

A Link Quality Prediction Method Based on External Attention and Prior Probability

Meng Xie, Weibin Shi, Yulai Lie, Wenfeng Xu

School of Optical-Electrical and Computer Engineering, University of Shanghai for Science & Technology, Shanghai, China
Email: 2399492587@qq.com

How to cite this paper: Xie, M., Shi, W.B., Lie, Y.L. and Xu W.F. (2026) A Link Quality Prediction Method Based on External Attention and Prior Probability. *Open Journal of Applied Sciences*, **16**, 1103-1116. <https://doi.org/10.4236/ojapps.2026.164065>

Received: March 11, 2026

Accepted: April 7, 2026

Published: April 10, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). <http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Link quality estimation is a critical foundation for path selection in routing protocols of wireless sensor networks. Affected by multipath fading, noise, and interference in wireless channels, wireless links typically exhibit nonlinear and non-stationary characteristics, which pose challenges to efficient and accurate link quality prediction. To address the issues of error accumulation and lack of parallel computing in existing autoregressive methods, a Transformer-based link quality prediction method named LEAPP is proposed. A multi-head external attention mechanism is introduced in the encoder to reduce computational complexity and enhance global modeling capability. The decoder uses packet success rate (PSR) as input and constructs a non-autoregressive prediction model, achieving effective integration of prior probability and deep learning models. Experimental results show that compared with baseline models, the MAE and RMSE of LEAPP are reduced by 22.1% and 16.3%, respectively, and the MAE drops to 0.0092 on the public dataset. Meanwhile, the non-autoregressive inference mode reduces the inference delay by approximately 82% compared with traditional methods, significantly improving the real-time performance and practicality of online link quality prediction.

Keywords

Transformer, Link Quality Prediction, External Attention, Prior Probability

1. Introduction

Wireless Sensor Network (WSN) are an important component of the IoT [1] [2]. Due to the adoption of low-power communication technologies in WSN, they are susceptible to factors such as multipath fading, environmental interference, and noise, resulting in unstable link states. Link quality exhibits non-stationary, asymmetric, and irregular fluctuation characteristics in both time and

spatial domains [3]. How to accurately predict link quality is a key problem that needs to be solved.

Early studies mostly adopted theoretical or empirical model-driven link quality estimation methods, which evaluate link quality by calculating predefined variables. Such methods are often only applicable to specific scenarios and difficult to generalize to complex and variable real-world environments [4]. In recent years, data-driven methods have gradually become mainstream and can be subdivided into three categories based on the technologies adopted: statistical models, machine learning models, and deep learning models. Statistical models realize link quality estimation by establishing the mapping relationship between link metrics and Packet Reception Ratio (PRR), which have the advantages of low computational overhead and easy deployment. However, they are highly dependent on measurement data collected under specific scenarios. Once the node deployment environment changes, the prediction performance of the original mapping model may degrade significantly. Machine learning methods improve the robustness and accuracy of models to a certain extent by fusing multi-dimensional link indicators and introducing fuzzy logic, regression, or classification models. Nevertheless, their performance is still limited by manual features, resulting in poor adaptability to dynamic channel environments and a bottleneck in prediction accuracy. Deep learning methods can automatically learn complex link features from raw data, significantly improving the prediction accuracy of PRR. However, their application in WSN faces multiple challenges: first, the model training and inference processes are computationally intensive and energy-consuming, making it difficult to meet the resource constraints of low-power devices; second, it is hard to obtain complete and high-quality training samples in dynamic channel environments; third, the model decision-making process lacks interpretability, which is not conducive to system debugging and trusted deployment. These factors collectively restrict the practicalization of deep learning in highly dynamic and resource-constrained WSN.

To address the limitations of the aforementioned existing methods, this paper proposes a link quality prediction method based on Transformer [5], named LEAPP (Link Estimation based on External Attention and Prior Possibility). In terms of model structure design, LEAPP introduces the Multi-Head External Attention (MHEA) [6] mechanism in the encoder, and innovatively adopts the physical layer parameter PSR as the input in the decoder to construct a non-autoregressive prediction mechanism. By effectively fusing the prior knowledge of theoretical models with the data-driven capabilities of deep learning, LEAPP maintains high prediction accuracy while improving the generalization performance of the model across different deployment scenarios, providing an efficient and scalable link quality estimation solution for resource-constrained and highly dynamic wireless sensor networks.

The main contributions of this paper are as follows:

- 1) A link quality prediction model based on the Transformer framework is pro-

posed. Different from structures such as LSTM that rely on local temporal modeling, LEAPP enhances the global dependency modeling capability by introducing the multi-head external attention mechanism and utilizing externally learnable parameters in the encoder, while reducing computational complexity.

2) A non-autoregressive decoder with PSR as input is designed. Combined with the attention mechanism, it effectively solves the problem of gradual error accumulation during the inference phase, achieving efficient parallel inference while improving prediction accuracy. The model can still converge quickly in small-sample scenarios, featuring good data efficiency and interpretability.

3) Comprehensive comparative experiments, ablation studies, and sensitivity analyses are conducted based on multiple self-collected and public WSN datasets. The results show that LEAPP significantly outperforms existing methods in complex and dynamic environments, demonstrating outstanding advantages in prediction accuracy, stability, and generalization ability. This verifies its application potential in low-power WSN systems.

2. Design of the LEAPP Model

2.1. Overall Model Framework

In this paper, a Transformer based sequence-to-sequence model for link quality prediction is designed. The overall architecture is shown in **Figure 1**. The model takes the channel measurement metrics at the receiver as input and outputs the estimated value of the current PRR.

The encoder is responsible for extracting contextual semantic features from the input SINR (Signal to Interference and Noise Ratio) sequence. Specifically, the original SINR sequence is first dimensionally upgraded through linear mapping to match the embedding dimension of the model. Subsequently, it is processed by multiple stacked MHEA (Multi-Head External Attention) modules and feed-forward networks. Residual connections and layer normalization are introduced between each sub-layer to stabilize the training process and accelerate convergence. The decoder takes the PSR (Physical Layer Parameter) sequence as input. The input at each time step undergoes linear transformation, then captures temporal dependencies through the MHEA module, and further enhances the nonlinear expression capability via the feed-forward network. The output of the encoder serves as the contextual input for the intermediate layers of the decoder, participating in cross-sequence attention computation to realize information interaction between the input sequence and the target sequence. In the output stage, the hidden representation of the last layer of the decoder is compressed into a single-channel output through a linear projection layer, establishing the mapping relationship from the SINR sequence to the PRR predicted values. This design not only retains the powerful sequence modeling capability of Transformer but also achieves a balance between real-time performance, interpretability, and deployment feasibility in resource-constrained WSN scenarios through structural innovations.

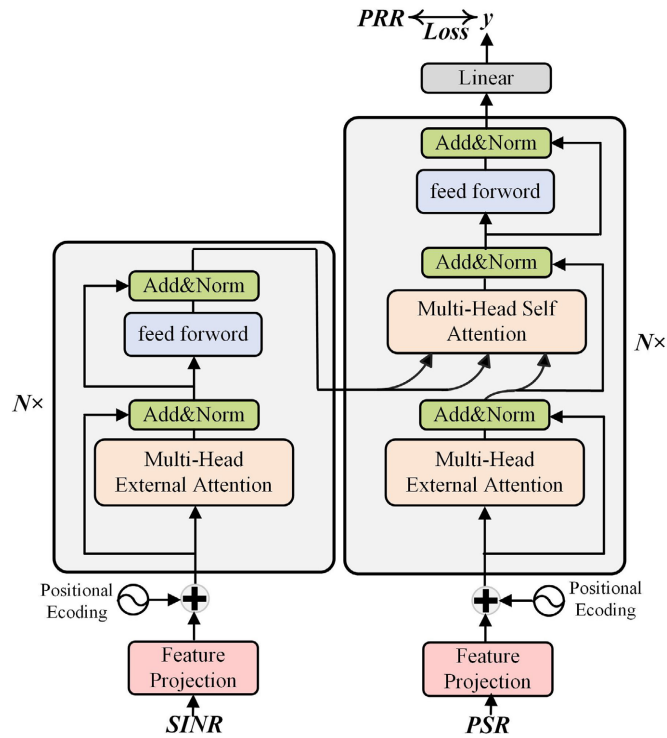


Figure 1. Model structure diagram.

2.2. Data Preprocess

According to communication theory, PSR is a function of the Signal to Interference plus Noise Ratio (SINR) [7]. Therefore, this paper selects SINR as the input feature of the encoder. For the decoder, the PSR calculated based on a modified theoretical model is adopted as the input sequence, with the calculation formulas shown in Equations (1) and (2).

$$BER = \frac{8}{15} \times \frac{1}{16} \times \sum_{k=2}^{16} (-1)^k \binom{16}{k} e^{\left(20 \times (SINR - \tau) \times \left(\frac{1}{k} - 1\right)\right)} \tag{1}$$

$$PSR = (1 - BER)^L \tag{2}$$

Among them, τ is the offset of SINR, which is measured through experiments. In the experiments of this paper, τ is set to 5.5 dB [8], L denotes the packet length, and is set to 37 in this work. The calculation formula of SINR is shown in Equation (3):

$$SINR(dB) = 10 \lg \left(\frac{P_s}{P_i + P_n} \right) \tag{3}$$

Among them, P_s denotes the received useful signal power, while P_i and P_n represent the interference signal power and noise power, respectively.

PSR denotes the probability of successfully receiving a packet under specific SNR conditions. PRR refers to the frequency of successful reception events among N transmissions. The probability that PRR takes a specific value R is determined by the PSR sequence within the statistical window, as shown in Equation (4):

$$\gamma = P\{PRR = R(p_1, p_2, \dots, p_n)\} \quad (4)$$

Among them, γ can be regarded as the prior probability of a successful reception event under specific SINR conditions, *i.e.*, under the condition of a specific PSR sequence. The expected value of PRR is shown in Equation (5):

$$PRR = \sum_{i=0}^N \gamma_i R_i \quad (5)$$

Among them, N denotes the window size. It is quite difficult to derive an analytical expression for γ through theoretical analysis, and machine learning provides an alternative solution—learning the mapping relationship between the PSR sequence and PRR via training.

When calculating PRR using a sliding window, LEAPP introduces the EWMA algorithm to optimize PRR estimation, where PRR labels are computed per packet. First, the local PRR mean is calculated based on a small time window, and then EWMA is applied to this sequence. This strategy retains the high sensitivity of small windows to sudden link changes while effectively alleviating the lag problem of the arithmetic mean of large windows. By virtue of exponentially decaying weights, EWMA can dynamically weight and smooth historical observations, thereby achieving a good balance between real-time performance and stability. The calculation methods of EWMA are shown in Equations (6) and (7).

$$y_0 = x_0 \quad (6)$$

$$y_k = (1 - \alpha)y_{k-1} + \alpha \cdot x_k, (k > 1) \quad (7)$$

Among them, α is the smoothing coefficient. In the experiments of this paper, the time window is set to 5 and $\alpha = 0.075$. The obtained EWMA of the window mean is approximately equal to the arithmetic mean calculated using a statistical window of size $N = 100$.

For data slicing, a sliding window mechanism is adopted to slice data in chronological order to construct fixed-length sequence samples. The SINR feature is mapped to the $[0, 1]$ interval using Min-Max normalization before being fed into the model, so as to improve training stability and convergence speed. The time series is split sequentially; to avoid inconsistent distributions between the training and test sets, samples are shuffled prior to splitting, and then divided into training, validation, and test sets at a ratio of 7:1:2.

2.3. External Attention Mechanism

To reduce the model complexity, LEAPP introduces the external attention mechanism. External attention constructs attention weights based on two lightweight, learnable parameter matrices, which can be implemented using only two cascaded linear layers and two normalization layers (as illustrated in **Figure 2(b)**). Compared with the inherent $O(n^2)$ computational complexity of traditional attention mechanisms (where n denotes the sequence length), external attention not only achieves linear computational complexity but also efficiently captures global feature correlations in the input data. The calculation method of external atten-

tion is shown in Equation (8):

$$External\ Attention(D_q, D_k, D_v) = Norm(D_q D_k^T) D_v \tag{8}$$

Among them, $D_q \in R^{n \times d}$, maps the input features to a matrix of fixed dimension., $D_k, D_v \in R^{S \times d}$ are externally learnable parameter matrices, and the computational complexity of external attention is $O(dSn)$.

The structure of multi-head external attention is illustrated in **Figure 2(b)**. It concatenates the outputs of h groups of external attention heads along the feature dimension and restores them to the model dimension through linear mapping to obtain the final output. The calculation method of multi-head external attention is shown in Equation (9):

$$MultiEHead(D_q, D_k, D_v) = Concat(head_1, head_2, \dots, head_h) W_O \tag{9}$$

where $head_i = ExternalAttention(D_q, D_k, D_v)$

By introducing external parameter matrices, the model retains the global parallel modeling capability of multi-head attention while significantly reducing computational complexity and storage overhead, thereby improving efficiency and stability in long-sequence modeling. This mechanism is particularly suitable for temporal tasks such as communication link quality prediction, which not only require the model to capture long-range temporal dependencies but also demand efficient and scalable inference on resource-constrained edge devices.

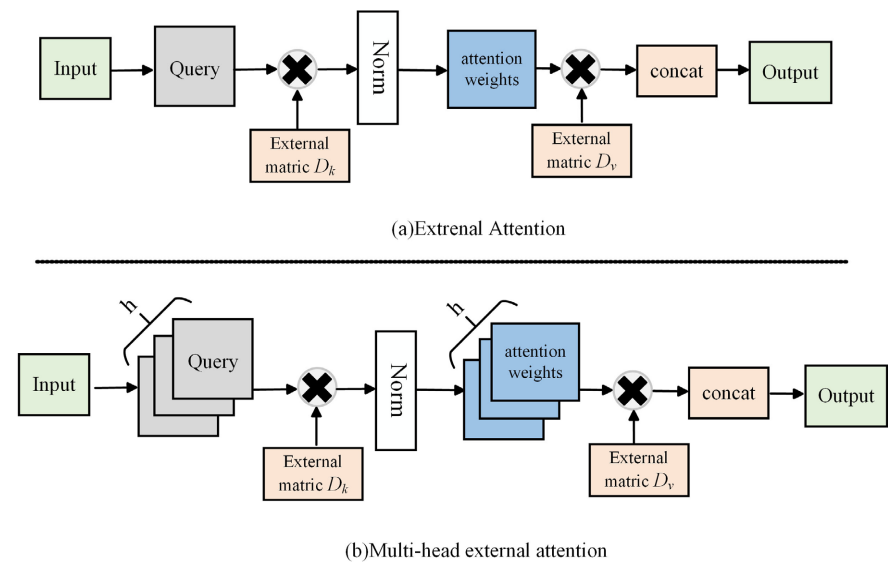


Figure 2. (a) External attention mechanism; (b) Multi-head external attention mechanism.

3. Experiments

3.1. Experimental Platform and Data Collection

To test the performance of the link quality prediction model, both self-collected datasets and public datasets are used in the experiments. The self-collected dataset is built based on a testbed constructed with CC2530 wireless sensor nodes, and

link quality-related data are collected in various environments¹. During the data collection process, two network topologies n-to-1 and 1-to-1 are adopted to cover the link dynamic characteristics under different communication scenarios. The experimental sites include the 9th floor of the Optoelectronics Building (indoor office environment), residential buildings (typical home environment), and parking lots (complex multipath and occlusion environment). The experimental conditions are shown in **Table 1**. In the names of the experimental datasets, suffixes “1” and “n” are used to denote the 1-to-1 and n-to-1 networks, respectively. For the data collected in parking lots, additional suffixes “1” and “2” are used to identify data with transmission powers of -22 dBm and -8 dBm, respectively.

Table 1. Experimental conditions.

Scenario	Dataset	Transmission Power/dBm	Transmission Period/ms	Communication Distance/m
Office Building	Oecb9-1/n	$-22 - 4.5$	100 - 2000	5 - 30
Parking	Parking-n-1/2	$-22, -8$	500	5 - 10
Residential	Resi-1/n	$-22 - 4.5$	50 - 2000	6

The public dataset used in the test experiments is due [9]. This dataset covers link quality data under various experimental conditions, including different communication distances and transmission powers. Even under a single distance condition, the number of data records collected in a single experiment reaches more than 2.4 million, providing sufficient and reliable support for evaluating the model’s performance across diverse scenarios.

The experiments in this study were conducted on a computing platform equipped with an NVIDIA GeForce RTX 4090 GPU, with Ubuntu 20.04 LTS as the operating system. The TensorFlow deep learning framework was adopted to implement model training and testing, and all codes were run in the Python 3.9 environment.

3.2. Ablation Experiments and Sensitivity Experiments

To verify the role of the external attention mechanism in the link quality prediction task, ablation experiments targeting the attention modules of the encoder and decoder were conducted on the Oecb9-n dataset. Only the form of attention was replaced in the experiments, while all other hyperparameters, training strategies, and data partitioning were kept consistent to ensure the comparability of results. Specifically, four model groups were constructed: both the encoder and decoder adopt Multi-Head Self-Attention (MHSA); the encoder adopts MHSA while the decoder adopts Multi-Head External Attention (MHEA); the encoder adopts MHEA while the decoder adopts MHSA; both the encoder and decoder adopt MHEA. The experiments used Mean Absolute Error (MAE) and Root

¹http://gitee.com/WING_USST/wndataset/tree/master.

Mean Square Error (RMSE) as performance evaluation metrics, and counted the FLOPs of the model in a single forward inference to measure computational complexity.

Table 2. Performance comparison of MHEA and MHSA.

enc_attention	dec_attention	FLOPs	MAE	RMSE
MHSA	MHSA	31843617	0.0315	0.0477
MHSA	MHEA	26605863	0.0299	0.0453
MHEA	MHSA	26625319	0.0278	0.0413
MHEA	MHEA	21387565	0.0251	0.0408

As can be seen from **Table 2**, the MAE and RMSE of the baseline model (MHSA-MHSA) are 0.0315 and 0.0477, respectively, with FLOPs of 31,843,617. When MHEA is introduced only in the decoder, the error decreases slightly; while replacing MHSA with MHEA in the encoder leads to a more significant improvement in model performance, indicating that external attention plays a greater role in capturing the global dependencies of input link features. When both the encoder and decoder adopt MHEA, the model achieves the optimal results, with reductions of approximately 22.1% and 16.3% in MAE and RMSE compared with the baseline, respectively, and the computational load is reduced by 10,456,052 FLOPs. This result demonstrates that multi-head external attention improves prediction accuracy while reducing computational overhead, with a particularly prominent effect at the encoder side, thus confirming its superior global modeling capability in the link quality prediction task.

To investigate the prediction accuracy of the proposed model under different hyper parameter settings, controlled variable experiments were conducted from four aspects: input sequence length (*len_seq*), number of rows of the external parameter matrix (*S*), dimension of the mapping layer (*d_model*), and number of attention heads (*num_heads*). All other hyperparameters remain unchanged, where the loss function is MAE, the optimizer is Adam, the learning rate is set to $1e-3$, the batch size is fixed, and the number of training epochs is 300.

During the experiments, only one parameter was adjusted at a time, while the remaining parameters were kept at their default configurations. All experiments were performed under the same training and inference strategies, with Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) as evaluation metrics to measure prediction accuracy and overall stability. As shown in **Table 3**, the model achieves good performance when $S = 32$, while the performance degrades when S increases to 64, indicating that an excessively large parameter matrix may introduce redundant information. Experiments show that $d_model = 128$ is an optimal configuration: an overly low dimension limits the model's expressive capacity, while an overly high dimension causes error rebound, indicating a tendency of over fitting in the model. The model performs best when the number of attention heads is 8, which balances global dependency modeling and computa-

tional overhead. In summary, the optimal configuration of the proposed model is concentrated at $S = 32$, $d_model = 128$, and $num_heads = 8$, demonstrating that the structure designed in this paper can stably achieve high-precision prediction under reasonable parameter conditions.

Table 3. Model sensitivity experiments under different hyperparameters.

S	d_model	num_heads	MAE	RMSE
10	128	8	0.0267	0.0436
20	128	8	0.0288	0.0464
32	128	8	0.0251	0.0408
64	128	8	0.0281	0.0440
32	32	8	0.0330	0.0489
32	64	8	0.0299	0.0451
32	128	8	0.0251	0.0408
32	256	8	0.0509	0.0724
32	128	2	0.0489	0.0691
32	128	4	0.0321	0.0490
32	128	8	0.0251	0.0408
32	128	16	0.0272	0.0440

As can be seen from the relationship between MAE and input sequence length shown in **Figure 3**, the MAE decreases significantly when the sequence length increases from 10 to 32; as the sequence length increases further, the magnitude of the MAE decreases gradually diminishes. This indicates that longer sequences help the model capture richer temporal characteristics of link quality, thereby improving prediction accuracy. However, excessively long sequences not only significantly increase computational overhead but also prolong the waiting time for data collection and inference, impairing the real-time performance of the system. Considering prediction performance, computational efficiency, and response speed comprehensively, this paper selects $len_seq = 32$ as the default configuration for input sequence length in subsequent experiments.

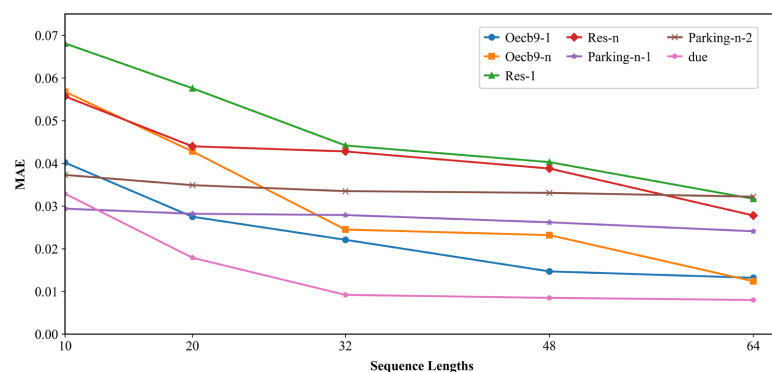


Figure 3. MAE for different input sequence lengths.

3.3. Evaluation of the Improved Decoder Input Sub-Layer

Traditional autoregressive decoders suffer from distribution mismatch between training and inference, which easily causes error accumulation and low inference efficiency. Scheduled sampling [10] dynamically adjusts the teacher-forcing probability to gradually align the training distribution with the inference distribution. Two-stage scheduled sampling further adopts dual-channel training to achieve a smoother transition between training and inference. To verify the effectiveness of the improved decoder input sub-layer proposed in this paper, comparative tests are conducted among LEAPP, the model with the original decoder input sub-layer (teacher-forcing in training and autoregressive mode in inference), and two models improved based on scheduled sampling. Inference time is also evaluated, and the timing includes the data preprocessing process.

The experimental results are shown in **Table 4**. The dataset used is Oecb9-n, which contains more than 10,000 data samples. Due to error accumulation in autoregressive inference, the model with the original decoder input sub-layer has a large prediction error. Models improved with scheduled sampling and two-stage scheduled sampling reduce the error by approximately 50% compared with the full teacher-forcing model. By using PSR as the decoder input, LEAPP eliminates the input discrepancy between training and inference and the resulting error accumulation. The prediction error is reduced by 82% compared with the teacher-forcing model, and by 62% and 61% compared with the scheduled sampling and two-stage scheduled sampling models, respectively. Since LEAPP supports parallel computation during inference, the inference time is reduced by 89% compared with the teacher-forcing model, and by 83% and 84% compared with the scheduled sampling and two-stage scheduled sampling models, respectively. This enables LEAPP to better meet the real-time requirements of application systems for link estimation tasks.

Table 4. Performance comparison of different input sub-layers for the decoder.

Decoder Input	Train method	Inference Time	MAE
PRR	teacher forcing	1.75 s	0.1370
	Scheduled Sampling	1.17 s	0.0641
	two-pass Scheduled Sampling	1.22 s	0.0625
PSR	non-autoregressive training	0.19 s	0.0245

3.4. Comparative Experiments

To comprehensively evaluate the performance of LEAPP, comparative experiments were conducted on multiple datasets between LEAPP and existing mainstream link quality prediction models. The compared models include traditional Recurrent Neural Network (RNN) [11], Long Short-Term Memory (LSTM) [12], Gated Recurrent Unit (GRU) [13], Transformer [5], and the hybrid convolutional and recurrent model CNN + LSTM (CL) [14]. To verify the effectiveness of differ-

ent attention mechanisms in link quality prediction, two extended models are added for comparison: CNN + LSTM + SA (CLSA) with self-attention and CNN + LSTM + EA (CLEA) with external attention. All experiments are carried out under unified hardware and training configurations, *i.e.*, $len_seq = 32$, $S = 32$, $d_model = 128$, $num_heads = 8$, to ensure the comparability of results.

Experimental results (see **Figure 4**) show that traditional recurrent neural network models achieve relatively low accuracy, indicating the limitations of using recurrent structures alone to capture complex link quality characteristics. The CL model achieves better performance than single recurrent neural network models by combining convolution and recurrent structures, demonstrating that joint modeling of local features and temporal dependencies can improve prediction accuracy. The two attention-based extended models, CLSA and CLEA, do not bring significant performance gains, suggesting that simply stacking multiple model components cannot effectively enhance the ability to model complex dependencies. As a typical model constructed with self-attention mechanism, Transformer achieves significantly superior prediction performance compared with traditional recurrent neural networks and CL-based models, reducing the average MAE by approximately 40% relative to RNN. This validates the effectiveness of self-attention in capturing long-range temporal dependencies. Further comparison reveals that models using the external attention mechanism achieve significantly lower average error than those using self-attention. LEAPP achieves the best performance on all datasets. For the public dataset due, the MAE of LEAPP is reduced to 0.0092, indicating that improving the decoder input sub-layer and adopting external attention instead of self-attention can significantly boost prediction accuracy and endow LEAPP with strong generalization ability.

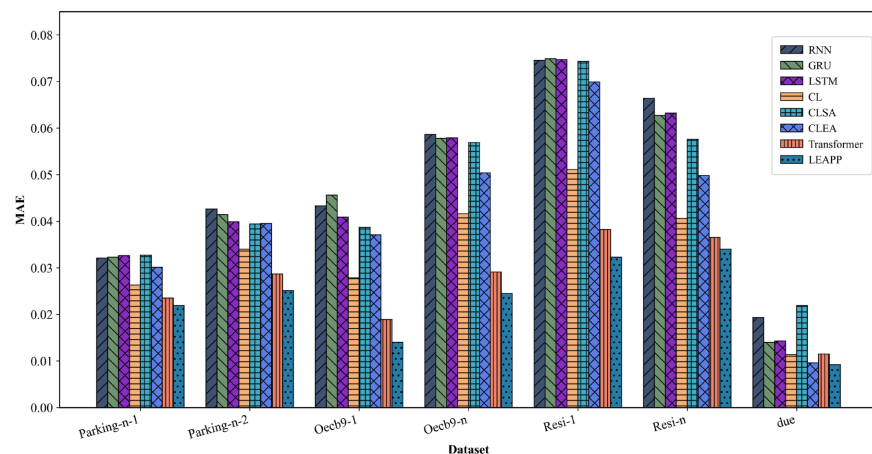


Figure 4. MAE of different models on different datasets.

Figure 5 depicts the temporal fluctuations of the predicted and true PRR values for all models. The experiment is conducted on the Oecb9-n dataset, and consistent patterns are observed across other datasets as well. It can be seen from **Figure 5** that the predicted values of all models generally follow the changing trend

of the true values, but differ in fitting accuracy. The Transformer achieves better fitting performance than traditional recurrent networks and convolutional combination models by virtue of the self-attention mechanism. Its prediction curve is closer to the true values, and the error is significantly lower than that of RNN, GRU, LSTM and CL-based models, demonstrating the superiority of self-attention in modeling temporal dependencies. Compared with the Transformer, LEAPP achieves a notably smaller prediction error, indicating the effectiveness of the external attention mechanism and the improved decoder sublayer in boosting model performance. Due to insufficient capability in modeling local dependencies, individual RNN, LSTM and GRU models suffer from large prediction errors. Combining LSTM with CNN reduces the prediction error to some extent, yet the error is still considerably larger than that of LEAPP.

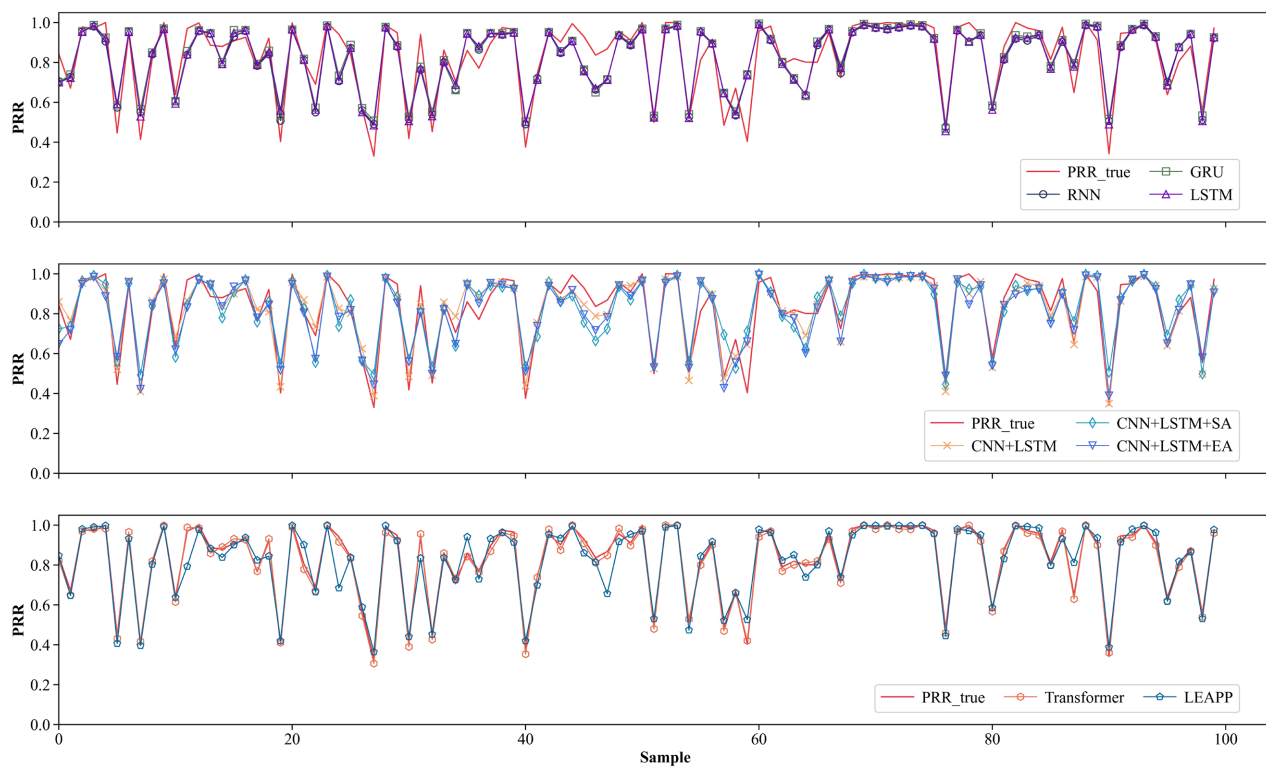


Figure 5. Comparison of errors between predicted results and actual values using different methods.

4. Conclusion

To address the shortcomings of existing methods in prediction accuracy and computational efficiency, a Transformer-based link quality estimation method named LEAPP is proposed. In terms of model design, a multi-head external attention mechanism is introduced to strengthen the global feature modeling capability and reduce model complexity. PSR is adopted as the decoder input to balance prediction accuracy and inference efficiency. Experimental results on multiple datasets demonstrate that compared with existing methods, LEAPP exhibits superior performance in terms of accuracy, stability, and generalization ability in complex environments.

Funding

Supported by the National Natural Science Foundation of China (Grant No. 61374040), and the Science and Technology Development Project of University of Shanghai for Science and Technology (Grant No. 2020KJFZ082).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Ali, A.A., Gharghan, S.K. and Ali, A.H. (2024) A Survey on the Integration of Machine Learning Algorithms with Wireless Sensor Networks for Predicting Diabetic Foot Complications. *AIP Conference Proceedings*, **3232**, Article ID: 040022. <https://doi.org/10.1063/5.0236289>
- [2] Yick, J., Mukherjee, B. and Ghosal, D. (2008) Wireless Sensor Network Survey. *Computer Networks*, **52**, 2292-2330. <https://doi.org/10.1016/j.comnet.2008.04.002>
- [3] Baccour, N., Koubâa, A., Mottola, L., Zúñiga, M.A., Youssef, H., Boano, C.A., et al. (2012) Radio Link Quality Estimation in Wireless Sensor Networks. *ACM Transactions on Sensor Networks*, **8**, 1-33. <https://doi.org/10.1145/2240116.2240123>
- [4] Cerar, G., Yetgin, H., Mohorcic, M. and Fortuna, C. (2021) Machine Learning for Wireless Link Quality Estimation: A Survey. *IEEE Communications Surveys & Tutorials*, **23**, 696-728. <https://doi.org/10.1109/comst.2021.3053615>
- [5] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017) Attention Is All You Need. arXiv: 1706.03762.
- [6] Guo, M., Liu, Z., Mu, T. and Hu, S. (2022) Beyond Self-Attention: External Attention Using Two Linear Layers for Visual Tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **45**, 5436-5447. <https://doi.org/10.1109/tpami.2022.3211006>
- [7] IEEE (2006) IEEE Standard for Information Technology Part 15. 4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (WPANs).
- [8] Shi, J.J., Qiu, Y.H., Long, H.B., et al. (2024) Research On wireless Network Link Quality Estimation Method Based on IEEE 802.15.4 Physical Layer. *Modeling and Simulation*, **13**, 4019-4034. <https://doi.org/10.12677/mos.2024.133365>
- [9] Fu, S., Zhang, Y., Jiang, Y., Hu, C., Shih, C. and Marron, P.J. (2015) Experimental Study for Multi-Layer Parameter Configuration of WSN Links. 2015 *IEEE 35th International Conference on Distributed Computing Systems*, Columbus, 29 June-2 July 2015, 369-378. <https://doi.org/10.1109/icdcs.2015.45>
- [10] Mihaylova, T. and Martins, A.F.T. (2019) Scheduled Sampling for Transformers. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, Florence, 5-10 July 2019, 351-356. <https://doi.org/10.18653/v1/p19-2049>
- [11] Xu, M., Liu, W., Xu, J., Xia, Y., Mao, J., Xu, C., et al. (2022) Recurrent Neural Network Based Link Quality Prediction for Fluctuating Low Power Wireless Links. *Sensors*, **22**, Article 1212. <https://doi.org/10.3390/s22031212>
- [12] Kanto, Y. and Watabe, K. (2024) Wireless Link Quality Estimation Using LSTM Model. *NOMS 2024-2024 IEEE Network Operations and Management Symposium*, Seoul, 6-10 May 2024, 1-5. <https://doi.org/10.1109/noms59830.2024.10575638>

- [13] Liu, L.L., Xiao, T.Z., Shu, J., *et al.* (2022) Link Quality Prediction Based on Gated Recurrent Unit. *Advanced Engineering Sciences*, **54**, 51-58.
- [14] Fan, J.B., and Liu, L.L. (2023) A Hybrid Model with CNN-LSTM for Link Quality Prediction. 2023 *6th International Conference on Electronics Technology (ICET)*, Chengdu, 12-15 May 2023, 603-607.
<https://doi.org/10.1109/icet58434.2023.10211502>