



ARTICLE

# Interpretable Damage State Identification of Buried Pipelines under Rotary Tiller Loading Using a PSO–CatBoost Framework

Liqiong Chen<sup>1</sup>, Haoyu Jia<sup>1</sup>, Mailun Liu<sup>2</sup>, Kai Zhang<sup>1,\*</sup>, Song Yang<sup>1</sup> and Zongjun Jiang<sup>1</sup>

<sup>1</sup>Petroleum Engineering School, Southwest Petroleum University, Chengdu, China

<sup>2</sup>National Petroleum and Natural Gas Co., Ltd. Oil and Gas Regulation Center, Beijing, China

\*Corresponding Author: Kai Zhang. Email: 202421001031@stu.swpu.edu.cn

Received: 15 December 2025; Accepted: 08 April 2026; Published: 30 June 2026

**ABSTRACT:** Buried natural gas pipelines are critical components of energy infrastructure, and their durability and safe operation depend on effective structural health monitoring and the early identification of damage states. In farmland environments, rotary tillage imposes repeated and often concealed mechanical loads on buried pipelines, resulting in stress accumulation, progressive deterioration, and potentially structural failure. However, predictive and interpretable health monitoring approaches that explicitly incorporate rotary tiller-induced damage mechanisms remain scarce. In this study, a physics-informed and interpretable hybrid framework is proposed for the structural health monitoring of buried pipelines subjected to rotary tiller loading. A three-dimensional multiphysics-coupled finite element model of the rotary tiller-pipeline-soil system was developed to simulate the mechanical response and damage evolution of pipelines under varying wall thickness, internal pressure, blade number, operating speed, and soil density. Based on the simulation results, pipeline conditions were classified into three damage states, namely elastic deformation, plastic deformation, and failure, with the first two regarded as warning states. A multi-class CatBoost model optimized using Particle Swarm Optimization (PSO) was subsequently established for damage-state identification. On the test set, the model achieved an accuracy of 0.94, and the AUC values for all three classes reached 0.93. SHapley Additive exPlanations (SHAP) were further employed to interpret the model outputs and quantify the contribution of individual parameters. The results revealed critical risk thresholds associated with the transition from warning states to failure under the shallow-cover rotary tiller disturbance scenario considered in this study. In particular, the risk of failure increased markedly when the blade number exceeded eight and the internal pressure was greater than 8 MPa. These findings indicate that wall thickness and internal pressure govern the baseline structural resistance of pressurized pipelines, while the identified thresholds can support the screening of high-risk conditions and the operational control of shallow-cover farmland sections.

**KEYWORDS:** Structural health monitoring; buried pipelines; third-party interference; damage state identification; finite element analysis; explainable machine learning

## 1 Introduction

Driven by the growth in oil and gas demand, long-distance pipelines have become a vital infrastructure for national energy supply; by the end of 2022, the total mileage of oil and gas pipelines in China reached 155,000 km, including 93,000 km of natural gas pipelines [1]. To further enhance energy security and network coverage, the national plan projects that the total mileage of China's oil and gas pipelines will reach 240,000 km by the end of 2025, marking a substantial increase in infrastructure scale. However, the recurrent soil disturbance from agricultural machinery such as rotary tillers poses a significant risk to buried

pipelines, potentially leading to gas leaks. Such incidents substantially elevate explosion hazards, thereby directly endangering public safety in surrounding areas [2,3]. Currently, there remains a significant research gap concerning pipeline damage mechanisms and protective technologies within these specific operational contexts. Further investigation is imperative to underpin precise risk assessment, enable effective preemptive monitoring and early warning, and facilitate the development of robust emergency response frameworks.

Research on pipeline third-party damage frequently employs finite element simulation to examine the associated mechanical response. For example, Jing [4] employed ANSYS/LS-DYNA to simulate buried pipeline responses subjected to rockfall impacts. Jiang and Zhao [5] developed a three-dimensional stochastic large-deformation model to explore the effects of soil spatial variability on pipeline behavior under impact loading. Yao et al. [6] pioneered in 2009, the development of mechanical and finite element analysis models for the multi-system interaction involving rolling rock, soil, and buried steel pipeline, utilizing a three-dimensional dynamic contact surface algorithm. This work established a foundational model for investigating the dynamic response of this complex system under impact loading. Subsequently, Zhao et al. [7] formulated in 2016 a model for pipeline puncture damage induced by excavator operation, aiming to predict the puncture load on gas pipelines. The outcomes of this study provided a theoretical and empirical basis for quantitatively assessing the puncture threat posed to gas pipelines by excavation equipment. The application of advanced technologies, notably machine learning, has seen growing prevalence within the domain of pipeline risk management and control. In 2023, Hong et al. [8] incorporated the Particle Swarm Optimization (PSO) algorithm to refine the parameter selection in Probabilistic Neural Network (PNN) models. This approach established a technical basis for improving the accuracy of evaluating the resilience of natural gas pipelines to third-party damage. Xiao et al. [9] employed Bayesian algorithms to optimize neural networks, enabling accurate prediction of the failure probability in corroded pipelines. Hu [10] pioneered, in 2024, an interpretable ensemble framework leveraging SHAP values and accumulated local effects for pipeline corrosion rate prediction. Through a comparative analysis of multiple machine learning methods, Liu et al. [11] identified CatBoost as the top performer across tasks including pipeline defect detection, size prediction, and growth rate estimation. Xiang and Zhou [12] developed a Bayesian network model for estimating pipeline third-party damage probability, thereby enhancing pipeline integrity management via data-driven methodologies. In addition, recent studies have further advanced the understanding of crack propagation and mitigation in steel gas pipelines. For example, Zhangabay et al. [13] investigated temperature-dependent crack propagation behavior and corresponding crack-arrest methodologies in steel gas pipelines, highlighting the important role of thermal conditions and reinforcement parameters in controlling crack evolution. Zhangabay et al. [14] also proposed a composite-overlay-based method for suppressing crack propagation in steel gas conduits under pressure surge conditions. These studies provide valuable insights into the protection of the structural integrity of cracked steel gas pipelines. However, the present study focuses on rotary tiller-induced external disturbances to buried pipelines, with particular emphasis on damage-state identification, predictive monitoring, and interpretable risk assessment in shallow-cover farmland scenarios. Although existing methods offer useful perspectives for risk assessment, the selection and optimization of assessment approaches for scenario-specific conditions, such as rotary tiller disturbances, still require further investigation. Research on third-party pipeline damage has predominantly focused on damage caused by mechanical excavation, with primary attention given to the associated failure mechanisms. In contrast, studies specifically addressing rotary tillers, particularly pipeline damage induced by their rotational impact during operation, remain very limited.

To address these gaps, this study employs finite element simulation to develop a rotary tiller-pipeline-soil model for factor analysis, in which pipeline response and terminal failure states are identified based on the maximum pipeline stress. Within the proposed framework, elastic deformation and plastic deformation

are regarded as warning states, whereas pipeline failure represents the terminal rupture state. Given the computational complexity and time-consuming nature of finite element simulations in practical applications, a Categorical Boosting (CatBoost) model trained on simulation data is introduced. This model enables accurate prediction and supports mechanistic analysis of pipeline damage and failure states. Compared with conventional decision tree models, CatBoost provides higher predictive accuracy owing to its built-in mechanisms for preventing target leakage and its other algorithmic advantages. To further improve model performance, its hyperparameters are optimized using the Particle Swarm Optimization (PSO) algorithm. In addition, the SHapley Additive exPlanations (SHAP) method is employed to interpret the CatBoost model, thereby quantifying the contribution of individual feature variables to the predictive results. The resulting PSO-CatBoost hybrid model leverages readily available data to deliver highly accurate predictions while retaining interpretability through SHAP. In this way, it effectively alleviates the limitations of conventional numerical simulations, particularly their high computational cost and limited practical adaptability.

## 2 Research Methodology

### 2.1 Finite Element Model Development and Validation

#### 2.1.1 Finite Element Model Development

Grounded in the theoretical framework for pipeline impact failure [15], this study utilizes finite element simulation to analyze the impact loads sustained by pipelines under actual operating conditions. The von Mises yield criterion is adopted to characterize the stress-based state transition of the pipeline. In the present classification framework, pipeline failure denotes the terminal rupture/loss-of-integrity state rather than an intermediate warning level. To address the nonlinear contact interaction problem between the rotary tiller blade and the buried natural gas pipeline, a three-dimensional finite element model comprising the overburden soil layer, the gas pipeline, and the rotary tiller blade is constructed. The overburden soil layer is represented by a homogeneous solid model. The pipeline is modeled utilizing thin-shell theory, while the rotary tiller blade is geometrically modeled based on its actual operational parameters. The mechanical response characteristics of the soil-pipeline-blade system are investigated via multiphysics-coupled simulation.

Considering that the compressive yield strength of soil is much higher than its tensile strength [16], the D-P model [17] is adopted in this study to characterize the overburden soil properties, owing to its capability to adequately simulate large elastoplastic deformations while ensuring numerical stability. The soil parameters are detailed in Table 1. The subject of this study is an X70 steel pipeline located in an agricultural farmland section along the Shanxi-Beijing route, where hemp yam cultivation is practiced and crawler-type rotary tillers are commonly used in field operations. The relevant pipeline parameters are listed in Table 2. For the simulation, the gas transmission pipeline is modeled using an isotropic elastoplastic material model that follows the von Mises yield criterion. A crawler-type rotary tiller, as depicted in Fig. 1, is therefore selected as a representative agricultural machine for the model. The geometry is simplified, retaining only the front tracks and the blades. The parameters pertaining to the blades are summarized in Table 3.

Table 1: Soil parameters.

| Density<br>(kg/m <sup>3</sup> ) | Elastic<br>Modulus (kPa) | Poisson's<br>Ratio | Internal Friction Angle<br>of D-P Model (°) | Flow Stress<br>Ratio | Dilation<br>Angle (°) |
|---------------------------------|--------------------------|--------------------|---|----------------------|-----------------------|
| 1850                            | 20,000                   | 0.4                | 28.7  | 1.0                  | 0                     |

**Table 2:** X70 steel pipeline parameters.

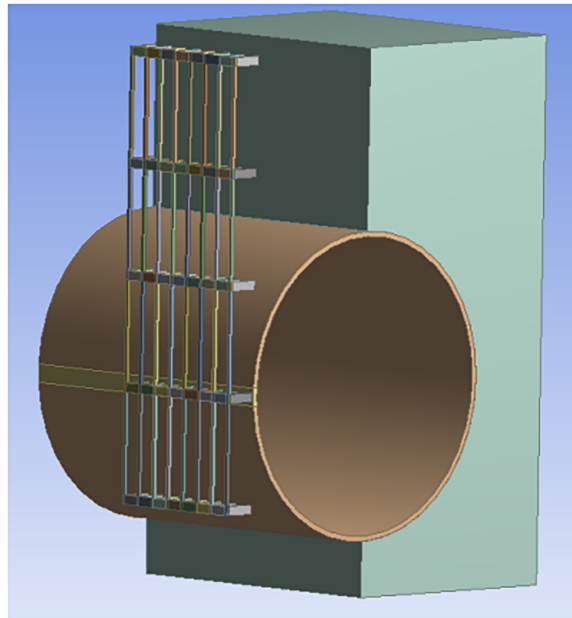
| Outer Diameter<br>(mm) | Wall Thickness<br>(mm) | Elastic Modulus<br>(GPa) | Yield Strength<br>(MPa) | Ultimate<br>Strength (MPa) | Poisson's<br>Ratio |
|------------------------|------------------------|--------------------------|-------------------------|----------------------------|--------------------|
| 1016                   | 17.5                   | 206                      | 487                     | 584                        | 0.3                |

**Figure 1:** The crawler-type rotary tiller.**Table 3:** Rotary tiller blade parameters.

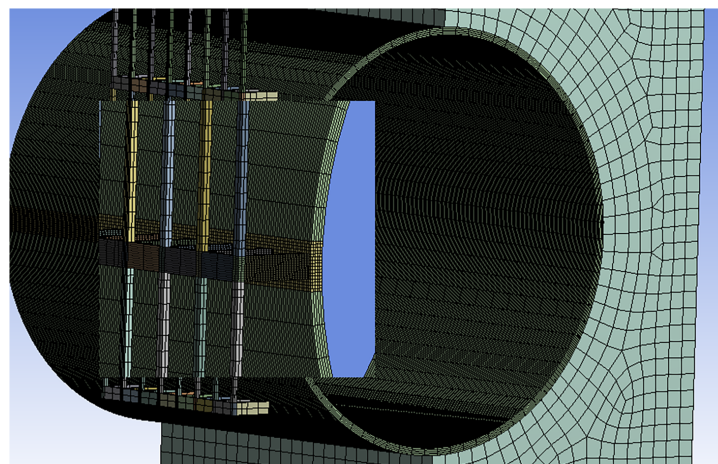
| Number of<br>Blade Rows | Blades per<br>Row | Blade Length<br>(mm) | Blade Width<br>(mm) | Blade Height<br>(mm) | Blade Spacing<br>(mm) | Density<br>(kg/m <sup>3</sup> ) | Elastic<br>Modulus<br>(GPa) |
|-------------------------|-------------------|----------------------|---------------------|----------------------|-----------------------|---------------------------------|-----------------------------|
| 5                       | 5                 | 300                  | 10                  | 100                  | 68                    | 7850                            | 210                         |

According to GB 50251-2015, the burial depth of buried pipelines in dryland areas should generally be no less than 600 mm. However, based on field investigation and overburden-loss observations along the Shanxi-Beijing pipeline corridor, it was found that the actual soil cover in some local sections has fallen below the code requirement due to long-term agricultural activities, surface erosion, and local overburden loss. The present study considers a specific hazardous scenario observed in an agricultural farmland section along the pipeline corridor, where rotary tillage is carried out within cultivation ditches/ridges and some local segments exhibit reduced soil cover. Under such conditions, the ditch depth together with the effective working depth of the rotary tiller blade substantially reduces the actual soil cover separating the blade from the buried pipeline. A 1:1 symmetric model was created, featuring a pipeline segment with a length of 1 m. An internal pressure of 6.0 MPa was applied to its inner wall. The soil body dimensions are 1000 mm × 800 mm × 1800 mm [18]. All translational degrees of freedom ( $U_x = U_y = U_z = 0$ ) were fixed at its upper, bottom, and rear surfaces. The front surface was assigned a free boundary condition to emulate the behavior of a semi-infinite soil mass. A two-stage prescribed motion was defined for the blade-soil interaction. During 0–0.6 s, the blade penetrated to the target working depth within the cultivation ditch corresponding to the shallow-cover hazard scenario; during 0.6–1.2 s, it advanced at a constant speed. Here, the prescribed penetration depth is governed by the realistic ditch-based tillage operation rather than the nominal pipeline burial depth. The finite element model of the machine–pipeline–soil system is illustrated in Fig. 2, and the corresponding mesh is shown in Fig. 3. All finite element simulations were performed using Abaqus/Explicit.

To improve the clarity and reproducibility of the numerical model, the mesh configuration is summarized in [Table 4](#). Specifically, the pipeline was discretized using 4-node reduced-integration shell elements (S4R), with a total of 2110 elements; the soil domain was discretized using 8-node reduced-integration brick elements (C3D8R), with a total of 7750 elements; and the rotary tiller was discretized using 10-node modified tetrahedral elements (C3D10M), with a total of 19,134 elements. Owing to the negligible deformation of the rotary tiller during operation, the discretized tiller part was further constrained as a rigid body to preserve geometric fidelity while reducing computational cost. In terms of meshing strategy, local mesh refinement was implemented in the blade–pipeline interaction region, whereas relatively coarser meshes were adopted in non-critical regions to balance accuracy and efficiency.



**Figure 2:** Machine-pipeline-soil finite element model.



**Figure 3:** Finite element model meshing.

**Table 4:** Mesh configuration of the finite element model.

| Component     | Element Type | Element Description                      | Number of Elements |
|---------------|--------------|--|--------------------|
| Pipeline      | S4R          | 4-node reduced-integration shell element | 2110               |
| Soil domain   | C3D8R        | 8-node reduced-integration brick element | 7750               |
| Rotary tiller | C3D10M       | 10-node modified tetrahedral element     | 19,134             |

Given the complexity inherent in the interaction among the pipeline, rotary tiller, and soil, several simplifying assumptions are adopted in the model to improve the efficacy of the numerical simulations:

- (1) The study considers an unfavorable hazard scenario in an agricultural farmland section along the Shanxi-Beijing pipeline corridor, where ditch-based rotary tillage and local overburden loss jointly reduce the effective soil cover above the pipeline. Under such conditions, the blade may directly contact the pipeline, and the resulting concentrated mechanical action is taken as the primary damage mechanism. Accordingly, the loading process is idealized as quasi-static, while secondary influences such as longitudinal vibration are neglected.
- (2) The shear force applied by the rotary tiller to the pipeline is assumed to be transformed into a rotational component.
- (3) Given the negligible deformation of the rotary tiller blades during operation, they are modeled as rigid bodies.
- (4) The soil interacts with the pipeline via surface-to-surface contact. Furthermore, the soil medium in the operational zone is idealized as a continuum with homogeneous density and isotropic properties.
- (5) The analysis focuses on the steady advancing stage of the rotary tiller; inertial effects associated with acceleration and vibration responses are not considered.

### 2.1.2 Simulation Scheme Design

Utilizing the developed finite element model, the mechanical response characteristics of the buried gas transmission pipeline subjected to rotary tiller operations were analyzed using Abaqus/Explicit. The influencing factors for damage are classified into two categories: pipeline body factors and external environmental factors. The former comprises wall thickness and operating pressure, whereas the latter covers the number of blades, operational speed, and soil density. Accordingly, a numerical simulation scheme was formulated, with specific parameters detailed in Table 5. In the present parametric study, soil density served as a secondary sensitivity parameter for spanning loose-to-compacted soil-constraint conditions surrounding the buried pipeline. The lower and upper bounds listed in Table 5 therefore cover variations in soil confinement and load-transfer capacity within the shallow-cover hazard scenario. All simulations in this study were conducted under the specific shallow-cover hazard scenario described in Section 2.1.1.

### 2.1.3 Finite Element Model Verification and Dynamic Validation

The rotary tiller-pipeline-soil problem involves transient contact, soil plasticity, and nonlinear structural response. To assess the credibility of the numerical dataset, a two-level verification and validation procedure was adopted. First, the internal-pressure loading step and the pipe constitutive response were verified under a quasi-static condition using Lamé's solution. Second, the dynamic response and soil-pipe interaction were validated against published drop-weight impact tests on buried steel pipelines. The first-level verification is presented below based on Lamé's solution.

$$\sigma_{eq} = \sqrt{\sigma_r^2 + \sigma_\theta^2 - \sigma_r \sigma_\theta} \quad (1)$$

$$\sigma_r = \frac{p_i r_0^2}{R^2 - r_0^2} \left( 1 - \frac{R^2}{r^2} \right) \quad (2)$$

$$\sigma_\theta = \frac{p_i r_0^2}{R^2 - r_0^2} \left( 1 + \frac{R^2}{r^2} \right) \quad (3)$$

where  $\sigma_{eq}$  is the effective stress, MPa;  $\sigma_r$  is the axial stress, MPa;  $\sigma_\theta$  is the hoop stress, MPa;  $p$  is the operating pressure, MPa;  $R$  is the outer diameter of the pipeline, mm;  $r$  is the inner diameter of the pipeline, mm.

**Table 5:** Numerical simulation scheme.

| Parameter                            | Constant | Variable  |
|--------------------------------------|----------|---|
| Wall Thickness (mm)                  | 175      | 8.75; 10.5; 12.25; 14; 15.75; 19.25; 21; 22.75; 24.5; 26.25 |
| Operating Pressure (MPa)             | 6        | 3; 3.6; 4.2; 4.8; 5.4; 6.6; 7.2; 7.8; 8.4; 9                |
| Number of Blades per Row             | 6        | 4; 5; 6; 7; 8; 9  |
| Rotary Tiller Operating Speed (mm/s) | 50       | 25; 30; 35; 40; 45; 55; 60; 65; 70; 75                      |
| Soil Density (kg/m <sup>3</sup> )    | 1850     | 925; 1110; 1295; 1480; 1665; 1850; 2035; 2220               |

**Table 6** compares the von Mises equivalent stress obtained from Lamé's thick-walled cylinder solution and the Abaqus simulation for a pipe segment ( $D = 1016$  mm,  $t = 17.5$  mm,  $p = 6.0$  MPa). The relative difference is 4.35%, which verifies that the internal-pressure boundary condition and the elastic-plastic material response of the pipe are implemented correctly. However, this comparison is not sufficient to validate the transient impact process, the soil plasticity model, or the contact interaction, which are addressed through the second-level dynamic validation described next. In this work, peak strain is adopted as the primary benchmarking metric, with the aim of examining the reasonableness of the developed model.

**Table 6:** Verification under internal pressure.

| Stress Type      | Theoretical Value (MPa) | Simulated Value (MPa) | Error (%) |
|------------------|-------------------------|-----------------------|-----------|
| Effective Stress | 221.78                  | 231.43                | 4.35      |

To validate the impact response and soil-pipe interaction captured in this study, the Abaqus/Explicit model was benchmarked against a published soil-box drop-weight test on a buried steel pipeline reported by Dong et al. [19]. A representative case, A114 × 2.5-d0.6-H2.0 in clay, was reproduced, in which the pipe had an outer diameter of 114 mm, a wall thickness of 2.5 mm, a burial depth of 0.6 m, and a drop height of 2.0 m. The experimental results reported peak mid-span strains of 816  $\mu\epsilon$  in the longitudinal direction and 603  $\mu\epsilon$  in the transverse direction. Under the same geometric configuration and material parameters reported in the benchmark study, the present simulation predicted peak strains of 774 and 574  $\mu\epsilon$  in the longitudinal and transverse directions, respectively, corresponding to relative errors of 5.17% and 4.87%. The close agreement between the numerical and experimental results demonstrates that the developed model can reliably capture the strain-level dynamic response of buried pipelines subjected to impact loading.

## 2.2 CatBoost Model

The Gradient Boosting Decision Tree (GBDT) is an ensemble learning algorithm extensively applied to both classification and regression problems [20]. GBDT employs an iterative strategy to build a sequence

of decision trees [21]. Each subsequent tree is trained to correct the prediction errors of its predecessors by fitting to the residuals. In classification tasks, the logarithmic loss function (Logloss) is typically employed as the optimization objective:

$$L(y, \hat{y}) = - \sum_{i=1}^N (y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)) \quad (4)$$

where  $y_i$  is the actual label of the  $i$ -th sample, and  $\hat{y}_i$  is the predicted probability for the  $i$ -th sample.

The loss function is employed to approximate the residuals for each decision tree, thereby training the newly generated decision tree  $h_m$  in the current iteration:

$$h_m(x) = \underset{h}{\operatorname{argmin}} \sum_{i=1}^n [L(y_i, F_{(m-1)}(x_i) + h(x_i))] \quad (5)$$

where  $F_{(m-1)}(x_i)$  is the predicted result of the model for the  $i$ -th sample after the  $(m - 1)$ -th iteration.

The trees are trained in a stepwise manner, with the model prediction  $F_m(x)$  being updated through the minimization of the loss gradient:

$$F_m(x) = F_{m-1}(x) + \eta \cdot h_m(x) \quad (6)$$

where  $\eta$  is the learning rate.

However, traditional GBDT models exhibit limitations including the curse of dimensionality and information loss in processing categorical features. The Categorical Boosting (CatBoost) algorithm, an enhanced variant of GBDT, innovatively incorporates an ordered boosting strategy. This strategy employs a permutation-driven mechanism for sequential tree construction, effectively mitigating the risk of target leakage associated with conventional approaches. This permutation-driven approach eliminates the prediction shift induced by target leakage, thereby substantially improving the predictive accuracy of the model [22,23]. When training the  $m$ -th tree, CatBoost calculates the gradient in the following manner:

$$g_i^{(m)} = \frac{\partial L(y_i, F(x_i) + h(x_i))}{\partial F(x_i)} \quad (7)$$

CatBoost integrates techniques such as target statistics transformation, symmetric tree structures, ordered learning, and target encoding to formulate a holistic scheme for feature processing and leakage prevention. Furthermore, it enhances robustness by employing gradient descent coupled with adaptive regularization. Compared to traditional GBDT, it demonstrates significant advantages in the efficiency of categorical feature processing and the predictive accuracy for high-dimensional data [24,25].

### 2.3 Particle Swarm Optimization Algorithm

The Particle Swarm Optimization (PSO) algorithm was proposed by Kennedy and Eberhart in 1995. It simulates the foraging behavior of bird flocks, where a swarm of particles iteratively approximates the global optimum through cooperative movement in a high-dimensional space. Throughout the iterative process, each particle's position and velocity are updated dynamically according to its own historical best position and the swarm's global best position [26]. Within the PSO algorithm, the velocity and position of a particle are updated according to the following formulas:

$$v_i(t+1) = w \cdot v_i(t) + c_1 \cdot r_1 \cdot (p_i(t) - x_i(t)) + c_2 \cdot r_2 \cdot (g(t) - x_i(t)) \quad (8)$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (9)$$

where  $w$  is the inertia weight, used to adjust the balance between exploration and exploitation during the particle update process.  $v_i(t)$ ,  $x_i(t)$ ,  $p_i(t)$ , and  $g(t)$  are the velocity, position, personal best, and global best of the particle at time  $t$ , respectively.  $c_1$  and  $c_2$  are learning factors that regulate the particle's reliance on its own historical best position and the global best position, respectively.  $r_1$  and  $r_2$  are random numbers that introduce randomness to the particle's movement.

The PSO algorithm is characterized by its few parameters and computational efficiency. Its mechanism for sharing information within the swarm facilitates rapid convergence, yielding notable advantages in tackling complex problems like hyperparameter tuning for neural networks and optimization of non-convex functions [27]. Owing to its remarkable convergence properties and user-friendliness, the PSO algorithm has established itself as a benchmark algorithm within the domain of swarm intelligence optimization [28].

## 2.4 Shapley Additive Explanations

SHAP (SHapley Additive exPlanations) is a model explanation method rooted in cooperative game theory's Shapley values. It conceptualizes features as players in a game, computes their marginal contributions across all possible feature subsets, and fairly allocates the predictive output's contribution among them, thereby providing precise quantification of each feature's impact [10,24]. The prediction of the SHAP model can be expressed as:

$$f(x) = \phi_0 + \sum_{i=1}^M \phi_i x_i \quad (10)$$

where  $\phi_0$  is the baseline value, representing the model output with no feature inputs;  $\phi_i$  is the Shapley value for feature  $i$ . A  $\phi_i$  greater than 0 indicates that feature  $i$  exerts a positive influence on the model.

SHAP supports both global and local explanations. Global explanations reveal the average contribution of each feature across the entire dataset, while local explanations target individual data points, demonstrating the contribution of features to a specific prediction [29,30]. CatBoost has a built-in method for calculating SHAP values, allowing direct acquisition of the SHAP value for each feature's contribution to the prediction result. The relationship between the SHAP values and the model's prediction can be formulated as follows:

$$\hat{y} = \phi_0 + \sum_{i=1}^M \phi_i \quad (11)$$

where  $\hat{y}$  is the predicted value of the model. Summation of all feature SHAP values plus the baseline expectation  $\phi_0$  reconstructs the model's prediction  $\hat{y}$ , thereby achieving the transformation from the explanatory SHAP values back to the original predictive output.

## 3 Results and Discussions

### 3.1 Finite Element Analysis

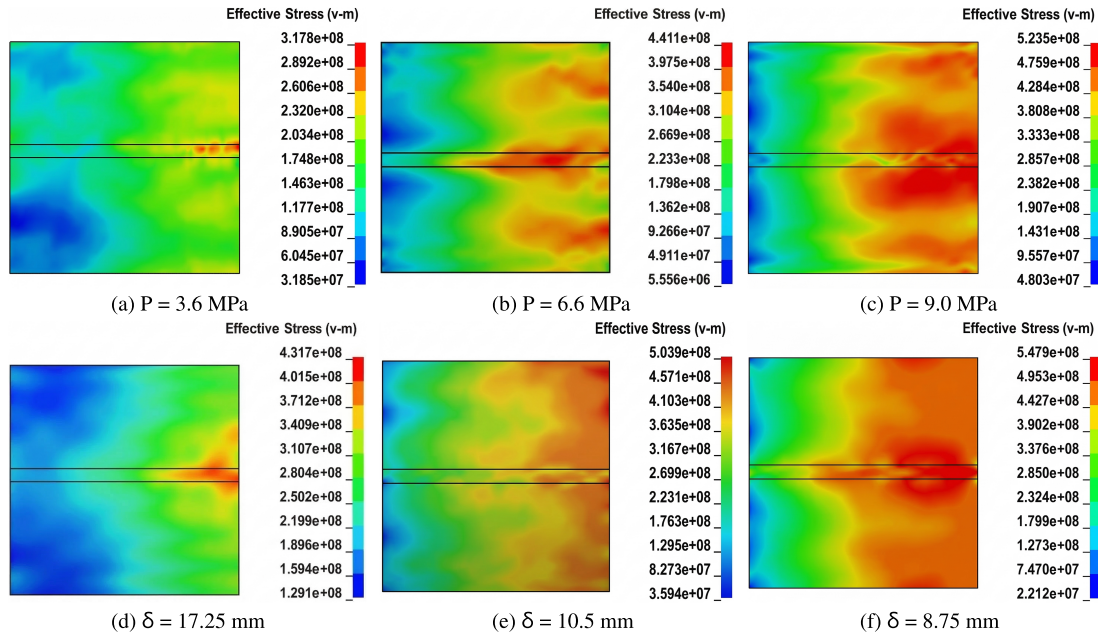
Pipeline damage states were evaluated primarily using the von Mises equivalent stress obtained from the Abaqus/Explicit simulations. Under combined internal-pressure and contact loading, the multiaxial stress state was converted into an equivalent uniaxial stress for engineering assessment, as defined by Eqs. (1)–(3). Based on the material yield strength and ultimate strength, the pipeline response was classified into three states: elastic deformation, plastic deformation, and failure. In this study, elastic deformation and

plastic deformation are interpreted as warning-indicative response states for pipeline integrity management, whereas pipeline failure denotes the terminal rupture/loss-of-integrity state of the pipeline. Within the present stress-based classification framework, the pipeline failure state is operationally identified by exceedance of the material ultimate-strength threshold. The specific classification criteria are detailed in Table 7.

**Table 7:** Pipeline response-state and failure evaluation criteria.

| Pipeline State      | Evaluation Criterion   |
|---------------------|--|
| Elastic Deformation | $\sigma_{eq} < \sigma_y$ ; warning-indicative low-risk response state.   |
| Plastic Deformation | $\sigma_y \leq \sigma_{eq} < \sigma_u$ ; warning-indicative elevated-risk response state.                                    |
| Pipeline Failure    | $\sigma_{eq} \geq \sigma_u$ ; terminal rupture/loss-of-integrity state in the present stress-based classification framework. |

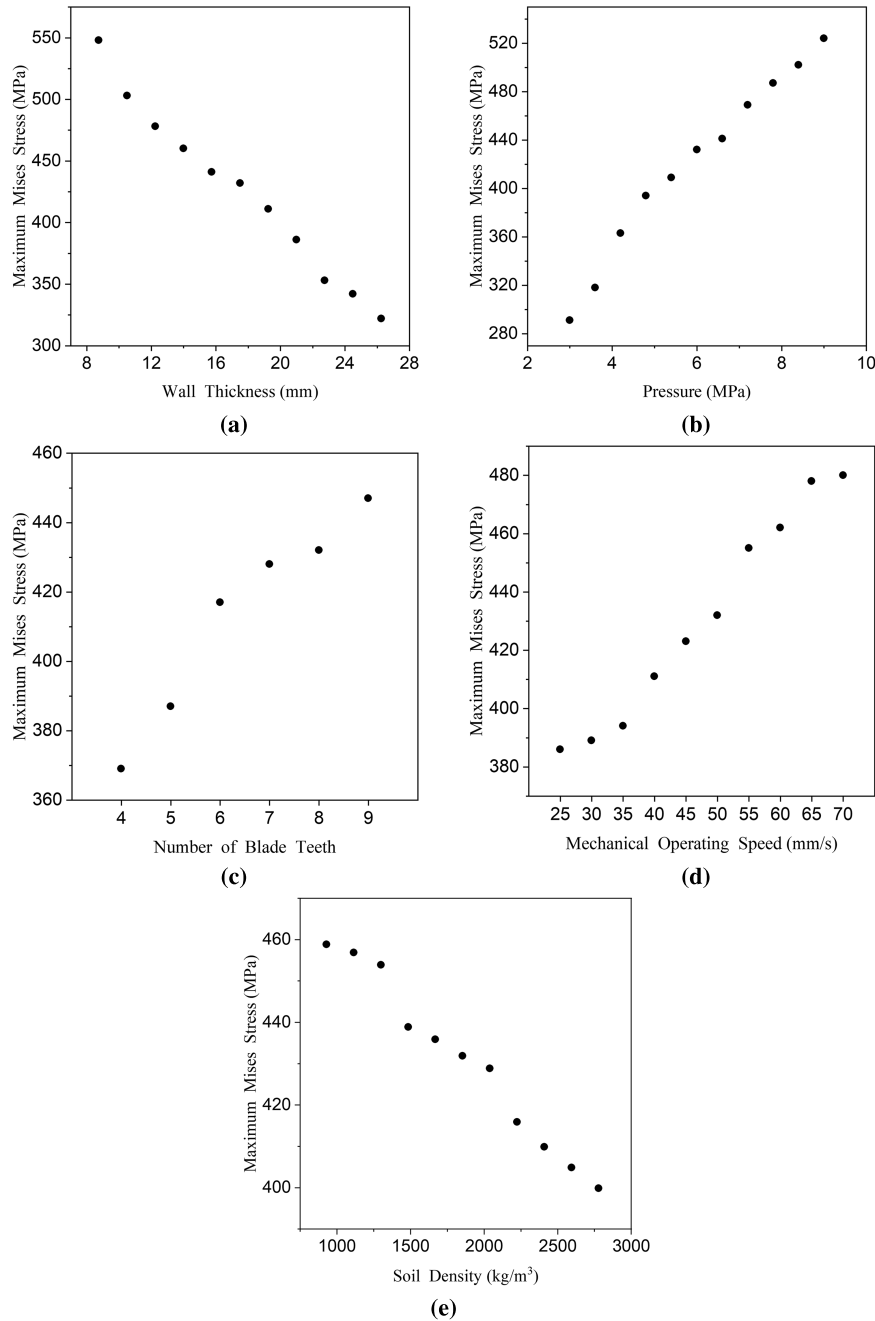
The finite element model for direct blade-pipeline interaction under the shallow-cover hazard scenario was solved using the Abaqus/Explicit solver with a time-stepping nonlinear contact algorithm. The von Mises stress contours of the pipeline under representative operating conditions are shown in Fig. 4. Subfigures (a,c) correspond to the warning states characterized by elastic deformation, while subfigures (b,d) represent the warning states associated with plastic deformation. Subfigures (e,f) illustrate pipeline failure, i.e., the terminal rupture or loss-of-integrity state considered in this study.



**Figure 4:** Maximum stress contour of pipeline under mechanical loads in selected working conditions.

Based on the results of multi-condition numerical simulations, this study analyzes the influence of key parameters on the maximum von Mises stress of the pipeline. The results for the discrete simulated cases are summarized in Fig. 5. The wall thickness of the pipeline exhibits a significant effect in suppressing stress concentration; its increase systematically reduces the maximum von Mises stress. Conversely, an increase in the internal pressure of the pipeline exacerbates the stress level in the pipe wall. In terms of external

loading conditions, an increase in either the number of blade teeth or the operational speed results in a linear escalation of the load intensity, thereby inducing higher stress. Within the examined parametric range, higher soil-density values correspond to a stronger confining and load-redistribution effect of the surrounding soil, which contributes to a reduction in the maximum stress of the pipeline. This trend reflects the sensitivity of pipeline stress to changes in the surrounding-soil constraint within the defined parameter space.



**Figure 5:** Effects of parameters on pipeline maximum stress. (a) Effect of wall thickness change on pipeline maximum stress. (b) Effect of operating pressure change on pipeline maximum stress. (c) Effect of blade number change on pipeline maximum stress. (d) Effect of operating speed change on pipeline maximum stress. (e) Effect of soil density change on pipeline maximum stress.

### 3.2 Failure Mode Prediction Model

#### 3.2.1 Dataset Construction

To fulfill the data volume requirement for machine learning models, five key parameters were selected to construct the feature vector, informed by engineering and field practice: the number of rotary tiller blade teeth, operational speed, pipeline internal pressure, wall thickness, and soil density. Detailed parameter configurations are provided in Table 8. Simulation calculations were then performed to establish the mapping relationship between this feature vector and the degree of pipeline damage under shallow-cover hazard scenarios. The constructed dataset was designed to support damage-state identification, high-risk condition screening, and operational control for buried gas pipelines subjected to rotary tiller disturbance in farmland environments. The wall thickness was strictly set according to the allowable gradient for X70 steel pipes specified in the GB 50251-2015 code, covering the range of  $17.5 \text{ mm} \pm 50\%$ , to systematically evaluate the differences in structural stiffness and load-bearing capacity from thin-wall buckling susceptibility to thick-wall high stiffness. The operating pressure adopted actual working-condition values, while the number of blades and operating speed were set within conventional ranges. Soil density served as a secondary sensitivity parameter for spanning the boundary conditions of loose-to-compacted soil constraint around the buried pipeline. The selected density levels therefore capture variations in soil confinement, shear transfer, and local load-redistribution capacity within the simplified continuum-soil framework [31]. In this setting, the density levels function as equivalent indicators of surrounding-soil constraint, enriching the feature space of the constructed dataset for the studied shallow-cover hazard scenario. The resulting prediction performance therefore corresponds to the defined parameter space covered by the dataset.

**Table 8:** Characteristic parameters of datasets.

| Parameter                            | Data                             |
|--------------------------------------|----------------------------------|
| Wall Thickness (mm)                  | 8.75; 10.50; 14.00; 17.50; 26.25 |
| Operating Pressure (MPa)             | 3.0; 6.0; 7.5; 8.0; 9.0          |
| Blades per Row                       | 4; 5; 6; 8; 9                    |
| Rotary Tiller Operating Speed (mm/s) | 25; 50; 60; 70; 75               |
| Soil Density ( $\text{kg/m}^3$ )     | 925; 1850                        |

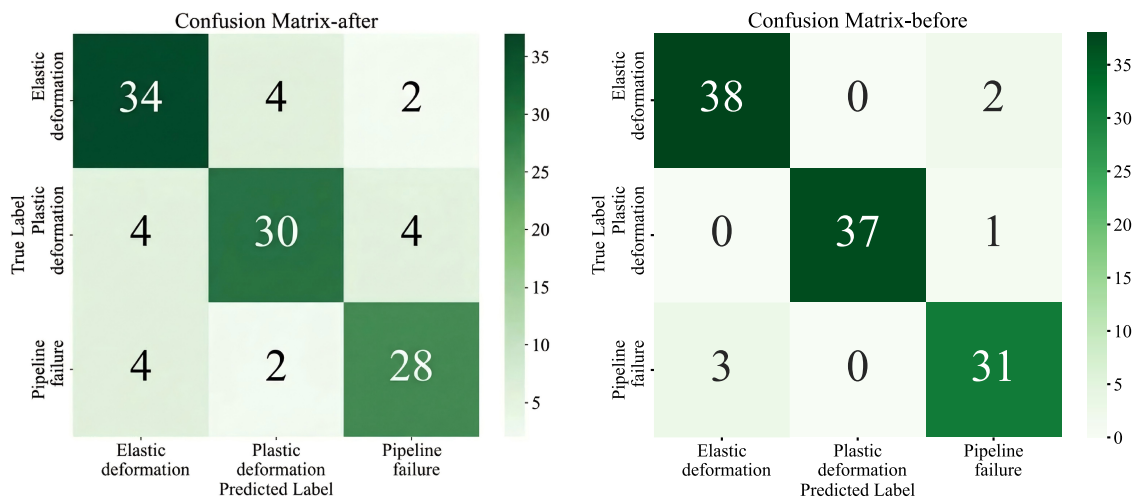
#### 3.2.2 Model Hyperparameter Optimization

Addressing the challenges inherent in navigating the high-dimensional parameter space for Cat-Boost hyperparameter optimization, a global optimization framework grounded in the Particle Swarm Optimization (PSO) algorithm was developed. This framework defines a search space covering critical hyperparameters including the learning rate, tree depth, L2 regularization coefficient, and subsampling rate, as summarized in Table 9. Using prediction accuracy as the objective function, the PSO algorithm's swarm collaboration mechanism is leveraged to achieve adaptive parameter exploration. The algorithm balances global exploration and local exploitation capabilities through its particle position update strategy, while dynamically adjusting the subsample rate based on the bootstrap method.

**Table 9:** Parameter space range and optimization results.

| Parameter        | Parameter Range     | Optimal Result |
|------------------|---------------------|----------------|
| learning_rate    | (0.005, 0.5)        | 0.479          |
| Depth            | (3, 12)             | 9              |
| l2_leaf_reg      | (0.1, 15)           | 1.960          |
| Subsample        | (0.6, 1.0)          | 0.8            |
| min_data_in_leaf | (1, 20)             | 11             |
| grow_policy      | Symmetric/Depthwise | Depthwise      |
| bootstrap_type   | Bayesian/Bernoulli  | Bayesian       |

To verify the effectiveness of the PSO algorithm in enhancing the prediction performance of the CatBoost model for natural gas pipeline damage modes under rotary tiller action, this study conducted experimental analysis by comparing confusion matrices before and after parameter optimization. As illustrated in Fig. 6, following PSO optimization, the model's accurate predictions for elastic deformation increased from 34 to 38, while instances misclassified as plastic deformation were reduced from 4 to zero. Concurrently, correct predictions for plastic deformation rose markedly from 30 to 37, corresponding to a 18.4% reduction in classification error. This optimization led to a decrease in the overall model misclassification rate and enhanced model stability. The results demonstrate that the PSO algorithm balanced the model's sensitivity to multiple damage modes through global optimization, increasing damage identification accuracy by 8.1%. The optimized parameter combination improved generalization capability by 15.4%, validating the algorithm's practical value in parameter tuning for complex scenarios.

**Figure 6:** Confusion matrix before and after parameter optimization.

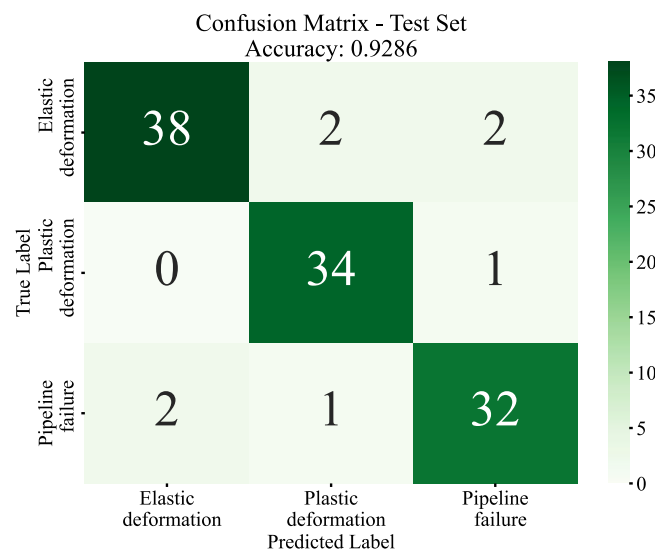
### 3.2.3 Prediction Model Development

This study develops a prediction model for natural gas pipeline damage under rotary tiller action. The number of rotary tiller blade teeth, operating speed, pipeline internal pressure, wall thickness, and soil density are selected as input features, with the damage level as the output variable. The dataset was divided into training and testing sets in an 8:2 ratio. During model training, a 5-fold cross-validation strategy was adopted. The modeling and hyperparameter tuning were implemented using the CatBoost framework integrated with

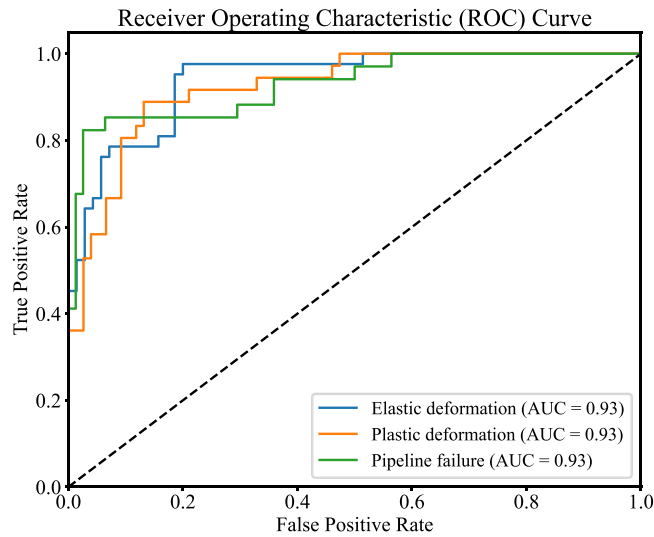
the PSO algorithm. The optimization targeted hyperparameters including the learning rate, tree depth, and L2 regularization coefficient, with the goal of maximizing classification accuracy. In addition, the post hoc model-interpretation analysis was conducted using the SHAP framework.

The PSO-CatBoost model was employed to predict three damage categories: elastic deformation, plastic deformation, and pipeline failure. Among these categories, elastic deformation and plastic deformation represent warning-indicative response states, whereas pipeline failure denotes the terminal rupture state. Model performance was assessed via a confusion matrix and Receiver Operating Characteristic (ROC) curves, with classification efficacy quantified by the Area Under the Curve (AUC) value. The classification performance on the test set, as depicted by the confusion matrix in Fig. 7, shows that 104 out of 112 samples were accurately predicted regarding their damage category. The model's ROC curves in Fig. 8 show that the area under the curve values for all three damage types reached 0.93, and their ROC curves are significantly higher than the random classification baseline. Among them, elastic deformation exhibits the best discriminative ability, while the other two also approach the ideal classification boundary. The results indicate that parameter optimization via the PSO algorithm substantially improved the multi-classification performance of the CatBoost model, endowing it with strong discriminative capability for all three damage categories under random-split evaluation. This validates the efficacy of the PSO algorithm for parameter tuning within the context of rotary tiller-pipeline interaction. Consequently, the PSO-CatBoost model demonstrated strong classification performance for both pipeline warning-state identification and terminal-failure recognition.

To further address the concern that random splitting may overestimate model performance, supplementary leave-one-speed-out validation was performed by treating the five operating-speed levels as independent groups. In each round, one speed group was held out for testing, while the remaining four groups were used for model training and hyperparameter tuning. The corresponding results are summarized in Table 10. The model achieved accuracies of 0.81, 0.83, 0.86, 0.84, and 0.88 across the five held-out speed groups, yielding a mean accuracy of  $0.844 \pm 0.027$ ; the Macro-F1 scores were 0.79, 0.81, 0.84, 0.81, and 0.86, with a mean value of  $0.822 \pm 0.028$ . These supplementary results indicate that, although the grouped-validation performance is lower than the original random-split result, the PSO-CatBoost model still maintains stable predictive capability under unseen speed conditions, demonstrating a certain degree of generalization ability.



**Figure 7:** Confusion matrix diagram of random test set.



**Figure 8:** Model ROC curve.

**Table 10:** Supplementary leave-one-speed-out validation results of the PSO-CatBoost model.

| Left-Out Speed (mm/s) | Accuracy        | Macro-F1        |
|-----------------------|-----------------|-----------------|
| 25                    | 0.81            | 0.79            |
| 50                    | 0.83            | 0.81            |
| 60                    | 0.86            | 0.84            |
| 70                    | 0.84            | 0.81            |
| 75                    | 0.88            | 0.86            |
| Mean +/- Std          | 0.844 +/- 0.027 | 0.822 +/- 0.028 |

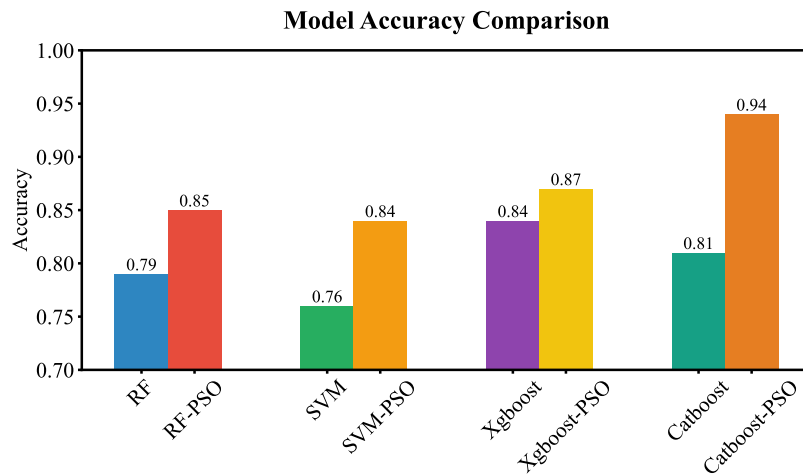
### 3.2.4 Comparative Verification of Algorithms

To validate the efficiency and effectiveness of the CatBoost-PSO model, this study conducted a comparative analysis with commonly used algorithms in pipeline failure mode prediction, including Random Forest, Support Vector Machine, and Xgboost. After PSO-based hyperparameter optimization, all four models showed improved predictive performance, as summarized in Fig. 9 and Table 11. Among them, PSO-CatBoost achieved the best overall results, with accuracy increasing from 0.81 to 0.94 and with lower MAE and RMSE than the competing models. These results indicate that PSO-based optimization improved model performance and that PSO-CatBoost provided the most accurate damage-state classification under the present random-split evaluation.

### 3.3 Model Interpretation

The deployment of machine learning algorithms for risk diagnosis is constrained by their inherent complexity and opaque “black-box” nature. In this context, model interpretability becomes paramount for fostering user trust. Based on the Shapley value from cooperative game theory, SHAP effectively resolves the model interpretability issue by quantifying each feature’s positive or negative contribution to a prediction. It offers key theoretical advantages including model-agnosticism, the ability to capture feature interactions, and

adherence to the property of additive consistency. Collectively, these attributes furnish a robust theoretical foundation for explaining complex models.



**Figure 9:** Model accuracy comparison chart.

**Table 11:** Evaluation metrics of various algorithms.

| Model        | Accuracy | MAE    | RMSE   |
|--------------|----------|--------|--------|
| RF           | 0.79     | 0.2589 | 0.6050 |
| RF-PSO       | 0.85     | 0.1964 | 0.5345 |
| SVM          | 0.76     | 0.3036 | 0.6547 |
| SVM-PSO      | 0.84     | 0.2143 | 0.5669 |
| Xgboost      | 0.84     | 0.2232 | 0.5901 |
| Xgboost-PSO  | 0.87     | 0.1875 | 0.5428 |
| Catboost     | 0.81     | 0.2339 | 0.5750 |
| Catboost-PSO | 0.94     | 0.0982 | 0.4119 |

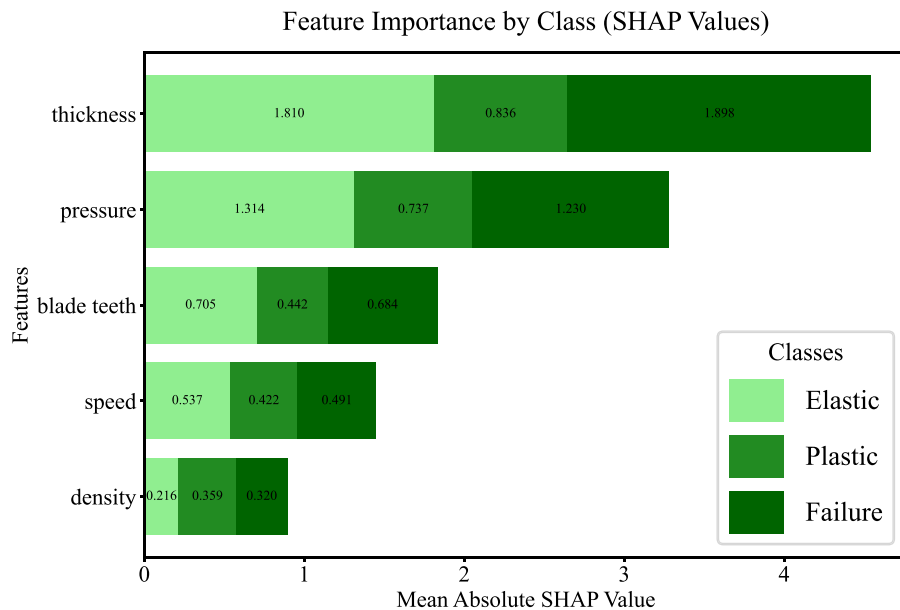
[Fig. 10](#) presents the global SHAP importance of the features used in the pipeline damage prediction model under rotary tiller loading. The color coding of the bars corresponds to different damage categories. The results show that wall thickness and internal pressure make strong contributions to the model output, which is consistent with their roles in governing the baseline stress state and structural resistance of pressurized pipelines. The global importance ranking provides a baseline sensitivity map for the damage prediction model. The SHAP analysis also highlights blade number as a key external variable because it directly controls the frequency and concentration of blade–pipeline contact under the shallow-cover hazard condition. These results establish the feature-sensitivity background for the subsequent identification of threshold-sensitive risk boundaries in [Fig. 11](#).

To analyze the transition mechanism of pipeline damage states during rotary tiller operation, this study uses SHAP dependence plots to identify threshold-sensitive risk boundaries associated with the transition from warning states to pipeline failure, as shown in [Fig. 11](#). The results show that increasing the blade count from 4 to 5 does not markedly change the predicted damage state, whereas the SHAP contribution to the Pipeline Failure class rises sharply when the blade count reaches 8–9. The results also show a clear

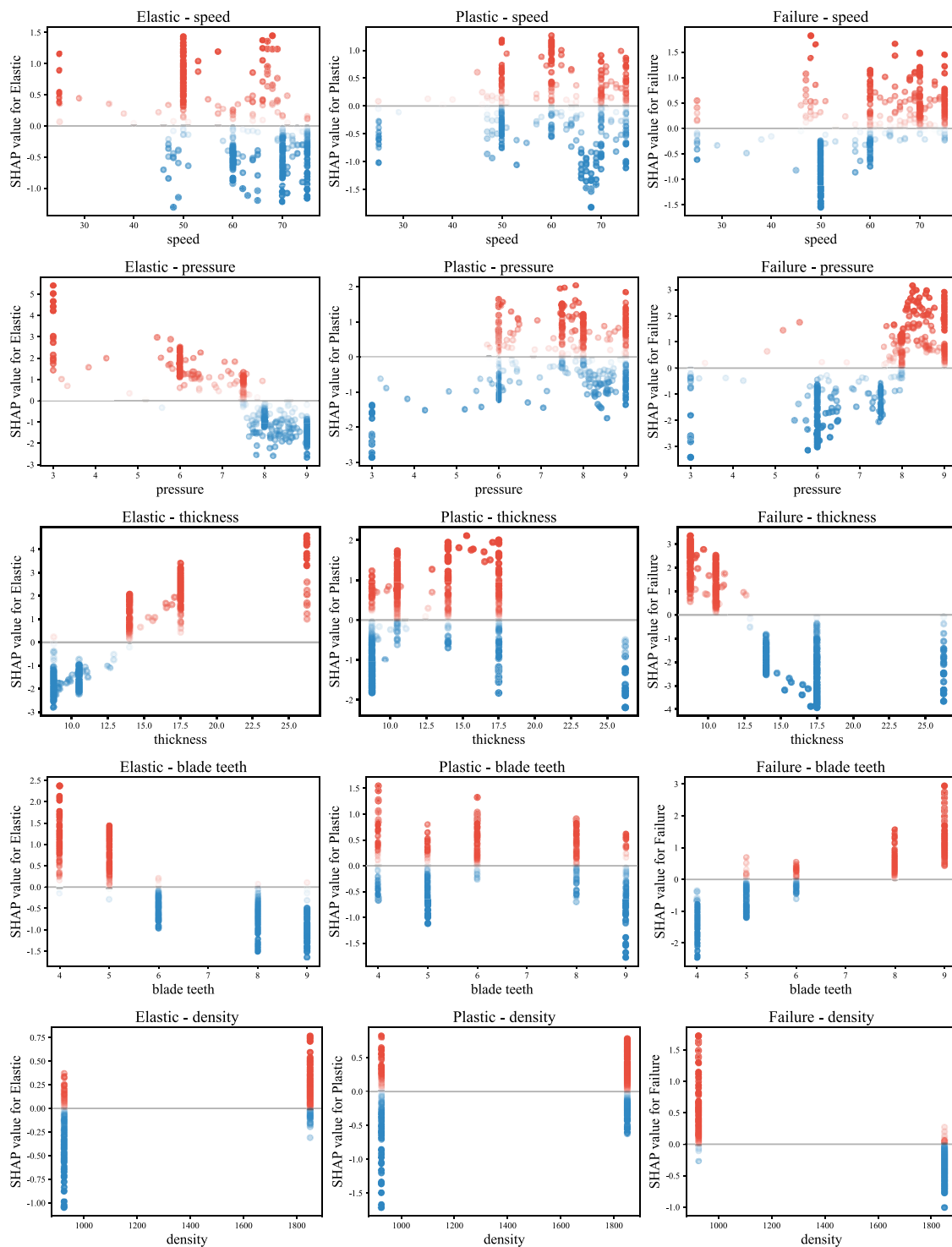
pressure threshold effect. Elastic deformation dominates when the internal pressure is 3–5 MPa, and the SHAP contribution to the Pipeline Failure class increases sharply when the pressure rises to 8–9 MPa. The operating-speed interval of 50–70 mm/s maintains a consistently elevated contribution to the Pipeline Failure class, indicating a high-risk operating window in the present dataset. These results identify blade count above 8, internal pressure above 8 MPa, and operating speed above approximately 50 mm/s as warning-level parameter ranges for rapid transition from elastic/plastic warning states to pipeline failure under the studied shallow-cover disturbance scenario.

Combining the SHAP dependence plots with structural mechanics theory, the identified risk boundaries can be interpreted as follows. A higher blade count increases the frequency and concentration of direct blade–pipeline contact under the shallow-cover hazard condition, which intensifies local stress concentration and accelerates the transition from warning states toward pipeline failure. High internal pressure increases the baseline hoop-dominated stress state of the pressurized pipeline, and the additional external contact load therefore drives the structure into yield and ultimately into pipeline failure more readily. Low wall thickness reduces bending stiffness and local load-carrying capacity, which determines the structural safety margin of the pipeline. Once the blade advancing speed exceeds the threshold range, the contact loading path becomes more severe and further promotes local instability. Soil density affects the likelihood of transition to pipeline failure through soil confinement, shear transfer, and local load redistribution under the studied operating conditions. These results provide a mechanical basis for identifying warning-level parameter ranges and for controlling high-risk operating conditions in shallow-cover pipeline sections before terminal rupture occurs.

To further examine the decision-making mechanism of the rotary tiller-induced pipeline damage prediction model, SHAP force plots were used to quantify the contribution of individual features. In the SHAP analysis, the horizontal axis represents the influence of each feature on the predicted probability of the pipeline failure class. Red indicates a positive contribution, corresponding to an increased probability of pipeline failure, whereas blue indicates a negative contribution, corresponding to a reduced probability.



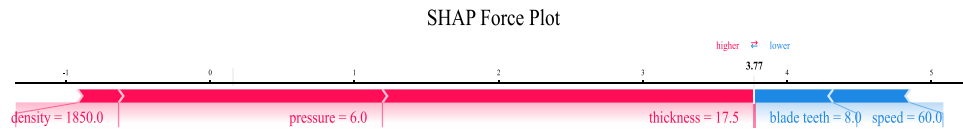
**Figure 10:** Model SHAP value bar chart.



**Figure 11:** SHAP dependence plots of features.

As shown by the SHAP force plot for the representative case in Fig. 12, the positive SHAP value associated with a soil density of  $1850 \text{ kg/m}^3$  suggests that higher soil density increases the likelihood of pipeline failure in this specific scenario. This may be attributed to the stronger constraint and load-transfer

effect of denser soil. Likewise, the positive SHAP values associated with an internal pressure of 6.0 MPa and a wall thickness of 17.5 mm indicate that these parameters contributed positively to the prediction of pipeline failure. In contrast, the negative SHAP values associated with blade count and lower operating speed suggest a reduced tendency toward pipeline failure.



**Figure 12:** Sample SHAP force plot.

The multiphysics-coupled analysis further indicates that the synergistic effect of internal pressure and wall thickness, together with the combined influence of blade count and operating speed on the contact loading path, jointly governs the damage evolution process. For this representative case, pipeline failure was predicted as the dominant outcome because the aggregate contribution of the positively contributing features outweighed that of the negatively contributing features.

#### 4 Conclusions

By integrating the Particle Swarm Optimization algorithm with the CatBoost model, this study developed a predictive model for rotary tiller-induced pipeline damage. Through the combined use of finite element simulation and machine learning, the proposed approach identified the key influencing factors and underlying damage mechanisms, and further established an interpretability-driven decision-support framework. The main findings are summarized as follows:

- (1) Based on the finite element simulation results, this study constructed an integrated dataset for rotary tiller-induced pipeline damage and employed SHAP-based interpretability analysis to identify threshold-sensitive risk boundaries associated with the transition from warning states to pipeline failure. The global feature importance was ranked as follows: pipeline wall thickness > internal pressure > blade number > operating speed > soil density. The results indicate that wall thickness and internal pressure determine the baseline structural vulnerability of the pipeline, whereas blade number and operating speed define the critical operating ranges associated with rapid risk escalation. The risk of transition to pipeline failure increases markedly when the blade number exceeds 8 and the internal pressure exceeds 8 MPa, while operating speeds above approximately 50 mm/s correspond to an elevated failure-prone range. These warning-level parameter ranges provide a direct basis for engineering risk screening and safety management prior to terminal rupture.
- (2) A benchmark comparison was conducted using models including Random Forest (RF), Support Vector Machine (SVM), and Xgboost. Experimental results demonstrate that the PSO-CatBoost model achieves superior overall performance, with an accuracy of 0.94, significantly higher than the accuracies of the non-optimized CatBoost (0.81) and PSO-optimized RF (0.85) and SVM (0.84) models. Furthermore, this model attains the lowest values in both Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). These findings substantiate the efficacy of the PSO algorithm for complex parameter optimization. In addition, supplementary leave-one-speed-out validation further indicates that the PSO-CatBoost model retains a certain degree of generalization ability under previously unseen operating-speed conditions.
- (3) An integrated CatBoost model based on Particle Swarm Optimization (PSO) was developed and proposed in this study. A global search for hyperparameters such as learning rate and tree depth

significantly improves the model's performance. The optimized PSO-CatBoost model demonstrates excellent performance on the test set. Specifically, it achieves an AUC value of 0.93 for all three damage categories—elastic deformation, plastic deformation, and pipeline failure—thereby validating its high-precision predictive capability for both warning-state and terminal-failure identification.

- (4) The PSO-CatBoost-based damage prediction framework provides a quantitative basis for optimizing rotary tiller operating parameters and screening high-risk working conditions, thereby reducing the likelihood of transition from warning states to pipeline failure under the studied shallow-cover hazard scenario. Combined with the warning-level parameter ranges identified through SHAP analysis, this framework supports safety assessment and operational control for pipeline sections exposed to insufficient effective soil cover in farmland environments. Its predictive performance is established on the constructed simulation dataset and the associated parameter space, and field-oriented application can be further strengthened through calibration with site-specific soil properties. Future work may extend the present framework by considering more realistic field operating/loading conditions, broader variations in surrounding-soil properties (e.g., soil density), and the interaction effects among multiple parameters.

**Acknowledgement:** Not applicable.

**Funding Statement:** The authors received no specific funding.

**Author Contributions:** The authors confirm contribution to the paper as follows: Conceptualization, Liqiong Chen; methodology, Liqiong Chen and Kai Zhang; software, HaoYu Jia and Song Yang; formal analysis, HaoYu Jia and Mailun Liu; data curation, Mailun Liu and Zongjun Jiang; writing—original draft preparation, Haoyu Jia; writing—review and editing, Liqiong Chen, Song Yang and Zongjun Jiang; supervision, Liqiong Chen and Kai Zhang. All authors reviewed and approved the final version of the manuscript.

**Availability of Data and Materials:** Restrictions apply to the availability of these data due to commercial confidentiality and proprietary restrictions. The data supporting the findings of this study were obtained from internal enterprise monitoring data and are not publicly available.

**Ethics Approval:** Not applicable. This study did not involve human participants, human data, or animals.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Gao ZY, Zhang HY, Gao P. New developments in China's oil and gas pipeline construction in 2022. *Int Petroecon.* 2023;73(3):16–23.
2. Zhao L, Yang R, Bao J, Ou H, Xing Z, Qi G, et al. Dynamic risk assessment model for third-party damage to buried gas pipelines in urban location class upgrading areas. *Eng Fail Anal.* 2023;154:107682. doi:10.1016/j.engfailanal.2023.107682.
3. Zhou Y, Teng Z, Chi L, Liu X. Buried pipeline collapse dynamic evolution processes and their settlement prediction based on PSO-LSTM. *Appl Sci.* 2023;14(1):393. doi:10.3390/app14010393.
4. Jing HY. Dynamic response analysis and numerical simulation of shallow-buried pipelines under rockfall impact [master's thesis]. Wuhan, China: China University of Geosciences; 2007.
5. Jiang F, Zhao E. Damage mechanism and failure risk analysis of offshore pipelines subjected to impact loads from falling object, considering the soil variability. *Mar Struct.* 2024;93(5):103544. doi:10.1016/j.marstruc.2023.103544.
6. Yao AL, Xing YF, Zeng XG, Wu ZP, Li YL, Zhao SP. Simulated analysis of the dynamic response of buried steel pipeline to the rolling stone impact. *J Saf Environ.* 2009;9(3):122–5. (In Chinese).

7. Zhao L, Yao A, Xu T, Zhao D. Study on prediction model for mechanical excavation puncture damage of gas pipelines. In: The 25th National Academic Conference on Structural Engineering; 2016 Aug 12–14; Baotou, China.
8. Hong B, Shao B, Zhou M, Qian J, Guo J, Li C, et al. Evaluation of disaster-bearing capacity for natural gas pipeline under third-party damage based on optimized probabilistic neural network. *J Clean Prod.* 2023;428:139247. doi:10.1016/j.jclepro.2023.139247.
9. Xiao R, Zayed T, Meguid M, Sushama L. Rapid failure risk analysis of corroded gas pipelines using machine learning. *Ocean Eng.* 2024;313:119433. doi:10.1016/j.oceaneng.2024.119433.
10. Hu J. Prediction of the internal corrosion rate for oil and gas pipelines and influence factor analysis with interpretable ensemble learning. *Int J Press Vessels Piping.* 2024;212:105329. doi:10.1016/j.ijpvp.2024.105329.
11. Liu W, Chen Z, Hu Y, Zhang J. Forecasting pipeline safety and remaining life with machine learning methods and SHAP interaction values. *Int J Press Vessels Piping.* 2023;205(4):105000. doi:10.1016/j.ijpvp.2023.105000.
12. Xiang W, Zhou W. Bayesian network model for predicting probability of third-party damage to underground pipelines and learning model parameters from incomplete datasets. *Reliab Eng Syst Saf.* 2021;205(3):107262. doi:10.1016/j.res.2020.107262.
13. Zhangabay N, Ibraimova U, Bonopera M, Suleimenov U, Avramov K, Chernobryvko M, et al. Novel methodologies for preventing crack propagation in steel gas pipelines considering the temperature effect. *Struct Durab Health Monit.* 2025;19(1):1–23. doi:10.32604/sdhm.2024.053391.
14. Zhangabay N, Suleimenov U, Bonopera M, Ibraimova U, Yeshimbetov S. A method for preventing crack propagation in a steel gas conduit reinforced with composite overlays. *Struct Durab Health Monit.* 2025;19(4):773–87. doi:10.32604/sdhm.2025.064980.
15. Xu D, Chen LQ, Yu C, Zhang S, Zhao X, Lai X. Failure analysis and control of natural gas pipelines under excavation impact based on machine learning scheme. *Int J Press Vessels Piping.* 2023;201(1):105000. doi:10.1016/j.ijpvp.2022.104870.
16. Luo X, Lu S, Shi J, Li X, Zheng J. Numerical simulation of strength failure of buried polyethylene pipe under foundation settlement. *Eng Fail Anal.* 2015;48(03):144–52. doi:10.1016/j.engfailanal.2014.11.014.
17. Jasoliya D, Untaroiu A, Untaroiu C. A review of soil modeling for numerical simulations of soil-tire/agricultural tools interaction. *J Terramech.* 2024;111(1):41–64. doi:10.1016/j.jterra.2023.09.003.
18. Ma WJ, Li HS. Experimental study on rockfall impact response of buried gas pipeline. *China Meas Test.* 2018;44(9):23–8. (In Chinese).
19. Dong F, Bie X, Tian J, Xie X, Du G. Experimental and numerical study on the strain behavior of buried pipelines subjected to an impact load. *Appl Sci.* 2019;9(16):3284. doi:10.3390/app9163284.
20. Berangi M, Lontra BM, Anupam K, Erkens S, Van Vliet D, Snippe A, et al. Gradient boosting decision trees to study laboratory and field performance in pavement management. *Comput Aided Civ Infrastruct Eng.* 2025;40(1):3–32. doi:10.1111/mice.13322.
21. Luo J, Yuan Y, Xu S. Improving GBDT performance on imbalanced datasets: an empirical study of class-balanced loss functions. *Neurocomputing.* 2025;634(22):129896. doi:10.1016/j.neucom.2025.129896.
22. Zhang Y, Ren W, Lei J, Sun L, Mi Y, Chen Y. Predicting the compressive strength of high-performance concrete via the DR-CatBoost model. *Case Stud Constr Mater.* 2024;21:e03990. doi:10.1016/j.cscm.2024.e03990.
23. Mesghali H, Akhlaghi B, Gozalpour N, Mohammadpour J, Salehi F, Abbassi R. Predicting maximum pitting corrosion depth in buried transmission pipelines: insights from tree-based machine learning and identification of influential factors. *Process Saf Environ Prot.* 2024;187(7):1269–85. doi:10.1016/j.psep.2024.05.014.
24. Li H, Li L, Chen X, Zhou Y, Li Z, Zhao Z. Addressing the inspection selection challenges of in-service pipeline girth weld using ensemble tree models. *Eng Fail Anal.* 2024;156(6):107852. doi:10.1016/j.engfailanal.2023.107852.
25. Fang J, Cheng X, Gai H, Lin S, Lou H. Development of machine learning algorithms for predicting internal corrosion of crude oil and natural gas pipelines. *Comput Chem Eng.* 2023;177:108358. doi:10.1016/j.compchemeng.2023.108358.
26. Peng S, Zhang Z, Liu E, Liu W, Qiao W. A new hybrid algorithm model for prediction of internal corrosion rate of multiphase pipeline. *J Nat Gas Sci Eng.* 2021;85(5):103716. doi:10.1016/j.jngse.2020.103716.

27. Abualigah L. Particle swarm optimization: advances, applications, and experimental insights. *Comput Mater Contin.* 2025;82(2):1539–92. doi:10.32604/cmc.2025.060765.
28. Allaoui M, Belhaouari SB, Hedjam R, Bouanane K, Kherfi ML. T-SNE-PSO: optimizing t-SNE using particle swarm optimization. *Expert Syst Appl.* 2025;269(1):126398. doi:10.1016/j.eswa.2025.126398.
29. Sepúlveda E, Vandervorst F, Baesens B, Verdonck T. Enhancing explainability in real-world scenarios: towards a robust stability measure for local interpretability. *Expert Syst Appl.* 2025;274(8):126922. doi:10.1016/j.eswa.2025.126922.
30. Fryer D, Strümke I, Nguyen H. Shapley values for feature selection: the good, the bad, and the axioms. *IEEE Access.* 2021;9:144352–60. doi:10.1109/ACCESS.2021.3119110.
31. Xie X, Yao Y, Liu J, Li P, Yang G. Mechanical behavior of unsaturated soils subjected to impact loading. *Shock Vib.* 2016;2016:4703981. doi:10.1155/2016/4703981.