TYPE Original Research
PUBLISHED 08 August 2025
DOI 10.3389/fmed.2025.1620257

OPEN ACCESS

EDITED BY
Ruida Hou,
St. Jude Children's Research Hospital,
United States

REVIEWED BY
Zhen Li,
Shenzhen MSU-BIT University, China
Yun Xin Teoh,
Sunway University, Malaysia

*CORRESPONDENCE
Kun Zhu
✉ bravo619@163.com

RECEIVED 29 April 2025
ACCEPTED 21 July 2025
PUBLISHED 08 August 2025

CITATION
Xu N, Yang G, Ming L, Dai J and Zhu K (2025) PSOA-LSTM: a hybrid attention-based LSTM model optimized by particle swarm optimization for accurate lung cancer incidence forecasting in China (1990–2021). *Front. Med.* 12:1620257.
doi: 10.3389/fmed.2025.1620257

PSOA-LSTM: a hybrid attention-based LSTM model optimized by particle swarm optimization for accurate lung cancer incidence forecasting in China (1990–2021)

Nannan Xu[1], Guang Yang[2], Linlin Ming[3], Jiefei Dai[3] and Kun Zhu[3]*

[1]Qiqihar First Hospital/Qiqihar Hospital Affiliated to Southern Medical University, Clinical Laboratory, Qiqihar, China, [2]Qiqihar First Hospital/Qiqihar Hospital Affiliated to Southern Medical University, Oral and Maxillofacial Surgery, Qiqihar, China, [3]The Third Affiliated Hospital of Qiqihar Medical College, Chest Surgery, Qiqihar, China

Background: Accurate forecasting of lung cancer incidence is crucial for early prevention, effective medical resource allocation, and evidence-based policymaking.

Objective: This study proposes a novel deep learning framework—PSOA-LSTM—that integrates Particle Swarm Optimization (PSO) with an attention-based Long Short-Term Memory (LSTM) network to enhance the precision of lung cancer incidence prediction.

Methods: Using the Global Burden of Disease 2019 (GBD 2019) dataset, the model predicts age- and gender-specific lung cancer incidence trends for the next 5 years. The proposed model was compared against traditional models including ARIMA, standard LSTM, Support Vector Regression (SVR), and Random Forest (RF).

Results: The PSOA-LSTM model achieved superior performance across five key evaluation metrics: mean squared error (MSE) = 0.023, coefficient of determination (R^2) = 0.97, mean absolute error (MAE) = 0.152, normalized root mean squared error (NRMSE) = 0.025, and mean absolute percentage error (MAPE) = 0.38%. Visualization results across 12 age groups and both genders further validated the model's ability to capture temporal trends and reduce prediction error, demonstrating enhanced generalization and robustness.

Conclusion: The proposed PSOA-LSTM model outperforms benchmark models in predicting lung cancer incidence across demographic segments, offering a reliable decision-support tool for public health surveillance, early warning systems, and health policy formulation.

KEYWORDS

lung cancer, healthcare forecasting, LSTM, attention mechanism, particle swarm optimization, time-series prediction

1 Introduction

Lung cancer is one of the deadliest cancers worldwide. Its incidence rate continues to rise, placing a heavy burden on public health systems. Predicting the long-term incidence trends of lung cancer across different age groups has become an important reference for disease warning, resource allocation, and prevention strategies (1). However, lung cancer incidence data exhibit strong time series characteristics and nonlinear fluctuations. Developing accurate and interpretable prediction models remains a key challenge (2).

In research on lung cancer incidence prediction, time series modeling methods have evolved continuously from traditional linear statistical models to machine learning and deep learning approaches. Early studies often used linear statistical methods such as the autoregressive integrated moving average (ARIMA) model. These methods are transparent in structure and easy to compute. They achieved good results when the data were relatively stationary (3–5). However, the incidence of lung cancer is influenced by multiple factors, including population aging, environmental exposures, and smoking behavior. These factors result in complex nonlinear growth, cyclical fluctuations, and differences across age groups. Therefore, traditional linear models face serious limitations in predictive performance under such conditions (6).

To address these issues, nonlinear machine learning methods such as support vector regression (SVR) and random forest (RF) have been introduced in medical prediction tasks (7, 8). These methods improve the model's ability to fit complex nonlinear patterns and have shown certain success in short-term prediction. However, they usually ignore the temporal dependencies in data, treating time series as unordered samples. As a result, it is difficult for them to model long-term dynamic processes (9–11).

With the development of deep learning, long short-term memory (LSTM) networks have become one of the main methods for medical time series prediction because of their strength in modeling long-term dependencies (12–14). LSTM uses gating mechanisms to retain important historical information and has been widely applied in medical fields such as chronic disease progression and epidemic forecasting (15–18). However, standard LSTM models assign equal weights to all time steps in the input sequence. This may cause the model to overlook critical periods, which can reduce prediction accuracy (19, 20).

The introduction of the attention mechanism helps to alleviate this problem to some extent (21). When the attention mechanism is integrated into the LSTM model, the model can assign higher weights to key time points in the input sequence. This improves its ability to recognize critical information and enhances model interpretability (22–24). Existing studies have shown that the Attention-LSTM structure outperforms the traditional LSTM model in predicting various disease risks. It also provides significant advantages in model transparency and clinical interpretability (25).

Nevertheless, the current Attention-LSTM models are still highly sensitive to hyperparameter settings, such as attention dimension, number of hidden layers, and learning rate (26). Manual tuning of these parameters is costly and can easily lead to underfitting or overfitting.

In recent years, particle swarm optimization (PSO), as a typical swarm intelligence optimization algorithm, has been increasingly applied to hyperparameter tuning in deep learning models (27). Compared to traditional grid search and random search, PSO offers stronger global search capability, faster convergence, and easier implementation. It is especially suitable for optimization problems in high-dimensional parameter spaces (28). Previous studies have successfully applied PSO optimization in tasks such as stroke prediction and lung function modeling, which has significantly improved model accuracy and stability (29, 30).

However, to date, there is still a lack of research that effectively combines the time modeling power of LSTM, the feature focusing ability of the attention mechanism, and the structural optimization strength of the PSO algorithm for lung cancer incidence prediction (31). Existing models find it difficult to simultaneously satisfy the requirements of nonlinear modeling, time dependency modeling, and automatic parameter tuning (32, 33). Therefore, this study proposes a particle swarm optimized attention-LSTM prediction model (PSOA-LSTM). By introducing the attention mechanism into the LSTM structure to strengthen modeling of critical time periods and using PSO for hyperparameter optimization, the model's prediction accuracy and robustness are improved. This research aims to provide an effective solution for modeling complex medical time series data, integrating accuracy, stability, and interpretability.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 introduces the experimental data, the structure of the proposed model, and the evaluation metrics. Section 4 describes the experimental design and performance evaluation. Section 5 discusses the model's performance, strengths and weaknesses, application prospects, and possible limitations. Section 6 concludes the paper and outlines future research directions.

2 Related work

Accurate prediction of cancer incidence is crucial for public health planning. Early studies mainly adopted traditional linear statistical models such as ARIMA due to their interpretability and computational simplicity. For example, Langat et al. (34) applied the ARIMA model to forecast cancer incidence in Kenya and found it effective for short-term prediction of relatively stable univariate series. Kong et al. (35) used an ARIMA-based approach for healthcare data prediction, confirming its utility for regular time series but noting its limited adaptability to structural changes and nonlinear patterns. With the increasing complexity of cancer epidemiological data, machine learning methods have been introduced. Ahmed et al. (36) compared several supervised learning algorithms for lung cancer classification using multi-dimensional datasets, demonstrating that machine learning models can improve prediction accuracy over traditional statistical approaches. Tuncal et al. (2) evaluated several machine learning algorithms for lung cancer incidence prediction and found that RF and SVR outperformed classical models in capturing complex nonlinear relationships. Wu et al. (37) further used random forest modeling to analyze lung cancer mortality associated with risk factors on a global scale, highlighting its effectiveness in variable selection and pattern recognition. More recently, deep learning models have gained attention for their ability to model long-term dependencies and handle high-dimensional data. Khan and Jie (38) developed an LSTM model to predict cancer incidence and mortality, reporting significant improvements in predictive accuracy compared to traditional and machine learning methods. Liu et al. (39) introduced an LSTM neural network combined with improved PSO and attention mechanisms for time series prediction in environmental monitoring, showing that the integration of attention and intelligent optimization substantially enhances model performance and robustness.

However, there remains a lack of studies that systematically integrate LSTM, attention mechanisms, and PSO-based hyperparameter optimization for age- and sex-stratified lung cancer incidence prediction using Global Burden of Disease(GBD) datasets. Most existing works either focus on traditional or machine learning models or lack benchmarking on stratified, real-world data. In response, this study proposes and systematically compares a PSOA-LSTM framework with representative models from the literature (ARIMA, SVR, RF, LSTM), providing an evaluation of its advantages and practical value in cancer incidence forecasting.

3 Materials and methods

To further validate the advantages of the reviewed methods and address the task of lung cancer incidence prediction, we designed a multi-sequence, attention-augmented PSOA-LSTM model to forecast the age-standardized incidence rate (ASIR) of lung cancer over the next 5 years. The model architecture consists of a sliding window input layer, an LSTM encoder, an attention mechanism, a fully connected output layer, and PSO hyperparameter optimization. This section introduces the data sources, model structure, evaluation metrics, and the overall algorithmic workflow.

3.1 Data source

This study obtained ASIR data for lung cancer in China from 1990 to 2021 using the GBD 2021 project through the GHDx platform (http://ghdx.healthdata.org/gbd-results-tool). The data are grouped by sex (male, female) and 5-year age intervals (40–44, 45–49, ..., 90–94, ≥95 years). ASIR represents the number of new cases per 100,000 people in each age group each year. The dataset provides annual estimates, covering 32 years, two sexes, and 12 age groups, for a total of 768 samples (2 × 12 × 32). Each record contains a unique ASIR value for a specific year, sex, and age group. This type of data can reflect risk differences among sexes and age groups, and provides an accurate basis for building time series models.

3.2 Model architecture

3.2.1 Sliding window input layer

Multi-sequence inputs are derived from 24 ASIR sub-series (by sex and age group), and a sliding window is used to extract the most recent 10 years of data ($w = 10$), resulting in an input dimension of (10, 24). The raw data undergoes normalization to ensure that the input values are within a similar scale, improving the model's convergence and stability. The data normalization process is given by:

$$X' = \frac{X - \mu}{\sigma} \tag{1}$$

where X is the original data, μ is the mean, and σ is the standard deviation. This ensures that all features contribute equally to the model, avoiding issues related to large variations in data values.

3.2.2 LSTM encoder

In this model, a single-layer LSTM encoder is responsible for transforming the 10-year sliding window of historical lung cancer ASIR multi-sequence data (dimension: 10, 24) into structured hidden representations with strong temporal dependencies. The core mechanism includes the input gate, forget gate, and output gate. These gating structures allow the model to selectively retain or discard information at each time step based on the input and previous hidden state, thus stably capturing long-term dependencies. The hidden vectors output at each time step preserve the historical context, providing high-quality features for the subsequent attention mechanism. Meanwhile, the number of hidden units in the LSTM encoder is automatically optimized by PSO, ensuring that the model capacity matches the data's dimensionality and complexity, and avoiding overfitting or underfitting. The joint multi-sequence encoding mechanism enables simultaneous modeling of data from multiple age groups and both genders, effectively leveraging cross-group information to improve overall learning efficiency and enhance the model's generalization ability.

Figure 1 illustrates the internal structure of a single LSTM unit, which consists of a core memory cell (the green circle, C_t) and three gating mechanisms: the input gate (i_t), the forget gate (f_t), and the output gate (o_t). Each gate is driven by the current input x_t and the previous hidden state h_{t-1}. After sigmoid activation, the gates produce control signals in the range of 0–1, dynamically regulating the flow of information. The input gate determines how much new information to write into the memory cell, the forget gate controls how much historical information to retain from the previous step, and the output gate decides how much information from the current memory cell should be output as the hidden state $h_{t.}$ Through element-wise operations, these gates precisely regulate both the input and output of the memory cell C_t. This gating mechanism enables the model to dynamically retain or forget information, filter out irrelevant noise, and focus on long-term trends and key turning points related to lung cancer incidence. LSTM is also effective in capturing nonlinear relationships and interactions among multiple subseries, such as age- and sex-specific ASIR data. This makes it an ideal choice for lung cancer incidence prediction tasks, as it can significantly improve prediction accuracy and enhance model stability and generalizability.

The LSTM cell consists of three main gates: the forget gate (f_t), the input gate (i_t), and the output gate (o_t). Next, we will present the training algorithm of LSTM.

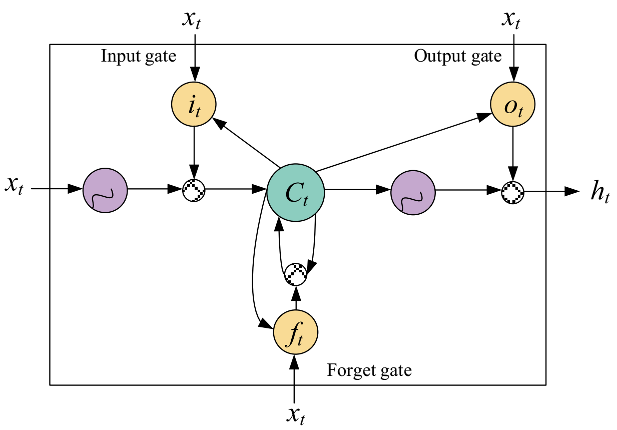

FIGURE 1
LSTM network structure unit.

The forget gate controls what information from the previous time step should be forgotten:

$$f_t = \sigma\left(W_f \cdot \left[h_{t-1}, x_t\right] + b_f\right) \tag{2}$$

The input gate decides what new information should be stored in the memory:

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right) \tag{3}$$

The output gate determines what information from the memory cell will be output:

$$o_t = \sigma\left(W_o \cdot \left[h_{t-1}, x_t\right] + b_o\right) \tag{4}$$

The memory cell C_t is updated as:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tanh\left(W_C \cdot \left[h_{t-1}, x_t\right] + b_C\right) \tag{5}$$

Finally, the hidden state h_t is calculated as:

$$h_t = o_t \cdot \tanh\left(C_t\right) \tag{6}$$

where σ denotes the sigmoid function, and W and b are the weights and biases of the network.

3.3.3 Attention mechanism

The attention mechanism is integrated into the LSTM to allow the model to focus on important time steps in the sequence. The attention-LSTM model mainly includes the input layer, LSTM layer, attention layer, and output layer. In this paper, the attention layer is added behind the LSTM layer, and the input layer of the attention layer is the feature vector output by the LSTM layer, as shown in Figure 2. The probability distribution value of the feature vector is calculated by the features learned by the LSTM layer according to the weight distribution principle, and better weight parameters are obtained by updating iteratively. Finally, through the fully connected layer, the final user power consumption forecast value is output.

The attention weight α_t is computed for each time step as:

$$\alpha_t = \frac{\exp\left(e_t\right)}{\sum_{t=1}^{T} \exp\left(e_t\right)} \tag{7}$$

where e_t is the attention score computed based on the LSTM hidden states h_t at each time step. The attention score is determined by:

$$e_t = v^T \cdot \tanh\left(W_a \cdot h_t + b_a\right) \tag{8}$$

The attention output a_t is then computed as a weighted sum of the hidden states:

$$a_t = \sum_{t=1}^{T} \alpha_t \cdot h_t \tag{9}$$

This allows the model to assign higher weights to the most relevant time steps and improve the prediction accuracy.

3.2.4 Fully connected output layer

After the attention fusion is completed, the concatenated vector is fed into a fully connected layer. This layer applies a linear transformation to the input vector and adds a bias term. The computation is defined as follows.

$$\hat{y}_t = W_y \cdot a_t + b_y \tag{10}$$

where W_y and b_y are the weights and biases for the output layer.

In this layer, the fully connected design ensures that each element of the input vector contributes directly to the output generation. This allows the model to fully exploit the resource information and learn its overall impact on each age-gender subsequence. The number of output nodes is set to 24 × 5, corresponding to the predicted incidence rates of 24 subsequences for each of the next 5 years. Compared with traditional step-by-step forecasting, the fully connected output layer enables single-shot multi-step forecasting. This approach reduces cumulative errors and allows the structural dependencies among subsequences to be jointly learned. For lung cancer ASIR prediction, it means the model can simultaneously forecast annual incidence rates for all age and gender groups, capturing potential co-movements among them.

3.2.5 Particle swarm optimization hyperparameter tuning

PSO is a population-based optimization algorithm that simulates the social behavior of birds flocking to find the best solution. Each particle in the swarm represents a potential solution (set of hyperparameters), and the swarm searches for the optimal set by iteratively updating the particle positions based on its own best-known position and the best-known position of the entire swarm.

Before model training, PSO was employed to automatically search for key hyperparameters. This ensures that the learning capacity of the LSTM encoder and the attention mechanism aligns

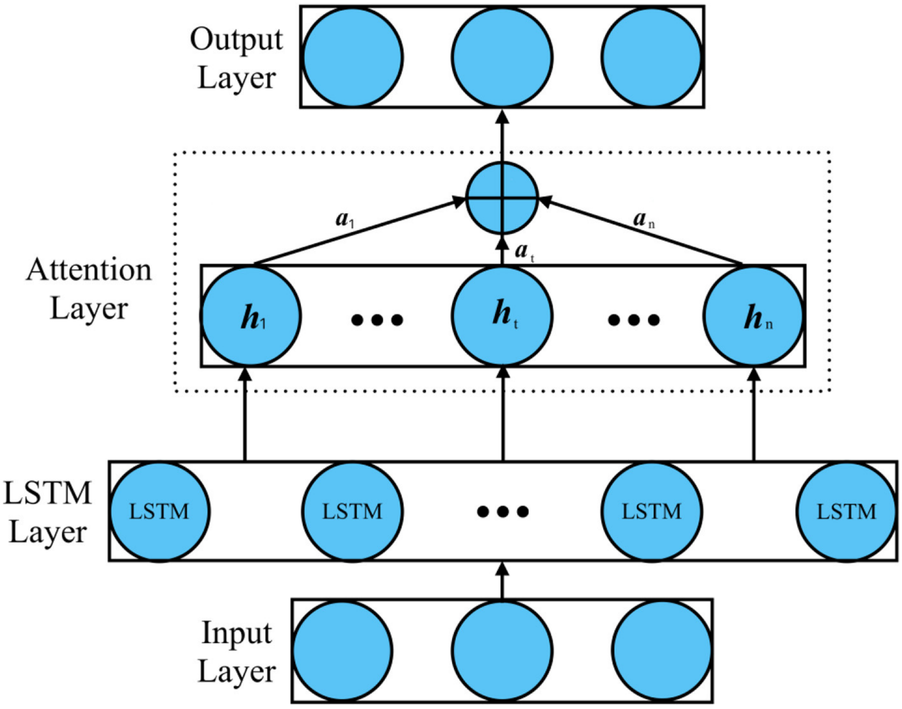

FIGURE 2
Attention-LSTM model structure.

well with the complexity of the lung cancer ASIR data. The hyperparameters tuned by PSO include the number of LSTM hidden units (16–64), dropout rate (0.0–0.4), learning rate (1×10^{-4} to 1×10^{-2} on a logarithmic scale), and batch size (16–64). The optimization was conducted using 10 particles over 50 generations. The inertia weight linearly decreased from 0.9 to 0.5, while both the cognitive and social learning factors were set to 2.0. A three-fold time-series cross-validation strategy was adopted for fitness evaluation: (1990–2005 → 2006–2010), (1990–2010 → 2011–2015), and (1990–2015 → 2016–2020). The objective was to minimize the mean squared error on the validation sets. The PSO process was executed once to avoid nested training and to enhance the reproducibility of the workflow. Given the strong structural trends in ASIR data and the complex interdependencies across age and gender subsequences, PSO allows adaptive configuration of model capacity. This reduces the risk of overfitting or underfitting caused by manual settings, thereby improving both predictive accuracy and model robustness.

The update equations for the particle positions and velocities are:

$$v_i(t+1) = \omega v_i(t) + c_1 \cdot r_1 \cdot \left(p_i - x_i(t)\right) + c_2 \cdot r_2 \cdot \left(g - x_i(t)\right)$$
$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (11)$$

where:

$v_i(t)$ is the velocity of particle i at iteration t,
$x_i(t)$ is the position (hyperparameters) of particle i,
p_i is the personal best position of particle i,
g is the global best position of the swarm,
ω is the inertia weight,
c_1 and c_2 are acceleration coefficients,
r_1 and r_2 are random numbers between 0 and 1.

PSO helps find the optimal hyperparameters by minimizing the loss function of the PSOA-LSTM model, improving its prediction accuracy.

3.3 Evaluation metrics

The performance of the PSOA-LSTM model is evaluated using five commonly used metrics in regression tasks: mean squared error (MSE), R-squared (R^2), mean absolute percentage error (MAPE), normalized root mean squared error (NRMSE), and mean absolute error (MAE).

The MSE is calculated as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \quad (12)$$

Where:
y_i is the ith actual value,
$\hat{y}_i$ is the ith predicted value,
n is the number of data points.
The R^2 value is calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \quad (13)$$

Where:
$\bar{y}$ is the mean of the actual values.

The MAPE is calculated as:

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100 \tag{14}$$

The NRMSE is calculated as:

$$\mathrm{NRMSE} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}}{\max\left(y_{true}\right) - min\left(y_{true}\right)} \tag{15}$$

Where:
$\max\left(y_{true}\right)$ is the maximum values of the actual data,
$\min\left(y_{true}\right)$ is the minimum values of the actual data.
The MAE is calculated as:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{16}$$

These five metrics comprehensively evaluate the accuracy and predictive power of the model.

3.4 Model algorithm flow

The PSOA-LSTM model algorithm flow is presented in Table 1.

4 Experimental design and performance evaluation

4.1 Experimental setup

This study designed a multi-sequence forecasting experiment based on the proposed PSOA-LSTM model. The dataset consists of lung cancer ASIR time series from 1990 to 2021, stratified by gender (male/female) and 12 5-year age groups (from 40–44 to ≥95 years), resulting in a total of 768 data points. Using a sliding window approach with a history length $w = 10$ years and a prediction horizon h = 5 years, we constructed 432 training samples. Each sample has an input shape of (w, 24), representing 24 age-gender subsequences, and an output structure corresponding to forecasts of these 24 subsequences over the next h years. To prevent information leakage, we employed time-series cross-validation using the TimeSeriesSplit method. A three-fold strategy was implemented (e.g., 1990–2005 → 2006–2010), ensuring that all training data strictly precedes the validation data in chronological order.

The model adopts a single-layer LSTM architecture with an attention mechanism for multi-step prediction of lung cancer ASIR subsequences (gender × 12 age groups). The key hyperparameters—number of LSTM hidden units, dropout rate, learning rate, and batch size—are tuned automatically before training using PSO. The optimization objective is to minimize the mean squared error on the validation set under a three-fold time series cross-validation scheme (Timeseries Split): Fold 1 (1990–2005 → 2006–2010), Fold 2 (1990–2010 → 2011–2015), and Fold 3 (1990–2015 → 2016–2020). PSO is configured with 10 particles

TABLE 1 PSOA-LSTM model algorithm flow.

```
PSOA-LSTM model algorithm flow

Input: ASIR, ASIR time series $X \in R^{T\times 24}$(years 1990–2021), sliding window length $L = 10$, forecast horizon h = 5, PSO hyperparameter search space: hidden_units $\in[16, 64]$, dropout_rate $\in[0.0, 0.4]$, learning_rate $\in[1\times10^{-4}, 1\times10^{-2}]$, batch_size $\in[16, 64]$, PSO settings: swarm_size = 10, max_iter = 50, inertia weight $\omega$, linearly decays from 0.9 to 0.5, acceleration coefficients $c_1 = c_2 = 2.0$, Early-stopping patience = 5 consecutive epochs with no improvement.

Output: Final model parameters $\theta$.

1. MODEL INITIALIZATION
       Normalize ASIR dataset $X$
2. SLIDING-WINDOW SAMPLE GENERATION
       For $t = L$ to $T–h$:
3.         X_input ← $X[t–L+1 : t, :]$ /* shape → ($L$ × 24) */
4.         Y_target ← $X[t+1 : t+h, :]$ /* shape → ($h$ × 24) */
5.       End For
6. PSO-BASED HYPERPARAMETER OPTIMIZATION
7.         Initialize swarm $\{x_i\}$, $i = 1\ldots n$, with random hyperparameter values
8.         best_global_score = $+\infty$
9.         For iter = 1 to max_iter do
10.            $\omega = 0.9 – 0.4 *$ (iter / max_iter) /* linear inertia decay */
11.            For each particle $x_i$ in swarm do
12.              Build Attention-LSTM model with hyperparameters $x_i$
13:              Perform 3-fold time-series CV → get validation MSE
14.              If MSE < particle_best_i then
15.                particle_best_i = MSE
16.              End if
17.              If MSE < best_global_score then
18.                best_global_score = MSE
19.                best_global_params = $x_i$
20.              End if
21.            End for
22.            For each particle $x_i$ do
23.              $v_i = \omega\cdot v_i$
24.                 + c1·rand()·(particle_best_i – $x_i$)
25.                 + c2·rand()·(best_global_params – $x_i$)
26.              $x_i = x_i + v_i$
27.            End for
28.            If best_global_score unchanged for 5 iterations then
29.              break
30.            End if
31.          End for
32.        $\sigma$ = best_global_params
33. FINAL MODEL TRAINING
34.          Initialize model using $\sigma$:
```

(Continued)

TABLE 1 (Continued)

PSOA-LSTM Model Algorithm Flow

```
        LSTM encoder with hidden_units and dropout_rate
        Dot-product attention layer
        Dense output layer ($h \times 24$) outputs
    Compile model:
        Loss = MSE
        Optimizer = Adam(lr = $\sigma$.learning_rate)
        Weight regularization = L2
    Set early-stopping (monitor validation MSE, patience = 5)
    Train on full training dataset:
        Batch size = $\sigma$.batch_size
        Max epochs = 100
Return: Final trained model parameters $\theta$
```

TABLE 2 PSO-optimized hyperparameter search space for PSOA-LSTM.

Parameters	Range of search	Types
Hidden units	16–64	Integer
Dropout rate	0.0–0.4	Floating point
Learning rate	$1e^{-4}-1e^{-2}$	Log floating point
Batch size	16–64	Integer
Number of PSO particles	10	–
Maximum number of PSO iterations	50	–
Criterion of convergence	No improvement or upper limit was reached for five consecutive generations	–

and a maximum of 50 generations. The inertia weight decreases linearly from 0.9 to 0.5, and both the cognitive and social learning factors are set to 2.0. The convergence criterion is defined as either no significant improvement in validation MSE over five consecutive generations or reaching the maximum number of iterations. The specific search space is listed in Table 2.

During the PSO-based hyperparameter optimization stage, the attention mechanism was activated. Positioned after the LSTM output, this mechanism learns the importance weights of different time steps, enabling the model to automatically focus on critical historical information from the 24 subsequences. This design integrates temporal dependency modeling with feature selection capability, thereby enhancing both the interpretability and accuracy of the predictions.

In this study, PSO was applied for one-time structural optimization before model training, without employing a nested training workflow, ensuring clarity in the overall methodology. The model implementation was based on the following open-source libraries and frameworks: TensorFlow 2.10 and Keras were used to construct the single-layer LSTM encoder and the attention mechanism. PSO hyperparameter tuning was performed using PySwarms (v1.3.0) with the following settings: n_particles = 10, max_iter = 50, inertia weight linearly decreasing from 0.9 to 0.5, and both cognitive and social coefficients (c1, c2) set to 2.0. The optimization was conducted before training using a three-fold Timeseries Split validation scheme (1990–2005 → 2006–2010, 1990–2010 → 2011–2015, and 1990–2015 → 2016–2020), aiming to minimize the validation mean squared error (MSE). Additional experiments were supported by scikit-learn (for SVR, RF, and ARIMA implementations), statsmodels, NumPy, and Pandas for data processing and evaluation tasks. A custom attention layer was implemented to learn time-step-level importance weights. The PSO-based parameter tuning was completed entirely before model training and did not involve nested optimization, ensuring full reproducibility. After tuning, the best hyperparameters were used for the final training phase. The training was set with a maximum of 100 epochs and an early stopping patience of five epochs (based on validation loss). The model typically converged between the 40th and 60th epochs. All experiments were conducted on a machine equipped with an NVIDIA RTX 3060 GPU and an Intel i7 CPU. Each epoch took ∼90 s, and the entire modeling process—including PSO optimization and final training—took about 1–1.5 h, achieving a balance between performance and computational efficiency.

4.2 Performance analysis of PSOA-LSTM predictive model

Figures 3, 4 present the forecasting results of the PSOA-LSTM model for male and female lung cancer ASIR across 12 age groups (from 40–44 to ≥95 years) during 1990–2021, showing comparisons between actual and predicted values. In each plot, the solid line represents Actual data, while the dashed line denotes the model's predictions. Based on a 10-year historical sliding window, the model performs multi-step forecasting over the next 5 years, outputting incidence rates for 24 age-gender subsequences per year. Across both sexes, the model successfully captures key temporal patterns, particularly in high-incidence middle-aged and elderly groups (60–79 years), where trends of increase, peak, and decline are well reflected. Even in groups with low incidence or data volatility (e.g., young adults and the oldest elderly), the model maintains stable forecasting performance. These results confirm that the PSOA-LSTM model offers strong robustness and generalization capabilities for structured health time series forecasting, and is suitable for age- and gender-specific ASIR prediction tasks.

4.3 Ablation study

To further validate the contribution of each component in the model, we conducted an ablation study by systematically removing or modifying parts of the model. The following variations were tested:

1. LSTM only (No Attention or PSO): in this configuration, we trained the model with only the LSTM layer, without any attention mechanism or PSO optimization. The model achieved an MSE of 0.042 and an R^2 of 0.91. While this model still performs reasonably well, it lacks the enhanced

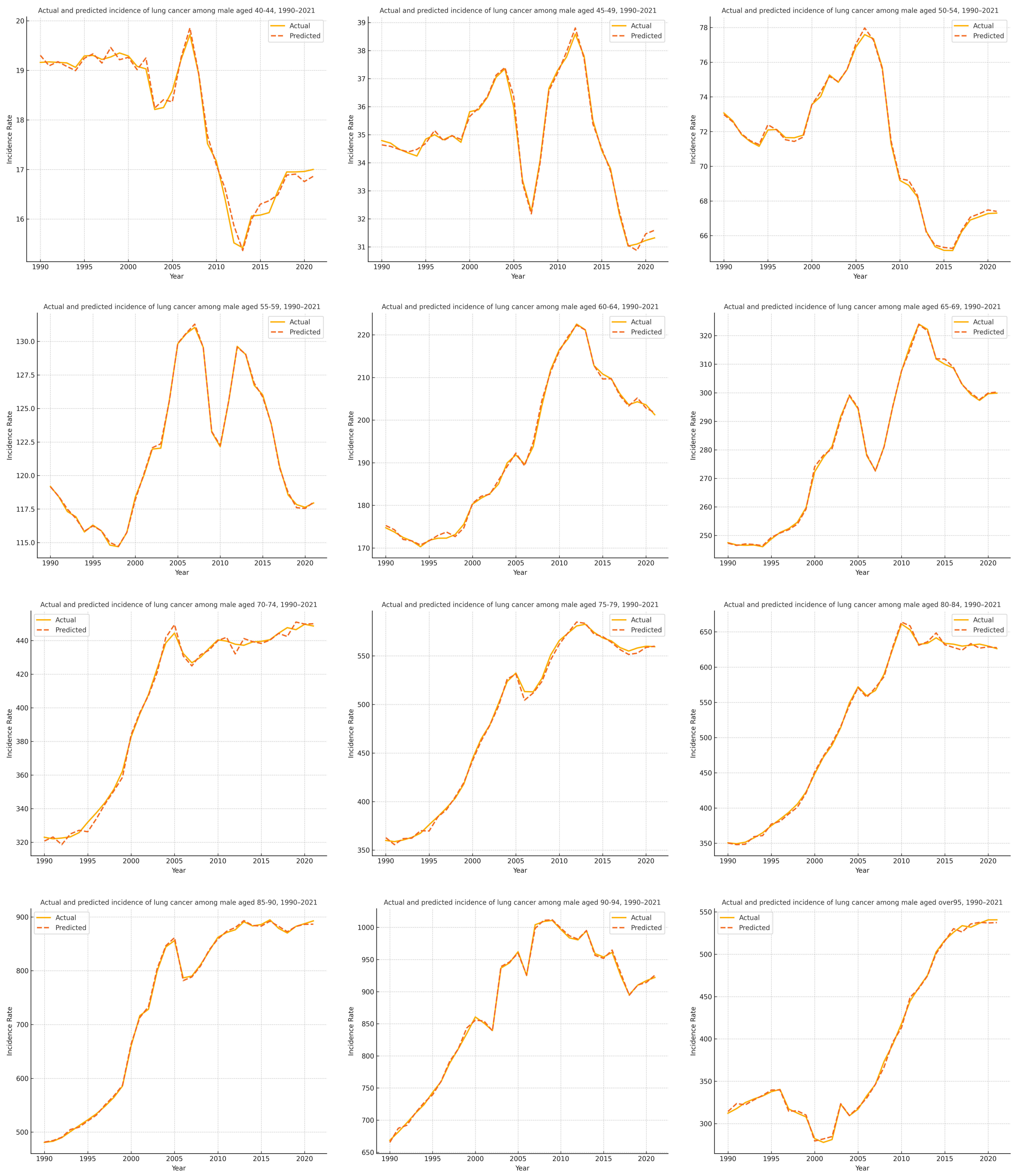

FIGURE 3
Comparison of actual and predicted lung cancer ASIR for male age groups (1990–2021) using the PSOA-LSTM model.

predictive capability provided by the attention mechanism and PSO optimization.

2. LSTM with attention (No PSO): in this setup, we added the attention mechanism to the LSTM model but kept the hyperparameters fixed, without PSO optimization. The model's MSE improved to 0.035, and R^2 increased to 0.93. The attention mechanism allowed the model to focus on more relevant time steps, resulting in a more interpretable and accurate model.

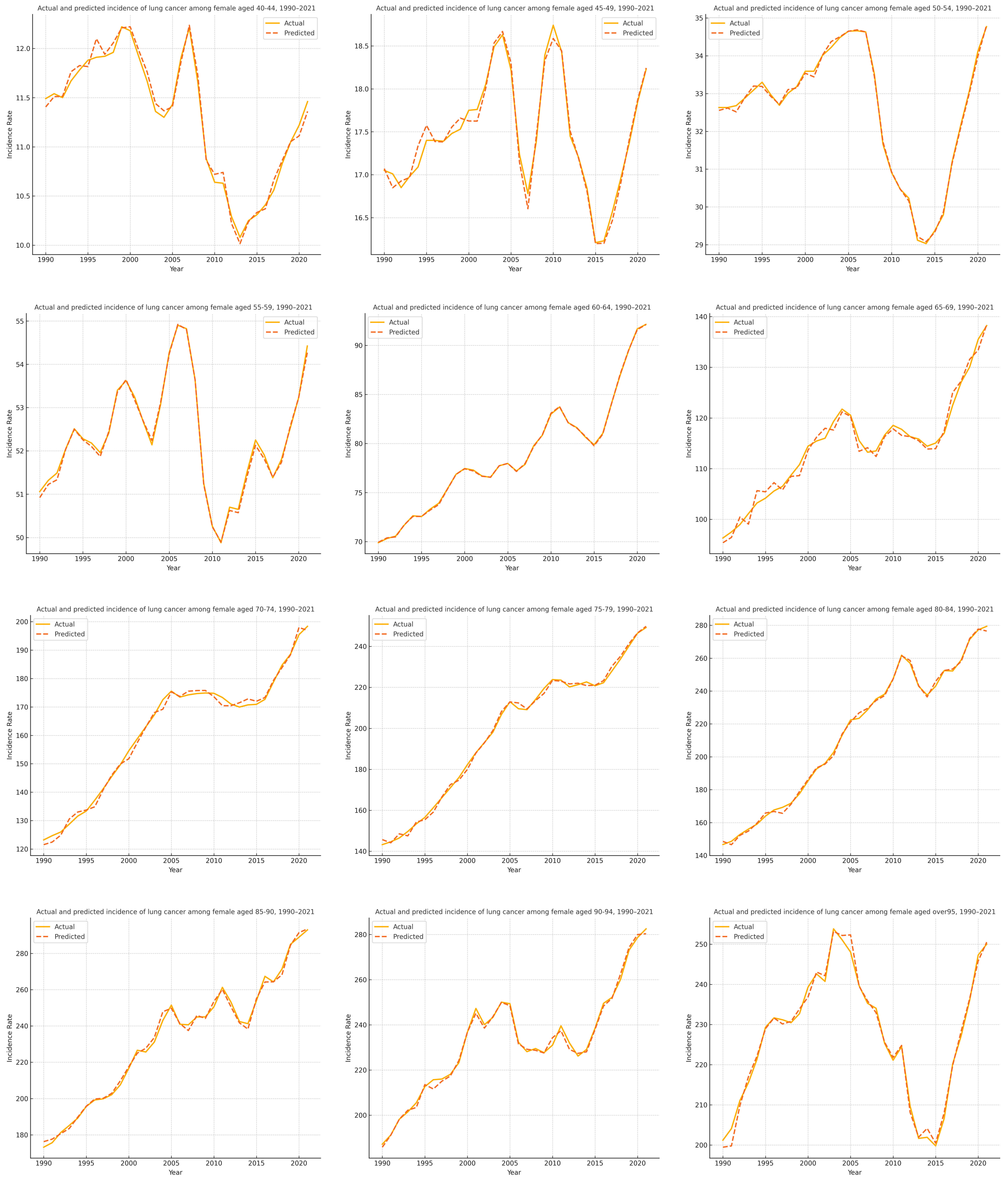

FIGURE 4
Comparison of actual and predicted lung cancer ASIR for female age groups (1990–2021) using the PSOA-LSTM model.

3. LSTM with PSO (no attention): in this variant, we removed the attention mechanism but applied PSO for hyperparameter optimization. The model achieved an MSE of 0.031 and an R^2 of 0.94. PSO helped the model converge more efficiently by tuning the LSTM units and learning rate, but without the attention mechanism, the model could not fully capture the most relevant time steps.
4. LSTM + attention + PSO (proposed model: PSOA-LSTM): the proposed model, which combines LSTM, attention, and PSO, achieved the best performance with an MSE of 0.023 and an R^2 of 0.97, as previously

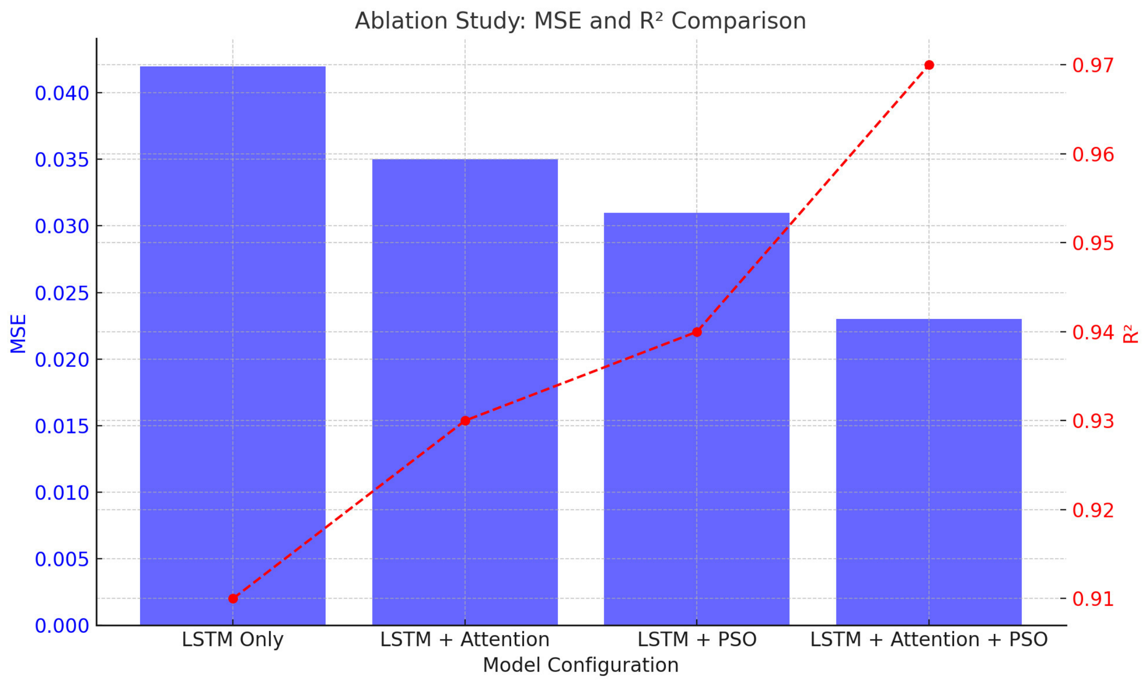

FIGURE 5
Ablation Study: MSE and R^2 Comparison for different models.

TABLE 3 PSOA-LSTM ablation study evaluation metrics.

Model variation	MSE	R^2	MAPE/%	NRMSE	MAE
LSTM only (no attention or PSO)	0.042	0.91	0.51	0.035	0.204
LSTM + attention (no PSO)	0.035	0.93	0.47	0.032	0.187
LSTM + PSO (no attention)	0.031	0.94	0.44	0.029	0.176
PSOA-LSTM	0.023	0.97	0.38	0.025	0.152

TABLE 4 Performance comparison between PSOA-LSTM and comparative models on lung cancer ASIR forecasting.

Model	MSE	R^2	MAPE/%	NRMSE	MAE
PSOA-LSTM	0.023	0.97	0.38	0.025	0.152
SVR	0.039	0.92	0.48	0.038	0.190
RF	0.043	0.90	0.52	0.041	0.205
ARIMA	0.056	0.85	0.66	0.047	0.597
LSTM	0.042	0.91	0.51	0.035	0.204

reported. This configuration shows that all components contribute to improving the model's ability to forecast lung cancer incidence.

The ablation study results are summarized in Table 3.

The results clearly demonstrate the advantage of combining LSTM with attention and PSO optimization. The ablation study reveals that each component of the model plays a vital role in improving prediction accuracy. The attention mechanism helps the model focus on critical time steps, while PSO optimization fine-tunes the hyperparameters, leading to better model performance.

Figure 5 visualizes the MSE and R^2 values for different model configurations, demonstrating the contribution of each component (LSTM, Attention, and PSO) in improving the performance.

4.4 Comparison with other models

To evaluate the forecasting performance of the proposed PSOA-LSTM model, we conducted comparative experiments against four baseline models: SVR, RF, ARIMA, and LSTM, as shown in Table 4. The configuration of each model is as follows: SVR: The RBF kernel function is used with a kernel parameter of 0.1, penalty term is 10; RF: Set to 100 decision trees, maximum depth = 10, and minimum samples split = 2; ARIMA: the setting was ($p = 0, d = 2, q = 0$); LSTM: Same architecture as PSOA-LSTM but without attention and PSO optimization.

Figure 6 presents a normalized heatmap of model performance across five key evaluation metrics (MSE, R^2, MAPE, NRMSE, and MAE), where green indicates the best performance and red indicates the worst. The PSOA-LSTM model consistently appears in dark green across all metrics, demonstrating its superior performance in multi-step lung cancer ASIR forecasting. In contrast, the ARIMA model is shown in red for all metrics, indicating the poorest performance—particularly in MAPE (0.660) and MAE (0.597)—highlighting its limitations in modeling nonlinear and structured time series. SVR and RF perform moderately, with some metrics in the mid-range but inconsistent across dimensions. The baseline LSTM performs better than RF and ARIMA on certain metrics like MSE and NRMSE, but falls short of PSOA-LSTM due to the absence of hyperparameter tuning and attention mechanisms. This heatmap provides a clear visual confirmation of PSOA-LSTM's comprehensive advantage and its robustness in structured health data forecasting.

The PSOA-LSTM model was employed to predict the annual ASIR of lung cancer in China for both females and males from 2022 to 2026. The predictions are stratified by 12 5-year age groups (from 40–44 to ≥95 years) and separated by gender, as shown in Tables 5, 6. The results indicate that lung cancer incidence rates increase markedly with age in both sexes, with males consistently exhibiting higher ASIR values than females in

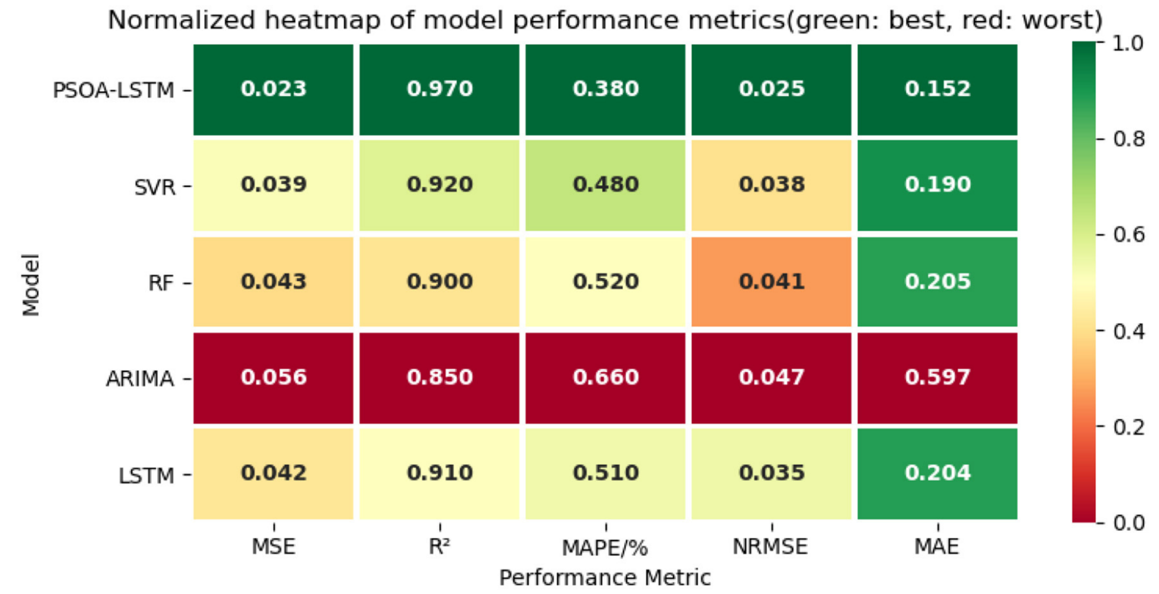

FIGURE 6
Scatter plot of predicted and actual values for ARIMA, LSTM, and PSOA-LSTM.

TABLE 5 The PSOA-LSTM model predicts the annual incidence of lung cancer (per 100,000 people) for Chinese males in each age group from 2022 to 2026.

Year	Age											
	40–44	45–49	50–54	55–59	60–64	65–69	70–74	75–79	80–84	85–89	90–95	Over 95
2022	13.89	44.41	77.63	121.86	195.81	284.58	401.70	484.19	514.60	740.71	869.91	396.24
2023	18.78	46.96	72.61	128.47	190.47	281.34	406.79	486.80	528.40	744.54	870.98	392.35
2024	23.17	41.52	65.58	116.5	193.65	283.5	399.57	487.98	525.66	746.28	877.45	393.46
2025	15.66	33.21	68.64	119.5	191.81	276.78	393.15	492.56	514.02	734.18	883.21	391.08
2026	5.59	22.21	78.85	120.45	194.17	288.68	404.44	490.33	531.82	734.77	878.16	388.13

TABLE 6 The PSOA-LSTM model predicts the annual incidence of lung cancer (per 100,000 people) for Chinese females in each age group from 2022 to 2026.

Year	Age											
	40–44	45–49	50–54	55–59	60–64	65–69	70–74	75–79	80–84	85–89	90–95	Over 95
2022	13.89	44.41	77.63	121.86	195.81	284.58	401.70	484.19	514.60	740.71	869.91	396.24
2023	18.78	46.96	72.61	128.47	190.47	281.34	406.79	486.80	528.40	744.54	870.98	392.35
2024	23.17	41.52	65.58	116.5	193.65	283.5	399.57	487.98	525.66	746.28	877.45	393.46
2025	15.66	33.21	68.64	119.5	191.81	276.78	393.15	492.56	514.02	734.18	883.21	391.08
2026	5.59	22.21	78.85	120.45	194.17	288.68	404.44	490.33	531.82	734.77	878.16	388.13

each corresponding age group. Notably, the incidence rises sharply among the elderly, reaching its peak in the ≥90 years group. This granular, gender- and age-specific forecasting provides a robust foundation for identifying high-risk subpopulations, supporting the rational allocation of medical resources, and informing the design of targeted prevention and intervention strategies in public health practice.

5 Discussion

While the PSOA-LSTM model demonstrates clear superiority in predictive accuracy across all evaluated metrics, a deeper inspection reveals several key aspects regarding model behavior and applicability. First, the substantial gain in performance over traditional models such as ARIMA highlights the critical role of capturing non-linear and long-term dependencies in lung cancer incidence data. The inclusion of the attention mechanism enables the model to dynamically focus on informative historical periods, enhancing the interpretability and relevance of learned patterns. Particle swarm optimization further ensures optimal hyperparameter selection, thus mitigating the risk of overfitting in a limited-sample context.

However, this study is not without limitations. Despite the use of stratified, multi-sequence input, the available annual data remains relatively sparse compared to many machine learning applications, which may constrain the maximum achievable model complexity and generalization. While PSOA-LSTM achieves an excellent fit on the current dataset, its extrapolative power beyond the training data—especially under scenarios of drastic epidemiological change (e.g., new screening or environmental interventions)—remains to be validated. Furthermore, the models rely on the availability and quality of age- and sex-specific incidence data, which may vary in completeness across regions and over time.

Practically, these findings underscore the need for robust, interpretable forecasting tools in cancer epidemiology. The clear performance gradient observed across model types suggests that hybrid deep learning approaches like PSOA-LSTM can significantly improve resource allocation, risk stratification, and early warning capabilities in public health systems. Yet, ongoing methodological refinement, external validation on different populations, and integration of additional risk factors (such as smoking prevalence or air pollution) will be essential for broadening the model's real-world impact.

6 Conclusion

In summary, this study developed and validated a PSOA-LSTM model for forecasting lung cancer incidence rates by age and sex in China. The proposed approach significantly outperformed conventional machine learning and statistical models, demonstrating superior accuracy and robustness. The findings provide an important foundation for targeted prevention, resource planning, and public health policy formulation in cancer control. Future work will focus on model generalization, external validation, and the incorporation of additional covariates to further enhance predictive capability and practical utility.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://vizhub.healthdata.org/gbd-results.

Author contributions

NX: Data curation, Writing – review & editing, Writing – original draft. GY: Conceptualization, Writing – review & editing. LM: Writing – review & editing, Methodology. JD: Validation, Writing – original draft. KZ: Writing – review & editing, Writing – original draft, Conceptualization, Methodology.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the Scientific Research Project of the Basic Scientific Research Business Expenses of Provincial Universities in Heilongjiang Province (2021-KYYWF-0357).

Acknowledgments

We appreciate the work of the Global Burden of Disease study 2021 collaborators.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. *CA Cancer J Clin.* (2024) 74:12–49. doi: 10.3322/caac.21820

2. Tuncal K, Sekeroglu B, Ozkan C. Lung cancer incidence prediction using machine learning algorithms. *J Adv Inf Technol.* (2020) 11:91–96. doi: 10.12720/jait.11.2.91-96

3. Tudor C. A novel approach to modeling and forecasting cancer incidence and mortality rates through web queries and automated forecasting algorithms: evidence from Romania. *Biology.* (2022) 11:857. doi: 10.3390/biology11060857

4. Tsan YT, Chen DY, Liu PY, Kristiani E, Nguyen KLP, Yang CT. The prediction of influenza-like illness and respiratory disease using LSTM and ARIMA. *Int J Environ Res Public Health.* (2022) 19:1858. doi: 10.3390/ijerph19031858

5. Li H, Zhao M, Fei G, Wang Z, Wang S, Wei P, et al. Epidemiological trends and incidence prediction of lung cancer in China based on the Global Burden of Disease study 2019. *Front Med.* (2022) 9:969487. doi: 10.3389/fmed.2022.969487

6. Bhargav AL, Ashokkumar C. AI-driven insights: a survey on innovative approach for lung cancer prediction utilizing machine learning and deep learning methods. In: *Proceedings of the 2024 3rd International Conference on Automation, Computing and Renewable Systems (ICACRS).* Piscataway, NJ: IEEE (2024). p. 1152–8. doi: 10.1109/ICACRS62842.2024.10841728

7. Farhatin N, Fadli M, Putranto AMY, Valerian J, Sihono DSK, Prajitno P. Prediction of radiation therapy dose for lung cancer IMRT technique using support vector regression model. *J Phys Conf Ser.* (2022) 2377:012030. doi: 10.1088/1742-6596/2377/1/012030

8. Bharati S, Podder P, Paul PK. Lung cancer recognition and prediction according to random forest ensemble and RUSBoost algorithm using LIDC data. *Int J Hybrid Intell Syst.* (2019) 15:91–100. doi: 10.3233/HIS-190263

9. Gupta S, Tran T, Luo W, Phung D, Kennedy RL, Broad A, et al. Machine-learning prediction of cancer survival: a retrospective study using electronic administrative records and a cancer registry. *BMJ Open.* (2014) 4:e004007. doi: 10.1136/bmjopen-2013-004007

10. She Y, Jin Z, Wu J, Deng J, Zhang L, Su H, et al. Development and validation of a deep learning model for non-small cell lung cancer survival. *JAMA Netw Open.* (2020) 3:e205842. doi: 10.1001/jamanetworkopen.2020.5842

11. Graves A. Long short-term memory. In: *Supervised Sequence Labelling with Recurrent Neural Networks.* Berlin; Heidelberg: Springer (2012). p. 37–45. doi: 10.1007/978-3-642-24797-2_4

12. Hochreiter S. Long short-term memory. *Neural Comput.* (1997) 9:1735–80. doi: 10.1162/neco.1997.9.8.1735

13. Huang S, Arpaci I, Al-Emran M, Kiliçarslan S, Al-Sharafi MA. A comparative analysis of classical machine learning and deep learning techniques for predicting lung cancer survivability. *Multimed Tools Appl.* (2023) 82:34183–98. doi: 10.1007/s11042-023-16349-y

14. Gao R, Huo Y, Bao S, Tang Y, Antic SL, Epstein ES, et al. Distanced LSTM: time-distanced gates in long short-term memory models for lung cancer detection. In: *Proceedings of the 10th International Workshop on Machine Learning in Medical Imaging (MLMI 2019).* Cham: Springer (2019). p. 310–8. doi: 10.1007/978-3-030-32692-0_36

15. Edara DC, Vanukuri LP, Sistla V, Kolli VKK. Sentiment analysis and text categorization of cancer medical records with LSTM. *J Ambient Intell Humaniz Comput.* (2023) 14:5309–25. doi: 10.1007/s12652-019-01399-8

16. Morid MA, Sheng ORL, Dunbar J. Time series prediction using deep learning methods in healthcare. *ACM Trans Manag Inf Syst.* (2023) 14:1–29. doi: 10.1145/3531326

17. Men L, Ilk N, Tang X, Liu Y. Multi-disease prediction using LSTM recurrent neural networks. *Expert Syst Appl.* (2021) 177:114905. doi: 10.1016/j.eswa.2021.114905

18. Thaventhiran C, Sekar KR. Target projection feature matching based deep ANN with LSTM for lung cancer prediction. *Intell Autom Soft Comput.* (2022) 31:1–10. doi: 10.32604/iasc.2022.019546

19. Zhang H, Xi Q, Zhang F, Li Q, Jiao Z, Ni X. Application of deep learning in cancer prognosis prediction model. *Technol Cancer Res Treat.* (2023) 22:15330338231199287. doi: 10.1177/15330338231199287

20. Rashid TA, Hassan MK, Mohammadi M, Fraser K. Improvement of variant adaptable LSTM trained with metaheuristic algorithms for healthcare analysis. In: *Research Anthology on Artificial Intelligence Applications in Security.* Hershey, PA: IGI Global (2021). p. 1031–51. doi: 10.4018/978-1-7998-7705-9.ch048

21. Vaswani A. Attention is all you need. *Adv Neural Inf Process Syst.* (2017) 30:1–11. doi: 10.5555/3295222.3295349

22. Gonçalves T, Rio-Torto I, Teixeira LF, Cardoso JS. A survey on attention mechanisms for medical applications: are we moving toward better algorithms? *IEEE Access.* (2022) 10:98909–35. doi: 10.21203/rs.3.rs-1594205/v1

23. Xiao L, Li M, Feng Y, Wang M, Zhu Z, Chen Z. Exploration of attention mechanism-enhanced deep learning models in the mining of medical textual data. *arXiv Preprint.* (2024) arXiv:2406.00016. doi: 10.1109/ICSECE61636.2024.10729303

24. Choi E, Bahadori MT, Sun J, Kulas J, Schuetz A, Stewart W, et al. RETAIN: an interpretable predictive model for healthcare using reverse time attention mechanism. *Adv Neural Inf Process Syst.* (2016) 29:1–9. Available online at: https://proceedings.neurips.cc/paper/2016/file/231141b34c82aa95e48810a9d1b33a79-Paper.pdf

25. Zhang Y. ATTAIN: attention-based time-aware LSTM networks for disease progression modeling. In: *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI-2019).* Palo Alto, CA: IJCAI Organization (2019). p. 4369–75. doi: 10.24963/ijcai.2019/607

26. Kennedy J, Eberhart R. Particle swarm optimization. In: *Proceedings of ICNN'95 - International Conference on Neural Networks.* Piscataway, NJ: IEEE (1995). p. 1942–8. doi: 10.1109/ICNN.1995.488968

27. Singh J. A comprehensive survey of PSO-ACO optimization and swarm intelligence in healthcare: implications for medical image analysis and disease surveillance. In: *Proceedings of the 2023 3rd Asian Conference on Innovation in Technology (ASIANCON).* Piscataway, NJ: IEEE (2023). p. 1–6. doi: 10.1109/ASIANCON58793.2023.10270025

28. Raghuvanshi SS, Arya KV, Patel V. PSbBO-Net: a hybrid particle swarm and Bayesian optimization-based DenseNet for lung cancer detection using histopathological and CT images. *Int J Electr Electron Res.* (2024) 12:1074–86. doi: 10.37391/ijeer.120343

29. Liao L, Li H, Shang W, Ma L. An empirical study of the impact of hyperparameter tuning and model optimization on the performance properties of deep neural networks. *ACM Trans Softw Eng Methodol.* (2022) 31:1–40. doi: 10.1145/3506695

30. Makarovskikh T, Abotaleb M, Albadran Z, Ramadhan AJ. Hyper-parameter tuning for the long short-term memory algorithm. In: *AIP Conference Proceedings.* Melville, NY: AIP Publishing (2022). p. 2977-1–9. doi: 10.1109/ITNT55410.2022.9848654

31. Quintiliano Bezerra Silva A. Predicting cervical cancer with metaheuristic optimizers for training LSTM. In: *Computational Science–ICCS 2019: 19th International Conference.* Cham: Springer International Publishing (2019). p. 642–55. doi: 10.1007/978-3-030-22750-0_62

32. Islam MS, Umran HM, Umran SM, Karim M. Intelligent healthcare platform: cardiovascular disease risk factors prediction using attention module based LSTM. In: *Proceedings of the 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD).* Piscataway, NJ: IEEE (2019). p. 167–75. doi: 10.1109/ICAIBD.2019.8836998

33. Praneeth VS, Gowtham N, RamaChandran S, Jansi R. Revolutionizing Alzheimer's disease prediction using EfficientNetB6. In: *Proceedings of the 2024 Tenth International Conference on Bio Signals, Images, and Instrumentation (ICBSII).* Piscataway, NJ: IEEE (2024). p. 1–7. doi: 10.1109/ICBSII61384.2024.10564023

34. Langat A, Orwa G, Koima J. Cancer cases in Kenya; forecasting incidents using Box & Jenkins ARIMA model. *Biomed Stat Inform.* (2017) 2:37–48. doi: 10.11648/j.bsi.20170202.11

35. Kong L, Li G, Rafique W, Shen S, He Q, Khosravi MR, et al. Time-aware missing healthcare data prediction based on ARIMA model. *IEEE/ACM Trans Comput Biol Bioinform.* (2022) 19:2345–53. doi: 10.1109/TCBB.2022.3205064

36. Ahmed SRA, Al Barazanchi I, Mhana A, Abdulshaheed HR. Lung cancer classification using data mining and supervised learning algorithms on multi-dimensional data set. *Period Eng Nat Sci (PEN).* (2019) 7:438–47. doi: 10.21533/pen.v7i2.483

37. Wu X, Denise BB, Zhan FB, Zhang J. Determining association between lung cancer mortality worldwide and risk factors using fuzzy inference modeling and random forest modeling. *Int J Environ Res Public Health.* (2022) 19:14161. doi: 10.3390/ijerph192114161

38. Khan R, Jie W. Using the TSA-LSTM two-stage model to predict cancer incidence and mortality. *PLoS ONE.* (2025) 20:e0317148. doi: 10.1371/journal.pone.0317148

39. Liu X, Shi Q, Liu Z, Yuan J. Using LSTM neural network based on improved PSO and attention mechanism for predicting the effluent COD in a wastewater treatment plant. *IEEE Access.* (2021) 9:146082–96. doi: 10.1109/ACCESS.2021.3123225