A hybrid deep learning approach by integrating extreme gradient boosting-long short-term memory with generalized autoregressive conditional heteroscedasticity family models for natural gas load volatility prediction

Huibin Zeng1 | Bilin Shao1 | Genqing Bian2 | Hongbin Dai1 | Fangyu Zhou3

1School of Management, Xi’an University of Architecture and Technology, Xi’an, China
2School of Information and Control Engineering, Xi’an University of Architecture and Technology, Xi’an, China
3School of Applied English, Chengdu Institute Sichuan International Studies University, Chengdu, China

Correspondence
Bilin Shao, School of Management, Xi’an University of Architecture and Technology, Xi’an 710055, China.
Email: sblin0462@sina.com

Funding information
National Natural Science Foundation of China, Grant/Award Number: 62072363; Yulin City Science and Technology Plan Project, Grant/Award Number: CXY-2020-063

Abstract
Natural gas load forecasting provides decision-making support for natural gas dispatch and management, pipeline network construction, pricing, and sustainable energy development. To explain the uncertainty and volatility in natural gas load forecasting, this study predicts the natural gas load volatility. As the natural gas load volatility has the time-series features, along with long-term memory, volatility aggregation, asymmetry, and nonnormality, this study proposes a natural gas load volatility prediction model by combining generalized autoregressive conditional heteroscedasticity (GARCH) family models, XGBoost algorithm, and long short-term memory (LSTM) network. The model first takes the GARCH family models parameters of sliding estimation and meteorological factors as the influencing factors of volatility, and then it screens these influencing factors through the extreme gradient boosting (XGBoost) algorithm. Finally, the selected important features are input into the LSTM network to predict the volatility, and the 90% confidence interval of the volatility is calculated. Compared with a variety of single and combined models, the model proposed in this study has an average reduction of 45.404% in the evaluation index of mean squared error. The experimental results show that the model proposed in this study has a good performance and accuracy in predicting the volatility of natural gas load.

Key words
GARCH family models, LSTM, natural gas load, volatility prediction, XGBoost
INTRODUCTION

With the implementation of the “carbon emission peak” (carbon emission peak refers to an inflection point of greenhouse gas emissions) and “carbon neutrality” (carbon neutrality refers to achieving net-zero carbon dioxide emissions) policy, natural gas, as a clean energy source, has become the spotlight in the energy field. In recent years, the global natural gas industry has developed rapidly in the background of the energy transition process. Forecasts show that global demand for natural gas is expected to continue to grow over the next 30 years and it is expected to peak in the mid-2030s and mid-2040s. Due to the COVID-19 pandemic in 2020, global energy development has been continuously and negatively affected, especially coal and oil, but the demand for natural gas is still increasing rapidly, and it is expected to reach 4.6 x 1012 cubic meters by 2030. The demand mainly comes from China, India, the Middle East, and Southeast Asian countries.

A reasonable estimate of the natural gas load can enable the natural gas industry to develop rapidly and orderly. In the long run (in the upcoming few years), natural gas producers and operators hope to obtain sufficient benefits from the production and sale of it; the natural gas management department hopes to formulate reasonable policies to help certain regions solve energy problems. In a relatively short period of time (in a few hours or a few days), natural gas operators hope to control the peak value on the basis of the predicted load value, so as to ensure the production and life of the people; The energy management department expects to obtain the predicted load data in advance, so as to reasonably dispatch related energy (such as electricity and coal) to maintain the balance of energy supply and demand in the region. It is clear that different functional departments have different requirements for natural gas load forecasting, and the orderly operation of society is inseparable from reliable natural gas load forecasting. In previous studies, it is a more traditional method to predict the load of natural gas at a certain time based on the historical load of natural gas. This is called point prediction, that is, predicting a certain single value at a certain point in time. However, in actual production and life, the load forecast of natural gas cannot maintain the ideal model, and the actual load of natural gas may deviate greatly from the predicted value. This difference may make it difficult for natural gas producers and operators to make optimal decisions. In addition, the predicted value of natural gas load cannot fully reflect the change of load, so it is difficult to reasonably measure the supply and demand of natural gas. Given the uncertainty of natural gas load changes, interval prediction of the natural gas load is more scientific and reasonable, and more flexible than point prediction. In natural gas load forecasting, volatility is a key influencing factor of uncertainty in natural gas load forecasting. Volatility complements point prediction and makes interval prediction feasible by determining the range of variation in load forecasting.

The natural gas load volatility effectively takes into account the concentration and dispersion of natural gas load, the specific value of natural gas load and the probability of occurrence of various load values, and the maximum and minimum natural gas load. By predicting the natural gas load volatility, the standard deviation of the natural gas load can be reasonably estimated, and the maximum and minimum load values in a certain period of time can be further calculated by means of probability statistics. This kind of interval prediction results can better provide data-based decision support for natural gas producers and operators. Therefore, this study focuses on the prediction of natural gas load volatility. Based on the following two points: (1) The natural gas load volatility is difficult to predict and the prediction accuracy of the volatility is low; (2) The factors affecting the natural gas load fluctuation are nonlinear and time-varying. This study proposes a new hybrid model based on parameters of the GARCH family and meteorological features (GFMF)-extreme gradient boosting (XGBoost)-long short-term memory (LSTM) for natural gas load volatility prediction. The model first sliding estimates the parameters of generalized autoregressive conditional heteroscedasticity (GARCH), threshold autoregressive conditional heteroskedasticity (TARCH), exponential generalized autoregressive conditional heteroscedasticity (EGARCH), and power autoregressive conditional heteroscedasticity (PARCH) models in the GARCH family of models, second the GARCH family of model parameters as well as meteorological factors are filtered by XGBoost for features, and the features with larger importance values are selected and input into the LSTM, thus predicting the natural gas load volatility. And it is further compared with other single and combined models such as autoregressive integrated moving average model (ARIMA), multiple linear regression (MLR), support vector regression (SVR), multi-layer perceptron (MLP), convolutional neural network (CNN), and so on. In this model, the parameters of the GARCH family models can effectively explain the volatility of the load, XGBoost can select features reasonably to improve the model’s prediction accuracy, and the LSTM network has strong learning and adaptive capabilities. The experimental results show that the model proposed in this study has a good predictive ability, and the predictive accuracy of the model is significantly improved.
The contributions of this study to existing research are as follows: (1) The calculation method of natural gas load volatility is more explicit, which distinguishes itself from the previous concept of volatility in the field of financial assets. In addition, a specific statistical description and analysis of natural gas load volatility are carried out, which is different from the previous volatility analysis of prices and stocks. (2) A hybrid deep learning model for natural gas load volatility forecasting is established based on the GARCH family and XGBoost-LSTM, which can effectively capture the nonlinearity and diversity of natural gas load volatility and its influencing factors. (3) In the previous literature, there is still a lack of research on natural gas load volatility prediction. To provide certain empirical evidence, we conducted an in-depth evaluation of different GARCH family models, a combination of models of GARCH family, and deep learning algorithms. The results show that the appropriate GARCH family parameters can effectively improve prediction accuracy. The GFMF-XGBoost-LSTM model performs best.

The organizational structure of this study is as follows. The second section briefly summarizes the current research status of natural gas load volatility. The third section analyzes the data used in this study and explains the construction of the model. The fourth section mainly describes the experiment from the division and normalization of the data set, the experimental results, and analysis. The fifth section summarizes the research results and limitations of this study.

2 | RESEARCH STATUS

Volatility forecasts are widely used in the financial asset industry. To model and predict volatility, Engle proposed an autoregressive conditional heteroscedasticity model (ARCH) to describe the clustering and persistence of financial market volatility. But it is difficult to reflect the characteristics of long-term memory and nonlinearity of actual data. In view of the shortcomings of the ARCH model, Bollerslev proposed a generalized autoregressive conditional heteroscedasticity model (GARCH). Although the GARCH model effectively improves the accuracy of volatility prediction, it still has certain limitations. For example, the GARCH model cannot describe the asymmetry of sequence fluctuations. For this reason, many scholars have conducted extensive research on ARCH and GARCH models, and gradually formed GARCH family models. Related research shows that the GARCH family models have the advantages of strong explanatory and good forecasting effect in the prediction of financial asset volatility. Cifter A uses Markov transform generalized autoregressive conditional heteroscedasticity model (MS-GARCH) to predict electricity prices. The research results show that the MS-GARCH model has a more accurate forecasting effect than the standard GARCH model. Cai and others established the HSMA-realized HAR GARCH model (Heteroskedastic single measure adaptive Realized HAR GARCH model), which effectively improves the model’s goodness of fit, volatility and VaR prediction accuracy compared to the Realized HAR GARCH model. Some of the above research results are mainly researched and improved on the basis of GARCH family models. However, with the development of economic and social informatization, people find that traditional financial time series methods are difficult to deal with large amounts of data and multiple features. As a result, intelligent prediction methods combined with machine learning or deep learning have developed rapidly in the field of volatility prediction. The network structure of deep learning is a deep neural network. These neural networks can tolerate erroneous data and find out nonlinear relationships among data. Compared with the econometric model, it does not have too many parameter predictions, so it can quickly and accurately learn the law from the data. Kim et al. used a new hybrid LSTM model to predict stock price fluctuations. This model combines the LSTM model with various generalized autoregressive heteroscedasticity (GARCH) models to effectively improve the prediction performance. Hu et al. combined the GARCH model with LSTM and ANN to propose a new hybrid method for predicting copper price fluctuations. Experimental results show that this method significantly improves the prediction performance of the model. Su used eight-parameter volatility forecasting models and eight combined volatility forecasting models to discuss whether neural network methods, leverage effects, and non-normal return distribution settings can improve the volatility forecasting effect. Vidal et al. added LSTM to CNN to predict the volatility of gold. The research results show that compared with GARCH and LSTM models, the hybrid model has a greater improvement. In the field of electricity forecasting, load forecasting has also been carried out by researchers with the help of relevant models in the GARCH family. Ghasemi-Marzbal A proposes a method for forecasting short-term electricity prices and loads that combines wavelet transform, least square support vector machines (LSSVM), generalized autoregressive conditional heteroscedasticity (GARCH), and the virus colony search algorithm (VCS). Zolfaghari et al. proposed a hybrid forecasting method combining wavelet neural network (WN), ARIMA-GARCH family models for electricity forecasting.
Bikcora et al. proposed a model for forecasting the probability distribution of daily electricity demand based on the ARMA-GARCH, CAViaR, and CARE models. These models and methods provide some reference for the prediction of natural gas load volatility.

The fluctuation of the natural gas load is similar to that of stocks and prices, but there are also inconsistencies. On the one hand, the change of natural gas load is also closely related to weather and many other factors. The forecast of load volatility is easily affected by many nonlinear factors. Therefore, the standard GARCH family models are difficult to reasonably predict the volatility of natural gas load. On the other hand, too many influencing factors will affect the effect of prediction, and appropriate influencing factors can effectively improve the prediction accuracy of the model. Kang et al. combined dynamic modeling (EMD), complex network (CN), and XGBoost, and improved the accuracy of urban traffic travel time prediction through feature screening. Zhang Y et al. used the XGBoost model to screen out key features to predict the concentration of PM2.5 pollutants in Beijing in the next 24 h. In addition, natural gas loads are susceptible to the influence of pipeline materials, and different materials may cause certain fluctuations in natural gas loads when external influences vary. Bajcar et al. studied the effects of traffic-induced vibrations on natural gas pipelines. Dynamic stresses from traffic vibrations affect the loading of natural gas pipeline materials and may have some impact on natural gas loads and pose some safety threat to the surrounding environment. Liang et al. investigated the mechanical behavior of common polyethylene natural gas pipelines when ground is overload. Their study shows that the point of maximum stress shifts from the top to the middle of the pipeline as the ground load increases. It can be easily found that the stresses to which the transport pipeline is subjected can have an effect on the natural gas load and influence the fluctuation of the load. At present, there are few studies related to natural gas load volatility prediction, but some similar studies have also been done, such as volatility analysis and prediction in the wind power industry. For example, Shen et al. examined the volatility prediction ability of different GARCH family wind power models. The results show that the use of the Markov state switching GARCH model to predict wind power volatility is significantly better than the traditional GARCH model. Agrawal and others first established an autoregressive model of high-frequency data on wind farms and calculated the volatility measurement of 5-min data through this model. Chen made systematic research on wind power forecasts based on the volatility model. The mature application of wind power industry volatility forecasting provides an effective reference for natural gas load volatility forecasting.

From the above literature, we can find: (1) The parameters of the GARCH family of models have a strong explanatory power for volatility analysis and volatility modeling. (2) Compared with the single GARCH family model, the performance of the hybrid model composed of the GARCH family and the neural network was improved. (3) Deep neural networks such as LSTM have the ability to learn complex nonlinear relationships from data sets. (4) The prediction of natural gas load volatility has strong practical significance, but there is a lack of relevant research. (5) Appropriate input features improve the prediction accuracy of the model, while overfull input features will have a negative impact on the prediction. XGBoost can effectively screen important features.

3 | DATA ANALYSIS AND MODEL CONSTRUCTION

3.1 | Data selection and statistical description

3.1.1 | Data source and preprocessing

In previous studies, the researchers directly forecast the natural gas load, ignoring the fluctuation of the natural gas load, which makes it difficult to reasonably estimate natural gas load intervals. In the study of this paper, the data are taken from the real load values of a metering branch under a natural gas station in Xi’an City. The natural gas station, located in Weiyang District, Xi’an, is an important node of the entire natural gas pipeline network in Xi’an. Weiyang District is one of the central urban areas of Xi’an, with a dense population and developed industry. Therefore, the natural gas station here is complicated, with one intake pipeline and three output pipelines. This station is one of the relay stations for natural gas supply in Xi’an, so the load data of a metering branch under this station is selected in this study. This metering branch provides mainly urban gas and is therefore representative and universal in terms of urban gas forecasting. The subject of this study is natural gas hourly load volatility, and the load values of a metering branch are used as the sample for the study, with a sample interval of approximately 20 s (366,920 sample points in total) from May 24 to August 17, 2021. This is shown in Figure 1, as can be seen herein, the real load data variation of natural gas has a certain cyclical variation pattern, but at the same time the variation fluctuates sharply. The transition between two peaks and
through load is not very gentle, but almost abruptly increased or decreased. From this load data, it is possible to calculate the gas hourly load fluctuation rate. In addition, the parameters of the GARCH family model were considered in addition to the construction of the feature data set. It was also considered that the load influencing factors of natural gas include meteorological factors such as temperature, weather, season, and wind direction,\textsuperscript{26–28} so we added these meteorological factors to the feature set of natural gas load volatility for the experiment. Therefore, the feature data set includes the main parameters of the GARCH family of models, weather, temperature, wind direction, season, and so on.

There are still some abnormal data in the unprocessed original data set. This is because errors in the collection and recording of these high-frequency data will result in missing or abnormal data. The use of unprocessed data may affect the accuracy of prediction, so the original data needs to be filtered and processed. This study uses the outlier detection method and the horizontal processing method to correct the abnormal data. The corrected data becomes relatively smooth and closer to the true value of the load, as shown in Figure 2 (special reference: the amount of data is huge, so only load values for 100 consecutive time points are shown).

3.1.2 | Definition of natural gas load volatility

The word “volatility” comes from the financial asset industry. In a narrow sense, volatility refers to the standard deviation of continuous compound interest returns within a certain period of time. In a broad sense, volatility refers to the degree of unpredictable changes in variables over time.\textsuperscript{29} The volatility in finance is a measure of the uncertainty of the return on assets and reflects the risk level of the return on assets. The natural gas load volatility reasonably reflects the changing law of natural gas load. In probability statistics, the average
value and standard deviation can reflect the most basic characteristics of variables, in which the average value can reflect the degree of concentration of the variable, and the standard deviation can reflect the degree of dispersion of the variable. The ratio of the standard deviation to the average value indicates the relative size of the dispersion of the variable. This ratio is called the coefficient of variation, which can well reflect the law of variation of the variable. Therefore, this study defines the natural gas load volatility based on the coefficient of variation, that is, the \( f \) (natural gas load volatility) is the ratio of \( s \) (standard deviation of the load) to \( \bar{x} \) (the average value of the load), where the standard deviation and the average value are calculated based on the number of samples \( n \). The formula is as follows:

\[
f = \frac{s}{\bar{x}},
\]

\[
s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2},
\]

\[
\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i,
\]

The calculation of the volatility of natural gas hourly load is to divide the standard deviation of the load by the average value of the load in a unit hour. Thereinto, the average hourly load is the arithmetic average within the hour, as shown in Figure 3. The standard deviation of the hourly load is the standard deviation based on a given sample within the hour, as shown in Figure 4.

### 3.1.3 Descriptive statistics

After data processing, it is necessary to further analyze the data characteristics of natural gas hourly load volatility, which can reflect the basic law of natural gas load volatility. The hourly load volatility from 00:00 on May 24, 2021, to 23:00 on August 17, 2021, is plotted as a time series diagram, as shown in Figure 5. From 7:00 to 9:00 and from 22:00 to 0:00 every day, the volatility is large and changes strongly, and the volatility is small and changes stably in other time periods.

As shown in Figure 6 and Table 1, the smaller the standard deviation, the weaker the dispersion of the sequence. The standard deviation of the sequence is 0.069631, indicating that the dispersion of the sequence is weaker. The skewness of the sequence is 3.944191, which is greater than 0, indicating that the sample has a right-skewed distribution. The kurtosis of the sequence is 20.64961, which is greater than the peak value 3 of the normal distribution, indicating that the sequence has a tall and thin distribution with the characteristics of sharp peaks and thick tails. The JB statistic is 32141.22, and the \( p \)-value is 0, so the sequence does not obey the normal distribution. In addition, through the distribution characteristics and changing laws of the above chart, we can see that the sequence distribution may be asymmetry, and there may be a certain phenomenon of volatility aggregation, which needs to be further verified. In summary, the hourly load volatility has the characteristics of right skewness, peak and thick tail, non-normality, and asymmetry.
FIGURE 4  Hourly load standard deviation

FIGURE 5  Time series diagram of the hourly volatility of natural gas load

FIGURE 6  Histogram of load volatility series
### 3.1.4 Stationarity test

In the analysis of time series data, the stationarity test is indispensable, which can avoid false regression problems. The commonly used stationarity test method (ADF test) is used here, in which the SIC criterion is selected as the test standard, and the ADF test results are shown in Table 2. Compared with the absolute value distribution of 1%, 5%, and 10%, the absolute value of the $T$-value is greater than the above three levels. It can be concluded that the original hypothesis is rejected at the above three levels, and there is no unit root in the sequence. And the test $p$-value of the load rate series is 0, indicating that the original hypothesis is completely rejected. Therefore, it is a stationary series.

### 3.1.5 Autocorrelation test

For the model $y_t = b_0 + b_1 x_{1t} + b_2 x_{2t} + \cdots + b_k x_{kt} + u_t$, if there is a correlation among the various period values of the random error term, the covariance $\text{cov}(u_t, u_s) \neq 0 (t \neq s, t, s = 1, 2, \ldots, n)$, then there is autocorrelation between the random error terms. As shown in Table 3, the AC (autocorrelation coefficient) value and PAC (partial autocorrelation coefficient) value of this sequence are not significantly different from zero value when the significance level equals 5%. The $p$-values of $Q$ statistics are all less than 5%. It can be seen that the null hypothesis is rejected at the 5% significance level, where the serial autocorrelation is considered. It shows that the hourly load volatility of natural gas has a long memory.

### 3.1.6 ARCH effect test

If the variance of the random error term $u_t$ in the model $y_t = b_0 + b_1 x_{1t} + b_2 x_{2t} + \cdots + b_k x_{kt} + u_t$ is not constant, it meets $\text{var}(u_t) = \sigma_t^2 \neq \text{constant}(t = 1, 2, \ldots, n)$, then the random error term $u_t$ has heteroscedasticity. The ARCH process is $\sigma_t^2 = a_0 + a_1 \sigma_{t-1}^2 + a_2 \sigma_{t-2}^2 + \cdots + a_p \sigma_{t-p}^2 + \epsilon_t$, and the original hypothesis is $H_0 : a_1 = a_2 = \ldots = a_p = 0$. Where $p$ is the order of the ARCH process and $\epsilon_t$ is the random error term.

After the stationarity test and autocorrelation test are performed on the sequence, the residual sequence obtained by the model estimation also needs to be an ARCH test. If the residual sequence passes the ARCH effect test, indicating that the residual sequence has an ARCH (autoregressive conditional heteroskedasticity) effect, and GARCH can be used to model and analyze the natural gas load volatility. On the contrary, it means that the sequence does not meet the GARCH modeling conditions.

(1) Graphical test of ARCH effect test:

Natural logarithm processing is performed on the sample sequence $\{s_p\}$, that is, the sequence $\{\ln(s_p)\}$ is estimated as the dependent variable, so the basic form of the estimation is:

$$\ln(s_p) = \mu + \rho \times \ln(s_{p-1}) + u_t. \quad (4)$$

Estimated by the ordinary least squares, the results are as follows:

$$\ln(s_p) = -4.763531 + 0.131559 \ln(s_{p-1}) + \hat{u}_t. \quad (5)$$

By drawing the residual sequence of the average value equation regression, it can be judged whether there is an ARCH effect according to the volatility cluster phenomenon presented in the time series diagram, as shown in Figure 7. Through the residual time series diagram, we can observe the clustering phenomenon of fluctuations: it is smaller in a period of time (e.g., the 625th hour to the 800th hour), and larger in other periods of time (e.g., the 1200th hour to the 1300th hour), which shows that the error term may have conditional heteroscedasticity.

---

**TABLE 1** Sequence statistics of natural gas load volatility

| Statistics | Std. dev | Skewness | Kurtosis | Jarque-Bera | Probability |
|------------|----------|----------|----------|-------------|-------------|
| Value      | 0.069631 | 3.944191 | 20.64961 | 32141.22    | 0           |

**TABLE 2** ADF inspection table of natural gas load volatility

| T-test statistics | Probability |
|-------------------|-------------|
| ADF statistics    | -42.46495   | 0.0000      |
| Significance level 1% test critical value | -3.433321 |
| Significance level 5% test critical value | -2.862739 |
| Significance level 10% test critical value | -2.567454 |
ARCH LM test:
The test of ARCH LM mainly applies $F$-statistic and chi-square statistics to judge whether there is an ARCH effect or not. As shown in Table 4, the probability $p$-values of the $F$-statistic and the chi-square statistic are both less than 0.05, rejecting the null hypothesis, and the sequence has an ARCH effect. Therefore, the GARCH model can better fit the fluctuation sequence.

3.2 Construction of hybrid forecasting model

These complex statistical characteristics mean that the natural gas load volatility is difficult to predict, and it is difficult for a single statistical method or machine learning algorithm to reasonably explain the characteristics of natural gas volatility. Through the above analysis, we know that the GARCH family models can explain the characteristics of natural gas load fluctuations well. Deep learning algorithms such as LSTM have the advantages of stronger learning ability, adaptability, portability, and so on, which not only conform to the long-term memory characteristics of natural gas load volatility but also can quickly and accurately predict natural gas volatility. Therefore, the combination of

| Autocorrelation coefficient | Partial autocorrelation coefficient | Q statistic | Probability |
|-----------------------------|-------------------------------------|-------------|-------------|
| 1 0.067                      | 0.067                               | 9.000       | 0.002       |
| 2 0.011                      | 0.007                               | 9.4661      | 0.009       |
| 3 −0.015                     | −0.016                              | 9.9158      | 0.019       |
| 4 −0.031                     | −0.029                              | 11.880      | 0.018       |
| 5 −0.009                     | −0.005                              | 12.042      | 0.034       |
| 6 −0.039                     | −0.038                              | 15.163      | 0.019       |
| 7 −0.002                     | 0.002                               | 15.173      | 0.034       |
| 8 0.073                      | 0.073                               | 26.111      | 0.001       |
| 9 0.041                      | 0.030                               | 29.581      | 0.001       |
| 10 0.026                     | 0.017                               | 30.943      | 0.001       |

| $F$ statistics | 6.604377 | Probability | 0.0000 |
|----------------|----------|-------------|--------|
| Chi-square statistics | 39.00756 | Probability | 0.0000 |
statistical methods and deep learning algorithms is of advanced significance.

3.2.1 | Analysis of predictability of load volatility

The change of natural gas load does not follow some strict rules and it is often affected by a variety of factors. Therefore, the change rule of the natural gas load is uncertain and nonlinear. From the characteristics of volatility and the load characteristics of natural gas, the characteristics of natural gas load fluctuations can be summarized:

(1) Time series:
The return on financial assets is similar to long-term memory, and the volatility of natural gas load also has long-term memory. According to the results of the autocorrelation test, it can be found that the volatility of natural gas load has a strong autocorrelation, indicating that historical information will affect the future volatility for a long time, and there is a correlation between the previous volatility and latter volatility.

(2) Long-term memory:
The return on financial assets is similar to long-term memory, and the volatility of natural gas load also has long-term memory. According to the results of the autocorrelation test, it can be found that the volatility of natural gas load has a strong autocorrelation, indicating that historical information will affect the future volatility for a long time, and there is a correlation between the previous volatility and latter volatility.

(3) Volatility aggregation
Volatility aggregation means a state that is from one unit of time to the next. Generally speaking, large fluctuations are often followed by large fluctuations, and small fluctuations are followed by smaller fluctuations. According to the above graph test of the ARCH effect, it can be found that there is fluctuation aggregation in the volatility of natural gas load.

(4) Asymmetry
The volatility of time series mostly has an asymmetry, that is, the impact of positive information and negative information has a different degree of influence on the fluctuation of future load. In the financial industry, facing the impact of information of the same intensity, negative information usually has a greater impact on asset price fluctuations than positive information. Through the above descriptive statistics of natural gas load volatility, it can be seen initially that there is also an asymmetry in natural gas load fluctuation. In the following study, asymmetry modeling will be used to further reflect the asymmetry of natural gas load fluctuation.

(5) Non-normality
When the kurtosis value of the sequence is greater than 3, it indicates that the sequence has the characteristic of sharp peaks and thick tails, which do not conform to the normal distribution. According to the above descriptive statistics, it can be judged that the natural gas load volatility conforms to a non-normal distribution.

3.2.2 | GARCH family models and parameter estimation

(1) GARCH model:
When the ARCH model describes the dynamic change process of sequence volatility, many parameters need to be estimated, which leads to a more complicated and changeable estimation of the ARCH model. The GARCH model has one more conditional variance term than the ARCH model, so the complexity of parameter estimation is reduced, the model is more long-memory, and the time-lag structure is more reasonable and flexible. The GARCH model used in general empirical testing is GARCH (1, 1), so this study mainly applies the GARCH (1, 1) model. Equation (6) is the average value equation and Equation (7) is the variance equation.

\[
y_t = x_t'\gamma + u_t, \quad t = 1, 2, 3, ..., T, \quad (6)
\]

\[
\sigma_t^2 = \omega + \alpha u_{t-1}^2 + \beta \sigma_{t-1}^2. \quad (7)
\]

Thereinto, in the average value equation, \(x_t = (x_{t1}, x_{t2}, ..., x_{tk})'\) is the vector of explanatory variables, and \(\gamma = (\gamma_1, \gamma_2, ..., \gamma_k)'\) is the vector of coefficients. In the variance equation, \(\omega\) is a constant term, \(u_{t-1}^2\) is an ARCH term, and \(\sigma_{t-1}^2\) is a GARCH term. Estimate the average value equation and variance equation through the GARCH (1, 1) model, and the results are as follows:

\[
\ln(sp_1) = -4.692999 + 0.145496\ln(sp_{t-1}) + \bar{u}_t. \quad (8)
\]

\[
\sigma_t^2 = 0.141673 - 0.00991u_{t-1}^2 + 0.94613\sigma_{t-1}^2. \quad (9)
\]

(2) TARCH model:
The TARCH model is an asymmetric GARCH model, which refers to the use of dummy variables to set a threshold to distinguish the influence of positive and negative shocks on the law of fluctuation. The commonly used TARCH model is TARCH (1, 1), which counts a dummy variable on the basis of GARCH (1, 1), and meets the following conditions:
\[ \begin{cases} d_{t-1} = 0, u_{t-1} \geq 0 \\ d_{t-1} = 1, u_{t-1} < 0 \end{cases} \]  \quad (10)

Then the variance equation of the TARCH model is shown in Equation (11).
\[ \sigma_t^2 = \omega + \alpha u_{t-1}^2 + \gamma u_{t-1}^2 d_{t-1} + \beta \sigma_{t-1}^2. \]  \quad (11)

Thereinto, \( d_{t-1} \) is a dummy variable. When \( u_{t-1} < 0, d_{t-1} = 1 \), indicating that the asymmetric effect occurs, and bad news will bring \((\alpha + \gamma)\) times the impact, otherwise \( d_{t-1} = 0 \), indicating that the asymmetric effect does not appear, and the good news brings \( \alpha \) times the impact. When \( \gamma \neq 0 \), it indicates that there is a leverage effect. When \( \gamma > 0 \), it shows that the main function of the asymmetric effect is to increase the fluctuation. When \( \gamma < 0 \), it shows that the asymmetric effect is to reduce the fluctuation.

The average value equation and variance equation are estimated by the TARCH (1, 1) model, and the results are as follows:
\[ \ln(s_p) = -4.686323 + 0.147987 \ln(s_{p-1}) + \bar{u}_t, \]  \quad (12)
\[ \sigma_t^2 = 0.17118 - 0.01014u_{t-1}^2 - 0.01958u_{t-1}^2 d_{t-1} + 0.93785\sigma_{t-1}^2, \]  \quad (13)

(3) EGARCH model:

The EGARCH model is also an asymmetric model,\(^{35}\) but the difference is that the variance equation analyzes not \( \sigma_t^2 \) but \( \ln \sigma_t^2 \). When the estimated value of \( \beta \ln \sigma_{t-1}^2 \) in the variance equation is a positive number, it indicates that the sequence has a persistent phenomenon of fluctuations (cluster phenomenon). The variance equation of the EGARCH (1, 1) model is as follows:
\[ \ln(\sigma_t^2) = \omega + \alpha \left( \frac{u_{t-1}}{\sigma_{t-1}} \right) + \gamma u_{t-1} + \beta \ln(\sigma_{t-1}^2). \]  \quad (14)

Thereinto, \( \ln \sigma_t^2 \) is the logarithm of the conditional variance, indicating that the leverage effect is exponential, so the predicted value of the conditional variance must be non-negative. When \( \gamma \neq 0 \), it indicates that there is a leverage effect.

Estimate the average value equation and variance equation through the EGARCH (1, 1) model, and the results are as follows:
\[ \ln(s_p) = -4.866808 + 0.110610 \ln(s_{p-1}) + \bar{u}_t, \]  \quad (15)
\[ \ln(\sigma_t^2) = 0.49664 + 0.19932 \frac{u_{t-1}}{\sigma_{t-1}} - 0.010216 \frac{u_{t-1}}{\sigma_{t-1}} + 0.19143 \ln(\sigma_{t-1}^2). \]  \quad (16)

(4) PARCH model:

The PARCH model is also an asymmetric model,\(^{36}\) and its variance equation analyzes not \( \sigma_t^2 \) but \( \sigma_t^h \) (h represents the power value). The variance equation of the PARCH (1, 1) model is as follows:
\[ \sigma_t^h = \omega + \sum_{i=1}^{p} \alpha_i (|u_{t-i}| - \gamma u_{t-i})^h + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^h. \]  \quad (17)

Thereinto, \( h > 0 \), when \( i = 1, 2, \ldots, r, |\gamma| \leq 1 \). When \( i > r, \gamma = 0, r \leq p \). When \( \gamma \neq 0 \), it indicates that there is a leverage effect.

Estimate the average value equation and variance equation through the PARCH (1, 1) model, and the results are as follows:
\[ \ln(s_p) = -4.694418 + 0.146879 \ln(s_{p-1}) + \bar{u}_t, \]  \quad (18)
\[ \sigma_t = 0.70268 - 0.006185(|u_{t-1}| - 0.30812u_{t-1}) + 0.93566\sigma_{t-1}. \]  \quad (19)

(5) Parameter sliding estimation:

The GARCH family models parameters used in this study include the main parameters of the GARCH model, TARCH model, EGARCH model, and PARCH model. As the parameters of the feature input, a fixed value is not used, but a sliding estimation method.\(^{37}\) That is, the model parameters at time \( t \) are estimated from the volatility data from time \( t-21 \) to time \( t \). The model parameters at time \( t + 1 \) are estimated from the volatility data from time \( t-20 \) to time \( t + 1 \), followed by sliding estimation. The sliding estimation process is shown in Figure 8, and some specific moments estimation parameters are shown in Table 5.

3.2.3 LSTM

To solve the problem of gradient explosion and gradient disappearance, LSTM introduces a gate mechanism on the basis of RNN (recurrent neural network).\(^{38}\) LSTM can effectively solve the problem of long-term dependence and reasonably control the increase and deletion of information in neurons.\(^{39}\) The LSTM network structure is shown in Figure 9.

In the LSTM unit structure, cell \( (c_t) \) is a memory unit that is in charge of the transmission of information. The
input gate \((i_t)\) is in charge of the increase of information, the forget gate \((f_t)\) is in charge of the deletion of information, and the output gate \((o_t)\) is in charge of the output of information. The complete training process of LSTM is that at each time \(t\), the three gates receive the input vector \(x_t\) at time \(t\) and the information of the hidden state \(h_{t-1}\) of the LSTM at time \(t-1\) and the memory unit \(c_t\). Then perform logical operations on the received information, and the logical activation function \(\sigma\) determines whether to activate it. Then the processing result of the input gate and the processing result of the forget gate is integrated to produce a new memory unit \(c_t\), and finally the final output result \(h_t\) is obtained through the nonlinear operation of the output gate. The calculation equation of each process is as follows:

\[
i_t = \sigma(W_{xi}^T x_t + W_{hi}^T h_{t-1} + b_i),
\]

\[
f_t = \sigma(W_{xf}^T x_t + W_{hf}^T h_{t-1} + b_f),
\]

\[
o_t = \sigma(W_{xo}^T x_t + W_{ho}^T h_{t-1} + b_o),
\]

\[
c_t = f_t \times c_{t-1} + i_t \times \tanh(W_{xc}^T x_t + W_{hc}^T h_{t-1} + b_c),
\]

\[
h_t = o_t \tanh(c_t).
\]

Thereinto, \(\sigma\) is generally a nonlinear activation function. \(W_{xi}, W_{xf}, W_{xo}, W_{xc}\) are the weight matrices of the nodes connected to the input vector at each layer. \(W_{hi}, W_{hf}, W_{ho}, W_{hc}\) are the weight matrices of each layer connected to the previous short-term state \(h_{t-1}\). \(b_i, b_f, b_o, b_c\) are the nodes of each layer the bias term.

### 3.2.4 XGBoost

XGBoost implements a general decision tree algorithm. Compared with the gradient lifting algorithm, the performance of the decision tree algorithm can be improved by 10 times. All tree models of XGBoost are CART models, and the tree integration model is shown in Equation (25).

\[
y_i = \sum_{k=1}^{K} f_k \left( x_i \right), f_k \in F.
\]

Thereinto, \(\hat{y}_i\) represents the predicted value of the \(i\)-th sample. \(K\) represents the number of trees, \(F\) is the set of regression trees, \(f_k\) represents \(k\) sub-models in \(F\), and \(x_i\) represents the input sample data. The objective function of XGBoost consists of two parts, a loss function and a regular term, as shown in Equations (26) and (27):

\[
L(\varphi) = \sum_{i=1}^{n} l(y_i, \hat{y}_i(t-1) + f_{t}(x_i)) + \Omega(f_{t}),
\]

\[
\Omega(f_{t}) = \gamma T + \frac{1}{2}\lambda \|\omega\|^2.
\]

Thereinto, \(L(\varphi)\) represents the objective function of the \(t\)-th iteration, \(\hat{y}_i(t-1)\) represents the predicted value of the \(t-1\) iteration. \(\Omega(f_{t})\) represents the regular term of the model of the \(t\)-th iteration, \(\Omega(f_{t})\) plays a role in reducing overfitting. \(\gamma\) and \(\lambda\) are penalty coefficients. \(T\) is the number of leaf nodes, and \(\omega\) is the score of leaf nodes. Perform a second-order Taylor approximation expansion of Equation (26), as shown in Equation (28):

\[
L(\phi) = \sum_{i=1}^{n} [g(t)(x_i) + \frac{1}{2} h(t)(x_i)] + \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2 + \frac{1}{2} \sum_{i=1}^{T} \sum_{j=1}^{T} \omega_i \omega_j.
\]

Thereinto, \(g(t)\) represents the first derivative of sample \(x_i\). \(h(t)\) represents the second derivative of sample \(x_i\). \(\omega_j\) represents the output value of the \(j\)-th leaf node, and \(\gamma T\) represents...
represents the sample subset of the value of the \( j \)-th leaf node.

It can be seen from Equation (28) that the objective function is a convex function. Taking the derivative of \( \omega_j \) and making the derivative function equal to zero value, the \( \omega_j \) that makes the objective function reach the minimum can be obtained, namely:

\[
\omega_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda},
\]

\[
L(\phi)_{\text{min}} = -\frac{1}{2} \sum_{i=1}^{T} \left( \frac{\left( \sum_{i \in I_j} g_i \right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T. \right)
\]

The smaller the value of the objective function, the better the model. In the training process, the greedy algorithm is used to divide the subtree, enumerate the feasible division points, and continuously calculate the node loss to select the leaf nodes with large gains. The gain calculation is shown in Equation (31).

\[
Gain = -\frac{1}{2} \left[ \frac{\left( \sum_{i \in I_j} g_i \right)^2}{\sum_{i \in I_j} h_i + \lambda} + \frac{\left( \sum_{i \in \mathcal{R}} g_i \right)^2}{\sum_{i \in \mathcal{R}} h_i + \lambda} \right] - \gamma.
\]

### 3.2.5 Establishment of hybrid model

Through analyzing previous research, we can find that single GARCH family models often perform poorly when dealing with multiple influencing factors and nonlinear sequences. A single LSTM model has a poor interpretation of volatility, therefore the prediction accuracy of volatility will not be too high. The combination of GARCH family models and LSTM models can effectively solve the shortcomings in volatility forecasting. However, it is difficult to achieve better prediction results simply by combining GARCH family models with LSTM networks, because too many feature inputs often have a negative impact on model predictions. So XGBoost algorithm is applied to the study because XGBoost is equivalent to the “catalyst,” which can effectively improve the prediction performance of the model.

In this study, the sliding estimated GARCH family of model parameters and meteorological factors are used as influential features of the volatility. Then use the XGBoost algorithm to perform feature screening on these influencing factors, and input the screened important features into the LSTM network for prediction.
Finally, the effectiveness of this model is verified by comparing it with a variety of single models and combined models. We call this model the GFMF-XGBoost-LSTM model and the model architecture is shown in Figure 10. The GARCH family models parameters can better explain the fluctuation characteristics of the load series. Thereinto, GARCH can better describe the long-term memory and volatility aggregation. TARCH, EGARCH, and PARCH can better characterize the asymmetry and non-normality of volatility. The sliding estimation parameters in these models are equivalent to extracting pre-training information for the deep neural network, effectively reducing model changes. XGBoost sorts the importance of features and can accurately screen out important features. Using these important features as input features can reduce the interference of irrelevant features and effectively improve the prediction accuracy of the model. The LSTM network has the ability to learn complex and nonlinear relationships from data sets, and can effectively process nonlinear, long-memory time series data.

\[ \text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2, \]  
\[ \text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2}. \]  

Thereinto, \( \hat{y}_i \) is the predicted value, \( y_i \) is the true value, and \( n \) is the number of test samples. RMSE is based on MSE, which is more intuitive in order of magnitude. The ranges of MSE and RMSE are both \([0, +\infty)\). Generally, the larger the value of MSE and RMSE, the greater the error and the lower the prediction accuracy of the model.

4 | EXPERIMENT

4.1 | Data set division and normalization

The data set needs to be divided before it can be input into the model for training, otherwise, the training results may be overfitted due to training on all the data. The experiment divides the data set into the training set, validation set, and test set. Thereinto, the training set data is 60%, the validation set data is 20%, and the test set data is 20%. The training set is used for model fitting data samples. The validation set is used to adjust the hyperparameters of the model and to initially evaluate the model’s ability. After the optimal parameters of the model are found, the data is then tested on the test set so as to evaluate the generalization ability of the model.

As the load volatility and the GARCH family models parameters have different meanings and dimensions, they will have an impact when they are input to the forecasting model. Therefore, this type of data needs to.
be normalized. The normalized data can reduce the training time of the model, accelerate the convergence speed of the model, and further improve the prediction accuracy of the model. The data normalization calculation equation is shown in Equation (34):

\[ x_{\text{norm}} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \] (34)

Thereinto, \( x_{\text{norm}} \) is the normalized value, \( x \) is the original data, \( x_{\text{min}} \) is the minimum value in the original data, \( x_{\text{max}} \) is the maximum value in the original data, and the normalized data size is constrained to be between [0, 1].

**4.2 Experimental results**

4.2.1 The prediction results of a single model

The experiment inputs data into GARCH family models (GARCH, TARCH, EGARCH, and PARCH), commonly used deep learning models (LSTM, MLP, and CNN), and some machine learning models (ARIMA, MLR, and SVR) for training and prediction. The purpose is to compare the forecasting effects of different models on load volatility. As shown in Figure 11, it is the fitting curve of the predicted value and the true value of every single model. It can be seen that the fitting effect of the LSTM and MLP models is better. As shown in Table 6, it is the value of the evaluation index of every single model. Thereinto, the value of MSE (0.0044710) and RMSE (0.0668658) of LSTM is the smallest. Therefore, the overall performance of the LSTM model is better. However, these GARCH family models and machine learning models did not show good prediction performance. This may be due to the non-normality, asymmetry, and heteroscedasticity of natural gas load volatility, which makes it difficult to predict natural gas load volatility.

4.2.2 The prediction results of the combined model

Through the above prediction, it is found that the LSTM model has a better fitting effect and higher prediction accuracy than the prediction results of a single GARCH family model and machine learning model. Furthermore, to improve the prediction accuracy of volatility, we rely on (1) GARCH family models can describe the characteristics of load fluctuations; (2) LSTM has a better prediction effect at the inflection point and surge point of volatility; (3) LSTM The neural network has a good explanation for the data. Through this experiment, we use the parameters of each model in the GARCH family as input features and input them into deep learning models such as LSTM. The prediction results are shown in Figure 12A–D. The evaluation results of the model are shown in Table 7.

Through the above prediction results and evaluation results, it can be found that compared with LSTM. The MSE and RMSE of GARCH-LSTM decreased by 6.48%
and 3.29% relatively. The MSE and RMSE of TARCH-LSTM decreased by 0.75% and 0.38% relatively. The MSE and RMSE of PARCH-LSTM decreased by 0.48% and 0.24% relatively. Therefore, in general, the prediction accuracy of GARCH-LSTM, TARCH-LSTM, and PARCH-LSTM models has improved. The prediction accuracy of other models has not improved significantly, and the prediction accuracy of some combined models has decreased significantly. It shows that the description effects of different GARCH family models parameters are different.

To test whether the GARCH family of parameters, as well as meteorological factors, have a positive impact on the forecasts or not, here is the process we used. We take a total of 15 sliding estimation parameters from the GARCH model, TARCH model, EGARCH model, and PARCH model as the input features of the LSTM network. Second, the meteorological factors are taken as the input features of the LSTM network, and their values are mapped to the [0, 1] interval and transformed into quantitative data. The quantitative results of influencing factors are shown in Table 8. Finally, the 15 parameters of the GARCH family model and the meteorological factors were simultaneously used as input features for the LSTM. The prediction results are shown in Figure 13 and the evaluation results are shown in Table 9.
From the prediction results and evaluation results, when GARCH family parameters are used as input features, it is clear that GARCH family-LSTM has higher prediction accuracy than GARCH family-MLP and GARCH family-CNN. When meteorological features are used as input features, MF-LSTM has higher prediction accuracy than MF-MLP and MF-CNN. When both GARCH family parameters and meteorological features are used as input features, GARCH family-MF-LSTM has higher prediction accuracy than GARCH family-MF-MLP and GARCH family-MF-CNN. However, comparing the prediction accuracy of these nine models with the above models, it was found that too many input features would instead reduce the prediction accuracy. This indicates that too many input features have a negative impact on prediction. All parameters or meteorological features of the GARCH family are used as input features, and the prediction accuracy is poor compared to that of a single GARCH family model, although the fit is better. Compared with a single deep learning model such as LSTM, the fitting effect and prediction accuracy are reduced. Compared with a single machine learning model, the fitting effect and prediction accuracy are reduced. Compared with the combined model of each model of the GARCH family and deep learning, the

**TABLE 7**  Evaluation results of the combined model of each model of the GARCH family and LSTM

| Model          | Mean squared error | Root mean square error |
|----------------|--------------------|-----------------------|
| GARCH-LSTM     | 0.0041812          | 0.0646624             |
| GARCH-MLP      | 0.0065278          | 0.0807947             |
| GARCH-CNN      | 0.0054056          | 0.0735225             |
| TARCH-LSTM     | 0.0044373      | 0.0666132             |
| TARCH-MLP      | 0.0052985          | 0.0727907             |
| TARCH-CNN      | 0.0073086          | 0.0854905             |
| EGARCH-LSTM    | 0.0054027          | 0.0735029             |
| EGARCH-MLP     | 0.0056218          | 0.0749784             |
| EGARCH-CNN     | 0.0102375          | 0.1011805             |
| PARCH-LSTM     | 0.0044494          | 0.0667042             |
| PARCH-MLP      | 0.0053169          | 0.0729172             |
| PARCH-CNN      | 0.0055692          | 0.0746271             |

Abbreviations: CNN, convolutional neural networks; EGARCH, exponential generalized autoregressive conditional heteroskedasticity; GARCH, generalized autoregressive conditional heteroskedasticity; LSTM, long short-term memory; MLP, multilayer perceptron; PARCH, power autoregressive conditionally heteroskedasticity; TARCH, threshold autoregressive conditional heteroskedasticity.

**FIGURE 12** The prediction results of the combined model of each model of the generalized autoregressive conditional heteroscedasticity (GARCH) family and the long short-term memory (LSTM). Panel (A) is GARCH-LSTM and other model prediction results. Panel (B) is threshold autoregressive conditional heteroskedasticity (TARCH)-LSTM and other model prediction results. Panel (C) is exponential generalized autoregressive conditional heteroskedasticity-LSTM and other model prediction results. Panel (D) is power autoregressive conditionally heteroskedasticity-LSTM and other model prediction results.
fitting effect and prediction accuracy are not improved. It can be concluded that too many parameters are used as input features, and the prediction accuracy is not significantly improved. What's more, it will lead to a decrease in the prediction effect of some models. Therefore, only appropriate parameters as input features can effectively improve the prediction accuracy.

4.2.3 Prediction results based on the GFMF-XGBoost-LSTM model

To select appropriate parameters as input features, the XGBoost method is used here to filter features and compare them with other feature selection methods. First, perform XGBoost modeling on the training set containing all the features, use five-fold cross-validation to find the optimal parameters, and sort the features according to $F$ score. Then filter the sorted feature set, use the $F$ score value to evaluate whether the feature can be retained, and remove the feature with the lowest $F$ score in turn; use the AUC (area under the curve) value of the verification set under the new feature subset to determine the prediction of the remaining feature. Whether the result is better. In the process of feature selection, we need to weigh the number of features and the effect of improving the model. However, some features have a limited improvement effect on the model. Therefore, the features that have a greater impact on prediction are selected as much as possible in the experiment. We may wish to select features by setting a threshold $h$ (the specific value of $h$ is set according to the experimental results). If the AUC value of the validation set is increased and greater than $h$, the features that have just been deleted are retained. If the AUC value of the validation set is increased and greater than $h$, we keep the features that were just deleted. If the AUC value of the validation set is increased and less than $h$ or decreased, we remove the deleted features. This algorithm can not only screen out the features that have a greater impact on the target variable but also reduce the redundancy between features. The 22 parameters were filtered by XGBoost and the top 10 parameters in terms of feature importance were finally selected as input features, as shown in Figure 14.
In all trees, the number of times a feature is used to split a node is Weight, and the total gain of a feature at each node split is Total_gain.

$$\text{FScore} = \frac{\text{Total_gain}}{\text{weight}}.$$  \hspace{1cm} (35)

The average gain value is calculated as follows:

$$\text{AverageGain} = \frac{\text{Total_gain}}{\text{Fscore}}.$$  \hspace{1cm} (36)

We sort the features according to their importance, the first tenth features are finally taken as input features, they are f0 (volatility), f3 (hours), f8 (EGARCH_ω), f11 (EGARCH_β), f10 (EGARCH_γ), f12 (GARCH_ω), f13 (EGARCH_α), f16 (PARCH_α), f21 (TARCH_α), f4 (Highest temperature). The optimal parameters of the LSTM network are finally determined through multiple training and verification of the training set. The main parameter settings are shown in Table 10. The prediction results are shown in Figure 15 and the evaluation results are shown in Table 11.

Note: The GFMF-XGBoost-LSTM model is that the GARCH family model parameters and meteorological factors are filtered by the XGBoost algorithm and fed into
the LSTM network for prediction. The GFMF-RF-LSTM model is that the GARCH family model parameters and meteorological factors are filtered by the RF algorithm and fed into the LSTM network for prediction. The GFMF-GBDT-LSTM model is that the GARCH family model parameters and meteorological factors are filtered by the GBDT algorithm and fed into the LSTM network for prediction, and so on.

According to the above prediction results and evaluation results, it can be found that appropriate input characteristics play a positive role in the prediction of the model. GFMF-XGBoost-LSTM model has higher prediction accuracy, less error with the real value, and a better fitting effect. And, to verify that XGBoost has better results in feature selection, the prediction results after XGBoost filtered features are compared with those after the random forest and GBDT filtering, and it can be found that the prediction results after XGBoost filtered features are more accurate, among which GFMF-XGBoost-LSTM has the highest prediction accuracy.

To confirm that the proposed model has obvious advantages over other comparison models, we compare the MSEs of the GFMF-XGBoost-LSTM model and several models with higher prediction accuracy in Tables 6, 7, 9, and 11 above. Compared with GFMF-XGBoost-MLP model, GFMF-XGBoost-CNN model, GARCH family-MF-LSTM model, MF-LSTM model, GARCH family-LSTM model, PARCH-LSTM model, and TARCH-LSTM model, GFMF-XGBoost-LSTM model has lower MSE and higher accuracy.

Table 11: GFMF-XGBoost-LSTM and other model evaluation results

| Model                  | Mean squared error | Root mean square error |
|------------------------|--------------------|-----------------------|
| GFMF-XGBoost-LSTM      | 0.0021462          | 0.0463269             |
| GFMF-XGBoost-MLP       | 0.0022471          | 0.0474035             |
| GFMF-XGBoost-CNN       | 0.0028854          | 0.0537156             |
| GFMF-RF-LSTM           | 0.0035279          | 0.0593963             |
| GFMF-RF-MLP            | 0.0036105          | 0.0600874             |
| GFMF-RF-CNN            | 0.0036590          | 0.0604898             |
| GFMF-GBDT-LSTM         | 0.0032324          | 0.0568702             |
| GFMF-GBDT-MLP          | 0.0033403          | 0.0577957             |
| GFMF-GBDT-CNN          | 0.0032973          | 0.0574225             |

Table 12: Model accuracy improvement

| Model                  | Mean squared error | Mean squared error (%) |
|------------------------|--------------------|------------------------|
| GFMF-XGBoost-LSTM      | 0.0021462          | 0                      |
| GFMF-XGBoost-MLP       | 0.0022471          | -4.490232              |
| GFMF-XGBoost-CNN       | 0.0028854          | -25.61863              |
| GARCH family-MF-LSTM   | 0.0051259          | -58.13028              |
| MF-LSTM                | 0.0051202          | -58.08367              |
| GARCH family-LSTM      | 0.0051408          | -58.25163              |
| PARCH-LSTM             | 0.0044494          | -51.76428              |
| TARCH-LSTM             | 0.0044373          | -51.63275              |
| GARCH-LSTM             | 0.0041812          | -48.67024              |
| LSTM                   | 0.0044710          | -51.99732              |

Abbreviations: CNN, convolutional neural networks; GBDT, gradient boosting decision tree; GFMF, parameters of the GARCH family and meteorological features; LSTM, long short-term memory; MLP, multilayer perceptron; RF, random forests; XGBoost, extreme gradient boosting.
TARCH-LSTM model, GATCH-LSTM model, and LSTM model, GFMF-XGBoost-LSTM model decreased by 4.490%, 25.619%, 58.130%, 58.252%, 51.764%, 51.633%, 48.670%, and 51.997%, respectively. As shown in Table 12. In short, the GFMF-XGBoost-LSTM model has an average reduction of 45.404% in MSE, which shows that the GFMF-XGBoost-LSTM model has higher prediction accuracy.

In natural gas load forecasting, if we ignore the fluctuation and heteroskedasticity of the load, the standard error and confidence interval of the mean gas load will be too narrow where the prediction of natural gas load volatility can effectively solve this problem. The prediction of natural gas load volatility can not only make the interval prediction of natural gas load feasible but also determine the probability of future load mean in a certain interval. In view of the large fluctuation, weak regularity, and strong uncertainty of load volatility, in the next research step, we also calculate the confidence interval of hourly natural gas load volatility, so that the estimated volatility is within the predictable range as far as possible, so as to prepare for the interval prediction of natural gas load in the next research step. Figure 16 shows the fitting curve of the predicted value, volatility within the 90% confidence interval and real volatility of the GFMF-XGBoost-LSTM model at the time, in which the average value of volatility in the figure is 0.0200 and the 90% confidence interval of the average value is (0.0157, 0.0244). It should be noted that in the process of calculating the confidence interval, due to the huge volatility, the lower limit of the 90% confidence interval has a negative value, and we treat the negative value as a zero value. This is because when the natural gas load has extremely small fluctuations (i.e., the volatility tends to zero), it will not seriously affect the safety of natural gas production and operation. But it is worth noting that people should pay enough attention when the natural gas load fluctuates greatly (i.e., the volatility is obvious). The confidence interval of volatility can continue to carry out the interval prediction research of natural gas load, and it also provides more reliable decision support for natural gas production and operation.

4.3 | Result analysis

The prediction results of a single LSTM model are better than a GARCH family model, MLP, CNN, ARIMA, MLR, and SVR. The MSE of the LSTM model is 14.939% lower than the GARCH model, 14.955% lower than the TARCH model, 14.937% lower than the EGARCH model, 14.963% lower than the PARCH model, 13.570% lower than the MLP model, 15.796% lower than CNN model, 7.592% lower than ARIMA model, 6.848% lower than MLR model and 30.852% lower than SVR model. As shown in Table 6.

The parameters of each model in the GARCH family are input into the depth neural network as information features, and the prediction results are quite different. Thereinto, the parameters of GARCH family models have a positive effect on the prediction accuracy of LSTM, while the prediction accuracy of MLP and CNN models has been reduced. Compared with the MSE of the LSTM model, the GARCH-LSTM model decreased by 6.931%, the TARCH-LSTM model decreased by 0.759%, EGARCH-LSTM model increased by 17.245% and PARCH-LSTM model decreased by 0.485%. It can be seen that the GARCH-LSTM model has the best prediction performance among these combined models, while the EGARCH-LSTM model has poor performance, as shown in Table 7.
Taking all the parameters of GARCH family models and Meteorological factors as input features significantly reduces the prediction accuracy. Thereinto, compared with MSE of GARCH-LSTM model, GARCH family-LSTM model increased by 18.667%, GARCH family-MLP model increased by 21.800%, GARCH family-CNN model increased by 46.063%, MF-LSTM model increased by 18.339%, MF-MLP model increased by 46.913%, MF-CNN model increased by 33.998%, GARCH family-MF-LSTM model increased by 18.430%, GARCH family-MF-MLP model increased by 41.944%, and GARCH family-MF-CNN model increased by 47.930%. It can be seen that inputting too many features will reduce the prediction accuracy, as shown in Table 9.

The feature filtering by the XGBoost algorithm will significantly improve the prediction accuracy of the combined model. The XGBoost filtered features were found to improve the prediction accuracy when compared with the Random Forest and GBDT algorithms. The prediction accuracy of GFMF-XGBoost-LSTM, GFMF-XGBoost-MLP, and GFMF-XGBoost-CNN were all significantly improved, with the GFMF-XGBoost-LSTM model showing the largest improvement in prediction accuracy, as shown in Table 11. The improved results of the model accuracy are shown in Table 12. It can be seen that the GFMF-XGBoost-LSTM model has the best prediction effect and the highest accuracy.

5 | CONCLUSION

For the problem that the natural gas load volatility is difficult to predict and the traditional volatility forecasting accuracy is low, this study proposes a natural gas load volatility hybrid prediction model based on the GARCH family, XGBoost algorithm, an LSTM network. First, the XGBoost algorithm is used to sort and screen the GARCH family parameters and meteorological factors by feature importance, and then the volatility is predicted by the LSTM network. Finally, it is compared with a variety of single models and combined models, and the confidence interval of the volatility is calculated. The experimental results show that the model proposed in this study effectively predicts the volatility of natural gas load, improves the prediction accuracy of volatility, and can provide decision-making support and useful references for natural gas dispatch and management, pipeline network construction, and pricing. Through experiments, it can be found that (1) using the parameters of the GARCH family as information features can effectively extract pre-training information and improve the prediction performance of volatility; (2) In the prediction of natural gas load volatility, the LSTM network performs better than the MLP network and the CNN network. The prediction fitting effect is better; (3) Using the XGBoost algorithm to filter the multi-element features can further improve the prediction accuracy. The most obvious advantages of the model proposed in this study are: (1) The load volatility is characterized by the parameters of the GARCH family model. (2) The appropriate parameters are selected as input features by XGBoost. (3) And the long-term memory sequence is predicted by LSTM. All of these have typical significance and can be applied to general load fluctuation forecasting research. In addition, due to the limited space and limited amount of data, this study still lacks further experiments to predict the natural gas load interval. This is a problem that needs to be addressed in following research.

ACKNOWLEDGMENTS

They also thank the editor and the anonymous reviewers for their valuable comments. This study was supported by the National Natural Science Foundation of China (Grant number 62072363) and the Yulin City Science and Technology Plan Project (Grant number CXY-2020-063).

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

Huibin Zeng is the author of the manuscript and the designer of the experiment. Bilin Shao grasps the writing direction of the whole manuscript. Genqing Bian is in charge of collecting the data. Hongbin Dai scrutinized all the work. Fangyu Zhou made valuable comments on the whole article.

DATA AVAILABILITY STATEMENT

The data cannot be made public because it is provided by the city’s natural gas company. However, data for this study is available on request from the corresponding author.

ORCID

Huibin Zeng http://orcid.org/0000-0001-5499-9217
Bilin Shao https://orcid.org/0000-0003-4043-4463
Genqing Bian https://orcid.org/0000-0002-6058-4832
Hongbin Dai https://orcid.org/0000-0002-8632-2515

REFERENCES

1. Zhang T, Bai H, Sun S. A self-adaptive deep learning algorithm for intelligent natural gas pipeline control. Energy Rep. 2021;7:3488-3496. doi:10.1016/j.egyr.2021.06.011
2. BP Energy Outlook 2020. https://www.bp.com/content/dam/bp/country-sites/zh_cn/china/home/reports/bp-energy-outlook/2020/energy-outlook-2020-press-release-cn.pdf
3. International Energy Agency. World Energy Outlook 2020. Accessed April 6, 2021. https://www.iea.org/reports/world-energy-outlook-2020

4. Zeng H, Shao B, Bian G, Song D, Li X. A survey of research progress and hot front of natural gas load forecasting from technical perspective. IEEE Access. 2020;8:222824-222840. doi:10.1109/ACCESS.2020.3044052

5. Shen Z, Ritter M. Forecasting volatility of wind power production. Appl Energ. 2016;176:295-308. doi:10.1016/j.apenergy.2016.05.071

6. Engle RF. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica. 1982;50(4):987-1007. doi:10.2307/1912773

7. Bollerslev T. Generalized autoregressive conditional heteroskedasticity. J Econometrics. 1986;31(3):307-327. doi:10.1016/0304-4076(86)90063-1

8. Cifter A. Forecasting electricity price volatility with the Markov-switching GARCH model: evidence from the Nordic electric power market. Electr Pow Syst Res. 2013;102:61-67. doi:10.1016/j.epsr.2013.04.007

9. Cai G, Wu Z. Research on Realized HAR GARCH model based on heteroscedasticity and time-varying fluctuation. Statist Decision. 2021;37(16):162-166. doi:10.13546/j.cnki.tjjyc.2021.06.037

10. Kristjanpoller W, Minutolo M. Forecasting volatility of oil price using an artificial neural network-GARCH model. Expert Syst Appl. 2016;65:233-241. doi:10.1016/j.eswa.2016.08.045

11. Kim HY, Won CH. Forecasting the volatility of stock price index: a hybrid model integrating LSTM with multiple GARCH-type models. Expert Syst Appl. 2018;103:25-37. doi:10.1016/j.eswa.2018.03.002

12. Hu Y, Ni J, Wen L. A hybrid deep learning approach by integrating LSTM-ANN networks with GARCH model for copper price volatility prediction. Physica A. 2020;557(14):124907. doi:10.1016/j.physa.2020.124907

13. Su JB. How to promote the performance of parametric volatility forecasts in the stock market? a neural networks approach. Entropy. 2021;23(9):1151. doi:10.3390/ent23091151

14. Vidal A, Kristjanpoller W. Gold Volatility Prediction using a CNN-LSTM approach. Expert Syst Appl. 2020;157:113481. doi:10.1016/j.eswa.2020.113481

15. Ghasemi-Marzbal A. A developed short-term electricity price and load forecasting method based on data processing, support vector machine, and virus colony search. Energy Effic. 2020;13(7):1524-1525. doi:10.1007/s12053-020-09898-w

16. Zolfaghari M, Sahabi B. A hybrid approach to model and forecast the electricity consumption by NeuroWavelet and ARIMAX-GARCH models. Energ Effic. 2020;12(8):2099-2122. doi:10.1007/s12053-019-09800-3

17. Bikcera C, Verheijen L, Welland S. Density forecasting of daily electricity demand with ARMA-GARCH, CAViaR, and CARE econometric models. Sust Ener Grids. 2018;13:148-156. doi:10.1016/j.jsegan.2018.01.001

18. Kang L, Hu G, Huang H, Lu W, Liu L. Urban traffic travel time short-term prediction model based on spatio-temporal feature extraction. J Adv Transport. 2020;2020(332):1-16. doi:10.1155/2020/3247847

19. Zhang Y, Zhang R, Ma Q, et al. A feature selection and multi-model fusion-based approach of predicting air quality. ISAT. 2020;100:210-220. doi:10.1016/j.isatra.2019.11.023

20. Tao R, Li X, Yudhanto A, Alfano M, Lubineau G. Laser-based interfacial patterning enables toughening of CFRP/epoxy joints through bridging of adhesive ligaments. Compos Part A Appl S. 2020;139:106094. doi:10.1016/j.compositesa.2020.106094

21. Hu P, Pulungan D, Tao R, Lubineau G. Influence of curing processes on the development of fiber bridging during delamination in composite laminates. Compos Part A Appl S. 2021;149:106564. doi:10.1016/j.compositesa.2021.106564

22. Bajcar T, Cimerman F, Širok B, Ameršek M. Impact assessment of traffic-induced vibration on natural gas transmission pipeline. J Loss Prev Proc. 2012;25(6):1055-1068. doi:10.1016/j.jlpp.2012.07.021

23. Liang Z, Yang Q, Zhang J, Zhu B. Mechanical analysis of buried polyethylene pipelines under ground overload. J Fail Anal Prev. 2019;19(1):193-203. doi:10.1007/s11668-019-00600-6

24. Agrawal MR, Boland J, Ridley B. Analysis of wind farm output: estimation of volatility using high-frequency data. Environ Model Assess. 2013;18(4):481-492. doi:10.1007/s10660-013-9357-3

25. Chen H Research on Wind Power Forecasting Based on Volatility Models. Southeast University; 2013.

26. Qiao W, Chen B, Wu S, et al. A forecasting model of natural gas daily load based on wavelet transform and LSSVM-DE. Nat Gas Ind. 2014;34(9):118-124. doi:10.3787/j.isnn.1000-0976.2014.09.019

27. Song J, Pan H. Gas daily load forecasting based on fruit fly optimization algorithm and SVM. Cont Eng China. 2017;24(10):1995-2002. doi:10.14107/j.cnki.kzgc.160468

28. Akpinar M, Yumusak N. Year ahead demand forecast of city natural gas using seasonal time series methods. Energies. 2016;9(7):727. doi:10.3390/en9070727

29. Zhang A. Modeling, Computing Methods of Implied Volatility and Its Application. Shanghai Jiao Tong University; 2009.

30. Gong W, Liu J, He X, et al. Load restoration considering load fluctuation rate and load complementary coefficient power system. Technology. 2014;38(9):2490-2496.

31. Tian X. The Empirical Analysis of the Volatility of the Stock Price Based on the Hybrid Model Integrating LSTM. Harbin Institute of Technology; 2019.

32. Wang L, Tang L, Qiu X, Zhang X, Ma R. The risk measurement of China’s carbon financial market: based on Garch and Var model. Appl Ecol Env Res. 2019;17(4):9301-9315. doi:10.15666/aeer/1704_93019315

33. Bali T, Theodossiou P. A conditional-SGT-VaR approach with alternative GARCH models. Ann Oper Res. 2007;151(1):241-267. doi:10.1007/s10479-006-0118-4

34. Lin Z. Modelling and forecasting the stock market volatility of SSE Composite Index using GARCH models. Future Gener Comp Sys. 2018;79:960-972. doi:10.1016/j.future.2017.08.033

35. Xu Y, Lien D. Forecasting volatilities of oil and gas assets: a comparison of GAS, GARCH, and EGARCH models. J Forecasting. 2021;41(2):259-278. doi:10.1002/for.2812

36. McMillan D, Garcia R. Does information help intra-day volatility? forecasts? J Forecasting. 2013;32(1):1-9. doi:10.1002/for.1243
37. Cao W, Ren S. Research on RMB exchange rate fluctuation prediction based on LSTM and GARCH hybrid model. *Appl Res Comput*. 2020;37(S1):79-82.
38. Kong W, Dong Z, Hill D, Luo F, Xu Y. Short-term residential load forecasting based on resident behaviour learning. *IEEE T Power Syst*. 2018;33(1):1087-1088. doi:10.1109/TPWRS.2017.2688178
39. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997;9(8):1735-1780. doi:10.1162/neco.1997.9.8.1735
40. Kong W, Dong Z, Jia Y, Hill D, Xu Y, Zhang Y. Short-term residential load forecasting based on LSTM recurrent neural network. *IEEE T Smart Grid*. 2019;10(1):841-851. doi:10.1109/TSG.2017.2753802
41. Yang Z, Wang Y, Li J, Liu L, Ma J, Zhong Y. Airport arrival flow prediction considering meteorological factors based on deep-learning methods. *Complexity*. 2020;2020:6309272. doi:10.1155/2020/6309272
42. Zhu J, Zou X, Xiong W,. Short-term power load forecasting based on prophet and XGBoost mixed model. *Modern Electric Power*. 2021;38(03):325-331.
43. Doan E. Analysis of the relationship between LSTM network traffic flow prediction performance and statistical characteristics of standard and nonstandard data. *J Forecasting*. 2020;39(8):1213-1228. doi:10.1002/for.2683
44. Dai H, Huang G, Zeng H, Yang F. PM2.5 Concentration prediction based on spatiotemporal feature selection using XGBoost-MSCNN-GA-LSTM. *Sustainability*. 2021;13(21):12071. doi:10.3390/su132112071

**How to cite this article:** Zeng H, Shao B, Bian G, Dai H, Zhou F. A hybrid deep learning approach by integrating extreme gradient boosting-long short-term memory with generalized autoregressive conditional heteroscedasticity family models for natural gas load volatility prediction. *Energy Sci Eng*. 2022;10:1998-2021. doi:10.1002/ese3.1122