Heliyon 9 (2023) e19870

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

A novel transformer-based method for predicting air absorbed dose rates in nuclear radiation environmental monitoring

Yizhi Cao [a], Zhaoran Liu [a], Yunlong Niu [b], Xinggao Liu [a,*]

[a] Zhejiang University, HangZhou, 310027, Zhejiang, China
[b] Radiation Environment Monitoring Technology Center, Ministry of Ecology and Environment, Hangzhou 310012, PR China

ARTICLE INFO

Keywords:
Air absorbed dose rate
Nuclear radiation prediction
Transformer
Long-time series forecasting

ABSTRACT

Many studies have used various methods to estimate future nuclear radiation levels to control radiation contamination, provide early warnings, and protect public health and the environment. However, due to the high uncertainty and complexity of nuclear radiation data, existing prediction methods face the challenges of low prediction accuracy and short warning time. Therefore, accurate prediction of nuclear radiation levels is essential to safeguard human health and safety. This study proposes a novel Mixformer model to predict future hourly nuclear radiation data. The seasonality and trend of nuclear radiation data are extracted by data decomposition. To address the slow speed problem common in traditional methods for long-time series prediction tasks, Mixformer simplifies the decoder with convolutional layers to speed up the convergence of the model. The experiments consider the air-absorbed dose rate of nuclear radiation data, spectral data, six climatic conditions, and two other conditions. We use MSE and MAE metrics to verify the effectiveness of Mixformer prediction. The results show that the Mixformer proposed in this paper has better prediction performance compared to the currently popular models. Therefore, the proposed model is a feasible method for industrial nuclear radiation data processing and prediction.

1. Introduction

Nuclear energy is an essential energy source, but the effects of nuclear radiation on humans and the environment remain a concern [1,2]. When nuclear radiation enters the body, it can damage cells and cause cancer or other health problems [3]. In addition, the effects of nuclear radiation on the environment and ecosystems are severe [4]. Radioactive materials may contaminate the environment through air, water, or soil, leading to radioactive contamination of soil and water sources [5]. Large amounts of radioactive material are released into the atmosphere, which impact climate change [6]. As a result, to protect the environment, monitoring and predicting nuclear radiation data is essential [4,7].

Nuclear weapon tests and nuclear accidents can lead to significant emissions of carbon dioxide and other gases, which can affect the global climate. In severe nuclear accidents, such as the 1986 Chornobyl nuclear leak [8] and the 2011 Fukushima nuclear accident in Japan [9], large numbers of deaths or long-term health problems occurred. Saenko described the accident at the Chornobyl nuclear power plant in detail. The accident released various radioactive isotopes, which led to radioactive contamination of a large area around the accident site and increased the vulnerability of the surrounding population to radiation-related thyroid cancer [10]. Wada

* Corresponding author.
E-mail address: lxg@zju.edu.cn (X. Liu).

https://doi.org/10.1016/j.heliyon.2023.e19870
Received 27 March 2023; Received in revised form 1 September 2023; Accepted 4 September 2023
Available online 9 September 2023

Fig. 1. The location of Yangtse North Road Station.

described the radionuclide contamination after the Fukushima Daiichi nuclear power plant (FDNPP) accident. The spatial and temporal trends of (137)Cs concentrations in different habitats and the radiation effects on freshwater fish are explored [11]. Rong used the Regional Ocean Model System (ROMS) to build a high-resolution unidirectional nested model to simulate and predict the potential impact of the Fukushima nuclear spill on the East China Sea (ECS), with a large domain covering the entire North Pacific Ocean [12]. Once a nuclear accident occurs, the adverse environmental effects are irreversible.

To prevent major nuclear accidents, various methods have been proposed to monitor and process nuclear radiation data. Hochbaum and Fishbain utilized a mobile distributed sensor network for nuclear threat detection. The presence of potential sources of danger and their approximate location is depicted by identifying a small area with a high concentration of reported risk [13]. Jeong provided a method for understanding changes in radiation levels using graphical comparisons. Whereas previous work used static sensor networks to detect and track radiation, mobile sensor networks offer a distinct advantage in tracking illegal radioactive materials [14]. Ogura conducted a discriminatory measurement of absorbed dose rates of natural and artificial radionuclides in the air in Namiie Town, Fukushima Prefecture. It is also expected to be used for radiation risk reporting by differentiating natural and artificial radionuclides to assess the annual effective external exposure dose [15]. Miles performed high-fidelity simulations of radiation transport within urban areas using the Monte Carlo N Particle (MCNP) method to infer the location and intensity of unknown sources [16]. Tsoulfanidis and Landsberger provided the most up-to-date and accessible introduction to radiation detector materials, systems, and applications [17]. These studies employ various data techniques to analyze the acquired nuclear radiation data, enable the processing of nuclear radiation data, and provide early warning of nuclear accidents.

However, nuclear radiation data are more complex than other data, and accurate prediction of nuclear radiation data is a challenging task [18]. Traditional nuclear radiation data analysis methods are no longer suitable for today's needs. Deep learning methods have recently been widely used for time-series data processing and prediction [19–21]. Choi and Lee proposed a machine learning model using consistency metrics for sensor error detection in nuclear power plant emergencies [22]. Wang proposed an LSTM-based anomalous event detection, identification, and isolation network to help maintain the safe operation of nuclear power plants [23]. Tsichrintzēs conducted a comprehensive survey and discussion of machine learning and data analysis methods related to nuclear safety [24]. Traditional condition monitoring methods do not provide operators with alerts or predictions of automatic shutdown in the event of abnormal conditions in nuclear power plants that do not immediately lead to an automatic shutdown. Chen presented an LSTM-based model for predicting the automatic scram assuming no operator intervention [25].

Although these deep learning methods have been successfully applied to nuclear radiation data processing, these methods have low prediction accuracy, slow response time, and do not sufficiently consider the problem of long-time prediction. These models do not apply to the prediction of long time series data, and the prediction accuracy decreases when the period of the input data is longer. Inspired by the problems of the existing model, we propose a new model called Mixformer, which retains the framework of the transformer. The encoder adds the decomposer module, which decomposes the data into seasonal and trend terms and learns the spatiotemporal relationship between the two data separately. The structure of the decoder is simplified to improve the model's prediction accuracy while speeding up the convergence of the model. The contributions of this study are summarized as follows: First, the problem of nuclear radiation data prediction is studied, which provides a research basis for early nuclear accident warning. Second, the prediction accuracy of nuclear radiation data is improved, and the results are better than other methods. Third, the nuclear accident warning time is long, and it can predict nuclear radiation data for a long time in the future. Fourth, the model speeds up the prediction response and saves computational resources and time. The experimental results show that Mixformer is more accurate in predicting complex nuclear radiation data, has longer prediction time steps, and runs faster than other prediction models.

The remaining sections of this paper are as follows. Section 2 describes the study area. The experimental data are analyzed and processed. Section 3 introduces Transformer, a popular method for timing prediction, and explains the structure and advantages of the new model Mixformer. Section 4 focuses on analyzing the results of the experiments and discussing the ablation studies. Section 5 summarizes the importance of the research in this paper to protect the environment. The paper experimental code has been uploaded

Table 1
Introduction of nuclear radiation data parameters at North Yangzi Road Station.

Index	Unit
Air-absorbed dose rate	$Gy \cdot h^{-1}$
Spectrometer value	eV
Temperature	$^{\circ}C$
Humidity	%
Air pressure	hPa
Wind direction	$^{\circ}$
Wind speed	m/s
Rainfall	mm
Spectrometer voltage	V
Spectrometer temperature	$^{\circ}C$

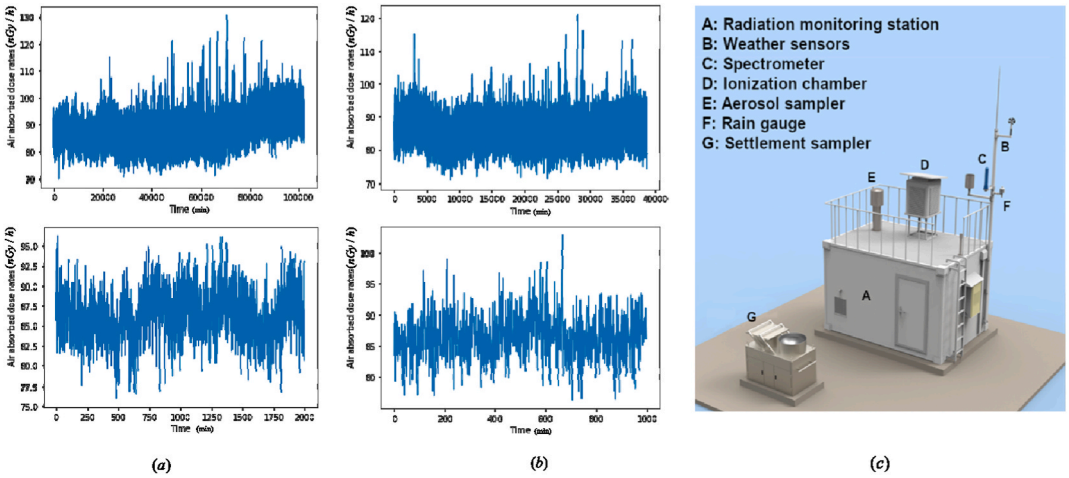


Fig. 2. The collection and display of study data. (a): Distribution of air absorbed dose rate data at Yangtze North Road Station. (b): Distribution of air absorbed dose rate data at Beijing Changchi Road Station. (c): The composition of a specific monitoring station.

to the github, https://github.com/1376588939/Mixformer_project.

2. Study area and data

2.1. Study area

The study area is the nuclear monitoring data of the Yangtse North Road Station in Yangzhou, Jiangsu, which is a relatively typical urban environmental monitoring station, and its radiation data can represent the radiation level in the urban environment and has a certain degree of representativeness. The nuclear radiation data used in the experiments were from January 1, 2020, to December 29, 2020. The nuclear power plant sampled monitoring data every 5 min, generating about 105,000 nuclear data. The location of the study area is shown in Fig. 1. Yangtse North Road Station is one of the national environmental monitoring stations, and its monitoring data has high accuracy and credibility, which can be used as a reliable data source for the study.

This study considered air-absorbed dose rates and nuclear radiation data measured by spectrometers. The air-absorbed dose rate is the ambient surface gamma radiation dose rate in gray-hour^{-1} (Gyh^{-1}) units before deducting the portion of the instrument response to cosmic rays. The background level of absorbed dose rate in the air mainly consists of two items: 1) the ambient surface γ radiation dose rate generated by natural radioactive materials contained in local strata and natural radioactive materials contained in surface buildings; 2) by the cosmic ray dose rate from space. The air-absorbed dose rate is a measurable quantity that can directly, quickly, and continuously reflect the environmental radiation level and is an integral part of environmental radiation monitoring. Nuclear leakage can lead to a significant and sustained increase in the absorbed dose rate of air, such as the nuclear accident in Fukushima, Japan, which led to a significant increase in the absorbed dose rate of air in its vicinity for a long time. As of early 2014, some sites' air-absorbed dose rates still reached tens of thousands of nGyh^{-1}. Monitoring the air-absorbed dose rate can obtain information about the environment's natural background γ radiation level and the environmental γ radiation level and its changes caused by human practice. It can also monitor the condition of nuclear facilities and other sources of radiation devices and issue alarms for abnormal releases or accident situations. Provide data for estimating the external exposure dose to the public caused by the γ radiation in the environment. In addition to this, the data include six climatic conditions: temperature, humidity, air pressure, wind direction, wind

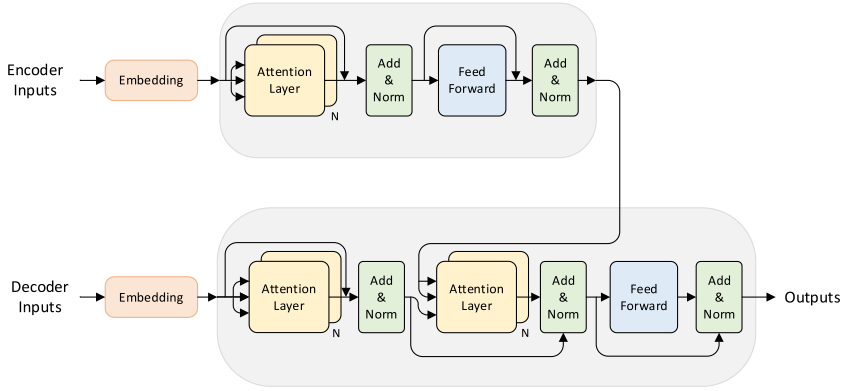


Fig. 3. Vanilla Transformer's encoder-decoder architecture. Top: The encoder includes an attention layer, norm layers and a feed forward layer. Bottom: The decoder includes two attention layers, norm layers and a feed forward layer.

speed, rainfall, and two other spectrometer operating conditions, including spectrometer voltage and spectrometer temperature. These two are the spectrometer variables, that is, the measured voltage and the measured temperature of the spectrometer. The output of the detector relates to the final measurement results, so the impact of these two variables cannot be ignored. Failure to consider the variations introduced by the measurement instruments may lead to erroneous prediction results. A detailed description of the reference indicators shows in Table 1.

2.2. Data processing

When it comes to monitoring instrument calibration, reset, and other routine maintenance behavior, or instrument and communication failure, power outages, and other phenomena, certain stations will have no monitoring or abnormal monitoring values for some time. We first pre-process the data before training. The missing values in the data are filled using averaging for the missing values, and the abnormal values are discarded. Because the data sampling interval is only 5 min, it results in a large amount of redundant information in the dataset. Therefore, by downsampling, we collect data for each hour to remove redundant information while relieving the pressure on the model training. The distribution of the absorbed air dose rate throughout the year is illustrated in Fig. 2 (a). We sliced the data to show the first two thousand data items, which makes it clearer to see the irregular data distribution and illustrates the difficulty of accurate prediction. The Radiation Monitoring Technology Center is equipped with a complete range of advanced domestic and radiation environment monitoring instruments, as shown in Fig. 2. (c). In order to verify the applicability of the model in different situations, the dataset collected at Changchi Road Station in Changping, Beijing, is also used in this experiment. This area is geographically different from Yangzhou and has a different climate. The dataset distribution is shown in Fig. 2 (b).

3. Methodology

Predicting future nuclear radiation data based on past observations at the Yangtze monitoring station is crucial. The general framework of the nuclear radiation monitoring early warning model proposed in this paper is shown in Fig. 5. The Mixformer model has three crucial modules: encoder, decoder, and decomposer. The raw input data is pre-processed and fed to the encoder. The role of the encoder is to extract the spatio-temporal correlation between different data features. Also, the raw input data is fed into the decoder to extract the seasonal and trend features of the nuclear radiation data. The predicted value is the combination of the encoder and decoder outputs.

3.1. Transformer for long-time series forecasting

To capture the spatiotemporal relationship between nuclear radiation data and the predictor variables, the prediction model proposed in this paper is based on Transformer [26] and its variant in time series prediction. The Transformer model is prevalent for time series tasks, and it achieves the best results on many tasks through the encoder-decoder structure and the attention mechanism. The structure of the vanilla transformer is shown in Fig. 3.

Vanilla Transformer uses the classic encoder and decoder architecture. The encoder consists of a position encoding layer, a self-attentive layer, a feedforward layer, and a normalization layer. The input layer converts the input time series into a vector of dimension D through a fully connected layer, which is input to the multi-headed attention layer after position embedding. Each attention layer extracts the correlation of different features of the input vector, which is then connected and normalized through the residuals layer. The obtained vector is the output of the encoder after passing through a fully connected feedforward layer.

The self-attentive layer is the critical part of the Transformer, which accepts the input x encoded by the position and obtains the matrices Q (query), K (key), and V (value) after the linear transformation W. After getting the matrices Q, K, and V, the output of the self-attention layer can be calculated,

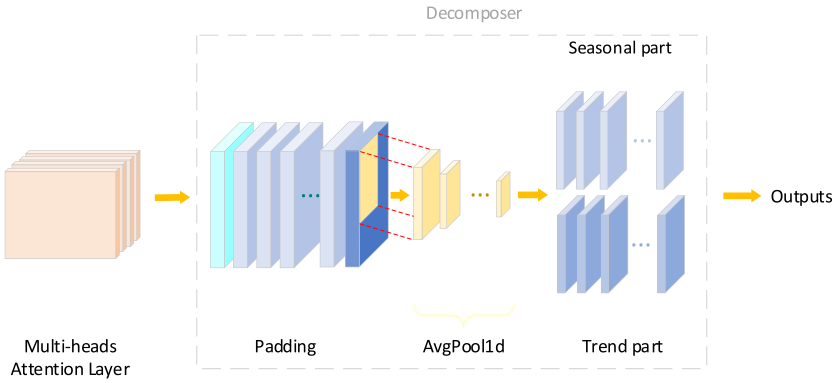


Fig. 4. The internal structure of the decomposer in Mixformer. Decomposer includes two steps: padding and 1-D AvgPool. The output of decomposer is divided into seasonal part and trend part.

$$\text{Attention}(Q,K,V)=\text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1}$$

Equation (1) determines the inner product of each row vector of matrices Q and K. The inner product is divided by the square root of d_k to keep it from being excessively huge, where d_k is the dimension of the input vector. The number of rows in the resulting matrix, generated by multiplying Q by K's transpose, is n, where n is the length of the temporal data. This matrix can be used to express how much attention was paid to each piece of data.

The output of the attention layer goes through the Add & Norm layer. The Add & Norm layer consists of two parts, Add module and Norm module, which are calculated as follows,

$$X_A=\text{LayerNorm}(X+\text{Attention}(X)) \tag{2}$$

X denotes the input to the attention layer. The Add module is a residual connection, typically used to solve the problem of training multilayer networks, that allows the network to focus only on the part of the current difference.

The Feed Forward layer is relatively simple, being a two-layer fully connected layer with ReLU as the activation function for the first layer and no activation function for the second layer, corresponding to the following equation,

$$\max\left(0,XW_{f_1}+b_{f_1}\right)W_{f_2}+b_{f_2} \tag{3}$$

The output of the feedforward layer also goes through the Add & Norm layer,

$$X_F=\text{LayerNorm}(X+\text{Feedforward}(X)) \tag{4}$$

X denotes the input to the feedforward layer. X_F is the final output of the encoder layer. The composition of the decoder and encoder layers is the same, except the decoder has one more attention layer. Notice that the input of the decoder's first attention layer is the decoder's original input data. Furthermore, the input of the second attention layer is the output of the encoder and the output of the first attention layer. The rest of the structure is the same as that of the encoder.

The transformer model achieves parallelized data training without the inherent sequential properties of previous temporal models such as RNN [27], LSTM [28], and GRU [29]. Moreover, the number of operations required by the Self-attention module to compute the association between two positions is distance independent. Therefore, Transformer has better results for the prediction of long-time series data. Nevertheless, the Transformer does not perform as well as it should regarding nuclear radiation data. Because of the problem of encoding location information in the Transformer, it focuses more on the spatial relationship of the data and ignores the temporal relationship of the data.

3.2. *Mixformer for long-time series forecasting*

Accurately predicting complex nuclear radiation data over long periods is a thorny problem. The new model proposed in this paper improves the Transformer, allowing the Mixformer to perform better when dealing with prediction tasks similar to nuclear radiation data. The structure of Mixformer is shown in Fig. 5.

As a whole, the Mixformer model retains the encoder and decoder structures in the Transformer model. Considering the complexity of nuclear radiation data, the Mixformer model simplifies the construction of the decoder. The complex model is more likely to cause overfitting to the training data, especially the nuclear radiation data, without apparent patterns. Therefore, the decoder of Mixformer consists of only the decomposer, convolutional layer, and feedforward layer.

First, the nuclear radiation data is preprocessed, and the resulting data is fed into the embedding layer. The embedding layer in Mixformer is different from that in the vanilla Transformer. It includes position embedding, token embedding, and temporal embedding. Position embedding uses sine and cosine functions to encode the input order of the data. Token embedding is a one-

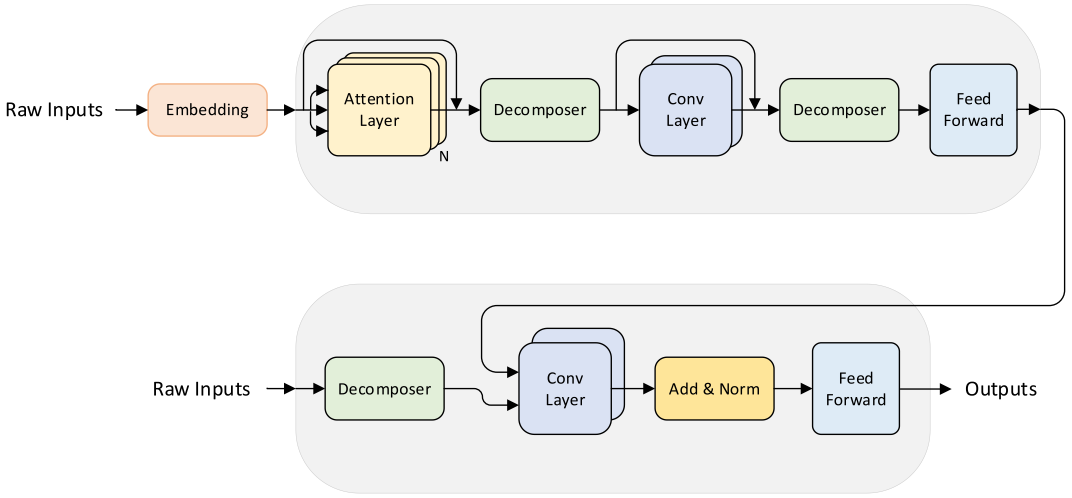


Fig. 5. The model structure of Mixformer. Top: The Encoder module includes N attention layers, two decomposers, convolution layers and a feed forward layer. Bottom: The decoder module includes a decomposer, convolution layers, feed forward layer and norm layer. The output of encoder is fed into the convolution layer in decoder.

dimensional convolution of the input data to expand the feature dimension of the data. Moreover, temporal embedding is responsible for further extracting the temporal relationship between the data. Each piece of nuclear radiation data is classed according to the month, day, hour, and minute because it is sampled yearly. Three embedding steps are necessary to reach the desired outcome.

$$X_{em} = \text{Token}(X_{in}) + \text{Position}(X_{in}) + \text{Temporal}(X_{in}) \tag{5}$$

After X_{em} is input to Encoder, the multi-headed self-attentiveness of the input sequence length is calculated first. The multi-headed self-attentiveness mechanism, which the following equation can express as,

$$\text{MultiHead}(Q, K, V) = \text{Concat}(h_1, h_2, \cdots, h_n)W \tag{6}$$

Where $h_i = \text{Attention}(Q_i, K_i, V_i)$ denotes the i th attention head, h denotes the number of heads, and W_i^Q, W_i^K and W_i^V denote the projection matrix of queries, keys, and values, respectively. W is the final mapping matrix.

Unlike other models, Mixformer will decompose the multi-headed attention layer's output to get the input data's seasonal and trend characteristics, effectively predicting nuclear radiation data. The patterns of nuclear radiation data itself are complex and difficult to predict. If the data is divided into two parts, it is easier to find the intrinsic relationships. The structure of the decomposer is shown in Fig. 4. The decomposer uses an average pooling layer to traverse the entire data time series and find its trend features. However, the feature extraction step is preceded by a padding operation on the input. The padding operation allows the input dimension to be the same as the output dimension. To better use the information at both ends of the time-series data, we first replicate the data at both ends and then pass the one-dimensional averaging pooling layer. Convolution kernels of different sizes also extract trend features of different lengths. The final seasonal and trend terms are obtained.

Then the features are learned by convolutional layers for seasonal and trend terms, respectively. The output is again passed through the decomposer to uncover the temporal correlation of the data further. Finally, the encoder output is obtained by a feedforward layer, either a convolutional layer or an MLP. The calculation process of the encoder is shown below,

$$\begin{aligned} X_a &= \text{Attention}(X_{em}), \\ S_{en}^1, T_{en}^1 &= \text{Decomposer}(X_a + X_{em}), \\ S_{en}^2, T_{en}^2 &= \text{Decomposer}\left(\text{Conv1d}\left(S_{en}^1\right) + S_{en}^1\right), \\ X_{en} &= \text{Linear}\left(\text{LayerNorm}\left(S_{en}^2\right)\right). \end{aligned} \tag{7}$$

The decoder structure of Mixformer has been simplified and improved. To retain the initial characteristics of the original input data, it enters the decoder directly without embedding processing. The first module of the decoder is the decomposer. The seasonal term S_{de} and the trend term T_{de} are obtained by convolutional layers. The output of the encoder is then summed with X_{en} and fed into the feedforward layer to obtain the decoder output. The calculation process of the decoder is shown below,

$$\begin{aligned} S_{de}^1, T_{de}^1 &= \text{Decomposer}(X_{de}), \\ S_{de}^2 &= \text{Conv1d}\left(S_{de}^1\right), \\ T_{de}^2 &= \text{Conv1d}\left(T_{de}^1\right), \\ X_{de} &= \text{Linear}\left(\alpha X_{en} + S_{de}^2 + T_{de}^2\right). \end{aligned} \tag{8}$$

Table 2
MSE and MAE results of Mixformer versus other methods for predicting air-absorbed dose rates. The input sequence length is fixed at 96h, and the predicted time lengths are 48h, 96h, 192h, 336h, and 720h. Each data is sampled at 1h intervals. Each method was run three times, and the average values were given. The best results are highlighted in bold, and the second best is underlined.

Method	Metric	Predictionlength : 48			Mean	Predictionlength : 96			Mean	Predictionlength : 192			Mean	Predictionlength : 336			Mean	Predictionlength : 720			Mean
Transformer	MSE	0.7178	0.6939	0.7325	0.7147	1.0089	0.9500	0.9928	0.9839	1.6108	1.6332	1.5794	1.6078		\	\	\	\		\	\
	MAE	0.6697	0.6587	0.6770	0.6685	0.7952	0.7692	0.7874	0.7839	1.1035	1.1227	1.0928	1.1063	\		\	\	\	\	\	
Autoformer	MSE	1.9269	0.7426	0.7476	1.1390	2.8329	2.8984	2.8645	2.8653	3.7865	3.6597	3.5572	3.6678	\	\	\	\	\	\	\	\
	MAE	0.8013	0.6728	0.6431	0.7057	0.9742	1.0808	1.0738	1.0429	1.6573	1.6097	1.5782	1.6151	\	\	\	\	\	\	\	\
Informer	MSE	**0.6939**	0.7548	**0.6982**	0.7156	0.8657	0.7801	0.8875	0.8444	1.6464	1.4041	1.4511	1.5005	\	\	\	\	\	\	\	\
	MAE	**0.6587**	0.6879	0.6606	0.6691	0.7316	0.6971	0.7434	0.7240	1.0511	0.9528	0.9709	0.9916	\	\	\	\	\	\	\	\
FEDformer-f	MSE	0.7210	**0.6885**	0.7013	**0.7036**	2.6035	2.6537	2.6663	2.6745	3.9035	3.8537	3.8663	3.8745	\	\	\	\	\	\	\	\
	MAE	0.6367	**0.6377**	**0.6398**	**0.6381**	1.0789	1.0743	1.0758	1.0763	1.6789	1.6743	1.6758	1.6763	\	\	\	\	\	\	\	\
LSTM	MSE	1.3235	1.3586	1.4907	1.3909	1.9219	1.5961	1.7891	1.7690	2.0186	2.1147	1.8325	1.9886	1.6651	1.6827	1.6995	1.6824	2.2170	2.2485	1.6911	2.0522
	MAE	0.9366	0.9524	1.0064	0.9651	1.1661	1.0326	1.1174	1.1054	1.2010	1.2338	1.1324	1.1891	1.0675	1.0752	1.0818	1.0748	1.2593	1.2692	1.0726	1.2004
Linear	MSE	0.8611	0.8547	0.8610	0.8589	1.0560	1.0630	1.0579	1.0590	1.2941	1.2953	1.2917	1.2937	1.4300	1.4328	1.4289	1.4306	1.7522	1.7498	1.7490	1.7503
	MAE	0.7325	0.7297	0.7326	0.7316	0.8173	0.8203	0.8184	0.8187	0.9157	0.9164	0.9149	0.9157	0.9706	0.9716	0.9699	0.9707	1.0923	1.0915	1.0911	1.0916
DLinear	MSE	0.8152	0.8165	0.8142	0.8153	1.0091	1.0189	1.0184	1.0155	1.2536	1.2479	1.2556	1.2524	1.3952	1.3941	1.3984	1.3959	1.7070	1.7176	1.6968	1.7071
	MAE	0.7119	0.7124	0.7115	0.7119	0.7985	0.8027	0.8026	0.8096	0.9018	0.8994	0.9026	0.9013	0.9600	0.9595	0.9612	0.9602	1.0783	1.0826	1.0743	1.0784
NLinear	MSE	0.8713	0.7714	0.7593	0.8007	0.8532	0.8554	0.8328	0.8471	1.0607	1.2178	1.2734	1.1840	1.4798	1.2970	1.7079	1.4949	1.5420	1.5276	1.5315	1.5337
	MAE	0.7091	0.6994	0.6834	0.6973	0.7076	0.7170	0.7042	0.7096	0.7990	0.8768	0.8987	0.8582	0.9783	0.9077	1.0691	0.9850	0.9467	0.8846	0.8973	0.9095
Mixformer	MSE	0.7081	0.7055	0.7095	0.7077	**0.7321**	**0.7344**	**0.7317**	**0.7327**	**0.7700**	**0.7698**	**0.7670**	**0.7689**	**0.7933**	**0.7923**	**0.7921**	**0.7926**	**0.8285**	**0.8281**	**0.8275**	**0.8280**
	MAE	0.6645	0.6632	0.6650	0.6642	**0.6760**	**0.6772**	**0.6759**	**0.6764**	**0.6937**	**0.6937**	**0.6923**	**0.6932**	**0.7048**	**0.7043**	**0.7043**	**0.7045**	**0.7212**	**0.7211**	**0.7209**	**0.7211**

[1] The '\' indicates failure for the out-of-memory.

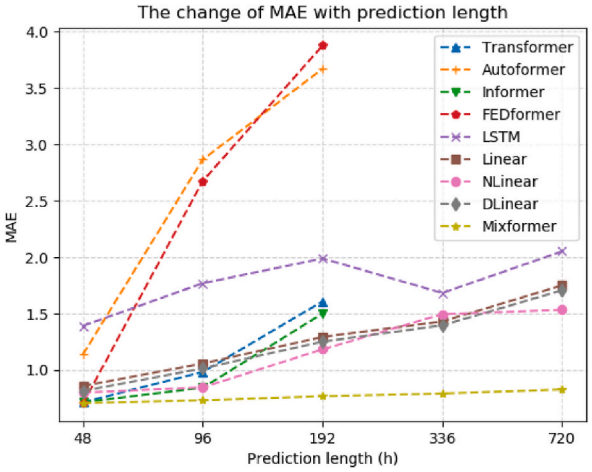


Fig. 6. The change of MSE for the prediction length: {48, 96, 192,336,720}.

X_{de} is the output of the decoder, which is the predicted value of the nuclear radiation data. α is the learning rate. The loss of the model is obtained by calculating the error between the predicted value $\widehat{y}$ and the actual value y, and then the network parameters are updated by backpropagation. The validity indicators of the model prediction performance are MSE and MAE, which are calculated by the formulas as follows,

$$\begin{aligned} \mathrm{MSE} &= \sum_{i=i}^{N} (\widehat{y}_i - y_i)^2 \\ \mathrm{MAE} &= \sum_{i=i}^{N} |\widehat{y}_i - y_i| \end{aligned} \tag{9}$$

4. Results and discussion

4.1. Experiment settings

Dataset. The data comes from the automatic monitoring results of the radiation environment automatic monitoring stations of the National Radiation Environment monitoring network, and the monitoring points include environmental quality monitoring points and nuclear power plant monitoring points. The monitoring station preprocessed the experimental data. Instrument measurements with a relative intrinsic error of less than 15% were selected for the monitoring station, and the test source decay correction was considered. The extended uncertainty of the ambient γ radiation dose rate measurement does not exceed 20%. The cosmic ray response values at the measurement points were corrected considering the correction factor for shielding cosmic rays by buildings. To ensure a fair comparison of model performance, we divided the data into a training, validation, and test set, with corresponding percentages of 70%, 10%, and 20%, respectively. The data include air-absorbed dose rate, spectral data, temperature, humidity, barometric pressure, wind direction, wind speed, rainfall, spectrometer voltage, and spectrometer temperature for the past year of nuclear radiation data at the North Yangzi Road station. The experimental results at Beijing Changchi Road station are presented in the appendix. The prediction is the future air-absorbed dose rate. This study belongs to the problem of multivariate prediction of univariate in long time series prediction.

Evaluation Metric. We use mean square error (MSE) and mean absolute error (MAE) as the core metric for comparing performance. Our method is trained with the L2 loss, using the Adam optimizer. The training process is early stopped within 10 epochs.

Compared Methods. We compare the LSTM [28] and four recent transformer-based methods: FEDformer [30], Autoformer [31], Informer [32], and Transformer [26]. In addition, we compare the latest State-Of-The-Art (SOTA) method Linear [33] and its two variants, which simplify the attention module in the Transformer, and we use it as an alternative implementation of the baseline. Since there are two variants of FEDformer, we compare the one with better accuracy (FEDformer-f by Fourier transform).

4.2. Comparison with transformers

For better comparison, we performed the same preprocessing on the input data for each model, where the length of the input data was 96. The lengths of the predicted data were 48, 96, 192, 336, and 720. Table 2 summarizes the multivariate time series prediction results of all models on the nuclear radiation data set. As shown in Table 2, Mixformer has the highest accuracy in predicting the air-absorbed dose rate for more extended periods in the future. When predicting the air-absorbed dose rate for the next short 48 h, Mixformer's results are similar to those of other transformer methods. FEDformer-f has the best result with an MSE of 0.7036, but

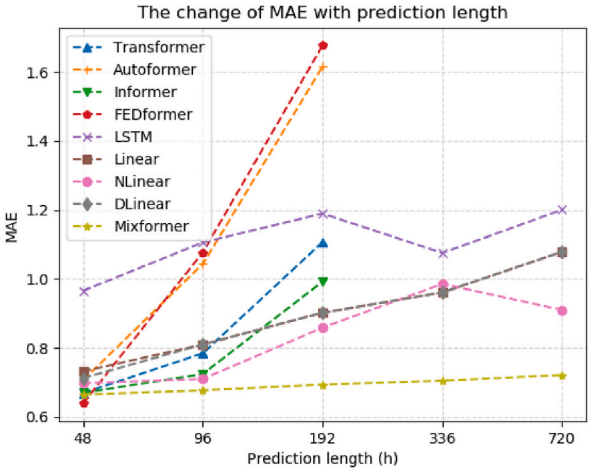


Fig. 7. The change of MAE for the prediction length: {48, 96, 192,336,720}.

Table 3
Network parameters (Param) and floating-point arithmetic numbers (Flops) of Mixformer compared to other Transformer-like models.

Method	Mixformer	FEDformer	Informer	Autoformer	Transformer
Param(10^5)	**86.30**	163.01	113.35	105.34	105.47
Flops(10^9)	**103.45**	152.16	140.00	152.16	152.23
Time(s)	**58.0**	380.1	81.1	138.9	76.2

Mixformer's MSE is only 0.7077. The prediction accuracy of the two methods is similar. The transformer class method outperforms the Linear and LSTM when predicting the absorbed air dose rate for shorter durations. But the situation changed when the length of the prediction increased. In particular, Mixformer's enhancements are very significant compared to Transformer-based models. Compared to the SOTA method FEDformer, Mixformer improves the overall accuracy by 72.6% relative to MSE. The reason for this anomaly is that the FEDformer model focuses more on the cyclical pattern of the data. It extracts the frequency domain features of the data by performing Fourier transform on the original data. Therefore, FEDformer performs well on data sets with significant periodic patterns. However, nuclear radiation datasets are affected by many complex and unstable factors and do not have obvious periodic patterns. These Transformer model-based methods do a lot of feature extraction on the raw data but do not perform as well as the simple Vanilla Transformer. Nevertheless, Transformer's performance is also inferior to Mixformer, with Transformer's MSE being 0.9839. Mixformer's effect is improved by 34.3%. The SOTA Transformer methods even suffer from GPU out-of-memory problems when predicting longer time steps like 336h and 720h. The visualization MSE and MAE results of Mixformer and other methods in predicting different time steps are shown in Fig. 6 & Fig. 7. It can be seen that the MSE and MAE of Mixformer increases slowly and smoothly with the increase of the prediction time step. When the prediction time is up to 720h, Mixformer's MSE is only 0.832, 45.8% lower than the sub-optimal NLinear. However, the MSE and MAE of other methods increase rapidly with the increase in prediction length. In particular, their model complexity affects FEDformer-f and Autoformer, resulting in poor final generalization. The experimental results show that the Mixformer model has better generalization capability when dealing with nuclear radiation data and can adapt to complex and variable air-absorbed dose rates and make reasonable predictions. In addition, compared to other Transformer class methods, Mixformer is faster and has fewer network parameters. Table 3 shows the network parameters, floating-point arithmetic numbers and the time required to run a single epoch.

Table 4
Results for Mixformer and variants predicting time step of 96.

Methods	Metrics	Prediction length: 96			Mean
Mixformer	MSE	0.7328	0.7554	0.7532	**0.7471**
	MAE	0.6876	0.7170	0.7142	**0.7063**
Mixformer-V1	MSE	0.9631	0.9094	0.9669	0.9465
	MAE	0.7689	0.7258	0.7426	0.7458
Mixformer-V2	MSE	1.1646	1.3601	1.3814	1.3020
	MAE	0.7256	0.7531	0.7078	0.7288

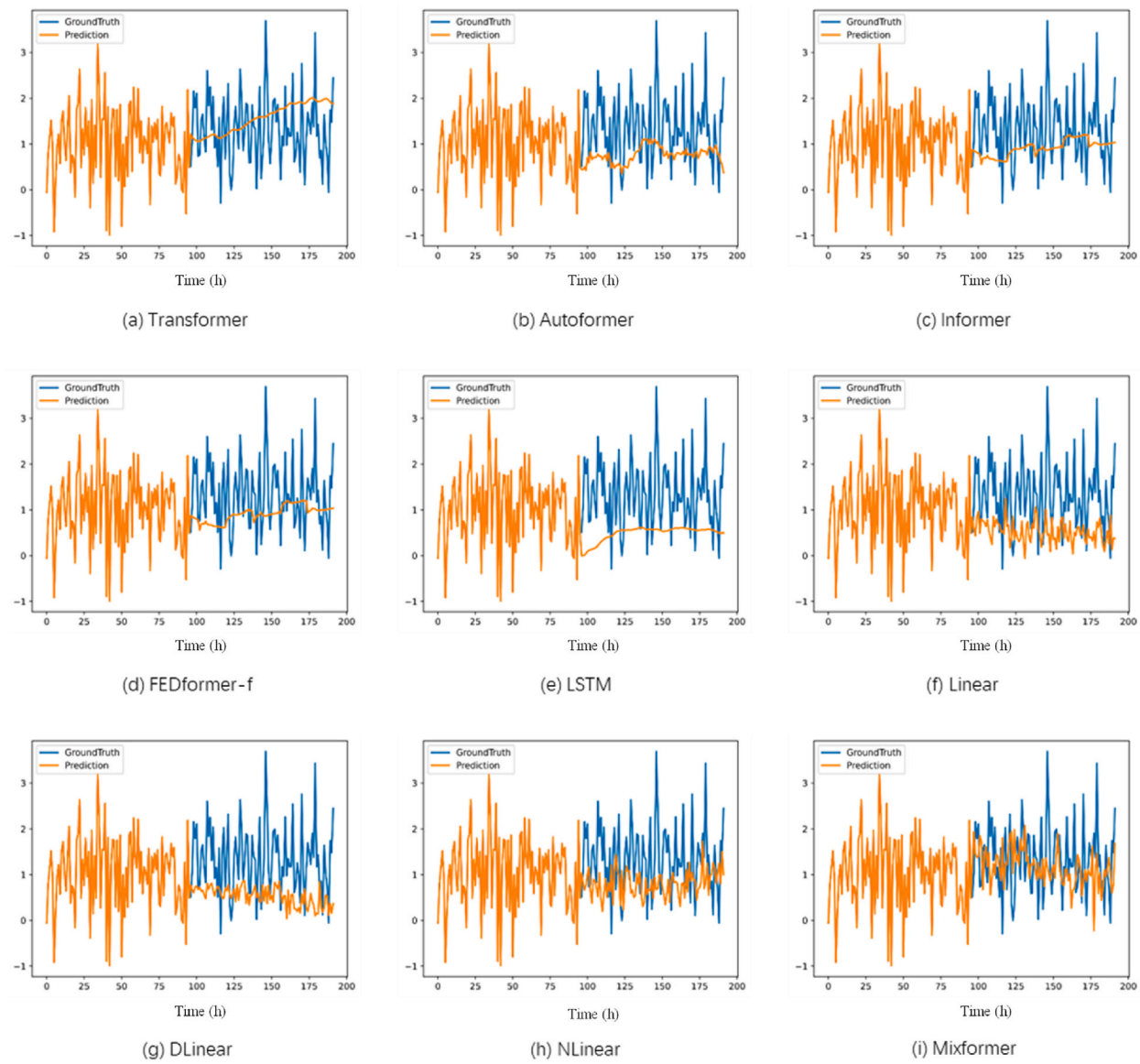
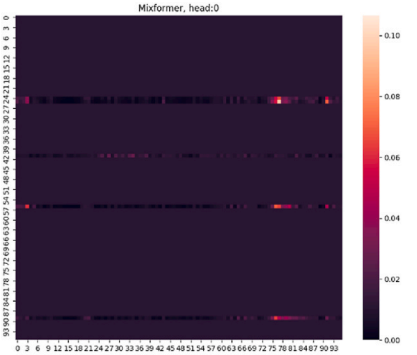
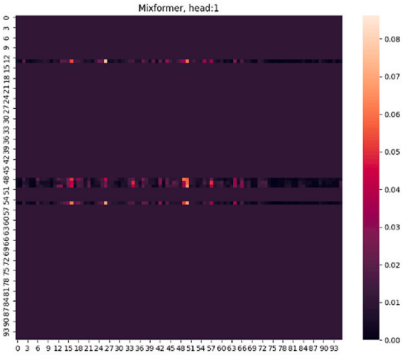


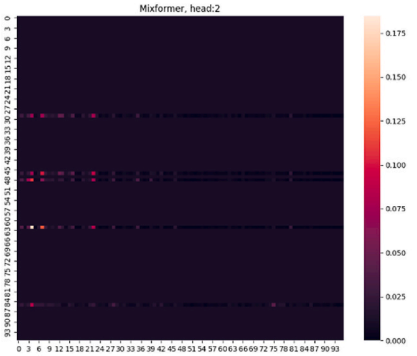
Fig. 8. Comparison of the predicted values of the air-absorbed dose rate with the actual values under different models. Where (a)–(h) are the predicted values of the comparison algorithm, and (i) is the predicted value of Mixformer. The input size of nuclear radiation data is 96h, and the predicted length is 96h.

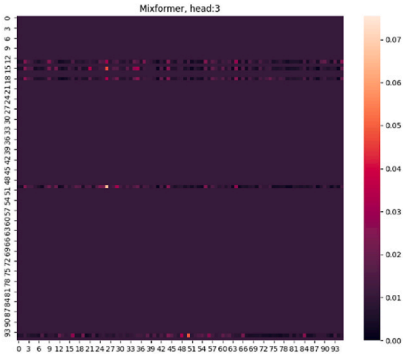
Table 5
Effect of averaging pooling layers in Mixformer on prediction results.

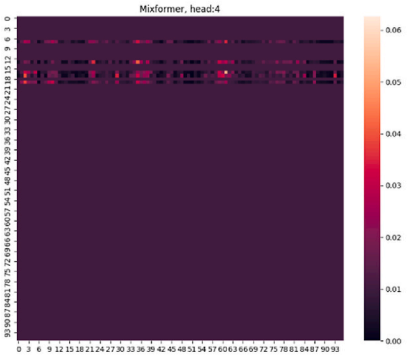
Kernel size	Metrics	Prediction length: 96			Mean
12	MSE	0.7824	0.7715	0.7593	0.7711
	MAE	0.6830	0.6794	0.6758	0.6794
24	MSE	0.7320	0.7344	0.7317	0.7327
	MAE	0.6761	0.6772	0.6760	0.6764
48	MSE	0.7345	0.7347	0.7352	0.7348
	MAE	0.6760	0.6770	0.6781	0.6770
96	MSE	0.7672	0.7485	0.7399	0.7519
	MAE	0.6948	0.6876	0.6845	0.6890
12, 24	MSE	0.7319	0.7335	0.7315	**0.7323**
	MAE	0.6753	0.6784	0.6751	**0.6762**
12, 24, 48	MSE	0.7360	0.7355	0.7369	0.7361
	MAE	0.6891	0.6786	0.6791	0.6823
12, 24, 48, 96	MSE	0.7360	0.7361	0.7369	0.7363
	MAE	0.6862	0.6863	0.6878	0.6868

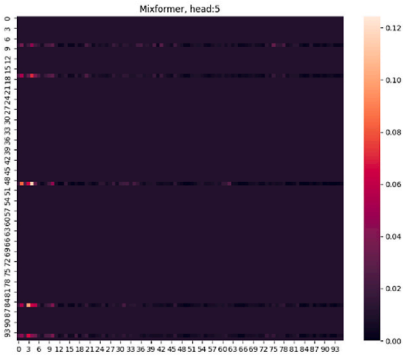


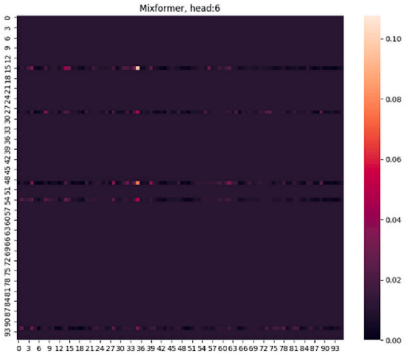


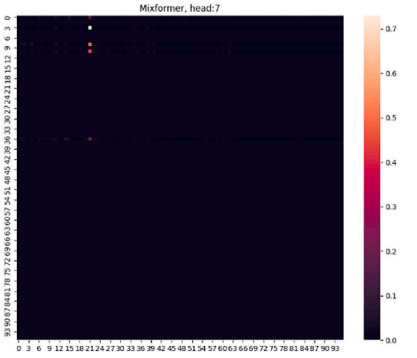












(caption on next page)

Fig. 9. The self-attentive heat map in Mixformer. The lighter part indicates a high correlation between the current two points, and the darker part indicates a low correlation. Mixformer uses a multi-headed attention mechanism with 8 different self-attention.

4.3. Ablation studies

In this section, we performed ablation experiments to verify the validity of the Mixformer model and its variants.

Although the Transformer-based approach performs poorly on nuclear radiation datasets, does the attention mechanism not play a role? Mixformer retains the attention mechanism in Transformer and removes the complex decoding process from the intrinsic structure. To verify the effectiveness of the attention mechanism, we compare other baseline algorithms that do not employ the attention mechanism, such as LSTM, with the latest method Linear for long-time series prediction tasks. The Linear algorithm has outperformed methods such as FEDformer on some public datasets. Two ablation variants of Mixformer were tested: 1) Mixformer without decomposer(Mixformer-V1): removing the decomposer from the encoder layer; 2) Mixformer with MLP(Mixforemer-V2): replacing the convolutional layer in the decomposer with a linear layer.

Two ablation variants of Mixformer were tested: 1) Mixformer without decomposer: removing the decomposer from the encoder layer; 2) Mixformer with MLP: replacing the convolutional layer in the decomposer with a linear layer.

The experimental results of the ablation experiments are shown in Table 4. The results of the Mixformer-V1 experiments show that the decomposer is crucial when processing nuclear radiation data. The decomposer can extract deep seasonal and trend features in complex data. This can also be seen from the results of DLinear in Table 2. Compared with Linear, DLinear with data decomposition is better. The Mixformer model extracts the inherent properties of the data twice using two decomposers in the Encoder layer, which allows for a better understanding of the seasonality and trend of the data. However, the number of decomposers should be manageable. Otherwise, it will easily cause an overfitting phenomenon and reduce the computing speed.In addition, the results of Mixformer-V1 outperform other Transformer class methods, verifying that the complex model does not apply to the prediction of nuclear radiation data. Even without the decomposer, the simplified decoder structure still improves performance and does not easily cause overfitting.

Similarly, Mixformer-V2 does not perform as well as Mixformer, verifying that convolutional layers are more capable of extracting the intrinsic connections of data than linear layers. The linear layer cannot extract spatiotemporal relationships for data sets without obvious periodic patterns.

4.4. Prediction of air-absorbed dose rates

Fig. 6 shows the difference between the predicted and actual values of the future air-absorbed dose rate when the input is a fixed 96 length of time-series data for a more visual comparison of the model predictions.

Because of space limitations, Fig. 8 only shows the results of predicting the future 96-time steps when the input data length is 96. (a)-(d) are the prediction results based on the Transformer and its variants. It can be seen that the Transformer learns the trend change of the data, and the prediction results are close to the mean value from the results. However, the prediction results are too flat and have no ups and downs. Such results are unsatisfactory, although the MSE and MAE appear small. Predicting the increasing or decreasing trend is not accurate enough for predicting nuclear radiation data. (e) is the result of LSTM, similar to the result of the Transformer class method. LSTM also cannot learn the change of data but only mimics the general trend of data. (f)-(h) are the forecasting results of Linear and its variants, which are very different from those of Transformer, as the forecasting results of Linear are more seasonal and less trendy. (i) is the forecast result of Mixformer, which combines the advantages of both methods, learns the seasonal and trend characteristics of historical data, and combines the results of both methods organically.

After sufficient comparative experiments, we found that the attention mechanism can extract the overall trend characteristics from complex nuclear radiation data. When the prediction results have significantly deviated, the self-attentiveness will pull the results toward the correct trend without any obvious misjudgment. The data prediction is also inseparable from the decomposer, which extracts different degrees of seasonal features depending on the size of the convolution kernel. In short, the decomposer allows the model to focus more on small changes, which is crucial for predicting nuclear radiation data because even small changes in air-absorbed dose rates can cause unpredictable safety hazards.

4.5. Convolution kernel analysis

As shown in Section 3.2, the decomposer comprises several averaging pooling layers. By varying the size of the average pooling layers, the effect of different sizes and the number of layers of the average pooling layers on the prediction results can be obtained. The results of the experiments are shown in Table 5.

To simplify the contrast test, we compare the average pooling layer's size and number of layers only when the fixed input time step is 96 and the prediction time step is 96. When using one layer of averaging pooling layers, the convolution kernel works best at one-quarter of the input time step. The size of the convolution kernel should also be proportional to the size of the input data. When the convolution kernel size is the same as the input data size, the MSE is 0.7519, a 2.6% increase compared to size 24. When using multiple layers of averaging pooling layers, more time and computational resources are consumed, although increasing the number of layers is less effective in improving the prediction accuracy. When dealing with complex and irregular air-absorbed dose rate data, too many decomposition layers cause overfitting. The more decomposition layers, the more accurately the detailed features of the data are

Table 6
MSE and MAE results of Mixformer versus other methods for predicting air-absorbed dose rates at Beijing Changchi Road Station. The input sequence length is fixed at 96h, and the predicted time lengths are 48h, 96h, 192h, 336h, and 720h. Each data is sampled at 1h intervals. Each method was run three times, and the average values were given. The best results are highlighted in bold, and the second best is underlined.

Method	Metric	Prediction length: 48			Mean	Prediction length:96			Mean	Prediction length:192			Mean	Prediction length:336			Mean	Prediction length:720			Mean
Transformer	MSE	1.174	1.1312	1.2076	1.1709	1.1766	1.2836	1.1375	1.1992	1.1403	1.1817	1.1666	1.1629	1.1609	1.245	1.2311	1.2123	1.1227	1.1172	1.149	1.1296
	MAE	0.8428	0.8306	0.8554	0.8429	0.8409	0.8796	0.8300	0.8502	0.8206	0.8337	0.8309	0.8284	0.8275	0.8614	0.8537	0.8475	0.8168	0.8167	0.8286	0.8207
Autoformer	MSE	1.0135	1.0369	1.0252	1.0252	1.1506	1.1543	1.0110	1.1053	1.1901	1.1857	1.1545	1.1768	1.1064	1.1789	1.1520	1.1458	\	\	\	\
	MAE	0.7834	0.7912	0.7885	0.7877	0.8347	0.8385	0.7789	0.8174	0.8469	0.8472	0.8361	0.8434	0.8119	0.8390	0.8311	0.8273	\	\	\	\
Informer	MSE	1.0314	1.0290	**0.6982**	0.9195	0.9371	0.9573	1.0715	0.9886	1.0258	0.9916	0.9271	0.9815	0.9743	0.9656	1.1128	1.0176	1.0462	1.0204	1.0688	1.0451
	MAE	0.7814	0.7784	**0.6606**	0.7401	0.7479	0.7575	0.7986	0.7680	0.7796	0.7684	**0.7457**	0.7646	0.7638	0.7588	0.8094	0.7773	0.7944	0.7804	0.8013	0.7920
FEDformer-f	MSE	0.9875	0.9882	0.9858	0.9872	1.1472	1.1412	1.1589	1.1491	1.2912	1.2696	1.2148	1.2585	1.1575	1.1926	1.1909	1.1803	1.2423	1.2480	1.2291	1.2398
	MAE	0.7732	0.7734	0.7721	0.7729	0.8370	0.8349	0.8413	0.8377	0.8940	0.8846	0.8632	0.8806	0.8334	0.8479	0.8470	0.8428	0.8699	0.8718	0.8652	0.8690
LSTM	MSE	1.0320	1.0446	1.0283	1.0350	1.0482	1.0596	1.0485	1.0521	1.1577	1.1601	1.1739	1.1639	1.2165	1.2190	1.2051	1.2135	1.4147	1.3436	1.3905	1.3829
	MAE	0.7835	0.7886	0.7813	0.7845	0.7892	0.7945	0.7879	0.7905	0.8492	0.8493	0.8537	0.8507	0.8695	0.8700	0.8659	0.8685	0.8648	0.8400	0.8546	0.8531
Linear	MSE	1.0080	1.0102	0.9992	1.0058	1.0808	1.0904	1.0814	1.0842	1.1211	1.1202	1.1242	1.1218	0.9577	0.9601	0.9739	0.9639	1.1238	1.1241	1.1234	1.1238
	MAE	0.7833	0.7297	0.7795	0.7642	0.7837	0.8128	0.8099	0.8021	0.8212	0.8210	0.8222	0.8215	0.7492	0.7493	0.7537	0.7507	0.8172	0.8177	0.8174	0.8174
DLinear	MSE	0.9643	**0.8165**	0.8142	0.8650	1.0427	1.0189	1.0184	1.0267	1.0819	1.2479	1.2556	1.1951	1.3952	1.3941	1.3984	1.3959	1.7070	1.7176	1.6968	1.7071
	MAE	0.7648	**0.7124**	0.7115	0.7296	0.7938	0.8027	0.8026	0.7997	0.8057	0.8994	0.9026	0.8692	0.9600	0.9595	0.9612	0.9602	1.0783	1.0826	1.0743	1.0784
NLinear	MSE	1.1100	1.1052	1.0936	1.1029	1.2281	1.2409	1.3012	1.2567	1.2932	1.2951	1.2734	1.2872	1.2765	1.2717	1.2766	1.2749	1.2957	1.2973	1.2985	1.2972
	MAE	0.8247	0.8226	0.8184	0.8219	0.8664	0.8709	0.8904	0.8759	0.8876	0.8883	0.8987	0.8915	0.8762	0.8741	0.8762	0.8755	0.8845	0.8853	0.8857	0.8852
Mixformer	MSE	**0.7931**	0.8979	0.8255	**0.8388**	**0.9040**	**0.9117**	**0.8968**	**0.9042**	**0.9745**	**0.9451**	**0.9199**	**0.9465**	**0.8951**	**0.9299**	**0.9126**	**0.9125**	**0.9906**	**0.9819**	**0.9554**	**0.9760**
	MAE	**0.7087**	0.7511	0.7195	**0.7264**	**0.7532**	**0.7523**	**0.7493**	**0.7516**	**0.7613**	**0.7518**	0.7592	**0.7574**	**0.7364**	**0.7529**	**0.7558**	**0.7484**	**0.7547**	**0.7517**	**0.7446**	**0.7503**

[1] The '\' indicates failure for the out-of-memory.

extracted, so the two-layer decomposer is suitable for the present research object.

4.6. Self-attention analysis

Mixformer consists of an encoder and a decoder, in which the self-attention of the input data needs to be calculated. To verify the effectiveness of the self-attention, Fig. 9 shows how the attention of the Encoder works during the prediction phase. As can be seen in the heat map, each attention head is focused on a different spatiotemporal location. The parts with highlighted colors indicate the high relevance of these time steps, and the parts with dark colors indicate that the attention does not focus on the information at these times. The size of the self-attentive heat map is related to the time step of the input data. When the input step size is N, the size of the self-attentive feature map is $N \times N$. Both horizontal and vertical axes are sorted from 0 to N time steps. Each point in the heat map represents the correlation between the time steps represented by the current coordinates. In head 0, it can be seen that even though points are often less relevant, there are many highlighted points at step 24. It indicates that the model pays more attention to the features of these time points as the data passes through this attention layer. A multiple attention layer will allow the model to focus on different time points. Even if the input is the same data, each attention head layer will extract different relevant features. Finally, the multi-headed attention layer combines the features for output. Therefore, the performance of Mixformer in extracting data features is improved with the multi-head attention mechanism, and the effectiveness of multi-head attention is verified.

5. Conclusion

The air-absorbed dose rate is a crucial nuclear radiation data indicator that provides a direct, rapid, and continuous measure of environmental radiation levels. The amount of radiation exposure to humans and the environment increases with the rate of air absorption dose, which may raise the risk of cancer and other radiation-related disorders, as well as nuclear accidents and environmental nuclear contamination. Therefore, accurate and rapid estimation of air-absorbed dose rates over long periods is critical for human health and environmental protection. This study proposes the Mixformer model and two variants to predict long-time air-absorbed dose rates using real-world data. It solves the problem of fast and accurate prediction of complex nuclear radiation data. The model learns the spatiotemporal relationship between the air-absorbed dose rate and other factors through a multi-headed attention mechanism. Furthermore, the seasonal and trend characteristics of the time-series data are resolved using a decomposer. The model's input variables contain ten observations, and the target output variable is the future air-absorbed dose rate. The mean MSE of the Mixformer model for predicting the future air absorbed dose rate for 96 h at the input step of 96 h is 0.7317, which is better than the SOTA model, including the Transformer class method and Linear class method. The study results show that the Mixformer model has a better predictive capability for nuclear radiation data and can provide a basis for research at nuclear monitoring stations. In addition, Mixformr's long-term prediction of air-absorbed dose rates and the efficiency of its calculations can provide earlier warnings of nuclear accidents and help people implement measures to reduce environmental damage.

Author contribution statement

Yizhi Cao: Conceived and designed the experiments; Performed the experiments; Wrote the paper.
Zhaoran Liu: Analyzed and interpreted the data; Wrote the paper.
Yunlong Niu: Xinggao Liu: Contributed reagents, materials, analysis tools or data.

Data availability statement

The data that has been used is confidential.

Additional information

No additional information is available for this paper.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work is supported by National Key R&D Program of China (Grant No. 2021YFC2101100), National Natural Science Foundation of China (62073288, 12075212, 12105246, 11975207), and their supports are thereby acknowledged.

A APPENDIX.

A.1 EXPERIMENTAL DETAILS

A.1.1 DATASETS

To verify the generalizability of the model, we provide an additional set of experiments in the Appendix section. Different from the dataset from the North Yangtze River station in the main text, the Beijing station has a smaller sample size and is, therefore, more difficult to predict the air absorbed dose rate. In addition to that, the data collected at the Beijing station are affected by a different climate. The monitoring stations provide both datasets. Therefore, they are reliable as experimental data.

A.1.1 RESULTS AND DISCUSSION

The experimental parameters were kept the same as in the text. Table 6 shows the experimental results. Overall, Mixformer achieves the best prediction results in all tasks. Because of the small amount of data in Beijing Changchi Road station, other Transformer class models do not easily lead to GPU memory overflow. When predicting long-time results, the Transformer class model predicts better than the linear class model. Taking the predicted length of 96 timestamps as an example, Mixformer outperforms other Transformer-based algorithms by reducing MSE by 8.5% and MAE by 2.1%, compared to the best-performing Informer model among them. Compared with DLinear, one of the linear-like methods, Mixformer improves MSE by 11.9% and MAE by 6.0%.

The specific prediction results are shown in Fig. 10, where we have selected the results with a prediction length of 96 time-stamps. The Transformer class method and the LSTM fit the flat results, which is unsatisfactory. However, the linear class method predicts more significant variation but inaccurate results. Mixformer combines the advantages of both and achieves the smallest MSE and MAE. Figs. 11 and 12 show the trend of MSE and MAE variation with different prediction lengths. The MSE and MAE variation of Mixformer is smoother compared to other methods.

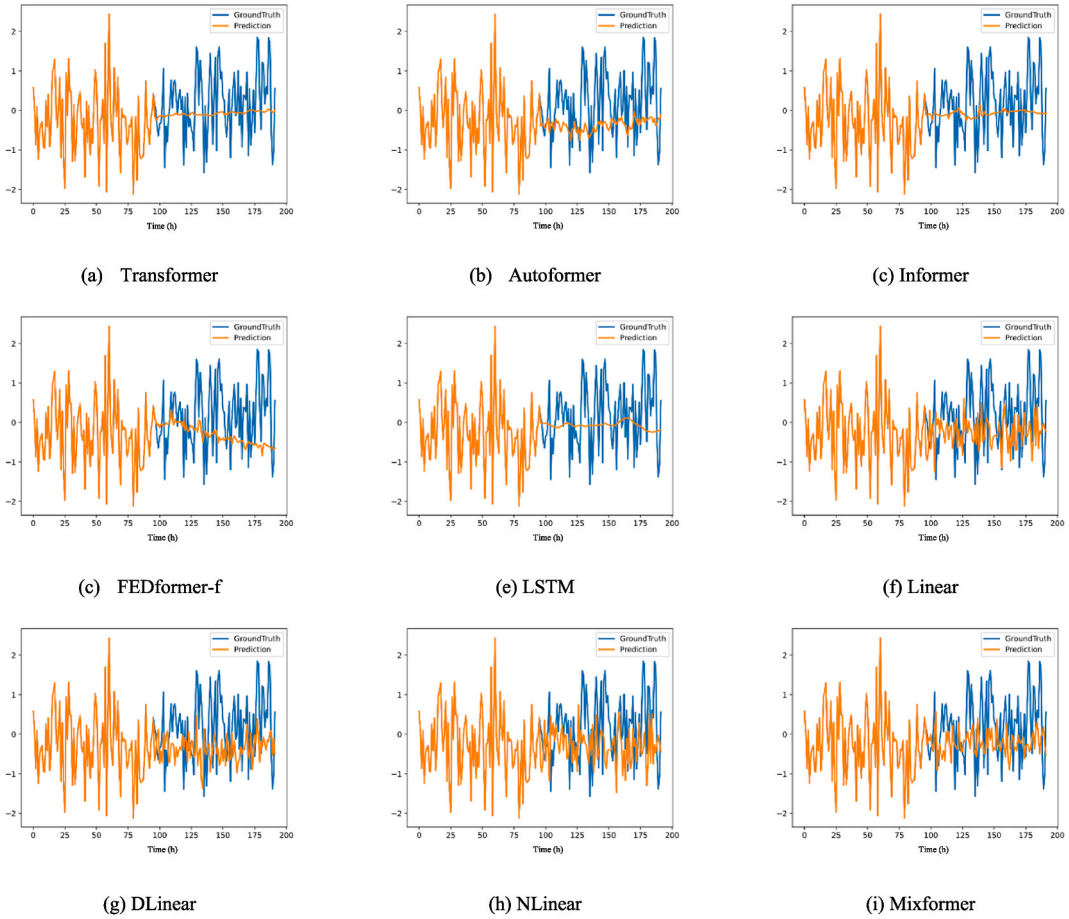


Fig. 10. Comparison of the predicted values of the air-absorbed dose rate with the actual values under different models at Beijing Changchi Road Station. Where (a)–(h) are the predicted values of the comparison algorithm, and (i) is the predicted value of Mixformer. The input size of nuclear radiation data is 96h, and the predicted length is 96h.

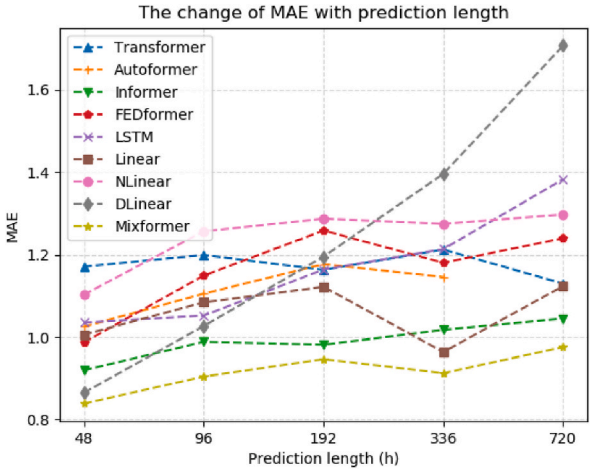


Fig. 11. The change of MSE for the prediction length: {48, 96, 192,336,720}.

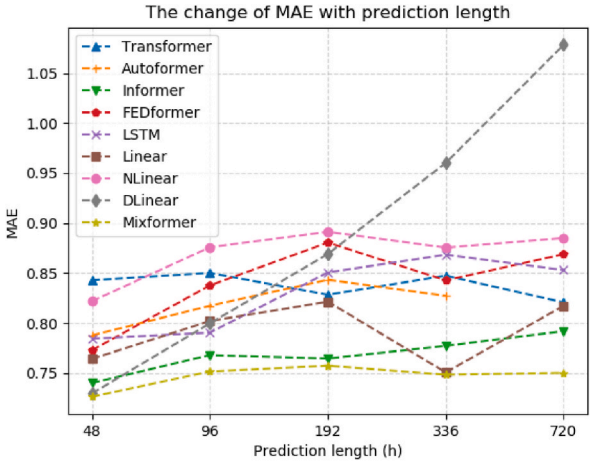


Fig. 12. The change of MAE for the prediction length: {48, 96, 192,336,720}.

References

[1] M.T. Kartal, The role of consumption of energy, fossil sources, nuclear energy, and renewable energy on environmental degradation in top-five carbon producing countries, Renew. Energy 184 (2022) 871–880, https://doi.org/10.1016/j.renene.2021.12.022.
[2] M.D. Mathew, Nuclear energy: a pathway towards mitigation of global warming, Prog. Nucl. Energy 143 (2022), 104080, https://doi.org/10.1016/j.pnucene.2021.104080.
[3] C. Busby, The Wings of Death: Nuclear Pollution and Human Health, Green Audit Books, Aberystwyth, 1995.
[4] F.W. Whicker, V. Schultz, Radioecology: Nuclear Energy and the Environment, CRC press, Boca Raton, FL, 1982.
[5] Z. Li, Y. He, C. Sonne, S.S. Lam, M.B. Kirkham, N. Bolan, J. Rinklebe, X. Chen, W. Peng, A strategy for bioremediation of nuclear contaminants in the environment, Environ. Pollut. (2022), 120964.
[6] W.C. Sailor, D. Bodansky, C. Braun, S. Fetter, B. van der Zwaan, A nuclear solution to climate change? Science 288 (2000) 1177–1178.
[7] M. Usman, M. Radulescu, Examining the role of nuclear and renewable energy in reducing carbon footprint: does the role of technological innovation really create some difference? Sci. Total Environ. 841 (2022), 156662.
[8] M. Hatch, E. Ron, A. Bouville, L. Zablotska, G. Howe, The Chernobyl disaster: cancer following the accident at the Chernobyl nuclear power plant, Epidemiol. Rev. 27 (2005) 56–66.
[9] K. Buesseler, M. Aoyama, M. Fukasawa, Impacts of the Fukushima nuclear power plants on marine radioactivity, Environ. Sci. Technol. 45 (2011) 9931–9935.
[10] V. Saenko, V. Ivanov, A. Tsyb, T. Bogdanova, M. Tronko, Yu Demidchik, S. Yamashita, The chernobyl accident and its consequences, Clin. Oncol. 23 (2011) 234–243, https://doi.org/10.1016/j.clon.2011.01.502.
[11] T. Wada, A. Tomiya, M. Enomoto, T. Sato, D. Morishita, S. Izumi, K. Niizeki, S. Suzuki, T. Morita, G. Kawata, Radiological impact of the nuclear power plant accident on freshwater fish in Fukushima: an overview of monitoring results, J. Environ. Radioact. 151 (2016) 144–155, https://doi.org/10.1016/j.jenvrad.2015.09.017.
[12] Y. Rong, X.S. Liang, A study of the impact of the Fukushima nuclear leak on East China coastal regions, Atmos.-Ocean 56 (2018) 254–267, https://doi.org/10.1080/07055900.2017.1421139.
[13] D.S. Hochbaum, B. Fishbain, Nuclear threat detection with mobile distributed sensor networks, Ann. Oper. Res. 187 (2011) 45–63.

[14] M.-H. Jeong, C.J. Sullivan, S. Wang, Complex radiation sensor network analysis with big data analytics, in: 2015 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC). Presented at the 2015 IEEE Nuclear Science Symposium and Medical Imaging Conference, NSS/MIC), IEEE, San Diego, CA, USA, 2015, pp. 1–4, https://doi.org/10.1109/NSSMIC.2015.7581760.
[15] K. Ogura, M. Hosoda, Y. Tamakuma, T. Suzuki, R. Yamada, R. Negami, T. Tsujiguchi, M. Yamaguchi, Y. Shiroma, K. Iwaoka, N. Akata, M. Shimizu, I. Kashiwakura, S. Tokonami, Discriminative measurement of absorbed dose rates in air from natural and artificial radionuclides in namie Town, Fukushima prefecture, IJERPH 18 (2021) 978, https://doi.org/10.3390/ijerph18030978.
[16] P.R. Miles, J.A. Cook, Z.V. Angers, C.J. Swenson, B.C. Kiedrowski, J. Mattingly, R.C. Smith, Radiation source localization using surrogate models constructed from 3-D Monte Carlo transport physics simulations, Nucl. Technol. 207 (2021) 37–53, https://doi.org/10.1080/00295450.2020.1738796.
[17] N. Tsoulfanidis, S. Landsberger, Measurement & Detection of Radiation, fifth ed., CRC Press, Boca Raton, 2021 https://doi.org/10.1201/9781003009849.
[18] R.R. Goulet, L. Newsome, H. Vandenhove, D.-K. Keum, J. Horyna, S. Kamboj, J. Brown, M.P. Johansen, J. Twining, M.D. Wood, others, Best practices for predictions of radionuclide activity concentrations and total absorbed dose rates to freshwater organisms exposed to uranium mining/milling, J. Environ. Radioact. 244 (2022), 106826.
[19] İ. Aksangür, B. Eren, C. Erden, Evaluation of data preprocessing and feature selection process for prediction of hourly PM10 concentration using long short-term memory models, Environ. Pollut. 311 (2022), 119973, https://doi.org/10.1016/j.envpol.2022.119973.
[20] N. Sarkar, R. Gupta, P.K. Keserwani, M.C. Govil, Air Quality Index prediction using an effective hybrid deep learning model, Environ. Pollut. 315 (2022), 120404, https://doi.org/10.1016/j.envpol.2022.120404.
[21] T. Yan, A. Zhou, S.-L. Shen, Prediction of long-term water quality using machine learning enhanced by Bayesian optimisation, Environ. Pollut. 318 (2023), 120870, https://doi.org/10.1016/j.envpol.2022.120870.
[22] J. Choi, S.J. Lee, Consistency index-based sensor fault detection system for nuclear power plant emergency situations using an LSTM network, Sensors 20 (2020) 1651, https://doi.org/10.3390/s20061651.
[23] M.-D. Wang, T.-H. Lin, K.-C. Jhan, S.-C. Wu, Abnormal event detection, identification and isolation in nuclear power plants using LSTM networks, Prog. Nucl. Energy 140 (2021), 103928.
[24] G. Tsichrintzēs, M. Birbu, L.C. Jain (Eds.), Advances in Machine Learning/deep Learning-Based Technologies, Selected Papers in Honour of Professor Nikolaos G. Bourbakis, Springer, Cham, 2022.
[25] H. Chen, X. He, P. Gao, S. Tan, X. Hou, Automatic scram forecasting during abnormal conditions of nuclear power plants based on LSTM, in: Volume 3: I&C, Digital Control, and Influence of Human Factors. Presented at the 2022 29th International Conference on Nuclear Engineering, American Society of Mechanical Engineers, Virtual, Online, 2022, https://doi.org/10.1115/ICONE29-90219. V003T03A010.
[26] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention Is All You Need, 2017.
[27] W. Zaremba, I. Sutskever, O. Vinyals, Recurrent Neural Network Regularization, 2015.
[28] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (1997) 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735.
[29] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling, 2014 arXiv preprint arXiv: 1412.3555.
[30] T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun, R. Jin, Fedformer: frequency enhanced decomposed transformer for long-term series forecasting, in: International Conference on Machine Learning, PMLR, 2022, pp. 27268–27286.
[31] M. Chen, H. Peng, J. Fu, H. Ling, Autoformer: searching transformers for visual recognition, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 12270–12280.
[32] H. Zhou, Shanghang Zhang, J. Peng, Shuai Zhang, J. Li, H. Xiong, W. Zhang, Informer: beyond efficient transformer for long sequence time-series forecasting, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2021, pp. 11106–11115.
[33] A. Zeng, M. Chen, L. Zhang, Q. Xu, Are Transformers Effective for Time Series Forecasting?, 2022 arXiv preprint arXiv:2205.13504.