IMPROVING PREDICTION QUALITY WITH XGBOOST MODEL FOR BENDING CAPACITY OF X65 DEFECTED PIPE

Van-Minh-Thanh Le[1], Van-Tuan Vu[1], Tien-Dung Nguyen[1], Van-Long Do[1], Hieu Chi Phan[2,*]

[1]*Class 53rd of Civil and Industry Construction Engineering, Le Quy Don Technical University*

[2]*Civil Engineering Department, Institute of Techniques for Special Engineering, Le Quy Don Technical University*

Abstract

There is a desire of a quality model predicting bending capacity of the defected pipe for pipeline integrity assessment. While the analytical model faces with the difficulty in modeling the local defect and corresponding local stress, the Finite Element method is a valuable alternative. A common approach for predicting interested variable is to scrape the result data and develop a data-driven model such as the classical linear regression, CART or XGBoost. Along with generating numerical database with FEM, this study illustrates the advance of XGBoost model compared with its counterparts in predicting the moment capacity of the defected X65 pipe.

Keywords: *Burst pressure; defected pipe; finite element method; machine learning; XGBoost.*

1. Introduction

Transporting materials such as water, oil and gas with pipeline system is an important method where the pipe works under hazard environment such as soil. Along with the applied internal pressure or axial force, the buried pipe, which suffers from the degradation of the corrosive defect [1-4], often faces to the thread of appearance of bending moment due to the soil-related phenomenon [5-9]. Estimating the capacity of the pipe under the bending moment is a critical duty for evaluating the functional of the pipeline system. In various cases, the bending moment is the key factor leading to the failure of the pipe [5, 6, 8].

The difficulty for developing an analytical model is obvious existing with the local defects on the pipe wall [10]. Available model with analytical approach usually contains the strong assumption on the idealization of the material. Consequently, available studies in literature are mostly based on Finite Element Method, FEM, with defects are commonly simulated with sharped edges [1, 11-14].

[*] Email: phanchihieu@lqdtu.edu.vn

https://doi.org/10.56651/lqdtu.jst.v4.n02.340.sce

A data-driven model, which relies heavily on the database, is a common approach for the results from FEM [15, 16]. Various models can be used with the different levels of performance. The paper focuses on the XGBoost model proposed by [17] to develop the predicting model for bending capacity of pipe contains single defect. Comparison among XGBoost and the simple linear regression and the molecular model, Classification and Regression Tree is conducted to illustrate the outstanding of chosen model.

2. Material and methodology

2.1. Establishment of Finite Element Model

Simulations on the Abaqus are conducted to provide the required database for the machine learning models for pipes made of X65 steel. A true stress-strain curve in Fig. 1 is used for assessing the burst pressure of the defected pipe.

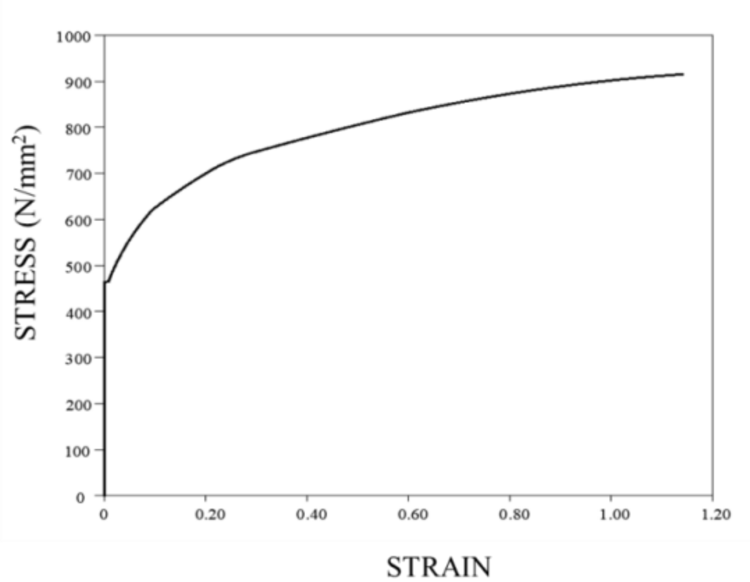


Figure 1. The true stress-strain curve of X65 steel (after [1]).

The critical inputs for the interested problems thus including the geometric inputs of pipe and defect: Outer diameter, D; wall thickness, t; depth of the defect, d; length of the defect, L; and width of the defect, w. Based on these critical inputs, half-model are established on commercial software, Abaqus, with a symmetric boundary to X-Y plane is set as in Fig. 2a. The cut is through the center of the defect modelled with sharp edge as in Fig. 2b. At the other end of the pipe is the Multiple Point Constrain pattern, MPC, (Fig. 2c) connects points in the pipe section to work as a unity beam. Center of this pattern is a reference point that applied moment is placed Fig. 2d.

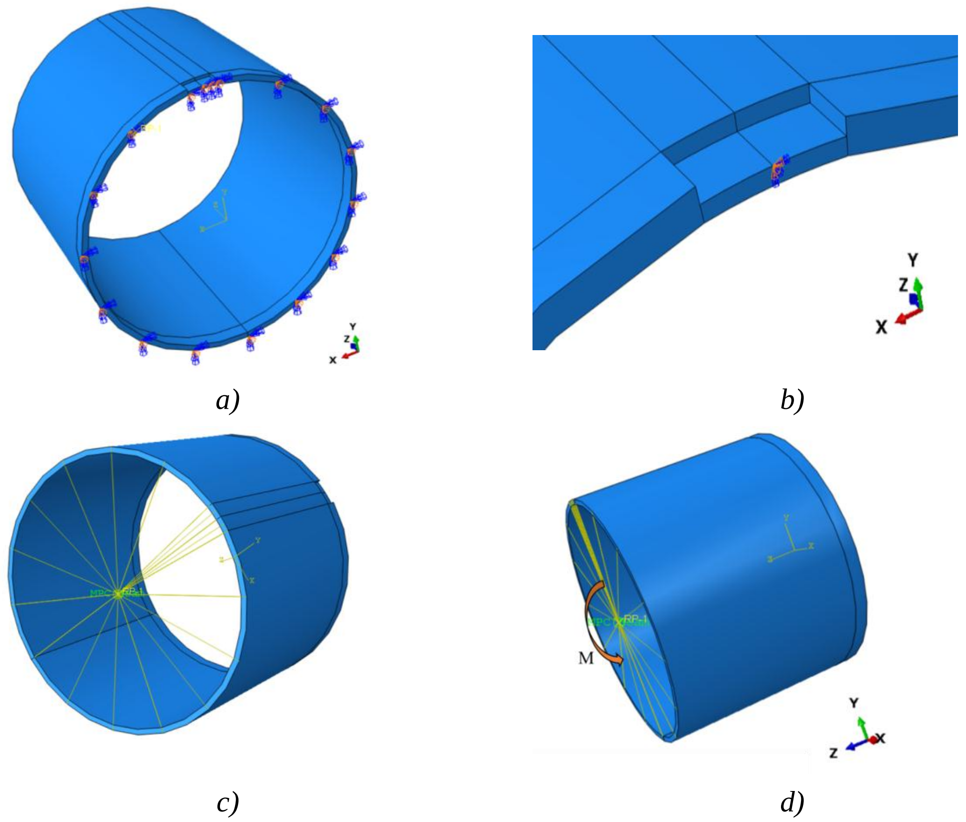


Figure 2. The half model for the defected pipe under bending moment.

A hybrid meshing system is used with the fine mesh located at the defects to capture the local stress appeared here. Meanwhile, the rest of the model uses a coarse meshing system with larger element size (Fig. 3).

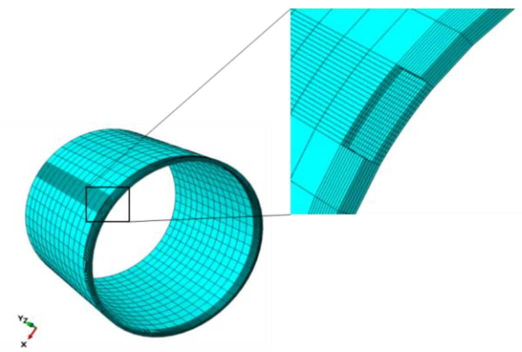


Figure 3. The meshing system for FE model.

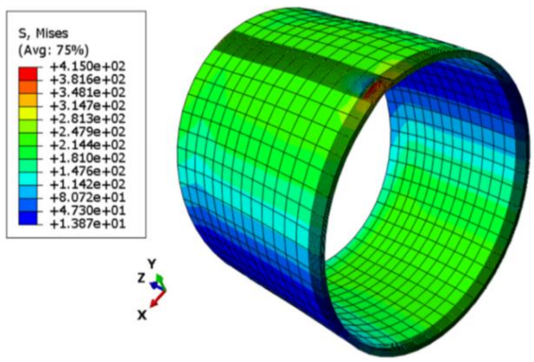


Figure 4. Von Mises distribution on the defected pipe.

The applied moment at the i^{th} step in Fig. 2 is increased linearly as the function of time step, t_i, as in Eq. (1a). If stress at any point in the pipe is reached, the critical time, t^*, and the bending capacity of pipe, M_{max}, is obtained as in Eq. (1b).

$$M_i = t_i \times \overline{M} \quad (1a)$$

$$M_{\max} = t^* \times \overline{M} \quad (1b)$$

where $\overline{M}$ is the maximum moment applied on pipe (i.e. M at time step = 1).

2.2. The XGBoost model

The conventional data splitting process with the ratio of 0.75/0.25 on train/test set is used in this study. The train set is used to develop model and the test set is for validating the developed model.

As earlier discussion, XGBoost model is used in the paper contains molecular weak models, the Classification and Regression Tree - CART proposed in [18]. The fundamental mechanism of CART is to split the inputs domains to seek for a structure that minimizes its errors to the actual values. An example of a CART is provided in Fig. 5.

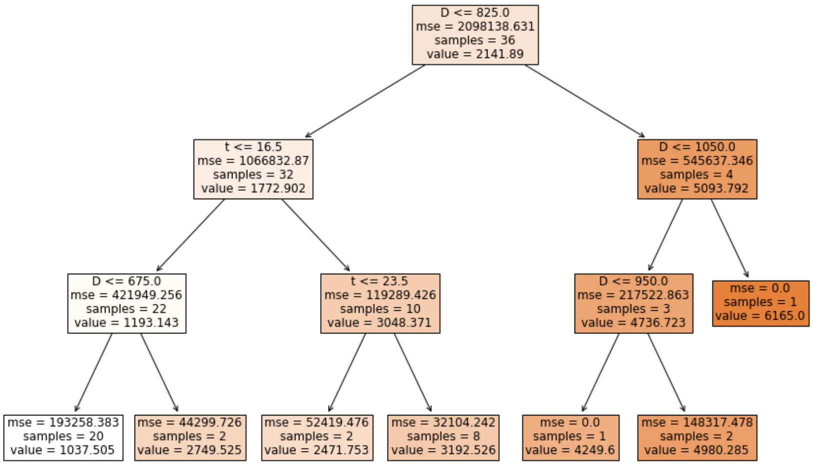


Figure 5. An example of the Classification and Regression Tree - CART.

Similar to CART or other machine learning models, the XGBoost model relies on the data (i.e., train set) to obtain the optimized configuration where the loss function on the train set is minimized. Instead of using a single tree for prediction, XGBoost is an ensemble model which based on a forest of CARTs and each CART is developed based on the subset of the train set. An interactive process of repeatedly finding the best objective function from the previous step is conducted to obtain the final XGBoost model.

To be specific, the Objective functions, OF, of the XGBoost at the m^{th} step can be mathematically expressed as in Eq. (2):

$$OF^m = L^m + \Omega^m \quad (2)$$

where OF^m is the hyper-parameters, L^m is the loss function, Ω^m is regularization term of the m^{th} interaction step.

In the case where Mean square error is used, the loss function in a given step can be written as in Eq. (3):

$$L^m = \sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 = \sum_{i=1}^{n}\left(y_i - \sum_{k=1}^{K} f_k^m\left(x_i\right)\right)^2 \tag{3}$$

where $\hat{y}_i$ is the predicted value and y_i is the actual value or results of FE analyses in this paper, $f_k^m\left(x_i\right)$ presented for the prediction of the model with inputs $\vec{x_i}$, K is the number of CARTs used.

The regularization term for the k^{th} tree is the function of number of leaves in tree, T, and weights value, w; and model hyper-parameters, λ and γ, Eq. (4). Intuitively, it is a penalty for the complexity of the model.

$$\Omega\left(f_k\right) = \gamma T + \frac{1}{2}\lambda \sum_{i=1}^{T} w^2 \tag{4}$$

The regularization term for the m^{th} step can be found in Eq. (5).

$$\Omega^m = \sum_{k=1}^{K}\Omega\left(f_k\right) \tag{5}$$

The OF of the training model at the step m^{th}, OF^m, is repeatedly improved by the previous prediction $\hat{y}^{m-1}$, Eq. (6).

$$OF^m = L^m\left(\hat{y}^{m-1}\right) + \Omega^m = \sum_{i=1}^{n}\left(y_i - \left(\hat{y}_i^{m-1} + f^m\left(\vec{x}_i\right)\right)\right)^2 + \Omega^m \tag{6}$$

3. Results

3.1. Results and database from FE

Observation on the behavior of defected pipe under the bending moment is provided in Fig. 6 and Fig. 7. Inputs of this case study are: D = 500 mm; t = 10 mm; d = 5 mm; L = 50 mm; w = 50 mm. The true stress-strain curve of X65 steel in Fig. 1 is used with the ultimate tensile strength σ_u is 563.8 MPa as in [19].

There are 8 reference points numbering from 1 to 8 as in Fig. 6. The stress developments of these points are given in Fig. 7 where Fig. 7a is for 6 on-section points and Fig. 7b is for 3 points on the outside of the defect. The manner of defining the failure step is provided in Fig. 7 with the applied moment M (Fig. 2d) linearly increases from time step 0 to 1, the $\overline{M}$ is 6×10^9 Nmm. With the given true stress-strain curve in Fig. 1 and the critical stress at any point reach 563.8 MPa is considered to be failure for the pipe, point 1 and 7 are simultaneously developed to this critical stress at the critical time t^* is 0.18 s. Consequently, the moment capacity of the pipe, M_{max}, is found by a simple transformation $M_{max} = \overline{M} \times t^* = 6\times10^9\times0.18 = 1.08\times10^9$ Nmm.

It can also be seen from Fig. 7 that the points 3 and 4 at the opposite side of the defect is slightly later reach the critical stress at around 0.20 s. The points at the surface of the defect (i.e., point 1, 2 and 7) have different the stress versus time step relationship but simultaneously reach the critical stress. Stresses at point 8 at the mount of the defect and point 5 and 6 close to the neural line are by far lower than point 1, 2 and 7.

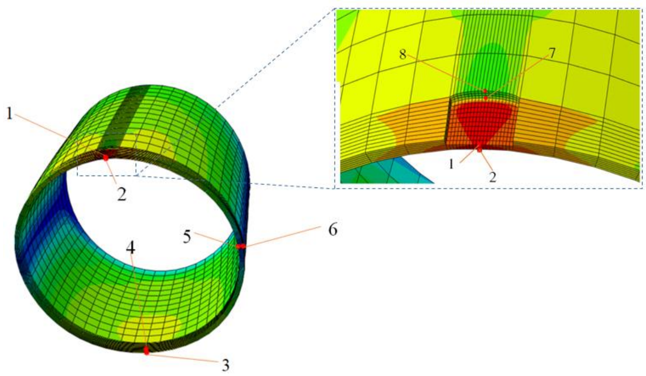
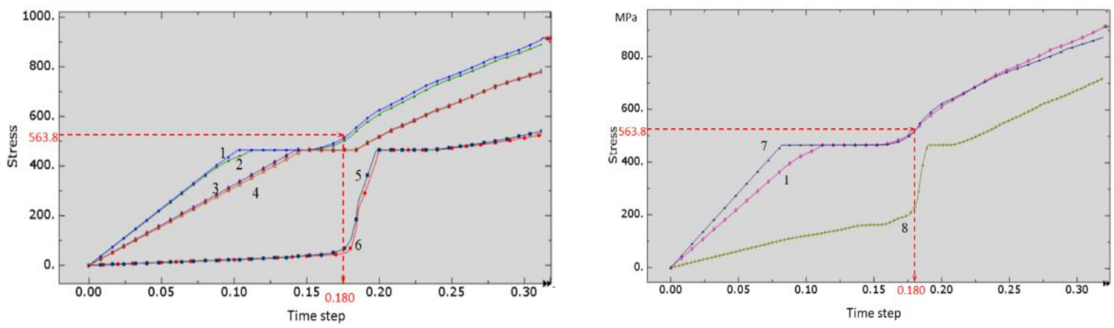


Figure 6. The observed points on pipes.



a) *b)*

Figure 7. Development of Von Mises stress at the 8 observed points.

The developed database from FE analyses results are selectively given in Table 1. The independent variables include the outer diameter, *D*; the wall thickness, *t*; the defect depth, length and width, *d*, *L*, *w*, respectively. A set of 48 samples is conducted to obtain the combined database for predicting the defected pipe made of X65 steel. Ranges of inputs are *D*~[100 : 1000] mm; *t*~[5 : 30] mm; *d*~[0 : 25] mm; *L*~[0 : 220] mm; *w*~[0 : 220] mm.

Table 1. The selected database

D	*t*	*d*	*L*	*w*	*M*
(mm)	(mm)	(mm)	(mm)	(mm)	(kNm)
500	10	0	0	0	1290
250	10	0	0	0	310
750	10	0	0	0	2960
1000	10	0	0	0	5365
…	…	…	…	…	…
100	10	5	50	50	34
500	7.5	5	50	50	826
500	12.5	5	50	50	1384
500	15	5	50	50	1679

3.2. The developed XGBoost model

The database in the previous section is used to develop 3 data-driven models, including the classical linear regression, CART and XGBoost models. The first 2 models are used for comparing with the developed model with XGBoost algorithm. Fig. 8 provides the predicted versus simulated values of these models (i.e. results from machine learning models and results from FE simulation, respectively). It can be seen that the XGBoost model has a clear concentration of data points around the 1:1 line. Meanwhile, the other models intuitively have larger residuals with the less focus of data points to 1:1 line.

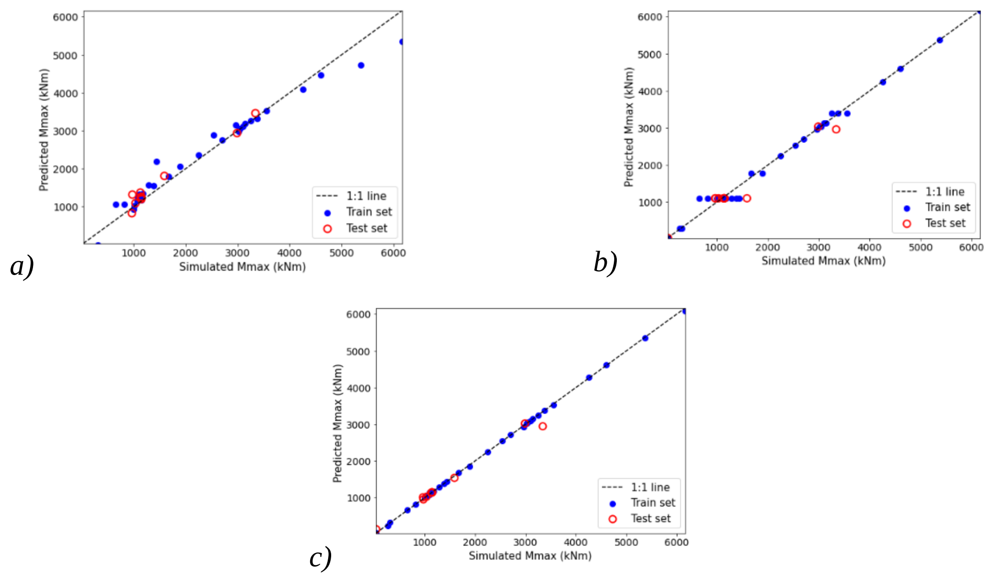


Figure 8. Predicted versus simulated values of
a) Linear regression; b) CART; and c) XGBoost models.

Official validating metrics on both train and test set of models are provided in Table 2. The XGBoost model has the “best” quality when its R^2 on train and test set are 0.9998 and 0.9807, respectively. Its molecular models, CART, are followed by a slightly lower R^2 values. In reverse, the linear regression has the slight over-fitting with R^2 drops from 0.9417 (lowest to other) to 0.8544. Errors (i.e. Mean Absolute Error, MAE, and Root Mean Square Error, RMSE) of XGBoost models are also observed to be consistently the lowest compared to their counterparts on both train and test set.

Table 2. Validating metrics for developed data-driven models

Model	**Data set**	**R^2**	**MAE**	**RMSE**
Linear regression	Train set	0.9417	223.5844	349.7949
	Test set	0.8544	210.5585	330.5265
CART	Train set	0.9921	72.5653	128.7489
	Test set	0.9519	120.9204	190.0083
XGBoost	Train set	0.9998	12.7645	19.3188
	Test set	0.9807	55.8139	120.4511

4. Conclusion

In this study, an FE model has established based on the true tress-strain for X65 steel to generate a database for the data-driven models predicting moment capacity of pipe. Linear regression, CART and XGBoost models are developed and validated on the test set to observe their prediction with unfamiliar data. XGBoost shows its advantage by observation on the predicted versus simulated the capacity of pipe. More official validation has numerically revealed the outstanding of XGBoost compared to other models.

Further work can be considered based on current results. For instance, expanding the data for other materials; incorporating other advanced techniques such as reliability assessment accounting to the uncertainty of the model; and application of methods improving the quality of the data-driven model, such as principal component analysis.

Acknowledgement

The paper is a result of science project for undergraduate student (the first four authors) titled as “Studying the mechanical behavior of defected pipe” of Le Quy Don Technical University.

Also, the study is a part of corresponding author’s project fund by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under Grant number: 107.02-2020.04.

References

[1] Mondal, B.C. and A.S. Dhar, "Burst pressure of corroded pipelines considering combined axial forces and bending moments," *Engineering Structures*, 186, 43-51, 2019.

[2] Velázquez, J., et al., "Predictive model for pitting corrosion in buried oil and gas pipelines," *Corrosion*, 65(5), 332-342, 2009.

[3] Keshtegar, B. and M.e.A.B. Seghier, "Modified response surface method basis harmony search to predict the burst pressure of corroded pipelines," *Engineering Failure Analysis*, 89, 177-199, 2018.

[4] Amaya-Gómez, R., et al., "Reliability assessments of corroded pipelines based on internal pressure - A review," *Engineering Failure Analysis*, 98, 190-214, 2019.

[5] Karamitros, D.K., C. Zoupantis, and G.D. Bouckovalas, "Buried pipelines with bends: analytical verification against permanent ground displacements," *Canadian Geotechnical Journal*, 53(11), 1782-1793, 2016.

[6] Uckan, E., et al., "Design Issues of Buried Pipelines at Permanent Ground Deformation Zones," *Disaster Science and Engineering*, 2(2), 53-58, 2016.

[7] Vazouras, P., S.A. Karamanos, and P. Dakoulas, "Mechanical behavior of buried steel pipes crossing active strike-slip faults," *Soil Dynamics and Earthquake Engineering*, 41, 164-180, 2012.

[8] Ni, P. and S. Mangalathu, "Simplified evaluation of pipe strains crossing a normal fault through the dissipated energy method," *Engineering Structures*, 167, 393-406, 2018.

[9] Trickey, S.A., I.D. Moore, and M. Balkaya, "Parametric study of frost-induced bending moments in buried cast iron water pipes," *Tunnelling and Underground Space Technology*, 51, 291-300, 2016.

[10] Mondal, B.C. and A.S. Dhar, "Improved Folias factor and burst pressure models for corroded pipelines," *Journal of Pressure Vessel Technology*, 140(1), 2018.

[11] Phan, H.C., et al., "An empirical model for bending capacity of defected pipe combined with axial load," *International Journal of Pressure Vessels and Piping*, 104368, 2021.

[12] Phan, H.C., A.S. Dhar, and B.C. Mondal, "Revisiting burst pressure models for corroded pipelines," *Canadian Journal of Civil Engineering*, 44(7), 485-494, 2017.

[13] Netto, T.A., Ferraz, U.S., and Estefen, S.F., "The effect of corrosion defects on the burst pressure of pipelines," *Journal of Constructional Steel Research*, 2005.

[14] Wang, N., and Zarghamee, M., "Evaluating fitness-for-service of corroded metal pipelines: Structural reliability bases," *Journal of Pipeline Systems Engineering Practice*, 2014.

[15] Silva, R., J. Guerreiro, and A. Loula, "A study of pipe interacting corrosion defects using the FEM and neural networks," *Advances in Engineering Software*, 38(11-12), 868-875, 2007.

[16] Ji, J., et al., "Prediction of stress concentration factor of corrosion pits on buried pipes by least squares support vector machine," *Engineering Failure Analysis*, 55, 131-138, 2015.

[17] Friedman, J.H., "Greedy function approximation: A gradient boosting machine", Annals of statistics, 1189-1232, 2001.

[18] Breiman, L., et al., "Classification and regression trees. Wadsworth & Brooks", Cole Statistics/Probability Series, 1984.

[19] Oh, C.-K., et al., "Ductile failure analysis of API X65 pipes with notch-type defects using a local fracture criterion", *International Journal of Pressure Vessels and Piping*, 84(8), 512-525, 2007.

NÂNG CAO CHẤT LƯỢNG DỰ BÁO KHẢ NĂNG CHỊU UỐN CỦA ỐNG CÓ KHUYẾT TẬT TỪ VẬT LIỆU THÉP X65 VỚI MÔ HÌNH XGBOOST

Lê Văn Minh Thành, Vũ Văn Tuấn, Nguyễn Tiến Dũng, Đỗ Văn Long, Phan Chí Hiếu

Tóm tắt: *Đạt được một mô hình dự báo khả năng chịu uốn của một ống có khuyết tật luôn là một yêu cầu cấp thiết đối với việc đánh giá tính toàn vẹn của ống. Trong khi các mô hình giải tích gặp phải khó khăn đối với việc mô hình các khuyết tật cục bộ và các ứng suất cục bộ tương ứng, phương pháp phần tử hữu hạn là một thay thế có giá trị. Một phương pháp phổ biến để dự báo các biến đầu ra là tận dụng các kết quả của phương pháp phần tử hữu hạn và phát triển một mô hình dựa trên dữ liệu như mô hình hồi quy tuyến tính cổ điển, CART hay XGBoost. Cùng với việc tạo ra bộ số liệu từ phương pháp phần tử hữu hạn, bài báo này minh họa tính ưu việt của mô hình XGBoost khi so sánh với các mô hình khác khi phát triển mô hình giúp dự báo khả năng chịu mô men của ống có khuyết tật làm từ vật liệu X65.*

Từ khóa: Áp lực thổi nổ; ống có khuyết tật; phần tử hữu hạn; học máy; XGBoost.

Received: 07/04/2021; *Revised:* 15/09/2021*; Accepted for publication:* 28/12/2021

❑