Augmented-LSTM and 1D-CNN-LSTM Models for Linearization of Wideband Power Amplifiers

Ambagahawela Rathnayake[1]*, Lesthuruge Silva[1], Hossein Rezaei[1], and Nandana Rajatheva[1]
[1]Centre for Wireless Communications, Univeristy of Oulu, Oulu, Finland
{ambagahawela.rathnayake, lesthuruge.silva, hossein.rezaei, nandana.rajatheva}@oulu.fi,

***Abstract*—Long short-term memory(LSTM) neural networks and convolutional neural networks (CNNs) have been used by modern researchers to compensate power amplifier (PA) distortions. In this research, two digital predistortion (DPD) models based on LSTM and CNN called augmented-LSTM and 1D-CNN-LSTM are proposed. The augmented-LSTM model effectively reduces the distortions in wideband PAs. The measurement results show that the augmented-LSTM model gives better linearization performance compared to other state-of-the-art models designed based on neural networks (NNs). On the other hand, the 1D-CNN-LSTM model is proposed to simplify the augmented-LSTM model by integrating a CNN layer prior the LSTM layer causing reduction in the number of input features to the LSTM layer. The measurement results show that the 1D-CNN-LSTM model offers low-complexity linearization for wideband PAs and provides comparable results to the augmented-LSTM model.**

***Index Terms*—Augmented-LSTM, digital predistortion, DPD, 1D-CNN, LSTM, power amplifier, PA, neural network.**

I. Introduction

In wireless communication systems, power amplifiers (PAs) are an essential component as they provide the necessary power for transmitting signals [1], [2]. However, the nonlinear characteristics and memory effects of PAs can lead to a phenomenon known as spectrum regrowth [3]–[5], which increases the interference with adjacent channels, and degrades the quality of the transmitted signal. This problem is particularly pronounced in 5G wireless communication systems due to the increased bandwidth and complexity of the transmitted signals [2].

To address this issue, a technique called digital predistortion (DPD) [2], [6] can be used. DPD predicts the distortion introduced by the amplifier and generates a pre-distorted version of the input signal that cancels out the distortion when amplified. Traditional DPD models include the Volterra series [7] model and its simplified versions, namely as the memory polynomial (MP) model [8], the generalized memory polynomial (GMP) model [9], and the simplified Volterra (SV) model [10]. These models have been shown to be effective in addressing the nonlinearities of PAs for low bandwidth signals. However, due to the high correlation between polynomial bases [2], these models become less effective as the signal bandwidth increases.

Another popular approach used by modern researchers to model the nonlinearities of PAs is the use of artificial neural networks (ANNs). The authors in reference [6] constructed a shallow Neural Network (NN) based predistorter to compensate for PA nonlinearity [2]. Subsequently, the authors in reference [11] have proposed a deep NN-based DPD. In addition to these models, there have been many ANN-based DPD models that have been proposed over the time. Among the ANN-based models, long short-term memory (LSTM) [12], [13] networks,recurrent neural networks (RNNs) [14] and convolutional neural networks (CNNs) [1], [2], [15] have been widely used in modeling DPDs. However, models such as DNNs, RNNs, and LSTM networks have high complexity problems. While CNN models have been proposed by the researchers as a solution to the complexity problem, they have less capacity to capture memory features [13].

To address these limitations, a augmented-LSTM model is proposed to model the predistorter of the wideband PAs with higher accuracy and low complexity compared to other DNN models. The proposed model leverages the ability of LSTM layer to exploit the time series information of the PA data [12], [13], [16]. The main difference between the LSTM-DNN model proposed by the authors in [12] and the proposed augmented-LSTM model is the input features that are utilized to model the inverse model of PA. In the latter part of the article, a 1D-CNN and LSTM-based model is proposed to simplify the proposed augmented-LSTM model. It utilizes a 1D-CNN layer before the LSTM layer to reduce the number of input features to the LSTM layer by capturing the important features from the same time stamp.The LSTM layer then exploits the time-series information. This approach aims to combine the strengths of both techniques to achieve a efficient model for wideband PA predistortion with low complexity.

The article is organized as follows. Section II discuss about the existing neural network based dpd models. In section III, a description of the proposed augmented-LSTM model is provided. Section IV elaborates the 1D-CNN-LSTM structure, which is proposed to simplify the augmented-LSTM model . Section V covers the training process for the proposed models. Section VI is the extension of the proposed model to the DPD design. Section VII presents the measurement results and compares the proposed DPD model to existing ANN-based DPD models. Finally, the results will be summarized in section VIII under the conclusion.

978-1-6654-6483-3/23/$31.00 ©2023 IEEE

II. Existing NN-based digital predistortion models

Both shallow neural networks with fewer hidden layers [6] and deep neural networks [11] with multiple hidden layers have been used for digital predistortion in recent years. Even though the simple network structure and training process can be beneficial, with the less number of hidden layers in shallow neural network based models reduce the ability to capture the nonlinear effect of the power amplifiers compared to DNN-based models with multiple hidden layers. However, high complexity of DNN-based model is an issue when it comes to high bandwidth situations. LSTM [12], [13] networks and RNN networks [14] have been introduced to capture the time series information of the input signal. The LSTM-DNN [12] model consists of a LSTM layer and two fully connected layers. To reduce the complexity of this model the LSTM-CNN model [13] was proposed.

III. The Proposed augmented-LSTM Model

The main difference between the LSTM-DNN model, which the authors proposed in [12] and the proposed augmented-LSTM model is the input features that are utilized in modeling the DPD model. The LSTM-DNN model considers only the real and imaginary (I/Q) components for predistortion modeling. However, to enhance the modeling performances, the proposed augmented-LSTM model utilizes the amplitude of the input signal in addition to I/Q components as in [1], [6]. The proposed augmented-LSTM model consists of an input layer, a LSTM layer, a fully connected layer and an output layer as shown in in Fig.1. The input layer serves as the initial point of data processing, where the input signal is normalized to a zero mean and a unity standard deviation. As mentioned in the reference [17], this normalization step is crucial for mitigating the potential problem of gradient explosion, which could prevent the model from converging to an ideal solution. The input to the model is a collection of real and imaginary (I/Q) components and envelop-dependent terms as shown below.

$$X_{in}(n) = \begin{bmatrix} I_{in}(n) & I_{in}(n-1) & \dots & I_{in}(n-m) \\ Q_{in}(n) & Q_{in}(n-1) & \dots & Q_{in}(n-m) \\ |x_{in}(n)| & |x_{in}n-1)| & \dots & |x_{in}(n-m)| \\ |x_{in}(n)|^2 & |x_{in}(n-1)|^2 & \dots & |x_{in}(n-m)|^2 \\ |x_{in}(n)|^3 & |x_{in}(n-1)|^3 & \dots & |x_{in}(n-m)|^3 \end{bmatrix} \tag{1}$$

Where $x_{in}(n)$ represents the current input signal and $I_{in}(n)$ and $Q_{in}(n)$ are the real and imaginary components of complex signal $x_{in}(n)$ respectively. $|x_{in}(n)|$ represents the amplitude of signal $x_{in}(n)$. $I_{in}(n-k)$ and $Q_{in}(n-k)$, k = 1,2,. . .,m represents the real and imaginary components of delayed samples and $|x_{in}(n-k)|$, k = 1,2,. . .,m denotes the amplitudes of delayed samples. m denotes the memory depth. The input layer passes the processed data to the LSTM layer which consists of a collection of LSTM units. The special internal structure of the LSTM unit [12], [13], [17] which as depicted in Fig.2 designed to address the vanishing gradient problem that is frequently experienced in RNN. The LSTM unit is composed of several components, including the input gate (i_t), forget gate (f_t), memory cells, and output gate. The input gate determines how much weight to assign to new input and how much of the previous memory cell state to retain. The forget gate decides how much of the previous memory cell state to keep or discard at each time step. Meanwhile, the memory cells store relevant information over time, and the output gate determines how much of the memory cell state to output as the final result of the LSTM unit. By combining these components, the LSTM layer can selectively retain or discard information from previous time steps, enabling it to handle sequential data and model long-term dependencies more effectively than a standard RNN unit.

The LSTM layer generates a corresponding output sequence from the current and delayed data input sequence by extracting the time-series information [12], [13]. Hyperbolic tangent activation function, $Tanh$ is used as the activation function. The output generated by the LSTM layer is then passed to the fully connected layer. Here, the last timestamp of the sequence output from the LSTM layer is selected and pass to the fully connected layer. The fully connected layer, also known as the dense layer, is a standard neural network layer where each neuron receives input from every neuron in the previous layer and is connected to every neuron in the following layers. Then the activation function $Tanh$ is used to the output of the fully connected layer. Finally, the desired in-phase and quadrature values are produced by the output layer, which consists of two neurons corresponding to real and imaginary components of the output signal.

IV. The Proposed 1D-CNN-LSTM Model

To simplify the proposed augmented-LSTM model, a convolution layer is utilized before the LSTM layer as shown in Fig. 3 to reduce the input size of the LSTM layer. By using a less number of input features to the LSTM layer, the complexity of the model can be reduced. Hence the idea of this 1D-CNN-LSTM model is to reduce the complexity and to achieve comparable linearization results. The proposed 1D-CNN-LSTM model is composed of several key layers; input layer, 1D-CNN layer, LSTM layer, fully connected layer and output layer. The Input layer processes and normalize the input data and passes the normalized data to 1D convolution layer. As shown in Fig. 4, the CNN layer uses a $3 * 1$ convolution kernel to extract the useful features in the same memory effect.

$$\begin{aligned} g_k &= X_{in}(n) \circledast h_k \\ &= \begin{bmatrix} I_{in}(n) & I_{in}(n-1) & \dots & I_{in}(n-m) \\ Q_{in}(n) & Q_{in}(n-1) & \dots & Q_{in}(n-m) \\ |x_{in}(n)| & |x_{in}n-1)| & \dots & |x_{in}(n-m)| \\ |x_{in}(n)|^2 & |x_{in}(n-1)|^2 & \dots & |x_{in}(n-m)|^2 \\ |x_{in}(n)|^3 & |x_{in}(n-1)|^3 & \dots & |x_{in}(n-m)|^3 \end{bmatrix} \circledast h_k \end{aligned} \tag{2}$$

where g_k represents the 1D convolution output and h_k represents the coefficients of the 1D convolution kernel. $\circledast$ denotes the convolution operation.

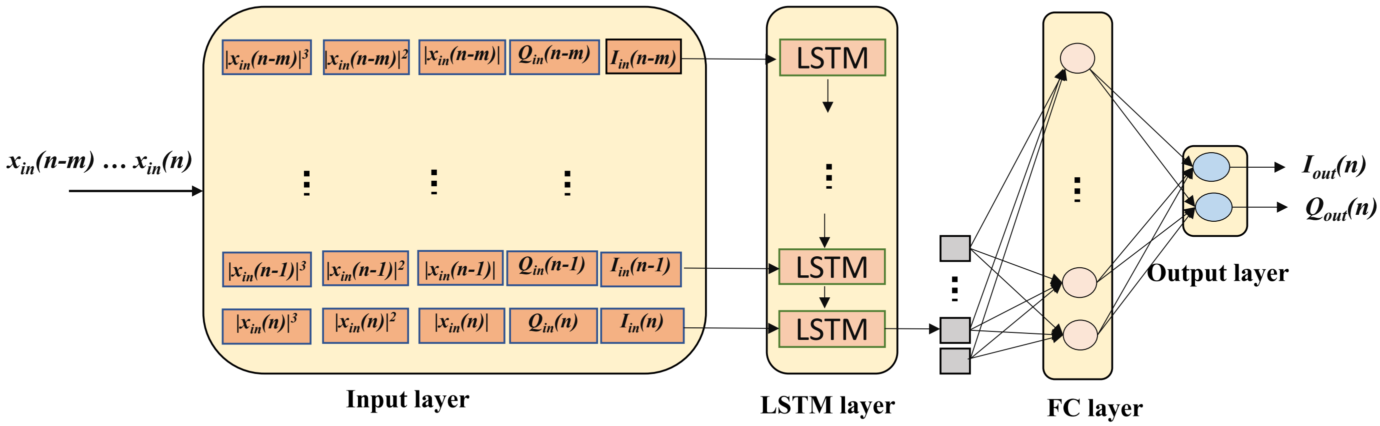

Fig. 1. Structure of the Proposed augmented-LSTM model

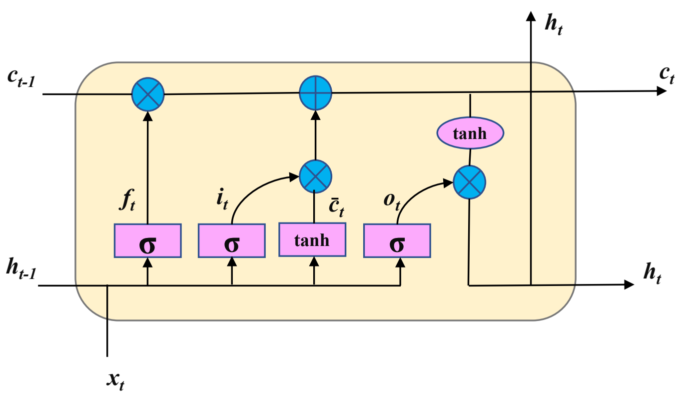

Fig. 2. Internal structure of LSTM unit

$$g_k = \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1(m+1)} \\ g_{21} & g_{22} & \cdots & g_{2(m+1)} \\ g_{31} & g_{32} & \cdots & g_{3(m+1)} \end{bmatrix} \quad (3)$$

where, g_{11} to $g_{3(m+1)}$ represent the convolution results as shown in Fig. 4. The output of the CNN layer is then passed to the LSTM layer to exploit time series information. The LSTM layer generates an output sequence by processing the input sequence. The last timestamp of this output sequence is then selected and passed to a fully connected layer. Finally, the output layer consisting of two neurons, provides the desired output related to I and Q components.

V. Training and Evaluation of the Proposed Models

Training and evaluation of the proposed algorithm were performed using the power amplifier measurements data set provided by the Mathworks, Inc. (www.mathworks.com) [18]. According to the reference [18], the data set was recorded using a NXP Airfast LDMOS Doherty power amplifier and the test signal employed was a 5G-like OFDM waveform, with each subcarrier carrying 16-QAM symbols. The specifications of the data set are listed in the Table I. In the training algorithm, Adam optimization algorithm [19] was employed as the optimizing function for parameter optimization of the model. Mean Squared Error (MSE) was utilized as the cost function to measure the error between actual and predicted values. To assess the performance of the training process and to prevent model from overfitting, a validation set was utilized. An early stopping criterion was implemented to further prevent overfitting, where the training process stops if the validation accuracy does not improve for a pre specified number of validations (validation patience). This ensures the generalizability of the model.

TABLE I
Characteristics of the Training and Testing Data

Parameter	Value
Operating frequency	3.6 - 3.8 GHz
Operating Bandwidth	100 MHz
Gain of the PA	29 dB

VI. Extension to DPD

In this work, the inverse modeling of PA's nonlinear function was performed to model the DPD. The indirect learning architecture [1] shown in Fig.5, a popular method for extracting DPD models was employed for this purpose. The inverse model was trained by using the output data of PA as the input data to the inverse model and input data of PA as the output data of the inverse model. Then using the trained DPD model, the main path DPD was updated to linearize the PA. Therefore, the input data matrix of PA's inverse model can be interpreted as below.

$$Y_n = \begin{bmatrix} I_{out}(n) & I_{out}(n-1) & \ldots & I_{out}(n-m) \\ Q_{out}(n) & Q_{out}(n-1) & \ldots & Q_{out}(n-m) \\ |y_{out}(n)| & |y_{out}(n-1)| & \ldots & |y_{out}(n-m)| \\ |y_{out}(n)|^2 & |y_{out}(n-1)|^2 & \ldots & |y_{out}(n-m)|^2 \\ |y_{out}(n)|^3 & |y_{out}(n-1)|^3 & \ldots & |y_{out}(n-m)|^3 \end{bmatrix} \quad (4)$$

Where $y_{out}(n)$ represents the output signal of the PA and $I_{out}(n)$ and $Q_{out}(n)$ are the real and imaginary components of complex signal $y_{out}(n)$.

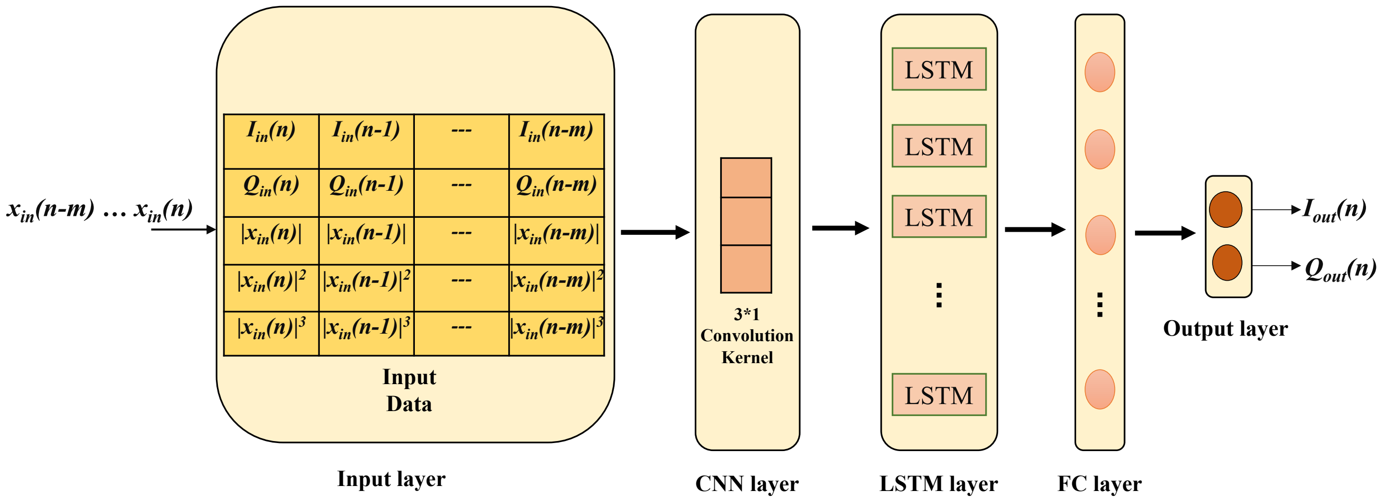

Fig. 3. Structure of the Proposed 1D-CNN-LSTM model

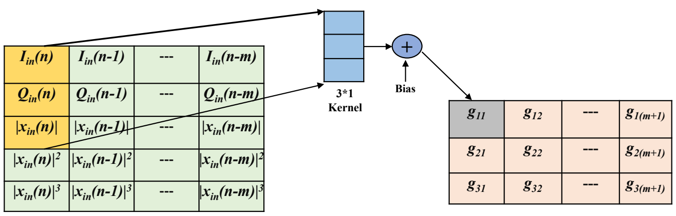

Fig. 4. Convolution diagram

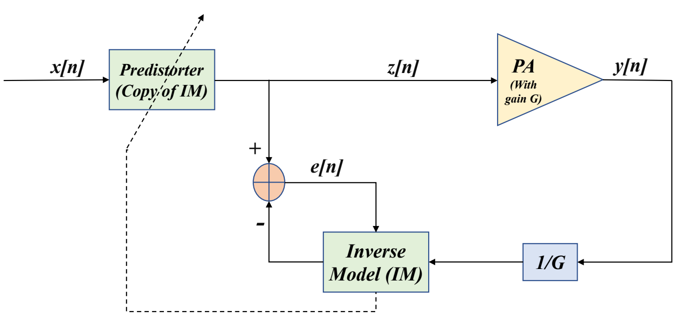

Fig. 5. Indirect Learning Architecture

VII. Measurement Results

To evaluate the proposed DPD model performances, the proposed models were built in PyTorch framework. The comparison of the Normalized Power Spectral Density (PSD) of the actual PA output signal and the PA output signal obtained with the proposed augmented-LSTM DPD model was done to evaluate the linearization performances of the DPD model. Fig. 6 displays normalized PSD graphs that demonstrate the linearization performances of the augmented-LSTM model. To further evaluate performance, Normalized Mean Squared Error(NMSE) and Adjacent Channel Power Ratio (ACPR) were calculated. The proposed DPD model enhances the NMSE from -22.17 dB to -33.49 dB and improves the ACPR performance from -26.69 dBc to -40.04 dBc (decibels relative to carrier). These results indicate that the proposed augmented-LSTM model has significant impact on reducing PA distortion.

To compare the linearization performance of augmented-

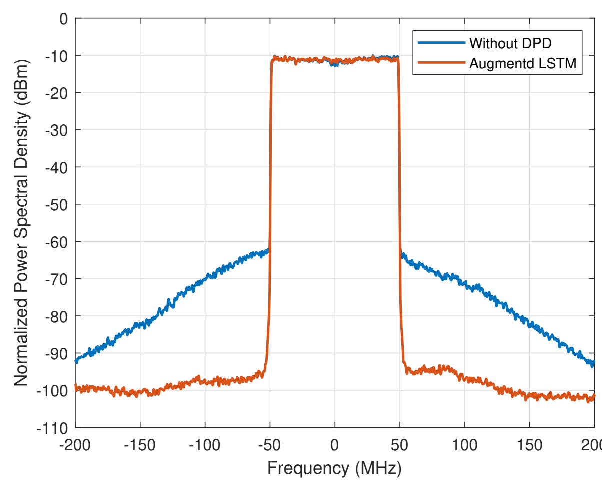

Fig. 6. Linearization performance of the augmented-LSTM model

LSTM model with other ANN models, DPD models including LSTM-DNN [12], DNN [11] and ARVTDNN [6] were also developed in the the PyTorch framework. As mentioned in the reference [12], number of neurons in the hidden layers of the LSTM-DNN network is [LSTM FC1 FC2] = [10 7 5]. Number of neurons in the hidden layer of ARVTDNN model is used as 17 according to the reference [6] and in the DNN model hidden layer structure is [17 17 17]. All the models were trained using the same algorithm used to train the augmented-LSTM model. Fig.7 compares the normalized PSD of the proposed augmented-LSTM model with other ANN DPD models and memory polynomial DPD. Table II further compares the linearization performances and complexity through the use of NMSE, ACPR and number of model coefficients. The results suggest that the proposed LSTM model has superior linearization performances compared to other ANN DPD models.

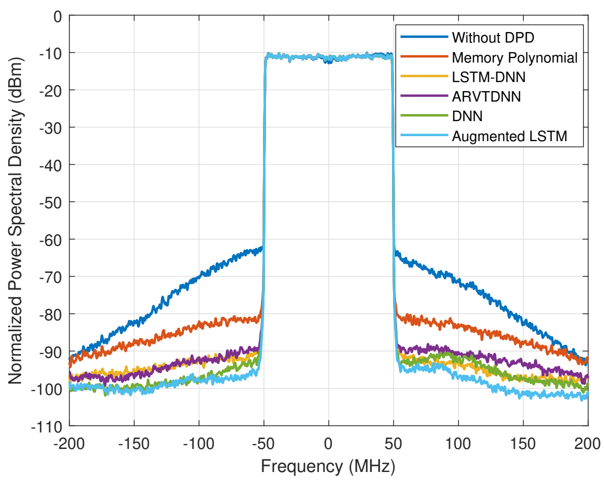

Fig. 7. Linearization performance of PA using various DPD

TABLE II
COMPARISON OF PERFORMANCES AND COMPLEXITY OF THE AUGMENTED-LSTM MODEL AND OTHER DPD MODELS

Model	**NMSE(dB)**	**ACPR (dBc) (-/+ 25 MHz)**	**Number of model coefficients**
Without DPD	-22.17	-26.19/-27.19	N/A
Memory Polynomial	-29.59	-34.76/-34.69	N/A
ARVTDNN [6]	-32.29	-38.54/-37.86	563
DNN [11]	-33.36	-40.15/-39.09	869
LSTM-DNN [12]	-32.79	-39.24/-38.45	689
Augmented-LSTM	-33.49	-41.01/-39.60	559

The 1D-CNN-LSTM DPD model, which was proposed to reduce the complexity achieved an NMSE value of -33.31 dB and improved the ACPR performance from -26.69 dBc to -39.79 dBc. Fig.8 displays the linearization performances of 1D-CNN-LSTM DPD model and the augmented-LSTM model. Table II compares the achieved NMSE, ACPR value and number of model coefficients for both the proposed 1D-CNN-LSTM model and augmented-LSTM model. These results indicates that the integrating the CNN layer before the LSTM layer maintains a comparable linearization results while reducing the complexity.

TABLE III
COMPARISON OF PERFORMANCES AND COMPLEXITY OF THE 1D-CNN-LSTM MODEL AND AUGMENTED-LSTM MODEL

Model	**NMSE(dB)**	**ACPR (dBc) (-/+ 25 MHz)**	**Number of model coefficients**
Without DPD	-22.17	-26.19/-27.19	N/A
Augmented-LSTM	-33.49	-41.01/-39.06	559
1D-CNN-LSTM	-33.31	-40.20/-39.37	466

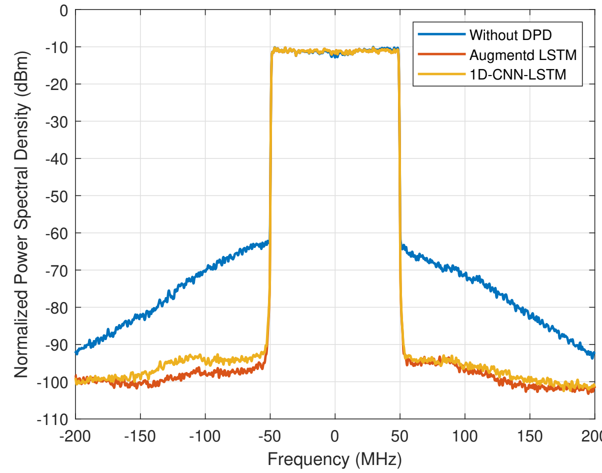

Fig. 8. Linearization performance of PA using 1D-CNN-LSTM DPD

VIII. CONCLUSION

There are two main outcomes in this paper. First, an augmented-LSTM model is proposed to model a DPD targeting wideband PAs. The proposed model uses a comprehensive basis set to enhance the modeling performance resulting in the NMSE improvement from -22.17 to -33.49 dB and the ACPR improvement from -26.69 to -40.04 dBc. Compared to other NN-based DPD models, the augmented-LSTM model shows a better linearization performance. The complexity of the augmented-LSTM model is then reduced using a 1D-CNN-LSTM model, where a CNN layer is employed before the LSTM layer to capture the useful features from the input data. This results in reducing the number of input features to the LSTM layer which translates to decreasing the complexity. The simulation results indicate that the less complex 1D-CNN-LSTM model achieves a comparable linearization performance as the augmented-LSTM model.

ACKNOWLEDGMENT

This research has been supported by the Academy of Finland, 6G Flagship program under Grant 346208.

REFERENCES

[1] Xin Hu, Zhijun Liu, Xiaofei Yu, Yulong Zhao, Wenhua Chen, Biao Hu, Xuekun Du, Xiang Li, Mohamed Helaoui, Weidong Wang, and Fadhel M. Ghannouchi, "Convolutional Neural Network for Behavioral Modeling and Predistortion of Wideband Power Amplifiers," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 8, pp. 3923–3937, 2022.

[2] Zhijun Liu, Xin Hu, Lexi Xu, Weidong Wang, and Fadhel M. Ghannouchi, "Low Computational Complexity Digital Predistortion Based on Convolutional Neural Network for Wideband Power Amplifiers," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 69, no. 3, pp. 1702–1706, 2022.

[3] Javier Reina-Tosina, Michel Allegue-Martínez, Carlos Crespo-Cadenas, Chao Yu, and Sergio Cruces, "Behavioral Modeling and Predistortion of Power Amplifiers Under Sparsity Hypothesis," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 2, pp. 745–753, 2015.

[4] Jessica Chani-Cahuana, Mustafa Özen, Christian Fager, and Thomas Eriksson, “Digital Predistortion Parameter Identification for RF Power Amplifiers Using Real-Valued Output Data,” *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 64, no. 10, pp. 1227–1231, 2017.

[5] Yue Li and Anding Zhu, “On-Demand Real-Time Optimizable Dynamic Model Sizing for Digital Predistortion of Broadband RF Power Amplifiers,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 68, no. 7, pp. 2891–2901, 2020.

[6] Dongming Wang, Mohsin Aziz, Mohamed Helaoui, and Fadhel M. Ghannouchi, “Augmented Real-Valued Time-Delay Neural Network for Compensation of Distortions and Impairments in Wireless Transmitters,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 1, pp. 242–254, 2019.

[7] R.J.P. de Figueiredo, “The Volterra and Wiener theories of nonlinear systems,” *Proceedings of the IEEE*, vol. 70, no. 3, pp. 316–317, 1982.

[8] Hyunchul Ku and J.S. Kenney, “Behavioral modeling of nonlinear RF power amplifiers considering memory effects,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, no. 12, pp. 2495–2504, 2003.

[9] D.R. Morgan, Z. Ma, J. Kim, M.G. Zierdt, and J. Pastalan, “A Generalized Memory Polynomial Model for Digital Predistortion of RF Power Amplifiers,” *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3852–3860, 2006.

[10] Tianjiao Du, Cuiping Yu, Jinchun Gao, Yuanan Liu, and Shulan Li, “A new accurate volterra-based model for behavioral modeling and digital predistortion of RF power amplifiers,” *Progress In Electromagnetics Research C*, vol. 29, pp. 205–218, 01 2012.

[11] Reina Hongyo, Yoshimasa Egashira, Thomas M. Hone, and Keiichi Yamaguchi, “Deep Neural Network-Based Digital Predistorter for Doherty Power Amplifiers,” *IEEE Microwave and Wireless Components Letters*, vol. 29, no. 2, pp. 146–148, 2019.

[12] Deepmala Phartiyal and Meenakshi Rawat, “LSTM-Deep Neural Networks based Predistortion Linearizer for High Power Amplifiers,” in *2019 National Conference on Communications (NCC)*, 2019, pp. 1–5.

[13] Wei Wang, Lu Sun, Haoming Liu, and Yibo Feng, “LSTM-CNN for Behavioral Modeling and Predistortion of 5G Power Amplifiers,” in *2022 IEEE 9th International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications (MAPE)*, 2022, pp. 28–32.

[14] Baitao Gong, Ziyang Feng, Cen Liu, Jun Wang, Chao Zhang, Changyong Pan, and Yonglin Xue, “A Memorized Recurrent Neural Network Design for Wide Bandwidth PA Linearization,” in *2022 International Conference on Electrical Engineering and Photonics (EExPolytech)*, 2022, pp. 114–117.

[15] Udara De Silva, Toshiaki Koike-Akino, Rui Ma, Ao Yamashita, and Hideyuki Nakamizo, “A Modular 1D-CNN Architecture for Real-time Digital Pre-distortion,” in *2022 IEEE Topical Conference on RF/Microwave Power Amplifiers for Radio and Wireless Applications (PAWR)*, 2022, pp. 79–81.

[16] Shijie Wang, MingQing Geng, Chao Yu, and Jialin Cai, “Improved behavioral modeling using augmented LSTM networks for ultra-broadband mmWave PA,” *Microwave and Optical Technology Letters*, 04 2022.

[17] Jinlong Sun, Wenjuan Shi, Zhutian Yang, Jie Yang, and Guan Gui, “Behavioral Modeling and Linearization of Wideband RF Power Amplifiers Using BiLSTM Networks for 5G Wireless Systems,” *IEEE Transactions on Vehicular Technology*, vol. 68, no. 11, pp. 10348–10356, 2019.

[18] Matworks Inc, “Data set provided by the MathWorks, Inc., MATLAB version 2022a,” http://www.mathworks.com, [Accessed on 01-Sep-2022].

[19] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” https://arxiv.org/abs/1412.6980v9, 2015.