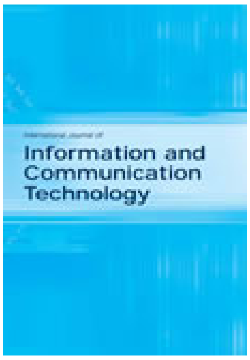
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Economic monitoring and early warning based on feature screening and hybrid neural network

Dongfang Dai

DOI: 10.1504/IJICT.2025.10071629

Article History:

Received:	15 April 2025
Last revised:	28 April 2025
Accepted:	29 April 2025
Published online:	20 June 2025

Economic monitoring and early warning based on feature screening and hybrid neural network

Dongfang Dai

College of Electronic Commerce,
Tangshan University,
Tangshang 063000, China
Email: daidongfang202312@163.com

Abstract: Intending to the issue that the existing study do not fully exploit features, the random forest algorithm (IRF) is improved first. The splitting feature screening function is simplified based on the principle of infinitesimal equivalence, and the Gini coefficient value of the non-category attribute is introduced to improve the computational efficiency of the algorithm. Then, public health economic impact variables are selected, and spatial features are extracted using a residual convolutional neural network. Temporal features are extracted using a gate rate unit (GRU), and a self-attention mechanism is incorporated to enhance the spatial and temporal features. Finally, the IRF filter is used to select the most important spatio-temporal features of the early warning results and map them to the monitoring and early warning results through nonlinear transformation. The experimental outcome indicates that the accuracy of the proposed model has been improved by 5.07%–14.85%.

Keywords: economic early warning; random forest; feature screening; residual convolutional neural network; gate rate unit; GRU.

Reference to this paper should be made as follows: Dai, D. (2025) 'Economic monitoring and early warning based on feature screening and hybrid neural network', *Int. J. Information and Communication Technology*, Vol. 26, No. 21, pp.23–38.

Biographical notes: Dongfang Dai received her Master's degree in Economics at Hebei University of Technology in 2007. She is currently an Associate Professor in the College of Electronic Commerce, Tangshan University. Her research interests include technology economics, regional innovation, and machine learning.

1 Introduction

In an era of accelerating globalisation, the impact of public health events has gone far beyond the scope of health, and has had a profound and widespread impact on the economy. Public health events are often accompanied by substantial investments in medical resources, including but not limited to the expansion of hospital facilities, the emergency procurement of medical supplies, and overtime compensation for medical staff (Szreter and Woolcock, 2004; Correia et al., 2022). Meanwhile, the spread of the disease will affect many industries. For example, the service industry will suffer greatly due to restrictions on the movement of people, while the manufacturing industry may face production stoppages due to supply chain disruptions. These direct and indirect economic impacts are intertwined, forming a complex network of public health economics (Lu et al., 2017). To better cope with the economic challenges posed by public health events, it is imperative to establish a comprehensive public health economic monitoring and early warning system. Through dynamic monitoring and early warning of the public health economy, decision-makers can predict in advance the extent of economic damage that may be caused by public health events, and then formulate more targeted and forward-looking economic policies to alleviate the pressure of economic downturns and maintain social and economic stability (Ebi and Schmier, 2005; Meckawy et al., 2016).

Ruck et al. (2021) introduced five public health economic ratio indicators into the model, completing the transformation of the early warning model from univariate to multivariate. Rappold et al. (2014) first considered using multiple linear discriminant analysis to study the problem of public health economic crises. This can also be referred to as a Z-score model, which discriminates risks based on the magnitude of the Z-score, which is inversely related to public health economic risks. Kolomoyets et al. (2021) used a logistic model to conduct a study on the medical industry and gave conditions for the effectiveness of the model's application in the field of public health economic early warning. Ma et al. (2023) constructed a Benford-Logistic model to study the issue of public health economic early warning, but there was a large error in the early warning results, but there was a large error in the early warning results.

In the field of early warning in public health economics, traditional statistical models and logistic regression have been widely used, but their limitations are becoming increasingly apparent. Machine learning models, with their more flexible data modelling capabilities, can effectively compensate for the shortcomings of traditional methods. When dealing with large amounts of complex data in public health economics, machine learning algorithms can better capture the non-linear characteristics of public health economics data, improving the accuracy of early warnings. Mao et al. (2020) used support vector machine (SVM) to conduct early warning for public health economics, and the experimental results proved the effectiveness of the method. Dou et al. (2023) used the Pearson coefficient to screen for characteristics of public health economic impacts, and implemented public health economic early warning through decision trees, but the accuracy of the early warning was not high. Yin et al. (2022) used PCA to reduce the dimensionality of the initially selected indicators, and then used a BP neural network to construct an early warning model, which has a good early warning effect on the public health economy. Botz et al. (2022) used random forests (RF) to screen for public health economic characteristics and built an early warning model using a classification and regression trees (CART) classification tree to improve the accuracy of early warning.

Machine learning-based public health economic monitoring and early warning models are difficult to use for mining the intrinsic characteristics of influencing indicators, resulting in low monitoring accuracy. Deep learning models have more powerful feature extraction and generalisation capabilities. Therefore, more and more scholars hope to solve the problem of public health economic early warning through deep learning. Zheng and Hu (2021) combined recurrent neural network (RNN), autoregressive moving average (ARMA) and exponential smoothing models to achieve public health economic early warning and achieved satisfactory prediction results. Sansone and Zhu (2023) used principal component analysis (PCA) to extract the principal components of impact indicators and convolutional neural network (CNN) to monitor and predict public health economics, which improved the performance of early warning. Devyatkin et al. (2021) propose a new public health economic early warning method that includes an long short-term memory (LSTM) module to prevent overfitting and an LSTM module for early warning, which can effectively reduce prediction errors. Although most current methods based on a single neural network are simple and effective, the extracted features are too monotonous. Hybrid neural networks make full use of the characteristics of different neural networks to extract deep and rich features and thereby improve the accuracy of early warning. Azadi et al. (2023) used a series and parallel connection of LSTM and CNN, respectively, to achieve public health economic early warning and improve the accuracy of early warning. Wang et al. (2024) constructed a method consisting of a CNN and bidirectional transformer to complete public health economic early warning.

The above research status shows that there is feature redundancy in studies based on hybrid neural networks, which leads to unsatisfactory early warning results. To this end, this paper first improves the RF algorithm by simplifying the split feature selection function according to the principle of equivalent infinitesimal, so as to improve the efficiency of function operation. The introduction of the non-class attribute in the simplified split feature selection function improves the calculation efficiency and screening performance of the algorithm by calculating the Gini coefficient value of this attribute. Then, public health economic impact variables were selected and pre-processed. Next, a hybrid neural network model was designed to extract the temporal and spatial characteristics of the variables. The residual structure was used to improve the CNN to fully extract the spatial characteristics of the variables. Meanwhile, the time features are extracted through GRU, and a self-attention mechanism is incorporated to enhance the spatiotemporal features. Finally, the IRF is used to filter the most important spatio-temporal features and perform a nonlinear transformation to map them to the final monitoring and early warning results. The experimental outcome indicates that the early warning accuracy of the proposed model is 93.36%, which can be better applied to the early warning of public health economic monitoring.

2 Relevant theoretical foundations

2.1 Convolutional neural network

CNN and RNN are both extended from traditional neural networks, but the neurons in CNN use local connections, which greatly reduces the difficulty of training (Kuo, 2016). CNN calculates by sliding a convolution kernel (filter) over local areas of the input data (such as 3 × 3 pixels) rather than being fully connected, significantly reducing the number of parameters (the same convolution kernel shares weights). RNN is suitable for time series prediction, but its feature extraction ability is weak. GAN is suitable for testing with fewer samples, requires additional design of a discriminator, and the computational cost is usually 2–3 times that of CNN. CNN generally consists of convolutional, pooling and fully connected levels, as shown in Figure 1. The roles of each level in CNN are shown below:

1 The convolutional level extracts features by collecting data and performing convolution operations using multiple convolution kernels. The three parameters are padding, stride, and kernel size. When the convolutional level extracts features, the convolution kernel represents the weight, and the convolution calculation is as bellow.

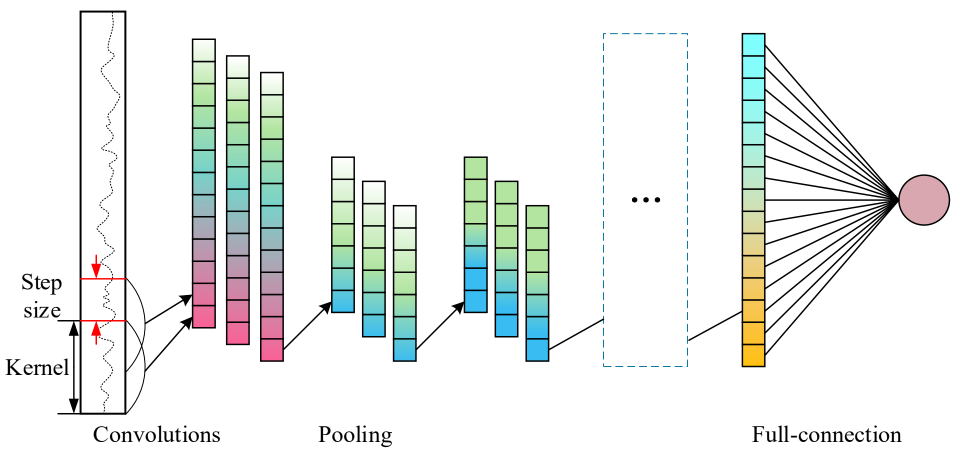
$$c_i = f\left(w \otimes x_i + b_i\right) \tag{1}$$

where x_i is the input, c_i is the feature output of the i^{th} layer, w is the weight matrix, $\otimes$ is the convolution operation, and is the bias parameter.

2 A pooling level is often adopted to remove redundant information and reduce the amount of computation.

3 The fully connected level is a multi-level perceptron with a Softmax activation function (Khan et al., 2020), which enables end-to-end learning.

Figure 1 General structure of CNN (see online version for colours)



2.2 Gated recurrent unit

GRU replaces the forget gate, input gate, and output gate in the LSTM model with an update gate and reset gate. The structure is simpler than LSTM, with fewer internal parameters, so training speed is improved compared to LSTM (Zarzycki and Ławryńczuk, 2022). The updated gate of GRU is used to determine which part of the information entered in x_t is retained and which part is discarded. The formula for calculating the output u_t of the gate is as follows.

$$u_t = \sigma_u\left(W_{xu}x_t + W_{hu}h_{t-1} + b_u\right) \tag{2}$$

The reset door determines the following formula for calculating the output r_t of the door: how much of the hidden state h_{t-1} at time t–1 is retained to the current time h_t.

$$r_t = \sigma_r\left(W_{xr}x_t + W_{hr}h_{t-1} + b_r\right) \tag{3}$$

Candidate hidden states n_t can be calculated from r_t. When $r_t = 0$, n_t only contains information about the current input x_t. n_t is calculated as follows.

$$n_t = \tanh\left(W_{xn}x_t + W_{hn}\left(r_t \odot h_{t-1}\right) + b_n\right) \tag{4}$$

Finally, the current neuron's hidden state h_t is updated through u_t and n_t, and the update formula is as follows.

$$h_t = \left(1 - u_t\right) \odot h_{t-1} + u_t \odot n_t \tag{5}$$

where $\odot$ is the multiplication operation, w, b, and σ are the weight matrix, bias vector, and activation function, respectively.

2.3 Random forest

The RF algorithm itself has the ability to filter features. The importance of the data features is calculated through RF (Paul et al., 2018), which is used to analyse the importance of all the data and then select the appropriate features as the input data for prediction. Feature selection is a critical step in machine learning, used to remove redundant or irrelevant features to improve model performance and interpretability. In addition to RF, filtering, wrapping, embedding, and other methods are also commonly used feature selection algorithms. RF quantifies feature importance by calculating the average reduction in impurity (Gini index or information gain) or the number of splits at which a feature acts as a splitting node across all trees. It does not require additional calculations and directly outputs an interpretable feature ranking. The algorithm process is shown in Figure 2. The specific steps are as follows.

1 First, the collected feature data is taken as input. It is assumed that there is feature variable $X_1, X_2, \ldots, X_j$, VIM_j is the average change of node splitting impurity of the j^{th} variable in all trees, and the Gini index of each feature is calculated as G_m (Speiser et al., 2019), as shown below.

$$G_m = \sum_{k=1}^{K} \hat{p}_{mk}\left(1 - \hat{p}_{mk}\right) \tag{6}$$

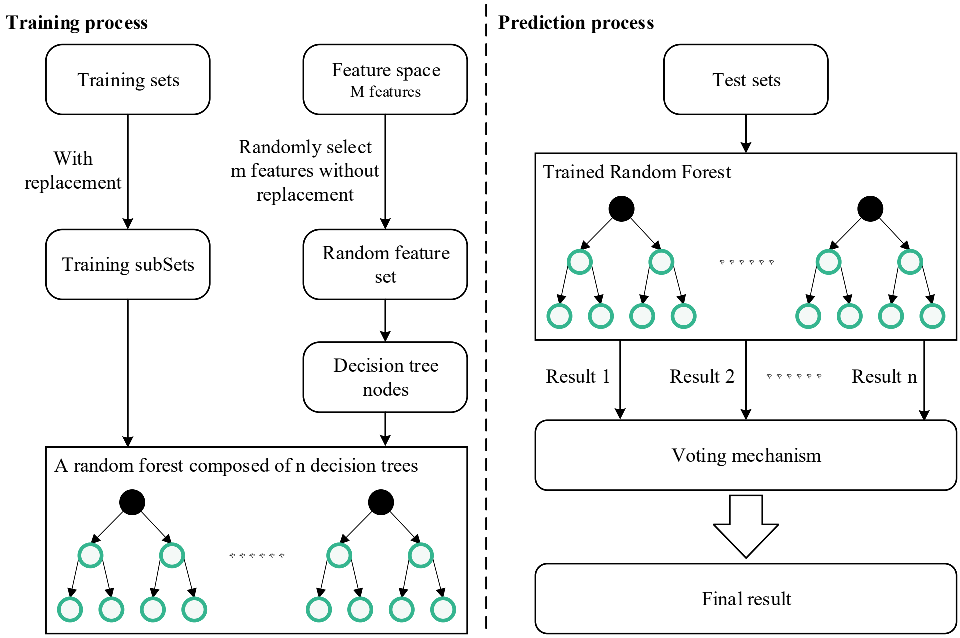
where G_m is the characteristic Gini index, K is the number of categories, m is the current number of nodes, and $\hat{p}_{mk}$ is the estimated probability that node m belongs to the k^{th} category.

2 The importance of the features at the nodes is shown below.

$$VIM_j^{(G)} = G_m - G_l - G_r \tag{7}$$

where G_l and G_r are the Gini indices of l and r nodes respectively.

Figure 2 The algorithm process of RF (see online version for colours)



3 An improved random forest feature selection algorithm based on the principle of equivalent infinitesimal

The RF algorithm has a very high computational complexity due to the large number of logarithmic functions that must be calculated during the calculation. This paper improves the RF algorithm (IRF), uses the Taylor formula (Bonfiglioli, 2009) to approximately replace the complex functions in the formula, changes the information gain ratio function to only need addition and subtraction operations to compare the impact of different feature selections on decision tree generation, and then introduces the mean value of the Gini coefficient to adjust the error caused by feature redundancy, improving the calculation efficiency and accuracy of the algorithm.

During the RF training feature selection process, the higher the information gain ratio, the more significant the feature is as a splitting feature in reducing the uncertainty of the sample. Therefore, During the RF training feature selection process, the higher the information gain ratio, the more significant the feature is as a splitting feature in reducing the uncertainty of the sample. Therefore, the feature with the highest information gain ratio should be selected as the classification feature for this node. For the Gini index gini, if the greater the redundancy between features, the greater the correlation between features, then the Gini index between this feature and other features will be smaller (Hasan et al., 2016), and the formula for the Gini index between features is as follows.

$$sum_g inisplit\left(A_F T\right) = \sum_{i=1}^{s} \sum_{j=1}^{x} \left[\frac{|T_{ij}|}{|T|} gini\left(T_{ij}\right) \right] \tag{8}$$

where S is the number of features, x is the number of features except DT, $|T_{ij}|$ is the number of samples when the j^{th} feature value is taken, and $gini(T_{ij})$ is the Gini index when the j^{th} feature is taken, as shown below.

$$gini\left(T_{ij}\right) = 1 - \sum_{k=1}^{n}\left(\frac{|TA_{ijk}|}{|T_{ij}|}\right)^2 \tag{9}$$

The mean of the gini sums of feature A and other characteristics is computed as bellow.

$$\overline{sum_g inisplit\left(A_F T\right)} = \frac{\sum_{i=1}^{s}\sum_{j=1}^{x}\left[\frac{|T_{ij}|}{|T|} gini\left(T_{ij}\right)\right]}{s} \tag{10}$$

Therefore, in the feature selection process based on RF, the gini mean value between features is added to improve the accuracy of feature selection. After this parameter is added, the information gain rate is calculated as a new split information (Prasetiyowati et al., 2021), as shown below.

$$GainR(A) = \frac{Gain(A)}{SplitInfo(AT) - \overline{sum_g inisplit\left(A_F T\right)}} \tag{11}$$

From the above formula, we can see that the greater the gini between features, the lower the correlation between features and the lower the redundancy, and the greater the value of $GainR(A)$ will be. Therefore, IRF is used to reduce the impact of variable redundancy on the accuracy of feature selection during feature splitting, thereby improving the accuracy of feature screening in the algorithm. In the process of calculating information gain, different feature selections correspond to different information gain values. Each function in $Gain(A)$ is broken down as follows.

$$\begin{aligned} Gain(A) &= \sum_{j=1}^{k}\frac{s_j}{s} \times \sum_{i=1}^{m} p_{ij}\log_2\left(p_{ij}\right) \\ &= \frac{1}{s}\sum_{j=1}^{k} s_j \times \left(\frac{s_{1j}}{s_j}\log_2\frac{s_{1j}}{s_j} + \ldots + \frac{s_{pj}}{s_j}\log_2\frac{s_{pj}}{s_j} + \ldots + \frac{s_{mj}}{s_j}\log_2\frac{s_{mj}}{s_j}\right) \\ &= \frac{1}{ln2 \times s}\sum_{j=1}^{k}\left(s_{1j} ln\left(\frac{s_{1j}}{s_j}\right) + \ldots + s_{mj} ln\left(\frac{s_{mj}}{s_j}\right)\right) \end{aligned} \tag{12}$$

To simplify the above formula, this paper mainly uses the principle of equivalent infinitesimal to approximately replace the formula. When $x \in (0, 1)$, $\ln(x) \approx (x-1) - \frac{1}{2}(x-1)^2 + \frac{1}{3}(x-1)^3$, equation (12) can be simplified as follows.

$$Gain(A) = \frac{1}{ln2 \times s}\sum_{j=1}^{k}\sum_{i=1}^{m}\left[\frac{s_{ij}\left(s_{ij} - s_j\right)\left(11s_j^2 + 211s_{ij}^2 - 7s_j s_{ij}\right)}{6s_j^3}\right] \tag{13}$$

where $ln2$ and s are both constants, and the magnitude of the constants does not change with different feature selections. Therefore, to further simplify the formula to improve calculation efficiency, the two constant terms in the formula can be removed to make the formula more concise. The simplified formula is as follows.

$$Gain(A)' = \sum_{j=1}^{k}\sum_{i=1}^{m}\left[\frac{s_{ij}\left(s_{ij}-s_j\right)\left(11s_j^2+211s_{ij}^2-7s_js_{ij}\right)}{6s_j^3}\right] \tag{14}$$

The final information gain rate is therefore simplified as follows.

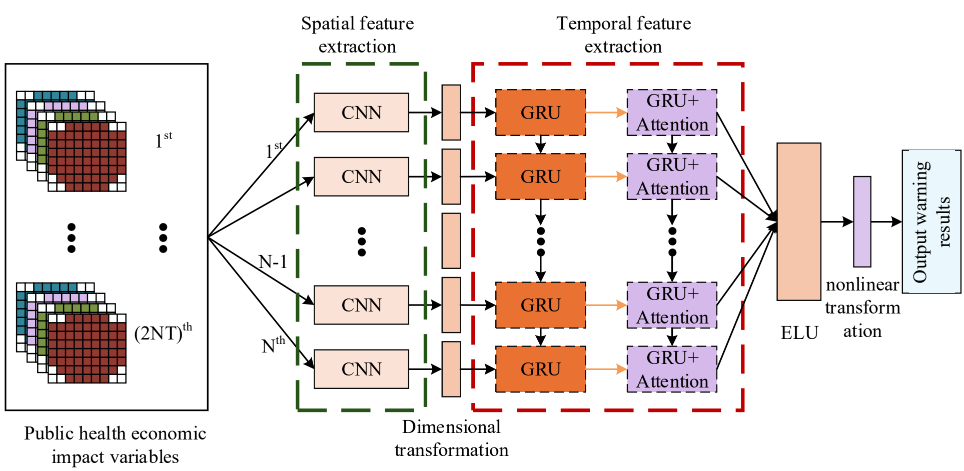
$$GainR(A)' = \frac{Gain(A)'}{SplitInfo(AT) - \overline{sum_g inisplit\left(A_F T\right)}} \tag{15}$$

4 Entity relationship recognition based on improved BiLSTM-CRF model and context awareness

4.1 Selection and pre-processing of variables affecting public health economics

To address the issue of insufficient feature extraction of a single neural network, which leads to low accuracy of early warning, this paper first selects public health economic impact variables and normalises them, and then introduces a residual CNN to extract the spatial features of the variables, uses a GRU to remember temporal features, and uses a self-attention mechanism (SAM) (Chen et al., 2023) to enhance key spatio-temporal features. Second, the IRF is used to screen for spatio-temporal features, and finally the screened spatio-temporal features are nonlinearly transformed and mapped to the monitoring and early warning results of public health economics. The overall model structure is shown in Figure 3.

Figure 3 The overall model structure (see online version for colours)



According to a thorough summary of previous research (Terza et al., 2008), the impact variables of public health economics include direct economic indicators, indirect economic indicators, and macroeconomic indicators. Direct economic indicators include medical expenditures, public health project inputs, etc. Indirect economic indicators include tourism, transportation, and services. Macroeconomic indicators include GDP growth rate and unemployment rate. These indicators are denoted as x_1, x_2, …, x_n. Due to the large difference in value ranges among the impact indicators, and in order to prevent overfitting caused by the differences, the impact variables are pre-processed by normalisation, as shown below.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{16}$$

where x is the impact indicator variable, min(x) and max(x) represent the maximum and minimum values of the impact indicator variable, respectively, and x' is the normalised impact indicator variable.

4.2 Spatio-temporal feature extraction of public health economic impact variables based on hybrid neural networks

After normalising the influencing variables, this paper uses the convolution layer of RC to perceive the spatial relationship between features and neighbouring features, and uses this to extract the spatial features between variables. Then a GRU is used to remember temporal information, and a SAM is used to extract temporal features.

1 Spatial feature extraction based on residual CNN. The CNN extracts adjacent data blocks from the normalised variables, and then performs spatial transformation and reorganisation on the data blocks by convolution kernel operation, so that each spatial position in the output features corresponds to a spatial position in the input features. Therefore, the convolutional level of CNN can well perceive the relationship between data points and adjacent data points, and capture the spatial relationship between features and adjacent features. The formula for calculating the convolution level is as follows.

$$x_j^n{}' = \sigma\left(\sum_{i \in M_j} w_{ij}^n x_i^{n-1}{}' + b_j^n\right) \tag{17}$$

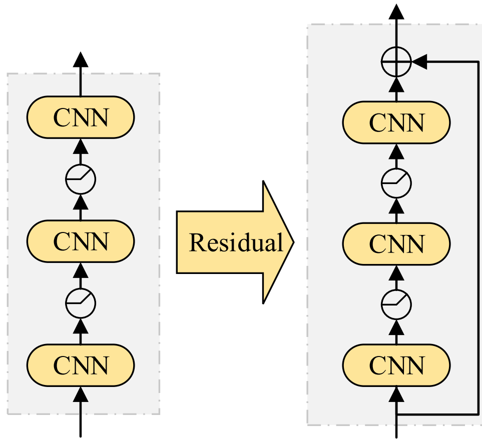
where $x_j^n{}'$ is the j^{th} element of the nth convolution; M_j is the feature matrix of the convolution input; w_{ij}^n is the weight matrix; $x_i^{n-1}{}'$ is the input variable; b_j^n is the bias matrix; and σ is the ReLU activation function.

However, although deepening the convolutional layer can extract deeper spatial features, deepening the network can cause the problem of gradient disappearance, which affects the accuracy of public health economic early warning. To solve this problem, this paper introduces residuals into the CNN. Using the residual structure, deeper networks can be constructed, allowing for better information flow and more stable deep spatial features (Asparouhov and Muthén, 2023). The formula for extracting spatial features from RC is as follows. The method of constructing RC through residuals is shown in Figure 4.

$$x^* = x' + C(x') \tag{18}$$

where x^* is the output of the RC; x' is the input to the influencing variable; and C is the output of the last CNN layer.

Figure 4 The method of constructing RC through residuals (see online version for colours)



2 Considering the time characteristics of public health economic variables, the GRU is selected to capture their time characteristics. Although the GRU can directly extract the time information in the feature input, there may be hidden information in the output after convolution processing. Therefore, the output of the RC is selected as the input of the GRU to extract more useful information to the greatest extent. The GRU is stacked from two GRU layers, and the SAM is introduced in the hidden layer of the GRU behind, so that useful features can be accurately extracted according to the importance of the information, and irrelevant information can be ignored (Li et al., 2020). The processing process inside the GRU can be expressed by the following equation.

$$u(t) = \sigma\left(W_u \odot \left[x_t^*, h_{t-1}\right] + b_u\right) \tag{19}$$

$$r_t = \sigma\left(W_u \odot \left[x_t^*, h_{t-1}\right] + b_r\right) \tag{20}$$

$$c_t = \tanh\left(W_c \left[x_t^*, (r_t \odot h_{t-1})\right] + b_c\right. \tag{21}$$

$$h_t = u_t * h_{t-1} + \left(1 - u_t\right) * c_t \tag{22}$$

where x_t^* is the spatial feature variable of the RC output, which is used as the input of the GRU at the current time; r_t is the reset gate; c_t is the current time's stored content; h_{t-1} is the hidden state of the previous time; and h_t is the output of the GRU hidden layer. Then, the attention probability distribution value a_k of the time characteristics of each variable and the feature variable v containing key information are calculated as follows.

$$v = \sum_{t=1}^{N} a_t h_t \tag{23}$$

$$a_t = \frac{\exp(e_t)}{\sum_{j=1}^{t} e_j} \tag{24}$$

$$e_t = w_t \tanh(W_t h_t + b_t) \tag{25}$$

where e_t is the spatiotemporal feature enhanced by SAM; w_t and W_t are the weight coefficients input at time t; and b_t is the corresponding bias at time t.

4.3 *Variable selection based on the temporal and spatial characteristics of the improved random forest algorithm*

After obtaining the spatio-temporal features of the influencing variables, this paper uses IRF to obtain the importance scores of the spatio-temporal feature variables and perform feature screening. Since IRF is an integration of several DTs and each DT contains many decision nodes, when using IRF to evaluate the importance of each spatio-temporal feature variable, the change in the Gini index before and after the decision node branch is used to measure the spatio-temporal feature variable. The specific steps for calculating the importance score of a temporal feature variable in IRF are given below.

1 Calculate the Gini index of decision node m in the i^{th} tree for e_j, $|y|$ is the total number of categories, and p_{mk} implies the proportion of the k^{th} category at decision node m.

$$gini_{im}(e_j) = \sum_{k-1}^{|y|} \sum_{k' \neq k} p_{mk}(1 - p_{mk}) = 1 - \sum_{k=1}^{|y|} p_{mk}^2 \tag{26}$$

2 Calculate the importance of the Gini index of e_j before and after the m branch of the above decision node. l represents the left branch and r represents the right branch.

$$VIM_{im}(e_j) = gini_{im}(e_j) - gini_{il}(e_j) - gini_{ir}(e_j) \tag{27}$$

3 Calculate the information gain rate $GainR(A)'$ of each tree, extract the features whose $GainR(A)'$ is 0, calculate the sum of the Gini index changes of e_j on all decision nodes in the i^{th} tree, and M is the set of all decision nodes.

$$VIM_i(e_j) = \sum_{m \in M} VIM_{im}(e_j) \tag{28}$$

4 Calculate the total change in the Gini index $VIM(e_j) = \sum_{i=1}^{n} VIM_i(e_j)$ of variable e_j over all DTs in the IRF, n is the number of decision trees.

5 Repeat step (1) to step (4) to calculate the total Gini index change of all spatio-temporal features on the IRF and sum them up. Normalise a spatio-temporal feature to finally obtain the importance score of the spatio-temporal feature in the IRF. A

feature subset is generated by removing one feature at a time from the sorted feature set. The accuracy of the feature subset is calculated, and the subset with the highest accuracy is finally selected as the screened feature subset.

$$VIM^{*}\left(e_j\right)=\frac{VIM\left(e_j\right)}{VIM\left(e_j\right)} \tag{29}$$

4.4 *Output of early warning results of public health economic monitoring*

Finally, the spatio-temporal feature data filtered by the IRF feature is used as the input of the output layer. At this time, the spatio-temporal feature data contains the correlation information of space and time. The relationship between spatio-temporal features is captured through a nonlinear transformation and mapped to the results of public health economic monitoring and early warning. The calculation equation is as follows.

$$y_i=\sigma\left(we_i+b_i\right) \tag{30}$$

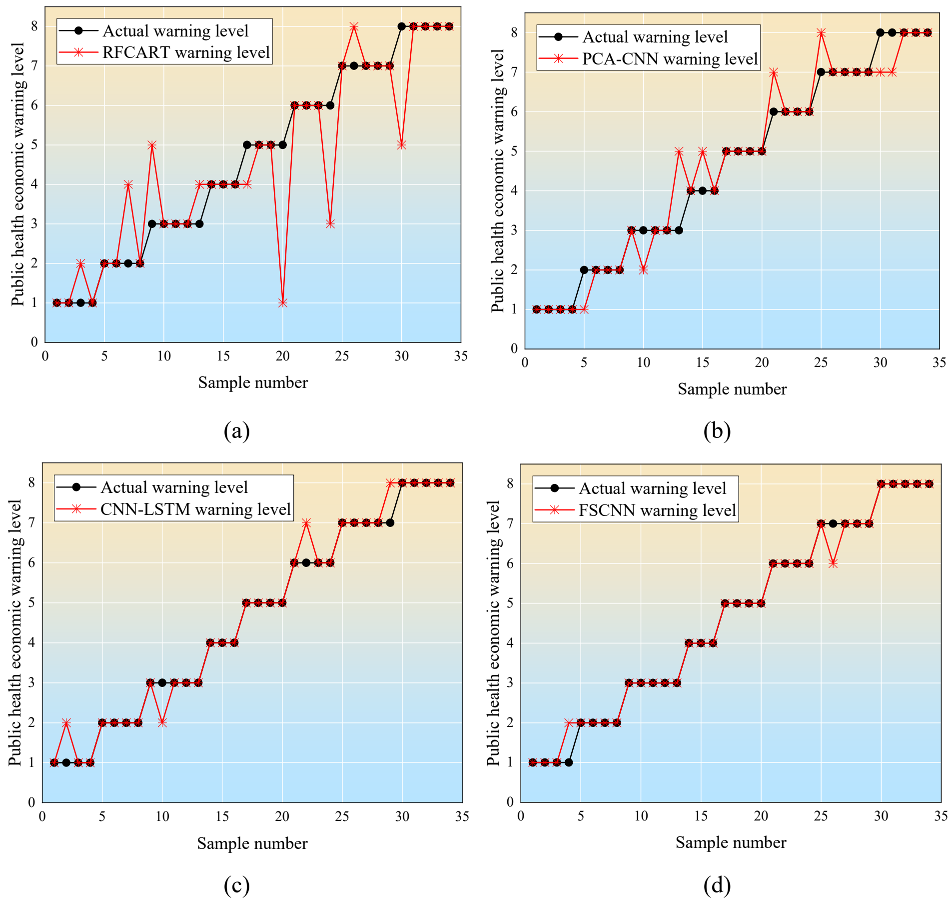
where y_i is the early warning value for public health economic monitoring; σ is the ReLU activation function; w is the weight matrix; and b_i is the bias matrix.

5 Experimental results and analyses

The dataset used in this article is from the global burden of disease (GBD) dataset. This dataset contains data on the global burden of different diseases, injuries and risk factors in countries around the world, such as disease incidence, mortality, disability-adjusted life years (DALY), etc., as well as various factors related to public health economics, such as direct economic indicators, indirect economic indicators and macroeconomic indicators. There are a total of 26,941 items. The dataset is divided into a training set, a validation set and a test set in a ratio of 7:2:1. This article classifies the results of public health economic monitoring and early warning into excellent, good, average, cold, light, medium, heavy and huge warning states, and labels their warning levels in order as 1, 2, 3, …, 8. All models are built using Python 3.8 and the Pytorch 1.6 framework. In addition, the experiment was performed on a computer configured with 16GB of RAM, an Intel i5-13400 F CPU and an RTX 4070 GPU. In the experiment, the optimiser used Adam, with an initial learning rate of 0.001, a training batch of 128, and a maximum epoch number set to 6,000.

To compare the results of public health economic early warning of different models, RFCART (Botz et al., 2022), PCA-CNN (Sansone and Zhu, 2023), CNN-LSTM (Azadi et al., 2023) and the proposed model FSCNN were selected for comparative experiments, as shown in Figure 5. RFCART and PCA-CNN have relatively high rates of misjudging the warning level. Among the 34 samples, the numbers of misjudgments are 7 and 8 respectively, and the accuracies are 79.41% and 76.47% respectively. The PCA-CNN had 4 false alarms and an accuracy of 88.23%, while the FSCNN had the best early warning results, with an accuracy of 94.12%. A longitudinal comparison shows the overlap between the predicted sample grades and the actual sample grades. A higher overlap indicates a better early warning effect. The grade recognition effect of FSCNN is better than the other three models. This shows that the FSCNN's public health economic monitoring and early warning results are better.

Figure 5 Public health economic early warning results from different models (a) monitoring and early warning results of RFCART (b) PCA-CNN (c) CNN-LSTM and (d) FSCNN (see online version for colours)



The loss curves of different models are shown in Figure 6. As can be seen from the figure, when the number of iterations reaches 1,500, the loss of FSCNN reaches about 0.05, while the loss of CNN-LSTM only reaches 0.05 after 2,000 iterations. The loss values of RFCART and PCA-CNN are generally lower than those of CNN-LSTM and FSCNN. FSCNN, a hybrid network model that incorporates an attention mechanism, has lower loss values during model training, better fitting results and faster training speed than the other three types of network models.

In addition to analysing the early warning results and loss functions of each model, this paper also uses quantitative indicators such as Accuracy, MAE, RMSE and the coefficient of determination R2 to fully evaluate the early warning performance of different models, as shown in Table 1. The FSCNN has an accuracy of 93.36%, which is an improvement of 14.85%, 12.31% and 5.07% over RFCART, PCA-CNN and CNN-LSTM, respectively. The MAE and RMSE of the FSCNN are 0.3518 and 0.7159 respectively, which is at least 11.21% lower than the other three models. Comparing R2 again, FSCNN's R2 is closest to 1, indicating its highest accuracy in early warning. Although RFCART screens the characteristics of public health economics through RF, it does not optimise RF. It only uses a single model to provide early warnings of the results of public health economic monitoring, resulting in large forecast losses. PCA-CNN is an early warning model based on a single neural network. It can only extract local features through the CNN and does not consider other features. CNN-LSTM does not filter features, so its early warning performance is not as good as that of the FSCNN model. Therefore, FSCNN has excellent early warning accuracy and precision, and can more accurately monitor and warn of public health emergencies.

Figure 6 The loss curves of different models (see online version for colours)

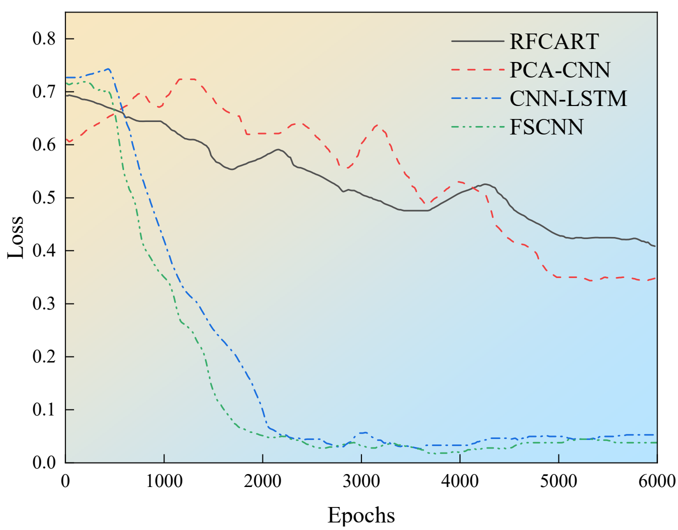


Table 1 Comparison of warning performance indexes of different models

Model	*Accuracy*	*MAE*	*RMSE*	*R2*
RFCART	78.51%	0.4861	0.8452	0.8381
PCA-CNN	81.05%	0.4289	0.8163	0.8936
CNN-LSTM	88.29%	0.3962	0.7841	0.9156
FSCNN	93.36%	0.3518	0.7159	0.9439

6 Conclusions

For the goal of dealing with the issue of insufficient feature extraction and feature redundancy in the current early warning method for public health economic monitoring, the RF split feature selection function is first simplified based on the principle of equivalent infinitesimal to improve the efficiency of function operation. The Gini coefficient value of the non-category attribute is introduced to overcome the experimental errors caused by variable redundancy and improve the calculation efficiency and screening performance of the RF algorithm. Then, public health economic impact variables are selected and normalised. Next, a hybrid neural network model is designed to extract the temporal and spatial features of the variables. A residual structure is introduced to improve the CNN to extract deeper spatial features. Meanwhile, the time features are extracted by GRU and integrated into SAM to enhance the temporal and spatial features. Finally, the IRF is used to further screen the spatio-temporal features to obtain the most important spatio-temporal features for the early warning results, and then perform a nonlinear transformation to map them to the monitoring and early warning results of public health economics. The experimental outcome indicates that the MAE and RMSE of the offered model are at least 11.21% lower than those of the other three models. It has high early warning accuracy and can achieve accurate monitoring and early warning of public health economics.

Acknowledgements

This work is supported by the Hebei Province Social Science Development Research Project (No. 202402327).

Declarations

All authors declare that they have no conflicts of interest.

References

Asparouhov, T. and Muthén, B. (2023) 'Residual structural equation models', *Structural Equation Modeling: A Multidisciplinary Journal*, Vol. 30, No. 1, pp.1–31.

Azadi, M., Yousefi, S., Saen, R.F. et al. (2023) 'Forecasting sustainability of healthcare supply chains using deep learning and network data envelopment analysis', *Journal of Business Research*, Vol. 154, p.113357.

Bonfiglioli, A. (2009) 'Taylor formula for homogenous groups and applications', *Mathematische Zeitschrift*, Vol. 262, pp.255–279.

Botz, J., Wang, D., Lambert, N. et al. (2022) 'Modeling approaches for early warning and monitoring of pandemic situations as well as decision support', *Frontiers in Public Health*, Vol. 10, p.994949.

Chen, Y., Wei, G., Liu, J. et al. (2023) 'A prediction model of student performance based on self-attention mechanism', *Knowledge and Information Systems*, Vol. 65, No. 2, pp.733–758.

Correia, S., Luck, S. and Verner, E. (2022) 'Pandemics depress the economy, public health interventions do not: evidence from the 1918 flu', *The Journal of Economic History*, Vol. 82, No. 4, pp.917–957.

Devyatkin, D., Otmakhova, Y., Usenko, N. et al. (2021) 'Deep learning approaches to mid-term forecasting of social-economic and demographic effects of a pandemic', *Procedia Computer Science*, Vol. 190, pp.156–163.

Dou, Q., Zhang, J. and Jing, B. (2023) 'A ML-based economic protection development level using decision tree and ensemble algorithms', *Soft Computing*, Vol. 27, No. 24, pp.18929–18947.

Ebi, K.L. and Schmier, J.K. (2005) 'A stitch in time: improving public health early warning systems for extreme weather events', *Epidemiologic Reviews*, Vol. 27, No. 1, pp.115–121.

Hasan, M.A.M., Nasser, M., Ahmad, S. et al. (2016) 'Feature selection for intrusion detection using random forest', *Journal of Information Security*, Vol. 7, No. 3, pp.129–140.

Khan, A., Sohail, A., Zahoora, U. et al. (2020) 'A survey of the recent architectures of deep convolutional neural networks', *Artificial Intelligence Review*, Vol. 53, pp.5455–5516.

Kolomoyets, A.V., Hbur, Z.V., Koshova, S.P. et al. (2021) 'Financial and economic effect for the healthcare institution from the introduction of logistics management methods', *Wiadomości Lekarskie*, Vol. 74, No. 6, pp.1499–1504.

Kuo, C-C.J. (2016) 'Understanding convolutional neural networks with a mathematical model', *Journal of Visual Communication and Image Representation*, Vol. 41, pp.406–413.

Li, W., Qi, F., Tang, M. et al. (2020) 'Bidirectional LSTM with self-attention mechanism and multi-channel features for sentiment classification', *Neurocomputing*, Vol. 387, pp.63–77.

Lu, Z.-N., Chen, H., Hao, Y. et al. (2017) 'The dynamic relationship between environmental pollution, economic development and public health: evidence from China', *Journal of Cleaner Production*, Vol. 166, pp.134–147.

Ma, R., Liu, J. and An, S. (2023) 'The early warning mechanism of public health emergencies through whistleblowing: a perspective based on considering the uncertainty of risk perception', Risk Management and Healthcare Policy, Vol. 5, pp. 503-523.

Mao, K., Zhang, K., Du, W. et al. (2020) 'The potential of wastewater-based epidemiology as surveillance and early warning of infectious disease outbreaks', *Current Opinion in Environmental Science and Health*, Vol. 17, pp.1–7.

Meckawy, R., Stuckler, D., Mehta, A. et al. (2022) 'Effectiveness of early warning systems in the detection of infectious diseases outbreaks: a systematic review', *BioMed Central Public Health*, Vol. 22, No. 1, p.2216.

Paul, A., Mukherjee, D.P., Das, P. et al. (2018) 'Improved random forest for classification', *IEEE Transactions on Image Processing*, Vol. 27, No. 8, pp.4012–4024.

Prasetiyowati, M.I., Maulidevi, N.U. and Surendro, K. (2021) 'Determining threshold value on information gain feature selection to increase speed and prediction accuracy of random forest', *Journal of Big Data*, Vol. 8, No. 1, p.84.

Rappold, A.G., Fann, N.L., Crooks, J. et al. (2014) 'Forecast-based interventions can reduce the health and economic burden of wildfires', *Environmental Science and Technology*, Vol. 48, No. 18, pp.10571–10579.

Ruck, D.J., Bentley, R.A. and Borycz, J. (2021) 'Early warning of vulnerable counties in a pandemic using socio-economic variables', *Economics and Human Biology*, Vol. 41, pp.100988.

Sansone, D. and Zhu, A. (2023) 'Using machine learning to create an early warning system for welfare recipients', *Oxford Bulletin of Economics and Statistics*, Vol. 85, No. 5, pp.959–992.

Speiser, J.L., Miller, M.E., Tooze, J. et al. (2019) 'A comparison of random forest variable selection methods for classification prediction modeling', *Expert Systems with Applications*, Vol. 134, pp.93–101.

Szreter, S. and Woolcock, M. (2004) 'Health by association? Social capital, social theory, and the political economy of public health', *International Journal of Epidemiology*, Vol. 33, No. 4, pp.650–667.

Terza, J.V., Bradford, W.D. and Dismuke, C.E. (2008) 'The use of linear instrumental variables methods in health services research and health economics: a cautionary note', *Health Services Research*, Vol. 43, No. 3, pp.1102–1120.

Wang, J., Qin, Z., Hsu, J. et al. (2024) 'A fusion of machine learning algorithms and traditional statistical forecasting models for analyzing American healthcare expenditure', *Healthcare Analytics*, Vol. 5, p.100312.

Yin, X., Li, J. and Huang, S. (2022) 'The improved genetic and BP hybrid algorithm and neural network economic early warning system', *Neural Computing and Applications*, Vol. 34, pp.1–10.

Zarzycki, K. and Ławryńczuk, M. (2022) 'Advanced predictive control for GRU and LSTM networks', *Information Sciences*, Vol. 616, pp.229–254.

Zheng, S. and Hu, X. (2021) 'Early warning method for public health emergency under artificial neural network in the context of deep learning', *Frontiers in Psychology*, Vol. 12, p.594031.