Journal of Mechanics Engineering and Automation 13 (2023) 127-135
doi: 10.17265/2159-5275/2023.05.001

Operation and Consideration of a Pipe Corrosion Inspection System Based on Human-in-the-Loop Machine Learning

Toshihiro Shimbo[1], Yousuke Okada[2] and Hitoshi Matsubara[3]

1. Mitsubishi Gas Chemical Company, Inc., Chiyoda-ku, Tokyo 100-8324, Japan

2. ABEJA, Inc, Minato-ku, Tokyo 108-0073, Japan

3. The University of Tokyo, Bunkyo-ku, Tokyo 113-8656, Japan

Abstract: The aim of this study is to improve the efficiency of external corrosion inspection of pipes in chemical plants. Currently, the preferred method involves manual inspection of images of corroded pipes; however, this places significant workload on human experts owing to the large number of required images. Furthermore, visual assessment of corrosion levels is prone to subjective errors. To address these issues, we developed an AI (artificial intelligence)-based corrosion-diagnosis system (AI corrosion-diagnosis system) and implemented it in a factory. The proposed system architecture was based on HITL (human-in-the-loop) ML (machine learning) [1]. To overcome the difficulty of developing a highly accurate ML model during the PoC (proof-of-concept) stage, the system relies on cooperation between humans and the ML model, utilizing human expertise during operation. For instance, if the accuracy of the ML model was initially 60% during the development stage, a cooperative approach would be adopted during the operational stage, with humans supplementing the remaining 40% accuracy. The implemented system's ML model achieved a recall rate of approximately 70%. The system's implementation not only contributed to the efficiency of operations by supporting diagnosis through the ML model but also facilitated the transition to systematic data management, resulting in an overall workload reduction of approximately 50%. The operation based on HITL was demonstrated to be a crucial element for achieving efficient system operation through the collaboration of humans and ML models, even when the initial accuracy of the ML model was low. Future efforts will focus on improving the detection of corrosion at elevated locations by considering using video cameras to capture pipe images. The goal is to reduce the workload for inspectors and enhance the quality of inspections by identifying corrosion locations using ML models.

Key words: HITL ML, collaboration between human and machine learning, diagnostic imaging, smart maintenance.

1. Introduction

Japanese chemical companies are facing several challenges, including overseas relocation of manufacturing facilities, retirement of skilled workers due to aging, and reevaluation of productivity through workstyle reforms. These challenges arise against the backdrop of intensifying competition with foreign companies, highlighting the need for improved productivity to enhance international competitiveness. Enhancing the operational reliability of facilities to maintain stable plant operations is essential to overcome these circumstances.

However, in Japanese chemical plants, the aging of equipment and reduction in the number of skilled maintenance personnel have increased the risk of accidents. This situation is exacerbated by the presence of aging plants and the impact of a declining workforce, necessitating the construction of systems that enable efficient work with a small number of personnel. This involves measures such as reducing workload, extending maintenance cycles, and improving

Corresponding author: Toshihiro Shimbo, M.Eng, Professional Engineer, research field: chemical engineering.

equipment reliability, with a focus on maintenance. According to industry experts, through these measures, Japanese chemical companies can enhance productivity in competition with foreign companies and increase their international competitiveness.

In recent years, the amount of available data has increased significantly owing to advances in measurement and communication technologies, such as the IoT (Internet of Things). To deal with this burgeoning volume of data, several AI (artificial intelligence)-based tools have been developed, and ML (machine learning) approaches are emerging. IoT and AI technologies can replace skilled workers. The introduction of AI in smart maintenance aims to construct ML models and improve their accuracy based on PoC (proof-of-concept) [2][3].

However, few such cases of AI implementation have been reported. One reason for this is the difficulty of developing highly accurate ML models based on PoC. In this context, this study utilized a HITL ML (human-in-the-loop machine learning) process during the implementation of ML models [1]. Thus, the challenge of constructing highly accurate ML models during the development phase is circumvented by developing an incomplete ML model and operating it with human support. As a case study, this study implements an AI corrosion-diagnosis system in a chemical company. To this end, we discuss the design concept for practicing HITL, verification results of actual operation, considerations on the collaboration between humans and AI, and future challenges and countermeasures.

2. Methods

2.1 Utilizing HITL ML

In this study, we defined HITL as a process involving human experts in the decision making of the ML model and operated the ML model while simultaneously improving it. Previous research on HITL has reported challenges in both the effectiveness of ML approaches, which are effective with well-defined problems and abundant data, and applicability in production environments, highlighting the need for further research on effective HITL methods within the ML loop [4]. However, to address the difficulty in developing highly accurate ML models during the PoC stage, we adopted HITL based on collaboration between humans and ML models, utilizing human experiential knowledge during operation. Thus, if the ML model exhibited an accuracy of 60% during development, human experts supported the remaining 40% of cases during operation.

Accuracy was measured in terms of the harmonic mean of recall and precision. Recall represents the proportion of cases identified as positive samples by both human experts and ML models. Precision represents the proportion of the ML model's positive predictions that agree with those of human experts.

2.2 Defining the Concept of an AI Corrosion-Diagnosis System Utilizing HITL ML

The Mitsubishi Gas Chemical, Niigata plant, faces severe external corrosion in its plants. This corrosion is primarily caused by salt content in the air from the Sea of Japan. The plant has exhibited a faster corrosion rate and more extensive corrosion than typical plants around the Japanese coast. Here, the external corrosion rate of carbon steel has been confirmed to be approximately 1.0 mm/Y, which is a significantly challenging environment compared to the average CUI (corrosion under insulation) rate of 0.2 mm/Y. Moreover, ESCC (pitting corrosion and environmentally assisted cracking) occurs in stainless steel, which necessitates a painted environment for stainless steel.

The conventional external corrosion-inspection method for chemical plant piping consists of the plant operators' first inspection of pipe photographs and the maintenance personnel's (skilled workers) second inspection for image confirmation and decision-making on remedies. In this inspection process, the operators record comments on the condition of the pipes and capture photographs during the first inspection, identifying areas where corrosion is progressing.

Subsequently, maintenance personnel conduct the second inspection, and if necessary, external contractors carry out precision inspections during regular repairs. While operators and maintenance personnel collaborate to address corrosion, the large number of captured images imposes a significant workload on both departments, making it crucial to prevent oversights in inspections. While the basic form of this process is established, visual judgment of the extent of corrosion produces subjective results, leaving room for improvement. Additionally, data accumulation and future analysis are deemed crucial, prompting the need for a more efficient and objective inspection method.

Fig. 1 illustrates the workflow of the traditional corrosion pipe inspection. This study addresses two challenges outlined in Fig. 1. Challenge 1 involves leveraging AI to reduce human workload and enhance the inspection quality, while Challenge 2 aims to design a system for centralized management of inspection information. Instead of pasting images into Excel files, this approach involves uploading them to a dedicated website for evaluation. Furthermore, the need for a system with search capabilities and the ability to create retraining data for ML models was recognized.

The desired state is depicted in Fig. 2.

The system image to be created, including the AI implementation and workflow efficiency, is depicted in Fig. 3.

A use-case diagram for the system of interest based on Fig. 3 is illustrated in Fig. 4. Fig. 4 demonstrates the functionality of the AI corrosion-diagnosis system, which includes automatic detection of corrosion based on images, linking images with relevant information, and retraining the ML model based on the correction results.

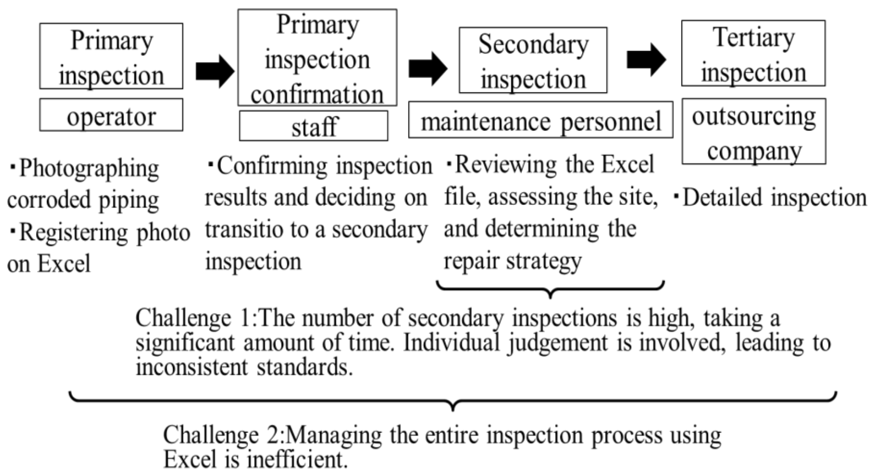

Fig. 1　Existing workflow and challenges.

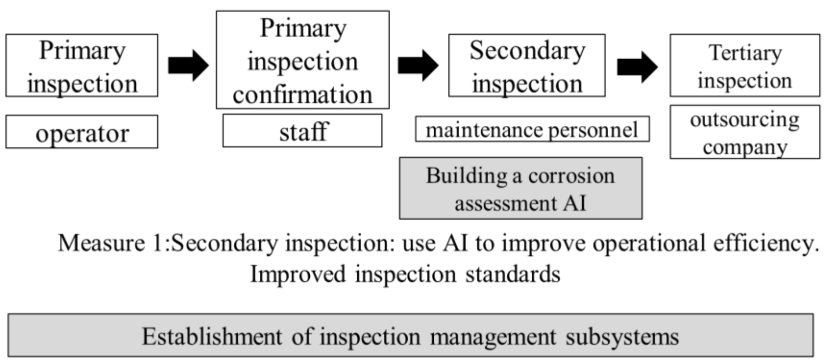

Fig. 2　The desired state.

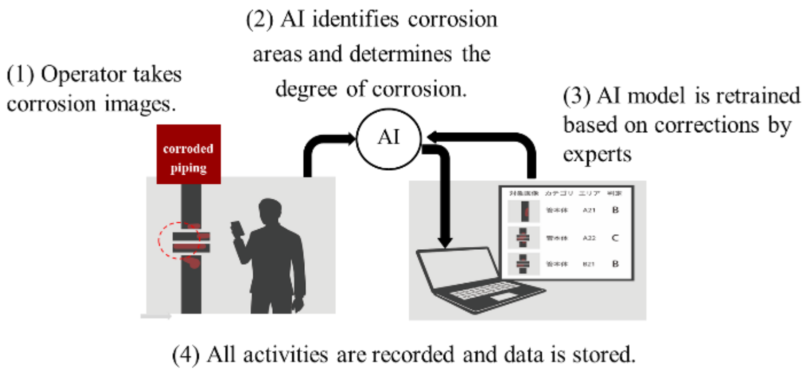

Fig. 3 Image of desired system.

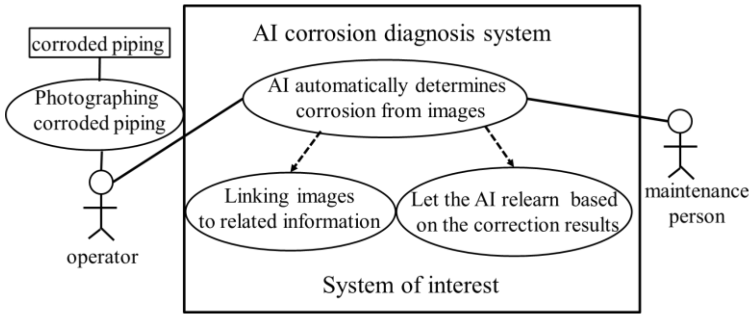

Fig. 4 Use-case diagram of the system of interest.

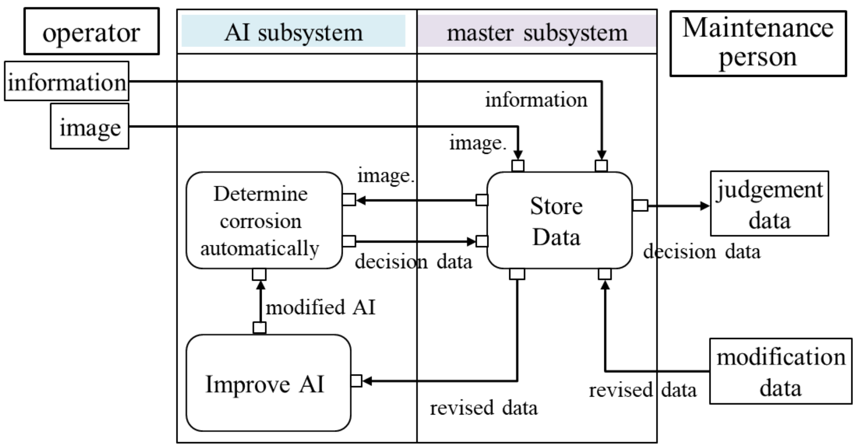

Fig. 5 Overview of the proposed system.

To achieve these functionalities, the AI corrosion-diagnosis system was designed with two subsystems—the AI subsystem and the master subsystem. Building upon Fig. 4, a functional flow diagram defining the interaction between humans and the system of interest is generated (Fig. 5). This incorporates HITL during the operational phase of the ML model.

First, images and information are transmitted from the driver to the master subsystem, where the data are stored. Subsequently, the images are transmitted to the AI Subsystem, which automatically detects corrosion. The assessment results are confirmed by maintenance personnel via the master subsystem, and they correct the errors as necessary. The corrected data are then returned to the AI subsystem through the master subsystem. The AI subsystem retrains the ML model based on these correction data. Thus, by incorporating the evaluations of maintenance personnel into the ML model through the master subsystem, the quality of inspections is improved. The AI corrosion-diagnosis system was designed, developed, and implemented based on the architecture illustrated in Fig. 5.

2.3 Points of Ingenuity in System Development

2.3.1 Standardization during the Creation of Learning Data

To create learning data for the ML model, evaluations by multiple maintenance experts had to be standardized. Over a period of three months, 11 maintenance experts performed the following activities:

(1) Conducted interviews to verbalize judgment methods.

(2) Identified differences in judgment criteria among skilled experts.

(3) Recorded verbalized information in the manual.

These activities elucidated the criteria and processes for manual decision-making, which enabled corrosion location and corrosion severity of the piping to be presented as learning data for the ML model. Corrosion severity was categorized into five stages in ascending order of severity—paint peeling, rusty appearance, mild corrosion, corrosion, and severe corrosion. However, the perceptions of the skilled workers were not entirely identical. An operation based on human-AI collaboration was necessary to account for this variation in training data. The corrosion judgment criteria for carbon steel created through these activities are presented in Table 1. Because of these activities, the criteria and processes for decision-making through manual work became lucid. This allowed for the presentation of corrosion locations and severity as learning data for the ML model.

2.3.2 Organization of Requirements for AI Subsystem and Master Subsystem

Requirements for the AI corrosion-diagnosis system were to develop a function that could identify corroded areas from images and determine the degree of corrosion using an ML model. Furthermore, the purpose of the system was to contribute to the improvement of plant safety management. It was intended to accumulate retraining data through corrections by maintenance personnel, enhancing the predictive accuracy of the ML model gradually. This would ensure that the accuracy of the ML model improved over time.

To achieve the workflow in Fig. 5, the requirements for the system had to be organized based on HITL to leverage human experience during operation. Coordination between the master subsystem managing image information and the AI subsystem conducting image diagnosis, as well as the creation of data for then ML model retraining in the master subsystem, was required.

The operators and maintenance personnel managing the corroded pipes demanded efficient execution of tasks from the AI system, as depicted in Fig. 1. Therefore, a master subsystem streamlining management, as depicted in Fig. 2, was necessary within the overall system. The requirements for the AI corrosion-diagnosis system are as follows:

- Integration of corrosion judgment by the ML model (primary inspection) with the ability to manage the implementation status of each inspection stage (allowing operators to easily confirm the number of unresolved cases).
- Seamless workflow (Fig. 1) from primary to tertiary inspection, as depicted in Fig. 2.
- Execution of ML model judgment for each item at the time of the operator's input during the primary inspection (improving the skills of the operator by allowing them to view ML model judgment results at the primary inspection stage).

Detailed requirements for the AI subsystem were defined as follows: having the ability to receive image data, save image data, submit corrosion degree determination to a mechanism, determine the degree of corrosion, evaluate judgment data, save judgment data, receive correction data learning requests, perform correction data learning requests, learn correction data, receive completion reports for correction data learning, save modified ML models, evaluate modified ML models, and deploy modified ML models. The system development was carried out in response to these requirements.

Table 1 Corrosion judgement criteria for carbon steel.

Valuation index	Significant corrosion	Corrosion	Minor corrosion	Rusting	Paint deterioration
Estimated amount of wall thinning	1.0 mm more	1.0 mm or less	0.2 mm or more	None	None
Rust bump size (5 times the amount of wall thinning)	5.0 mm more	5.0 mm or less	1.0 mm or less	None	None
Presence of rusting	Yes	Yes	Yes	Yes	None

3. Results and Discussion

3.1 Verification and Discussion of Results Obtained during Operation

An AI-based corrosion-diagnosis system was developed for carbon steel and stainless-steel piping without insulation. Fig. 6 depicts the AI-identified corrosion location and progression in corrosion severity based on photographs of pipes.

Table 2 summarizes the development of carbon steel and stainless-steel piping. The development of carbon steel piping required 3,800 annotation data points.

For stainless-steel piping, the proposed ML model exhibited a prediction accuracy of 74% based on 500 training data points, leveraging the experience of developing carbon-steel piping. The master subsystem focused on improvements during operation, such as rapid manual access to information, efficient image storage, and efficient creation of relearning data.

Table 3 lists the effects of the proposed AI-based corrosion-diagnosis system.

Given the nature of corrosion diagnosis, emphasis should be placed on recall, focusing on the diagnosis of significant corrosion severity while considering precision. In corrosion diagnosis, the ML model should provide strict diagnoses. In this context, while the recall rate of the system is approximately 70%, deviations indicating a severity level one step more significant than that by human diagnosis are considered operationally acceptable. Consequently, the safety level in the system's operation was maintained at an accuracy of approximately 90%.

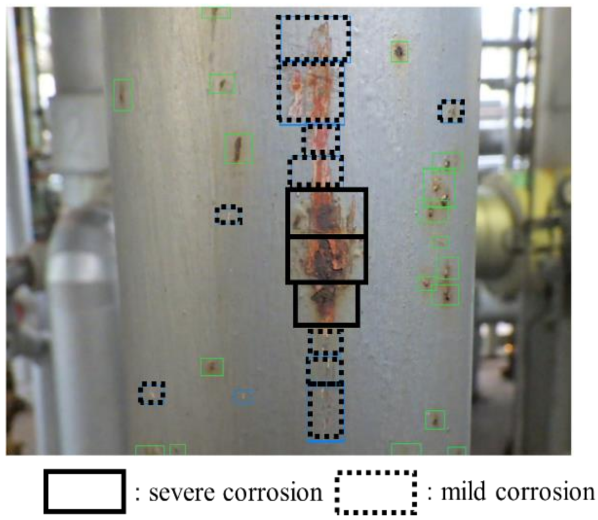

Fig. 6 AI-identified corrosion sites and progression in corrosion severity based on images of pipes.

Table 2 Development summary for carbon-steel and stainless-steel piping.

	Development period	Number of annotated images	Degree of corrosion determination	AI accuracy	Remarks
Carbon steel corrosion piping	2019.11-2021.3 (16 months)	3,800 sheets	5 steps	69%	Two developments were necessary: identifying corrosion sites and determining the degree of corrosion progression
Stainless steel corrosion piping	2021.9-2022.1 (4 months)	500 sheets	4 steps	74%	Carbon steel experience enabled the development of an operational level system in a short time.

Table 3 Effects of the AI-based corrosion-diagnosis system (initial implementation).

	Before implementing the system	After operating the system	Effects after introduction of the system
Operator	・ Captures corrosion images and uploads them on Excel	・ Captures corrosion images and uploads them to the system	・ 67% reduction in workload
Maintenance person	・ Corrosion images are captured and attached to Excel ・ Perform image diagnosis ・ Input diagnosis results into Excel	・ AI automatically performs corrosion assessment ・ Check AI results and relearn if different	・ Improved human inspection variability ・ AI accuracy of 69%, 30 reduction in workload

Overall, the workload is reduced by approximately 50%.

In the development stage, seven maintainers annotated 3,800 images. To ensure annotation consistency, attempts were made to standardize evaluation criteria and improve the quality of training data. However, distinguishing between Judgments B and C proved challenging, which led to some variation in the maintainers' diagnoses. Concerns were also raised about potential variations in evaluation criteria consistency over time. In the future, when accumulated correction data necessitate retraining, the criteria would need to be reevaluated and multiple maintainers would require confirming the annotations.

Interviews were conducted with 50 maintainers and operators to collate opinions based on actual work experiences. The results are summarized below:

(1) Even with the accuracy of the ML model being 69%, maintainers were able to use the system without issues.

(2) The image input time for operators was reduced from 3 to 1 min (reduction of 67%). This reduction was mainly attributed to the transition from Excel management to systematization, automating facility information input such as reducing image input tasks and shortening the time for diagnosis sheet preparation.

(3) Maintainers' workload decreased by 30%, contributing to a reduction in the overall workload by approximately 50% when combined with a reduction of 67% in operator workload. Despite a 70% diagnostic accuracy of the ML model, a final check by maintainers is still required, but this contributed to time saving.

(4) Clicking the "ML Model Judgment" button after operators input images resulted in the AI judgment results displayed in approximately 1-3 s. This early-stage ML model judgment enables operators to learn from the results, potentially improving the manual image-detection accuracy. The skilled maintainers' annotations contribute to the ML model's diagnosis, and the operators' image diagnostic skills improve over time, enhancing sensitivity to corrosion pipes.

(5) Both maintainers and operators have a high interest in ML model judgments, indicating a significant advancement in cooperation between humans and ML models. Further improvements in safety measures are expected.

From these results, the introduction of the AI corrosion-diagnosis system contributed to increased efficiency, education, and enhanced safety measures. This study demonstrated that the introduction and operation of the AI corrosion-diagnosis system can improve efficiency and accuracy of corrosion diagnosis. The concept of HITL proved crucial for efficient system operation, especially when the initial accuracy of the ML model was low. Comparing ML model judgments with human diagnostic results provides learning opportunities during operation, promoting continuous system improvement.

Additionally, the reduction in workload resulting from the system's introduction significantly simplified the tasks of both operators and maintainers. The transition from Excel management to systematization greatly contributed to data management efficiency. The high level of interest in the ML model by operators and maintainers should contribute to future collaborative efforts and improved safety measures.

Based on operational experience in this study, simplifying the corrosion-diagnosis classification into three stages was suggested to reduce variability in human diagnosis and further improve the ML model's accuracy. This should enhance the overall efficiency and effectiveness of the system.

Finally, the vast amount of data obtained from operating this system could be utilized in the future for analyzing and predicting corrosion trends within the factory. This can significantly contribute to long-term maintenance management strategies and risk reduction.

3.2 Insights obtained from Human-AI Collaboration

The AI corrosion-diagnosis system was developed based on the meticulous examination of 3,800 learning data by skilled maintainers. During operational retraining, the ML model relearned 110 data points that were incorrectly predicted at first. However, post-retraining, the predictive accuracy of the ML model remained nearly constant. This is attributed to the insufficient quantity of retraining data compared to the original learning data. To address this issue, the relationship between the number of learning images and the ML model accuracy was investigated. The application of 1,000 learning images resulted in an accuracy of 60%. As the number of learning data points increased to 1,500, the accuracy improved to 67%; thereafter, it plateaued. As the current ML model was trained with 3,800 images, without collecting a substantial amount of retraining data, the retraining is unlikely to be effective. Additionally, human involvement in creating retraining data makes achieving perfect accuracy challenging. Ongoing collaboration between humans and the ML model, with feedback on ML model judgments, is crucial for sustained improvement in the operation of the AI corrosion-diagnosis system.

Practical operation of HITL revealed inconsistencies between ML judgments and human diagnoses, particularly related to differences in identifying detachment areas. Variances in diagnosing coating- and thickness-reduction detachments were suggested. Where humans could distinguish between the two based on images, the ML model struggled. In actual operation, humans confirmed cases of detachment detected by the ML model and determined the severity of corrosion. Developing a ML model capable of distinguishing different detachment types should contribute to further accuracy improvements.

3.3 Future Challenges and Countermeasures

After the commencement of AI corrosion-diagnosis system operations, new challenges emerged. Differences in image quality between the development and operational stages—ideal corrosion images obtained from maintainers (experts) during development versus actual corrosion images from operators during operation—led to no improvement in accuracy during retraining. The operational images included pipes that were unclear and difficult to distinguish. Such differences in image quality are considered noise. Future efforts should focus on improving the method of capturing images.

Furthermore, in the operation of the ML model targeting stainless-steel pipes, degrees of corrosion observed differed from the four stages defined during development. To address these drawbacks, the addition of a new model to the existing ML model is currently under consideration.

Future research aims to utilize actual pipe thickness measurement data to enhance the accuracy of the ML model. This pursuit seeks to optimize system operation based on the coordination between the ML model and humans, achieving an efficient and reliable corrosion-diagnosis system.

Following the initiation of actual system operations, the oversight of corrosion during the primary inspection of pipes in elevated locations became a significant issue. The difficulty of visual confirmation from the ground and the requirement for intense focus posed a significant burden on operators. To address this issue, the exploration of a new method involving the use of video cameras to capture pipes and identify corrosion locations through the ML model is being considered. Using data from video cameras, a new ML model capable of identifying corrosion in high-risk areas (A, B ranks) is being developed. The primary goal of this model is to eliminate oversight in detecting corrosion. The ML model will handle corrosion identification, and inspectors will adopt a new operational method of capturing pipes with video cameras. This approach aims to reduce the workload on operators and improve the quality of inspections at elevated locations. This method is independent of the knowledge and experience of operators, and future applications involving drone-based video capture and ML model judgments are anticipated.

4. Conclusions

In this study, we developed an AI corrosion-diagnosis system that utilized HITL for visual inspection of corroded pipes in a chemical plant. The model was trained based on the expertise of maintenance personnel (skilled workers), achieving a prediction accuracy of 70%. Through the collaborative operation of human and ML models, we successfully reduced human workload by approximately 50%. The operational version of the AI system was constructed and implemented in an actual factory. Although the current state of using retraining processes for the ongoing improvement of the ML model's accuracy has not yielded significant results, the involvement of maintenance personnel for final confirmation of the model's determinations remains crucial. Hence, a portion of the HITL approach is being practiced.

The results of experiments showed an improvement in the accuracy from 60% with 1,000 learning images to 67% with 1,500 images. This suggests that, as retraining data accumulate, accuracy increases.

The collaboration between human and ML models provided valuable insights into the safety operations of the plant. Corrosion-diagnosis systems should be enhanced with AI through collaboration, addressing specific challenges, and incorporating human experience into operations. By validating the diagnosis and safe operation of the ML model in practical use, we demonstrated the importance of coordination between human and ML models in real-world applications, showcasing the new value it brings.

In the future, we aim to reduce the workload of inspectors and improve inspection quality by exploring methods such as using video cameras to capture images of pipes at elevated locations and employing ML models to identify corrosion spots.

Acknowledgements

The authors would like to express their gratitude to ABEJA Corporation and the Niigata Plant of Mitsubishi Gas Chemical Company, Inc., for their contributions to the development of the AI-based corrosion-diagnosis system.

The authors would like to express their gratitude to Professor Hidekazu Nishimura of the Graduate School of System Design and Management, Keio University, for his valuable advice in defining the system architecture.

References

[1] DoD Modeling and Simulation (M&S) Glossary, DoD 5000.59-M, DoD. 1998. "Directives Division. Archived from the original (PDF) on 2007-07-10. Retrieved 2009-04-22.

[2] Hata, M. 2019. "Corrosion Diagnosis System for Pipe Outer Surface by Image Data Analysis." Presented at the Accenture, 49th Petroleum and Petrochemical Conference (Yamagata Conference). (in Japanese).

[3] Tanaka, T., & Takeda, K. 2020. "Development of a Management Target Selection Method by Automatic Recognition of Pipes Using AI." Presented at the 63rd Joint Automatic Control Conference. (in Japanese)

[4] Christos, E., Petros, P., Luka, B., Vassilis, K., Apostolos, F., Christos, K., & Cristobal, R. 2019. "Enabling the Human in the loop: Linked Data and Knowledge in Industrial Cyber-Physical Systems." *Annual Reviews in Control* 47: 249-65.