Neurocomputing 581 (2024) 127470

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A new feature selection method based on importance measures for crude oil return forecasting

Yuan Zhao [a,1], Yaohui Huang [b,c,1], Zhijin Wang [c], Xiufeng Liu [d,*]

[a] School of Business Administration, South China University of Technology, Wushan Road 381, 510630, Guangzhou, China
[b] College of Electronic Information, Guangxi Minzu University, Daxue East Road 188, 530006 Nanning, China
[c] College of Computer Engineering, Jimei University, Yinjiang Road 185, 361021 Xiamen, China
[d] Department of Technology, Management and Economics, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

ARTICLE INFO

Communicated by L. Zhang

Keywords:
Crude oil return forecasting
Feature selection
Importance measures
Gaussian noise
Prediction errors

ABSTRACT

This paper introduces a novel feature selection method, called Feature Selection based on Importance Measures (FS-IM), to enhance the forecasting of crude oil returns. FS-IM innovatively combines active learning with the application of Gaussian noise to input features and selects the most relevant features using an optimal threshold value. The paper applies a ridge regression (RR) model based on FS-IM (FS-RR) to identify the factors that have important information for crude oil return forecasting. The paper compares FS-IM with other dimension reduction methods such as Principal Component Analysis (PCA), Kernel Principal Component Analysis (KPCA), and Independent Component Analysis (ICA). The results show that FS-IM can significantly improve model accuracy, demonstrating its effectiveness in finding key features. Moreover, FS-IM is more stable and consistent than other dimension reduction methods in enhancing the prediction accuracy in different scenarios, indicating its superior capability in capturing complex relationships between input and output variables. Furthermore, this study compares FS-RR model with other 13 prediction models by conducting experiments using a series of evaluation metrics, different statistical tests, and different step-ahead predictions and training sets. The results confirm that the RR model based on FS-IM can consistently outperform other model in terms of predictive performance and economic value, proving its effectiveness and robustness. This study contributes to the literature on crude oil price forecasting by addressing the challenges of high-dimensional and complex data, and by providing a robust, practical tool for professionals in energy economics and finance.

1. Introduction

Crude oil is a crucial resource in the global economy, significantly affecting sectors such as energy, transportation, manufacturing, and finance [1,2]. The price of crude oil is determined by the balance of supply and demand, as well as factors including geopolitical events, macroeconomic conditions, market sentiments, and projections of future trends. Research by He et al. [3], Zhang et al. [4] and Zhang and Wang [5] shows that forecasting crude oil returns is essential. These forecasts are critical for investors planning their portfolio allocation and hedging strategies, for policymakers in developing economic and energy policies, and for researchers in proposing new theories and models [2]. The growing focus among academics on improving the accuracy of crude oil price predictions [6–8] highlights its importance in making informed decisions and reducing risks.

Crude oil returns forecasting are influenced by many factors such as macroeconomic, supply, geopolitical, demand, and so on [9,10]. For example, Zhang et al. [4] use a large number of factors to predict the crude oil returns. However, the high dimensionality of the data poses computational and statistical challenges, such as the curse of dimensionality, overfitting, multicollinearity, and information redundancy. Thus, forecasting crude oil returns is a challenging task, as it involves dealing with high-dimensional and complex characteristics. Therefore, it is essential to select the most relevant features from a large feature set that can capture the salient information and patterns in the data and improve the forecasting performance of the models.

Feature selection is a vital step that aims to reduce the dimensionality and complexity of the data by selecting the most relevant features from a large feature set. Feature selection can help enhance the prediction performance of forecasting models, as well as reduce the computational cost and avoid overfitting [11]. However, feature selection is not a trivial task, as it requires a proper evaluation of

* Corresponding author.
E-mail addresses: bmyzhao@outlook.com (Y. Zhao), yhhuang5212@gmail.com (Y. Huang), zhijinecnu@gmail.com (Z. Wang), xiuli@dtu.dk (X. Liu).
[1] Equal contributor.

https://doi.org/10.1016/j.neucom.2024.127470
Received 2 September 2023; Received in revised form 24 January 2024; Accepted 4 March 2024
Available online 7 March 2024
0925-2312/© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Nomenclature

List of abbreviations and their corresponding full names

Abbreviations

ARMA	Auto-Regressive Moving Average Model
BR	Bagging Regression
CRBSCI	Commodity Research Bureau's Commodity Index
CSFE	Cumulative Squared Forecast Error
CER	Certainty Equivalent Return
CW	Clark West test
EIA	U.S. Energy Information Administration
FS-IM	Feature Selection-Importance Measures
FS-RR	Feature Selection-Ridge Regression
GBR	Gradient Boosting Regression
HA	Historical Average
ICA	Independent Component Analysis
ICA-RR	Independent Component Analysis-Ridge Regression
IM	Importance Measures
KPCA	Kernel Principal Component Analysis
KPCA-RR	Kernel Principal Component Analysis-Ridge Regression
LASSO	Least absolute shrinkage and selection operator
MLR	Multiple linear Regression
MAE	Mean Absolute Error
MCS	Model Confidence Set
PCA	Principal Component Analysis
PCA-RR	Principal Component Analysis-Ridge Regression
RDI	Real Dollar Index
RF	Random Forest
RMSE	Root Mean Square Error
RR	Ridge Regression
Std	Standard Deviation
SVR	Support Vector Regression
UG	Utility Gain
WTI	West Texas Intermediate
LSTM	Long Short-Term Memory
GRU	Gated Recurrent Unit

the importance of each feature in relation to the forecasting objective. Importance measures are quantitative indicators that reflect the contribution of each feature to the system reliability or performance. Different importance measures may capture different aspects of feature importance, such as sensitivity, correlation, causality, or information content.

Existing methods for dimension reduction, such as principal component analysis (PCA) [5], kernel principal component analysis (KPCA) [12], and independent component analysis (ICA) [13], have some limitations when applied to crude oil return forecasting. First, these methods are based on linear or nonlinear transformations of the original features, which may result in a loss of interpretability and information. For instance, PCA transforms the original features into orthogonal components that explain the maximum variance in the data, but these components may not have any physical meaning or relevance to the forecasting objective. KPCA extends PCA by using kernel functions to map the original features into a higher-dimensional space where they become linearly separable, but this may introduce additional complexity and parameters that need to be tuned. ICA transforms the original features into independent components that are assumed to be generated by latent sources, but these sources may not be identifiable or observable in reality. Second, these methods do not consider the importance of individual features or their interactions with other features, which may affect the forecasting performance. For example, PCA selects the components based on their eigenvalues that reflect their variance contribution to the data, but this may not reflect their predictive power or relevance to the output variable. KPCA selects the components based on their kernel coefficients that reflect their similarity to other features in the kernel space, but this may not reflect their causal or nonlinear effects on the output variable. ICA selects the components based on their mutual information that reflects their independence from other features in terms of information content, but this may not reflect their sensitivity or correlation with the output variable. Third, these methods do not account for the uncertainty and noise in the data, which may lead to inaccurate or unstable results. For instance, PCA assumes that the data follows a multivariate normal distribution with zero mean and constant covariance matrix, but this may not hold inwhich may exhibit non-Gaussianity, heteroscedasticity, and structural breaks. KPCA assumes that the kernel function is positive definite and symmetric, but this may not hold in practice when dealing with noisy or incomplete data. ICA assumes that the latent sources are statistically independent and non-Gaussian distributed, but this may not hold in practice when dealing with dependent or Gaussian distributed sources. Moreover, these methods do not capture the effects of various factors that may cause large fluctuations or structural breaks in crude oil prices. These factors include the impact of wars, natural disasters, pandemics, or policy changes on crude oil supply and demand, such as the 1973 oil crisis, the 2011 Fukushima nuclear disaster, the 2020 COVID-19 pandemic, and the 2021 OPEC+ agreement [2,14,15]. These factors also include the advantages and disadvantages of using different data frequencies and intervals for capturing short-term or long-term patterns and trends in crude oil prices, such as daily, weekly, monthly, quarterly, or yearly data. These factors illustrate how crude oil prices are influenced by various sources of uncertainty and heterogeneity, as well as by nonlinear and dynamic relationships with other variables. Therefore, there is a need for a more robust and reliable feature selection method that can handle high-dimensional and complex data and enhance prediction performance for crude oil return forecasting.

Many studies use Ridge Regression (RR) model for crude oil price forecasting. For example, Hao et al. [16] and Wang et al. [17] use RR model for crude oil price predictions. Moreover, RR model often exhibits higher computationally efficient, especially on large data sets, than Least Absolute Shrinkage and Selection Operator (LASSO), Long Short-Term Memory (LSTM), Random Forest (RF) and so on. Therefore, we use RR as the base model for feature selection and make predictions.

In this paper, we propose a novel feature selection method based on importance measures for crude oil return forecasting. Our method uses active learning to add Gaussian noise to the input features and measure their importance scores based on the prediction errors. Our method also uses an optimal threshold value to select the most relevant features with prediction ability from the original feature set. Our method has several advantages over existing methods. First, our method can effectively select the most relevant features from a large feature set without losing information or interpretability. Second, our method can capture the interactive effects of features on the forecasting performance by using importance measures derived from prediction errors. Third, our method can handle uncertainty and noise in the data by using Gaussian noise to perturb the input features and active learning to update the importance scores.

The main contributions of this paper are as follows:

- We introduce a new feature selection method that uses active learning to add Gaussian noise to the input features and measure their importance scores based on the prediction error change of a

base forecasting model. We also use an optimal threshold value to select the features with crucial information and strong prediction ability.
- We evaluate our method using different-step-ahead forecasting and different training sets, and compare it with other dimension reduction methods and models without feature selection. We show that our method can improve the forecasting accuracy of ridge regression model, and outperform other methods such as principal component analysis, kernel principal component analysis, and independent component analysis.
- We assess the economic value of our method for investors and policymakers, and show that our method can increase the utility gain and certainty equivalent return of crude oil trading compared to the historical average model.

The rest of this paper is organized as follows: Section 2 reviews the related work. Section 3 introduces the methodology of our proposed feature selection method in this study. Section 4 describes the data and experimental settings used in this study. Section 6 concludes the paper and suggests future research directions.

2. Related work

In this section, we review the existing literature on crude oil return forecasting, which can be divided into three main research streams: feature selection methods, dimensionality reduction methods, and forecasting models. Moreover, we also review the related literature on feature selection based on importance scores. Feature selection methods aim to select the most relevant and informative features from a large and diverse feature set that contains various economic, financial, and geopolitical factors affecting crude oil prices. Dimensionality reduction methods aim to reduce the number of features and extract the most salient and representative features or components from the original feature set. Forecasting models aim to predict the future returns of crude oil prices based on the selected or reduced features or components. We discuss the advantages and disadvantages of different methods in each research stream, and compare them with our proposed method. We also identify the main challenges and gaps in the existing literature, and how our paper contributes to addressing them.

2.1. Impact factors of crude oil return forecasting

Crude oil is one of the most important commodities in the global economy, and its price fluctuations have significant implications for various sectors and regions. Therefore, forecasting crude oil returns is a crucial task for researchers, policymakers, and investors. However, forecasting crude oil returns is challenging due to the complex and dynamic nature of the crude oil market, which is influenced by various economic, financial, and geopolitical factors [18]. In this subsection, we review the existing literature on how these factors affect crude oil prices and returns, and how different data frequencies and intervals may influence the forecasting results.

The dynamics of crude oil prices are influenced by a confluence of factors, primarily governed by the supply and demand conditions in the global market. Supply considerations encompass the production decisions of major oil producers, such as OPEC and non-OPEC countries, and the availability and cost of alternative energy sources like shale oil and renewable energy. Demand is shaped by the economic activity and growth of major oil consumers, including China, India, and the US, and is further influenced by the energy efficiency and conservation policies of various nations (e.g., [10,19–21]). Another significant determinant of crude oil price fluctuations is the financial market dynamics and investor behavior. The financialization of crude oil markets, characterized by the increased influence of financial investors such as hedge funds, pension funds, and index traders, has altered the structure of crude oil markets. These investors engage in trading crude oil futures and options contracts based on their expectations, risk preferences, and portfolio strategies, potentially affecting the spot price movements of crude oil. Furthermore, understanding the perspective of market investors on forecasting crude oil return volatility is crucial, as it impacts general market activities through macroeconomic channels [22]. Additionally, the interconnection between financial markets and crude oil markets through channels such as stock market performance, currency exchange rates, and commodity index funds has been examined (e.g., [23–25]). A third major factor influencing crude oil prices is the geopolitical events and uncertainties that impact the stability and security of crude oil production and transportation. The market's sensitivity to political events, policy decisions, international relations, and conflicts in key oil-producing or consuming regions, such as the Middle East, Africa, and Latin America, may result in supply disruptions or demand shocks, leading to price variations. Moreover, these events may influence the expectations and sentiments of market participants, further affecting price dynamics (e.g., [9,26,27]). The intricate interplay of these factors necessitates a comprehensive understanding of their individual and collective impacts on crude oil prices and returns.

The choice of data frequency and interval is an essential issue in forecasting crude oil returns [3]. Different data frequencies and intervals may reflect different aspects and patterns of crude oil price dynamics. For example, daily or intraday data may capture more short-term fluctuations and noise in crude oil prices than monthly or quarterly data. Moreover, different data frequencies and intervals may also affect the selection and performance of forecasting models [28]. For instance, employing a principal component analysis combination approach has shown promise in forecasting crude oil futures market returns [5]. Additionally, a reduced-rank approach to forecasting has been recognized for its practical importance and contributions to academic research in this field [7]. Several studies have used various data frequencies and intervals, such as daily, weekly, monthly, quarterly, or yearly data, to forecast crude oil returns using different models (e.g., [8,29–31]). Notably, employing the default return spread (DFR) has also been shown to be a powerful predictor of crude oil price returns [32]. These studies have found that the forecasting accuracy may vary depending on the data frequency and interval used. Therefore, it is essential to select a suitable data frequency and interval for forecasting crude oil returns.

The above provides the background and motivation for selecting a large and diverse feature set for crude oil return forecasting. By considering various economic, financial, and geopolitical factors, as well as different data frequencies and intervals, we aim to capture the complex and dynamic nature of the crude oil market and improve the forecasting performance.

2.2. Dimensionality reduction methods for crude oil return forecasting

Dimensionality reduction methods are useful for reducing the number of features and extracting the most relevant and informative ones for forecasting crude oil returns [33]. However, different methods have different strengths and weaknesses, and may not capture all the important information or factors that affect crude oil prices. In the following, we briefly compare three popular dimensionality reduction methods, namely PCA, KPCA, and ICA, and highlight their advantages and disadvantages in terms of information loss, interpretability, computational cost, and prediction performance.

PCA is a linear method that transforms a set of correlated features into a set of uncorrelated features called principal components (PCs), which are ordered by the amount of variance they explain in the data [34]. PCA can reduce noise, redundancy, and multicollinearity in the data, improve computational efficiency, and facilitate visualization and interpretation of the data. However, PCA may lose some information due to dimensionality reduction, assume linearity in the data, be sensitive to outliers and scaling, and ignore higher-order dependencies among the features. Several studies have used PCA for crude

oil return forecasting (e.g., [3,5,7]). KPCA is a nonlinear extension of PCA that uses a kernel function to map the original features into a higher-dimensional feature space, where linear PCA is applied. KPCA can capture nonlinear relationships among the features that are not visible in the original space, be more flexible in choosing different kernel functions, and be more robust to outliers and scaling [12]. However, KPCA may be more computationally expensive than PCA, be less interpretable than PCA, and require parameter tuning for the kernel function. Several studies have used KPCA for crude oil return forecasting (e.g., [6,35,36]). ICA is another nonlinear method that transforms a set of features into a set of statistically independent components (ICs), which are assumed to be non-Gaussian sources that generate the observed data through a linear or nonlinear mixing process [13]. ICA can separate sources that are not orthogonal or uncorrelated, capture higher-order dependencies among the features, and be more suitable for analyzing complex systems with multiple factors. However, ICA may lose some information due to dimensionality reduction, assume independence among the sources, be sensitive to noise and scaling, and require parameter tuning for the mixing model and the non-Gaussianity measure. Several studies have used ICA for crude oil return forecasting (e.g., [37,38]).

Combining dimensionality reduction methods with neural networks has been explored to address the challenges of managing a large number of predictors in crude oil return forecasting. This combination helps mitigate overfitting risks and enhances out-of-sample forecasts, particularly crucial in financial applications [39]. A Reduced-Rank Approach (RRA) is another significant advancement, proving its robustness and effectiveness by outperforming traditional methods and offering valuable insights to investors [7]. Furthermore, the use of uncertainty indicators in tandem with dimensionality reduction techniques has shown promise in boosting predictive accuracy for crude oil volatility [40]. However, these methods are not without their limitations. Common challenges in dimensionality reduction for crude oil forecasting include handling rare or unexpected events that can lead to significant market fluctuations [3]. Additionally, these methods may overlook feedback effects and causal relationships between different market variables, potentially impacting the dynamics and interactions of crude oil prices. There is also a concern about the uncertainty and variability of the components or features extracted, that can affect the confidence and robustness of the forecasting models. Moreover, compatibility and consistency with various forecasting models, especially those with specific input requirements, remain key considerations. To address these challenges, there is a pressing need for innovative feature selection methods that can effectively harness a large and diverse feature set without significant information loss, noise, or bias. Such methods should handle nonlinear and dynamic data, integrate domain knowledge, combine various techniques, and align with different forecasting models. This study aims to propose and develop a novel method based on importance measures to enhance crude oil return forecasting performance.

2.3. Forecasting model for crude oil returns

The task of forecasting crude oil future returns is intricate, with methodologies evolving from traditional statistical techniques to cutting-edge machine learning algorithms. Traditional statistical methods, grounded in historical data analysis, have utilized regression [4] and time series analysis [40] to construct predictive models.

The rise of machine learning has introduced non-linear models like support vector regression (SVR) and random forests (RF), which have significantly advanced the forecasting of crude oil returns [41,42]. These methods excel in capturing the complex dynamics inherent in crude oil markets. Neural computing has recently emerged as a transformative force in forecasting, with neural networks such as Long Short-Term Memory (LSTM) models demonstrating exceptional ability to identify temporal patterns and non-linear relationships in time series data [43]. LSTM models, in particular, have outperformed traditional models like ARIMA, offering superior generalization and precision in forecasting [44]. Innovative hybrid models that combine LSTM with Convolutional Neural Network (CNN) have been proposed, utilizing ARIMA and GARCH outputs as features to predict Brent Crude Oil return, showing remarkable performance improvements over conventional models [45]. Additionally, the predictability of crude oil spot price movements has been enhanced by considering information from the term structure of oil futures prices, with LSTM models highlighting non-linear dependencies within the dataset [46]. Furthermore, the application of artificial neural networks has been explored to understand the impact of monetary policy and other major drivers on crude oil prices, considering the exhaustible nature of crude oil [47]. These studies underscore the potential of neural computing in providing more accurate and computationally efficient forecasts.

While this research does not integrate neural computing methods into our model, it acknowledges their growing importance in the literature. The advancements in neural computing offer promising directions for future research in enhancing the predictive accuracy of crude oil return models. This paper focuses on the application of ridge regression (RR) due to its interpretability and computational efficiency, which are particularly valuable for real-time decision-making in financial markets [5,16,48].

2.4. Feature selection methods based on importance measures

Feature selection, a critical phase in forecasting, seeks to identify the most pertinent and informative features from an extensive and multifaceted feature set [49]. This process often leverages importance measures, which gauge the significance of each feature to system reliability or performance according to specific criteria. Various importance measures such as correlation [50], mutual information [51], and chi-square test [52] can be employed for this purpose. In the following, we will explore existing methods that utilize these importance measures for feature selection, contrasting them with our proposed approach.

Correlation-based methods, such as Pearson's correlation coefficient [53], Spearman's rank correlation coefficient [54], and Kendall's rank correlation coefficient [55], offer a fast and effective way to assess linear or nonlinear relationships. However, their limitations lie in possibly overlooking rare or unexpected events, failing to account for feedback effects or causal relationships, and being susceptible to outliers or data scaling. Some examples of correlation-based methods for feature selection are [56,57] and [58]. Mutual information-based methods, on the other hand, delve into the dependency between features and the target variable through mutual information, a measure that quantifies shared information between two random variables. While adept at capturing nonlinear relationships, these methods may encounter challenges such as information loss due to dimensionality reduction, sensitivity to noise and scaling, and the need for parameter tuning [59–61]. Lastly, chi-square test-based methods utilize the chi-square test to evaluate the association between categorical features and the target variable. Though simple and straightforward, they may be constrained by their inapplicability to continuous or ordinal features, assumptions of independence, sensitivity to sample size and distribution, and neglect of higher-order dependencies [62–65]. The nuanced strengths and weaknesses of these methods underscore the importance of method selection tailored to the specific characteristics of the data and the problem domain.

Our proposed method introduces an novel approach that integrates active learning, Gaussian noise, prediction errors, and an optimal threshold value for feature selection. By combining active learning with Gaussian noise, our method generates importance scores for each feature based on their impact on prediction errors. An optimal threshold value, grounded in standard deviation, then aids in selecting the most relevant features from a broad and diverse set. Our method's strengths lie in its ability to handle nonlinear and dynamic data, incorporate domain expertise, synergize various methods, and align with diverse forecasting models, all while preserving information and minimizing noise or bias.

3. Methodology

In this section, we will first formulate the research problem of crude oil return forecasting, then give an overview of our proposed feature selection method based on importance measures (FS-IM), and finally describe the components of our method in detail

3.1. Problem formulation

The problem of crude oil return forecasting can be formulated as follows: Given a dataset $D = (X, Y)$ of input features and output variable, where $X = \{x_1, x_2, \dots, x_n\}$ is a set of n features and $Y = \{y_1, y_2, \dots, y_m\}$ is a set of m corresponding output values, the goal is to select a subset of relevant features $U_S \subseteq X$ that can improve the forecasting accuracy of a base forecasting model f for crude oil return forecasting. The input features include various economic, financial, and geopolitical factors that may affect the crude oil market. The output variable is the crude oil return. The objective function is to minimize the prediction error E of the base forecasting model f on the validation sets $Va = \{Va_1, Va_2, \dots, Va_k\}$ obtained by k-fold cross-validation, given by

$$E = \frac{1}{k}\sum_{j=1}^{k} E^j,$$

where E^j is the prediction error of f on the validation set Va_j. The constraint is to select the optimal number of features $|U_S|$ by using an optimal threshold value τ that minimizes the prediction error E, given by

$$\tau = \arg \min_{\tau \in [\frac{1}{2n} : \frac{1}{n} : \frac{1}{6n}]} E.$$

The main challenges of this problem are the high dimensionality and complexity of the input data, which can cause overfitting, multicollinearity, and noise issues for the forecasting models. Therefore, we propose a novel feature selection method based on importance measures that can overcome some of these challenges and enhance the prediction performance.

3.2. Overview of proposed method

This subsection presents our proposed method for crude oil return forecasting, which consists of two novel feature selection steps based on importance measures and the optimal threshold value. Fig. 1 illustrates the main steps of our method.

Our method aims to select a subset of relevant and informative features from a large number of candidate features that may affect the crude oil return. We use a normalized dataset of input features and output variable, where the input features include various economic, financial, and geopolitical factors, and the output variable is the crude oil return. We divide the dataset into training sets and validation sets by k-fold cross-validation. The first feature selection step evaluates the importance of each feature by adding Gaussian noise to it and observing the change in the prediction error of a base forecasting model. The intuition is that the more the prediction error changes, the more important the feature is. The second feature selection step determines how many features to select by using a range of possible threshold values and selecting the one that minimizes the mean prediction error of the base forecasting model on the validation sets. After obtaining the reduced feature set, we use it to train and test a base forecasting model for crude oil return forecasting. Ridge regression (RR) is a linear regression model with L2 regularization that can reduce overfitting and improve generalization by penalizing large coefficients. RR is often used to forecast the returns of finance assets, such as stock return and crude oil return [16,66]. We choose RR as our base forecasting model because of its advantages in financial markets and its effectiveness in crude oil market.

3.3. Feature selection based on importance measures

3.3.1. Importance measures of features

In the domain of machine learning, feature selection plays a vital role in improving the performance and interpretability of predictive models. Feature selection methods aim to select a subset of relevant features from a large number of candidate features, which can reduce the dimensionality, complexity, and computational cost of the model, as well as enhance the generalization ability and accuracy of the model. Feature selection methods can be broadly classified into three categories: filter methods, wrapper methods, and embedded methods. Filter methods evaluate the features based on some statistical measures, such as correlation, mutual information, or chi-square test, and select the features that have a strong relationship with the target variable. Wrapper methods use a predefined learning algorithm to assess the quality of selected features by their predictive power. Embedded methods incorporate the feature selection process into the model construction and optimize a predefined criterion function.

In this subsection, we propose a novel feature selection method based on importance measures (FS-IM) to select the most relevant features for crude oil return forecasting. The main idea of FS-IM is to use active learning to evaluate the importance of each feature by adding Gaussian noise to it and observing the change in the prediction error of a base forecasting model. The more the prediction error changes, the more important the feature is. Formally, let $D = (X, Y)$ be a normalization dataset of input features and output variable, where $X = \{x_1, x_2, \dots, x_n\}$ is a set of n features and $Y = \{y_1, y_2, \dots, y_m\}$ is a set of m corresponding output values. Let f be a base forecasting model, such as linear regression, support vector regression, or artificial neural network. Let k be the number of folds for cross-validation, and σ be the standard deviation of Gaussian noise. Motivated by random forest [67] and mean impact value [68], we use the change of prediction errors to calculate the importance scores of features. We define the importance score of feature x_i as follows:

Definition 3.1. The importance score of feature x_i, denoted by IM_i, is the average absolute difference between the prediction errors of the original feature and the noisy feature on the validation sets, given by

$$IM_i = \frac{1}{k}\sum_{j=1}^{k} |E^j_{x'_i} - E^j_{x_i}|, \tag{1}$$

where $E^j_{x_i}$ and $E^j_{x'_i}$ are the prediction errors of feature x_i and its noisy version x'_i on the validation set Va_j, respectively.

Algorithm 1 describes the procedural workflow for implementing our FS-IM based feature selection. The algorithm takes as input a normalized dataset of input features and output variable, the number of folds for cross-validation, the standard deviation of Gaussian noise, and the base forecasting model. The algorithm outputs a set of importance scores for each feature, which can be used to select the most relevant features for crude oil return forecasting. A distinctive aspect of this algorithm is its strategic design to mitigate overfitting risks. By employing a cross-validation mechanism and iteratively evaluating feature importance across different subsets of data, the algorithm ensures a robust selection process. This structure not only enhances the model's generalization capabilities but also strengthens its predictive reliability by preventing the skewing of results due to over-reliance on specific data patterns or anomalies. The algorithm consists of the following steps:

- **Step 1**: Set the initial feature set U_S and not-selected feature set $U_{\bar{S}}$. The algorithm initializes the selected feature set and the not-selected feature set as empty sets, and sets the not-selected feature set as the original feature set. This step prepares the feature sets for the feature selection process.

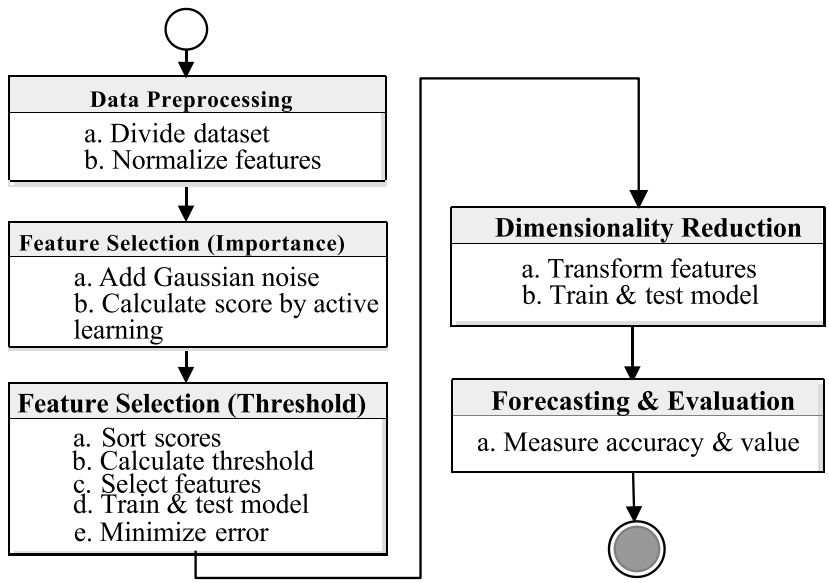

Fig. 1. Overview of the proposed method.

- **Step 2**: Set different training sets and validation sets by k-fold cross-validation. The algorithm divides the dataset into training sets and validation sets by k-fold cross-validation. This step splits the data into different subsets for training and validation purposes.
- **Step 3**: Calculate the importance I of each feature. We add Gaussian noise to each feature of the not-selected feature set and train the model f for each training set according to original feature and the feature with noise respectively. Then, we calculate the prediction error $E_{x'_{ij}}$ and $E_{x_{ij}}$ using each test set of original feature and the feature with noise respectively. We calculate the importance score of each feature using the definition given in Eq. (1). The importance score is the average absolute difference between the prediction errors of the original feature and the noisy feature on the validation sets. The intuition is that the more the prediction error changes, the more important the feature is. This step uses active learning to perturb the features and measure their importance based on the prediction errors of the base forecasting model on the validation sets.
- **Step 4**: The algorithm selects the feature with the highest importance score and adds it to the selected feature set, and removes it from the not-selected feature set. The algorithm repeats steps 3 and 4 until all features are selected. This step selects the most relevant features from the original feature set in a greedy manner, based on their importance.
- **Step 5**: The algorithm normalizes the importance scores by dividing each element by the sum of all elements. This step scales the importance scores to the range of [0, 1] for comparison and interpretation purposes.

The algorithm returns the importance scores of the features, which can be used to select the most relevant features for crude oil return forecasting.

The proposed FS-IM has several advantages over existing methods. First, it can effectively address issues such as over-fitting, multicollinearity, and redundancy. To achieve this, we calculate the importance of features by adding noise, which can measure the influence of features on the prediction by the error change. We also extract the importance information and reduce the number of features to eliminate the redundant information, which can help the model to avoid over-fitting and multicollinearity. Moreover, we use k-fold cross-validation to calculate the average prediction error in FS-IM, which can further mitigate the overfitting problem [69]. Second, it can capture the nonlinear and interactive effects of the features on the output variable by measuring the importance of each feature based on the prediction error of the base forecasting model. Third, it can select the optimal number of features by using a range of possible threshold values and minimizing the mean prediction error of the base forecasting model on the validation sets. This is a novel and effective way to determine the optimal feature subset size, which is often a challenging and subjective task in feature selection. We will explain how to select the optimal threshold value and how it affects the feature selection process and the forecasting performance in the next subsection. Fourth, it can be applied to any base forecasting model and any feature set without any prior assumptions or constraints. Therefore, the proposed FS-IM is a novel and effective method for crude oil return forecasting.

3.3.2. *The optimal threshold value*

After obtaining the importance scores of all features, we need to select a subset of features that can best forecast the crude oil return. A simple way is to set a threshold value and select all features whose importance scores are above the threshold. However, different threshold values may lead to different results, and finding an optimal threshold value is not trivial. We determine the optimal threshold value based on the minimum of the average prediction error. Algorithm 2 shows the procedure of selecting the optimal threshold value and the corresponding features based on importance measurement.

The main objective and challenge of selecting the optimal threshold value is to balance the number of features and the forecasting accuracy. A low threshold value may result in selecting too many features, which may increase the dimensionality, complexity, and computational cost of the model, as well as introduce noise, bias, or over-fitting. A high threshold value may result in selecting too few features, which may lose important information or reduce the explanatory power of the model. Therefore, the optimal threshold value is the one that can select the most relevant and informative features that can enhance the forecasting performance of the base forecasting model without sacrificing much information or introducing much noise or bias.

Let $IM = \{IM_1, IM_2, \ldots, IM_n\}$ be a set of importance scores for each feature, where n is the number of features. Let f be a base forecasting model, and k be the number of folds for cross-validation. We define the optimal threshold value as follows:

Definition 3.2. The optimal threshold value, denoted by T^*, is the value that minimizes the mean prediction error of the base forecasting

Algorithm 1: Calculating the importance scores of the features using active learning

Data: $D = (X, Y)$, a normalization dataset of input features and output variable
Result: $IM = \{IM_1, IM_2, ..., IM_n\}$, a set of importance scores for each feature
Input: k, the number of folds for cross-validation; σ, the standard deviation of Gaussian noise; f, the base forecasting model
Output: IM, the importance scores of features

Initialize U_S and IM as empty sets, and $U_{\tilde{S}} = X$, where U_S represents the selected representative features, and $U_{\tilde{S}}$ represents the not-selected features;
Divide D into k training sets $Tr = \{Tr_1, Tr_2, ..., Tr_k\}$ and corresponding validation sets $Va = \{Va_1, Va_2, ..., Va_k\}$ by k-fold cross-validation;
for i from 1 to n **do**
  Initialize R_i as an empty set to collect results;
  for feature x_{ij} in $U_{\tilde{S}(i)}$ **do**
    Initialize I_{ij} as zero;
    Generate Gaussian noise $\epsilon_{ij} \sim N(0, \sigma^2)$ and add it to feature x_{ij} to get noisy feature x'_{ij};
    for m from 1 to k **do**
      Train model f on Tr_m using the selected feature set U_S, merged separately with the original features x_{ij} and the noisy features x'_{ij};
      Test model f on Tr_m using the selected feature set U_S, merged separately with the original features x_{ij} and the noisy features x'_{ij};
      Calculate errors E_x and $E_{x'}$ using absolute errors or squared errors;
      Update I_{ij} by adding $|E_{x'_{ij}} - E_{x_{ij}}|$;
    end
    Get the average importance I_{ij} of feature x_{ij};
    Add I_{ij} to R_i;
  end
  Determine the maximum value $\tilde{I}_{ij}$ from R_i;
  Update U_S by appending the feature x_{ij} corresponding to $\tilde{I}_{ij}$;
  Update $U_{\tilde{S}}$ by removing the feature x_{ij} corresponding to $\tilde{I}_{ij}$;
  Add $\tilde{I}_{ij}$ to IM;
end
Normalize IM by dividing each element by the sum of all elements;
Return IM;

Algorithm 2: Feature selection base on *IM*

Data: $IM = \{IM_1, IM_2, ..., IM_n\}$, a set of importance scores for each feature
Input: n, the number of features
Output: $\tilde{X}^*$, the reduced feature set

The importance scores $IM = \{IM_1, IM_2, ..., IM_n\}$ of the input features are sorted in descending order, resulting in the importance score $IM' = \{IM'_1, IM'_2, ..., IM'_n\}$ and the feature $\tilde{X} = \{\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n\}$;
Calculate the range of threshold value TV, that is, $TV = [\frac{1}{2n} : \frac{1}{n}, \frac{1}{6n}]$;
L=len(TV);
for j from 1 to L **do**
  if $IM' > TV_j$ **then**
    Add feature $\tilde{x}_i$ to X^*;
  end
  for t from 1 to k **do**
    Train model f using the Tr_j including the feature set X^*;
    Get the prediction results of Va_j using the feature set X^* and model f;
  end
  Calculate the mean prediction errors E^*_j of all validation sets ;
  Add E^*_j to Er;
end
Get the optimal threshold value T and select the features $\tilde{X}^*$ by finding the minimum of Er;
Return $\tilde{X}^*$.

model on the validation sets, given by

$$T^* = \arg\min_{T \in TV} E^*(T), \tag{2}$$

where TV is a range of possible threshold values, and $E^*(T)$ is the mean prediction error of the base forecasting model using the features whose importance scores are above T on the validation sets. The validation sets are generated by k-fold cross-validation.

The FS-IM method consists of four steps:

- **Step 1**: Sort the importance scores in descending order and obtain the corresponding features; According to Section 3.3.1, we get the importance score of all features. The features are sorted in descending order of importance scores.
- **Step 2**: Calculate a range of possible threshold values based on the number of features; The range of threshold value is $[\frac{1}{2n} : \frac{1}{n}, \frac{1}{6n}]$, where n is the number of features and $\frac{1}{6n}$ is the increment value. As the number of features increases, keeping the threshold range constant may result in fewer selected features. In other words, this can lead to the selected features with less information. Therefore, we adjust the threshold range based on the changing number of features. On one hand, the optimal threshold can ensure that the crucial factors for prediction are selected. On the other hand, adjusting the threshold range helps maintain the important information content of the features.
- **Step 3**: Calculate the prediction errors based on the each threshold value. For each threshold value, we select the features whose importance scores are above it to train a base forecasting model and calculate the prediction errors of validation sets. According to k-fold cross validation, we get the mean prediction absolute errors of each threshold value. In the paper, $k = 5$
- **Step 4**: Select the optimal threshold value based on the minimum error of Step 3. We can get the reduced feature set $\tilde{X}^*$ by the optimal threshold value.

The proposed FS-IM has several benefits over existing methods. First, it can optimize the trade-off between the number of features and the forecasting accuracy by minimizing the mean prediction error of the base forecasting model on the validation sets. Second, it can adapt the range of possible threshold values based on the changing number of features, which can help preserve the information content of the features. Third, it can be applied to any base forecasting model and any feature set without any prior assumptions or constraints. Therefore, the proposed FS-IM is a novel and effective method for crude oil return forecasting.

3.4. Dimensionality reduction methods

Dimensionality reduction is a process of reducing the number of features or variables in a dataset while preserving as much information

as possible. Dimensionality reduction can improve the efficiency, performance, and interpretability of predictive models, as well as reduce the risk of overfitting and multicollinearity. Dimensionality reduction methods can be divided into two categories: feature extraction and feature selection. Feature extraction methods transform the original features into a new set of features with lower dimensionality by using some mathematical techniques, such as linear or nonlinear transformations. Feature selection methods select a subset of features from the original set based on some criteria, such as relevance, redundancy, or importance. In this subsection, we compare our proposed feature selection method based on importance measures (FS-IM) with three feature extraction methods: principal component analysis (PCA) [34], kernel principal component analysis (KPCA) [12], and independent component analysis (ICA) [13]. These methods are widely used in crude oil return forecasting and have different advantages and disadvantages.

PCA is a linear transformation method that projects the original features onto a new orthogonal coordinate system, where each coordinate axis is called a principal component (PC). The PCs are ordered by their variances, which reflect the amount of information they contain. PCA aims to find a few PCs that can explain most of the variance in the original data. The advantages of PCA are that it can reduce noise, redundancy, and correlation among features, as well as simplify the data structure and visualization. The disadvantages of PCA are that it may lose some important information that is not captured by variance, such as nonlinear relationships or outliers. Moreover, PCA may not be suitable for forecasting problems that require high accuracy and interpretability, since the PCs are linear combinations of the original features and may not have clear physical meanings.

KPCA is a nonlinear extension of PCA that uses kernel functions to map the original features into a higher-dimensional feature space, where linear PCA is applied. The kernel functions can capture nonlinear relationships among features that are not visible in the original space. KPCA can overcome some limitations of PCA by finding more informative PCs in the kernel space. The advantages of KPCA are that it can handle nonlinear data and improve the prediction performance of linear models. The disadvantages of KPCA are that it may increase the computational complexity and memory requirement due to the kernel matrix calculation, as well as introduce some parameters that need to be tuned, such as the kernel function type and parameters.

ICA is another nonlinear transformation method that assumes that the original features are linear mixtures of some latent independent components (ICs). ICA aims to find a linear transformation that can recover the ICs from the observed features by maximizing their statistical independence. The advantages of ICA are that it can reveal hidden factors or sources that generate the observed data, as well as separate noise or interference from signals. The disadvantages of ICA are that it may not preserve the order or scale of the ICs, which may affect their interpretation and comparison. Moreover, ICA may not be applicable to data that do not satisfy the linear mixture assumption or have more features than observations.

We apply these three feature extraction methods to our dataset of crude oil return and 33 input features, and compare their results with our proposed feature selection method based on importance measures (FS-IM). We use the same base forecasting model, which is ridge regression (RR) [70], to evaluate the prediction accuracy and economic value of the reduced features. RR is a linear regression model with L2 regularization, which can be expressed as follows:

$$\min_{\boldsymbol{w}} \frac{1}{2N} \sum_{i=1}^{N} (\boldsymbol{w}^T \boldsymbol{x}_i - y_i)^2 + \lambda \|\boldsymbol{w}\|_2^2, \tag{3}$$

where $\boldsymbol{x}_i$ is an input feature vector, y_i is an output value, $\boldsymbol{w}$ is a weight vector, N is the number of observations, and λ is a regularization parameter.

3.5. Evaluation metrics

To measure the prediction performance and economic value of our proposed method and other comparison models, we use the following evaluation metrics:

3.5.1. Forecasting performance metrics

We use root mean square forecasting error (RMSE) and mean absolute forecasting error (MAE) to measure the prediction error of the models. They are defined as:

$$\begin{aligned} RMSE &= \sqrt{\frac{1}{N} \sum_{t=1}^{n} \left(\hat{y}_t - y_t\right)^2}, \\ MAE &= \frac{1}{N} \sum_{t=1}^{n} \left|\hat{y}_t - y_t\right|, \end{aligned} \tag{4}$$

where N is the number of observations, $\hat{y}_t$ and y_t are the predicted and the actual values respectively. A lower RMSE or MAE indicates a higher prediction accuracy of the model.

Following Wang et al. [71], we consider the cumulative squared forecast error (CSFE) to examine whether the superiority of proposed model is robust over time, which is defined as

$$CSFE = \frac{1}{N} \sum_{i=1}^{N} \left[\left(y_i - \hat{r}_{C,i}\right)^2 - \left(y_i - \hat{r}_{B,i}\right)^2 \right]. \tag{5}$$

The negative value of CSFE indicates that the competing model is better than the benchmark and has stable superiority over time.

In addition, following Tan et al. [72] and Zhang et al. [4], we use the out-of-sample R^2 (R^2_{OS}) statistic to assess the accuracy of the forecasts. R^2_{OS} are expressed as,

$$R^2_{OS} = \left(1 - \frac{MSE_C}{MSE_B}\right) * 100\%, \tag{6}$$

respectively, where $\hat{r}_i$ denotes the forecasting value of return at time i, MSE_C, and MSE_B represents the out-of-sample mean squared error of the competing and benchmark models respectively. The positive R^2_{OS} means that the competing model has smaller prediction errors than the benchmark model.

We also use two statistical test to compare the prediction performance of different models: Clark West test [4,73] and model confidence set (MCS) [39,74]. The CW test is used to the superiority of the proposed model in statistical significance. The hypothesis of CW test that the competing and benchmark models have the same prediction performance, and the alternative hypothesis is that the competing model has better prediction performance than the benchmark model. The MCS test is used to identify a set of models that are statistically indistinguishable from the best model in terms of predictive ability.

3.5.2. Economic value metrics

We assume that investors allocate assets in crude oil and risk-free assets. Following Zhang et al. [4], we consider that investors pay a portion of the total value of crude oil to meet margin requirements, so investors' gains and losses are magnified by the leverage ratio L. The leverage ratio L is inversely proportional to the margin. Crude oil spot trading also requires margin, so it has a similar leverage effect.

To maximize the investor's utility, in the tth month, the investor should allocate the following share ω_t of total assets to crude oil in the $(t+1)$th month

$$\omega_t = \frac{1}{\gamma} \frac{L(\hat{r}_{t+1} - r^f_{t+1}) + (L-1) r^f_{t+1}}{L^2 \hat{\sigma}^2_{t+1}}, \tag{7}$$

The portfolio return for the $(t+1)$th month is then:

$$R_{p,t+1} = \omega_t L \hat{r}_{t+1} + \left(1 - \omega_t\right) r^f_{t+1}. \tag{8}$$

Table 1
Parameter range of models.

Model	Parameters	Numerical range	Component number
RR	alpha	logspace(−5, 1, 200, base = 3)	200
LASSO	alpha	logspace(−5, 1, 200, base = 3)	200
SVR	C	arange(1.03,2.1,0.05)	110
	γ	[5e−1,1e−2, 5e−2, 1e−3, 5e−3]	
RF	max_depth	[4,8,12,16]	16
	n_estimators	[10,20,40,50]	
GBR	max_depth	[4,8,12,16]	24
	n_estimators	[40,60,80]	
	learning_rate	[0.005,0.015,0.025]	
BR	n_estimators	[8,10,12,14,16,18,20]	7

Note: This table reports the parameter range of models. The *logspace* function is used to construct the geometric progression, and the *arange* is used to construct the arithmetic progression.

We employ the certainty equivalent return (CER) as a utility-based metric to evaluate the economic value of prediction results. The formula for CER is:

$$CER_p = \omega L\hat{r} + (1-\omega)r^f - \frac{1}{2}\gamma\omega^2 L^2\sigma^2. \tag{9}$$

We compare the CER of comparison models with that of the historical average (HA) model and calculate the utility gain (UG) as follows:

$$UG = CER_{p,c} - CER_{p,HA}. \tag{10}$$

$CER_{p,c}$ and $CER_{p,HA}$ are the CER of the comparison and HA model respectively. A higher UG indicates a higher economic value of comparison model compared to HA model.

4. Experiments

4.1. Experimental settings

We conducted our experiments on a computer equipped with an Intel(R) Core(TM) i7-9700F CPU, which is a desktop processor with 8 cores and 8 threads. It has a base frequency of 3.0 GHz and a maximum turbo frequency of 4.7 GHz. We used Python 3.7 for the modeling and prediction process, and Matlab for the evaluation process. We use 5-fold cross validation to determine the parameters of RR, LASSO, SVR, GBR and BR models, and the range of the related parameters is shown in Table 1. The parameters of the MLR and ARMA models can be directly estimated, so the parameter settings of these two models are not shown in Table 1. The epochs of LSTM and GRU, and their neuron number are both 30.

4.2. Data description

We use monthly data of crude oil price and related indicators from January 2000 to December 2021, mainly from the U.S. Energy Information Administration (EIA).[2] We choose West Texas Intermediate (WTI) crude oil spot price as the output variable, as it is one of the most important benchmarks for crude oil prices in the world market [75,76]. We calculate the return of WTI by taking the natural logarithm of the ratio of current price to previous price. We use 34 input features including lag H-order crude oil return, supply, demand, economic, financial, and geopolitical factors. In test set, the input of lag H-order crude oil return is updated by the prediction results. We normalize both the input features and output variable by subtracting their means and dividing by their standard deviations. To prevent information leakage, the training and test sets are standardized respectively. This makes the data more comparable and reduces the influence of outliers or extreme values. We divide the dataset into training sets and validation sets by 5-fold cross-validation. This helps us evaluate the performance and generalization ability of our forecasting models. We also add Gaussian noise to each input feature to evaluate its importance by using our proposed feature selection method based on importance measures. This helps us reduce the dimensionality and complexity of our data.

Table 2 shows the crude oil price and related indicators that we use in our data analysis. The table categorizes the variables into four groups: crude oil return, supply, demand, and macroeconomics. The table also provides the name, meaning, and data source of each variable. Table 3 shows the statistic description of variables that we use in our data analysis. The table provides the mean, standard deviation, median, minimum, maximum, and correlation coefficient of each variable. The table shows that WTI has a positive mean and median, indicating an upward trend over time. It also shows that WTI has a high standard deviation and a large range between the minimum and maximum values, indicating high volatility and uncertainty. It also shows that WTI has a strong positive correlation with CRBSCI (0.852), which is a commodity index, and a strong negative correlation with RDI (−0.376), which is a real dollar index.

To test the robustness of our proposed model, we use different H-step-ahead predictions and different training sets to predict the same test set. Many studies examine the forecasting ability of models in one-step-ahead or multi-step-ahead prediction [77]. Following Lv and Qi [78], we set $H = 1$, 3, and 6. These values correspond to one-month-ahead, one-quarter-ahead, and six-month-ahead forecasting respectively. For the different-step-ahead forecasting experiment, the training set covers January 2000 to December 2017, and the test set covers January 2018 to December 2021. For the training sets experiment, we use three groups of training sets that cover January 2000 to December 2017, January 2001 to December 2017, and January 2002 to December 2017 respectively, and we use the same test set for all groups, i.e., January 2018 to December 2021.

4.3. Prediction result analysis

To verify the effectiveness of the FS-RR model, we establish various models for making predictions and comparisons. We consider both linear and nonlinear models. For linear models, we include RR, Multiple Linear Regression (MLR), LASSO, and Autoregressive Moving Average (ARMA) models. For nonlinear models, we consider Support Vector Regression (SVR), Random Forest (RF), Gradient Boosting Regression (GBR), Bagging Regression (BR), LSTM, and Gated Recurrent Unit (GRU). These models are widely used in financial forecasting studies [4,66,79–81]. Additionally, we also apply PCA, KPCA, and ICA for dimension reduction, which helps to demonstrate the superiority of the proposed feature selection method in dimensionality reduction.

4.3.1. The results of different-step-ahead forecasting

In this section, we build models and make H-step-ahead forecasts based on the training set 1, that is,

$$y_t = f(X_{t-H}), \tag{11}$$

in which, $f(.)$ is the base model, y_t is the output variable at t time, and X_{t-H} is the input variable set at $t-H$. This means that we use H-order lag input variables to establish model.

Table 4 shows the prediction errors in different-step-ahead forecasting. We use RMSE and MAE as the prediction error metrics, as shown in Eq. (4). Smaller RMSE and MAE values indicate smaller differences between predicted and true values, implying a smaller prediction error for the model. When $H = 1$, FS-RR achieves a RMSE of 0.1484 and a MAE of 0.0896, and RR achieves a RMSE of 0.1585 and a MAE of 0.0914. For $H = 3$ and $H = 6$, FS-RR also has smaller RMSE and MAE than RR. This demonstrates that the FS-RR model produces predictions closer to the actual data, thereby improving the accuracy of the RR

[2] https://www.eia.gov/.

Table 2
Crude oil price and related indicators.

Factor group	Variable description	Data source
Crude oil return	WTI: Return of WTI crude oil spot price	EIA
Supply	USPP: Relative change in U.S. (50 States) Petroleum Production	EIA
	OPECPP: Relative change in OPEC Petroleum Production	EIA
	WTPP: Relative change in World Total Petroleum Production	EIA
	NOPECPP: Relative change in Non-OPEC Petroleum Production	EIA
	USPU: Relative change in U.S. Percent Utilization of Refinery Operable Capacity	EIA
	USTCOI: Change in U.S. (50 States) Total Crude Oil Inventory	EIA
	WTCOI: Change in World Total Crude Oil Inventory	EIA
Demand	WTPC: Relative change in World Total Petroleum Consumption	EIA
	USPC: Relative change in U.S. (50 States) Petroleum Consumption	EIA
	OECDPC: Relative change in OECD Petroleum Consumption	EIA
	NOECDPC Relative change in Non-OECD Petroleum Consumption	EIA
	USPCI: Change in US Petroleum Net Imports	EIA
	USPPCI: Change in US Petroleum Product Net Imports	EIA
	USRSP: Relative change in US Raw Steel Production	EIA
Macroeconomics	SER: Relative change in Spot Exchange Rate:USD to CNY	Wind
	MSCIWI: Relative change in MSCI World index	Bloomberg
	RDI: Relative change in Real Dollar Index	Wind
	FFR: Relative change in Federal Funds Rate	Wind
	USTBR: Relative change in U.S. Treasury Bill Rate	Wind
	GDP: Relative change in Real GDP	EIA
	NFER: Relative change in Non-farm Employment Rate	EIA
	IPI: Relative change in Industrial Production Index	EIA
	CPI: Relative change in Consumer Price Index	EIA
	EPPI: Relative change in European 27: PPI	Wind
	USPMI: Relative change in US PMI index	Wind
	EPCI: Relative change in Economic Policy Uncertainty Index	https://www.policyuncertainty.com/
	MPCI: Relative change in Money Policy Uncertainty Index	https://www.policyuncertainty.com/
Finance market	SP500: Relative change in S&P 500	Wind
	VIX: Relative change in VIX index	Wind
	CRBSCI: Relative change in CRB Spot Commodity Index	Wind
	NG: Relative change in Natural gas price	Wind
	GSCINEI: Relative change in S&P GSCI Non-Energy index	Bloomberg
Geopolitics	GRI: Relative change in Geopolitical Risk Index	Wind

Note: This table reports the various factors required for modeling, including their factor group, description, and data source. WTI crude oil return (the change of crude oil price) is as the dependent variable and it is written by $r_t = \log \frac{p_t}{p_{t-1}}$, in which r_t and p_t is crude oil price and return at time t respectively. We take the change of the remaining factors as the independent variable, and the crude oil return with H-order lag is also added to the independent variables.

Table 3
Statistic description of variables.

Variable	Mean	Std	Med	Min	Max	Cor	Variable	Mean	Std	Med	Min	Max	Cor
WTI	0.004	0.106	0.016	−0.568	0.546		RDI	0.000	0.012	0.001	−0.034	0.063	−0.376***
USPP	0.003	0.031	0.003	−0.203	0.158	−0.033	FFR	−0.016	0.222	0.002	−2.554	0.715	0.439***
OPECPP	0.001	0.019	0.001	−0.184	0.059	−0.230***	USTBR	−0.011	0.168	−0.007	−1.444	0.594	0.309***
WTPP	0.001	0.011	0.002	−0.121	0.036	−0.179***	GDP	0.002	0.009	0.000	−0.094	0.073	0.312***
NOPECPP	0.001	0.011	0.002	−0.087	0.051	−0.077	NFER	0.000	0.010	0.001	−0.146	0.033	0.433***
USPU	0.000	0.037	0.001	−0.167	0.147	0.099	IPI	0.000	0.013	0.001	−0.146	0.061	0.447***
USTCOI	0.003	0.720	−0.010	−2.440	2.190	0.137**	CPI	0.002	0.004	0.000	−0.019	0.015	0.445***
WTCOI	0.015	2.173	−0.030	−9.680	15.330	0.282***	EPPI	0.002	0.007	0.002	−0.021	0.049	0.487***
WTPC	0.001	0.022	0.002	−0.119	0.055	0.224***	USPMI	0.000	0.040	−0.002	−0.168	0.199	0.426***
USPC	0.000	0.030	0.002	−0.238	0.100	0.277***	EPCI	0.003	0.269	−0.020	−0.813	1.198	−0.194***
OECDPC	0.000	0.029	0.000	−0.212	0.083	0.238***	MPCI	0.004	0.504	−0.046	−1.345	1.767	−0.097
NOECDPC	0.003	0.021	0.003	−0.087	0.067	0.166***	SP500	0.004	0.044	0.010	−0.186	0.119	0.333***
USPCI	−0.038	0.607	−0.070	−1.660	1.860	0.035	VIX	0.000	0.210	−0.016	−0.614	0.853	−0.330***
USPPI	−0.020	0.391	−0.050	−1.010	1.490	0.095	CRBSCI	0.001	0.044	0.008	−0.223	0.086	0.852***
USRSP	−0.001	0.050	0.003	−0.341	0.211	0.375***	NE	0.002	0.147	−0.006	−0.714	0.680	0.165***
SER	−0.001	0.007	0.000	−0.025	0.038	−0.168***	GSCINE	0.004	0.040	0.003	−0.205	0.120	0.250***
MSCIWI	0.003	0.045	0.011	−0.211	0.119	0.239***	GRI	0.001	0.213	−0.012	−0.509	1.839	−0.049

Note: This table reports the statistic description of variables. Std, Min, and Max are the standard deviation, minimum, and maximum of the dataset. Corr is the correlation coefficient between the independent variable and WTI. "*", "**" and "***" are statistically significant at the 10%, 5%, and 1% levels, respectively.

model through our feature selection method. In comparison with MLR, LASSO, ARMA, SVR, RF, GBR, BR, LSTM, and GRU, the FS-RR model exhibits lower RMSEs and MAEs across different H-step-ahead forecasting scenarios, indicating superior prediction performance. Furthermore, the FS-RR model consistently shows smaller prediction errors than PCA-RR, KPCA-RR, and ICA-RR models, suggesting that our feature selection method is more effective at extracting useful information than PCA, KPCA, and ICA. Consequently, FS-RR consistently achieves the smallest RMSE and MAE among these 14 models in various step-ahead forecasting scenarios, indicating that it is closest to the real data. Based on the above analysis, the proposed feature selection method significantly enhances the prediction accuracy of the RR model. Additionally, the FS-RR model demonstrates superior prediction accuracy compared to RR, MLR, LASSO, ARMA, SVR, RF, GBR, BR, LSTM, and GRU. Moreover, our feature selection method proves to be more effective in information extraction and dimensionality reduction than PCA, KPCA, and ICA.

Table 4
Prediction errors in different-step-ahead forecasting.

Model	H = 1		H = 3		H = 6	
	RMSE(1e−2)	MAE(1e−2)	RMSE(1e−2)	MAE(1e−2)	RMSE(1e−2)	MAE(1e−2)
FS-RR	14.841	8.964	16.197	9.810	16.217	9.818
RR	15.845	9.135	16.550	10.388	16.920	10.685
MLR	15.319	9.810	19.491	12.369	18.979	11.984
LASSO	15.689	9.551	16.232	9.843	16.233	9.866
ARMA	16.284	9.944	16.251	10.103	16.390	10.181
SVR	16.014	9.374	16.298	9.999	16.266	9.843
RF	17.217	9.920	16.657	10.327	16.638	10.802
GBR	16.096	9.766	16.435	9.934	16.273	9.987
BR	16.011	9.560	16.754	10.856	16.397	10.572
LSTM	16.335	10.512	16.908	11.631	22.477	15.312
GRU	17.312	11.569	19.170	12.779	21.476	13.093
PCA-RR	17.196	9.747	16.526	10.009	17.437	11.248
KPCA-RR	15.931	9.606	16.278	9.907	16.349	10.032
ICA-RR	16.264	9.625	16.293	9.859	16.451	10.254

Note: This table reports the prediction errors in different-step-ahead forecasting. RMSE and MAE are shown in Eq. (4).

Table 5 shows the MCS test in different-step-ahead forecasting. We use MSE and MAE as the loss functions for the MCS test. The p-value and ranking are connected on a one-to-one basis, where a higher p-value corresponds to a higher rank. When p-values are the same, a lower loss function leads to a higher rank. A greater p-value or higher rank means a better prediction performance [82]. Using MSE as the loss function, we find that FS-RR has p-value of 1 and rank of 1 in three different H-step-ahead forecasting scenarios ($H = 1$, $H = 3$, and $H = 6$). This means that FS-RR is consistently superior to other models in terms of MSE. When using MAE as the loss function, FS-RR consistently has a p-value of 1 and a rank of 1. Regardless of whether MSE or MAE is used as the loss function, FS-RR consistently has a higher p-value than the other models in different H-step-ahead forecasting scenarios, which demonstrates that FS-RR model outperform RR model in statistical significance. Regardless of using MSE or MAE as the loss function, the rank of FS-RR model is higher than RR, MLR, LASSO, ARMA, SVR, RF, GBR, BR, LSTM and GRU, indicating that FS-RR model has better prediction accuracy. The p-value of FS-RR model is always 1, but the p-values of other models is fluctuant in three different H-step-ahead forecasting. This means that FS-RR model has stable prediction performance. On the other hand, PCA-RR, KPCA-RR, and ICA-RR models do not show consistent or significant improvement over RR. For example, PCA-RR, KPCA-RR, and ICA-RR models do not consistently have higher p-values or ranks than RR model, and PCA-RR model has greater RMMSE and MAE than RR model. This demonstrates that PCA, KPCA, and ICA cannot obviously improve the prediction accuracy of RR model in statistical significance. Therefore, according to MCS test FS-RR has better and more stable prediction performance than other models in statistical significance. Summarizing the finding from Tables 4 and 5, it can be demonstrated that FS-RR model is effective and stable in feature selection and improving prediction accuracy.

Table 6 reports the prediction evaluation of different models in different-step-ahead forecasting. Following Wang et al. [71] and Tan et al. [72], we use CSFE, R^2_{OS} and CW test to further verify the effectiveness and superiority of FS-RR model. The negative CSFEs means that the competing model has less prediction errors. The greater R^2_{OS} and CW statistics means that prediction accuracy of the competing model is more significantly superior to the benchmark model. By taking the FS-RR model as the competing model and the other 13 models as benchmarks, we calculate these three indicators. Taking RR and FS-RR as the benchmark and competing models respectively, CSFE less than 0 and R^2_{OS} greater than 0 indicate that our feature selection method enhance the prediction accuracy of RR model. Moreover, CW test reject the null hypothesis at 10% or better significance level when $H = 3$ and $H = 6$. This indicates that FS-RR can extract effective information and significantly improve the prediction accuracy of RR model. When taking MLR, LASSO, ARMA, SVR, RF, GBR, BR, LSTM, and GRU as the benchmark models, all CSFE values are less than 0 across different-step-ahead prediction, indicating that FS-RR model has excellent and stable forecasting performance over time. All R^2_{OS} values are more than 0 in three H-step-ahead prediction, and most of CW test results reject the null hypothesis, indicating that FS-RR model has better predictive performance. Using PCA-RR, KPCA-RR and ICA-RR models as the benchmark models, the negative CSFE and positive R^2_{OS} values show that FS-RR model has more superior prediction performance. Moreover, most of CW test reject the null hypothesis at 10% or better significance level. These CSFE, R^2_{OS} and CW test imply that our feature selection can extract more effective information than PCA, KPCA and ICA. In contrast to the RR model, the absolute values of CSFE and R^2_{OS} for the PCA-RR, KPCA-RR, and ICA-RR models are not consistently less than those of the RR model, meaning that PCA, KPCA and ICA do not have stable performance in improving the prediction accuracy of RR model. Thus, Table 6 again verifies that PCA-RR, KPCA-RR, and ICA-RR models do not show consistent or significant improvement over RR model. CSFE, R^2_{OS} and CW test also demonstrate the superiority and stabilityn of FS-RR model again.

4.3.2. The prediction results using different training sets

To verify the robustness of FS-RR model, we also employ different training sets to build model. We use three different training sets in this section, such as 2000.1–2017.12, 2001.1–2017.12 and 2002.1–2017.12, to establish model. We set $H = 1$ and predict the same test set based on these three different training sets.

Table 7 shows the prediction errors using different training sets. All RMSEs and MAE of FS-RR model are less than these of other models in three training sets, indicating that its prediction results are closet to the real data. Using three training sets, all RMSEs of FS-RR model are less than 15.4, and their MAEs are less than 9.2. However, all RMSEs of RR model are more than 15.8. Meanwhile, MAEs of FR-RR also exhibit more superior performance than these of RR model. This means that FS-RR model can effectively extract the predictor information and has less prediction errors than RR model. RMSEs and MAEs of FS-RR model are obviously less than these of MLR, LASSO, ARMA, SVR, RF, GBR, BR, LSTM and GRU, indicating the effectiveness and superiority of FS-RR model. RMSEs and MAEs of PCA-RR model are more than RR model using different training sets, suggesting that PCA cannot extract effective feature information. KPCA-RR and ICA-RR do not consistently show less RMSEs and MAEs based on different training set, implying that KPCA and ICA cannot reliably improve the prediction accuracy of RR model. Contrasted with PCA, KPCA and ICA, our feature selection method by extracting the most related features can get more effective predictors to enhance the prediction performance of RR models. Contrasted with MLR, LASSO, ARMA, SVR, RF, GBR, BR, LSTM and GRU, FS-RR model has significant superiority in crude oil return forecasting.

Table 5
MCS test in different-step-ahead forecasting.

Model	Loss function: MSE						Loss function: MAE					
	H = 1		H = 3		H = 6		H = 1		H = 3		H = 6	
	p-value	Rank	*p*-value	Rank	*p*-value	Rank	*p*-value	Rank	*p*-value	Rank	*p*-value	Rank
FS-RR	1.000	1	1.000	1	1.000	1	1.000	1	1.000	1	1.000	1
RR	0.882	4	0.925	5	0.240	13	1.000	2	0.642	9	0.280	9
MLR	0.965	2	0.900	7	0.583	7	0.766	6	0.819	7	0.502	6
LASSO	0.836	6	0.915	6	0.969	4	0.364	10	0.993	4	–	–
ARMA	0.587	13	0.998	3	0.304	11	0.315	13	0.448	12	–	–
SVR	0.805	8	0.860	9	0.832	5	0.159	14	0.484	11	1.000	2
RF	0.688	10	0.517	12	0.462	9	0.842	4	0.643	8	0.150	11
GBR	0.664	11	0.938	4	0.994	3	0.755	7	1.000	2	0.952	3
BR	0.743	9	0.474	13	0.998	2	0.866	3	0.378	13	0.724	4
LSTM	0.916	3	1.000	2	0.279	12	0.767	5	0.959	5	0.209	10
GRU	0.524	14	0.686	11	0.322	10	0.348	11	0.372	14	0.337	8
PCA-RR	0.860	5	0.405	14	0.600	6	0.345	12	0.934	6	0.545	5
KPCA-RR	0.603	12	0.892	8	0.217	14	0.528	9	0.567	10	0.113	12
ICA-RR	0.807	7	0.858	10	0.466	8	0.606	8	1.000	3	0.485	7

Note: This table reports MCS test in different-step-ahead forecasting. In MCS test, the statistics is TR, and its loss functions is mean square error (MSE) and MAE. The greater *p*-value means the better prediction performance of models. '–' means that the model is eliminated at the significant level of 0.1.

Table 6
Prediction evaluation in different-step-ahead prediction.

Model	H = 1			H = 3			H = 6		
	CSFE	R^2_{OS} (%)	CW test	CSFE	R^2_{OS} (%)	CW test	CSFE	R^2_{OS} (%)	CW test
RR	−0.148	12.271	1.140	−0.055	4.218	1.438*	−0.112	8.132	2.609***
MLR	−0.069	6.152	1.411*	−0.564	30.942	1.682**	−0.467	26.987	1.592*
LASSO	−0.124	10.519	1.510*	−0.005	0.424	1.597*	−0.003	0.199	0.811
ARMA	−0.216	16.946	1.497*	−0.008	0.664	1.039	−0.027	2.097	1.490*
SVR	−0.174	14.12	1.590*	−0.016	1.226	0.972	−0.008	0.596	1.052
RF	−0.366	25.703	1.472*	−0.072	5.441	2.181**	−0.066	4.999	2.600***
GBR	−0.186	14.994	2.217**	−0.037	2.87	1.570*	−0.009	0.685	0.971
BR	−0.173	14.083	1.395*	−0.088	6.536	3.103***	−0.028	2.183	1.357*
LSTM	−0.224	17.456	2.010**	−0.113	8.223	2.197**	−1.163	47.945	3.675***
GRU	−0.381	26.514	2.773***	−0.505	28.607	2.653***	−0.952	42.981	3.059***
PCA-RR	−0.362	25.516	1.085	−0.052	3.940	1.814**	−0.197	13.507	1.813**
KPCA-RR	−0.161	13.217	1.667**	−0.013	0.989	1.453*	−0.021	1.613	2.649***
ICA-RR	−0.213	16.741	1.842**	−0.015	1.166	1.242	−0.037	2.828	2.332***

Note: This table reports the prediction evaluation in different-step-ahead forecasting. In CW test, the null hypothesis is that the benchmark and competing model has the same prediction performance, and the alternative that the competing model has better prediction performance. Taking the FS-RF model as the competing model and others as the benchmark models respectively, we can get CSFE, R^2_{OS} and CW test. '***', '**', and '*' indicate the statistical significance level of 1%, 5%, and 10% levels respectively.

Table 8 shows the MCS test using different training sets. We use MSE and MAE as the loss functions for the MCS test. A greater *p*-value or higher rank indicates higher prediction accuracy. Based on three different training sets, the *p*-values and ranks of FS-RR model are always 1 in both MSE and MAE loss functions. The rank of FS-RR model is consistently higher than that of RR model, indicating that FS-RR model outperforms RR model in statistical significance. Using MSE as the loss function, FS-RR model has higher rank than MLR, LASSO, ARMA, SVR, RF, GBR, BR, LSTM and GRU across three training sets. Using MAE as the loss function, FS-RR model also ranks higher than other models. The higher rank of FS-RR model means that it is statistically superior to other models in terms of both MSE and MAE. Taking MAE as the loss function, PCA-RR is eliminated when we use training set 2 and 3 to train model, and RR is always remained in MCS test, which means that PCA cannot extract the useful information from original features. Although KPCA-RR and ICA-RR do not be eliminated based on three training sets, their *p*-values and ranks are less than these of RR model. This imply that KPCA and ICA cannot enhance the prediction performance of RR model in statistical significance. Based on the MCS test, it can be concluded that for different training sets, FS-RR has better prediction performance than other 13 models in statistical significance. This further proves the effectiveness and superiority of our feature selection method, and also confirms the prominent and stable prediction accuracy of the FS-RR model in different training scenarios.

Table 9 shows the prediction evaluation using different training sets. Using the FS-RR model as the competing model and other 13 models as the benchmark respectively, we get the evaluation metrics in Table 9. Using RR and FS-RR as the benchmark and competing models respectively, all CSFEs are less than 0, and R^2_{OS} are more than 0, meaning that FS-RR model outperforms RR model and its prediction results are stable over time for different training scenarios. In addition, most of CW tests reject the null hypothesis at 10% or better significance level, which verifies that FS-RR is significantly superior to RR model in statistical significance. Taking MLR, LASSO, ARMA, SVR, RF, GBR, BR, LSTM, and GRU as the benchmark model, all CSFEs less 0 and R^2_{OS} over 0 mean that the competing model (FS-RR) has smaller prediction errors in different training scenarios. Moreover, the results of CW test are consistent with R^2_{OS}, verifying the excellent prediction accuracy of FS-RR model again. Taking PCA-RR, KPCA-RR, and ICA-RR as the benchmark model, the negative CSFEs and positive R^2_{OS} indicate that FS-RR model has better prediction accuracy. Using three training sets, all absolute values of CSFE and R^2_{OS} of PCA-RR model are more than RR model, which indicates that PCA cannot improve the prediction performance of RR model. Contrasting RR, KPCA-RR, and ICA-RR, KPCA-RR and ICA-RR do not consistently have less absolute values of CSFE and R^2_{OS} than RR model. This implies that KPCA and ICA cannot extract the effective information to enhance the prediction accuracy of RR model. However, the negative CSFE and positive R^2_{OS} mean that FS-RR has better prediction accuracy than PCA-RR, KPCA-RR and ICA-RR. Thus, compared with PCA, KPCA, and ICA, the proposed feature selection can not only reduce the dimension by extracting the

Table 7
Prediction errors using different training sets.

Model	Training set 1		Training set 2		Training set 3	
	RMSE(1e−2)	MAE(1e−2)	RMSE(1e−2)	MAE(1e−2)	RMSE(1e−2)	MAE(1e−2)
FS-RR	14.841	8.964	15.321	9.167	15.236	9.015
RR	15.845	9.135	16.417	9.388	15.965	9.224
MLR	15.319	9.810	15.730	9.631	15.374	9.428
LASSO	15.689	9.551	15.827	9.647	16.191	9.770
ARMA	16.284	9.944	16.229	9.851	16.132	9.796
SVR	16.014	9.374	16.107	9.376	16.039	9.451
RF	17.217	9.920	16.761	9.568	16.438	9.506
GBR	16.096	9.766	16.032	9.743	16.026	9.764
BR	16.011	9.560	17.075	9.931	16.632	9.528
LSTM	16.335	10.512	15.691	11.292	15.691	11.292
GRU	17.312	11.569	15.714	10.620	15.714	10.620
PCA-RR	17.196	9.747	17.022	9.731	16.479	9.536
KPCA-RR	15.931	9.606	15.932	9.607	15.964	9.593
ICA-RR	16.264	9.625	16.193	9.603	16.068	9.539

Note: This table reports the prediction errors in different training sets. RMSE and MAE are shown in (4).

Table 8
MCS test using different training sets.

Model	Loss function: MSE						Loss function: MAE					
	Training set 1		Training set 2		Training set 3		Training set 1		Training set 2		Training set 3	
	p-value	Rank	*p*-value	Rank	*p*-value	Rank	*p*-value	Rank	*p*-value	Rank	*p*-value	Rank
FS-RR	1.000	1	1.000	1	1.000	1	1.000	1	1.000	1	1.000	1
RR	0.882	4	0.859	5	0.874	5	1.000	2	0.999	2	0.986	3
MLR	0.965	2	0.883	4	1.000	3	0.766	6	0.992	3	0.994	2
LASSO	0.836	6	0.855	6	0.840	8	0.364	10	0.466	11	0.328	10
ARMA	0.587	13	0.486	13	0.811	11	0.315	13	0.461	12	0.533	8
SVR	0.805	8	0.763	10	0.750	13	0.159	14	0.933	5	0.318	11
RF	0.688	10	0.784	8	0.854	7	0.842	4	0.990	4	0.931	4
GBR	0.664	11	0.709	12	0.775	12	0.755	7	0.675	9	0.283	12
BR	0.743	9	0.108	14	0.723	14	0.866	3	0.605	10	0.924	5
LSTM	0.916	3	1.000	2	1.000	2	0.767	5	0.862	6	0.737	6
GRU	0.524	14	1.000	3	0.999	4	0.348	11	0.811	7	0.635	7
PCA-RR	0.860	5	0.826	7	0.870	6	0.345	12	–	–	–	–
KPCA-RR	0.603	12	0.756	11	0.813	10	0.528	9	0.447	13	0.262	13
ICA-RR	0.807	7	0.772	9	0.821	9	0.606	8	0.679	8	0.448	9

Note: This table reports MCS test using different training sets. In MCS test, the statistics is TR, and its loss functions is mean square error (MSE) and MAE. The greater *p*-value means the better prediction performance of models. '–' means that the model is eliminated at the significant level of 0.1.

most related and useful factors but also steadily improve the prediction accuracy. Summarizing RMSE, MAE, MCS test, CSFE, R^2_{OS} and CW test, FS-RR model consistently has the smallest predication errors among these 14 models in different training scenarios.

According to Tables 4, 5, 6, 7, 8 and 9, FS-RR always has the best prediction accuracy in different H-step-ahead prediction and different training windows. This verified the prominent prediction and robustness of FS-RR model. Compared to RR model, FS-RR model improves the prediction accuracy, which proves that the proposed method can effectively extract the factors related to crude oil return and improve the prediction accuracy. Compared with other linear and nonlinear models, FS-RR also has better and more stable prediction accuracy. Compared to the models with PCA, KPCA, and ICA, the proposed model can consistently enhance the prediction performance of RR model, which verifies that the proposed method may have certain advantages in information extraction end dealing with the complex relationship of the input and output variables. It can be considered that FS-RR model with robustness can significantly improve the predictive performance of the RR model and our proposed feature selection method is more effective than PCA, KPCA and ICA. Therefore, the proposed model is robust, and it can not only extract the features related to crude oil return forecasting effectively but also capture the complex relationship to get excellent prediction performance.

4.4. Economic value analysis

In Section 4.3, we focus on the prediction performance of models, but it does not reflect their actual significance. In this section, we apply the forecasting results to make investments and calculate the economic values. We assume that the assets are crude oil and risk-free bills, and we find their optimal allocation by using the mean–variance model [83]. We do not consider the borrowing and short sell in the portfolio of crude oil and risk-free bills, that is, $\omega \in [0, 1]$. There is a leverage effect in crude oil futures or spot markets. Following Zhang et al. [4], we set that the margin accounts remain at 20%, that is the leverage ratio is $L = 5$. We set risk aversion coefficients $\gamma = 4$ and 6 to reflect different risk aversion of investors. Following [4,72,84], we compare the economic values of models using the historical average (HA) model as the benchmark model.

Table 10 shows the economic values of different-step-ahead forecasting. We use CER as a utility-based metric to evaluate the economic value of prediction results, as shown in Eq. (9). We also calculate UG as the difference between CER of comparison models and CER of HA model, as shown in Eq. (10). A higher CER indicates that the model has a higher economic value. CER over 0 means that the model can help investors to gain benefits. A higher UG means a higher economic value of the comparison model compared to HA model. As can be seen from Table 10, for $\gamma = 4$, the CERs of all models except MLR, LSTM, and GRU are positive in different-step-ahead forecasting, meaning that these model can help investors to make money. The negative CERs of MLR, LSTM, and GRU means that these model cannot help investors to gain returns in different H-step-ahead forecasting. For $\gamma = 4$, we find that FS-RR, RR, LASSO, ARMA, SVR, RF, BR, KPCA-RR, and ICA-RR models consistently have positive UGs in different-step-ahead predictions, implying that these models have more economic values

Table 9
Prediction evaluation in different training sets.

Indicators	Training set 1			Training set 2			Training set 3		
	CSFE	R^2_{OS} (%)	CW test	CSFE	R^2_{OS} (%)	CW test	CSFE	R^2_{OS} (%)	CW test
RR	−0.148	12.271	1.140	−0.167	12.908	0.958	−0.109	8.927	1.159
MLR	−0.069	6.152	1.411*	−0.061	5.140	1.359*	−0.020	1.787	0.769
LASSO	−0.124	10.519	1.510*	−0.076	6.301	0.931	−0.144	11.452	1.179
ARMA	−0.216	16.946	1.497*	−0.138	10.885	1.271	−0.135	10.799	1.240
SVR	−0.174	14.120	1.590*	−0.119	9.521	1.310*	−0.121	9.763	2.004**
RF	−0.366	25.703	1.472*	−0.222	16.447	1.322*	−0.183	14.089	1.526*
GBR	−0.186	14.994	2.217**	−0.107	8.679	1.981**	−0.119	9.621	1.933**
BR	−0.173	14.083	1.395*	−0.273	19.490	1.507*	−0.214	16.084	1.645*
LSTM	−0.224	17.456	2.010**	−0.055	4.660	2.361***	−0.068	5.713	2.290**
GRU	−0.381	26.514	2.773***	−0.059	4.938	1.466*	−0.071	5.987	1.508*
PCA-RR	−0.362	25.516	1.085	−0.264	18.989	0.965	−0.189	14.523	1.055
KPCA-RR	−0.161	13.217	1.667**	−0.092	7.533	1.313*	−0.109	8.918	1.259
ICA-RR	−0.213	16.741	1.842**	−0.132	10.482	1.734**	−0.125	10.086	1.496*

Note: This table reports the prediction evaluation in different training sets. In CW test, the null hypothesis is that the benchmark and competing model has the same prediction performance, and the alternative that the competing model has better prediction performance. Taking the FS-RF model as the competing model and other models as the benchmark model respectively, we can get CSFE, R^2_{OS} and CW test. '***', '**', and '*' indicate the statistical significance level of 1%, 5%, and 10% levels respectively.

than HA model. The negative UGs of MLR, GBR, LSTM, GRU, and PCA-RR indicate these models gain less returns than HA model. For $\gamma = 6$, the CER of HA model is negative, indicating that HA model does not help investors to achieve returns. Moreover, only FS-RR model, LASSO, ARMA, SVR, KPCA-RR, and ICA-RR model consistently have positive CERs for different-step-ahead prediction, which means that only these models can always earn money. While RR, MLR, RF, GBR, BR, LSTM, GRU, and PCA-RR models with CERs less than 0 may make investors lose money. When $\gamma = 6$, all FS-RR, LASSO, ARMA, SVR, KPCA-RR, and ICA-RR models have positive UGs in different-step-ahead predictions, meaning that they can get more benefits than HA models. Whatever $\gamma = 4$ or $\gamma = 6$, FS-RR model, LASSO model, ARMA model, SVR model, KPCA-RR model, and ICA-RR model always have positive gains (CERs) and UGs in different H-step-ahead predictions. This indicates that they have stable and excellent performance in economic values. Furthermore, for different-step-ahead predictions and different risk aversion coefficients, the CERs and UGs of FS-RR model are always in the top 4 among these models. For instance, when $H = 1$ and $H = 3$, FS-RR model has the highest CERs and UGs for $\gamma = 4$ or $\gamma = 6$. When $H = 6$, the CERs and UGs are always in the top 4. This shows that FS-RR model has outstanding and stable economic values in different-step-ahead predictions.

Table 11 shows the economic values for different training sets. For $\gamma = 4$, all models except LSTM have positive CERs for different training sets, meaning that they consistently help investors earn money. Taking HA as the benchmark model, the UGs of all models except MLR, LSTM, and GRU are consistently over 0 for different training sets, meaning that they have greater CERs than HA model. Moreover, only FS-RR model has CERs and UGs that are always among the top 4 of these models for different training sets. For $\gamma = 6$, the CERs of FS-RR, RR, LASSO, ARMA, SVR, RF, BR, PCA-RR, KPCA-RR, and ICA-RR models are consistently more than 0 for different training scenarios, indicating that these models can help investors to make favorable decisions. While the CERs of MLR, LSTM and GRU models are not always over 0, meaning that they cannot consistently help investors to gain money. In addition, we can conclude the same conclusion as CERs through UGs. Regardless of $\gamma = 4$ or $\gamma = 6$, the CERs and UGs of only FS-RR model are always in the top 4 for different training sets. For example, when $H = 1$, FS-RR model has the highest CERs and UGs for $\gamma = 4$ or $\gamma = 6$. When $H = 3$ and $H = 6$, the CERs and UGs are always in the top 4. These also verify that FS-RR model has more obvious advantages in economic values than other models of Table 11.

For different-step-ahead forecasting, different training sets, and different investors with risk aversion, the economic value of FS-RR model is always positive and in the top 4 among these models. This suggests that FS-RR model has good robustness, and it has high profits in the portfolio, which can help investors to gain income or avoid risks.

5. Discussion

From the experiments, our findings are consistent with some previous studies that have demonstrated the benefits of feature selection for crude oil return forecasting, such as [2,14]. These studies have shown that feature selection can reduce the dimensionality and complexity of the data, enhance the generalization ability and accuracy of the models, and reveal the key factors affecting the oil market dynamics. However, our findings also differ from some previous studies that have used different dimension reduction methods, models, or data sets for crude oil return forecasting, such as [2,15,85]. These studies have reported mixed results on the performance of PCA, KPCA, ICA, LASSO, SVR, RF, GBR, and ARMA for crude oil return forecasting, depending on the data characteristics, model assumptions, and evaluation metrics. One possible reason for the differences is that our method can handle high-dimensional and complex data better than other methods by using our importance measures, which are derived from prediction errors and can reflect the nonlinear and non-Gaussian relationships between the input features and the output variable. Another possible reason is that our method can select the optimal number of features by using a threshold value that minimizes the mean prediction error of the base forecasting model, which can avoid overfitting or underfitting problems. A third possible reason is that our method can use RR as the base forecasting model, which is a simple and robust model that can reduce overfitting and improve generalization by penalizing large coefficients.

Our method has excellent and stable prediction performance in different-step-ahead forecasting scenarios and different training sets. Our method can adapt to different market conditions and data characteristics, and provide reliable forecasts for crude oil returns. Our method can also handle different-step-ahead forecasting horizons. Our method has high economic value for investors who allocate assets in crude oil futures or spot markets. Our method can help investors gain income or avoid risks by providing optimal portfolio allocation based on the forecasted returns. Our method can also provide valuable information for policymakers who are concerned about the impact of crude oil price fluctuations on the economy. Our method proposes a novel feature selection method for neurocomputing that can enhance the prediction performance of AI models for crude oil return forecasting, a challenging and relevant problem for the field. We also cover practical aspects with contributions on advances in software development environments for neurocomputing, such as identifying important features.

Our feature selection method can also be applied to other domains besides energy, such as signal processing and image processing, which are relevant to the field of neurocomputing. For example, our method can be used to select the most informative features from a large set

Table 10
Economical values of different-step-ahead forecasting.

Model	H = 1				H = 3				H = 6			
	CER (%)		UG (%)		CER (%)		UG (%)		CER (%)		UG (%)	
	$\gamma = 4$	$\gamma = 6$	$\gamma = 4$	$\gamma = 6$	$\gamma = 4$	$\gamma = 6$	$\gamma = 4$	$\gamma = 6$	$\gamma = 4$	$\gamma = 6$	$\gamma = 4$	$\gamma = 6$
HA	0.019	−0.009			0.019	−0.009			0.019	−0.009		
FS-RR	0.062	0.033	4.333	4.256	0.044	0.016	2.491	2.541	0.041	0.013	2.212	2.203
RR	0.055	0.014	3.673	2.370	0.020	−0.022	0.151	−1.299	0.044	0.001	2.579	1.031
MLR	0.018	−0.096	−0.057	−8.690	−0.195	−0.423	−21.335	−41.375	−0.083	−0.272	−10.126	−26.242
LASSO	0.045	0.015	2.591	2.479	0.043	0.015	2.430	2.447	0.040	0.012	2.152	2.149
ARMA	0.038	0.010	1.929	1.936	0.042	0.015	2.363	2.414	0.035	0.005	1.651	1.495
SVR	0.056	0.022	3.751	3.139	0.039	0.010	2.003	1.971	0.052	0.024	3.299	3.364
RF	0.041	0.008	2.226	1.738	0.021	−0.012	0.185	−0.237	0.037	0.003	1.810	1.238
GBR	0.030	−0.002	1.161	0.767	0.017	−0.017	−0.188	−0.784	0.038	0.010	1.958	1.951
BR	0.052	0.012	3.285	2.167	0.021	−0.011	0.214	−0.152	0.025	−0.011	0.668	−0.141
LSTM	−0.006	−0.098	−2.457	−8.823	0.016	−0.080	−0.270	−7.029	0.075	−0.024	5.585	−1.433
GRU	0.011	−0.047	−0.785	−3.799	−0.012	−0.096	−3.044	−8.639	0.021	−0.107	0.214	−9.752
PCA-RR	0.062	0.019	4.329	2.811	0.035	0.004	1.603	1.387	0.008	−0.065	−1.041	−5.551
KPCA-RR	0.043	0.014	2.465	2.329	0.038	0.010	1.927	1.955	0.037	0.008	1.815	1.791
ICA-RR	0.039	0.009	2.006	1.894	0.035	0.008	1.623	1.711	0.038	0.007	1.901	1.650

Note: This table reports economical values of different-step-ahead forecasting. CER can be got by Eq. (9), and UG can be got by Eq. (10). Taking HA and other model as the benchmark and competing models, we get UG.

Table 11
Economical value of different training sets.

Model	Training set 1				Training set 2				Training set 3			
	CER (%)		UG (%)		CER (%)		UG (%)		CER (%)		UG (%)	
	$\gamma = 4$	$\gamma = 6$	$\gamma = 4$	$\gamma = 6$	$\gamma = 4$	$\gamma = 6$	$\gamma = 4$	$\gamma = 6$	$\gamma = 4$	$\gamma = 6$	$\gamma = 4$	$\gamma = 6$
HA	0.019	−0.009			0.019	−0.009			0.019	−0.009		
FS-RR	0.062	0.033	4.333	4.256	0.054	0.023	3.503	3.229	0.056	0.023	3.692	3.276
RR	0.055	0.014	3.673	2.370	0.057	0.015	3.834	2.464	0.060	0.020	4.167	2.927
MLR	0.018	−0.096	−0.057	−8.690	0.026	−0.060	0.699	−5.101	0.055	−0.005	3.601	0.441
LASSO	0.045	0.015	2.591	2.479	0.044	0.018	2.544	2.743	0.049	0.021	2.986	3.093
ARMA	0.038	0.010	1.929	1.936	0.033	0.005	1.433	1.486	0.027	0.001	0.821	1.083
SVR	0.056	0.022	3.751	3.139	0.056	0.022	3.687	3.111	0.055	0.024	3.662	3.310
RF	0.041	0.008	2.226	1.738	0.049	0.016	3.016	2.589	0.048	0.018	2.952	2.794
GBR	0.030	−0.002	1.161	0.767	0.033	0.001	1.458	1.059	0.048	0.015	2.900	2.405
BR	0.052	0.012	3.285	2.167	0.047	0.011	2.798	2.012	0.092	0.058	7.349	6.751
LSTM	−0.006	−0.098	−2.457	−8.823	0.018	−0.071	−0.051	−6.199	0.018	−0.071	−0.051	−6.199
GRU	0.011	−0.047	−0.785	−3.799	0.026	−0.039	0.744	−2.919	0.026	−0.039	0.744	−2.919
PCA-RR	0.062	0.019	4.329	2.811	0.060	0.016	4.173	2.540	0.064	0.023	4.508	3.204
KPCA-RR	0.043	0.014	2.465	2.329	0.043	0.014	2.470	2.333	0.048	0.018	2.886	2.796
ICA-RR	0.039	0.009	2.006	1.894	0.039	0.010	2.048	1.955	0.045	0.016	2.619	2.528

Note: This table reports economical values of different-step-ahead forecasting. CER can be got by Eq. (9), and UG can be got by Eq. (10). Taking HA and other model as the benchmark and competing models, we get UG.

of signals or images, and improve the performance of neural networks or other AI models for classification, segmentation, or recognition tasks. Some references that have used feature selection methods for signal processing and image processing are [65,86,87]. Our feature selection method offers a novel perspective in feature selection by integrating principles from neurocomputing, such as active learning and Gaussian noise perturbation, which are inspired by the biological neural network modeling and machine learning. Active learning is a learning strategy that selects the most informative samples to train a model, which mimics the human brain's ability to focus on the most relevant information and ignore the irrelevant ones. Gaussian noise perturbation is a technique that adds random noise to the input data, which simulates the biological neural network's robustness to noise and uncertainty. These principles can enhance the prediction performance of AI models by capturing the complex and nonlinear relationships between the features and the output variable, which are essential for forecasting. Our feature selection method is not only a data processing technique, but also a crucial step in the development of efficient and effective AI models. By reducing data dimensionality and complexity, our method improves the computational efficiency and performance of AI models, which is a key concern in the field of neurocomputing. Our method can also help to avoid the curse of dimensionality, which is a problem that occurs when the data has too many features and the model becomes overfitted or underfitted. Our method can also help to improve the interpretability and explainability of AI models, which is a challenge that arises when the models are too complex and opaque.

Despite the promising results of our study, we acknowledge that there are some limitations that may affect the interpretations and implications of our findings. One limitation is the data quality, as we used monthly data from different sources that may have different definitions, measurements, and accuracies, which may affect the feature selection and forecasting performance. Another limitation is the model assumptions, as we assumed that the relationship between the input features and the output variable is linear and additive in RR model, which may not capture the complex and nonlinear dynamics of the oil market. A third limitation is that we implement static feature selection rather than dynamic feature selection. We select the same features in both the training set and the test set according to FS-IM, which cannot reflect the time-varying effects of the features on the prediction. These limitations may affect the accuracy, validity, and causality of our results. For example, the data quality may introduce some biases or errors in the feature selection and forecasting process, which may reduce the confidence and precision of our results. The model assumptions may limit the explanatory power and the predictive power of our results, as they may not account for the nonlinear and non-Gaussian features of the data. Over time, the characteristics and distribution of the data may change. Static feature selection cannot adapt to this change, and then affect the prediction accuracy. To

address or mitigate these limitations, we suggest some possible directions for future research. First, we recommend using more reliable and consistent data sources to improve the data quality. Second, we suggest relaxing the model assumptions by using some nonlinear and non-Gaussian models, such as neural networks, support vector machines, or Gaussian mixture models, to capture the complex and nonlinear relationships between the input features and the output variable. Third, a new dynamic feature selection strategy should be proposed based on FS-IM to quantify the varying influence.

6. Conclusions

This paper presented a novel feature selection method based on importance measures (FS-IM) to enhance the forecasting of crude oil returns. FS-IM innovatively combined active learning with Gaussian noise distribution to assess the importance of each feature by the change in the prediction errors of a ridge regression (RR) model. FS-IM also determined the optimal number of features by using a threshold value that minimized the mean prediction error of the RR model. We applied FS-IM to a West Texas Intermediate crude oil spot price dataset, which consisted of 34 input features and one output variable. We compared FS-RR model with PCA-RR, KPCA-RR, and ICA-RR model using RMSE, MAE, CSFE, R^2_{OS}, and statistical tests (MCS and CW tests). The results show that FS-RR model has much better prediction accuracy for different training and prediction windows and more stable economic value for different investors and policymakers. Our findings verified that FS-IM could also extract more useful information than principal component analysis, kernel principal component analysis, and independent component analysis and could capture the complex relationships between the features and the output variable, which are essential for forecasting. Compared with RR model, FS-RR model has evident superiority in prediction accuracy and economic profits. Our finding demonstrated that FS-IM could effectively select the most relevant features from a large feature set and improve the forecasting accuracy of the RR model. Moreover, we also compared FS-RR model with other prediction models, such as MLR, LASSO, ARMA, SVR, RF, GBR, BR, LSTM, and GRU using different-step-ahead forecasting, different training sets and various evaluation metrics. The superiority and robustness of FS-RR model are also proved. According to the comparative analysis, we confirm its effectiveness in feature selection and its merit in improving prediction performance and economic values.

CRediT authorship contribution statement

Yuan Zhao: Conceptualization, Data curation, Methodology, Software, Validation, Writing – original draft, Writing – review & editing. **Yaohui Huang:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Zhijin Wang:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Xiufeng Liu:** Conceptualization, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

We thank the editors and anonymous reviewers for their valuable comments and suggestions, which greatly benefited this paper. This work was supported by the National Social Science Foundation of China (Grant No. 20BJY246) and the Fundamental Research Funds for the Central Universities (Grant No. 30923011034).

References

[1] Y. Zhao, W. Zhang, X. Gong, C. Wang, A novel method for online real-time forecasting of crude oil price, Appl. Energy 303 (2021) 117588.
[2] C. Zhang, D. Pan, M. Yang, Z. Pu, et al., A lead-lag relationship and forecast research between China's crude oil futures and spot markets, Complexity 2022 (2022).
[3] M. He, Y. Zhang, D. Wen, Y. Wang, Forecasting crude oil prices: A scaled PCA approach, Energy Econ. 97 (2021) 105189.
[4] Y. Zhang, F. Ma, Y. Wang, Forecasting crude oil prices with a large set of predictors: Can LASSO select powerful predictors? J. Empir. Financ. 54 (2019) 97–117.
[5] Y. Zhang, Y. Wang, Forecasting crude oil futures market returns: A principal component analysis combination approach, Int. J. Forecast. 39 (2) (2023) 659–673.
[6] Y. Yang, J. Guo, S. Sun, Y. Li, Forecasting crude oil price with a new hybrid approach and multi-source data, Eng. Appl. Artif. Intell. 101 (2021) 104217.
[7] Y. Song, M. He, Y. Wang, Y. Zhang, Forecasting crude oil prices: A reduced-rank approach, Int. Rev. Econ. Finance 88 (2023) 698–711.
[8] J. Manickavasagam, S. Visalakshmi, N. Apergis, A novel hybrid approach to forecast crude oil futures using intraday data, Technol. Forecast. Soc. Change 158 (2020) 120126.
[9] J. Zhao, Exploring the influence of the main factors on the crude oil price volatility: An analysis based on GARCH-MIDAS model with Lasso approach, Resour. Policy 79 (2022) 103031.
[10] Q. Lu, S. Sun, H. Duan, S. Wang, Analysis and forecasting of crude oil price based on the variable selection-LSTM integrated model, Energy Inform. 4 (2) (2021) 1–20.
[11] G. Hafeez, K.S. Alimgeer, A.B. Qazi, I. Khan, M. Usman, F.A. Khan, Z. Wadud, A hybrid approach for energy consumption forecasting with a new feature engineering and optimization framework in smart grid, IEEE Access 8 (2020) 96210–96226.
[12] J.-M. Lee, C. Yoo, S.W. Choi, P.A. Vanrolleghem, I.-B. Lee, Nonlinear process monitoring using kernel principal component analysis, Chem. Eng. Sci. 59 (1) (2004) 223–234.
[13] J.V. Stone, Independent component analysis: an introduction, Trends Cogn. Sci. 6 (2) (2002) 59–64.
[14] P. Bahrami, F. Sahari Moghaddam, L.A. James, A review of proxy modeling highlighting applications for reservoir engineering, Energies 15 (14) (2022) 5247.
[15] J. Amboko, K. Ocharo, Global Oil Price Volatility and Foreign Direct Investment Inflows in Kenya (Ph.D. thesis), KENYATTA UNIVERSITY, 2021.
[16] X. Hao, Y. Zhao, Y. Wang, Forecasting the real prices of crude oil using robust regression models with regularization constraints, Energy Econ. 86 (2020) 104683.
[17] J. Wang, H. Zhou, T. Hong, X. Li, S. Wang, A multi-granularity heterogeneous combination approach to crude oil price forecasting, Energy Econ. 91 (2020) 104790.
[18] K. He, C. Xie, S. Chen, K.K. Lai, Estimating VaR in crude oil market: A novel multi-scale non-linear ensemble approach incorporating wavelet analysis and neural network, Neurocomputing 72 (16–18) (2009) 3428–3438.
[19] R. Bollapragada, A. Mankude, V. Udayabhanu, Forecasting the price of crude oil, Decision 48 (2) (2021) 207–231.
[20] N. Nonejad, The price of crude oil and (conditional) out-of-sample predictability of world industrial production, J. Commod. Mark. 23 (2021) 100167.
[21] A. Sarmah, D.P. Bal, Does crude oil price affect the inflation rate and economic growth in India? A new insight based on structural VAR framework, Indian Econ. J. 69 (1) (2021) 123–139.
[22] L. Liu, Q. Geng, Y. Zhang, Y. Wang, Investors' perspective on forecasting crude oil return volatility: Where do we stand today? J. Manage. Sci. Eng. 7 (3) (2022) 423–438.
[23] Y. Wei, S. Qin, X. Li, S. Zhu, G. Wei, Oil price fluctuation, stock market and macroeconomic fundamentals: Evidence from China before and after the financial crisis, Finance Res. Lett. 30 (2019) 23–29.
[24] B. Anand, S. Paul, Oil shocks and stock market: Revisiting the dynamics, Energy Econ. 96 (2021) 105111.
[25] B. Abuzayed, N. Al-Fayoumi, Risk spillover from crude oil prices to GCC stock market returns: New evidence during the COVID-19 outbreak, N. Am. J. Econ. Financ. 58 (2021) 101476.
[26] J. Bouoiyour, R. Selmi, S. Hammoudeh, M.E. Wohar, What are the categories of geopolitical risks that could drive oil prices higher? Acts or threats? Energy Econ. 84 (2019) 104523.
[27] Z. Zhang, Y. Wang, J. Xiao, Y. Zhang, Not all geopolitical shocks are alike: Identifying price dynamics in the crude oil market under tensions, Resour. Policy 80 (2023) 103238.
[28] N. Kourentzes, F. Petropoulos, J.R. Trapero, Improving forecasting by estimating time series structural components across multiple frequencies, Int. J. Forecast. 30 (2) (2014) 291–302.
[29] Y.-Y. Zhang, Y.-J. Zhang, The impact of institutional analyst forecast divergence on crude oil market: Evidence from the mixed frequency models, Int. Rev. Financ. Anal. 84 (2022) 102418.

[30] A. Dutta, E. Bouri, D. Roubaud, Modelling the volatility of crude oil returns: Jumps and volatility forecasts, Int. J. Finance Econ. 26 (1) (2021) 889–897.

[31] X. Li, Y. Wei, X. Chen, F. Ma, C. Liang, W. Chen, Which uncertainty is powerful to forecast crude oil market volatility? New evidence, Int. J. Finance Econ. 27 (4) (2022) 4279–4297.

[32] Q. Han, M. He, Y. Zhang, M. Umar, Default return spread: A powerful predictor of crude oil price returns, J. Forecast. (2023).

[33] X. Yan, J. Bai, X. Li, Z. Chen, Can dimensional reduction technology make better use of the information of uncertainty indices when predicting volatility of Chinese crude oil futures? Resour. Policy 75 (2022) 102521.

[34] R. Bro, A.K. Smilde, Principal component analysis, Anal. Methods 6 (9) (2014) 2812–2831.

[35] Y. Yifan, G. Ju'e, S. Shaolong, L. Yixin, A new hybrid approach for crude oil price forecasting: Evidence from multi-scale data, 2020, arXiv preprint arXiv: 2002.09656.

[36] I. Ghosh, T.D. Chaudhuri, E. Alfaro-Cortés, M. Gámez, N. García, A hybrid approach to forecasting futures prices with simultaneous consideration of optimality in ensemble feature selection and advanced artificial intelligence, Technol. Forecast. Soc. Change 181 (2022) 121757.

[37] L. Fan, S. Pan, Z. Li, H. Li, An ICA-based support vector regression scheme for forecasting crude oil prices, Technol. Forecast. Soc. Change 112 (2016) 245–253.

[38] J. Henríquez, W. Kristjanpoller, A combined Independent Component Analysis–Neural Network model for forecasting exchange rate variation, Appl. Soft Comput. 83 (2019) 105654.

[39] A. Bucci, L. He, Z. Liu, Combining dimensionality reduction methods with neural networks for realized volatility forecasting, Ann. Oper. Res. (2023) 1–29.

[40] X. Li, C. Liang, Z. Chen, M. Umar, Forecasting crude oil volatility with uncertainty indicators: New evidence, Energy Econ. 108 (2022) 105936.

[41] A.B.R. Costa, P.C.G. Ferreira, W.P. Gaglianone, O.T.C. Guillén, J.V. Issler, Y. Lin, Machine learning and oil price point and density forecasting, Energy Econ. 102 (2021) 105494.

[42] Q. Qin, Z. Huang, Z. Zhou, C. Chen, R. Liu, Crude oil price forecasting with machine learning and Google search data: An accuracy comparison of single-model versus multiple-model, Eng. Appl. Artif. Intell. 123 (2023) 106266.

[43] G. Lai, W.-C. Chang, Y. Yang, H. Liu, Modeling long-and short-term temporal patterns with deep neural networks, in: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, 2018, pp. 95–104.

[44] S. Siami-Namini, N. Tavakoli, A.S. Namin, A comparison of ARIMA and LSTM in forecasting time series, in: 2018 17th IEEE International Conference on Machine Learning and Applications, ICMLA, IEEE, 2018, pp. 1394–1401.

[45] S.M.A. Sajadi, P. Khodaee, E. Hajizadeh, S. Farhadi, S. Dastgoshade, B. Du, Deep learning-based methods for forecasting brent crude oil return considering COVID-19 pandemic effect, Energies 15 (21) (2022) 8124.

[46] H. Zhu, S. Zhou, M. Wang, Forecasting short-term crude oil price movements using futures market dynamics: Crude oil forecasting with futures market properties, in: Proceedings of the 2022 13th International Conference on E-Education, E-Business, E-Management, and E-Learning, 2022, pp. 542–554.

[47] S. Ramyar, F. Kianfar, Forecasting crude oil prices: A comparison between artificial neural networks and vector autoregressive models, Comput. Econ. 53 (2019) 743–761.

[48] G.E. Box, G.M. Jenkins, G.C. Reinsel, G.M. Ljung, Time Series Analysis: Forecasting and Control, John Wiley & Sons, 2015.

[49] C.-H. Su, C.-H. Cheng, A hybrid fuzzy time series model based on ANFIS and integrated nonlinear feature selection method for forecasting stock, Neurocomputing 205 (2016) 264–273.

[50] M.A. Hall, Correlation-Based Feature Selection for Machine Learning (Ph.D. thesis), The University of Waikato, 1999.

[51] J.R. Vergara, P.A. Estévez, A review of feature selection methods based on mutual information, Neural Comput. Appl. 24 (2014) 175–186.

[52] Y. Zhai, W. Song, X. Liu, L. Liu, X. Zhao, A chi-square statistics based feature selection method in text classification, in: 2018 IEEE 9th International Conference on Software Engineering and Service Science, ICSESS, IEEE, 2018, pp. 160–163.

[53] K. Mei, M. Tan, Z. Yang, S. Shi, Modeling of feature selection based on random forest algorithm and pearson correlation coefficient, in: Journal of Physics: Conference Series, vol. 2219, IOP Publishing, 2022, 012046.

[54] H.Y. Song, S. Park, An analysis of correlation between personality and visiting place using Spearman's rank correlation coefficient, KSII Trans. Internet Inf. Syst. (TIIS) 14 (5) (2020) 1951–1966.

[55] F. Wang, J. Ai, Z. Zou, A cluster-based hybrid feature selection method for defect prediction, in: 2019 IEEE 19th International Conference on Software Quality, Reliability and Security, QRS, IEEE, 2019, pp. 1–9.

[56] H. Tan, G. Wang, W. Wang, Z. Zhang, Feature selection based on distance correlation: a filter algorithm, J. Appl. Stat. 49 (2) (2022) 411–426.

[57] M. Prasad, R.K. Gupta, S. Tripathi, A multi-level correlation-based feature selection for intrusion detection, Arab. J. Sci. Eng. 47 (8) (2022) 10719–10729.

[58] T. Xu, Z.-H. Feng, X.-J. Wu, J. Kittler, Learning adaptive discriminative correlation filters via temporal consistency preserving spatial feature selection for robust visual object tracking, IEEE Trans. Image Process. 28 (11) (2019) 5596–5609.

[59] X. Gu, J. Guo, L. Xiao, C. Li, Conditional mutual information-based feature selection algorithm for maximal relevance minimal redundancy, Appl. Intell. 52 (2) (2022) 1436–1447.

[60] U.F. Siddiqi, S.M. Sait, O. Kaynak, Genetic algorithm for the mutual information-based feature selection in univariate time series data, IEEE Access 8 (2020) 9597–9609.

[61] Z. Sun, J. Zhang, L. Dai, C. Li, C. Zhou, J. Xin, S. Li, Mutual information based multi-label feature selection via constrained convex optimization, Neurocomputing 329 (2019) 447–456.

[62] R. Sawhney, R.R. Shah, V. Bhatia, C.-T. Lin, S. Aggarwal, M. Prasad, Exploring the impact of evolutionary computing based feature selection in suicidal ideation detection, in: 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), IEEE, 2019, pp. 1–6.

[63] I.S. Thaseen, C.A. Kumar, Intrusion detection model using fusion of chi-square feature selection and multi class SVM, J. King Saud Univ.-Comput. Inf. Sci. 29 (4) (2017) 462–472.

[64] C. Yan, B. Wu, J. Ma, G. Zhang, J. Luo, J. Wang, H. Luo, A novel hybrid filter/wrapper feature selection approach based on improved fruit fly optimization algorithm and chi-square test for high dimensional microarray data, Curr. Bioinform. 16 (1) (2021) 63–79.

[65] Y. Wang, C. Zhou, Feature selection method based on chi-square test and minimum redundancy, in: Emerging Trends in Intelligent and Interactive Systems and Applications: Proceedings of the 5th International Conference on Intelligent, Interactive Systems and Applications (IISA2020), Springer, 2021, pp. 171–178.

[66] S. Gu, B. Kelly, D. Xiu, Empirical asset pricing via machine learning, Rev. Financ. Stud. 33 (5) (2020) 2223–2273.

[67] R. Genuer, J.-M. Poggi, C. Tuleau-Malot, Variable selection using random forests, Pattern Recognit. Lett. 31 (14) (2010) 2225–2236.

[68] G. Xu, P. Schwarz, H. Yang, Determining China's CO2 emissions peak with a dynamic nonlinear artificial neural network approach and scenario analysis, Energy Policy 128 (2019) 752–762.

[69] Y.-C. Chen, W.-C. Huang, Constructing a stock-price forecast CNN model with gold and crude oil indicators, Appl. Soft Comput. 112 (2021) 107760.

[70] G.C. McDonald, Ridge regression, Wiley Interdiscip. Rev. Comput. Stat. 1 (1) (2009) 93–100.

[71] J. Wang, X. Guo, X. Tan, J. Chevallier, F. Ma, Which exogenous driver is informative in forecasting European carbon volatility: Bond, commodity, stock or uncertainty? Energy Econ. 117 (2023) 106419.

[72] X. Tan, K. Sirichand, A. Vivian, X. Wang, Forecasting European carbon returns using dimension reduction techniques: Commodity versus financial fundamentals, Int. J. Forecast. 38 (3) (2022) 944–969.

[73] X. Gong, W. Zhang, J. Wang, C. Wang, Investor sentiment and stock volatility: new evidence, International Review of Financial Analysis 80 (2022) 102028.

[74] X. Gong, X. Ye, W. Zhang, Y. Zhang, Predicting energy futures high-frequency volatility using technical indicators: the role of interaction, Energy Economics 119 (2023) 106533.

[75] C. Baumeister, L. Kilian, What central bankers need to know about forecasting oil prices, Internat. Econom. Rev. 55 (3) (2014) 869–889.

[76] C. Baumeister, L. Kilian, Real-time forecasts of the real price of oil, J. Bus. Econom. Statist. 30 (2) (2012) 326–336.

[77] T. Zhang, Z. Tang, J. Wu, X. Du, K. Chen, Multi-step-ahead crude oil price forecasting based on two-layer decomposition technique and extreme learning machine optimized by the particle swarm optimization algorithm, Energy 229 (2021) 120797.

[78] W. Lv, J. Qi, Stock market return predictability: A combination forecast perspective, Int. Rev. Financ. Anal. 84 (2022) 102376.

[79] H.W. Teng, Y.-H. Li, S.-W. Chang, Machine learning in empirical asset pricing models, in: 2020 International Conference on Pervasive Artificial Intelligence, ICPAI, IEEE, 2020, pp. 123–129.

[80] M. Bagnara, Asset pricing and machine learning: A critical review, J. Econ. Surv. (2022).

[81] O.B. Sezer, M.U. Gudelek, A.M. Ozbayoglu, Financial time series forecasting with deep learning : A systematic literature review: 2005–2019, Appl. Soft Comput. 90 (2020) 106181.

[82] Y. Liang, Y. Lin, Q. Lu, Forecasting gold price using a novel hybrid model with ICEEMDAN and LSTM-CNN-CBAM, Expert Syst. Appl. 206 (2022) 117847.

[83] C.J. Neely, D.E. Rapach, J. Tu, G. Zhou, Forecasting the equity risk premium: the role of technical indicators, Manage. Sci. 60 (7) (2014) 1772–1791.

[84] J.Y. Campbell, S.B. Thompson, Predicting excess stock returns out of sample: Can anything beat the historical average? Rev. Financ. Stud. 21 (4) (2008) 1509–1531.

[85] Z.-G.Z. Lu-Tao Zhao, Oil price forecasting using a time-varying approach, Energies 13 (6) (2020) 1403, http://dx.doi.org/10.3390/en13061403.

[86] S. Mahajan, A.K. Pandit, Hybrid method to supervise feature selection using signal processing and complex algebra techniques, Multimedia Tools Appl. (2021) 1–22.

[87] B. Chen, L. Chen, Y. Chen, Efficient ant colony optimization for image feature selection, Signal Process. 93 (6) (2013) 1566–1576.

Yuan Zhao is a Ph.D. candidate in Management Science and Engineering with the School of Business Administration, South China University of Technology, Guangzhou, China. Her research mainly focuses on machine learning, time series analysis and forecasting methods. He has published in top-tiered journals, including Applied Energy, IEEE Transactions on Sustainable Energy, Energy Conversion and Management, Renewable Energy.

Yaohui Huang is a master student with the College of Electronic Information, Guangxi Minzu University, Nanning, China. His research interests include time series analysis, load forecasting, and graph neural networks.

Zhijin Wang is an associate professor with the College of Computer Engineering, Jimei University, Xiamen, China. He received the Ph.D. degree from the Department of Computer Science and Technology, East China Normal University, Shanghai, China, in 2016. His research interests include recommendation systems, time series forecasting, and artificial intelligence in health/healthcare.

Xiufeng Liu is a Senior Researcher with the Department of Technology, Management and Economics, Technical University of Denmark. He received the Ph.D. degree in computer science from Aalborg University, Denmark, in 2012, and did post-doctoralresearch with the University of Waterloo, Canada, from 2013 to 2014. His research interests include smart meter data analysis, data warehousing, energy informatics, and big data.