Radar Emitter Classification With Attention-Based Multi-RNNs

Xueqiong Li, Zhangmeng Liu, Zhitao Huang, and Weisong Liu

***Abstract*—Analyzing and recognizing radar signals are important tasks for effective Electronic Support Measurement (ESM) system operation. The electromagnetic environment is highly complex nowadays, however, resulting in non-uniformed distributed pulse streams. The high-dimensional features of the radar emitters are also overly complicated. Isolating useful information of the pulse streams and removing noise can assist in the emitter classification process. This letter proposes an attention-based approach for radar emitter classification using recurrent neural networks (RNNs). Several RNNs assigned to individual features exploit the intrinsic patterns of the radar pulse streams via supervised learning; the learned patterns are then used to identify patterns of interest in the test pulse streams and place them into different categories. The attention mechanism demonstrates effective treatment of high missing and spurious pulse ratios, especially in cases of multiple consecutive missing pulses and multifunctional radar pulses. Simulation results also show that the proposed model outperforms other state-of-the-art neural network structures.**

***Index Terms*—Radar emitter classification, pulse streams, Electronic Support Measurement (ESM), recurrent neural networks (RNNs), attention mechanism.**

I. INTRODUCTION

RADAR emitter classification serves to categorize various radar emitters from received radar pulse samples [1], which is crucial for effective Electronic Support Measurement (ESM) system operation. The correct identification is necessary for further platform identification, target tracking, missile guidance, and other processes [2]. Complex radar features and high signal density make the ESM environment very complicated, however, which makes the radar emitter classification of pulse streams a difficult task [3].

Traditional methods for categorizing pulses are based on the pulse description words (PDWs), which consist of several statistical features such as pulse width (PW), time-of-arrival (TOA), and angle-of-arrival [4]. The pulse repetitive interval (PRI), which reflects the first-order difference in TOA, characterizes the temporal regularity of pulse streams. Most researchers utilize these statistical features for emitter classification. Traditional signal recognition techniques rely on complex algorithms, which are computationally intensive and require manual validation [5]. Machine learning methods have been introduced for radar emitter classification to remedy this. Zhang *et al.* [6], for example, used support vector machines (SVMs) as a hierarchical classifier structure to recognize radar emitter signals. Ren *et al.* [7] used fuzzy support vector machines (FSVMs) for signal classification based on mutual information. Convolutional neural networks (CNNs) have also been used as an encoding technique to manage inconsistent features [8]. Petrov *et al.* [9] used a feed-forward network to classify three different types of radar signals; other have used multilayer perceptrons (MLPs) [10] for emitter identification with interval-value data. Recurrent neural networks (RNNs) have been introduced for the classification of the pulse streams with PW and PRI [11].

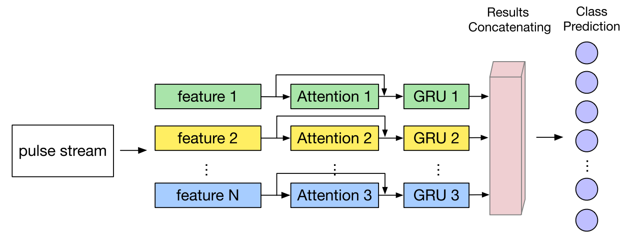


Fig. 1. Proposed network structure.

The above approaches fail in several scenarios that do frequently emerge in electronic warfare (EW) environments. For instance, when the ratios of missing and spurious pulses are very high – and especially when there are multiple consecutive missing pulses – these approaches are inapplicable. They also fail when the differences among PDWs of the emitters are small, e.g., when the PRI values of different emitters are very close. Multifunctional radars may work in the same mode with the same PDWs, while other times they work in different modes with different respective PDWs. This also renders traditional methods ineffective.

Based on the scenarios mentioned above, we introduce the attention mechanism to the ESM domain as a novel method for efficiently processing pulse streams. The Attention mechanisms can reveal unobvious differences among emitters in noisy environments. It has been applied in neural networks for some time, especially in image recognition [12] and Natural Language Processing (NLP) [13], [14]. The basic principle of attention-mechanism-based noise reduction is to isolate parts of the scene without noise. This network requires an attention mechanism to function properly, as noise cannot be located in ordinary neural networks; the entire input is treated equally, resulting in poor performance. Attention mechanisms have been tested in a variety of noisy environments and are known to perform well [15].

In this study, we developed an attention-based multi-RNN scheme for radar emitter classification (Fig. 1). We leveraged the Gated Recurrent Unit (GRU) for each feature of pulse streams to obtain an intermediate representation useful for

Manuscript received March 4, 2020; revised April 12, 2020; accepted May 17, 2020. Date of publication May 20, 2020; date of current version September 12, 2020. The research is supported by the Program for Innovative Research Groups of the Hunan Provincial Natural Science Foundation of China. The associate editor coordinating the review of this letter and approving it for publication was M. Oner. *(Corresponding author: Xueqiong Li.)*

The authors are with the Department of Electronic and Science, National University of Defense Technology, Changsha 410073, China (e-mail: lixueqiong13@nudt.edu.cn).

Digital Object Identifier 10.1109/LCOMM.2020.2995842

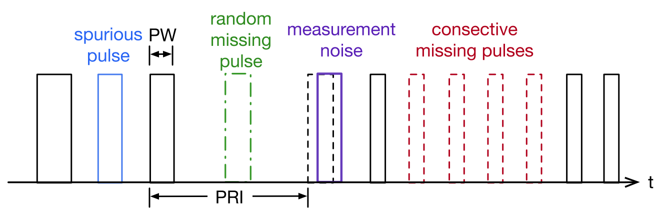


Fig. 2. Noise in EW environment.

classification, then deployed an attention mechanism to resolve complex noise problems. Simulations were run based on three common scenarios to compare the proposed model to extant, state-of-the-art deep learning approaches. Besides, the pulses of the target radar represent 50% or more of all the pulses of the stream in the simulations.

II. Problem Formulation

Sequential patterns are intrinsic characteristics that distinguish pulse streams from random noise. These patterns contain statistical features of each pulse such as the PW, frequency, and TOA. The PRI is a feature that indicates how pulses emerge along the time axis, which is a major feature of the classification problem [11], [16].

In this study, we exploited the PRI as the main feature to categorize pulse streams. We took the PW into consideration as well to show how other features can be jointly used with the PRI dor classification. The pulse streams can be described as $P = [p_1, p_2, \ldots, p_i, \ldots, p_L]$, where L is the number of the intercepted pulses and each pulse is of the form $p_i = (pri_t, pw_t)$. pri_t and pw_t represent the tth PRI and PW value in the pulse stream. Each pulse stream P is associated with a class label $c \in [0, C]$, where C is the number of the classes. Finally, the pulse streams and the classes form a dataset $(P_1, c_1), (P_2, c_2), \ldots, (P_i, c_i), \ldots, (P_N, c_N)$ of N samples that are categorized into C classes. The purpose of the classifier $f(\cdot)$ with the output prediction $y = f(P)$ is to minimize the loss between the output y and the truth class label c, so that $y = c$ for as many (P_i, c_i) as possible.

We ran simulations with three types of noise based on the scenarios mentioned in Section I, as shown in Fig. 2. The rectangles in the figure are individual pulses received by the radar; solid rectangles are the pulses being detected and dotted rectangles are the missing pulses which are not detected. The width of the rectangle is the PW and the interval between two successive pulses is the PRI.

A. Spurious Pulses

Spurious pulses are irrelevant pulses which are incorrectly detected by the receiver. They are indicated by blue solid rectangles on the figure. Spurious pulses are inevitable due to imperfections in the deinterleaving process before the classification (or pulse splitting).

B. Missing Pulses

Missing pulses occur when the receiver fails to detect the pulses or the pulses randomly overlap with other noise. They are represented by green dotted rectangles in the figure. An uneven pulse stream distribution may also emerge when the radar is not able to fully observe the objects or the antenna is facing away from the radiation sources. Several consecutive pulses are missing in this case, which makes the analysis and identification progress very challenging. Missing pulses are marked by red dotted rectangles in the figure.

C. Measurement Errors

Measurement errors are fairly common as the devices are not very precise. This error is considered as Gaussian noise in each pulse stream, and is marked by purple rectangles in the figure.

Noise in the actual electronic warfare environment makes classification a challenging endeavor as it can destroy the PRI patterns and break the regularity of the pulse streams. Traditional methods typically operate on simulation data with very little noise and thus cannot be generalized to real-world environments. In this study, we developed an attention-based multi-RNNs model to solve radar emitter classification problems. The RNN is suitable for processing time-series data with variable length. The proposed model can effectively manage joint PDW features, as each RNN cell is utilized for one feature of the pulse stream. The attention mechanism, as discussed above, also readily resolves noisy-environment-related problems.

III. Attention-Based Multi-RNNs Structure for Radar Emitter Classification

The proposed attention-based multi-RNNs classification scheme is presented in this section. The PRI and PW vectors are input into the model, then output predictions are supervised by a loss function and back-propagated to train the network via stochastic gradient descent method.

A. Network Architectures

Each feature channel has an individual RNN network for pattern extraction while temporal dependencies are maintained (Fig. 1). The GRU [17] manages long pulse sequences as an alternative to the vanilla RNN; it shows better performance and lower computation consumption on long-term data [18]. The pulse stream $P = [p_1, p_2, \ldots, p_i, \ldots, p_L]$ is taken as the input vector $\mathbf{x}_n$ for the GRU. The detailed procedures are shown in Eqs. (1)-(4).

$$\mathbf{z}_n = \sigma(\mathbf{W}^{(u)}\mathbf{x}_n + \mathbf{U}^{(u)}\mathbf{h}_{n-1} + \mathbf{b}^{(u)}), \tag{1}$$

$$\mathbf{r}_n = \sigma(\mathbf{W}^{(r)}\mathbf{x}_n + \mathbf{U}^{(r)}\mathbf{h}_{n-1} + \mathbf{b}^{(r)}), \tag{2}$$

$$\mathbf{f}_n = tanh(\mathbf{W}\mathbf{x}_n + \mathbf{r}_n \odot \mathbf{U}\mathbf{h}_{n-1} + \mathbf{b}), \tag{3}$$

$$\mathbf{h}_n = \mathbf{z}_n \odot \mathbf{f}_n + (\mathbf{1} - \mathbf{z}_n) \odot \mathbf{h}_{n-1}. \tag{4}$$

There are four main vectors in the GRU structure: the update gate $\mathbf{z}_n$, the reset gate $\mathbf{r}_n$, the new memory $\mathbf{f}_n$, and the hidden state vector $\mathbf{h}_n$ at time instant n. Other parameters include the input $\mathbf{x}_n$, the bias vector $\mathbf{b}$, and the weight matrices $\mathbf{W}$ and $\mathbf{U}$. The update gate $\mathbf{z}_n$ decides how much of $\mathbf{h}_{n-1}$ will pass to the next state. The reset gate identifies how important $\mathbf{h}_{n-1}$ is to the new memory $\mathbf{f}_n$. The memory vectors $\mathbf{f}_n$ are also obtained by $\mathbf{W}$, $\mathbf{U}$ and $\mathbf{b}$ but via a hyperbolic tangent function $tanh(\cdot)$. The hidden states of GRU $\mathbf{h}_n$ are finally updated by the other three vectors. $\mathbf{1}$ is an all-one vector in the equation, and $\odot$ represents an element-wise multiplication.

There is a fully connected layer following the GRU structure that provides the probability distribution over the PRI class. Assuming that the K PRI modulation mode is to be recognized, the probability of the input PRI sequences is $\hat{\mathbf{p}} = s(\mathbf{W}^{(o)}\mathbf{h}_n + \mathbf{b}^{(o)})$, where $\mathbf{W}^{(o)}$ are weight matrices, $\mathbf{b}^{(o)}$ are bias vectors, and $s(\cdot)$ is the softmax function.

B. Attention Mechanism Principle

Only the true pulses of the entire pulse stream are useful in solving the classification problem. Noise or pulses with measurement errors not conducive to this task. However, the GRU lacks the ability to adaptively focus on certain areas or locations; it may result in redundancy or lost information during the learning process. In this study, we introduced an attention mechanism to focus the GRU on the real pulses that contribute to the classification. The PW is a statistical feature and has independent values for each pulse. The PRI value has an intrinsic pattern with several consecutive pulses. We only apply the attention mechanism to the RNN for the PRI here.

The attention mechanism, in neural networks context, is based on the human attentional visual mechanism [19]. In the classification task, it provides large weight values for certain pulses of the entire pulse stream and ignores pulses with incorrect values [20].

The attention mechanism links to the related parts of the PRI values and assigns a higher weight to the corresponding pulses of the input $\mathbf{x}_n$ [21]. The attention gate receives the sequence information $\mathbf{x_n}$ and hidden state at the last moment $\mathbf{h_{n-1}}$ to learn a weight matrix $\mathbf{A}_n$ that can express the importance of this information.

$$\alpha_n = \tanh(\mathbf{h}_{n-1}, \mathbf{x}_n). \tag{5}$$

$$\mathbf{A}_n = \frac{\exp(\alpha_n)}{\sum_n \exp(\alpha_n)}, \tag{6}$$

$$\hat{\mathbf{h}}_n = \tanh\left(\mathbf{A}_n \left[\mathbf{x}_n, \mathbf{h}_{n-1}\right]\right), \tag{7}$$

where $\mathbf{x}_t$ is the input vector at time n and $\mathbf{h}_{n-1}$ is the hidden state at the last moment. This attention mechanism can be considered to construct a fixed length of the input value $\hat{\mathbf{h}}_n$ by calculating an adaptive weighted average of the state sequences $\mathbf{h}_{n-1}$ and the input $\mathbf{x}_n$.

A constant PRI example was operated to observe the effects of the attention mechanism. The PRI sequences were imported into the proposed attention-based RNNs structure for training. Once the RNNs converged, a pulse stream of constant PRI value with several spurious pulses and their respective attention weights α_i were drawn in Fig. 3. The value of the attention weight is between $[-1, 1]$, so we took the absolute value of the attention weight and multiplied it by a proper positive number to make the regularity more obvious.

The blue line in Fig. 3 is a curve of a constant PRI sequence containing several additional spurious pulses (falling portions of the curve). The red line represents the corresponding attention weights for this PRI sequence. A larger attention weight indicates greater contribution of the corresponding PRI value to the recognition results. Since spurious pulses are not conducive to the recognition, the corresponding attention weights should be small. Conversely, the attention weights of the normal PRI value at 500 should be large. In Fig. 3, the falling parts of the red line exactly match the falling parts of the blue line.

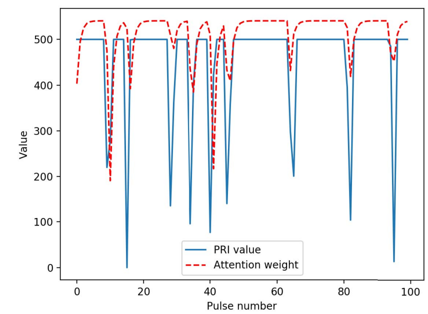


Fig. 3. PRI values and corresponding attention weights.

C. Training the Proposed Model

The attention-based multi-RNN model can be trained to categorize pulse streams from radar emitters. The entire structure consists of GRUs, attention mechanisms, and a classifier. The detailed parameter settings we used to train and validate samples with these layers are described below.

- GRU. There are two GRUs for two features of the pulse streams. Each GRU has three layers of 256 hidden size.
- Attention mechanism. We only applied an attention mechanism on the GRU for the PRI in the study. The input and the output sizes of the attention layer are both the number of the pulses.
- Classifier. A fully connected layer with a sigmoid function was added to the above layers as a classifier.

The loss function in the proposed model follows the cross entropy criterion [22] to reveal the difference between the predicted label and the ground truth:

$$\text{loss} = -\sum_{k=1}^{K} \left[c_k \log\left(y_k\right) + (1 - c_k)\log(1 - y_k)\right], \tag{8}$$

where c_k is the kth class of the truth label and y_k is the predicted class of the pulse stream.

IV. Simulations

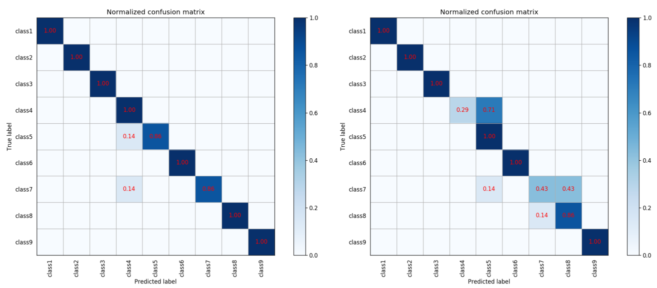
A. Dataset and Setup

We simulated nine classes of pulse streams with different PW and PRI to test the performance of the proposed model in the environments discussed above. Detailed feature values are listed in Table I.

We simulated $5,000$ pulse samples for each class: $4,500$ for training and 500 are for validation, with around 100-150 pulses in one sample. Test set are simulated with different levels of error ratios.

TABLE I
SIMULATED PULSE STREAM PARAMETERS

No.	PW Mean (μs)	PRI Type	PRI Mean (μs)
class1	3	constant	300
class2	3	constant	320
class3	2	constant	300
class4	3	stagger	[100, 300, 500, 700]
class5	3	stagger	[100, 320, 500, 700]
class6	2	stagger	[100, 300, 500, 700]
class7	3	mixed	300*n1, 320*n2
class8	3	mixed	300*n1, [100, 300, 500, 700]*n2
class9	2	mixed	300*n1, 320*n2



(a) (b)

Fig. 4. Confusion matrix of (a) proposed model; (b) model without attention and 70% lost pulses, 50% spurious pulses, and up to 10% measurement error.

Three types of noise were considered in the simulation. The ratios of spurious and missing pulses were set up to 50% and 70% and to randomly appear in the entire pulse stream set. We set several consecutive missing pulses (up to 90% of the entire stream) in $class1$ to $class3$ to test how the proposed model performs an uneven observation of pulse streams. Additive Gaussian noise was applied in increasing quantities to the pulse sequences reaching up to 10% of the signal magnitude. These noise settings are consistent with the noise conditions in most real EW environments [3].

The proposed model was trained on the Pytorch [23] platform with a batch size of 64 and a learning rate of $\eta = 0.001$. Twenty epochs is typically enough for the model to converge.

B. Contribution of Attention Mechanisms

The first simulation was run to assess the proposed model's classification performance for pulse streams with the nine classes mentioned in Table I. Confusion matrices (Fig. 4) were also drawn to show the recognition accuracy per class and the main reasons for ambiguity having increased the difficulty of the classification problem. The first confusion matrix belongs to the proposed model; the matrix shown in Fig. 4(b) is of the same model without the attention mechanism. The attention mechanism appears to markedly enhance the model's performance under noisy conditions.

The confusion matrix shows where most prediction categories in our case were located on a diagonal overlapping the ground truth categories. However, when the differences between $class4$ and $class5$, $class7$ and $class8$, and $class4$ and $class7$ were relatively small, they grew more confusing. When the PWs changed while the PRIs remained the same, the proposed model still accurately identified them. The multi-RNN structure is suitable, to this effect, for real-world classification problem.

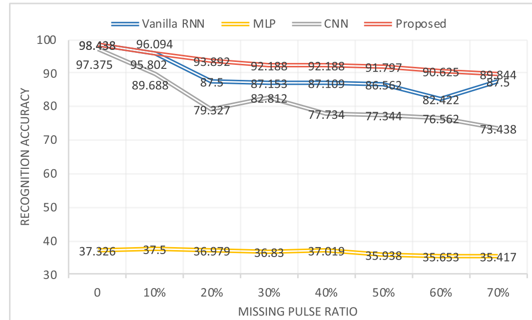


Fig. 5. Performance for various techniques with different missing pulses ratios.

C. Comparison of Different Models

We compared the performance of the proposed model with different DL-based emitter classification techniques: the single vanilla RNN model proposed by Liu and Philip [11], which we gave the same hyper parameters as the proposed RNN structure but without an attention mechanism, an MLP [9], and a CNN model as a baseline model, where the PRI and PW were imported as a 2D matrix. The CNN is effective for image classification problems on 2D inputs. We used the same dataset throughout this comparison to ensure an objective baseline. The results are shown in Fig. 5.

We found that the proposed model outperforms the other baseline DL methods we tested in noisy environments. The MLP did not resolve the classification problem under the complex simulation conditions. As the proportion of lost pulses increased, the proposed model maintained relatively stable recognition accuracy while the other DL methods grew less effective.

V. CONCLUSIONS

In this study, we developed an attention-based multi-RNNs model for radar emitter classification. This model functions well in noisy environments and on complex pulse stream patterns. It maintains this strong performance even when differences among the emitters are small. Simulation results validated the effectiveness of our method in terms of accuracy over other state-of-the-art DL techniques. The proposed model can be used for radar emitter classification with complex radar patterns and in complicated environments.

REFERENCES

[1] J. Liu, J. P. Y. Lee, L. Li, Z.-Q. Luo, and K. M. Wong, "Online clustering algorithms for radar emitter classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1185–1196, Aug. 2005.
[2] A. E. Spezio, "Electronic warfare systems," *IEEE Trans. Microw. Theory Techn.*, vol. 50, no. 3, pp. 633–644, Mar. 2002.
[3] D. Adamy, *EW 101: A First Course in Electronic Warfare*, vol. 101. Norwood, MA, USA: Artech House, 2001.
[4] R. G. Wiley, *ELINT: The Interception Analysis Radar Signals*. Norwood, MA, USA: Artech House, 2006.

[5] G. B. Willson, "Radar classification using a neural network," *Proc. SPIE Appl. Artif. Neural Netw.*, vol. 1294, pp. 200–210, Aug. 1990.

[6] G. Zhang, W. Jin, and L. Hu, "Radar emitter signal recognition based on support vector machines," in *Proc. 8th Control, Autom., Robot. Vis. Conf.*, vol. 2, 2004, pp. 826–831.

[7] M. Ren, J. Cai, Y. Zhu, and M. He, "Radar emitter signal classification based on mutual information and fuzzy support vector machines," in *Proc. 9th Int. Conf. Signal Process.*, Oct. 2008, pp. 1641–1646.

[8] J. Sun, G. Xu, W. Ren, and Z. Yan, "Radar emitter classification based on unidimensional convolutional neural network," *IET Radar, Sonar Navigat.*, vol. 12, no. 8, pp. 862–867, Aug. 2018.

[9] N. Petrov, I. Jordanov, and J. Roe, "Radar emitter signals recognition and classification with feedforward networks," *Procedia Comput. Sci.*, vol. 22, pp. 1192–1200, Dec. 2013.

[10] C.-S. Shieh and C.-T. Lin, "A vector neural network for emitter identification," *IEEE Trans. Antennas Propag.*, vol. 50, no. 8, pp. 1120–1127, Aug. 2002.

[11] Z.-M. Liu and P. S. Yu, "Classification, denoising, and deinterleaving of pulse streams with recurrent neural networks," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 4, pp. 1624–1639, Aug. 2019.

[12] K. Xu *et al.*, "Show, attend and tell: Neural image caption generation with visual attention," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 2048–2057.

[13] A. M. Rush, S. Chopra, and J. Weston, "A neural attention model for abstractive sentence summarization," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Lisbon, Portugal, Sep. 2015, pp. 379–389.

[14] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2016, pp. 1480–1489.

[15] B. Zhao, H. Lu, S. Chen, J. Liu, and D. Wu, "Convolutional neural networks for time series classification," *J. Syst. Eng. Electron.*, vol. 28, no. 1, pp. 162–169, Feb. 2017.

[16] H. K. Mardia, "New techniques for the deinterleaving of repetitive sequences," *IEE Proc. F Radar Signal Process.*, vol. 136, no. 4, pp. 149–154, 1989.

[17] R. Zhao, D. Wang, R. Yan, K. Mao, F. Shen, and J. Wang, "Machine health monitoring using local feature-based gated recurrent unit networks," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1539–1548, Feb. 2018.

[18] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, *arXiv:1409.0473*. [Online]. Available: http://arxiv.org/abs/1409.0473

[19] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 3104–3112.

[20] F. Ma, R. Chitta, J. Zhou, Q. You, T. Sun, and J. Gao, "Dipole: Diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2017, pp. 1903–1911.

[21] S. Seo, J. Huang, H. Yang, and Y. Liu, "Interpretable convolutional neural networks with dual local and global attention for review rating prediction," in *Proc. 11th ACM Conf. Recommender Syst.*, Aug. 2017, pp. 297–305.

[22] J. Wu, Z. Wang, and B. Vojcic, "Partial iterative decoding for binary turbo codes via cross-entropy based bit selection," *IEEE Trans. Commun.*, vol. 57, no. 11, pp. 3298–3306, Nov. 2009.

[23] N. Ketkar, "Introduction to PyTorch," in *Deep Learning With PyTorch*. Berkeley, CA, USA: Apress, 2017.