = 80$, $w_o = [start = 10, end = 80, step = 10]$, and $s = 1$, where label width (w_o) varies from 10 to 80 with a step size of 10. As seen in Table 9, transformer provides good prediction (R2 score over 0.9) with label with $w_o > 20$. When label width $w_o = 10$, R2 score starts to decrease slightly below 0.9. In general, transformer achieves good performance accuracy when adjusting label width. The shorter the label width is, the more unlikely trend of the target/label becomes clear; leading to a reduction in performance accuracy. The results are in Table 9, the minimum length of label width to obtain the R2 score above 0.9 is 20 in the case of $w_i = 80$ and $s = 1$. The prediction on original-scaled data of the Transformer

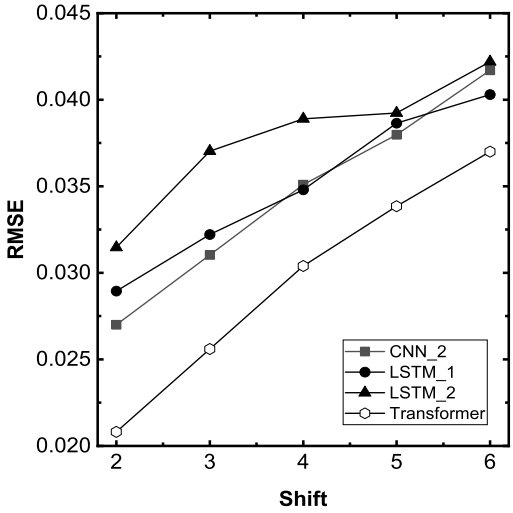


(a) Window 1

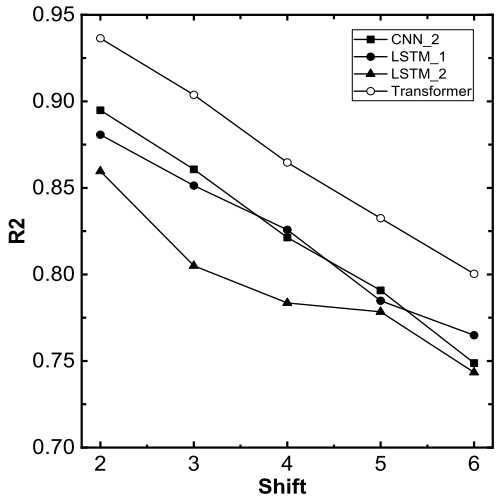


(b) Window 2

FIGURE 14. Prediction on original-scaled data with Transformer model in the case of $w_i = w_o = 80$ and $s = 1$.



(a) RMSE vs shift



(b) R2 vs shift

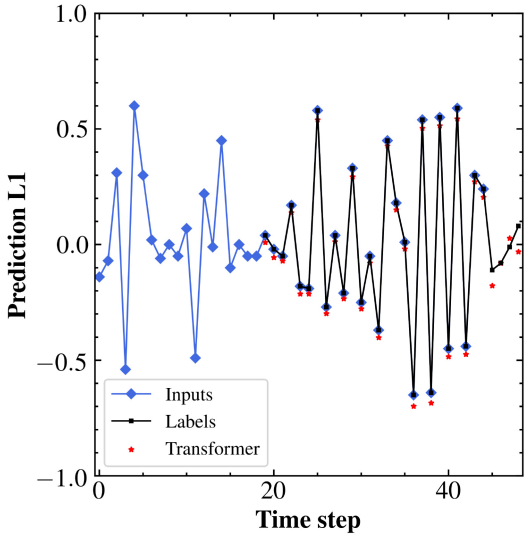
FIGURE 15. Performance accuracy w.r.t shift.

model for the first two windows in the case of $w_i = w_o = 80$ and $s = 1$ is presented in Fig. 14.

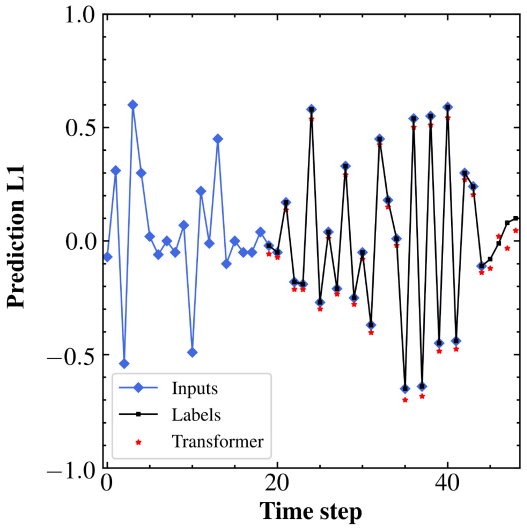
4) PERFORMANCE ACCURACY W.R.T SHIFT

In the previous sections, all experiments were performed under the condition of $s = 1$. However, in this section, we examine the performance accuracy of four models - CNN_2, LSTM_1, LSTM_2, and transformer - with respect to the variation of shift. The input width is set to $w_i = 45$, label width to $w_o = 30$, and shift to $s = [start = 1, end = 6, step = 1]$. As shown in Fig. 15(a), shift is the parameter that has the most significant impact on the performance accuracy of the four models. Specifically, a slight increase in shift leads to a notable increase in RMSE and a decrease in R2 score.

The graph in Fig. 15(b) indicates that the transformer model performs better than the other three models, including CNN_2, LSTM_1, and LSTM_2, in terms of both RMSE and R2 score for all values of shift. For example, when shift is 2, all four models achieve a performance of over 0.85, whereas when shift is 5, only the transformer model attains an R2 score value of nearly 0.8. The transformer model's prediction with shift $s = 4$ is depicted in the figure, which closely follows the trend of the label, as shown in the two windows of Fig. 16. The performance of the three models, CNN_1, CNN_2, and LSTM_2, is compared in Fig. 16. The results demonstrate that LSTM_2 performs the worst in terms of both RMSE and R2 score, whereas CNN_1 and LSTM_1 have relatively similar performance metrics for both RMSE and R2 score. The reason for this lies in that



(a) Window 1



(b) Window 2

FIGURE 16. Prediction on original-scaled data with Transformer model in the case of $w_i = w_o = 80$ and $s = 4$.

the LSTM_2 model only returns the last output in the output sequence, which may reduce its forecasting performance. The intersection observed between CNN_2 and LSTM_1 at the shift of 5 in Figure 15 is influenced by the variation of loss during training process. These results are consistent with those in Section VI-C2.

To summarize, when it comes to forecasting RF-EMF time-series, using MISO and MIMO methods yields better performance than SISO prediction. In multi-step input prediction, the transformer model generally performs well across various time-series parameters such as input width, output width, and shift, when compared to other models based on DNN, CNN, and LSTM architectures. Based on the simulation results, the transformer model demonstrates significantly better prediction performance compared to the other models when the number of input width (w_i) is high, given that $w_i = w_o$. In terms of LSTM architecture, LSTM_1 model that returns the total sequence in the output sequence, performs better than LSTM_2 model that only returns the last output in the output sequence in both MISO and MIMO cases. LSTM_2 model also performs worse than the other DL models in multi-step input prediction. When comparing LSTM and CNN architecture, LSTM_1 outperforms CNN_1 in the case of MISO, but only for input widths lower than 80. When the input width reaches 80, LSTM_1 no longer performs better than CNN_1 in terms of prediction performance.

On the other hand, LSTM_2 and CNN_2 exhibit only minor differences in their prediction performance in the MIMO case. It is evident that the LSTM architecture outperforms the CNN architecture only when the input width is relatively short, i.e., within the range of $[20, 60]$, and the output width is a single step. Therefore, for time-series forecasting of RF-EMF data with short input and output widths, i.e., $w_i = w_o = 20$ with a shift of 1, both CNN and LSTM architectures can be used. However, when the input and output widths increase to a higher value, i.e., $w_i = w_o = 80$, the transformer model is a good option that offers high prediction performance, especially with an R2 score score above 0.8. Furthermore, all architectures are sensitive to increases in the length of shift, where the prediction accuracy of CNN and LSTM models significantly decreases (R2 score below 0.8) at $s = 5$. In contrast, the transformer model maintains an acceptable prediction accuracy even when $s = 6$. In summary, the transformer model is a suitable candidate for long input and label widths, particularly when the shift increases.

Transformer outperforms other DL models especially when the input width, output width and the shift increase in this paper. Transformer architecture does not contain sequential characteristics as in LSTM in which the future time step does not depend on the previous hidden states. The entire sequence is processed simultaneously at the same time instead of processed each time step. In other words, transformer model shows its superiority in capturing distant or long-range time steps and understanding dependencies within data. This capability enables the learning of the relationship between time steps in a longer time-series sequence. Therefore, the transformer is not at the risk of forgetting or losing the past information. This ability can be achieved by using multi-head attention which contributes valuable insights into the relationships between different information in different time steps [23]. In the mean while, LSTM and CNN, information must navigate through numerous processing steps to cover a considerable distance, resulting in challenges in learning [27]. Furthermore, there is no assumption of temporal or spatial relationship in Transformer. By learning the entire sequence simultaneously Transformer is well-suits for a long time-series sequence [28].

On the contrary, parallel training is not feasible for LSTM. Encoding the information in the second time step requires the computation of information in the first step beforehand. The point is that the encoding of a specific step information is retained only for the next time step, which means that the encoding of information strongly affects only the representation of the step information in the next step. Consequently, the influence of the information in the previous step is quickly lost after a few time steps. While this issue can be addressed by implementing deeper processing of hidden states through specific units or utilizing a bi-directional LSTM model, the problem is fundamentally associated with recursion. This drawback causes LSTM to experience a decline in performance as the sequence length increases [27]. In the meanwhile, CNN deals with the dependence between information in each time-series sequence by applying different kernels to the same sequence. For example, kernel size 2 can learn the relationship of information between a pair of time steps. Kernel size 3 would capture the relationship for 3 time steps and so on. The challenge with CNN in handling time-series data lies in the need for a substantial number of kernels to comprehend relationships in lengthy time-series sequences [27].

In sum, transformer is superior to other counterpart DL models in RF-EMF forecasting based on applying the mechanism of multi-head attention. By doing so, transformer can avoid the recursion problem in LSTM or an issue of the large number of kernels in CNN. While the transformer architecture proposed in this paper may not be optimized for extremely lengthy sequences, it is demonstrated to be effective with the length of the RF-EMF dataset that is observed in this study.

VII. CONCLUSION

This study has utilized a dataset of long-term instantaneous RF-EMF measurements taken in the city center of Ordu city, Turkey. The study presented a comprehensive analysis of the time-series data and data processing techniques, aimed at addressing time-series forecasting problems in three scenarios - SISO, MISO, and MIMO. Various DL models and their architectures were applied to the RF-EMF dataset, and their performance was evaluated in each scenario. The results indicated that DL models such as DNN, CNN, and LSTM perform better in terms of prediction accuracy when the input or output steps are short, around 20. However, the Transformer model has shown good prediction performance with higher R2 score values even when the length of input steps is increased, and the output steps and shift are varied. Overall, the transformer model has demonstrated its robustness in addressing time-series forecasting problems.

In future work, it is necessary to explore various architectures of Transformer that are suitable to address the RF-EMF forecasting problem in different scenarios. Such scenarios can be based on wireless networks in indoor (e.g., WiFi), 5G New Radio (beamforming focus), reconfigurable intelligent surface enabled networks and how knowledge of RF-EMF forecasting can be used to develop RF-EMF constraints future networks. Apart from refining the architecture of DL models, it may be beneficial to introduce an enhanced cost function that could reduce prediction errors.

REFERENCES

[1] H. H. H. Mahmoud, A. A. Amer, and T. Ismail, "6G: A comprehensive survey on technologies, applications, challenges, and research problems," *Trans. Emerg. Telecommun. Technol.*, vol. 32, no. 4, 2021, Art. no. e4233.

[2] M. Liyanage, "Explainable eAI for B5G/6G: Technical aspects, use cases, and research challenges," 2021, *arXiv:2112.04698*.

[3] B. Thors, D. Colombi, Z. Ying, T. Bolin, and C. Törnevik, "Exposure to RF EMF from array antennas in 5G mobile communication equipment," *IEEE Access*, vol. 4, pp. 7469–7478, 2016.

[4] M. A. Saeidi, H. Tabassum, and M.-S. Alouini, "Multi-band wireless networks: Architectures, challenges, and comparative analysis," 2022, *arXiv:2212.07606*.

[5] "Accurately assessing exposure to radio frequency electromagnetic fields from 5G networks," Ericsson, Stockholm, Sweden, White Paper, 2021. [Online]. Available: https://www.ericsson.com/en/reports–papers/white-papers/accurately-assessing-exposure-to-radio-Freq.-Electromagn.-fields-from-5g-Netw.

[6] W. H. Bailey et al., "Synopsis of IEEE Std C95.1™-2019 'IEEE standard for safety levels with respect to human exposure to electric, magnetic, and electromagnetic fields, 0 Hz to 300 GHz,"' *IEEE Access*, vol. 7, pp. 171346–171356, 2019.

[7] C. Kurnaz and M. Mutlu, "Comprehensive radiofrequency electromagnetic field measurements and assessments: A city center example," *Environ. Monit. Assess.*, vol. 192, pp. 1–14, May 2020.

[8] H. M. Madjar, "Human radio frequency exposure limits: An update of reference levels in Europe, USA, Canada, China, Japan and South Korea," in *Proc. Int. Symp. Electromagn. Compat.*, 2016, pp. 467–473.

[9] Y. A. Sambo, M. Al-Imari, F. Héliot, and M. A. Imran, "Electromagnetic emission-aware schedulers for the uplink of OFDM wireless communication systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1313–1323, Feb. 2017.

[10] H. Ibraiwish, A. Elzanaty, Y. H. Al-Badarneh, and M.-S. Alouini, "EMF-aware cellular networks in RIS-assisted environments," *IEEE Commun. Lett.*, vol. 26, no. 1, pp. 123–127, Jan. 2022.

[11] A. Zappone and M. Di Renzo, "Energy efficiency optimization of reconfigurable intelligent surfaces with electromagnetic field exposure constraints," *IEEE Signal Process. Lett.*, vol. 29, pp. 1447–1451, 2022.

[12] H. Jalilian, M. Eeftens, M. Ziaei, and M. Röösli, "Public exposure to radiofrequency electromagnetic fields in everyday microenvironments: An updated systematic review for Europe," *Environ. Res.*, vol. 176, Sep. 2019, Art. no. 108517.

[13] L. Chiaraviglio et al., "EMF exposure in 5G standalone mm-Wave deployments: What is the impact of downlink traffic?" *IEEE Open J. Commun. Soc.*, vol. 3, pp. 1445–1465, 2022.

[14] L. Chiaraviglio, S. Turco, G. Bianchi, and N. Blefari-Melazzi, "Do dense 5G networks increase exposure to electromagnetic fields? [Point of view]," *Proc. IEEE*, vol. 109, no. 12, pp. 1880–1887, Dec. 2021.

[15] Y. Kiouvrekis, A. Alexias, Y. Filipopoulos, V. Softa, C. D. Tyrakis, and C. Kappas, "Unsupervised machine learning and EMF radiation in schools: A study of 205 schools in Greece," 2020, *arXiv:2007.13208*.

[16] M. R. Bakcan et al., "Measurement and prediction of electromagnetic radiation exposure level in a university campus," *Tehnički Vjesnik*, vol. 29, no. 2, pp. 449–455, 2022.

[17] Z. Pala, "Examining EMF time series using prediction algorithms with R," *IEEE Can. J. Electr. Comput. Eng.*, vol. 44, no. 2, pp. 223–227, Jun. 2021.

[18] D. Colombi, P. Joshi, B. Xu, F. Ghasemifard, V. Narasaraju, and C. Törnevik, "Analysis of the actual power and EMF exposure from base stations in a commercial 5G network," *Appl. Sci.*, vol. 10, no. 15, p. 5280, 2020.

[19] S. Faye et al., "A survey on EMF-aware mobile network planning," *IEEE Access*, vol. 11, pp. 85927–85950, 2023.

[20] L. Chiaraviglio, C. Di Paolo, and N. Blefari-Melazzi, "5G network planning under service and EMF constraints: Formulation and solutions," *IEEE Trans. Mobile Comput.*, vol. 21, no. 9, pp. 3053–3070, Sep. 2022.

[21] B. Yin, W. Joseph, and M. Deruyck, "RIS-aided mmWave network planning towards connectivity enhancement and minimal electromagnetic field exposure," *IEEE Access*, vol. 11, pp. 115911–115923, 2023.
[22] P. Lara-Benítez, M. Carranza-García, and J. C. Riquelme, "An experimental review on deep learning architectures for time series forecasting," *Int. J. Neural Syst.*, vol. 31, no. 3, 2021, Art. no. 2130001.
[23] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–15.
[24] S. Li et al., "Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–14.
[25] Y. Li, C. Yang, and Y. Sun, "Dynamic time features expanding and extracting method for prediction model of sintering process quality index," *IEEE Trans. Ind. Informat.*, vol. 18, no. 3, pp. 1737–1745, Mar. 2021.
[26] S. Ahmed, I. E. Nielsen, A. Tripathi, S. Siddiqui, G. Rasool, and R. P. Ramachandran, "Transformers in time-series analysis: A tutorial," 2022, *arXiv:2205.01138*.
[27] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–9.
[28] J. Grigsby, Z. Wang, and Y. Qi, "Long-range transformers for dynamic spatiotemporal forecasting," 2021, *arXiv:2109.12218*.

CHI NGUYEN received the B.S. degree in electronics and telecommunications engineering from the University of Transport and Communications, Hanoi, Vietnam, in 2011, and the M.S. degree in telecommunication engineering from the Hanoi University of Science and Technology in 2016. She is currently pursuing the Ph.D. degree with the University of Ulster. She is currently with Queen's University Belfast, U.K. Her research interests include channel modeling, DL, non-orthogonal multiple access, reconfigurable intelligent surface, physical layer security, underwater communications, and visible light communications.

ADNAN AHMAD CHEEMA (Member, IEEE) received the B.Sc. degree from COMSATS University, Pakistan, in 2006, the M.Sc. degree from King's College London, U.K., in 2008, and the Ph.D. degree from Durham University, U.K., in 2015. From 2015 to 2017, he was a Postdoctoral Research Associate with Durham University and was involved in 5G channel measurements and modeling (sub-6 GHz and 24-90 GHz). In 2017, he joined Ulster University, U.K., as a Lecturer of Electronics Engineering and currently leading the SenComm Research Lab. His research interests include channel modeling, reconfigurable intelligent surface, non-orthogonal multiple access, and machine learning for wireless communications.

CETIN KURNAZ received the B.S., M.S., and Ph.D. degrees in electrical and electronics engineering from Ondokuz Mayis University, Samsun, Turkey, in 1999, 2002, and 2009, respectively, where he is an Associate Professor with the Department of Electrical and Electronics Engineering. His current research interests include wireless communication, antennas and wave propagation, and electromagnetic fields.

ARDAVAN RAHIMIAN (Member, IEEE) received the M.Eng. degree (Hons.) in electronic and communications engineering from the University of Birmingham, U.K., in 2009, and the Ph.D. degree in electronic engineering from the Queen Mary University of London, U.K., in 2018, where he was a Postdoctoral researcher with the School of EECS from 2018 to 2019, working on an industrial research project supported by Huawei Technologies. Since 2019, he has been a Lecturer (Assistant Professor) of Electronic Engineering with the School of Engineering, Ulster University, Belfast, U.K. He regularly serves as a reviewer for leading journals, flagship conferences, book proposals, and funding applications. His research interests include applied electromagnetics, communication systems, energy networks, and their multidisciplinary applications. He has authored or coauthored several papers in these areas. He was the winner of the Rohde & Schwarz Technology Prize in 2009.

CONOR BRENNAN (Senior Member, IEEE) received the B.A. degree (Mod) in mathematics and the Ph.D. degree from Dublin University (Trinity College), Ireland, in 1994 and 1998, respectively. In 2003, he joined the School of Electronic Engineering, Dublin City University, where he is currently an Associate Professor. His research interests are in computational electromagnetics, wireless propagation modeling, and channel modeling. He is an Associate Editor for the IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION and currently serves as the Irish representative on the Management Committee of COST Action CA20120 INTERACT and as a member of the European Association of Antennas and Propagation working group on propagation.

TRUNG Q. DUONG (Fellow, IEEE) is a Canada Excellence Research Chair and a Full Professor with the Memorial University of Newfoundland, Canada. He is also the adjunct Chair Professor of Telecommunications with Queen's University Belfast, U.K. His current research interests include quantum communications, wireless communications, signal processing, machine learning, and realtime optimization.

Dr. Duong received the Best Paper Award at the IEEE VTC-Spring 2013, IEEE ICC 2014, IEEE GLOBECOM 2016, 2019, and 2022, IEEE DSP 2017, IWCMC 2019 and 2023, and IEEE CAMAD 2023. He has received the two prestigious awards, including the Research Chair of the Royal Academy of Engineering from 2021 to 2025 and the Royal Academy of Engineering Research Fellowship from 2015 to 2020. He is the recipient of the prestigious Newton Prize 2017. He has served as an Editor/Guest Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE COMMUNICATIONS LETTERS, IEEE WIRELESS COMMUNICATIONS LETTERS, IEEE WIRELESS COMMUNICATIONS, *IEEE Communications Magazines*, and IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS.