JOURNAL OF SOCIAL COMPUTING
ISSN 2688-5255 02/06 pp206−218
Volume 3, Number 3, September 2022
DOI: 10.23919/JSC.2022.0012

A Novel Hybrid Model for Gasoline Prices Forecasting Based on Lasso and CNN

Hu Yang*, Xinlu Tian, Xin Jin, and Haijun Wang*

Abstract: Gasoline is the lifeblood of the national economy. The forecasting of gasoline prices is difficult because of frequent price fluctuations, its complex nature, diverse influencing factors, and low accuracy of prediction results. Previous studies mainly focus on forecasting gasoline prices in a single region by single time series analysis which ignores the daily price co-movement of different series from multiple regions. Because price co-movement may contain useful information for price forecasting, this paper proposes the Lasso-CNN ensemble model that combines statistical models and deep neural networks to forecast gasoline prices. In this model, the Least Absolute Shrinkage and Selection Operator (Lasso) screens and chooses the correlated time series to enhance the performance of forecasting and avoid overfitting, while Convolutional Neural Network (CNN) takes the selected multiple series as its input and then forecasts the gasoline prices in a certain region. Forecasting results of gasoline prices at the national level and regional levels by using the new method demonstrate that the new approach provides more accurate results for the predictions of gasoline prices than those results generated by alternative methods. Thus, the relevant series can enhance the performance of forecasting and help to gain better results.

Key words: gasoline prices; forecasting; Lasso-CNN; multiple time series

1 Introduction

Since the outbreak of the COVID-19 pandemic in December 2019, a lot of countries around the world have adopted the lockdown policy and restricted social contact to prevent the spread of the disease, which not only reduces people's consumption of goods but also decreases the need for travel. As a result, there is a drop in the demand for gasoline and the price of gasoline falls accordingly. Recently, with the conflict between Russia and Ukraine, energy prices have risen sharply because of insufficient energy supplies. Such unexpected events undoubtedly have significant impacts on the demand and supply of crude oil, resulting in great fluctuations in gasoline prices. Gasoline is extremely important to the economy and functions as a key factor in the Consumer Price Index (CPI). Sharp fluctuations in gasoline prices would affect the CPI, bring about inflation, and hurt the economy. Meanwhile, gasoline prices play a role in macroeconomic activities. Since an oil price shock is strongly correlated with the corresponding economic recession, gasoline prices are more likely to explain the recession than gasoline spending does[1]. Gasoline prices also affect home prices and home foreclosure rates[1, 2] . Consumer activities responding to the changes in gasoline prices have been studied to explain a variety of economic phenomena, such as demand for automobiles[3, 4] , choice of transportation[5], search behavior[6], and sticky prices[7]. Furthermore, gasoline price predictions play a key role in depicting the automobile market through microeconomic models and

• Hu Yang, Xinlu Tian, and Xin Jin are with the School of Information, Central University of Finance and Economics, Beijing 100872, China. E-mail: hu.yang@cufe.edu.cn.
• Haijun Wang is with the School of Economics, Beijing Wuzi University, Beijing 101149, China. E-mail: wanghaijun2005@126.com.
∗ To whom correspondence should be addressed.
† Hu Yang and Xinlu Tian contribute equally to this paper.
Manuscript received: 2022-05-04; revised: 2022-09-22; accepted: 2022-09-28

analyzing environmental policies[8, 9]. As the lifeblood of the nation's economy, the balance between gasoline prices and others in the market affects the normal operation of the nation's economic activities, which directly impacts people's life. As a result, fluctuations in gasoline prices would directly affect the development of the global economy. Thus, forecasting gasoline prices has important practical value.

At present, researchers have conducted in-depth analyses and research on oil prices by econometric methods, mathematical methods, and a combination of different models. In these studies of oil prices, economic time series methods are mostly used, including the Granger causality test[10], the cointegration theory proposed by Engle and Granger, the Autoregressive Conditional Heteroskedastic (ARCH) model, the ARCH family proposed by Engle, the Vector Autoregression (VAR) model, impulse response function, etc., which are used to predict the oil prices in the international oil market[11, 12]. For oil price predictions performed by mathematical methods, Abramson and Finizza[13] used the belief networks to forecast oil prices; Tang and Hammoudeh[14] proposed an oil prices prediction model to analyze and predict oil price fluctuations based on the target zone model. The combination of the aforementioned methods and other forecasting methods can grasp the factors of commodity price changes and get better forecasting results. Abramson and Finizza[13] proposed a combined prediction model of belief network and probability model to predict oil prices; Feng et al.[15] combined the Autoregressive Integrated Moving Average (ARIMA) model with the neural network to predict oil prices; Fan et al.[16] used the wavelet neural network to predict oil prices while optimizing the network input by principal component analysis. The results show that the combined forecasting method has higher accuracy than each of the methods. Although the methods of forecasting international crude oil prices can be extended to gasoline prices forecasting, most of the methods can only be applied to gasoline prices forecasting in a single region. And, those methods are not suitable for gasoline price forecasting in multiple regions.

The gasoline prices system is essentially a nonlinear and non-stationary complex system. The flow of gasoline through trade subjects forms a gasoline prices network, which contains the interrelationship among gasoline trade subjects and interaction information among them. When forecasting gasoline prices of target regions, one can consider gasoline prices in different regions as nodes of a complex network. Thus, the interrelationship of gasoline prices among different regions can be fully considered to assist the prediction. The challenges of employing the interrelationship and interaction information among multiple series include (1) which of the correlated series should be considered to implement the forecasting model for a target series; and (2) how to combine correlated series to enhance the forecasting model's performance. To resolve these two issues, this paper proposes a new hybrid approach, Lasso-CNN, a combination of statistical models and deep neural networks. Taking the task of forecasting gasoline prices in multiple regions in the US as an example, the new method screens and selects a set of relevant series by Least Absolute Shrinkage and Selection Operator (Lasso), and then it uses Convolutional Neural Network (CNN) to implement the gasoline prices forecasting. The remaining parts of this paper are as follows. Section 2 provides a review of related research. In Section 3, we propose an ensemble model, which illustrates the processes of modeling step-by-step. Section 4 demonstrates the application of the Lasso-CNN ensemble model by using a dataset of real-world US gasoline prices. Finally, Section 5 draws the conclusion.

2 Related Studies

2.1 Time series forecasting methods

Time Series Forecasting (TSF) methods are widely used in various fields. Similar to other TSF tasks, the forecasting of gasoline prices can be considered as the problem of TSF. TSF methods are of great importance in the real world, which can solve the following problems, such as network traffic, weather or pollution forecasting, and the stock markets[17]. The prediction process of time series is mainly divided into two types of computation, the first one involves statistical methods and the second one involves soft computing methods, which include machine learning and deep learning.

Generally, the goal of gasoline prices forecasting is to predict the value of gasoline price at $t+h$ using available observations from a time series at time t. Suppose there is only a single time-dependent variable that is available, the problem can be reframed by using the Univariate Time Series (UTS) analysis methods, which can be formulated as follows:

$$\widehat{y}_{t+h} = f(y_t, y_{t-1}, \ldots, y_{t-k}; \theta) \quad (1)$$

where $Y_t = \{y_t, y_{t-1}, \ldots, y_{t-k}\}$ refers to time series, θ is the parameter such as autoregression coefficient, $\widehat{y}_{t+h}$ is the forecasting value at $t+h$, k is the number of inputs, and $h = 1, 2, \ldots$ is any positive integer. With other time series, the problem of gasoline price forecasting becomes a problem of Multivariate Time Series (MTS) analysis, which can be formulated as

$$\widehat{y}_{j,t+h} = f(Y_{1,t}, Y_{2,t}, \ldots, Y_{m,t}; \theta) \quad (2)$$

where $Y_{1,t}, Y_{2,t}, \ldots, Y_{m,t}$ refer to multi-target series, $Y_{j,t} = \{y_{j,t}, y_{j,t-1}, \ldots, y_{j,t-k}\}$, and m is the number of multiple series for $j = 1, 2, \ldots, m$.

Intuitively, gasoline prices are in the form of time series, and gasoline prices are the feedback of the market. So, one can use the time series method to predict gasoline prices. The main idea is to predict some uncertain values in the future based on historical information. In the past, some linear and non-linear statistical methods are used to solve the TSF problem, such as linear regression, Historical Average (HA), Autoregressive Integrated Moving Average (ARIMA) model[18], Spatio-Temporal Autoregressive Integrated Moving Average (STARIMA) model[19], and Threshold Autoregressive (TAR) model[20]. In addition to those traditional statistical methods, machine learning models, such as Decision Trees (DTs)[21], Support Vector Machines (SVMs)[22], and Hidden Markov Models (HMMs)[23], can also model and learn time series forecasting problems because of the close connection between the nature of time series prediction and the regression analysis of machine learning.

Inspired by the notable achievements accomplished by deep learning in many fields, such as natural language processing[24], image classification[25], and reinforcement learning[26], several Artificial Neural Network (ANN) algorithms have drawn people's attention and established that they are strong contenders among statistical methods in the forecasting community due to their higher accuracies of prediction[27]. Significantly different from machine learning methods that require hand-crafted features, deep learning methods have a great potential to learn complex non-linear temporal feature and their interactions among multiple series. Because deep learning methods automatically learn complex data representations of an MTS, it reduces the amount of work for manual feature engineering and model design[28, 29] . Moreover, deep learning methods can learn the linear and nonlinear patterns of data better. They can help us to better understand the data and make accurate time series forecasting models, thus they can be used to effectively extract the hidden feature of the original data[30]. Therefore, gasoline price forecasting can also be implemented by state-of-the-art deep learning methods, such as CNN and Deep Belief Nets (DBN)[31], which have accomplished remarkable achievements in the fields of image recognition, speech processing, and natural language processing. In addition, the Recurrent Neural Network (RNN) can store a lot of information about the past situation and allow updates to the hidden state dynamically[32−34]. To address the weakness of RNNs in terms of managing long-term dependencies, the Long-Short Term Memory (LSTM)[35], which is a variant of RNN that is capable of learning long-term dependence, has also been employed for series forecasting[36]. LSTM comprises a separate autoencoder and forecasting sub-models. Different from RNN, an RNN architecture can resolve the issue of vanishing gradient. The Gate Recurrent Unit (GRU)[37] is also an important variant of RNN: its basic idea of learning long-term dependence is consistent with LSTM; however, it only has one reset gate and one update gate.

2.2 Lasso regression for variable selection

Although including many time series in the forecasting process reduces the training error and helps to obtain an unbiased model, it will increase the variance of forecasting and generate inaccurate prediction results. The variable selection method is an effective tool to solve the issue: one can only select the relevant variables to build the prediction model and discard other variables. Under the sparsity assumption, estimates are sparse if many of their components are zero or approximate to zero. The shrinkage methods are useful and efficient to implement variable selection by shrinking some of the coefficients to zero. Among these methods, the Lasso or the l_1 penalty is one of the most popular and easiest methods proposed for variable selection[38]. The Lasso can help some statistical and machine learning models to select relevant variables, make the model more stable, and enhance the interpretability of the models. Lasso has been widely used in a variety of fields. For instance, in previous studies, researchers used Lasso to screen for effective predictors for crude oil prices forecasting[39] and used machine learning methods with relief and feature selection techniques to make an efficient prediction model for cardiovascular disease analysis[40]. Similarly, researchers used Lasso-WOA to explore the carbon emissions from commercial building operations. And, they utilized the Lasso regression to estimate the results. The results show that the major driving forces of carbon emissions from commercial buildings in China are the population size and energy intensity of carbon emissions[41]. Besides, variable selection is a useful tool in other studies. For instance, EEMD-LASSO-QRNN method, which consists of data preprocessing, feature selection, prediction, and data post-processing, is used to forecast the short-term usage of wind power[42]. With the help of Lasso variable selection, some supervised algorithms can be used to build expert cancer classification models to identify different stages of deadly cancer[43]. MSGP-LASSO is a new Multi-Stage Genetic Programming (MSGP) technique that improves modeling accuracy by coupling the MSGP and multiple LASSO regression together[44].

2.3 Motivation

Although some deep learning methods perform well in mining and learning the linear and nonlinear characteristics of either UTS or MTS, most of them take a single series as a forecasting target and ignore the interactions among multiple series. These interactions may contain the co-movement information which is helpful for forecasting. For instance, gasoline prices in multiple regions can be viewed as multiple time series, which cannot be modeled simultaneously by using previous methods. Thus, the potential relationships among them cannot be revealed for forecasting purposes. Therefore, the goal of this paper is to take advantage of the interactions among time series to design an interpretable deep learning model, named the Lasso-CNN ensemble model. Moreover, we use the new model to make accurate predictions of gasoline prices in multiple regions in the US. This paper has two contributions to its field: one is that we develop a screening and selecting process, where relevant series are used for forecasting a target series based on the complex networks and Lasso methods, which helps the deep learning model to filter out invalid information and avoid overfitting; the other contribution is that we use selected relevant series for forecasting and implement the Lasso-CNN ensemble model on them, which enhances the target series forecasting ability by combining relevant series at the national level and regional levels.

3 Hybrid Approaches Combining Statistical Models with Deep Neural Networks

Although a variety of external factors can affect gasoline prices, the forecasting is usually carried out by time series analysis, which only learns and models the changes to gasoline prices without considering other auxiliary exogenous series. As a result, this forecasting model is inaccurate and unstable. In practice, the national gasoline prices always move along with gasoline prices in some regions. Furthermore, there is a similar pattern of the regional gasoline prices among those regions that have a similar degree of economic development and consumption of goods. It indicates that gasoline prices in some certain regions are informative to the prediction of the gasoline prices of the entire nation and other regions. We propose the Lasso-CNN ensemble model that utilizes the above-mentioned information by implementing the combined model of Lasso and deep neural networks. Because of Lasso’s good performance in variable selection and CNN’s strong ability of capturing the local features of the dynamical system, the new combined model can learn and utilize the interactions of gasoline prices in different regions. Thus, we can get more accurate and robust forecasting results for both national gasoline prices and regional gasoline prices. The framework of the Lasso-CNN ensemble model constructed in this paper is shown in Fig. 1.

3.1 Relevant time series selection

Although including multiple series in the forecasting model may improve the performance of forecasting a target series, involving redundant series causes the overfitting problem. Therefore, in the forecasting model,

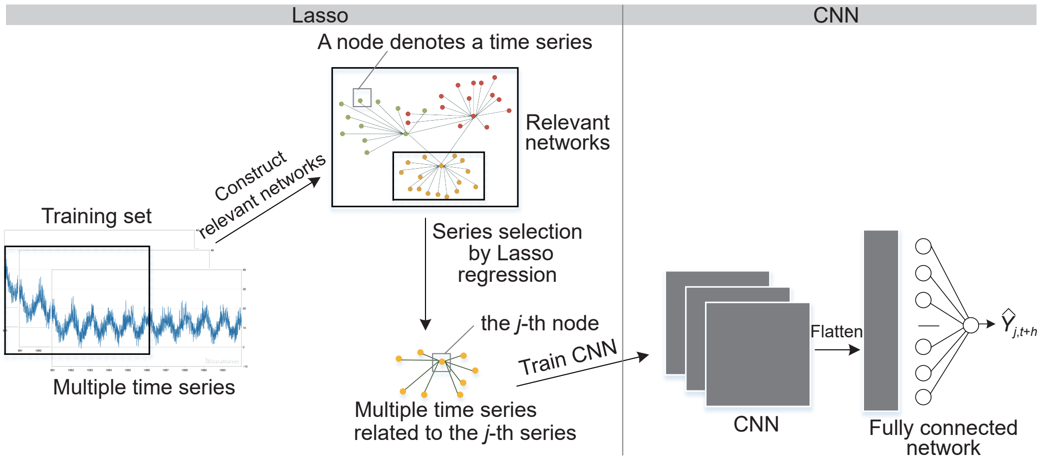


Fig. 1 Framework of the Lasso-CNN model.

we only need to add those relevant series that are highly correlated to the target series. The process of screening and selecting relevant series that should be used to predict the target series includes the flowing steps. Firstly, we calculate the similarity of any two UTSs of the target series to estimate their correlations by using the Euclid distance. For example, given any two UTSs of gasoline prices Y_u and Y_v, their similarity is defined as $d_{uv} = \| Y_u - Y_v \|_2^2$. We prefer the Euclid distance to the correlation coefficient because of the significant differences in the scale of gasoline prices in various regions. Secondly, after the computation of the similarities, the hierarchical clustering algorithm is employed to divide J UTSs into different groups, where the sequence of multiple MTSs $\{Y_{1,t}, Y_{2,t}, \ldots, Y_{J,t}\}$ can be grouped into K sections. Taking the g-th group as an example, it contains g_J UTSs, denoted as $\left\{Y_{g_j,t}\right\}_{g_j=1}^{g_J}$, g_j is the index; $\left\{Y_{g_j,t}\right\}_{g_j=1}^{g_J} \cap \left\{Y_{q_j,t}\right\}_{q_j=1}^{q_J} = \varnothing$ if $g \neq q$ and $g, q = 1, 2, \ldots, K$, and $\sum_{g=1}^{K} g_J = J$.

In essence, the Lasso regression model is the extension of the Ordinary Least Square (OLS) method, which penalizes coefficients in the process of parameter estimation. That is, the absolute value function is used as the penalty term to constrain the regression coefficients. So, the regression coefficient minimizes the residual sum of the squares, where some coefficients have shrunken to zero. By doing so, we complete the process of variable selection. In practical implementations, Lasso regression is usually used to select the most important variables or influencing factors that relate to the response variable, which not only ensures high accuracy of the prediction but also reduces the computational costs and further simplifies the complexity of the model.

Consider the linear regression model,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon \tag{3}$$

where β_0 is an intercept, $\beta_1, \beta_2, \ldots, \beta_p$ are coefficients, and ε is the random error. Let $\left\{Y_{g_j,t}\right\}_{g_j=1}^{g_J}$ be the observed value of the input multi-target time series at t, Y_t be the target value, $t = 1, 2, \ldots, T$. At first, we screen related multi-target time series by using the Lasso regression, which is

$$\widehat{\beta} = \operatorname{argmin}\left\{\sum_{i=1}^{n}\left(Y_t - \beta_0 - \sum_{j=1}^{g_J} Y_{g_j,t}\beta_j\right)^2 + \lambda \sum_{j=1}^{g_J} |\beta_j|\right\} \tag{4}$$

where $\beta_1, \beta_2, \ldots, \beta_{g_J}$ are the effects of $\left\{Y_{g_j,t}\right\}_{g_j=1}^{g_J}$ on Y_t, $\lambda \sum_{j=1}^{g_J} |\beta_j|$ is the Lasso penalty, and λ is a non-negative tuning parameter. Selecting how many time series are determined by the value of the tuning parameter λ. If λ is non-zero, some $\widehat{\beta}$ will be shrunk to zero, where the corresponding series will not be selected or used in the forecasting model. Thus, we implement the selection process in the forecasting model to select relevant series that are related to the target series of gasoline prices in relevant regions.

3.2 Multi-target series forecasting

Taking the subset time series of $\left\{Y_{g_j,t}\right\}_{g_j=1}^{g_J}$ as inputs, CNN is used to implement the process of forecasting. CNN consists of multiple convolutional layers, pooling layers, and fully connected layers in different ways. The convolutional layer can learn and extract the nonlinear characteristics of the input data through the convolutional operation. And then, the convolutional layer outputs the new representation of inputs. Assuming the selected subset time series of $\{Y_{g_j,t}\}_{g_j=1}^{g_J}$ is $\{Y_{g_j,t}\}_{g_j=1}^{m_J}$, and $m_J \leqslant g_J$. Firstly, we use multiple CNN layers and multiple Relu layers to learn the new representation of multiple series, which is

$$\begin{gathered} H_{l+1,1} = f(H_{l,1}w_{l,1} + b_{l,1}), H'_{l+1,1} = \max(0, H_{l+1,1}); \\ H_{l+1,2} = f(H_{l,2}w_{l,2} + b_{l,2}), H'_{l+1,2} = \max(0, H_{l+1,2}); \\ \cdots \\ H_{l+1,m_J} = f\left(H_{l,m_J}w_{l,m_J} + b_{l,m_J}\right), H'_{l+1,m_J} = \\ \max\left(0, H_{l+1,m_J}\right) \end{gathered} \tag{5}$$

where H_{l+1,g_j} is the new representation of input data, $H_{l,g_j} = Y_{g_j,t}$ when $l = 0$, for $j = 1, 2, \ldots, m_J$. w_{l,g_j} and b_{l,g_j} represent the weight and the bias of the convolution kernel, respectively. The advantages of the convolutional layer are weight sharing and local perception. We use Relu as the activation function for CNN here because it can improve the nonlinear fitting ability of the network. In contrast, the Sigmoid activation function in the BP algorithm is likely to cause the problem of gradient disappearance or gradient explosion.

After obtaining the new representation $H'_{l+1,1}$, $H'_{l+1,2}, \ldots, H'_{l+1,m_J}$, the pooling layer, which performs as an under-sampling layer, is used to integrate new representations to forecast the target series. The pooling operation is defined as

$$H_{l+1,target} = pooling\left(H'_{l+1,1}, H'_{l+1,2}, \ldots, H'_{l+1,m_J}\right) \tag{6}$$

We pool the embeddings from the new representation of the target time series and obtain the new features $H_{l+1,target}$. The pooling operation includes max-pooling, mean-pooling, or stochastic pooling. Whereas, we only use max-pooling operation in our analysis.

After the convolution operation and pooling, we take the outputs of the pooling layer as the inputs to train the prediction model. The objective of series prediction is to reconstruct the relationship between the input and the output. A one-layer feedforward neural network is used as the prediction function, which performs like linear regression. For the j-th UTS, let $H_{l+1,target}$ be the representation learned by the former pooling layer, the output of a target series is given by

$$\widehat{Y}_{target,t+h} = \sigma\left(H_{l+1,target}W_o + b_o\right) \tag{7}$$

where W_o and b_o are trainable parameters or the coefficients of linear regression; $\sigma(\cdot)$ is an identity function; and $\widehat{Y}_{target,t+h}$ is the predicted value of the series.

3.3 Loss function

Let $Y_{target,t+h}$ be a set of test series, and $\widehat{Y}_{target,t+h}$ be the set of the corresponding predicted series, whose loss function is calculated by the Mean Square Error (MSE) as follows:

$$\text{MSE} = \frac{1}{H}\sum_{h=1}^{H}\left(Y_{target,t+h} - \widehat{Y}_{target,t+h}\right)^2 \tag{8}$$

where H is the length of predicted values of the target series.

3.4 Implementation

In practice, we implemented the new method in Keras. Following the processes in the previous section, we have used Lasso to filter out regions, to which gasoline prices are related to. And then, the one-dimensional convolution is used to process the gasoline prices time series, where the size of the convolution kernel is 1. And, a pooling layer is added after the convolution layer, thereby reducing the parameters of the learned representation. A one-layer feedforward neural network is used as the prediction model or the output layer, which only has one hidden neuron. When we train CNN, we utilize the Adam algorithm[45] to estimate the unknown weight matrix and bias vectors so as to reduce the operation time and simplify the parameter adjustment, thereby improving the efficiency of operation. The training epochs are set to 50 and the batch size is set to 32. Based on these settings, the proposed method is described as the following processes.

4 Experiments and Results

4.1 Dataset

The US gasoline prices data are from the weekly data published by the Energy Information Administration. The dataset is a multiple time series consisting of 28 univariate time series, including 10 cities, 9 states, 8 territories, and 1 nation. There are 968 pieces of gasoline prices data (dollar per gallon), spanning 17 years from the last week of May 2003 to the first week of December 2021. 70% of the data are used as training samples, and 30% of the data are used as test samples.

4.2 Experiment settings

The experiments are conducted based on the analysis of the gasoline prices in the US using the Keras platform[46]. Before feeding the input data into the new neural networks, the multiple MTSs are normalized first. And then, we embed CNN and GRU into our framework for forecasting a target series with relevant series selected by Lasso. In comparison, we only use CNN and GRU to forecast the future values without considering any other series. We use two kinds of measures to evaluate the performance of the model. One is the Root Mean Square Error (RMSE), which weighs the average squared difference between the estimated value and the actual value; the other is the Mean Absolute Percentage Error (MAPE), which is one of the commonly used metrics that measure the forecasting accuracy based on 100 repeated random experiments. Both of them are defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \widehat{y_i})^2},$$

and

$$\text{MAPE} = \frac{100\%}{N}\sum_{i=1}^{N}\left|\frac{y_i - \widehat{y_i}}{y_i}\right|.$$

4.3 Result

4.3.1 Relevant regions selection

Before implementing the model for forecasting gasoline prices of each region and the entire nation, it is necessary to screen out relevant time series of those regions related to the target series. The Lasso regression is used to achieve the goal. Given different λ values, the model will find a set of optimal regression coefficients to minimize the loss. We tune the hyperparameter λ by cross-validation, choose the optimal λ to make cross-validation (cv) smallest, fix the value of λ, and run Lasso regression to obtain those relevant time series of gasoline prices, whose coefficients are non-zero. The relevant regions related to the target time series screened by the Lasso regression are shown in Table 1.

4.3.2 National level gasoline prices forecasting

In order to evaluate the performance of different forecasting models, we first compare the proposed Lasso-CNN model with different information fusion models, such as max-pooling and concatenation operations, which are different ways to integrate representations, denoted as Lasso-CNN (Max) and Lasso-CNN (Conc), respectively. In addition, for the Lasso-CNN model, we can replace CNN with GRU to obtain the new models, Lasso-GRU (Max) and Lasso-GRU (Conc), to learn the new features of input time series. Besides, we compare the proposed model with GRU and CNN, which only take UTS of target series as input while forecasting. And, we compare the proposed model with CNN (all), which takes all the series for forecasting the target series without screening them. All of these models are used to predict gasoline prices at the national level and regional levels in the US. Table 2 shows that the Lasso-CNN ensemble model outperforms all the alternative methods in terms of both RMSE and MAPE. According to Fig. 2, the US gasoline prices forecasting based on the Lasso-CNN ensemble model is superior to other methods. It illustrates that gasoline prices at the regional level could amplify the signal and help to predict gasoline prices at the national level. And, we found that the effect of using the auxiliary information after the encoding is not as good as the effect of direct use of the auxiliary information, which shows that the coding process would cause the loss of some information so that the enhancement effect brought by the auxiliary information is weakened.

4.3.3 Regional levels gasoline prices forecasting

By the same token, at the regional levels, we compare the performance of the new method with other alternative methods. For each region, we predict the gasoline prices by using other regions' information based on the new method, and we report the average predictive RMSE and MAPE. The experimental results are presented in Table 3 and Fig. 3.

Table 3 indicates that the RMSE of the Lasso-CNN ensemble model proposed in this paper is better than other alternative methods for predicting regional gasoline prices except for New England, Rocky Mountain, Colorado, Chicago, San Francisco, and Seattle. Moreover, similar to the predictions at the national level, the auxiliary information filtered out by Lasso enhances the prediction ability, but the use of the coded information results in the loss of effective information. Intuitively, Fig. 3 shows that the proposed method outperforms other alternative models in terms of MAPE in most cases. It also illustrates that there is a

Table 1 Linkage of regional gasoline prices.

Region	Relevant region
US	East Coast, Central Atlantic, Lower Atlantic, Midwest, Gulf Coast, Rocky Mountain, West Coast
East Coast	New England, Central Atlantic, Lower Atlantic
New England	East Coast, Central Atlantic, Lower Atlantic, Rocky Mountain
Central Atlantic	East Coast, New England, West Coast
Lower Atlantic	East Coast, Gulf Coast
Midwest	Gulf Coast, Rocky Mountain, West Coast
Gulf Coast	New England, Central Atlantic, Lower Atlantic, Midwest, Rocky Mountain, West Coast
Rocky Mountain	Midwest, Gulf Coast, West Coast
West Coast	New England, Central Atlantic, Midwest, Gulf Coast, Rocky Mountain
California	Florida, Ohio, Texas, Washington
Colorado	Minnesota, Texas, Washington
Florida	California, New York, Texas
Massachusetts	California, New York, Washington
Minnesota	Colorado, Ohio, Texas, Washington
New York	Florida, Massachusetts
Ohio	Minnesota, Texas
Texas	California, Colorado, Florida, Massachusetts, Minnesota, Ohio
Washington	California, Colorado, Massachusetts
Boston	Cleveland, New York City, Seattle
Chicago	Cleveland, Denver, Houston, Los Angeles
Cleveland	Boston, Chicago, Houston
Denver	Chicago, Houston, Seattle
Houston	Chicago, Cleveland, Denver, Miami, New York City, San Francisco
Los Angeles	Miami, San Francisco
Miami	Houston, New York City
New York City	Boston, Houston, Miami
San Francisco	Denver, Los Angeles, Miami, Seattle
Seattle	Boston, Denver, San Francisco

Table 2 MAPE and RMSE of forecasting gasoline prices of the entire nation based on both the new method and alternative methods.

Method	MAPE (%)	RMSE
Lasso-GRU (Max)	2.3887	0.0824
Lasso-GRU (Conc)	8.0075	0.2297
Lasso-CNN (Max)	1.9283	0.0653
Lasso-CNN (Conc)	2.3838	0.0807
CNN(all)	3.4134	0.1134
uni-GRU	5.6793	0.1817
uni-CNN	4.6230	0.2234

certain linkage among gasoline prices in different regions and that the use of information from other relevant regions can enhance the prediction ability of gasoline prices forecasting in specific regions. The new method has a better performance of forecasting than alternative methods do.

Therefore, the experimental results show that the univariate CNN can only process independent gasoline price information in a certain region for prediction purposes. In contrast, the Lasso-CNN method proposed in this paper takes into account the information of other related regions for modeling, which potentially enhances the performance of forecasting gasoline prices in the target region. And, the use of relevant information from other regions after encoding is not as effective as the direct use of such relevant information, which may be caused by the loss of some valid information in the process of advanced coding. We can reach the same conclusion on the GRU model. Meanwhile, comparing all the experimental results of adding information of all regions at the same level, we found that the Lasso-CNN performs better for gasoline prices prediction in terms of RMSE, which shows that adding information about relevant regions can enhance

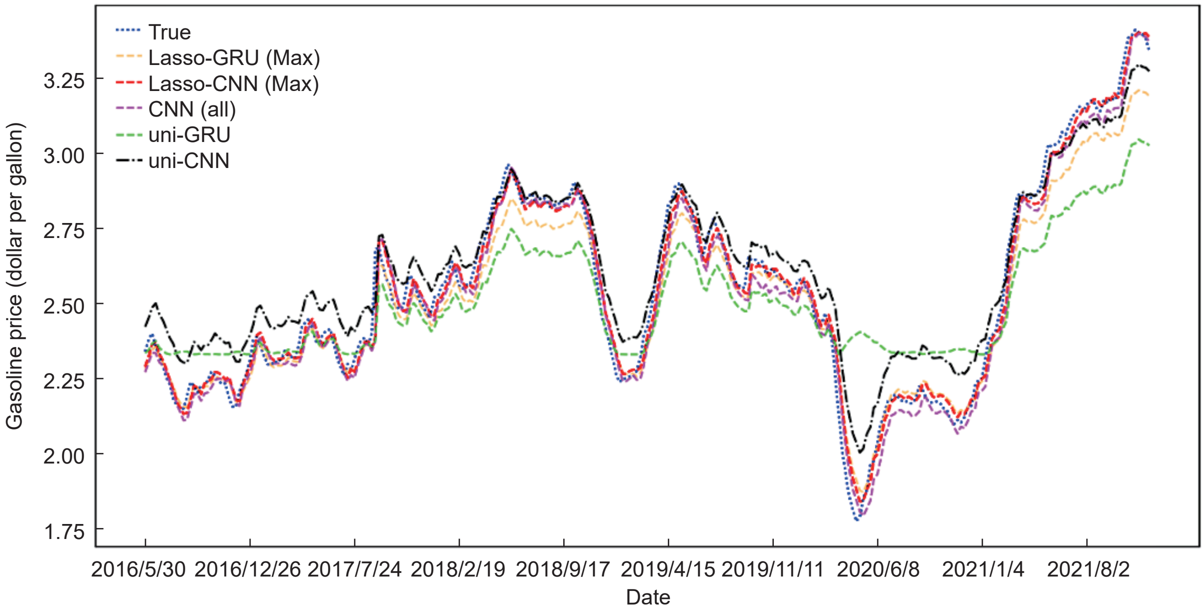


Fig. 2 Gasoline prices of the entire nation and their prediction based on the new method and alternative methods.

Table 3 RMSE of forecasting gasoline prices in various regions based on the new method and alternative methods.

Region	RMSE						
	Lasso-GRU (Max)	Lasso-GRU (Conc)	Lasso-CNN (Max)	Lasso-CNN (Conc)	CNN (all)	uni-GRU	uni-CNN
East Coast	0.0920	0.3456	0.0534	0.0624	0.2110	0.1863	0.2327
New England	0.0794	0.3278	0.0874	0.0801	0.2166	0.1937	0.2401
Central Atlantic	0.1042	0.3581	0.0645	0.0828	0.2679	0.1705	0.2893
Lower Atlantic	0.1075	0.3958	0.0557	0.0749	0.0973	0.2077	0.2071
Midwest	0.1397	0.3942	0.1112	0.1345	0.2401	0.2171	0.2476
Gulf Coast	0.1128	0.2862	0.1114	0.1512	0.1401	0.236	0.2625
Rocky Mountain	0.1686	0.3745	0.1177	0.1167	0.1474	0.1986	0.2219
West Coast	0.2575	0.3275	0.1294	0.1828	0.2224	0.2502	0.1888
California	0.3309	0.4200	0.1790	0.2000	0.3535	0.2844	0.2969
Colorado	0.1761	0.3853	0.1191	0.1071	0.2733	0.2138	0.2348
Florida	0.1319	0.4046	0.0872	0.1053	0.1082	0.2036	0.2502
Massachusetts	0.1328	0.3871	0.1306	0.1601	0.2708	0.1846	0.2249
Minnesota	0.1129	0.3688	0.1032	0.1260	0.1243	0.2187	0.2349
New York	0.0986	0.4136	0.0559	0.0699	0.3029	0.1984	0.2615
Ohio	0.1496	0.4560	0.1117	0.1327	0.1747	0.2384	0.2653
Texas	0.1281	0.3053	0.1065	0.1448	0.1276	0.2426	0.2676
Washington	0.1793	0.3890	0.1247	0.1384	0.2026	0.2203	0.2154
Boston	0.1025	0.3628	0.0789	0.1047	0.1352	0.1886	0.3123
Chicago	0.1273	0.3812	0.1280	0.1335	0.1909	0.2050	0.2533
Cleveland	0.1368	0.3955	0.1194	0.1319	0.1485	0.2206	0.2567
Denver	0.1769	0.3717	0.1419	0.1441	0.2482	0.2165	0.2381
Houston	0.1241	0.3106	0.1118	0.1430	0.2222	0.2306	0.2633
Los Angeles	0.2095	0.4945	0.1272	0.1846	0.3290	0.2783	0.2857
Miami	0.1359	0.4022	0.1098	0.1454	0.1839	0.2297	0.2549
New York City	0.1148	0.3429	0.0871	0.1007	0.1818	0.1882	0.2284
San Francisco	0.3148	0.4218	0.2857	0.1985	0.3855	0.3042	0.3099
Seattle	0.2282	0.3944	0.1527	0.0814	0.2954	0.2573	0.1697
Total	4.2551	10.4467	3.1564	3.5184	5.9147	6.1656	6.9372

Note: Table 3 reports the average predictive performance.

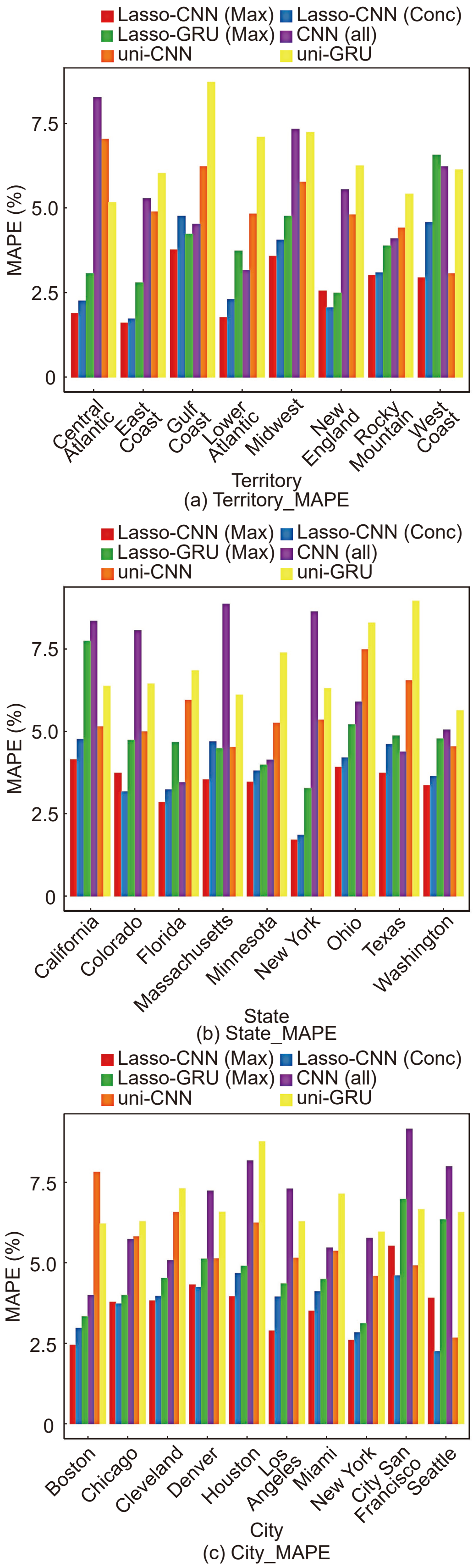


Fig. 3 Forecasting performance comparison of different methods.

the prediction ability while adding information about irrelevant regions can reduce the accuracy of prediction. Using Lasso to select effective information for modeling, the influence of useless noise on prediction is avoided, and the model has better performance.

5 Conclusion

The Lasso-CNN ensemble model proposed in this paper is a new gasoline prices prediction method, where the Lasso method is used to select regions that gasoline prices are related to, and the information of selected regions is used as its input, and then predict gasoline prices through CNN. The results show that the use of the Lasso method can effectively screen out the relevant regions, which enhances the prediction ability, the new approach outperforms alternative methods for predicting gasoline prices at the national level and regional levels. It also shows that the gasoline prices network contains information about the relationships and interactions among gasoline trade entities, and there is a certain linkage among gasoline prices in different regions. Since other similar time series may have valuable information that might improve the prediction accuracy, predicting a time series by taking into account its linkage with other time series is a better strategy. The new method could also be applied to other areas, such as global crude oil prices, carbon emissions, climate change forecasting, and more.

In the future, on one hand, factors affecting the price of gasoline, such as gasoline production, economic index, and crude oil price, should be considered to improve the prediction accuracy. On the other hand, the hierarchical structure can be integrated into the model, and by optimizing the global objective function, we can simultaneously predict the gasoline price time series in each region and achieve the optimal overall prediction.

Acknowledgment

We are grateful to the editors and reviewers for many constructive and insightful comments. This work was supported by the National Natural Science Foundation for Distinguished Young Scholars of China (No. 71701223), the National Statistical Science Foundation of China (No. 2018LZ08), the Central University of Finance and Economics Young Talents Training Support Project (No. QYP2014), Fundamental Research Funds for the Central Universities (China):

the Central University of Finance and Economics Scientific Research and Innovation Team Support Project, the Strategic Economy Interdisciplinarity (Beijing Universities Advanced Disciplines Initiative (No. GJJ2019163)), and the Emerging Interdisciplinary Project of CUFE (No. 020659919002).

References

[1] J. D. Hamilton, Causes and consequences of the oil shock of 2007-08, NBER Working Paper Series, National Bureau of Economic Research, Cambridge, MA, USA, https://www.nber.org/papers/w15002, 2009.

[2] R. S. Molloy and H. Shan, The effect of gasoline prices on household location, *Review of Economics and Statistics*, vol. 95, no. 4, pp. 1212–1221, 2013.

[3] M. R. Busse, C. R. Knittel, and F. Zettelmeyer, Pain at the pump: The differential effect of gasoline prices on new and used automobile markets, NBER Working Paper Series, National Bureau of Economic Research, https://econpapers.repec.org/paper/nbrnberwo/15590.htm, 2009.

[4] H. Allcott and N. Wozny, Gasoline prices, fuel economy, and the energy paradox, *Review of Economics and Statistics*, vol. 96, no. 5, pp. 779–795, 2014.

[5] C. R. Knittel and R. Sandler, Carbon prices and automobile greenhouse gas emissions: The extensive and intensive margins, NBER Working Paper 16482, National Bureau of Economic Research, https://ideas.repec.org/p/nbr/nberwo/16482.html, 2010.

[6] M. S. Lewis and H. P. Marvel, When do consumers search, *Journal of Industrial Economics*, vol. 59, no. 3, pp. 457–483, 2011.

[7] S. Borenstein and A. Shepard, Sticky prices, inventories, and market power in wholesale gasoline markets, NBER Working Paper 5468, National Bureau of Economic Research, https://ideas.repec.org/p/nbr/nberwo/5468.html, 1996.

[8] M. R. Busse, C. R. Knittel, and F. Zettelmeyer, Are consumers myopic? Evidence from new and used car purchases, *American Economic Review*, vol. 103, no. 1, pp. 220–256, 2013.

[9] M. R. Busse, C. R. Knittel, J. Silva-Risso, and F. Zettelmeyer, Who is exposed to gas prices? How gasoline prices affect automobile manufacturers and dealerships, *Quantitative Marketing and Economics*, vol. 14, pp. 41–95, 2016.

[10] C. W. J. Granger, Investigating causal relations by econometric models and cross-spectral methods, *Econometrica*, vol. 37, no. 3, pp. 424–438, 1969.

[11] A. F. Alhajji and D. A. Huettner, OPEC and world crude oil markets from 1973 to 1994: Cartel, oligopoly, or competitive? *The Energy Journal*, vol. 21, no. 3, pp. 31–60, 2000.

[12] V. Brémond, E. Hache, and V. Mignon, Does OPEC still exist as a cartel? An empirical investigation, *Energy Economics*, vol. 34, no. 1, pp. 125–131, 2012.

[13] B. Abramson and A. J. Finizza, Using belief networks to forecast oil prices, *International Journal of Forecasting*, vol. 7, no. 3, pp. 299–315, 1991.

[14] L. Tang and S. Hammoudeh, An empirical exploration of the world oil price under the target zone model, *Energy Economics*, vol. 24, no. 6, pp. 577–596, 2002.

[15] C. Feng, J. Wu, and F. Jiang, Portfolio forecasting studies of oil prices, (in Chinese), *Journal of the University of Petroleum*, no. 1, pp. 12–14, 2004.

[16] L. Fan, J. Dai, and J. Yin, Oil price forecasting based on an improved wavelet neural network, *Statistics & Decision*, doi: 10.13546/j.cnki.tjyjc.2017.12.020.

[17] J. G. D. Gooijer and R. J. Hyndman, 25 years of IIF time series forecasting: A selective review, Monash Econometrics and Business Statistics Working Papers 12/05, Monash University, Department of Econometrics and Business Statistics, https://ideas.repec.org/p/msh/ebswps/2005-12.html, 2005.

[18] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*. New York, NY, USA: Springer, 1991.

[19] R. L. Martin and J. E. Oeppen, The identification of regional forecasting models using space: Time correlation functions, *Transactions of the Institute of British Geographers*, no. 66, pp. 95–118, 1975.

[20] H. Tong, *Threshold Models in Nonlinear Time Series Analysis*. New York, NY, USA: Springer, 1983.

[21] P. S. Game, V. Vaze, and M. Emmanuel, Induction of decision trees based on gray wolf optimizer for heart disease classification, *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, pp. 168–173, 2019.

[22] E. E. Osuna, *Support Vector Machines: Training and Applications*. Cambridge, MA, USA: Massachusetts Institute of Technology, 1997.

[23] L. E. Baum and T. Petrie, Statistical inference for probabilistic functions of finite state markov chains, *Annals of Mathematical Statistics*, vol. 37, no. 6, pp. 1554–1563, 1966.

[24] J. Devlin, M. -W. Chang, K. Lee, and K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv: 1810.04805, 2019.

[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ImageNet classification with deep convolutional neural networks, *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.

[26] D. Silver, A. Huang, C. Maddison, A. Guez, L. Sifre, G. V. D. Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, et al., Mastering the game of go with deep neural networks and tree search, *Nature*, vol. 529, no. 7587, pp. 484–489, 2016.

[27] G. Zhang, B. E. Patuwo, and M. Y. Hu, Forecasting with artificial neural networks: The state of the art,

International Journal of Forecasting, vol. 14, no. 1, pp. 35–62, 1998.

[28] Y. Bengio, A. Courville, and P. Vincent, Representation learning: A review and new perspectives, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[29] B. Lim and S. Zohren, Time-series forecasting with deep learning: A survey, *Philosophical Transactions of the Royal Society A*, vol. 379, no. 2194, p. 20200209, 2021.

[30] M. Elleuch, N. Tagougui, and M. Kherallah, Optimization of DBN using regularization methods applied for recognizing Arabic handwritten script, *Procedia Computer Science*, vol. 108, pp. 2292–2297, 2017.

[31] G. E. Hinton and R. R. Salakhutdinov, Reducing the dimensionality of data with neural networks, *Science*, vol. 313, pp. 504–507, 2006.

[32] C. Rodriguez-Antona and M. Ingelman-Sundberg, Cytochrome P450 pharmacogenetics and cancer, *Oncogene*, vol. 25, no. 11, pp. 1679–1691, 2006.

[33] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, Learning representations by back-propagating errors, *Nature*, vol. 323, pp. 533–536, 1986.

[34] P. J. Werbos, Backpropagation through time: What it does and how to do it, *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1550–1560, 1990.

[35] S. Hochreiter and J. Schmidhuber, Long short-term memory, *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[36] F. A. Gers, J. Schmidhuber, and F. Cummins, Learning to forget: Continual prediction with LSTM, *Neural Computation*, vol. 12, no. 10, pp. 2451–2471, 2000.

[37] R. Dey and F. M. Salem, Gate-variants of gated recurrent unit (GRU) neural networks, in *Proc. 2017 IEEE 60th International Midwest Symposium on Circuits and Systems* (*MWSCAS*), Boston, MA, USA, 2017, pp. 1597–1600.

[38] R. Tibshirani, Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society. Series B* (*Methodological*), vol. 58, no. 1, pp. 267–288, 1996.

[39] Y. Zhang, F. Ma, and Y. Wang, Forecasting crude oil prices with a large set of predictors: Can LASSO select powerful predictors, *Journal of Empirical Finance*, vol. 54, pp. 97–117, 2019.

[40] P. Ghosh, A. Karim, S. T. Atik, S. Afrin, and M. Saifuzzaman, Expert cancer model using supervised algorithms with a LASSO selection approach, *International Journal of Electrical and Computer Engineering*, vol. 11, no. 3, pp. 2631–2639, 2021.

[41] X. Xiang, X. Ma, Z. Ma, and M. Ma, Operational carbon change in commercial buildings under the carbon neutral goal: A LASSO–WOA approach, *Buildings*, vol. 12, no. 1, p. 54, 2022.

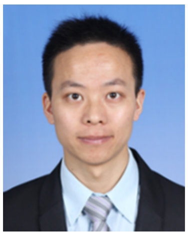
[42] Y. He and Y. Wang, Short-term wind power prediction based on EEMD-LASSO-QRNN model, *Appl. Soft Comput.*, vol. 105, p. 107288, 2021.

[43] P. Ghosh, S. Azam, M. Jonkman, A. Karim, F. M. J. M. Shamrat, E. Ignatious, S. Shultana, A. R. Beeravolu, and F. D. Boer, Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques, *IEEE Access*, vol. 9, pp. 19304–19326, 2021.

[44] A. D. Mehr and A. H. Gandomi, MSGP-LASSO: An improved multi-stage genetic programming model for streamflow prediction, *Inf. Sci.*, vol. 561, pp. 181–195, 2021.

[45] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv: 1412.6980, 2015.

[46] N. Ketkar, Introduction to keras, in *Deep learning with Python*. Berkeley, CA, USA: Apress, 2017, pp. 97–111.



Hu Yang received the BEng degree from Dalian Maritime University, Dalian, China in 2005, the MEng degree from National University of Defense Technology, Changsha, China in 2007, and the PhD degree from Renmin University of China, Beijing in 2014. He has been a visiting scholar at Aarhus University in Denmark, University of Minnesota in the United States, and Hokkaido University in Japan. He is currently an associate professor at the School of Information, Central University of Finance and Economics. His research projects include big data analysis, social computing, social networks analysis, and biostatistics. He has published some academic papers in well-known journals, such as *Statistics in Medicine*, *Information Processing & Management*, and so on. He serves as an associate editor of the *Journal of Social Computing* and reviewer of many journals and international conferences.

Xinlu Tian received the BS degree in computer science and technology from Jiangxi University of Finance and Economics, Nanchang, China in 2021 and she is currently pursuing the MS degree in electronic information at the School of Information, Central University of Finance and Economics, Beijing, China.

Xin Jin received the PhD degree from Donghua University, Shanghai, China in 2004. He is now a full professor in Central University of Finance and Economics, Beijing, China. He has published hundreds of academic papers, eight books, and three patents. He is a senior member of China Computer Society. His current research mainly focuses on fintech and business intelligence.

Haijun Wang is an associate professor and researcher (concurrently) at the School of Economics, Beijing Wuzi University, and a postdoctoral researcher in economics. He has been engaged in financial investment, risk management, and strategic research in large financial institutions and listed companies for more than ten years. His research interests include financial risk, debt risk, foreign investment risk, financial technology, etc. He has successively worked in the following fields: financial research, financial and trade research, reform, international trade, industrial economy research, economic and management research, world economic and political forum, world economic research, international financial research, economic and social system comparison, foreign economy and management, Nankai economic research, economic journal, economic management, and operations research and management. He has published more than 100 papers in important journals at home and abroad, such as *Journal of Wuhan University* (Zheshe Edition), *Frontiers of Economics in China*, *Journal of Social Computing*, *International Review of Economics & Finance*, *Frontiers in Psychology*, etc.