

Article

Intelligent Risk Early Warning Model for Coupling Risk of Oil Pump Pipeline System in Station Under Soft Soil Foundation Conditions Based on ABC-XGBoost Algorithm

Shengyang Yu ^{1,2}, Xiangsong Feng ¹, Liwen Chen ^{1,3}, Qingqing Xu ^{1,*}  and Shaohua Dong ¹

¹ College of Safety and Ocean Engineering, China University of Petroleum (Beijing), Beijing 102249, China; yushengyang@cnpicnt.com (S.Y.); 2023310514@student.cup.edu.cn (X.F.); 2022211037@student.cup.edu.cn (L.C.); shdong@cup.edu.cn (S.D.)

² China National Oil and Gas Exploration and Development Corporation, Beijing 102100, China

³ Hefei General Machinery Research Institute Co., Ltd., Hefei 230031, China

* Correspondence: xuqq@cup.edu.cn

Abstract

With rapid economic development in China's coastal regions, more oil stations are being built on soft soil foundations, facing risks such as foundation settlement and pipeline failures. Mechanical vibrations of oil pumps can induce resonance in pipelines, leading to rupture, leakage, and fire or explosion, threatening both safety and sustainable operation. Traditional monitoring methods, relying on physical models or data-driven approaches alone, are limited in capturing these coupled risks. This study proposes an ABC-XGBoost hybrid risk warning model, where the artificial bee colony algorithm optimizes XGBoost hyperparameters (iteration number, tree depth, learning rate) to improve predictive accuracy. By using multidimensional data—such as internal pressure, vibration amplitude, and ground settlement—the model evaluates stress and resonance risks in real time, supporting sustainable safety management. Validation with real station data shows an accuracy of 95.22%, 2.61% higher than the unoptimized model, demonstrating effective early warning and contribution to sustainable pipeline operation.

Keywords: artificial bee colony algorithm; soft soil foundation station; risk early warning; XGBoost algorithm; sustainable operation

1. Introduction

In recent years, with the vigorous development of China's coastal economic belt, large-scale land reclamation projects have also been steadily advancing. In this context, many stations have to be built in soft soil foundation areas along the coast. The high compressibility of soft soil foundation results in significant compressive deformation when subjected to surrounding structural loads; its rheological properties lead to the continuous development of soil deformation over time, and even if the load remains constant, the foundation will still settle [1]. Ground settlement is mainly divided into uniform settlement and uneven settlement, which poses many hidden dangers to the safe operation of pipelines in the station due to geological movements. The natural gas pipeline explosion in the United States in 2010 and the pipeline burst in Dalian in 2014 were both caused by ground subsidence [2].

During the operation of the oil pump in the station, the high-speed rotation of the impeller will generate strong mechanical vibration, which will be transmitted along the



Academic Editor: Xuelong Li

Received: 26 December 2025

Revised: 31 January 2026

Accepted: 1 February 2026

Published: 9 March 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

pipeline to the supporting structure, resulting in an increase in stress on the pipeline. When the vibration frequency of the oil pump meets specific conditions with the natural frequency of the pipeline, the oil pump pipeline system will undergo resonance response, and the vibration amplitude will significantly increase. Long-term resonance can easily cause cracks in pipelines, leading to medium leakage and even serious fire and explosion accidents, resulting in huge economic losses and casualties.

At present, pipeline risk warning technology is divided into physical model-based evaluation methods, data-driven intelligent warning methods, traditional monitoring and detection technologies, and comprehensive risk assessment and decision-making methods.

The evaluation technology based on physical models mainly evaluates the safety status of pipelines through residual strength assessment, stress–strain analysis, and leakage diffusion analysis. The current research status of pipeline safety assessment technology based on physical models is detailed in Table 1.

Table 1. Research status of pipeline safety assessment technology based on physical models.

Research Category	Representative Scholars	Research Methodology	Research Object/Core Model	Main Research Content and Conclusion
Remaining Strength Assessment	Cui Mingwei et al. [3]	Comparative analysis of theoretical models	Remaining Strength Model of Corroded Pipeline	The system analyzed and compared various residual strength assessment methods for corroded pipelines, clarifying the applicability and limitations of different models under corrosion depth, morphology, and loading conditions, providing a theoretical basis for pipeline integrity assessment.
Leakage and diffusion analysis	Dongfen et al. [4]	Numerical simulation	Hazardous Chemical Leakage and Diffusion Model	The semi lethal concentration distribution characteristics of hazardous chemicals under different working conditions were predicted through leakage diffusion simulation, providing quantitative support for accident consequence assessment and emergency decision-making.
Stress and strain analysis	Zhandi et al. [5]	Finite Element Analysis (FEM)	Mechanical model of pipeline structure	Constructing a finite element model to analyze the Von Mises equivalent stress during pipeline operation, accurately predicting the location of maximum stress distribution, and providing a basis for structural failure determination.
Multi-factor modeling	Lai Lennian et al. [6]	Multiple linear regression analysis	Pipeline stress prediction model	Taking into account various influencing factors, a pipeline stress prediction model was established, revealing the influence relationship of each variable on stress evolution and improving the practicality of physical model prediction.

The data-driven intelligent warning method mainly uses machine learning classification, time series prediction, and anomaly detection algorithms. In practical engineering applications, multiple methods are often combined to fully leverage the advantages of each method. These methods have evaluated the safety of station pipelines from different perspectives and achieved good results. The current research status of data-driven pipeline intelligent warning methods is detailed in Table 2.

Table 2. Research status of data-driven pipeline intelligent warning methods.

Research Category	Representative Scholars	Research Methodology	Research Object/Core Model	Main Research Content and Conclusion
Stress prediction	Ding Shifeng [7]	LSTM + GRU	Time series of stress in ship structure	By combining LSTM and GRU networks, high-precision prediction of structural stress has been achieved, verifying the advantages of deep learning models in complex temporal modeling.
Trend prediction optimization	Wu Zemin [8]	LSTM + Particle Swarm Optimization	Model of pipeline stress development trend	The optimization of LSTM parameters through particle swarm optimization algorithm significantly improved the accuracy of stress trend prediction and model stability.
Network structure improvement	ZHANG Z et al. [9]	Weight sharing LSTM	Pipeline status warning model	Propose a weight sharing mechanism to improve the LSTM grid structure, effectively reducing model complexity and improving warning accuracy.
Leak detection	Tan Zhen et al. [10]	RNN	Pipeline leakage detection model	Constructing a leakage detection model based on RNN to achieve multi-scale leakage signal recognition enhances the detection capability for small anomalies.
Comprehensive warning	Abdulnaser M et al. [11]	Machine learning methods	Long-distance pipeline safety warning model	Propose a machine learning based early safety warning method for long-distance pipelines to achieve early identification of potential risks, which has strong engineering application value.

Traditional pipeline monitoring and detection technologies mainly include manual inspection, SCADA-based real-time operational monitoring, ultrasonic corrosion testing, and magnetic flux leakage inspection. These approaches enable the acquisition of basic pipeline status information, such as pressure, flow rate, temperature, and localized defect characteristics, and allow operators to identify abnormal fluctuations or apparent degradation phenomena.

However, in conventional applications, the data collected by SCADA systems and field sensors are primarily used for threshold-based alarming and post-event analysis, with limited capability for extracting latent correlations among multi-source variables or identifying complex, evolving risk patterns. As a result, although traditional systems provide essential raw data and first-level safety assurance, their potential for proactive risk assessment and early warning remains largely unrealized without the support of intelligent data-driven analysis methods.

The comprehensive method evaluation and risk warning generally adopt evaluation methods such as risk matrix method, accident tree analysis method, and analytic hierarchy process, covering various evaluation methods such as qualitative, semi quantitative, and quantitative. Wu Yanhua et al. [12] established a railway disaster monitoring and early warning system based on the risk matrix method. Zhu Lina [13] constructed a safety production accident risk warning model based on a risk index system combined with fuzzy evaluation. AYYILDIZ et al. [14] constructed a warehouse risk warning indicator system based on fuzzy logic and Bayesian networks. Wang Xin [15] established a real-time monitoring and early warning model for mine water inrush by constructing a BP neural network. Sun Yi et al. [16] analyzed the various methods used in the safety warning process of storage tank areas.

In view of the difficulty in determining early warning levels for station oil pump pipeline systems under the coupled effects of multiple operating conditions and environ-

mental factors, single-model approaches exhibit inherent limitations in feature representation and decision boundary construction, making it challenging to accurately characterize the complex nonlinear relationships governing warning level evolution. To address this issue, an ensemble learning framework is introduced to enhance the comprehensive representation of multi-source and multi-scale influencing factors. By integrating multiple base learners, ensemble learning effectively alleviates model bias and variance, thereby improving generalization performance in complex engineering scenarios.

On this basis, this study constructs an intelligent risk early warning model for station oil pump pipeline systems under soft soil foundation conditions by coupling the artificial bee colony (ABC) optimization algorithm with XGBoost. XGBoost (eXtreme Gradient Boosting) is an extreme gradient boosting tree algorithm, which belongs to a type of ensemble learning [17]. The primary contribution of this work lies in the application-oriented integration of ABC-optimized XGBoost tailored to the risk warning problem of oil pump pipelines in soft soil environments, rather than in proposing a new learning algorithm. Specifically, the ABC algorithm is employed to adaptively optimize key hyperparameters of XGBoost, including the number of iterations, learning rate, and tree depth, thereby improving the model's sensitivity to nonlinear coupling effects induced by soft soil foundation conditions [18]. The proposed framework enables more accurate and robust early warning of coupling risks in station oil pump pipeline systems, demonstrating its effectiveness for intelligent safety assessment in complex geotechnical environments.

The second part of this article introduces the construction process of the coupled risk warning model for the oil pump pipeline system in the station under soft soil foundation conditions. The third part introduces the process of optimizing XGBoost using the ABC algorithm. The fourth part focuses on conducting case studies to verify the accuracy of the model.

2. Risk Early Warning Model for Coupling Risk of Oil Pump Pipeline System in Station Under Soft Soil Foundation Conditions

In order to avoid stress deformation of the oil pump pipeline coupling system under external loads and ensure safe operation of the station, accurate judgment of the real-time status of the system should be made, and a system coupling risk warning model should be established.

2.1. Establishment of Risk Warning Indicator System

2.1.1. Stress Warning Indicators

The warning level of pipeline stress can be determined by using the stress value on the pipeline. This article divides stress values into two categories: stress values obtained through on-site monitoring and stress values calculated through numerical simulation.

The settlement displacement of the pipeline is displayed by displacement sensors, and the strain at any point on the circumference of the pipeline can be obtained by Formula (1). By taking the derivative and using Formulas (2) and (3), the maximum and minimum strains of the pipeline can be determined.

$$\varepsilon_{(x,y)} = \frac{\varepsilon_1 + \varepsilon_3}{2} + \frac{\varepsilon_3 - \varepsilon_1}{2} \cdot \frac{x}{r} - \frac{\varepsilon_1 + \varepsilon_3 - 2\varepsilon_2}{2} \frac{y}{r} \quad (1)$$

$$\frac{x}{r} = \pm \frac{\varepsilon_3 - \varepsilon_1}{\sqrt{2\varepsilon_1^2 + 2\varepsilon_3^2 - 4\varepsilon_1\varepsilon_2 - 4\varepsilon_2\varepsilon_3 + 4\varepsilon_2^2}} \quad (2)$$

$$\frac{y}{r} = \pm \sqrt{1 - \left(\frac{x}{r}\right)^2} \quad (3)$$

ε_1 : maximum tensile strain, ε_2 : axial strain, ε_3 : minimum compressive strain, x, y : direction vector.

Therefore, by converting monitoring values, the maximum stress of the pipeline section and the settlement displacement of the foundation can be obtained in real time. The maximum stress value on the pipeline cross-section reflects the actual strength state of the pipeline under current operating and geotechnical conditions.

The maximum von Mises stress of the pipeline is calculated by integrating real-time monitoring data, including oil pump vibration signals, internal pipeline pressure, and ground settlement displacement, transmitted via the on-site SCADA system from monitoring devices installed in the upstream section. These measured pressure and displacement data are used to dynamically update the boundary conditions of the numerical stress model, enabling time-synchronized stress estimation. To verify the consistency between monitoring data and numerical predictions, key stress-sensitive response indicators derived from vibration and pressure signals are compared with model-predicted responses.

The discrepancy between monitored responses and model-predicted stress is quantified using a normalized deviation indicator, defined as

$$D_t = \frac{\left| f_t^{mon} - \hat{f}(\sigma_t^{sim}) \right|}{\max(f_t^{mon}, \lambda)} \quad (4)$$

f_t^{mon} : measured stress-sensitive feature at time t , $\hat{f}(\sigma_t^{sim})$: corresponding response predicted from the simulated maximum von Mises stress σ_t^{sim} , λ : a small positive constant to avoid numerical instability.

Based on the magnitude of the discrepancy, a confidence weight factor is introduced as

$$\omega_t = \exp(-\alpha D_t) \quad (5)$$

α : a deviation sensitivity coefficient controlling the influence of data–model inconsistency. This weight is then used to regulate the contribution of the simulation-derived stress to the final risk assessment.

Specifically, the corrected risk index is expressed as

$$R_t^{final} = R_t^{base} + \omega_t \cdot R_t^{\sigma} \quad (6)$$

R_t^{base} : baseline risk index obtained from the data-driven early warning model, R_t^{σ} : the stress-related risk component derived from the numerical simulation.

Through this mechanism, larger discrepancies between monitoring data and numerical predictions lead to a conservative adjustment of the stress contribution, thereby preventing underestimation of potential hazards. Under this integrated framework, the maximum von Mises stress obtained from numerical simulation conveys the theoretical stress state of the pipeline in real time and serves as a reliable supplementary stress indicator when on-site stress sensors are unavailable or temporarily fail. By explicitly quantifying data–model discrepancies and incorporating them into the risk evaluation process, the proposed approach reduces uncertainty propagation and enhances the robustness and credibility of pipeline safety assessment under complex operating conditions.

2.1.2. Resonance Warning Indicators

When the vibration frequency of the oil pump approaches the natural frequency of the pipeline, the oil pump pipeline system will produce a strong vibration response with a sharp increase in amplitude. Extract the natural frequencies of the oil pump and pipeline,

calculate the frequency ratio between the oil pump and pipeline, reflect the resonance safety status of the system, and determine the resonance warning level.

2.1.3. Warning Level Determination

Obtain data on pipeline internal pressure, ground settlement displacement, pipeline amplitude, pipeline stress, vibration frequency of oil pumps, and pipeline vibration frequency to calculate stress warning level and resonance warning level. And any warning indicator reaching the warning value represents that the station oil pump pipeline system has met the conditions for failure. Therefore, the total warning level of the system is set to the highest level among the three warning indicators mentioned above. The calculation process is shown in Figure 1.

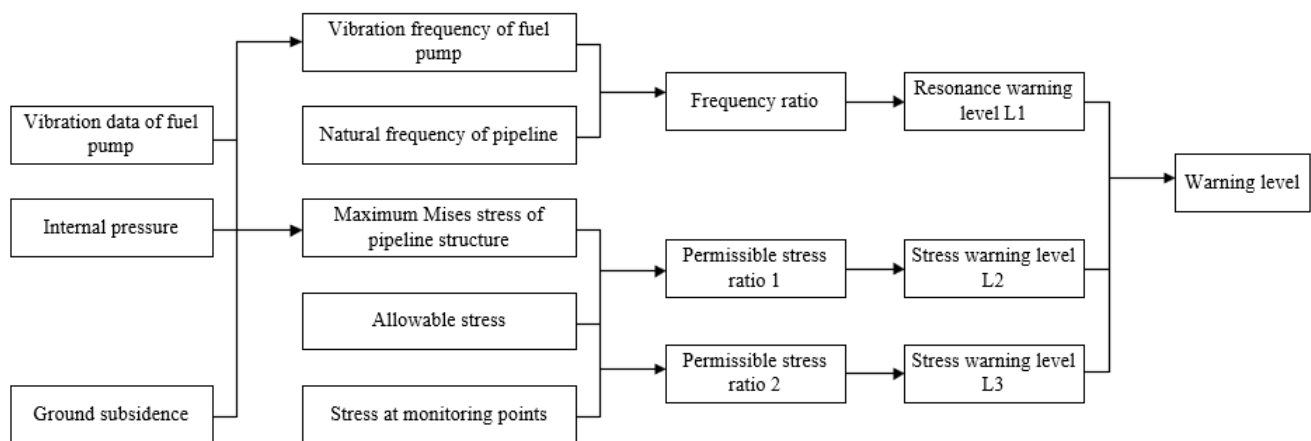


Figure 1. Pipeline risk early warning flow chart of soft soil foundation station.

Calculate the system warning level based on the resonance warning level and stress warning level calculated for each part using the following formula:

$$L = \max\{L_1, L_2, L_3\} \quad (7)$$

L: warning level, L1: resonance warning level, L2: monitoring stress warning level, L3: finite element calculation of stress warning level.

$$\begin{cases} 0.8f < f_p < 1.2f \\ 0.8f_p < f < 1.2f_p \end{cases} \quad (8)$$

f : Vibration frequency of fuel delivery pump, Hz; f_p : The natural frequency of the pipeline, Hz.

Define frequency ratio γ :

$$\gamma = \frac{f_p}{f} \quad (9)$$

Based on the recommendations provided in ISO 10816 [19] and API 618 [20], the resonance risk is categorized into three levels based on the frequency ratio:

$$\begin{cases} \gamma < 0.8 \text{ or } \gamma > 1.2 & \text{Safety} \\ 0.8 \leq \gamma < 0.9 \text{ or } 1.1 < \gamma \leq 1.2 & \text{Warning} \\ 0.9 \leq \gamma \leq 1.1 & \text{Dangerous} \end{cases} \quad (10)$$

2.2. Soft Soil Foundation Station Oil Pump Pipeline Warning Model Based on ABC-XGBoost Algorithm

The overall process of the ABC-XGBoost soft soil foundation base station pipeline risk warning model is shown in Figure 2.

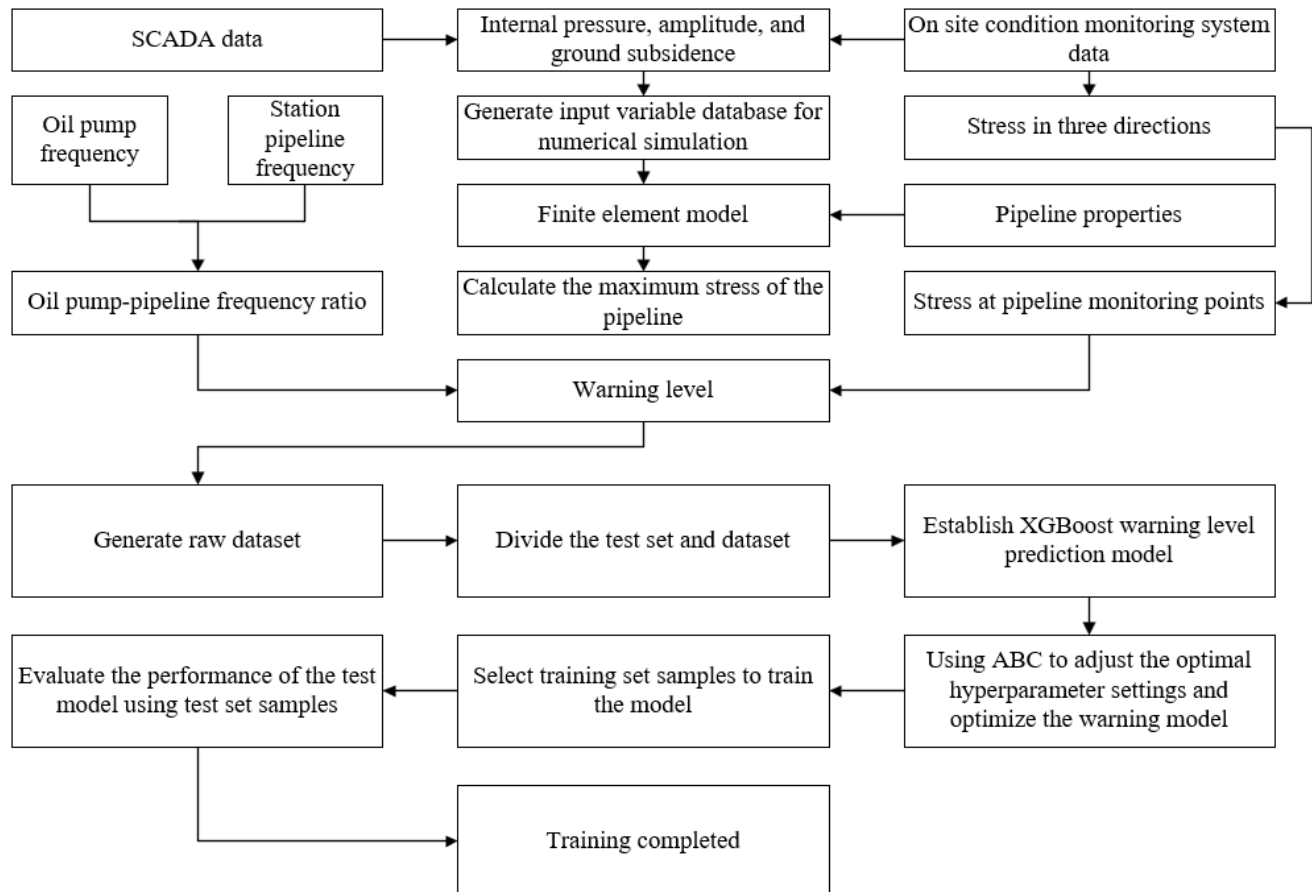


Figure 2. Early warning model construction process.

3. Optimize Hyperparameters of XGBoost Model

The hyperparameters of the XGBoost model, including the maximum number of iterations, tree depth, and learning rate, are optimized using the global search capability of the artificial bee colony (ABC) algorithm. The search ranges of these hyperparameters are determined based on a combination of empirical experience from related studies and preliminary exploratory experiments, ensuring that the selected intervals cover both conservative and aggressive learning configurations while avoiding unrealistic parameter settings.

The population of the ABC algorithm is initialized such that each individual corresponds to a candidate set of XGBoost hyperparameters, and the maximum number of ABC iterations is set to 100 to balance optimization effectiveness and computational efficiency. During the iterative optimization process, employed bees, onlooker bees, and scout bees adjust the hyperparameter combinations according to their respective search strategies, and the updated parameters are subsequently applied to train the XGBoost model, forming the ABC-XGBoost risk early warning framework for station pipeline systems under soft soil foundation conditions.

To evaluate the robustness of the proposed model with respect to the selected hyperparameter ranges, a sensitivity analysis is conducted by moderately expanding and

contracting the predefined intervals. The results indicate that the model performance exhibits limited sensitivity to small variations in the hyperparameter ranges, confirming the adequacy of the selected search spaces and the stability of the optimization process. The overall procedure of XGBoost hyperparameter optimization using the ABC algorithm is illustrated in Figure 3.

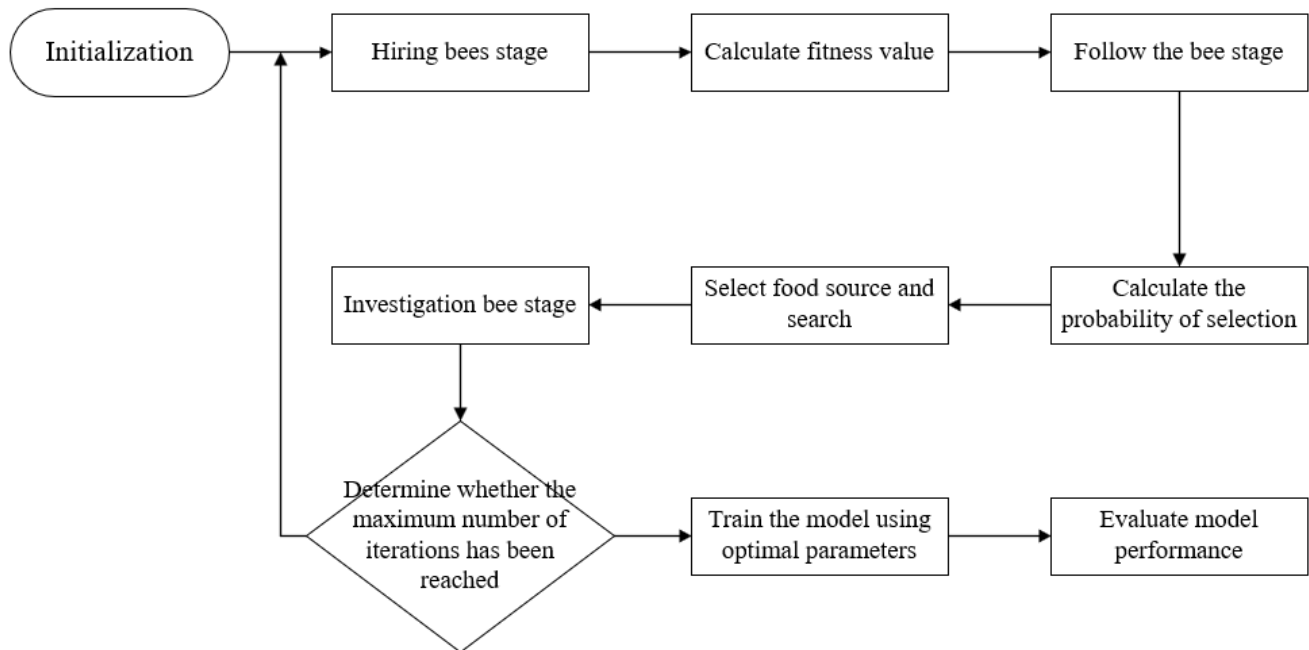


Figure 3. The process of ABC optimizing XGBoost hyperparameters.

In the process of building decision trees, XGBoost uses greedy algorithms to traverse features and splitting points, calculate the splitting method with the highest selection gain, and to process large-scale data and improve efficiency; it also uses feature parallel computing to increase the speed of data training.

XGBoost is an ensemble model based on classification regression trees, as shown in Equation (11):

$$\begin{cases} \hat{y}_i = \sum_{k=1}^T f_k(x_i), f_k \in F \\ F = \left\{ f(x) = w_{q(x)} \right\} (q: R^m \rightarrow L, w \in R^T) \end{cases} \quad (11)$$

In the formula, T: the number of trees; L: the number of leaf nodes in the tree; q: the structure of trees; w: leaf node weight vector.

As shown in Equation (12), the XGBoost objective function consists of a loss function and a regularization term. The loss term l is used to measure the difference between the predicted value and the target value, while the penalty term Ω can limit the model complexity and avoid overfitting.

$$\begin{cases} Obj = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \\ \Omega(f) = \gamma L + \frac{1}{2} \lambda ||w||^2 \end{cases} \quad (12)$$

The training adopts an iterative approach, and the iterative update method is shown in Equation (13):

$$\hat{y}_i^{(t)} = \sum_{k=1}^t f_k(x_i) = y_i^{(t-1)} + f_i(x_i) \quad (13)$$

Therefore, the objective function for predicting the value $\hat{y}_i^{(t)}$ of the t -th iteration of x_i is Formula (14), and the second-order Taylor expansion is shown in Formula (15):

$$Obj^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_i(x_i)) + \Omega(f_i) \quad (14)$$

$$Obj^{(t)} = \sum_{i=1}^n \left[l(y_i, \hat{y}_i^{(t-1)}) + g_i f_i(x_i) + \frac{1}{2} h_i f_i^2(x_i) \right] + \Omega(f_i) \quad (15)$$

Therefore, the objective function after removing the constant term is shown in Equation (16), the first derivative is shown in Equation (17), and the second derivative is shown in Equation (18):

$$Obj^{(t)} \simeq \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)}) + g_i f_i(x_i) + \frac{1}{2} h_i f_i^2(x_i) + \Omega(f_i) \quad (16)$$

$$g_i = \partial_{y_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)}) \quad (17)$$

$$h_i = \partial_{y_i^2}^2 l(y_i, \hat{y}_i^{(t-1)}) \quad (18)$$

Therefore, the objective function for the t -th iteration is

$$\begin{cases} Obj^{(t)} \simeq \sum_{i=1}^n \left[g_i f_i(x_i) + \frac{1}{2} h_i f_i^2(x_i) \right] + a(f_i) \\ Obj^{(t)} \simeq \sum_{i=1}^n \left[g_i w_{q(x_i)} + \frac{1}{2} h_i w_{q(x_i)}^2 \right] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^t w_j^2 \\ Obj^{(t)} \simeq \sum_{i=1}^n \left[\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma T \end{cases} \quad (19)$$

If I_j is defined as the set on each leaf node, then the problem of finding the best predicted value and optimal objective function for node j can be transformed into an extremum problem of a quadratic equation, as shown in Formula (20):

$$\begin{cases} G_j = \sum_{i \in I_j} g_i \\ H_j = \sum_{i \in I_j} h_i \end{cases} \rightarrow \begin{cases} w_j^* = -\frac{G_j}{H_j + \lambda} \\ Obj_j^* = -\frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} \end{cases} \quad (20)$$

The gain function for segmentation is defined as Equation (21), and the gain is used to determine whether to segment and the candidate nodes for segmentation.

$$Gain = \frac{1}{2} \left(\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right) - \lambda \quad (21)$$

During the training process of the XGBoost model, hyperparameters such as iteration count, learning rate, and tree depth have a significant impact on the model's predictive performance. If the number of iterations is too small, it can easily lead to underfitting of the model and insufficient learning of data features; on the contrary, if the number of iterations is too high, it may lead to overfitting of the model and overfitting to the noise and details in the training data, thereby reducing its generalization ability. The learning rate is too low, and the model steps slowly during gradient descent, resulting in a significant decrease in convergence speed; when the learning rate is too high, the model is prone to crossing the optimal value during gradient descent, making it difficult to converge to the ideal solution.

The ABC optimization algorithm consists of five steps.

- (1) Generate initial population:

$$x_{i,j} = x_j^{\min} + \text{rand}(0,1) \cdot (x_j^{\max} - x_j^{\min}) \quad (22)$$

In the formula, i : the sequence number of the solution; j : dimensions of solution; $[x_j^{\max}, x_j^{\min}]$: the upper and lower limits of the j -th dimension of the decision variable; $\text{Rand}(0,1)$: a uniformly distributed random number in the closed interval $[0, 1]$.

- (2) Dispatch hired bees to search near the honey source and generate a new honey source according to Equation (20):

$$v_{i,j} = x_{i,j} + \phi_{i,j} \cdot (x_{i,j} - x_{k,j}) \quad (23)$$

In the formula, v_i : the new solution generated by the neighborhood of the current solution; x_k : a randomly selected solution from the population; $\phi_{i,j}$: scalability factor.

- (3) Follow the bees through roulette wheel to select promising honey sources for further search. See Equations (21) and (22):

$$p_i = \frac{\text{fit}(X_i)}{\sum_{i=1}^{SN} \text{fit}(X_i)} \quad (24)$$

$$f = \begin{cases} \frac{1}{1+f_i} & f_i \geq 0 \\ 1 + |f_i| & f_i < 0 \end{cases} \quad (25)$$

- (4) When the honey source has not been updated after multiple iterations, it is considered to have been mined and a new honey source is initialized in space using Equation (18).
- (5) In the termination condition stage, if the maximum number of iterations is reached or the fitness converges to a certain accuracy, the iteration is stopped. Otherwise, steps 2 to 5 are repeated.

4. Case Study

4.1. Dataset Preparation

The selection of input and output parameters is crucial when building predictive models. The correct input parameters can improve the predictive ability of the model, while the output parameters are the target of model prediction. Considering that the warning level characteristics are derived from measured data at the station, the dataset is formed by selecting internal pressure, amplitude, ground settlement, stress 1, stress 2, stress 3, oil pump vibration frequency, and pipeline natural frequency as inputs, and safety level as output parameters. Encode the security level as shown in Table 3.

Table 3. Security level coding.

Status	Encoding	Status	Encoding	Status	Encoding
Safety	1	Warning	2	Dangerous	3

4.2. Pipeline Model and Dataset Construction

According to the actual operating conditions of the station, a multi-factor finite element model is established by explicitly incorporating the soft soil foundation, support pier constraints, pipeline structure, and oil pump loading. The model is employed to simulate

the stress state of the pipeline under three representative settlement scenarios, including uniform foundation settlement, local settlement of the buried pipeline foundation, and localized settlement at the pipeline–oil pump connection.

To ensure the reproducibility and numerical reliability of the simulation results, the finite element model is constructed using structured meshing for the pipeline and refined local meshes in settlement-sensitive regions, while appropriate boundary conditions are applied to represent support pier constraints and soil–structure interaction. The oil pump load and internal pressure are imposed as external and internal loads, respectively. Mesh convergence is verified by monitoring the stability of the maximum von Mises stress with progressive mesh refinement, and the solution is accepted when further refinement results in negligible stress variation. Based on the validated model, characteristic stress data are extracted for subsequent analysis.

Calculate the risk warning level of station pipeline under different states using on-site feature data. Construct 200 pipeline models from the input parameter set, and combine them with other parameters to create a total of 766 sets of data. The feature data sample table is shown in Table 4.

Table 4. Data sample table.

Feature 1	Feature 2	Feature 3	Feature 4	Feature 5	Feature 6	Feature 7	Feature 8	Category
347.00	418.00	3.79	1.32	11.85	54.37	73.84	46.47	2
55.00	845.00	3.77	3.86	2.84	97.75	70.24	43.20	1
375.00	552.00	3.71	4.50	3.13	67.13	1.60	108.63	1
367.00	814.00	3.25	5.13	13.00	103.96	112.00	18.16	1
285.00	316.00	3.82	2.35	7.80	53.96	67.05	112.09	3

In Table 4, feature 1 represents pipeline frequency, feature 2 represents oil pump frequency, feature 3 represents internal pressure, feature 4 represents amplitude, feature 5 represents settlement, feature 6 represents stress data 1, feature 7 represents stress data 2, and feature 8 represents stress data 3.

4.3. Data Training

In order to eliminate dimensional differences and accelerate model convergence, the feature data is normalized. Randomly allocate 70% of the dataset as the training set and 30% of the dataset as the testing set. The normalized dataset was input into the XGBoost-based risk warning model for training, with a training set accuracy of 100% and a testing set accuracy of 95.21%, as shown in Figure 4. By using the artificial bee colony optimization algorithm to optimize the hyperparameters of the XGBoost model, the hyperparameters of the bee colony optimization algorithm and the training results of the ABC XGBoost risk warning model were obtained, as shown in Figure 5. X-axis: predicted samples, Y-axis: predicted results. The detailed comparison of the training effectiveness of risk early warning models is presented in Table 5.

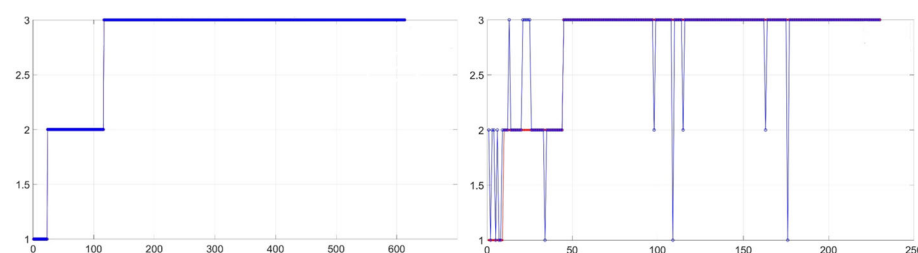


Figure 4. The training effect of risk early warning model based on XGBoost.

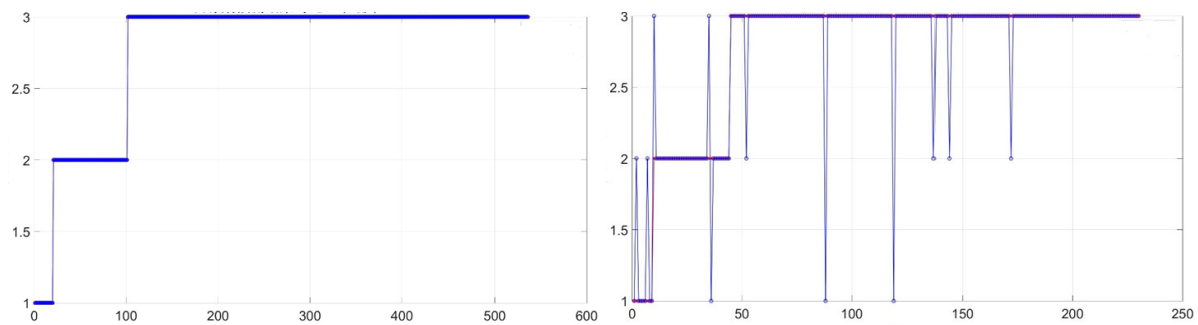


Figure 5. The testing effect of risk early warning model based on ABC-XGBoost.

Table 5. Comparison of training effectiveness of risk warning models.

Types of Risk Warning Models	Test Set Prediction Accuracy
XGBoost	92.60%
ABC-XGBoost	95.21%

4.4. Result Analysis

The training accuracy of the risk warning model based on ABC-XGBoost reaches 95.22%, representing an improvement of 2.61% compared with the unoptimized XGBoost-based model, which demonstrates the effectiveness of hyperparameter optimization. As shown by the confusion matrix, the training dataset consists of 536 samples with an overall classification accuracy of 100%, while the test dataset contains 230 samples with an accuracy of 95.22%. Specifically, two safety samples are misclassified as warning states; among the warning samples, one is misclassified as safe and two are correctly identified as danger; among the danger samples, two are misclassified as safe and four are correctly identified.

The multi-factor finite element model constructed based on the actual operating conditions of the station effectively captures the stress and response characteristics of the pipeline system under the coupled effects of soft soil foundation, pier constraints, pipeline structure, and oil pump loading. The multidimensional feature data extracted under different settlement scenarios—including uniform foundation settlement, local settlement of buried pipeline sections, and localized settlement at the pipeline–oil pump connection—provide a reliable data basis for the training and validation of the subsequent risk warning models. By normalizing frequency, internal pressure, vibration amplitude, settlement, and multi-source stress features, the influence of dimensional inconsistency on model training is eliminated, thereby improving model convergence and numerical stability.

The initial XGBoost-based risk warning model exhibits strong fitting capability on the training dataset and achieves high recognition accuracy on the test dataset, indicating that the selected feature set can effectively characterize the pipeline risk state. On this basis, the introduction of the artificial bee colony algorithm for global hyperparameter optimization further enhances the classification performance of the ABC-XGBoost model, leading to a notable improvement in test accuracy and overall recognition capability.

An analysis of the classification errors provides additional engineering insights. The misclassification between safety and warning states is mainly attributed to the overlap of feature distributions under early-stage abnormal conditions, where variations in vibration amplitude and internal pressure remain within moderate ranges and have not yet resulted in significant stress concentration. Similarly, the confusion between warning and danger states in a small number of samples is primarily caused by localized settlement conditions, under which the stress response exhibits strong nonlinearity and temporal variability, making the boundary between warning and danger states less distinct. Furthermore,

uncertainties in monitoring data, such as sensor noise and short-term fluctuations in operating conditions, may also contribute to occasional misjudgments.

Despite these misclassifications, the overall error rate remains within an acceptable range for engineering applications. Overall, the proposed ABC–XGBoost risk warning model demonstrates good generalization ability and practical applicability under multi-condition and multi-feature coupling scenarios, and can provide effective technical support for graded early warning and safety management of pipeline operational risks in station environments.

5. Conclusions

This study proposes an intelligent risk early warning model for oil pump and pipeline systems in soft soil foundation stations, taking into account the characteristic operating and geotechnical conditions of such environments. By employing the artificial bee colony (ABC) algorithm to optimize key hyperparameters of the XGBoost model, including the maximum number of iterations, tree depth, and learning rate, the prediction performance of the risk warning model is effectively improved, enabling reliable discrimination of risk warning levels for oil pump pipeline systems under complex operating conditions.

From an engineering application perspective, the proposed model integrates numerical simulation results, finite element stress responses, and real-time monitoring data to support data-driven safety state identification and early warning of pipeline systems during operation. It should be noted that the current study is based on a relatively limited dataset, which may introduce potential risks of overfitting and constrain the generalization capability of the model. However, through parameter optimization and validation procedures, the model demonstrates stable performance and a consistent ability to capture the dominant risk patterns associated with soft soil foundation effects and multi-condition coupling.

At the present stage, the size of the dataset primarily affects the reliability of early warning in boundary cases and rare operating conditions, where insufficient samples may limit the model's sensitivity to extreme or evolving risk states. Nevertheless, the results indicate that the proposed framework is suitable for feasibility validation and comparative risk identification under existing data conditions. With the continuous accumulation of long-term monitoring data and the enrichment of operational scenarios, the robustness and reliability of the model are expected to be further enhanced, allowing the framework to evolve toward a more comprehensive lifecycle-oriented safety assessment. Ultimately, this study provides a practical and extensible technical basis for intelligent safety management and risk early warning of oil pump pipeline systems in station environments.

Author Contributions: Conceptualization, S.D. and Q.X.; methodology, S.D. and Q.X.; software, L.C.; validation, S.Y., X.F. and L.C.; formal analysis, Liwen Chen; investigation, S.Y.; resources, S.Y.; data curation, S.Y.; writing—original draft preparation, S.Y. and X.F.; writing—review and editing, S.D. and Q.X.; visualization, L.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by China Petroleum Science and Technology Innovation Fund, grant number 2021DQ02-0801.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: Author Shengyang Yu was employed by the company China National Oil and Gas Exploration and Development Corporation. Author Liwen Chen was employed by the company Hefei General Machinery Research Institute Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Yan, Z. Study on Consolidation Settlement Deformation Law of Foundation Soil at Magang Raw Material Yard along the Yangtze River. Master's Thesis, Guilin University of Technology, Guilin, China, 2023.
2. Zhang, Y.N. Safety Evaluation of Station Pipeline Settlement Based on Multi-Source Data Fusion. Master's Thesis, China University of Petroleum (Beijing), Beijing, China, 2023. [CrossRef]
3. Cui, M.; Cao, X. Comparison of residual strength assessment methods for corroded pipelines with different steel grades. *Oil Gas Storage Transp.* **2012**, *31*, 486–490.
4. Ye, D.; Ye, Q.; Luo, W. Calculation and implementation of hazardous chemical leakage areas based on Gaussian diffusion model. *Comput. Appl. Chem.* **2012**, *29*, 195–199. [CrossRef]
5. Zhan, D.; Ma, X. Finite element analysis and prediction of stress in coastal buried natural gas pipelines under non-uniform settlement. *Hot Work. Technol.* **2023**, *in press*. [CrossRef]
6. Lai, L.N.; Lin, T.; Hao, D.; Ma, X. Intelligent prediction of stress in offshore gas pipelines under non-uniform settlement. *Oil-Gasfield Surf. Eng.* **2022**, *41*, 83–90.
7. Ding, S.; Li, F.; Han, S.; Zhou, L.; Dong, W. Stress prediction analysis of ship structures based on LSTM-GRU model. *Chin. J. Ship Res.* **2023**, *64*, 146–158.
8. Wu, Z.; Zhou, L.; Leng, J. Pipeline stress prediction method based on particle swarm optimized LSTM model. *Press. Vessel Technol.* **2021**, *38*, 76–80.
9. Zhang, Z.; Ye, L.; Qin, H.; Liu, Y.; Wang, C.; Yu, X.; Yin, X.; Li, J. Wind speed prediction method using shared-weight long short-term memory network and Gaussian process regression. *Appl. Energy* **2019**, *247*, 270–284. [CrossRef]
10. Tan, Z.; Guo, X.L.; Li, J.; Guo, Y.; Pan, J. Pipeline leakage detection model based on multi-scale convolutional neural networks. *J. Hydraul. Eng.* **2023**, *54*, 220–231. [CrossRef]
11. Abdulnaser, M.; Hitham, A.; Said, J.; Jagadeesh, A. Prediction of oil and gas pipeline failures through machine learning approaches: A systematic review. *Energy Rep.* **2023**, *10*, 1313–1338. [CrossRef]
12. Wu, Y.; He, F.; Wang, F.; Li, P. Research on railway disaster monitoring and early warning system based on disaster risk assessment model. *China Railw. Sci.* **2012**, *33*, 121–125.
13. Zhu, L. Risk Early Warning of Safety Production Accidents in Power Generation Enterprises Based on Bayesian Networks. Master's Thesis, North China Electric Power University, Beijing, China, 2019. [CrossRef]
14. Ayyildiz, E.; Erdogan, M.; Gul, M. A comprehensive risk assessment framework for occupational health and safety in pharmaceutical warehouses using Pythagorean fuzzy Bayesian networks. *Eng. Appl. Artif. Intell.* **2024**, *135*, 108763. [CrossRef]
15. Wang, X. Theoretical Research on Real-Time Monitoring and Early Warning of Mine Water Inrush. Ph.D. Thesis, China University of Mining and Technology, Xuzhou, China, 2020. [CrossRef]
16. Sun, Y.; Peng, S.T.; Wang, X.L.; Guan, W. Research progress on safety early warning of tank areas in petrochemical terminals. *J. Waterw. Harb.* **2014**, *35*, 438–444.
17. Lian, K.Q. Research and Analysis of Boosting-Based Ensemble Tree Algorithms. Master's Thesis, China University of Geosciences (Beijing), Beijing, China, 2018. Available online: https://kns.cnki.net/kcms2/article/abstract?v=A-1EuXenf_puy1Gdu37K49sGx9s1MG8RPz4Ixt05BWqkE8Ilve6NXBAWYf4RbFURIQKTqdaGqE8PmyounoVLJwWp9g6EFey4QeS4cl-kc-P1ys5ruiwe77j0il8v1Nz4zhNYONnQ1UhnPrK4qNSdbpBHU6UkKYIstmFXo7mwTo8=&uniplatform=NZKPT (accessed on 25 July 2025).
18. Lin, S.; Dong, C.; Chen, M.; Zhang, F.; Chen, J. Review of novel swarm intelligence optimization algorithms. *Comput. Eng. Appl.* **2018**, *54*, 1–9.
19. ISO 10816-3: 2014; Mechanical Vibration–Evaluation of Machine Vibration by Measurements on Non-Rotating Parts. International Organization for Standardization: Geneva, Switzerland, 2014.
20. API 618-2019; Reciprocating Compressors for Petroleum, Chemical, and Gas Industry Services. American Petroleum Institute: Washington, DC, USA, 2019.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.