Nuclear Science and Techniques (2026) 37:107
https://doi.org/10.1007/s41365-026-01929-y

Optimizing measurements of the rate of radon exhalation from porous media using a hybrid convolutional neural network-transformer model

Feng Xiao[1] · Yu-Xi Xie[1] · Hong-Bo Xu[1] · Chen-Xi Zu[1] · Xian-Fa Mao[1] · Shi-Cheng Luo[1] · Xin-Yue Yang[1] · Hao-Yu You[1] · Hao You[1] · Yi Liu[1] · Cheng Luo[1] · Jia Liu[1] · Hong-Zhi Yuan[1] · Yan-Liang Tan[1]

Received: 2 January 2025 / Revised: 10 March 2025 / Accepted: 10 April 2025 / Published online: 26 March 2026

Abstract
The rate of exhalation of radon from porous materials is a key parameter used to study the migration of radon and to assess environmental radiation hazards. Measurements of radon concentrations are often affected by statistical fluctuations and instrument errors, which lead to reduced measurement accuracy. To address this issue, we propose a deep learning method based on a hybrid model that combines a convolutional neural network with a transformer (CNN-Transformer). The proposed method is designed to optimize radon concentration data and improve the accuracy of radon exhalation rate fitting. The proposed method combines the advantages of one-dimensional convolutional neural networks in local feature extraction with the capabilities of the transformer model to represent time-series data. After training the model on 6 million samples of simulated radon concentration data, it exhibited strong generalization capability and high prediction accuracy across different radon concentration ranges. The optimized results from fitting the curves were close to the theoretical true values, with the coefficient of determination (R^2) remaining stable between 0.97 and 0.99. These results significantly outperformed simulated instrument measurements and reduced the uncertainty of the data. In an experimental evaluation of the performance of the proposed approach for practical applications, optimization predictions were generated by combining the proposed solid radon source exhalation reference model with measurement data collected from the RAD7 radon detector. The fitting results for each experimental group showed a clear improvement. The optimized radon exhalation rate was much closer to the reference value (43 ± 0.48 mBq m^{-2} s^{-1}), with a significant reduction in deviation. The optimized fitting coefficient of determination (R^2) remained stable between 0.988 and 0.996, which was significantly higher than the results obtained from measurement instruments (R^2, which ranged from 0.851 to 0.973). The proposed CNN-Transformer hybrid model provides an innovative approach to optimize data on the rate of radon exhalation from various materials. Thus, the results show that the proposed approach improved measurement accuracy while reducing errors and showed significant potential for practical applications.

Keywords RAD7 · Radon exhalation rate · Radon concentration prediction · CNN-Transformer hybrid model · Solid radon source exhalation reference model

1 Introduction

Radon (Rn-222) is a colorless, odorless, inert gas widely found in rocks, soil, and groundwater. It is a radioactive product of the uranium-238 decay series [1–3]. Owing to its chemical inertness, radon can migrate freely through soil pores and enter building environments through cracks or cavities [4, 5]. Prolonged exposure to radon and its progeny through inhalation into the lungs (in which alpha decay occurs) can damage lung tissue and increase the risk of lung cancer [6, 7]. The International Agency for

This work was supported by Postgraduate Scientific Research Innovation Project of Hunan Province (No. CX20231259), Natural Science Foundation of Hunan Province (No. 2023JJ50091), and Project of Hunan Provincial Department of Education (No. 23A0516).

✉ Yan-Liang Tan
hytyl@163.com

[1] College of Physics and Electronic Engineering, Hengyang Normal University, Hengyang 421008, China

Research on Cancer (IARC) and the World Health Organization (WHO) have both classified radon as the second leading cause of lung cancer after tobacco use [8, 9].

Radon has attracted attention not only owing to its potential health risks but also because of its relevance in fields such as environmental science [10], geological exploration [11], and radioactive contamination assessment [12]. As a natural tracer, radon helps scientists study various environmental processes such as greenhouse gas emissions, atmospheric circulation, groundwater flow, and volcanic activity [13–16]. Accurate measurements of the concentration and exhalation of radon provide crucial data to support the study of these natural phenomena.

The exhalation rate of radon refers to the quantity of radon released per unit area of a given medium per unit time. This metric is critical to assess the hazards associated with radon exposure in various environments where humans are expected to be present. Accurate measurement of the rate of radon exhalation not only helps to identify hazardous areas with high levels of radon but also provides essential data for atmospheric and environmental pollution assessments. Currently, several techniques are commonly used to measure the rate at which radon is exhaled to ensure safety, including the closed-loop method (also known as the accumulation method) [17, 18] as well as an open-loop method [19, 20], activated charcoal adsorption [21, 22] and solid-state nuclear track detectors (SSNTDs) [23, 24]. However, in complex real-world applications, conventional methods still involve some limitations of note in terms of the accuracy of the obtained data and the efficiency of the measurement process. Additionally, the results are often associated with uncertainty due to statistical fluctuations in nuclear measurements and instrument errors. Errors become more pronounced on surfaces with low exhalation rates. This uncertainty can lead to discrepancies between estimated and true radon exhalation rates. Therefore, the development of methods to improve the accuracy with which the rate of radon exhalation can be measured has been recognized as a key issue in research on radon monitoring.

In recent years, with the rapid development of machine learning and deep learning technologies, these techniques have gradually been applied in environmental data analysis. Conventional machine learning methods such as support-vector machines (SVMs) [25], random forest (RF) models [26], and eXtreme gradient boosting (XGBoost) [27], have already shown promising results in anomaly detection and spatial prediction of radon concentration. Using external parameters and manually designed features, machine learning methods can effectively handle the relationships between known variables and thus provide significant support for analyses of the concentration of radon in complex environments.

With the increased scale of data and advancements in computing power, deep learning technologies have attracted increasing attention owing to their powerful capability to automatically extract features and model nonlinear relationships. Huang et al. successfully identified anomalous radon sources and quantified the radon precipitation rate by combining particle swarm optimization (PSO) and a long short-term memory network (LSTM) [28]. McKenzie et al. modeled groundwater tracer radon in coastal areas using a one-dimensional convolutional neural network (CNN), recurrent neural network (RNN) model (1D-CNN-RNN), and fully connected neural network (FCNN) models [29]. Luo et al. proposed a calibration method based on fast local outlier factor (FastLOF) and nonuniform popularity-similarity optimization (NPSO) with backpropagation (BP) (NPSO-BP) neural network models, which significantly improved the accuracy of radon concentration measurements [30]. In addition, a reduced-order model [31] and an intelligent prediction method [32] proposed by He et al. based on deep learning have made significant progress in applications within the nuclear industry. Deep learning has demonstrated great potential for handling complex data and improving the accuracy of radon monitoring. The development of integrated and hybrid tools based on artificial intelligence may offer a new approach for environmental radioactivity modeling. Complex nonlinear relationships can be better captured by combining multiple deep learning models, which can significantly enhance the accuracy and timeliness of the obtained predictions. The development of integrated and hybrid artificial intelligence tools may offer a new approach for environmental radioactivity modeling. By combining multiple deep learning models, complex nonlinear relationships can be captured more effectively, which can significantly enhance prediction accuracy and timeliness.

Temporal variations in radon concentration typically exhibit significant nonlinear characteristics that encompass both local rapid fluctuations (such as statistical noise and instrumental disturbances) and long-term dynamic equilibrium trends (such as the saturation phase of radon exhalation). Conventional methods have both advantages and limitations in modeling such data as time series. Long short-term memory (LSTM) and gated recurrent unit (GRU) models rely on recurrent structures for sequential computation to capture long-term dependencies effectively. However, their sequential nature limits parallelization, which makes them less efficient for large-scale data processing and sensitive to local noise. Pure convolutional neural networks (CNNs) extract features using local receptive fields, which makes them well-suited for short-term pattern recognition. However, their fixed convolutional window constrains their ability to model long-range dependencies. Transformer models excel in global temporal modeling and capture long-range dependencies effectively. However, they may overlook local patterns on short time scales. Consequently, a single deep learning model cannot fully accommodate the complex characteristics of radon concentration data. Thus, leveraging the strengths of different models is crucial to enhance prediction accuracy and stability.

In the field of deep learning, the hybrid models combining CNN structures with transformer models have demonstrated outstanding performance in handling complex tasks, which makes these techniques an effective tool for processing intricate time-series data. For example, Xu et al. achieved efficient real-time semantic segmentation by using a transformer model to extract semantic information during training while using a single-branch CNN for inference [33]. Additionally, Peng et al. proposed a feature-coupling unit that integrates the local feature extraction capability of a CNN with the global representation learning of the transformer to significantly improve image classification and object detection performance [34]. Compared to conventional multilayer perceptrons (MLPs), transformers, and pure CNNs, the CNN-Transformer hybrid model not only retains the ability of CNNs to learn local patterns but also leverages the transformer to capture long-range dependencies for better computational efficiency and parallelization [35, 36].

Based on this background, we propose an optimization method for measured data on radon exhalation rates based on a hybrid model that combines CNN and transformer architectures. The hybrid model combines the advantages of one-dimensional convolutional neural networks (1D-CNNs) in local feature extraction [37, 38] and the relative advantages of transformer models in modeling time-series data [39]. Complex nonlinear relationships can be captured without relying on external parameters by directly modeling variations in time-series data representing the concentration of radon. By learning from a large-scale dataset, the proposed models optimized the measurement results to reduce the uncertainties caused by statistical fluctuations and instrument errors and improve the goodness of fit of the data on exhalation rates.

2 Materials and methods

2.1 Data set acquisition

The core of deep learning technology lies in training models on large datasets to construct accurate and complex predictive models. However, acquiring data on nuclear radiation is challenging, especially with regard to the collection and compilation of large-scale and diverse datasets that conventional methods often fail to meet. Therefore, we adopted a strategy to generate simulated data. A large-scale dataset containing 6 million samples of data on radon concentrations was constructed to train the model based on a numerical model of radon exhalation on the surface of the medium and the principle of statistical fluctuation. The model was tested and validated using actual measurement data.

The measurement chamber of a radon monitor based on the electrostatic collection method (e.g., RAD7) is conventionally used to measure the rate of radon exhalation on the surface of a medium using the accumulation method. During radon decay, the rate at which alpha particles are emitted by ^{218}Po can be expressed by the following formula [40].

$$n(t) = \eta_{\mathrm{Po}} V_{\mathrm{R}} \frac{JS}{\lambda_{\mathrm{eff}} V} \left(1 - \frac{\lambda_{\mathrm{eff}} \mathrm{e}^{-\lambda_{\mathrm{Po}} t} - \lambda_{\mathrm{Po}} \mathrm{e}^{-\lambda_{\mathrm{eff}} t}}{\lambda_{\mathrm{eff}} - \lambda_{\mathrm{Po}}} \right), \quad (1)$$

where $n(t)$ represents the counting rate of alpha particles emitted by ^{218}Po at time t, η_{Po} is the detection efficiency of the planar semiconductor detector in the radon monitor, V_{R} is the volume of the radon monitor chamber in units of m^3, J is the radon exhalation rate on the surface of the medium in units of Bq m^{-2} s^{-1}, S is the area at the bottom of the radon accumulation chamber in units of m^2, V is the volume of the radon accumulation chamber in units of m^3, λ_{eff} is the effective decay constant in units of s^{-1}, which includes the decay constant and leakage constants λ, diffusion constant λ_{b}, and chamber leakage constant λ_{leak}; λ_{Po} is the decay constant of ^{218}Po (with a value of 0.0037 s^{-1}).

Typically, the calibration factor of a radon monitor (i.e., k_{p}) is obtained through a calibration process in a standard radon chamber. Therefore, the true concentration of radon as a function of time can be expressed by the following formula [41].

$$C_{\mathrm{Po}}(t) = n(t) k_{\mathrm{p}}, \quad (2)$$

where $C_{\mathrm{Po}}(t)$ represents the true concentration of radon as a function of time and k_{p} is the calibration factor of the radon monitor.

Assuming that the period of each measurement cycle is T, the rate at which alpha particles are emitted by ^{218}Po during the n-th measurement cycle can be expressed by the following formula.

$$N_{\mathrm{Po}}(nT) = \int_{(n-1)T}^{nT} \eta_{\mathrm{Po}} V_{\mathrm{R}} \frac{JS}{\lambda_{\mathrm{eff}} V} \left(1 - \frac{\lambda_{\mathrm{eff}} \mathrm{e}^{-\lambda_{\mathrm{Po}} t} - \lambda_{\mathrm{Po}} \mathrm{e}^{-\lambda_{\mathrm{eff}} t}}{\lambda_{\mathrm{eff}} - \lambda_{\mathrm{Po}}} \right) \mathrm{d}t, \quad (3)$$

where $N_{\mathrm{Po}}(nT)$ is the number of alpha particles emitted by ^{218}Po during the n-th measurement period.

Based on the above equations, the theoretical true radon concentration values required to train a learning model during the simulation of the measurement cycles can be obtained using the following formula.

$$C_{\mathrm{Po}}(nT) = \frac{k_{\mathrm{p}}}{T}\int_{(n-1)T}^{nT} \eta_{\mathrm{Po}} V_{\mathrm{R}} \frac{JS}{\lambda_{\mathrm{eff}} V} \left(1 - \frac{\lambda_{\mathrm{eff}} \mathrm{e}^{-\lambda_{\mathrm{Po}} t} - \lambda_{\mathrm{Po}} \mathrm{e}^{-\lambda_{\mathrm{eff}} t}}{\lambda_{\mathrm{eff}} - \lambda_{\mathrm{Po}}}\right) \mathrm{d}t = \frac{k_{\mathrm{p}}}{T} N_{\mathrm{Po}}(nT), \tag{4}$$

where $C_{\mathrm{Po}}(t)$ represents the concentration of radon in the theoretical true measurement period.

Two kinds of uncertainty caused by the error are primarily considered in simulating the true radon concentration measured by the measuring instrument, including the statistical fluctuation uncertainty and the measurement error of the instrument itself (±1 counting error).

The uncertainty caused by statistical fluctuations can be derived by calculating the square root of the number of measurement events [42]. For radon measurements with a ^{218}Po count recorded during the n-th measurement cycle of N, a statistical uncertainty of one-sigma is given by $\sqrt{N}$. The following function can be utilized to generate pseudo-random integers using the Poisson distribution, because the output of the measurement instrument is an integer value.

$$N(nT) = \mathrm{RandomInteger}\left[\mathrm{PoissonDistribution}\left[N_{\mathrm{Po}}(nT)\right], 1\right], \tag{5}$$

where $N(nT)$ represents the count rate for alpha particles emitted by ^{218}Po after accounting for statistical fluctuations. PoissonDistribution[$N_{\mathrm{Po}}(nT)$] represents a Poisson distribution, the mean parameter of which is $N_{\mathrm{Po}}(nT)$, which is the true numerical value given by the theory. Using the RandomInteger[.] function, a random integer can be generated from this Poisson distribution to simulate the count fluctuations caused by statistical fluctuations in the measurement.

The measurement error of the instrument itself (±1 count error) also introduces some additional uncertainty that results from a mismatch between the precise timing of the gate opening and closing and the alpha particle events on the detector [43]. A slight misalignment between the timing window of the counter and the occurrence of a particle event can cause a count difference between the readings, which can thus affect the accuracy of the resulting measurements. Therefore, the analog count can be further represented by

$$N_1 = N(nT) + \mathrm{RandomChoice}\left[\{-1, 1\}, 1,\right], \tag{6}$$

where N_1 represents the rate of alpha particles detected in ^{218}Po emissions after incorporating both statistical fluctuations and measurement errors. RandomChoice[$\{-1, 1\}$, 1] indicates a random selection between −1 and 1 to simulate the ±1-count deviation caused by gate timing errors during the measurement process.

Based on Eqs. (1) to (6), the final simulated radon concentration measured by the instrument, considering the effects of statistical fluctuations and measurement errors, can be calculated using the following equation:

$$C(t) = \frac{k_{\mathrm{p}}}{T}\left(N_1 \pm \sqrt{N_1}\right) = \frac{1}{\eta_{\mathrm{Po}} V_{\mathrm{R}} T}\left(N_1 \pm \sqrt{N_1}\right), \tag{7}$$

where $C(t)$ is the radon concentration measured using the simulated instrument.

The number of cycles was set to 12, and the values of key parameters, such as the radon exhalation rate and effective decay constant, were randomly combined as outlined in Table 1. Theoretical true radon concentration values were generated based on Eq. (4), and simulated radon concentration values designed to mimic those obtained in real measurements were generated using Eq. (7). A total of 6 million samples of simulated radon concentration data were generated across a wide range of concentrations from zero to several hundred thousand.

The dataset was divided into training and validation sets with a ratio of 8:2. The training set was used to train the parameters of the model, whereas the validation set was used to evaluate the generalization performance of the model to ensure its predictive capability with input data that were not present in the training dataset. To improve the stability of the learning process and accelerate convergence, the raw radon concentration data were standardized before being input into the model. Standardization was performed by calculating the mean and standard deviation of the sample data and converting each feature into a zero-mean unit-variance form. This process can be expressed as follows.

$$\tilde{x} = \frac{x - \mu}{\sigma}, \tag{8}$$

where $\tilde{x}$ represents the standardized time series of radon concentrations, x is the original time series of radon concentrations, and μ and σ, respectively, represent the overall mean and standard deviation of all the data. This standardization ensured a more stable training process with less instability and helped to improve the convergence speed.

Table 1 Value range of each parameter

Fixed parameter	Value	Random parameter	Value range
η_{Po}	0.3	λ_{eff}	$1 \times 10^{-3} \sim 5 \times 10^{-5}$
V_{R}	0.007	V	$0.001 \sim 0.01$
S	0.04	J	$0.005 \sim 1$
k_{p}	4.7619×10^3	T	240

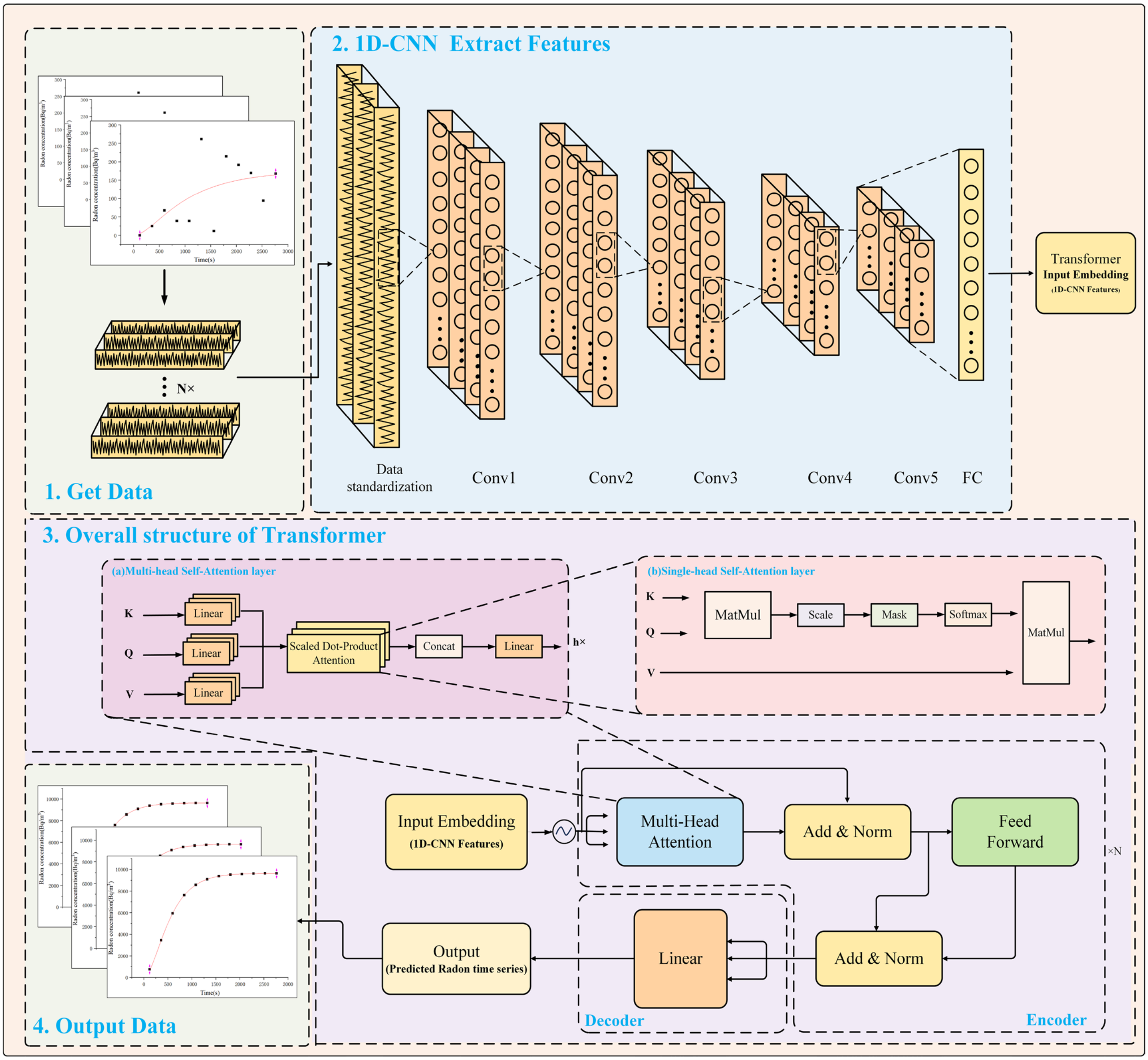

Fig. 1 (Color online) CNN-Transformer hybrid model overall network structure

2.2 Model construction

In the process of measuring radon concentration using the accumulation method, the radon concentration data showed significant nonlinear changes over time, which mainly manifested as a rapid growth stage in the early stage and a stable dynamic equilibrium in the later stage. These complex timing changes are subject to a combination of statistical fluctuations and instrumental errors. To meet the requirements of optimizing radon exhalation rate measurement data, we propose a deep learning framework based on a CNN-Transformer hybrid model to optimize radon concentration data and improve the accuracy with which measured radon exhalation rate can be fit to the expected data for a given condition. The model was trained iteratively on a desktop system equipped with an NVIDIA RTX 3060 GPU. The implementation was developed using the Python 3.9 programming language using the PyTorch library, and the source code was managed and developed using PyCharm.

The proposed CNN-Transformer hybrid model adopts a concatenated fusion approach [44], fully combines the advantages of local feature extraction and global time series modeling, and predicts radon concentrations optimally based on an input time series. The network structure of the model is shown in Fig. 1. In this hybrid model, the 1D-CNN extracts local features from the time series through sliding window operations and multilayer convolutional stacking. This is particularly effective in capturing dynamic change patterns during the initial rapid growth phase to provide high-quality input for global modeling. The transformer module uses the multi-head self-attention mechanism to model long- and short-term dependence relationships of the time series from a global perspective. Especially in the dynamic equilibrium stage, the results show that the transformer model was able to capture the global characteristics of the change in the concentration becoming stable accurately, and it further improved the fitting accuracy and modeling capability of existing methods.

2.2.1 One-dimensional convolutional neural network (1D-CNN)

The core task of the 1D-CNN module is to extract key local features from the input radon concentration time series to provide basic support for improving the radon exhalation rate-fitting accuracy. Through the convolution operation of the sliding window, 1D-CNN is able to capture local patterns in the time series, such as the change trend and fluctuation characteristics of the radon concentration over time. Furthermore, the introduction of the ReLU activation function enhances the nonlinear expression capability of the network so that it can learn complex data relations more effectively. This module consists of five convolutional layers and one fully connected layer (FC).

Convolutional layers are the core component of 1D-CNN models, and their main function is to extract local features from input data through a sliding window operation. The convolution operation enables local perception of an input sequence by sliding a window across it. The extent of each local perception is determined by the size of the convolution kernel. The model progressively extracts local dynamic features of the time series, ranging from short-term fluctuations to high-level feature patterns, by stacking multiple convolutional layers. For the input time-series $\mathbf{X} = [x_1, x_2, \dots, x_T]$ of data on radon concentrations, each layer of the convolution operation can be expressed as:

$$Z_i = \mathrm{Conv1D}\left(Z_{i-1}, W_i, \mathrm{b}_i\right), \tag{9}$$

where Z_i is the output feature map of the i-th layer, and the initial input $Z_0 = \mathbf{X}$ is the radon concentration time series, and W_i and b_i, respectively, denote the weight and bias parameters of the convolutional kernel. These parameters were dynamically adjusted during training to adapt to varying data distributions to enhance the capability of the model to capture a diverse range of feature patterns.

The output of the convolution operation is processed by the nonlinear activation function ReLU to introduce nonlinear capability, the formula of which is given as follows.

$$A_i = \mathrm{ReLU}\left(Z_i\right), \tag{10}$$

where A_i represents the high-dimensional features of the convolutional stack from the i-th layer. The ReLU activation function is defined as $\mathrm{ReLU}(x) = \max(0, x)$, which means that for any input value x, if $x > 0$, the output is x, and if $x \leq 0$, the output is 0. The introduction of ReLU effectively increases the nonlinear representation capability of the model, enables the network to capture key patterns from complex feature relationships, reduces the problem of gradient disappearance, and further improves the training efficiency and feature extraction capability of the network.

With the progressive processing of multilayer convolution and activation operations, the feature dimension gradually expands, and the network can extract higher-level patterns and structures. After all the convolution operations are completed, the convolution stacked high-dimensional feature An (the output of the last convolution layer) is input to the Fully Connected layer (FC). The fully connected layer integrates these features through linear transformations and maps them to the target space as expressed by the following formula:

$$Y = W_{\mathrm{FC}} \cdot A_n + b_{\mathrm{FC}}, \tag{11}$$

where W_{FC} and b_{FC} represent the weight and bias parameters of the fully connected layer, respectively. Y is the feature representation generated after the processing in the previous layer. This was used as the input for the subsequent transformer module to provide a high-quality feature basis for global modeling. This approach ensures that both long- and short-term dependencies, as well as the global trends of the data, can be captured more accurately in the subsequent steps.

2.2.2 Transformer

In this study, the primary role of the transformer module was to perform global temporal modeling on local feature sequences extracted by a one-dimensional convolutional neural network. The core design of the transformer aims to capture the long-range dependencies between different time steps in the input data, thereby generating features with a global representation. This module comprises two key components: an encoder stack based on a self-attention mechanism and a simplified decoder output structure.

The transformer model included a four-layer encoder stack, with each layer comprising a multi-head self-attention mechanism, a feedforward neural network (FFN), and layer normalization.

The multi-head self-attention mechanism updates the feature representations by calculating the correlation between each time step and the other time steps in the feature sequence. The formula for each attention head is given as follows.

$$\mathrm{Attention}(Q, K, V) = \mathrm{Soft\,max}\left(\frac{QK^{\mathrm{T}}}{\sqrt{d_{\mathrm{k}}}}\right)V, \tag{12}$$

where the result of the self-attention mechanism is denoted as attention (Q, K, V), Q represents the query matrix, K represents the key matrix, and V represents the value matrix. Softmax(·) denotes the normalization applied to the scaled results to obtain the weight matrix. K^{T} is the transpose of the key matrix used to calculate the dot-product similarity between the query and key vectors, whereas d_{k} represents the dimensions of the keys that control the scaling of the values. By performing parallel computations across multiple attention heads and concatenating the results, the model learns to capture the complex interactions between different time steps in the data.

After the attention mechanism, the features of each time step pass through a two-layer feedforward neural network (FFN), the nonlinear transformation of which is expressed as follows.

$$\mathrm{FFN}(x) = \mathrm{ReLU}\left(xW_1 + b_1\right)W_2 + b_2, \tag{13}$$

where FFN(x) represents the output of the feedforward neural network, x represents the input features from the previous layer, W_1 and W_2 are the weight matrices for the feedforward network, and b_1 and b_2 are the corresponding bias terms. The FFN enables a nonlinear mapping of the time-step features to enhance the model's expressive power and introduce greater complexity.

Layer normalization improves the stability of the gradient flow by standardizing the inputs and outputs of each layer. The formula is given as follows.

$$\mathrm{LayerNorm}(x) = \frac{x - \mu}{\sigma} \cdot \gamma + \beta, \tag{14}$$

where LayerNorm(x) represents the output of the layer normalization and x is the output from the previous layer, which serves as the input feature for the current layer, μ is the mean of the input vector x and σ is the standard deviation, γ is the learned scale parameter used to scale the normalized result, and β is the learned bias parameter used for the translation of the normalization results. Layer normalization significantly accelerates the convergence of the model while enhancing its adaptability to different data distributions and improving its overall robustness.

In contrast to the conventional complex transformer decoder, we designed a simplified decoder structure specifically tailored to the requirements of radon concentration time-series data by removing the conventional autoregressive mechanism. Because the input features have already been fully extracted by the 1D-CNN and transformed into a nonsequential structure, the original sequence-to-sequence structure is simplified to nonsequential sequence mapping. Consequently, the decoder consists of a single linear layer that directly maps the high-dimensional feature output from the encoder to the target space. The formula is given as follows.

$$\widehat{Y} = W_{\mathrm{out}}H + b_{\mathrm{out}}, \tag{15}$$

where $\widehat{Y}$ represents the final prediction result of the hybrid model, H is the feature matrix output of the encoder, W_{out} and b_{out} are the weights and biases of the linear layer, respectively. Through this simplified design, the model retains the capability of time-series modeling, avoids the high computational consumption of the autoregressive step in a conventional decoder, and significantly reduces the complexity and computational cost of the model.

2.3 Training process and performance evaluation

In this study, the training process of the CNN-Transformer hybrid model used the mean squared error (MSE) as the loss function to measure the difference between the predicted and true values. The formula is given as follows.

$$L(\theta) = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2, \tag{16}$$

where $L(\theta)$ represents the loss function, i.e., the mean squared error (MSE), θ denotes the set of model parameters, such as the weights and biases of the neural network, N is the number of samples in the predicted radon concentration time series; $\hat{y}_i$ represents the predicted value of the radon concentration time series, and y_i represents the true value of the radon concentration time series. MSE effectively reflects the deviation between the predicted results and true observations. By minimizing MSE, the parameters of the model can be optimized to improve the accuracy of the obtained predictions.

During the optimization process, a stochastic gradient descent (SGD) algorithm is used as the base optimization method [45]. The SGD algorithm iteratively computes the gradient of the loss function and updates the parameters of the model to reduce the loss and enhance the learning capability of the model. However, relying solely on SGD may lead to overfitting, especially with complex models. Therefore, we incorporated an L2 regularization (weight decay) term into the loss function to limit the excessive growth of the parameters and control the complexity of the model [46]. The final optimization objective function is expressed as follows.

$$L_{\mathrm{data}}(\theta) = L(\theta) + \lambda \sum_{i=1}^{N} \theta_i^2, \tag{17}$$

where $L_{\mathrm{data}}(\theta)$ represents the optimized total loss function; λ is the regularization strength coefficient, which controls the impact of the L2 regularization on the total loss, $\sum_{i=1}^{N} \theta_i^2$ denotes the L2 regularization term (also known as weight decay), which is the sum of the squares of all model parameters used to penalize excessively large parameter values in the model. By limiting the growth of parameter values, L2 regularization reduces the complexity of the model and helps prevent overfitting.

In each epoch, the update rule for the parameters is given as follows.

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta L_{\mathrm{data}}(\theta), \tag{18}$$

where θ_t and θ_{t+1} represent the current and next rounds of parameters, η is the learning rate that controls the step size of each parameter update, and $\nabla_\theta L_{\mathrm{data}}(\theta)$ is the gradient of the loss function with respect to the model parameters. When L2 regularization is introduced, the update rule also includes the gradient of the regularization term to ensure that the number of parameters does not increase excessively.

We also adopted a cosine-annealing learning rate scheduler with a warm-up strategy to further enhance the stability of the training process and accelerate the convergence [47]. This combined approach optimizes the adjustment of the learning to enable the model to update parameters more efficiently during different stages of training.

During the initial few training epochs, a warm-up strategy was applied that gradually increased the learning rate from a small initial value to the preset initial learning rate. This phase was designed to stabilize the parameter updates and prevent instability caused by an excessively large learning rate. By gradually increasing the learning rate, the model can enter an appropriate parameter space smoothly. The learning rate follows a linear growth pattern expressed by the following formula.

$$\eta_t = \eta_{\min} + \frac{t}{T_{\mathrm{warmup}}} \cdot \left(\eta_0 - \eta_{\min}\right), \tag{19}$$

where η_t represents the learning rate of the t-th cycle, $\eta_{\min}$ is the initial minimum learning rate, η_0 is the initial learning rate, and T_{warmup} is the total number of cycles in the warm-up phase.

After the warm-up phase, the learning rate decayed periodically according to a cosine function, starting from the initial learning rate η_0 and gradually decreasing toward the set minimum learning rate $\eta_{\min}$. This approach enabled the model to fine-tune the parameters more precisely in the later stages of training. The updated formula for the cosine-annealing learning rate scheduler is given as follows.

Table 2 Optimal performance indicators of CNN-Transformer hybrid model under 1000 epochs

Performance index	Training set	Validation set
Loss	0.002096	0.002093
RMSE (Bq m^{-3})	171.0977	171.893
MAPE (%)	0.0569	0.057
MAE (Bq m^{-3})	103.1781	103.5404
R^2	0.9979	0.9979
r	0.9996	0.9996

$$\eta = \eta_{\min} + \frac{1}{2}\left(\eta_0 - \eta_{\min}\right)\left(1 + \cos\left(\frac{t - T_{\mathrm{warmup}}}{T_{\mathrm{cosine}}}\pi\right)\right), \tag{20}$$

where η represents the learning rate of each period in the cosine-annealing stage, and T_{cosine} is the total number of periods in the cosine-annealing phase.

The combination of the warm-up and cosine-annealing strategies for learning rate adjustment ensured stable training in the early stages to prevent rapid parameter updates that could lead to poor convergence. In the later stages, the gradual decay of the learning rate enhanced the precision of the optimization, which helped to avoid local minima and ultimately improved the overall effectiveness of the training process and the performance of the model.

After the training process was completed, in addition to the evaluation dependent on the loss function, a series of comprehensive evaluation indicators were introduced to comprehensively measure the performance of the model and ensure its performance in terms of correction effect and generalization capability. These metrics included the mean absolute percentage error (MAPE), root mean square error (RMSE), mean absolute error (MAE), coefficient of determination (R^2), and correlation coefficient (r). By evaluating the predictive performance of the model from multiple perspectives, these metrics effectively validate the accuracy and reliability of the model in predicting the concentration of radon to provide important references for further optimization.

3 Results and discussion

In this study, the CNN-Transformer hybrid model was trained for 1000 epochs, with optimal results achieved at the 990th epoch. The specific performance metrics of the training and validation sets are listed in Table 2. Figure 2 presents the changes in the loss function for the training and validation sets with respect to the number of epochs during the training process. This intuitively illustrates the training process of the model and its convergence characteristics.

As indicated by Table 2 and Fig. 2, the performance of the CNN-Transformer hybrid model was highly consistent across both the training and validation sets, with no signs of overfitting. The loss of this model, root mean square error (RMSE), mean absolute percentage error (MAPE), and mean absolute error (MAE) were all relatively low, indicating high prediction accuracy and stability. Additionally, the coefficient of determination (R^2) and correlation coefficient (r) were close to the ideal values, further confirming the strong linear correlation between the predicted and actual values. During the training process, the loss value decreases gradually, especially in the early stage, which indicates that the model can quickly capture the main features of the data. In the subsequent fine-tuning phase, the loss reduction becomes more gradual, reflecting the model's optimization of the finer details. Overall, the gap between the training and validation losses was always small, and the model had a good generalization capability.

To further verify the performance of the CNN-Transformer hybrid model in the optimization and prediction of radon concentration data, this study adopted the above-mentioned empirical formula to generate additional radon concentration data as a test set, which was not included in the training or validation sets. After using the trained CNN-Transformer model to predict and optimize the test set data, the performance evaluation results showed that the mean absolute percentage error (MAPE), root mean square error (RMSE), and MAE were 0.320%, 306.49 Bq m^{-3}, and 165.01 Bq m^{-3}, respectively. Compared to the training and validation sets, the error of the test set was improved, which may be related to the small amount of test data and the diversity of the data distribution in different concentration ranges. However, the determination coefficient ($R^2 = 0.9979$) and correlation coefficient ($r = 0.9999$) of the model were close to the ideal values, indicating that the fitting capability and prediction accuracy of the model on the test data remained high.

To optimize the radon concentration data and improve the radon exhalation rate-fitting accuracy, further analysis was conducted on the data from different concentration ranges in the test set. Nonlinear fitting of the simulated instrument measurement data, model-predicted optimized data, and theoretical true data was performed using Origin software, with the fitting performed according to Eq. (2). The coefficients of determination (R^2) and other related performance metrics are summarized in Table 3, and Fig. 3 displays the curve plots after fitting, illustrating the variation in the radon concentration over time.

As indicated by Table 3 and Fig. 3, the CNN-Transformer hybrid model demonstrates strong predictive capabilities across different ranges of radon concentration data. In terms of the fitting effect of the radon exhalation rate, the optimized data of the model prediction are superior to the simulated instrument measurement data; particularly, the optimization effect was more obvious at higher concentrations, and the fitted curve was closer to the theoretical real value. Moreover, the coefficient of determination (R^2) for all predicted data exceeded that of the simulated

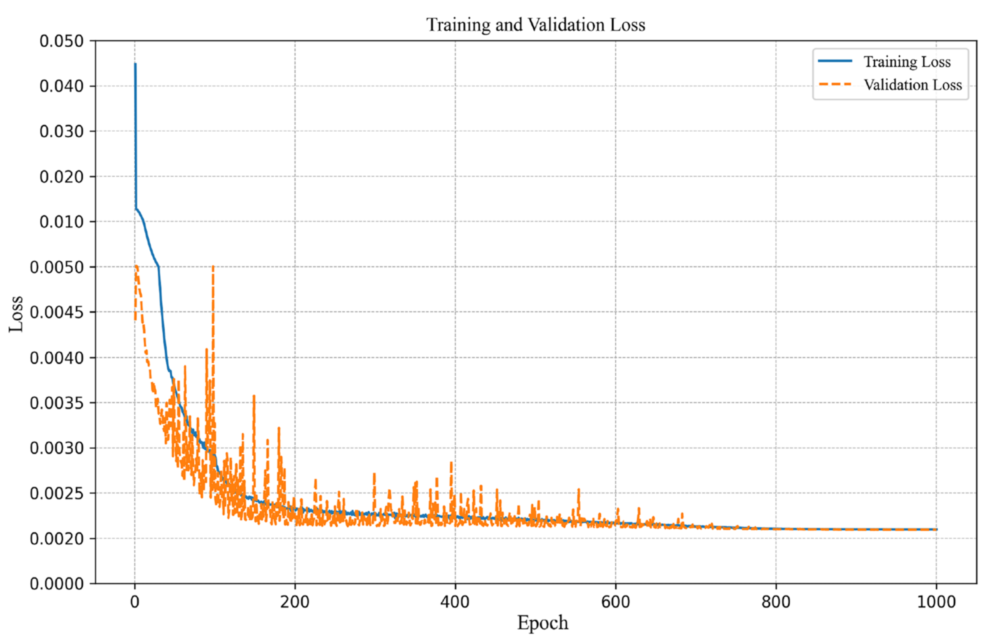

Fig. 2 (Color online) CNN-Transformer hybrid model training and verification loss curve

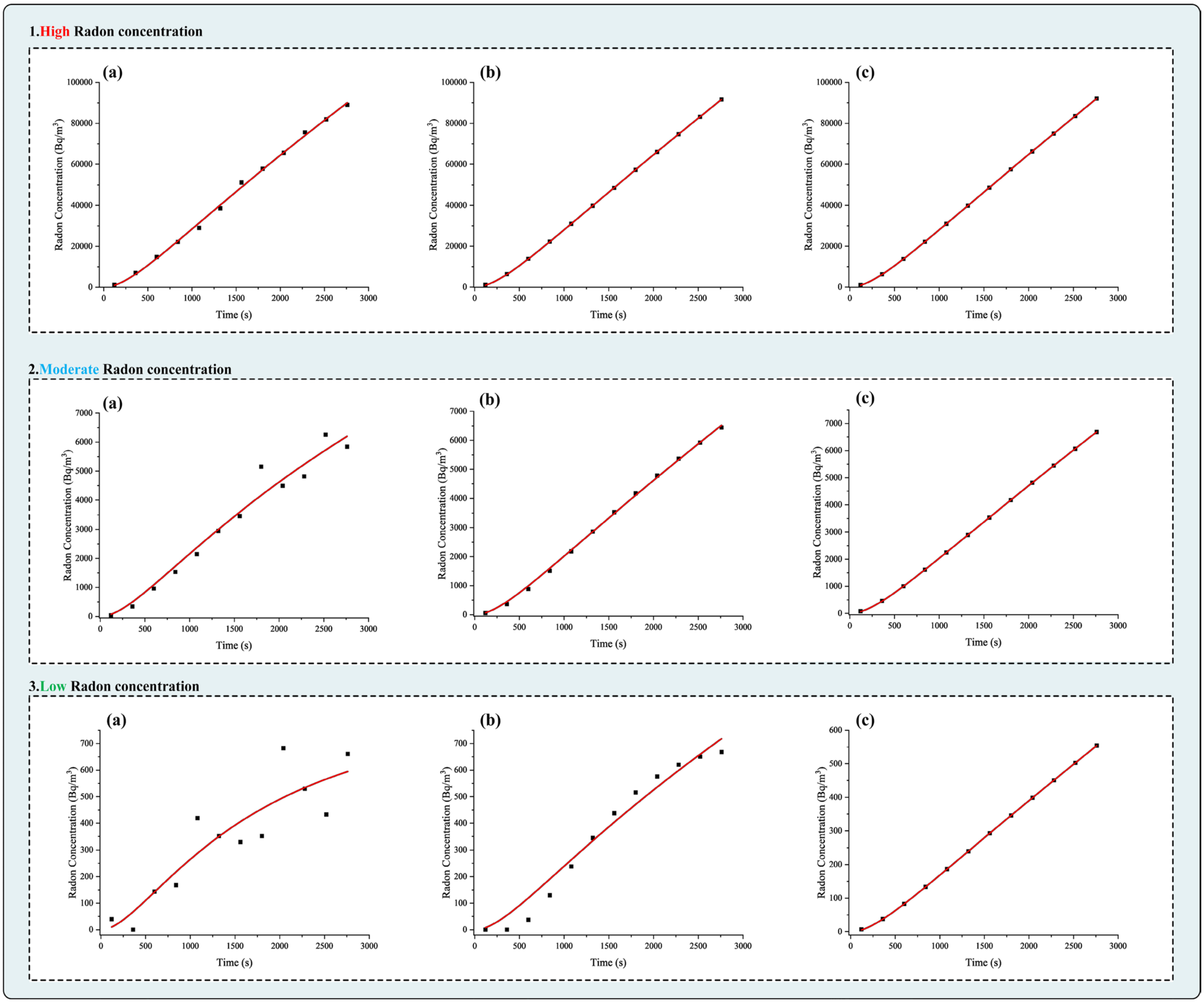

Fig. 3 Nonlinear least squares curves of the data before and after the modification of the model compared with the theoretical values for three randomly selected groups with different levels of radon concentrations (fitting based on Eq. (2)): **a** Simulated instrument measurement data; **b** data predicted by the hybrid model; **c** simulation of theoretical real data

Table 3 (Color online) Nonlinear fitting results of data and theoretical values before and after model modification within one standard deviation ($\pm 1\sigma$) for three randomly selected groups with different radon concentration levels

Experimental class	Simulated instrument measurement data			Hybrid model-predicted data			Simulation of theoretical true data		
	$J \pm \sigma$ (mBq m^{-2} s^{-1})	$\lambda_{eff} \pm \sigma$ ($\times 10^{-5}$) (s^{-1})	R^2	$J \pm \sigma$ (mBq m^{-2} s^{-1})	$\lambda_{eff} \pm \sigma$ ($\times 10^{-5}$) (s^{-1})	R^2	$J \pm \sigma$ (mBq m^{-2} s^{-1})	$\lambda_{eff} \pm \sigma$ ($\times 10^{-5}$) (s^{-1})	R^2
High	997.18±30.78	7.90±3.13	0.998	966.88±2.29	3.97±0.24	0.999	964.61±1.63	3.38±0.17	0.999
Moderate	79.61±10.5	20.46±14.23	0.966	70.01±1.50	5.41±2.14	0.999	70.04±0.12	3.38±0.17	0.999
Low	11.25±3.39	57.25±39.11	0.820	8.61±1.12	14.52±13.62	0.970	5.80±0.01	3.38±0.17	0.999

Table 4 Nonlinear fitting results of RAD7 measured radon concentration data before and after optimization

Experimental Number	Instrument measurement data			Hybrid model-predicted data		
	$J \pm \sigma$ (mBq m^{-2} s^{-1})	$\lambda_{eff} \pm \sigma$ ($\times 10^{-4}$) (s^{-1})	R^2	$J \pm \sigma$ (mBq m^{-2} s^{-1})	$\lambda_{eff} \pm \sigma$ ($\times 10^{-4}$) (s^{-1})	R^2
1	45.72 ± 4.13	2.18 ± 0.45	0.973	44.51 ± 1.90	2.04 ± 0.21	0.994
2	43.95 ± 10.14	2.18 ± 1.14	0.851	43.33 ± 1.97	2.12 ± 0.22	0.993
3	44.08 ± 7.82	1.86 ± 0.84	0.911	43.03 ± 2.68	1.72 ± 0.29	0.988
4	57.29 ± 7.91	3.38 ± 0.79	0.937	48.02 ± 2.13	2.39 ± 0.22	0.993
5	35.76 ± 8.39	0.25 ± 0.90	0.875	45.60 ± 1.61	1.38 ± 0.16	0.996
6	51.19 ± 10.90	2.90 ± 1.15	0.872	46.22 ± 1.92	2.40 ± 0.21	0.994

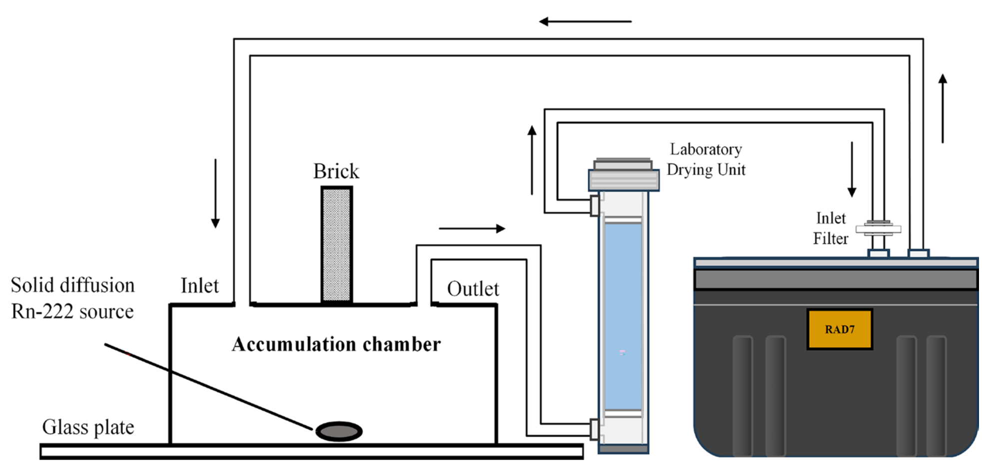

Fig. 4 (Color online) Scheme diagram for measuring radon concentration with reference model

instrument measurement data, with R^2 values consistently ranging from 0.97 to 0.99. Thus, the results confirm the high stability and reliability of the proposed model in processing different concentration data.

To further validate the optimization performance of the model in practical applications, this study employed a solid radon source exhalation reference model [48] with a reference radon exhalation rate of 43 ± 0.48 mBq m^{-2} s^{-1}. The test setup and the connections of the experimental instruments are shown in Fig. 4. The experimental apparatus primarily included a drying tube, filter membrane, brick, glass plate, RAD7 radon detector, a radon accumulation chamber, and a solid radon source. These components were connected and sealed with silicone tubing to ensure that no gas leaked during the experiment.

In practical measurements, a set of measurements is usually taken in ten-minute cycles to minimize the experimental error, rather than the four-minute cycles used to train the model. Therefore, to enhance the practical applicability of the model, radon concentration data were collected over a period of 10 min. Data acquisition was performed using a RAD7 instrument combined with an accumulation method. A total of six sets of repeated measurements were carried out to ensure the reliability of experimental results. Subsequently, the designed CNN-Transformer deep learning hybrid model was applied to optimize and predict the

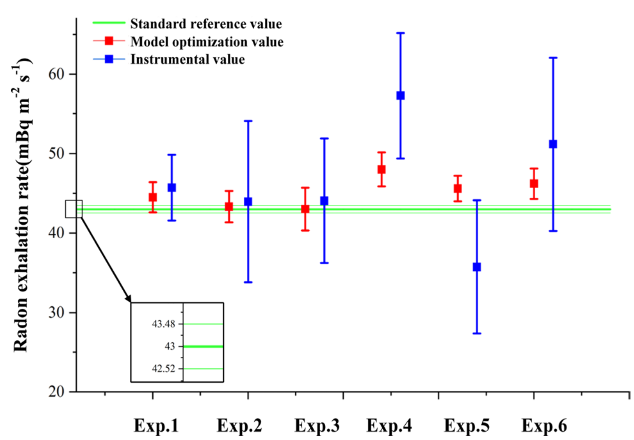

Fig. 5 (Color online) Radon exhalation rate obtained by nonlinear fitting of different measurement data: Green line: reference value of solid radon source exhalation model; Red line: CNN-Transformer hybrid model optimization value; Blue line: Instrument measurement value

measurement data. A nonlinear fitting analysis of the measurement results was conducted using the Origin software, and the fitting results are listed in Table 4.

As shown in Fig. 5, after optimization and prediction using the CNN-Transformer hybrid model, the fitting results for each experimental group showed a significant improvement. The optimized radon exhalation rate (J) was much closer to the reference value, with a significant reduction in deviation. Furthermore, the optimized fitting coefficients (R^2) were generally higher than those for the unoptimized data, ranging from 0.988 to 0.996, which significantly outperformed the fitting results of the instrument measurements (R^2, which ranged from 0.851 to 0.973). These results indicate that the performance of the proposed CNN-Transformer hybrid model in terms of optimization exhibited high stability and reliability in practical tests. This effectively reduced the uncertainty in the experimental data and thus improved the accuracy with which the radon exhalation rate was fitted to the data.

4 Conclusion

In this study, we have proposed a deep learning approach based on a CNN-Transformer hybrid model to optimize radon concentration data and enhance the accuracy of radon exhalation rate fitting. The model demonstrated high stability and good generalization capability during both training and validation, with strong learning capabilities. In the testing phase, the optimized radon exhalation rate closely matched the theoretical values, and the coefficient of determination (R^2) remained stable between 0.97 and 0.99, which outperformed the simulated instrument measurement data. The model exhibited high prediction accuracy and stability across a range of different radon concentrations, with the optimization effect becoming more pronounced as the radon concentration increased.

In the practical application test, the solid radon source exhalation reference model, combined with the measurement data collected by the RAD7 radon detector, was used to optimize the prediction. The fitting results from the various experimental groups showed significant improvements, with the optimized radon exhalation rate (J) approaching the reference value (43 ± 0.48 mBq m^{-2} s^{-1}), and the deviation was notably reduced. Further, the optimized R^2 values were stable between 0.988 and 0.996, which was significantly better than the instrumental measurement data (R^2 ranging from 0.851 to 0.973). These results indicate that the CNN-Transformer hybrid model offers high stability and reliability in optimizing radon exhalation rate measurement data. The proposed approach reduced the impact of statistical fluctuations and instrument errors and significantly improved the accuracy with which the obtained predictions of radon exhalation rates fit the measured data.

Author Contributions All authors contributed to the study conception and design. Feng Xiao was involved in conceptualization, methodology, and writing—original draft. Yu-Xi Xie and Hong-Bo Xu contributed to writing—review and editing and methodology. Chen-Xi Zu and Xian-Fa Mao were involved in resources and data curation. Shi-Cheng Luo and Xin-Yue Yang contributed to supervision and validation. Hao-Yu You and Hao You were involved in investigation and visualization. Yi Liu and Cheng Luo contributed to software and visualization. Jia Liu and Hong-Zhi Yuan were involved in formal analysis and validation. Yan-Liang Tan contributed to funding acquisition, project administration, and writing—review and editing. Feng Xiao wrote the first draft of the manuscript. All authors commented on previous versions and approved the final manuscript.

Data Availability The data that support the findings of this study are openly available in Science Data Bank at https://cstr.cn/31253.11.sciencedb.j00186.00957 and https://www.doi.org/10.57760/sciencedb.j00186.00957.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. C. He, Z. Zeng, L. Zhang et al., A new-designed system for continuous measurement of radon in water. Appl. Radiat. Isot. **187**, 110320 (2022). https://doi.org/10.1016/j.apradiso.2022.110320
2. Q.J. Guo, Research progress on environmental radionuclides by the Radiation Protection Research Group. Chin. Sci. Bull. **68**, 1104–1111 (2023). https://doi.org/10.1360/TB-2022-1118. (in Chinese)
3. M. Xia, Y.J. Ye, S.Y. Liu, Numerical simulations for radon migration and exhalation behavior during measuring radon exhalation rate with closed-loop method. Nucl. Sci. Tech. **35**, 9 (2024). https://doi.org/10.1007/s41365-024-01362-z
4. W.C. Ding, Y.J. Wang, F. Fang et al., A practical soil radon (^{222}Rn) measurement method. Nucl. Sci. Tech. **21**(3), 182–186 (2010). https://doi.org/10.13538/j.1001-8042/nst.21.182-186
5. N. Wang, J. Yang, H. Wang et al., Characteristics of indoor and soil gas radon, and discussion on high radon potential in Urumqi, Xinjiang, NW China. Atmosphere **14**(10), 1548 (2023). https://doi.org/10.3390/atmos14101548
6. J. Yang, L. Zhang, K. Abdumomin et al., Study on peak shape fitting method in radon progeny measurement. Radiat. Prot. Dosim. **167**(1–3), 115–120 (2015). https://doi.org/10.1093/rpd/ncv226
7. V. Zag'a, M.S. Cattaruzza, P. Martucci et al., The "polonium in vivo" study: polonium-210 in bronchial lavages of patients with suspected lung cancer. Biomedicines **9**(1), 4 (2020). https://doi.org/10.3390/biomedicines9010004
8. V.J. Cogliano, R. Baan, K. Straif et al., Preventable exposures associated with human cancers. J. Natl Cancer Inst. **103**(24), 1827–1839 (2011). https://doi.org/10.1093/jnci/djr483
9. J.Y. Yoon, J.D. Lee, S.W. Joo et al., Indoor radon exposure and lung cancer: a review of ecological studies. Ann. Occup. Environ. Med. **28**, 15 (2016). https://doi.org/10.1186/s40557-016-0098-z
10. P. Huang, W. Lv, R. Huang et al., Impact of environmental factors on atmospheric radon variations at China Jinping Underground Laboratory. Sci. Rep. **14**(1), 31402 (2024). https://doi.org/10.1038/s41598-024-82936-0
11. H. Kang, Y.L. Chen, D.P. Li et al., Deep-penetrating geochemistry for concealed sandstone-type uranium deposits: A case study of Hadatu uranium deposit in the Erenhot Basin. North China. J. Geochem. Explor. **211**, 106464 (2020). https://doi.org/10.1016/j.gexplo.2020.106464

12. Y. Lou, Y. Liu, G. Peng et al., Radioactivity risk assessment of radon and gamma dose at one uranium tailings pond in China. Paper presented at the IOP Conference Series: Earth and Environmental Science (Xi'an, China, 28–30 Dec. 2018)
13. I. Levin, U. Karstens, S. Hammer et al., Limitations of the radon tracer method (RTM) to estimate regional greenhouse gas (GHG) emissions–a case study for methane in Heidelberg. Atmos. Chem. Phys. **21**(23), 17907–17926 (2021). https://doi.org/10.5194/acp-21-17907-2021
14. K. Zhang, H. Wan, M. Zhang et al., Evaluation of the atmospheric transport in a GCM using radon measurements: sensitivity to cumulus convection parameterization. Atmos. Chem. Phys. **8**(10), 2811–2832 (2008). https://doi.org/10.5194/acp-8-2811-2008
15. X. Lu, C. Wu, X. Gong et al., Lake-groundwater recharge fluxes during dry season in plain lakeland based on radon isotopes. J. Lake Sci. **36**(3), 939–950 (2024). https://doi.org/10.18307/2024.0345. (in Chinese)
16. Y. Zhang, Y. Chen, D. Li et al., Tracing sources of geochemical anomalies in a deeply buried volcanic-related hydrothermal uranium deposit: the Daguanchang deposit, Northern Hebei Province. North China Craton. J. Earth Sci. **35**(4), 1186–1195 (2024). https://doi.org/10.1007/s12583-021-1597-6
17. S. Liu, T. Hu, F. Lin et al., A new model to accurately measure the radon exhalation rate from soil using AlphaGUARD. J. Environ. Radioact. **266**, 107226 (2023). https://doi.org/10.1016/j.jenvrad.2023.107226
18. L. Zhang, X. Lei, Q. Guo et al., Accurate measurement of the radon exhalation rate of building materials using the closed chamber method. J. Radiol. Prot. **32**(3), 315 (2012). https://doi.org/10.1088/0952-4746/32/3/315
19. Y. Tan, D. Xiao, H. Yuan et al., A method to simultaneously and continuously measure the ^{222}Rn and ^{220}Rn exhalation rates of soil in an open loop. Isot. Environ. Health Stud. **50**(4), 531–537 (2014). https://doi.org/10.1080/10256016.2014.916705
20. S. Liu, F. Lin, Z. Fan et al., Comparison of radon exhalation rate measurements on reference device in open and closed loop by AlphaGUARD in flow-through mode. Radiat. Prot. Dosim. **199**(10), 1151–1157 (2023). https://doi.org/10.1093/rpd/ncad063
21. Y.C. Song, J.L. Wang, B. Shang et al., Study on a new charcoal closed chamber method for measuring radon exhalation rate of building materials. Radiat. Meas. **134**, 106308 (2020). https://doi.org/10.1016/j.radmeas.2020.106308
22. L. Lv, Z. He, S. Qiu et al., Evaluation and measurement methods for the surface radon exhalation rate of buildings. Indoor Built Environ. **31**(10), 2378–2385 (2022). https://doi.org/10.1177/1420326X221109754
23. Z. Fan, J. Shan, F. Lin et al., Developing a radon monitor for simultaneous measurement of Rn-222 and Rn-220 with less influence of humidity based on electrostatic collection and CR-39 detector. Nucl. Instrum. Methods Phys. Res. Sect. A **1052**, 168285 (2023). https://doi.org/10.1016/j.nima.2023.168285
24. D.H. Fan, W.H. Zhuo, B. Chen et al., Uncertainty of an automatic system for counting alpha tracks on CR-39. Nucl. Sci. Tech. **28**(11), 164 (2017). https://doi.org/10.1007/s41365-017-0314-8
25. E. Petermann, H. Meyer, M. Nussbaum et al., Mapping the geogenic radon potential for Germany by machine learning. Sci. Total Environ. **754**, 142291 (2021). https://doi.org/10.1016/j.scitotenv.2020.142291
26. D. Vienneau, S. Boz, L. Forlin et al., Residential radon–Comparative analysis of exposure models in Switzerland. Environ. Pollut. **271**, 116356 (2021). https://doi.org/10.1016/j.envpol.2020.116356
27. M. Rafique, A.D.K. Tareen, A.A. Mir et al., Delegated regressor, a robust approach for automated anomaly detection in the soil radon time series data. Sci. Rep. **10**(1), 3004 (2020). https://doi.org/10.1038/s41598-020-59881-9
28. D. Huang, Y. Liu, Y. Liu et al., Identification of sources with abnormal radon exhalation rates based on radon concentrations in underground environments. Sci. Total Environ. **807**, 150800 (2022). https://doi.org/10.1016/j.scitotenv.2021.150800
29. T. McKenzie, H. Dulai, J. Lee et al., Using deep learning to model the groundwater tracer radon in coastal waters. Water Resour. Res. **59**(3), e2022WR033870 (2023). https://doi.org/10.1029/2022WR033870
30. Q.B. Luo, L. Li, Y.X. Yang et al., Research on radon concentration measurement value correction based on FASTLOF and NPSO-BP neural network model. Radiat. Meas. **177**, 107257 (2024). https://doi.org/10.1016/j.radmeas.2024.107257
31. S. He, M. Wang, J. Zhang et al., A deep-learning reduced-order model for thermal hydraulic characteristics rapid estimation of steam generators. Int. J. Heat Mass Transfer **198**, 123424 (2022). https://doi.org/10.1016/j.ijheatmasstransfer.2022.123424
32. S. He, Y. Ye, M. Wang et al., A machine learning and CFD based approach for fouling rapid prediction in shell-and-tube heat exchanger. Nucl. Eng. Des. **432**, 113759 (2025). https://doi.org/10.1016/j.nucengdes.2024.113759
33. Z. Xu, D. Wu, C. Yu et al., SCTNet: Single-branch CNN with transformer semantic information for real-time segmentation. Paper presented at the 38th AAAI Conference on Artificial Intelligence (Vancouver, Canada, 20–27 Feb. 2024)
34. Z. Peng, W. Huang, S. Gu et al., Conformer: Local features coupling global representations for visual recognition. Paper presented at the IEEE/CVF International Conference on Computer Vision (ICCV 2021) (Montréal, Canada, 11–17 Oct. 2021)
35. Y. Zhao, G. Wang, C. Tang et al., A battle of network structures: An empirical study of CNN, Transformer, and MLP (2021). arXiv2108.13002
36. J. Fang, H. Lin, X. Chen et al., A hybrid network of CNN and transformer for lightweight image super-resolution. Paper presented at the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022) (New Orleans, Louisiana, USA, 18–24 Jun. 2022)
37. S. Kiranyaz, O. Avci, O. Abdeljaber et al., 1D convolutional neural networks and applications: A survey. Mech. Syst. Signal Process. **151**, 107398 (2021). https://doi.org/10.1016/j.ymssp.2020.107398
38. S. Guessoum, S. Belda, J.M. Ferrandiz et al., The short-term prediction of length of day using 1D convolutional neural networks (1D CNN). Sensors **22**(23), 9517 (2022). https://doi.org/10.3390/s22239517
39. A. Vaswani, Attention is all you need. Adv. Neural Inf. Process. Syst. (2017). arXiv:1706.03762
40. Y. Tan, S. Tokonami, M. Hosoda, On the calibration of a radon exhalation monitor based on the electrostatic collection method and accumulation chamber. J. Environ. Radioact. **144**, 9–14 (2015). https://doi.org/10.1016/j.jenvrad.2015.02.012
41. F. Lin, R. Xie, Z. Fan et al., Comparison of two models for measuring radon exhalation rate with a ventilation chamber by numerical simulation. J. Instrum. **17**(10), P10007 (2022). https://doi.org/10.1088/1748-0221/17/10/P10007
42. S. Pomme, R. Fitzgerald, J. Keightley, Uncertainty of nuclear counting. Metrologia **52**(3), S3 (2015). https://doi.org/10.1088/0026-1394/52/3/S3
43. Electronics Notes. Frequency Counter Accuracy: Resolution & Errors, https://www.electronics-notes.com/articles/test-methods/frequency-counter-timer/accuracy-resolution-errors.php [accessed 25 December 2024]
44. A. Khan, Z. Rauf, A. Sohail et al., A survey of the vision transformers and their CNN-transformer based variants. Artif. Intell. Rev. **56**(Suppl 3), 2917–2970 (2023). https://doi.org/10.1007/s10462-023-10595-0

45. Y. Tian, Y. Zhang, H. Zhang, Recent advances in stochastic gradient descent in deep learning. Mathematics **11**(3), 682 (2023). https://doi.org/10.3390/math11030682
46. R. Moradi, R. Berangi, B. Minaei, A survey of regularization strategies for deep models. Artif. Intell. Rev. **53**(6), 3947–3986 (2020). https://doi.org/10.1007/s10462-019-09784-7
47. O.V. Johnson, X.Y. Chew, K.W. Khaw et al., ps-CALR: Periodic-Shift Cosine Annealing Learning Rate for Deep Neural Networks. IEEE Access. (2023). https://doi.org/10.1109/ACCESS.2023.3340719
48. H. Li, R. Xie, Z. Fan et al., Verification of the applicability of the RAD7 radon monitor to the solid radon source exhalation reference model. Nucl. Eng. Technol. **57**, 103341 (2025). https://doi.org/10.1016/j.net.2024.11.043