ARTICLE TEMPLATE

Detecting Market Corners in Crude Oil Futures: A Robust Hybrid Anomaly Detection Approach

Dun Li[a], Noel Crespi[a], Roberto Minerva[a], Reza Farahbakhsh[a], Kuan-Ching Li[b], and Meng-Yen Hsieh[b]

[a]Samovar, Telecom SudParis, Institut Polytechnique de Paris, 91120 Palaiseau, France;
[b]Dept. of Computer Science and Information Engineering, Providence University, Taiwan

ARTICLE HISTORY
Compiled April 18, 2024

ABSTRACT
The price of crude oil is one of the most critical factors of the world economy, as it is volatile and sensibly affected by the macro-economic, thus attracting large-scale speculative activities. The vulnerability of the West Texas Intermediate (WTI) crude oil market, influenced by external events such as financial crises and oil trade wars and potentially by manipulative practices, underscores the need for robust market corner risk detection. This study aims to elucidate the characteristics of market cornering in the crude oil futures market and to identify latent market corner risks within WTI. To achieve this, we introduce a hybrid anomaly detection approach. The research begins with a comprehensive analysis of market corner characteristics and definitions, followed by the extraction of pertinent features from actual market data. Utilizing the Local Outlier Factor (LOF) algorithm, we initially identify market corners exhibiting anomalous pricing and trading volumes, indicative of potential market manipulation. Next, the detected results are used as pseudo-labels, and the entire month's trading behaviour is trained and classified through the Support Vector Machine (SVM) and Random Forest (RF) algorithms to identify potential market corners. Experimental results show that the proposed model has excellent accuracy, precision, recall, and F-score, indicating that the model is feasible and has strong robustness. Furthermore, based on the successful detection of potential market corner risk, the model can be further used for individual risk control and overall supervision of the crude oil futures market.

KEYWORDS
Crude Oil; WTI; Price Manipulation; Anomaly Detection; Hybrid Algorithm

1. Introduction

West Texas Intermediate (WTI) is one of the crucial crude oil price variables and economic indexes in today's economy, as its price level is closely correlated with energy markets worldwide, providing trading opportunities under almost all market conditions. The drastic fluctuation of crude oil prices has always been a significant problem for economists and investors [1,2]. The primary daily price trend is shown in Fig 1(a), and it can be seen that, due to the financial crisis in 2008, together with other commodity economies and indexes, the price of WTI sharply fell [3]. According to the closing price data of WTI, the corresponding median X_M, upper quartile Q_1 and

Corresponding Author Meng-Yen Hsieh. Email: mengyen@pu.edu.tw

lower quartile Q_3 are calculated, as shown in the box chart in Fig 1(b). In 2015, due to the breakthrough of shale oil and gas technology in the US [4], the traditional crude oil market suffered a small impact, also caused price fluctuations and a substantial drop [5].

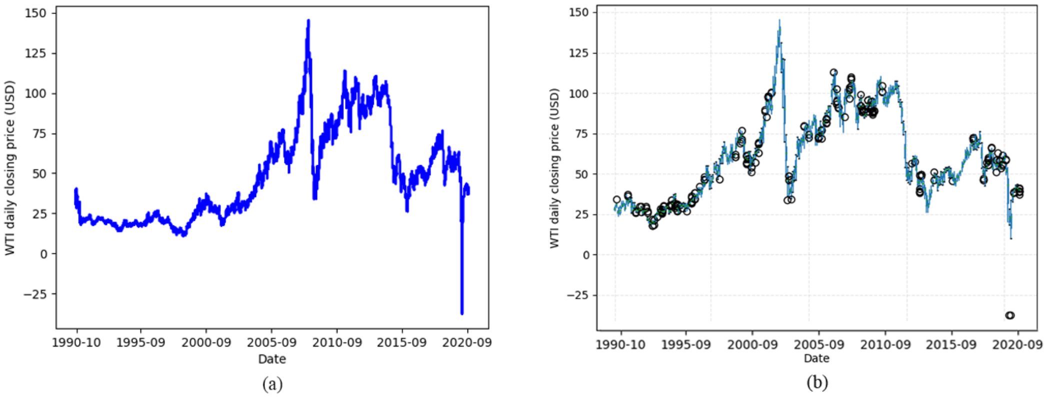

Figure 1. WTI daily price.

Inevitably, the impact of these relevant events on crude oil prices is systemic. However, other than the significant impact caused by black swan events, futures trading has a large number of market corner behaviours, making the price critically deviate from the market in a very short period [6,7], contrary to the practical market theory. A large portion of these deviations is caused by the delivery date effect and the corresponding market corner risk. Because of the market corner, it tends to be inevitable that the futures price fluctuates suddenly and abruptly [8]. Furthermore, the rising complexity of financial transactions involving oil prices and financial products, along with the increased frequency of financial transactions between financial products and commodities, has led to a notable escalation in corresponding financial risks. For instance, between 2009 and 2012, the Houston subsidiary of Total Gas & Power North America, and Western Electric Corporation engaged in uneconomical transactions to manipulate real market prices on multiple occasions [9]. Additionally, JPMorgan faced a \$285 million fine for similar misconduct [10]. These instances underscore the heightened financial risks in the crude oil futures market.

Notably, there are always financial risks in the crude oil futures market. whenever the expiration day of a futures contract is approaching, the long and short parties involved in futures trading utilize various methods to impose on futures and even spot prices and obtain a contract delivery price that is beneficial to themselves. The spot index critically determines the final delivery price in stock index futures, leading many funds to enter the spot market and cause substantial index volatility, known as the expiration-day effect. The market corner, essentially a form of trading manipulation by either short or long positions, is the primary driver behind these effects. Although these trading orders generally comply with current market regulations, their deliberate, speculative nature geared towards profit can lead to significant negative repercussions [11]. In light of these dynamics, the implementation of an effective risk management framework is imperative [12].

To date, many scholars have begun to notice the negative impact brought by the fluctuation of crude oil futures prices. For example, Herrera *et al.* [13] expressed that, from the analysis of the crude oil price, the application of the machine learning algorithm is superior and precise if compared to the traditional econometric method. The anomaly detection algorithm in machine learning has been widely used in the stock market and cryptocurrency market to detect abnormal price trends and identify price manipulation risk [14–19]. To ensure the proper functioning of the market, it is an efficient and challenging topic to enhance the identification of excessive market corner risks. The most formidable challenge associated with this subject matter lies in the presence of consistent market corner patterns and characteristics, despite the absence of readily accessible labelled datasets [20–39].

Motivated by the abovementioned issues, this article proposes a hybrid anomaly detection approach based on the combination of Local Outlier Factor (LOF), Support Vector Machine (SVM), and Random Forest (RF) algorithms. First, the outliers are separated and marked to solve the no-label problem due to abnormal prices and volumes. Next, based on the market behaviour before the expiration day, SVM & RF are conducted to identify whether there will be a market corner risk on the expiration day. The two-step design of the model can identify the risk based on its performance characteristics and achieve the aims of predicting in advance through the training data and the segmented structure of time nodes. Consequently, it allows regulators and investors to be forewarned and facilitate timely adjustments and responses. The main contributions of this article are summarized as follows.

- To the best of our knowledge, this is the first work applying an anomaly detection algorithm to the crude oil futures market to identify market corner risk in time and predict possible price deviation in advance, assisting in reducing the chain reaction caused by the crisis and strengthen risk monitoring.
- We propose a hybrid detection approach for supervised learning classification based on pseudo-labels generated by an unsupervised learning algorithm according to the patterns of forced position risk. The method achieves excellent applicability and robustness.
- An original data set obtained in this work and the data set after each step of processing and feature selection are provided.

The remainder of this article is organized as follows. In Section Section 2, some related research to this investigation is presented. Next, we depict a typical case study in Section 3, followed by the problem definition in Section 4, the proposed model and the overall detection framework presentation in Section 5, and discussions on the performance and scalability of the proposed model in Section 6. Finally, the concluding remarks and the future work are presented in Section 7.

2. Related Work

In this section, we survey the related works in three aspects: the corresponding speculative impact of significant events in Section 2.1, traditional econometric methods to identify price anomalies in Section 2.2, and anomaly detection algorithm to identify price anomalies in Section 2.3.

2.1. Price Volatility of Crude Oil Futures Market

Crude oil is one of the prominent representatives of bulk commodities, because of the price discovery function and substantial directional effect of crude oil futures price, there have been many studies on the price volatility and anomalies. Ji *et al.*[8] studied the price spillover and found that the crude oil futures price is mainly affected by demand factors. The block swan event and other severe market fluctuations systematically impact the price. However, many speculative trading behaviours will increase the impact of volatility in this process. Fong *et al.* [40] further confirmed that the volatility mechanism is strongly related to significant events affecting oil supply and demand. Kaufmann *et al.* [41] found that, market fundamentals are the essential factors causing oil price changes in the long run, but when speculators realize that the possibility of oil prices rising over time increases, they will start to speculate, which aggravates the sharp short-term fluctuations of oil prices. In [42], Dulaimi *et al.* presented that, any medium and long-term price forecast is complex and uncertain. The only thing that can be confirmed seems to be the significant fluctuation of oil market prices under the influence of some military, political, and economic events in the future.

2.2. Identifying Price Anomalies with Traditional Statistical Methods

Crude oil futures often experience significant price deviations due to the financial pressures exerted by short and long positions, leading to rapid discrepancies between spot and futures prices. In the financial realm, abnormal trading behaviour poses substantial challenges to market oversight and risk management. Furthermore, seasoned traders collaborate to form collusion teams, employing similar trading strategies to deceive other investors and maximize profits. Wang *et al.* [43] proposed a method to detect potential collusion groups in futures market instruments. The connected components in multiple sparse weighted graphs are combined to form a risk monitoring and management tool based on the futures market, proposed as a pilot application. Caporale *et al.* [44] uses statistical techniques and trading simulation methods to investigate the weekly effect in the cryptocurrency market, detecting whether there are significant differences between price changes and stochastic outcomes. Sun *et al.* [45] analyzed the trading records of several stocks and found a high degree of averaging trading networks and a low correlation between price returns and the proportion of buyers and sellers.

2.3. Identifying Market Manipulation with Anomaly Detection Algorithm

In recent years, researchers have increasingly utilized diverse classification techniques to identify financial risks proactively. Consequently, the establishment of suitable classifiers, either individually or as an ensemble, has become a requisite step in the context of financial risk prediction. Li *et al.* [46] highlighted the suitability of supervised machine learning methods for detecting market manipulation in daily transaction data, though they underperformed on tick data. Cao *et al.* [47] introduced Ahmmas, a method that combines wavelet transform for price anomaly detection and manipulation identification in real-time. Additionally, research has extended to applying anomaly detection algorithms in the cryptocurrency market. Sayadi *et al.* [48] proposed a transaction anomaly detection model with high accuracy, while Morgia *et al.* [49] tackled pump & dump behaviour in cryptocurrency and presented a real-time fraud detection approach.

3. Negative Settlement Price of WTI: A Case Study

This section focuses on a critical case study: the adverse oil price event at West Texas Intermediate (WTI). This study scrutinizes instances where observed prices notably diverged from the established normative price trajectory. To elucidate this phenomenon more effectively, the K-line charts depicting WTI's price trends during the years 2008, 2015, and 2020 are presented as Fig 2 (a), Fig 2 (b), and Fig 2 (c), respectively. Additionally, a box chart illustrating the WTI price trend for the year 2020 is included as Fig 2 (d), providing a comprehensive visual analysis of these anomalous pricing events.

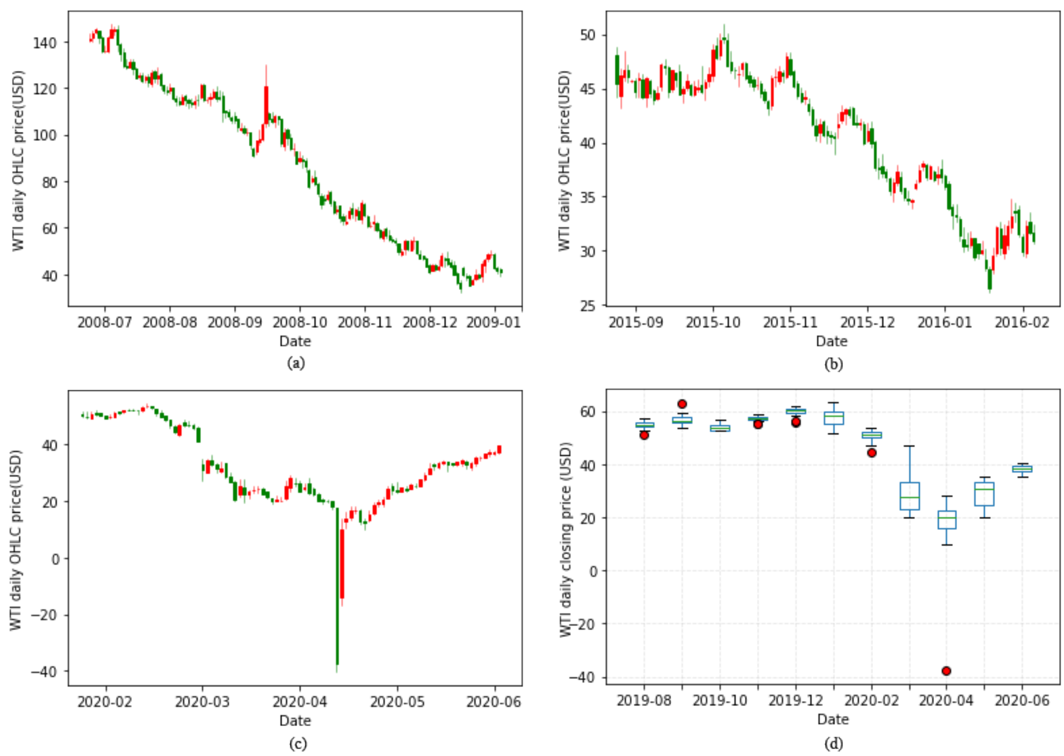

Figure 2. Intense price fluctuation .

The negative oil price settlement event (April 20, 2020) of the May futures contract occurred in the WTI market, causing large-scale losses at investment institutions and severe adverse effects as shown in Fig 2 (c). On this day, the WTI May futures contract settled at $-37.63\$$ per barrel, an unprecedented level of volatility in the history of the crude oil market.

To exacerbate the situation, CME responded to the negative energy options prices by announcing on April 21, 2020, the launch of option contracts with negative exercise prices. This decision followed technical adjustments made on April 8, transitioning from the Black-Scholes Option Pricing Model (BS model) to the Bachelier model, as illustrated in Fig 3. Under Bachelier's formula, the expressions for call and put options are defined by Eq 1 and Eq 2.

CME Group Advisory Notice

TO: Clearing Member Firms
Chief Financial Officers
Back Office Managers

FROM: CME Clearing

ADVISORY #: 20-171

SUBJECT: Switch to Bachelier Options Pricing Model

DATE: April 21st, 2020

Pursuant to Clearing Advisory 20-152 that was published on April 8th, the clearing house will switch the options pricing and valuation model to Bachelier to accommodate negative prices in the underlying futures and allow for listing of option contracts with negative strikes for the set of products specified below.

The switch will be effective for the margin cycle run at the end of trading tomorrow April 22, 2020 and will remain in place until further notice.

Figure 3. CME group announcement.

$$C(P,T) = PN\left(\frac{P-X}{\sigma\sqrt{D}}\right) - XN\left(\frac{P-X}{\sigma\sqrt{D}}\right) + \sigma\sqrt{D}\left(\frac{P-X}{\sigma\sqrt{D}}\right) \tag{1}$$

$$P(P,D) = XN\left(\frac{X-P}{\sigma\sqrt{T}}\right) - PN\left(\frac{X-P}{\sigma\sqrt{D}}\right) + \sigma\sqrt{D}\left(\frac{P-X}{\sigma\sqrt{D}}\right) \tag{2}$$

where P is the underlying asset price, D the duration, and σ the volatility. Unlike the BS model, the basic principles of the Bachelier model are as follows.

- Assuming that the market is fully liquid and the underlying asset price is continuously changing,
- The underlying asset fluctuates around the real price,
- The underlying asset price changes relatively gently.

The Bachelier model begins with the assumption of standard Brownian motion. Under this model, $S(T)$ follows a normal distribution, allowing for both positive and negative asset prices. This implies that negative oil prices can be calculated as settlement prices. To visually assess the deviation of prices on April 20, we constructed a box chart using data from August 2019 to June 2020 (Fig 2(d)). The settlement prices deviated significantly, indicating clear signs of a market corner. Surprisingly, neither regulators nor investors had taken preventive measures against this risk. This event garnered substantial media attention and resulted in significant losses for both institutional and individual investors. However, early warnings to regulators and investors could have potentially averted such losses and risks.

4. Data Description and Problem Definition

Based on the historical WTI data, this section aims to detect anomalous prices in the WTI trading process. section 4.1 describes the data used, while section 4.2 presents the problem definition.

4.1. Data Description

This article focuses on the price anomalies caused by market corners and is not intended to discuss the reasons behind the trend or the manipulation initiated by any institution or individual. The data used comes from the data related to the market and price *Investing* [1]. The downloaded data includes daily trading data from April 1983 to October 2020, containing opening price(*open*), closing price(*close*), highest price(*high*), lowest price(*low*), change rate(*rate*), and trading volume(*volume*).

4.2. Problem Definition

Under the efficient market theory, the long-term price trend obeys the law of supply and demand [50]. Thus, the black swan event primarily impacts short and medium-term prices, while price manipulation affects prices briefly. In anomaly detection, set anomalies correspond to black swan events, and point anomalies identify individual price manipulation instances, often associated with market corners. These corners typically exhibit abnormal price and trading volume behaviour, especially on maturity dates. Due to data limitations, we approximate settlement day data using the closest available variable, the closing price. The dataset comprises daily closing prices (*close*) and trading volumes (*volume*). Examples of real data used in the model are shown in Tab 1.

Table 1. Examples of real data used for market corner detection.

date	1983-4-4	1983-4-5	1983-4-6	1983-4-7	1983-4-8	1983-4-11	1983/4/12	···	2020-10-15
close	29.44	29.71	29.90	30.17	30.38	30.25	30.83	···	40.96
volume	0.16K	0.18K	0.39K	0.82K	0.37K	0.27K	0.47K	···	248.25K

A pseudo label corresponding to the market corner can be generated based on the anomaly detection results. The closing date for trading in the current delivery month is the third business day before the 25th day of the previous month. Based on the design mechanism of the settlement price, the settlement date is not fixed, so we programmed all the predicted points in time to be the settlement date based on the actual situation in the model. The pseudo-code is shown in Algorithm 1.

The occurrence of market corners is not a temporary intention but needs to be arranged in advance. Moreover, the event of abnormal prices and trading volume is not accidental, but the factors of planning and layout in advance, and these activities will eventually be reflected in the market trading behaviour. Therefore, the data we use are the market change data generated by trading activities on all trading days before the settlement date of each month. The label we used in fitting is the first step, divided into abnormal for '1' and non-abnormal for '0'. The examples of real data used in the model are shown in Tab 2.

Table 2. Examples of real data for market corner classification.

date	**close**	**open**	**high**	**low**	**volume**	**rate**
2020-4-1	20.31	20.10	21.55	19.90	703.29K	1.04%
2020-4-2	25.32	21.22	27.39	20.76	1.10MK	19.32%
2020-4-3	28.34	24.81	29.13	23.520	1.01M	14.23%
2020-4-6	26.08	26.09	28.24	25.28	752.71K	-0.04%
2020-4-7	23.63	26.34	27.24	23.54	797.71K	-10.29%
···	···	···	···	···	···	···
2020-4-17	20.31	20.10	21.55	19.90	703.29K	-8.65%
2020-4-20			**1**			

[1] https://cn.investing.com/

Algorithm 1 Get settlement date

Require: List_mate, which is stored in months;
Ensure: y_train, which is stored in months;

```
list_copy ← [];
y_train ← [];
x_train ← [];
n1 ← 0
n2 ← 0
for var 0 to len(list_mate) step 1 do
    d ← list_mate[i] ;
    n2 ← n2 + 1 ;
    dnx ← d [close, open, high, low, rate, volume] ;
    dny ← d[y] ;
    xnp ← np.array(dnx) ;
    ynp ← np.array(dny) ;
    xnp2 ← np.array(xnp[len(xnp) - 20 len(xnp) - 4])
    if len(xnp2) = 0 then then
        y_train ← np.append(y_train, ynp[len(ynp) - 4]) # have 25 ;
        x_train ← np.append(x_train, xnp2) ;
    elsen1 ← n1 + 1 ;
    end if
end for
return [x_train, y_train]
```

In summary, this approach begins by detecting market corner risk based on specific characteristics and subsequently generates pseudo-labels derived from this identification process. Recognizing that market corners are not random, spontaneous events but rather premeditated actions, we proceed to classify them based on market behaviour and the labels established in the initial step

5. Methodology and Proposed Model

This section first introduces the used unsupervised anomaly detection method in section 5.1 and section 5.2, then presents the model evaluation metrics in 5.3, and finally presents the overall framework for the research model in 5.4.

5.1. LOF

In the first step, the unsupervised learning anomaly detection algorithm detects anomalous volume and price changes. LOF is a typical high-precision outlier detection method based on density [51]. As Fig 4 shows, the detection results are based on calculating the local density deviation of a given data point relative to its neighbourhood. Each data point is assigned an outlier factor in this method, depending on the neighbourhood density, and then determines whether the data point is an outlier. Therefore, we apply LOF to detect the outliers in volume and price changes, which aligns with the problem definition.

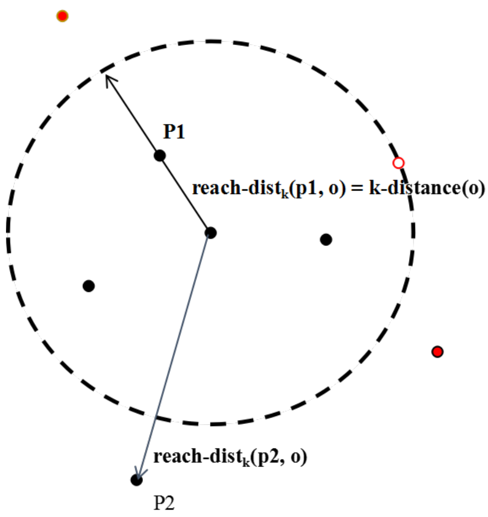

Figure 4. k-distance.

5.2. SVM&RF

In this step, we fit and train a supervised learning model based on the labels derived in the previous step. According to the relevant research results in the model selection [46] [52] [53], SVM and RF algorithms show strong robustness and applicability in financial time series. Due to this, they are chosen to solve this problem. SVM is a binary classification model, and the samples are classified hyper-parametrically using a separated hyper-plane $\boldsymbol{w} \cdot x + b = 0$. RF is a classifier that uses multiple decision trees to train and predict samples. The Equation is as follows.

$$Predict = \mathrm{RF}(feature_1, feature_{2,......}feature_n) \tag{3}$$

5.3. Performance Evaluation Metrics

To evaluate the performance of the proposed model in anomaly detection, we use a confusion matrix approach, consisting of four parts: TP, FP, FN, and TN, where $TP = truepostive$, $TN = truenegative$, $FP = falsepositive$, $FN = falsenegative$. On this basis, we calculate the results of $Accyracy$, $Precision$, $Recall$, $F - score$, AUC, and ROC as evaluation metrics, and the equations are defined as follows.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$F - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{7}$$

5.4. Detecting Framework

According to the problem definition and methodology description, the model consists of three parts, as shown in Fig 5.

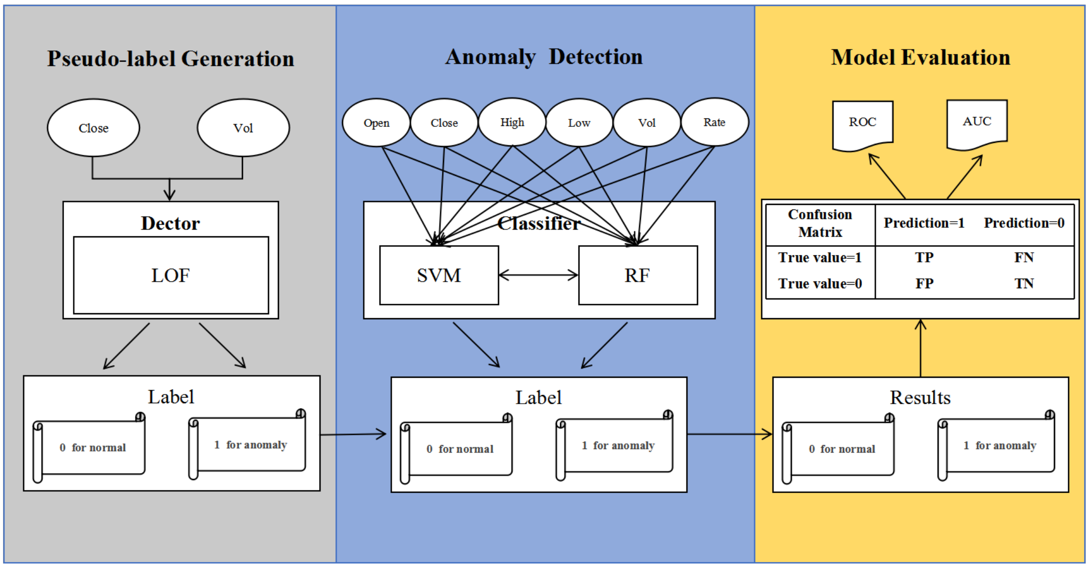

Figure 5. Model Framework.

In the first part of the model, the unsupervised learning-based LOF is used to detect dates with significant price and volume anomalies, and all identified dates are marked with 0 as normal and 1 as an anomaly. Based on the WTI's settlement mechanism, all trading data from the 1st day of each month to the day before the settlement day are used as features in the second part of the model. Two supervised learning algorithms, SVM and RF, are used as classifiers. As we use the pseudo-labels generated in the first part to train the classifier, comparisons on the predicted value with the real value are performed, and the model is evaluated next.

6. Experimental Results

6.1. Data Processing

The data selected in this article are the real transaction data of WTI as mentioned in Section 4.1. We use the median interpolation method to fill in the missing value for the lost value of trading volume in the data. Then, to update the original data more quickly, we use a linear function to transform the original data into the range of [0-1] and normalize the data. The normalization is as Eq 8.

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{8}$$

In this method, the original data is scaled equivalently. Based on the data set, we use the data between the trading day of each month and the trading day of each month to predict whether there will be any abnormality in the trading day. Among them, the data between two trading days is the training data value, whether the trading day is abnormal is tagged, taking 0 and 1. There are 422 data in total, among which the ratio between the abnormal trading day and the normal trading day is 121:301. When training SVM and RF, the training and test sets were randomly divided in a 358:64 ratio, preventing the training set from being randomly divided into only one classification or too few of the other.

6.2. Feature Selection

As mentioned in Section 4.2, in the first step of the anomaly detection algorithm, we select the features of daily settlement price (*close*) and volume (*vol*). Next, we take the month as the unit and select the market trading data of all trading days before the settlement date, including opening price(open), closing price(*close*), highest price(*high*), lowest price(*low*), change rate(*rate*) and trading volume(*vol*) to train the classifier. After data processing and feature selection, the final modelling data examples in step 1 are shown in Tab 3, and data examples in step 2 are shown in Tab 4.

Table 3. Examples of modeling data in step 1.

date	1983-4-4	1983-4-5	1983-4-6	1983-4-7	1983-4-8	1983-4-11	1983-4-12	⋯	2020-10-15
close	0.161905	0.161785	0.161982	0.161929	0.161891	0.161915	0.161857	⋯	0.163518
volume	0.148788	0.148013	0.148321	0.148148	0.147733	0.147767	0.147791	⋯	0.343165

Table 4. Examples of modeling data in step2.

Feature1	**Feature2**	**Feature3**	**Feature4**	**Feature5**	**Feature6**	⋯	**Feature96**	**Label**
0.161476	0.161490	0.161524	0.161442	0.147208	0.148696	⋯	0.1498046	0
0.161673	0.161644	0.161683	0.161630	0.147516	0.148181	⋯	0.1509607	1

6.3. Anomaly Detection

In the initial step, the LOF is applied to identify outliers based on daily trading volume (*vol*) and closing price (*close*) characteristics, followed by the retrieval and marking of abnormal point coordinates. Three thresholds of 6%, 12%, and 18% are employed, as shown in Fig 6.

Subsequently, kernel density estimation is employed to fit the LOF factor distribution for each point under different thresholds, as shown in Fig 7. While selecting an optimal threshold based solely on distribution fitting is challenging, the 12% threshold strikes a balance between the distribution of abnormal and normal points, aligning with practical trading experience and the related works [52,53]. Hence, the detection results under the 12% threshold are chosen for further investigation.

6.4. Parameter Sensitivity

According to the results in step 1, we have obtained the label corresponding to the sample. Besides, according to the definition of the problem, we take the abnormal results of the first step on settlement day as the training label and the market data of the trading day before the settlement day of each month as the training characteristics, so that SVM and RF are conducted to this dataset. In the model training, we use the ratio of training set: test set as 85%:15 % to train the model. We need to adjust the model parameters to obtain a better training model.

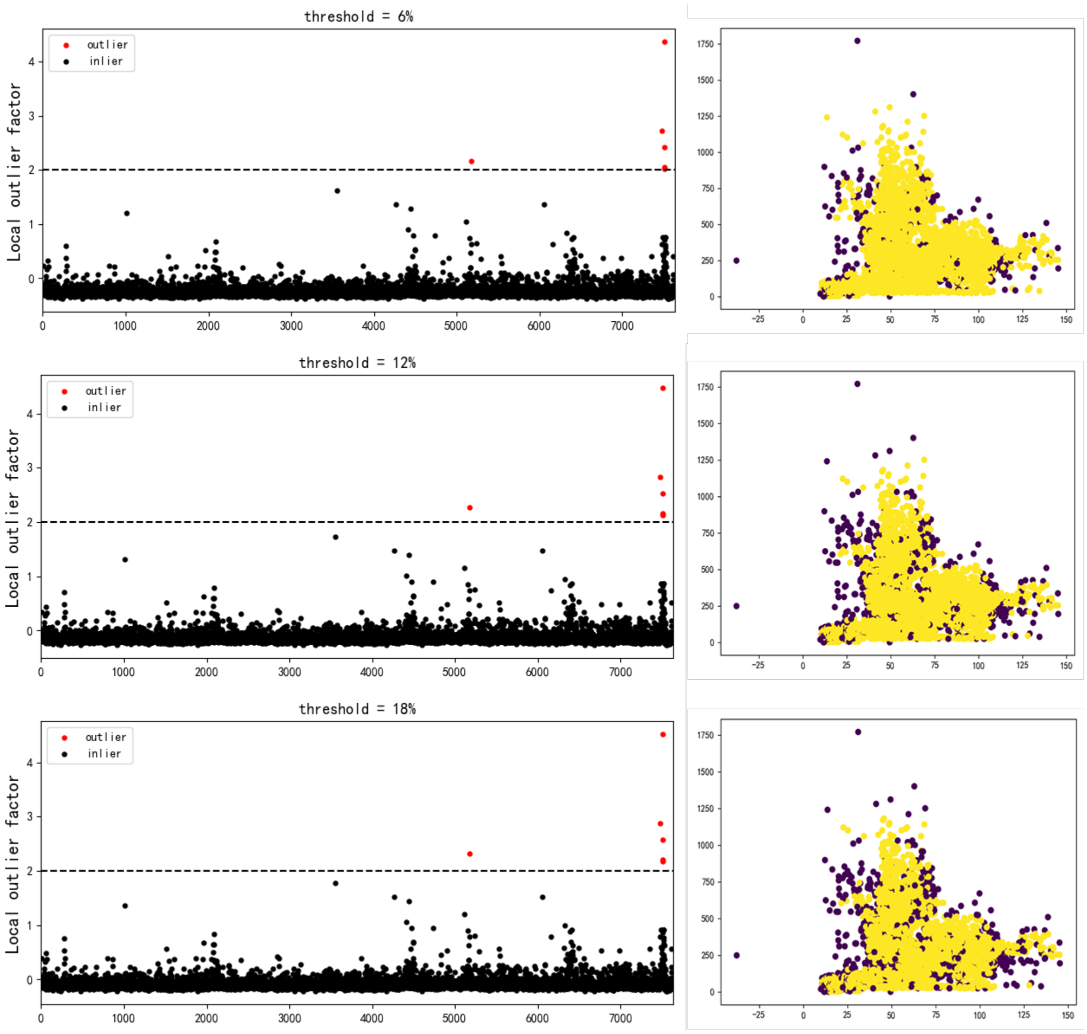

Figure 6. Detection results under the thresholds of 6%, 12%, and 18%.

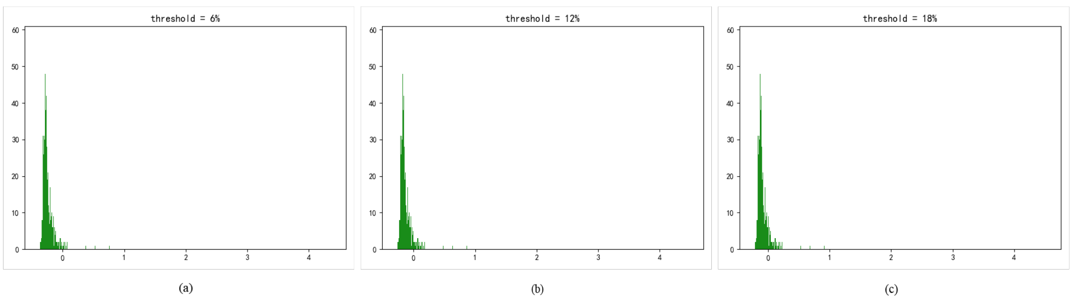

Figure 7. Distribution of LOF factors under the thresholds of 6%, 12%, 18%.

6.4.1. Optimal Parameters in SVM

The most frequently adjusted parameters of SVM are cost($-c$) and gamma($-g$). $-c$ is the penalty coefficient. $-g$ is a parameter of the RBF function after it is selected as the kernel. We use the grid-optimization method to find the optimal $-c$ and $-g$. The optimal parameters of SVM are $-c = 1.0$, $-g = 0.5$, and the code is shown in Algorithm 2.

Algorithm 2 Identify the best parameters in SVM

Require: x_train, which is the data set; y_train, which is the list set;
Ensure: best_estimator, which are the best parameters

```
x,y,z ← []
for var C from 1to 10 in step 1 do
    auc =
cross_val_score(SVC(C=C,kernel='rbf',gamma=gamma/10),
x_train,y_train,cv=5,scoring='roc_auc').mean();
    x.append(C)
    y.append(gamma/10)
    z.append(auc)
end for
x ← np.array(x).reshape(9,10)
y ← np.array(y).reshape(9,10)
z ← np.array(z).reshape(9,10)
best_estimator ← [x, y, z]
return best_estimator
```

6.4.2. Optimal Parameters in RF

The main parameters of RF include the number of subtrees(*n_estimators*), the maximum growth depth of trees(*max_depth*), the minimum number of samples of leaves(*min_samples_split*), the minimum number of samples of branch nodes(*min_samples_split*), and the maximum number of selected features(*max_features*). Based on the actual needs, we select three parameters to optimize, from which the two most influential parameters are *n_Estimators*, *max_depth*, and *max_deatures*. The optimal parameters are given as *n_estimators*= 85, *max_depth*= 14, and *max_deatures*= 18.

6.5. Performance Evaluation

To evaluate the performance of the proposed model, we use the confusion matrix to assess the performance of SVM and RF on the verification set. The results are shown in Tab 5.

Table 5. Performance comparison of SVM & RF.

Method	Accuracy	Precision	Recall	F-score	AUC Score
SVM	0.812500	0.833333	0.909091	0.869565	0.754545
RF	0.875000	0.860000	0.977273	0.914894	0.859091

Based on the dataset shown in Tab 5, the ROC curve on the verification set is drawn.

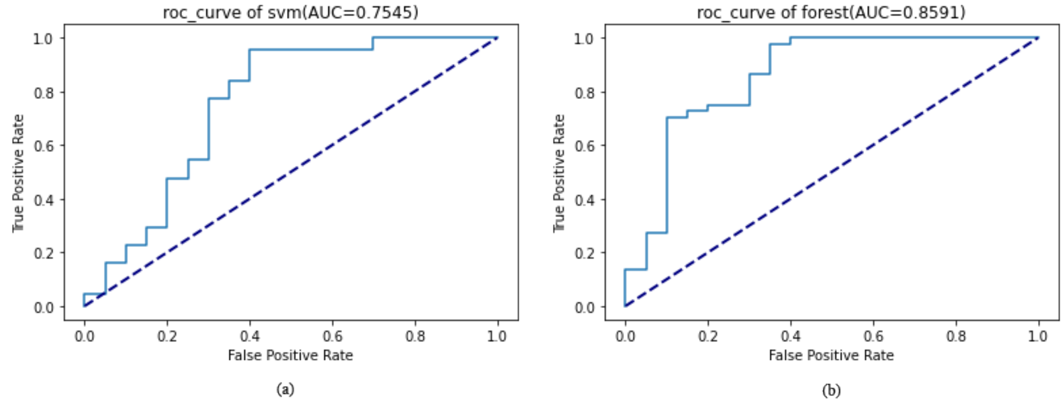

Figure 8. ROC curve of SVM&RF.

As Fig 8 shows, both the SVM and the RF are feasible and have strong robustness. Nevertheless, the performance of RF is better.

6.6. Scalability

While the current model has demonstrated exceptional performance, it possesses the limitation of detecting abnormal prices only one day before the settlement day in practical scenarios. Given this short lead time, there may be insufficient opportunity for preventive measures or adjustments. Often, on the settlement day, contracts remain untraded due to the absence of counter offers. To address this issue, we propose extending the model's forecasting horizon to seven trading days in advance while excluding data near the settlement date. This approach aims to mitigate the limitations, and you can find illustrative data examples in Tab 6.

Table 6. Examples of real data.

date	**close**	**open**	**high**	**low**	**volume**	**rate**
2020-4-1	20.31	20.10	21.55	19.90	703.29K	1.04%
2020-4-2	25.32	21.22	27.39	20.76	1.10MK	19.32%
2020-4-3	28.34	24.81	29.13	23.520	1.01M	14.23%
2020-4-6	26.08	26.09	28.24	25.28	752.71K	-0.04%
2020-4-7	23.63	26.34	27.24	23.54	797.71K	-10.29%
2020-4-8	25.09	24.30	26.45	23.74	823.55K	6.18%
2020-4-9	20.31	26.28	28.36	22.57	1.12M	-9.29%
2020-4-20	**1**					

Then, we applied the processed dataset to fit the model, adjusted the parameters, and tested it in the validation set. The performance is shown in Tab 7.

The ROC curve on the verification is shown in Fig 9. In the forecast performed, the accuracy and F-score of both SVM and RF declined since some samples were removed near the settlement date. Nevertheless, they still have a certain degree of prediction. Moreover, the performance of RF is slightly better than SVM.

Table 7. Performance comparison of SVM & RF.

Method	Accuracy	Precision	Recall	F-score	AUC Score
SVM	0.788235	0.797619	0.985294	0.881579	0.692907
RF	0.843750	0.862069	0.961538	0.909091	0.633814

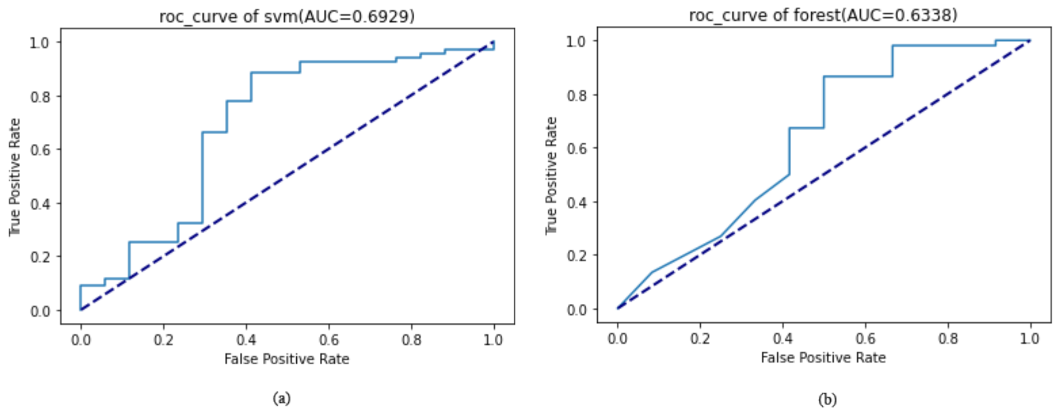

Figure 9. The ROC curve of SVM & RF.

6.7. Results Discussion

The analysis of the market corner, characterized by abnormal trading volumes and price fluctuations, reveals a significant detection of abnormal values, all of which have been meticulously identified and marked. This is evident from the disproportionate ratio of abnormal values on expiration days compared to non-expiration days, with $R_1 = 1{:}2.63$. This ratio is notably higher than the 1:21.53 ratio of expiration to non-expiration days, represented as R_2. It's noteworthy that expiration days constitute only 4.439% (R_3) of the total trading days. However, the proportion of abnormal points on these days is unexpectedly high, accounting for 27.548% (R_4) of the total abnormal points. In a market without corners and related maturity effects, the ratio of R_4/R_3 would theoretically approximate 1. However, according to our model's results, the likelihood of anomalies during maturity is 6.2 times higher than during non-maturity, as indicated by $R_4/R_3 = 6.2$. This stark contrast in the data compellingly supports the hypothesis of a significant market corner presence.

In the subsequent phase of the study, we focus on predicting price deviations caused by the market corner a day ahead. The model exhibits exceptional performance and robustness in this regard. To better reflect real-world market dynamics, we also aim to predict potential market corner risks one week ahead. While this extension does lead to a slight reduction in the model's precision and robustness due to data constraints, the approach remains fundamentally sound and feasible. These experimental findings affirm the model's capability to identify potential price manipulation and mitigate the risk of market corners.

7. Conclusions and Future Work

In this article, we introduce a novel hybrid model that combines LOF with SVM and RF to address the challenges posed by the unpredictable nature of international crude oil price fluctuations. This model leverages key features of the futures market, specifically abnormal closing prices and trading volumes, to conduct initial detection. This initial phase scrutinizes these factors and utilizes the findings to generate labels that are essential for the next stage of the process. In the subsequent step, we acknowledge that market corner behaviour is typically premeditated and manifests in market trading patterns. Therefore, we employ market trading data to fine-tune our model. This training, grounded in the labels derived from the earlier phase, enables the model to make informed predictions. These predictions are particularly focused on the intent and potential impacts of abnormal price fluctuations. Experimental results show that the proposed model exhibits exceptional performance and robustness, offering valuable insights for practical market regulation and future research initiatives. Notably, the exploration of the relationship between the definition of price manipulation and label composition presents significant implications for further investigation.

Nevertheless, there is potential for further enhancement of this study, which includes:

(1) Exploring datasets with finer time granularity, as the current study relies on daily data that may not precisely align with the settlement mechanism.
(2) Updating and integrating other computational models to augment the efficiency of the proposed framework.
(3) Extending the application of this model to other financial market sectors to validate its effectiveness and adaptability.

Data Availability

The authors confirm that the data supporting the findings of this study are available within the article [and/or] its supplementary materials.

References

[1] Wang Q, Sun X. Crude oil price: Demand, supply, economic activity, economic policy uncertainty and wars–from the perspective of structural equation modelling (sem). Energy. 2017;133:483–490.
[2] Zhang YJ, Zhang L. Interpreting the crude oil price movements: Evidence from the markov regime switching model. Applied Energy. 2015;143:96–109.
[3] Wang GJ, Xie C. Cross-correlations between wti crude oil market and us stock market: A perspective from econophysics. Acta Physica Polonica B. 2012;43(10).
[4] Kilian L. The impact of the shale oil revolution on us oil and gasoline prices. Review of Environmental Economics and Policy. 2016;10(2):185–205.
[5] Liu H, Li J. The us shale gas revolution and its externality on crude oil prices: a counterfactual analysis. Sustainability. 2018;10(3):697.
[6] Wei Y, Wang Y, Huang D. Forecasting crude oil market volatility: Further evidence using garch-class models. Energy Economics. 2010;32(6):1477–1484.
[7] Charles A, Darné O. Forecasting crude-oil market volatility: Further evidence with jumps. Energy Economics. 2017;67:508–519.

[8] Ji Q, Fan Y. How does oil price volatility affect non-energy commodity markets? Applied Energy. 2012;89(1):273–280.
[9] Ledgerwood SD, Carpenter PR. A framework for the analysis of market manipulation. Review of Law & Economics. 2012;8(1):253–295.
[10] Markham JW. Regulating the too big to jail financial institutions. Brook L Rev. 2017; 83:517.
[11] Pirrong C. Energy market manipulation: definition, diagnosis, and deterrence. Energy LJ. 2010;31:1.
[12] Chkili W, Hammoudeh S, Nguyen DK. Volatility forecasting and risk management for commodity markets in the presence of asymmetry and long memory. Energy Economics. 2014;41:1–18.
[13] Herrera GP, Constantino M, Tabak BM, et al. Long-term forecast of energy commodities price using machine learning. Energy. 2019;179:214–221.
[14] Li M, Han D, Yin X, et al. Design and implementation of an anomaly network traffic detection model integrating temporal and spatial features. Security and Communication Networks. 2021;2021.
[15] Zhang X, Zhang L, Li D. Transmission line abnormal target detection based on machine learning yolo v3. In: 2019 International Conference on Advanced Mechatronic Systems (ICAMechS); IEEE; 2019. p. 344–348.
[16] Li D, Han D, Zhang X, et al. Panoramic image mosaic technology based on sift algorithm in power monitoring. In: 2019 6th International conference on systems and informatics (ICSAI); IEEE; 2019. p. 1329–1333.
[17] Li M, Han D, Li D, et al. Mfvt: an anomaly traffic detection method merging feature fusion network and vision transformer architecture. EURASIP Journal on Wireless Communications and Networking. 2022;2022(1):39.
[18] Cai S, Han D, Li D, et al. An reinforcement learning-based speech censorship chatbot system. The Journal of Supercomputing. 2022;78(6):8751–8773.
[19] Li D, Han D, Crespi N, et al. A blockchain-based secure storage and access control scheme for supply chain finance. The Journal of Supercomputing. 2023;79(1):109–138.
[20] Gou H, Zhang X, Yang J, et al. Three-way fusion measures and three-level feature selections based on neighborhood decision systems. Applied Soft Computing. 2023;:110842.
[21] Aghaee R, Momeni M, Moallem P. A fusion-based approach to improve hyperspectral images' classification using metaheuristic band selection. Applied Soft Computing. 2023; :110753.
[22] Diao C, Zhang D, Liang W, et al. A novel spatial-temporal multi-scale alignment graph neural network security model for vehicles prediction. IEEE Transactions on Intelligent Transportation Systems. 2022;24(1):904–914.
[23] Hu N, Zhang D, Xie K, et al. Multi-graph fusion based graph convolutional networks for traffic prediction. Computer Communications. 2023;210:194–204.
[24] Hu N, Zhang D, Xie K, et al. Multi-range bidirectional mask graph convolution based gru networks for traffic prediction. Journal of Systems Architecture. 2022;133:102775.
[25] Cai J, Liang W, Li X, et al. Gtxchain: A secure iot smart blockchain architecture based on graph neural network. IEEE Internet of Things Journal. 2023;.
[26] Long J, Liang W, Li KC, et al. A regularized cross-layer ladder network for intrusion detection in industrial internet of things. IEEE Transactions on Industrial Informatics. 2022;19(2):1747–1755.
[27] Liang W, Li Y, Xie K, et al. Spatial-temporal aware inductive graph neural network for c-its data recovery. IEEE Transactions on Intelligent Transportation Systems. 2022;.
[28] Liang W, Yang Y, Yang C, et al. Pdpchain: A consortium blockchain-based privacy protection scheme for personal data. IEEE Transactions on Reliability. 2022;.
[29] Xu Z, Liang W, Li KC, et al. A time-sensitive token-based anonymous authentication and dynamic group key agreement scheme for industry 5.0. IEEE Transactions on Industrial Informatics. 2021;18(10):7118–7127.
[30] Liang W, Li Y, Xu J, et al. Qos prediction and adversarial attack protection for distributed

services under dlaas. IEEE Transactions on Computers. 2023;.
[31] Long J, Luo C, Chen R, et al. A cross-layered cluster embedding learning network with regularization for multivariate time series anomaly detection. The Journal of Supercomputing. 2023;:1–25.
[32] Xiao L, Han D, Li KC, et al. Uetopsis: A data-driven intelligence approach to security decisions for edge computing in smart cities. ACM Transactions on Sensor Networks. 2024;.
[33] Han D, Zhou H, Weng TH, et al. Lmca: a lightweight anomaly network traffic detection model integrating adjusted mobilenet and coordinate attention mechanism for iot. Telecommunication Systems. 2023;84(4):549–564.
[34] Li Y, Liang W, Xie K, et al. Lightnestle: quick and accurate neural sequential tensor completion via meta learning. In: IEEE INFOCOM 2023-IEEE Conference on Computer Communications; IEEE; 2023. p. 1–10.
[35] Zhou S, Li K, Xiao L, et al. A systematic review of consensus mechanisms in blockchain. Mathematics. 2023;11(10):2248.
[36] Gong Y, Li K, Xiao L, et al. Vaserp: an adaptive, lightweight, secure, and efficient rfid-based authentication scheme for iov. Sensors. 2023;23(11):5198.
[37] Zhang S, Hu B, Liang W, et al. A trajectory privacy-preserving scheme based on transition matrix and caching for iiot. IEEE Internet of Things Journal. 2023;.
[38] Zhang S, Hu B, Liang W, et al. A caching-based dual k-anonymous location privacy-preserving scheme for edge computing. IEEE Internet of Things Journal. 2023;.
[39] Zhang S, Yan Z, Liang W, et al. Baka: Biometric authentication and key agreement scheme based on fuzzy extractor for wireless body area networks. IEEE Internet of Things Journal. 2023;.
[40] Fong WM, See KH. A markov switching model of the conditional volatility of crude oil futures prices. Energy Economics. 2002;24(1):71–95.
[41] Kaufmann RK, Ullman B. Oil prices, speculation, and fundamentals: Interpreting causal relations among spot and futures prices. Energy Economics. 2009;31(4):550–558.
[42] Al Dulaimi HAM. The collapse of crude oil prices: Cyclical evolution or market manipulation? Economic Insights-Trends & Challenges. 2014;66(4).
[43] Wang J, Zhou S, Guan J. Detecting potential collusive cliques in futures markets based on trading behaviors from real data. Neurocomputing. 2012;92:44–53.
[44] Caporale GM, Plastun A. The day of the week effect in the cryptocurrency market. Finance Research Letters. 2019;31.
[45] Sun XQ, Cheng XQ, Shen HW, et al. Distinguishing manipulated stocks via trading network analysis. Physica A: Statistical Mechanics and its Applications. 2011;390(20):3427–3434.
[46] Li A, Wu J, Liu Z. Market manipulation detection based on classification methods. Procedia Computer Science. 2017;122:788–795.
[47] Cao Y, Li Y, Coleman S, et al. Adaptive hidden markov model with anomaly states for price manipulation detection. IEEE transactions on neural networks and learning systems. 2014;26(2):318–330.
[48] Sayadi S, Rejeb SB, Choukair Z. Anomaly detection model over blockchain electronic transactions. In: 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC); IEEE; 2019. p. 895–900.
[49] La Morgia M, Mei A, Sassi F, et al. Pump and dumps in the bitcoin era: Real time detection of cryptocurrency market manipulations. arXiv preprint arXiv:200506610. 2020; .
[50] Jovanovic F, Andreadakis S, Schinckus C. Efficient market hypothesis and fraud on the market theory a new perspective for class actions. Research in International Business and Finance. 2016;38:177–190.
[51] Breunig MM, Kriegel HP, Ng RT, et al. Lof: identifying density-based local outliers. In: Proceedings of the 2000 ACM SIGMOD international conference on Management of data; 2000. p. 93–104.

[52] Wu J, Yuan Q, Lin D, et al. Who are the phishers? phishing scam detection on ethereum via network embedding. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 2020;52(2):1156–1166.
[53] Kamps J, Kleinberg B. To the moon: defining and detecting cryptocurrency pump-and-dumps. Crime Science. 2018;7(1):18.