Sensors and Materials, Vol. 37, No. 12 (2025) 5651–5667
MYU Tokyo

S & M 4270

Building a Monitoring Mechanism for Water Pipe Spatial Status in a Cyber-physical System

Zhong-Wei Ye,[1] Rong-San Lin,[1] Kun-Yi Huang,[1]
Gwo-Jiun Horng,[1*] and Yeou-Jiunn Chen[2]

[1]Department of Computer Science and Information Technology, Southern Taiwan University of Science and Technology, No. 1, Nantai St., Yungkang Dist., Tainan City 710301, Taiwan
[2]Department of Electrical Engineering, Southern Taiwan University of Science and Technology, No. 1, Nantai St., Yungkang Dist., Tainan City 710301, Taiwan

(Received November 9, 2025; accepted December 15, 2025)

Keywords: water pipes, intelligent image recognition, abnormal sound detection

In this paper, we propose a cyber-physical sensing framework for the real-time monitoring of water pipe integrity using multimodal detection technologies. The system integrates an unmanned vehicle equipped with an embedded camera for internal visual inspection and external microphones for acoustic sensing. The mobile platform enables flexible navigation inside pipes and captures high-quality images for detecting cracks and structural defects. Concurrently, acoustic signals collected from the pipe surface are transformed into the frequency domain via fast Fourier transform, and machine learning classifiers are employed to distinguish normal flow conditions from leakage-induced anomalies. The vision module utilizes a You Only Look Once v9-based detection model trained on an augmented dataset to enhance the accuracy of crack and rupture identification, particularly in large-diameter pipes. By combining acoustic and visual sensing within a unified cyber-physical architecture, the proposed system achieves efficient, low-cost, and scalable pipe condition monitoring. This approach provides the timely detection and localization of damage, offering practical value for reducing water loss and supporting more sustainable urban water infrastructure management.

1. Introduction

Global water scarcity continues to intensify as rapid population growth, urbanization, and aging infrastructure increase pressure on water distribution systems. Water leakage within pipes is one of the leading causes of water loss worldwide. Undetected leaks not only waste resources but also undermine the stability and safety of water supply networks. Traditional manual inspection and fixed-point sensor monitoring often fail to identify early-stage internal damage due to limited coverage, high cost, and insufficient sensing resolution.(1)

To address these challenges, intelligent sensing systems have emerged as a core component of modern pipe monitoring. These systems leverage multimodal sensing technologies—including

*Corresponding author: e-mail: grojium@stust.edu.tw
https://doi.org/10.18494/SAM6012

ISSN 0914-4935
https://myukk.org/

audio, vibration, image recognition, and cyber-physical integration—to improve leak recognition accuracy and provide continuous monitoring.[2,3] However, existing studies typically focus on single-modality sensing (either acoustic or image-based), lack flexible inspection platforms, or require specialized high-cost devices.

Prior research on pipe monitoring has mainly focused on different approaches. Acoustic leak detection uses fast Fourier transformer (FFT) features, time–frequency maps, or underwater acoustic signals. Image-based inspection relies on closed-circuit television devices or deep-learning-based crack detection.[3–11] Internet of Things and a cyber-physical system (CPS) improve distributed sensing and automated data collection.[12,13] However, most existing approaches still exhibit limitations, such as fixed sensor placement, inability to localize leaks at multiple positions, or lack of integration between different sensing modalities. Few studies combine both acoustic and image data in a unified CPS framework.

In this study, we propose an integrated CPS for the spatial monitoring of water pipe conditions. The system combines two major sensing subsystems: (i) an audio-based anomaly detection module using frequency-domain analysis and machine learning classification, and (ii) a vision-based pipe inspection module using an unmanned vehicle equipped with embedded sensors and a You Only Look Once (YOLO) v9 object detection model. Through real-time multimodal sensing, this CPS platform provides enhanced spatial awareness inside water pipes and supports the rapid detection of structural anomalies.[14–16]

The proposed system aligns closely with the research domain of intelligent sensing systems, sensor networks, and flexible CPS architectures. Microphone-based acoustic sensing enables the monitoring of pressure-related vibration signatures, whereas the image recognition module acts as a smart visual sensor for structural defect analysis. The unmanned vehicle serves as a mobile sensing agent, forming a dynamic and flexible sensor array along the pipe's inner surface. Together, these components form an intelligent/smart sensing system capable of comprehensive pipe condition assessment.

In this study, we address these gaps through the following contributions: (i) a dual-modal CPS sensing architecture, combining FFT-based acoustic anomaly detection with YOLO v9-based visual inspection; (ii) a modular, adjustable pipe testbed enabling controlled leak simulation at multiple spatial positions (holes 1–3); (iii) a flexible unmanned vehicle platform with an onboard camera and an embedded system for image acquisition, movement control, and real-time data transmission; (iv) a comprehensive machine learning evaluation, comparing naïve Bayes (NB), neural network (NN), random forest (RF), support vector machine (SVM), and *K*-nearest neighbor (KNN) across 300 and 600 Hz high-pass filtering conditions; and (v) a complete end-to-end sensing workflow, from data collection and preprocessing to classification and visualization. These novelties position the proposed system as an advanced CPS-based sensing solution for water infrastructure monitoring.

2. System Model

The proposed CPS water-pipe monitoring system integrates acoustic sensing and vision-based sensing into a unified architecture (Fig. 1), creating a lightweight yet effective multimodal

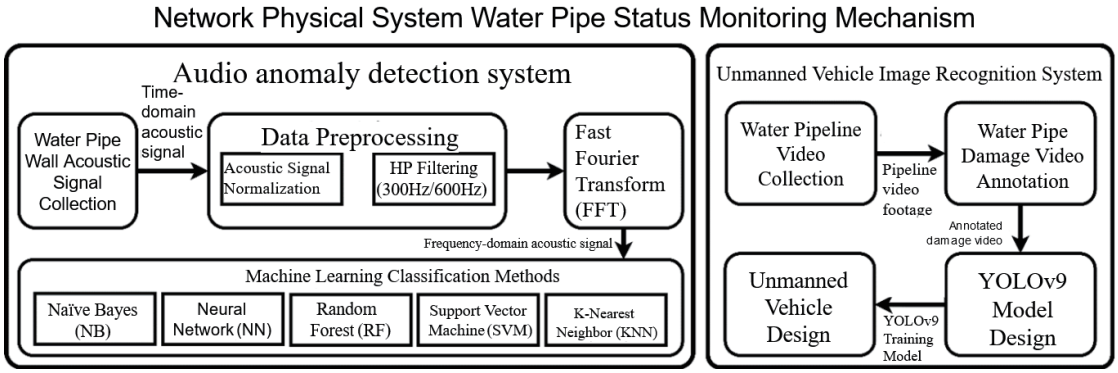


Fig. 1. System diagram of the proposed CPS.

platform capable of evaluating water pipe conditions. Instead of relying on traditional fixed sensors, this system combines external acoustic measurements with internal visual inspection to generate a comprehensive understanding of pipe integrity. By combining spectral acoustic sensing with deep-learning-driven visual perception in a unified cyber-physical architecture, the proposed system establishes a multimodal intelligent sensing framework. This integration enhances situational awareness inside pipes, improves anomaly detection robustness under diverse operating conditions, and demonstrates a significant advancement in CPS-based infrastructure monitoring. The system provides a scalable, low-cost, high-perception-fidelity solution that strengthens the role of intelligent sensing in future smart-city water management.

2.1 System architecture

The proposed system consists of two complementary sensing mechanisms working in coordination. The audio anomaly detection module attaches microphones to the outer pipe wall to record flow-induced vibration signals. After normalization and optional high-pass filtering, the signals are transformed into the frequency domain through FFT, enabling the extraction of characteristic spectral features associated with pipe ruptures. These features are then classified using machine learning algorithms to distinguish between normal pipe conditions and various hole positions. Through this process, the acoustic subsystem provides a low-cost, noninvasive means of identifying the presence and approximate location of structural anomalies.

In the data classification part, common machine learning classifiers such as NB, NN, RF, SVM, and KNN are used, and after each classifier completes the classification, we compare the various indicators to determine which method is most suitable for this data.[4–6] The classification results will include four categories: normal water pipe, hole 1 water pipe, hole 2 water pipe, and hole 3 water pipe. By using different features in the frequency domain, the water pipe data will also be different. By this method, the position of the water pipe hole can be located.

In parallel, the vision-based monitoring module employs a tracked unmanned vehicle equipped with an embedded controller and a camera to navigate inside the pipe and capture real-time images of the pipe's internal wall. The images are labeled and augmented to construct a

training dataset for the YOLO v9 model, which learns to recognize cracks and holes with high accuracy. During inspection, the trained model performs inference on captured images and transmits detection results wirelessly, enabling remote visualization and analysis. This mobile imaging approach enhances coverage and flexibility, overcoming the limitations of static monitoring devices.

The flowchart of the audio anomaly detection system in Fig. 2 illustrates the complete workflow of the water pipe hole detection system, covering multiple steps from the collection of audio data to the final detection results. The audio anomaly detection subsystem functions as a perceptual front-end that continuously acquires structural vibration signatures from the pipe surface through an externally mounted microphone. After normalization, the signals are filtered using a 300 or 600 Hz high-pass filter to suppress low-frequency hydraulic noise and enhance rupture-related spectral components. The refined signals are transformed to the frequency domain via FFT, enabling the extraction of high-resolution spectral features that reflect subtle variations in pipe-wall dynamics. These features are fed into multiple machine learning classifiers including NB, NN, RF, SVM, and KNN to discriminate normal states from multiple rupture locations.

The visual perception subsystem complements the acoustic sensing layer through image-based structural assessment. A tracked unmanned vehicle equipped with an embedded camera traverses the pipe and captures visual data reflecting internal surface textures, crack morphology, and potential faults. After labeled data augmentation, the images are used to train a YOLO v9 model, with hyperparameters tuned for defect-sensitive learning. Through iterative optimization, the model acquires robust perceptual representations of crack geometry and material discontinuities. Deployed on the vehicle, the model performs real-time inference, enabling the immediate detection and localization of visual anomalies. Results are transmitted via Wi-Fi to a remote station for maintenance decision-making.

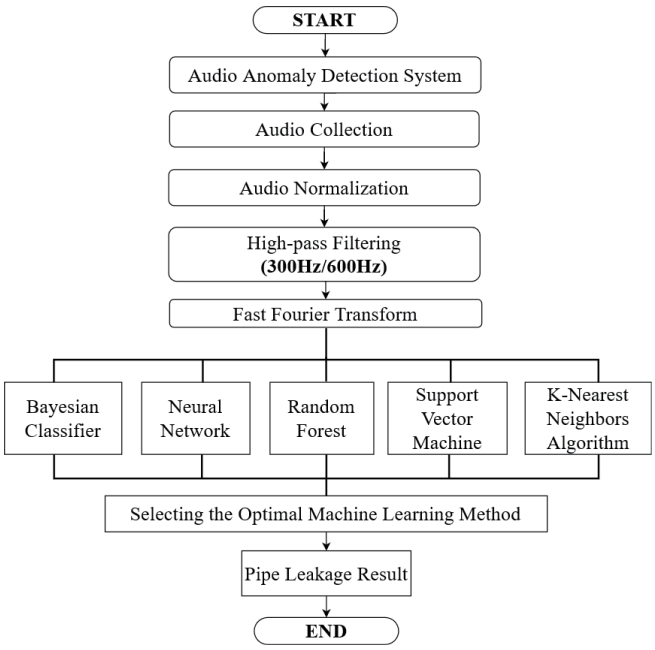
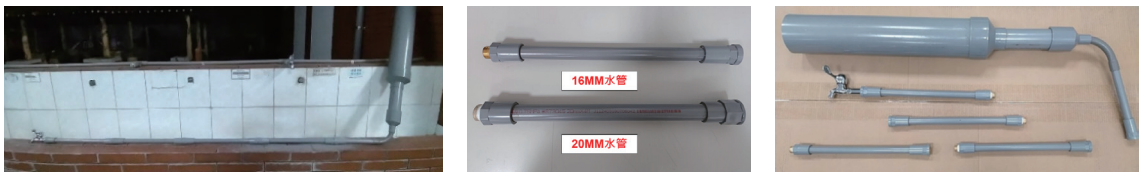


Fig. 2. Flowchart of proposed audio anomaly detection system.

2.2 Audio anomaly detection system

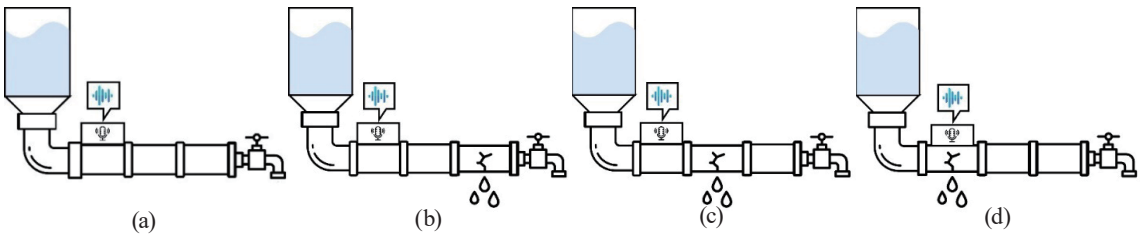
The audio anomaly detection system, as shown in Fig. 3(a), can adjust the actual diagram of the water pipe. This system uses a microphone to stick to the wall of the water pipe, records the sound inside the water pipe, and identifies the unique sound characteristics when it ruptures in the frequency domain. To achieve data collection, a set of adjustable water pipe systems has been designed in this part, as shown in Figs. 3(b) and 3(c), and as shown in the solid diagram of different pipe diameters in Fig. 3. Water pipes with internal diameters of 16 and 20 mm are designed, as shown in Fig. 4, which is a detachable water pipe with sections every 1.5 m. Each section is a movable water pipe, which is designed to facilitate the collection of sound characteristics of different positions of holes for the collection of time-domain data. The collected time-domain data are combined with FFT to convert the time-domain data into the frequency domain to capture the characteristics of the sound of hole breaking. Moreover, through machine learning classification techniques, it is possible to effectively distinguish the differences in frequency domain data between normal water pipes and damaged water pipes in various parts, to achieve the efficient detection and localization of abnormal water pipe states.

In this study, we analyzed the sound generated by the flow of water in a water pipe and used a microphone attached to the pipe wall for audio analysis. Therefore, schematic diagrams of water pipe holes are designed as shown in Fig. 4. Each section has a hole, and taking one hole as an example, the hole closest to the faucet is hole 1, the middle water pipe section is hole 2, and the hole closest to the water storage pipe is hole 3. By filling the left water storage pipe with water, the microphone is attached to the outside of the pipe wall for audio collection. In addition, the audio collected by the microphone is sampled at a rate of 8000 points per second, with a bit resolution of 16 bits and the number of channels being monophonic to accurately collect



(a) (b) (c)

Fig. 3. (Color online) Tested pipe configurations: (a) adjustable water pipe diagram, (b) different pipe diameters, and (c) detachable water pipe.



(a) (b) (c) (d)

Fig. 4. (Color online) Diagram of water pipe in (a) no holes, (b) hole 1, (c) hole 2, and (d) hole 3.

experimental data. This experimental design helps simulate sound scenarios in real water pipe environments, thereby improving the reliability of experimental results. By placing the microphone near the pipe wall, it can more sensitively capture the faint sound produced by the water flow.

The acoustic monitoring procedure consists of three main stages: signal acquisition, preprocessing, and feature extraction. First, vibration-induced audio signals are collected from the pipe wall using a microphone mounted externally. Each signal is sampled at 8 kHz with a 16-bit resolution to ensure sufficient temporal fidelity. Next, the raw signals undergo preprocessing, which includes normalization and high-pass filtering. A fifth-order high-pass filter with cutoff frequencies of 300 and 600 Hz is applied to suppress low-frequency hydraulic noise while retaining rupture-related components. This step enhances feature separability by eliminating irrelevant operational noise. Finally, the preprocessed signals are transformed into the frequency domain via FFT. The resulting spectral representation highlights leakage-induced energy spikes across multiple frequency bands. These spectral features serve as input for machine-learning classifiers in subsequent stages, enabling the accurate detection and localization of pipe anomalies.

Machine learning is applied to frequency-domain acoustic features to distinguish normal from damaged pipes and to localize leakage positions. Treating each spectral component as a feature enables the accurate characterization of rupture-related signatures and supports the real-time classification of pipe conditions. This approach enhances the reliability of leakage detection and provides actionable decision support for water infrastructure monitoring. The audio detection process begins with the acquisition of 16-bit monophonic acoustic data sampled at 8000 samples per second from microphones attached to the pipe surface. These raw signals are converted into the frequency domain via FFT, which highlights the high-energy spectral components produced by ruptures while suppressing irrelevant noise through preprocessing. The resulting spectral data are then used as input to five machine learning classifiers including NB, NN, RF, SVM, and KNN to evaluate their effectiveness in distinguishing normal conditions from different hole positions. This streamlined procedure eliminates redundant operations and focuses on the essential stages required for accurate leak detection.

In this study, we expect to use five machine learning algorithms for water pipe perforation and localization classification, including NB, NN, RF, SVM, and KNN. As shown in Fig. 2, after completing the data collection, the data in the frequency domain is subjected to preprocessing and FFT, and the data is segmented as input data for the machine learning algorithm. The commonly used 8:2 ratio is utilized for training and testing segmentation and training. After training, the output part consists of four categories: normal water pipe, hole 1 water pipe, hole 2 water pipe, and hole 3 water pipe. These categories can be used for water pipe hole detection and localization.

2.3 Image recognition of unmanned vehicles

The objective of this study is to develop an efficient unmanned vehicle platform and an intelligent real-time image recognition system for in-pipe inspection. The platform integrates a

camera module, an embedded controller, and a wireless communication unit, enabling autonomous navigation, image acquisition, and real-time data transmission. As shown in Fig. 5(a), the system is designed to capture high-resolution visual information from the inner pipe walls, providing essential data for structural assessment. The embedded hardware ensures stable operation and supports flexible movement within confined pipe environments, enabling continuous monitoring and responsive feedback during inspection.

Figure 5(b) illustrates the architecture of the proposed real-time image recognition system. The unmanned vehicle traverses the pipe and collects visual data used to identify cracks and defects. Captured images are labeled and processed to train a YOLO v9 detection model, which is optimized to recognize various forms of internal damage with high accuracy. This integration of mobile sensing and deep-learning-based analysis offers a comprehensive and scalable solution for pipe health assessment. The system enhances the speed and precision of defect detection, reduces manual inspection burdens, and supports cost-effective maintenance planning. By enabling the timely identification of structural issues, the proposed architecture contributes to more efficient water pipe management and promotes the sustainable utilization of water resources.

The unmanned vehicle-based vision subsystem follows a similarly concise process. As the vehicle moves through the interior of the pipe, it continuously captures images of the wall surface. These images are then labeled and augmented to expand dataset diversity, ensuring the robust training of the YOLO v9 detection model at a resolution of 640 × 640 over 120 epochs. Once trained, the model is applied during real-time inspection, enabling the immediate identification of cracks or holes and the wireless transmission of detection results. This integrated workflow supports efficient visual analysis without the unnecessary repetition of operational steps.

3. Experimental Results and Discussion

The proposed approach can be divided into two systems: an audio detection system and unmanned vehicle image recognition. The audio detection system will use five classifiers for

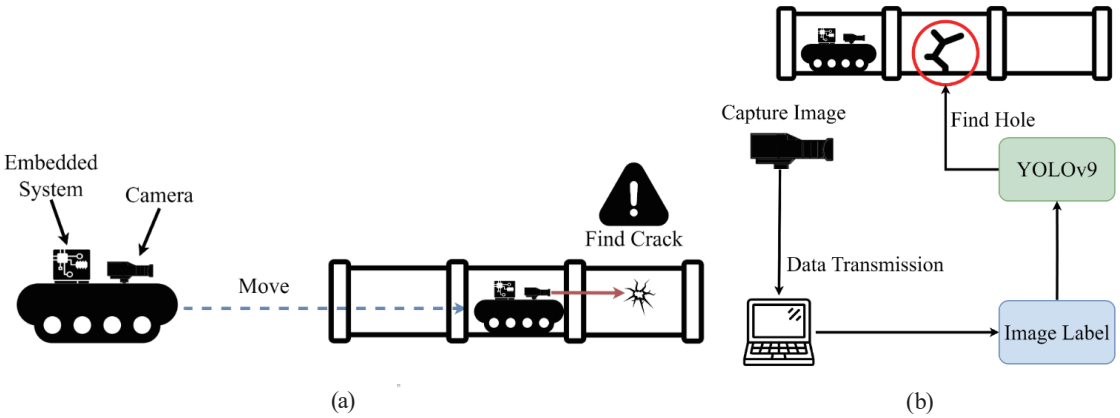


Fig. 5. (Color online) (a) Schematic diagram and (b) system architecture of unmanned vehicle operation.

experimentation, including NB, NN, RF, SVM, and KNN, trained on the same audio data. The image recognition of unmanned vehicles will be trained using YOLOv9, with a uniform image input size of 640 × 640 to ensure consistency in model input and improve prediction accuracy. Setting the batch size to 12 helps to maximize the use of GPU computing power while maintaining model stability. Setting the number of training steps (Epochs) to 120 provides sufficient learning time for the model to fully grasp the features in the data. The experimental software environment is Python version 3.6.10, Pytorch version 1.7.1, and Cuda version 1.0.130, and the hardware accessory environment is NVIDIA GeForce RTX 2080Ti GPU and Intel Core I9-9900K CPU @ 3.60 GHz.

The subsequent experiments are organized into three major components to comprehensively evaluate the proposed cyber-physical sensing framework. First, model evaluation metrics are introduced to establish the performance criteria for both audio-based and image-based detection systems. Second, the audio anomaly detection system is examined under two preprocessing configurations, applying high-pass filters at 300 and 600 Hz to suppress low-frequency noise and enhance rupture-related spectral features. Five machine learning classifiers, namely, NB, NN, RF, SVM, and KNN, are used to classify four categories: normal water pipe, hole 1, hole 2, and hole 3. Finally, the unmanned vehicle image recognition subsystem is assessed using a YOLO v9-based detection model trained on an augmented dataset of internal pipe images.

3.1 Model evaluation metrics

Model performance was evaluated using standard metrics, including Accuracy, Recall, Precision, F1-score, Average Precision (*AP*), and mean Average Precision (*mAP*). These indicators provide a comprehensive assessment of classification and detection reliability for both audio-based anomaly detection and image-based recognition tasks.

Accuracy reflects overall correctness and is defined as

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{1}$$

where *TP*, *TN*, *FP*, and *FN* are correctly identified positive cases, correctly identified negatives, incorrectly predicted positives, and missed positive cases, respectively. Recall measures the ability to correctly identify positive cases and is defined as

$$Recall = \frac{TP}{TP + FN}. \tag{2}$$

High Recall reduces missed detections, which is critical for safety and reliability. Precision evaluates the correctness of positive predictions and is defined as

$$Precision = \frac{TP}{TP + FP}. \tag{3}$$

High Precision minimizes false alarms, reducing unnecessary maintenance actions. *F*1-*score* provides a harmonic mean of Precision and Recall and is defined as

$$F1\text{-}score = \frac{Precision \times Recall}{Precision + Recall}. \tag{4}$$

For object detection tasks, *AP* quantifies the area under the Precision–Recall curve for each defect category. *AP* is defined as

$$AP = \sum_n \left(R_n - R_{n-1} \right) P_n, \tag{5}$$

where P_n and R_n are the accuracy and recall at the nth threshold, respectively. *mAP* represents the mean *AP* across all categories under various Intersection over Union thresholds. *mAP* is defined as

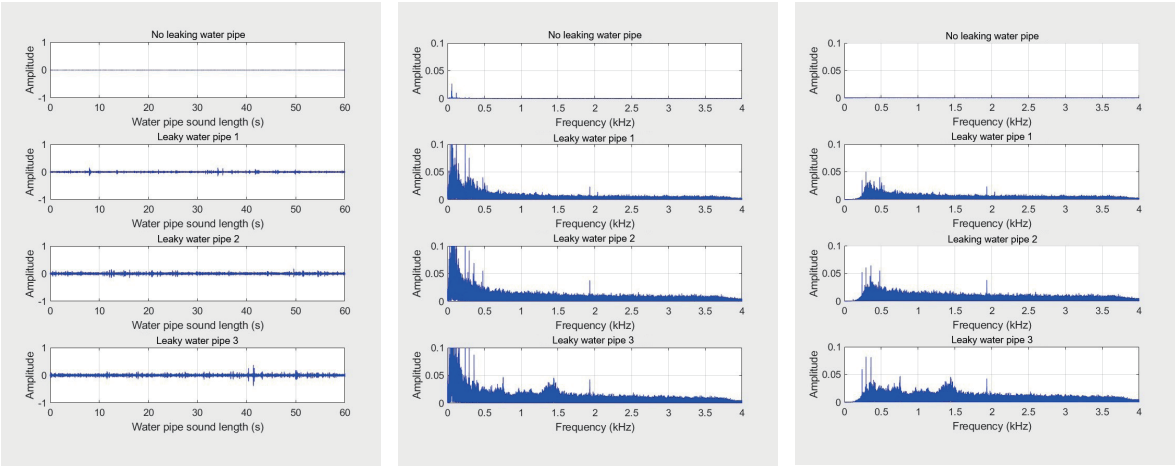
$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i. \tag{6}$$

High *mAP* values indicate robust detection and localization performance across diverse conditions, ensuring practical applicability in real-world pipe monitoring.

3.2 Experimental results of audio anomaly system

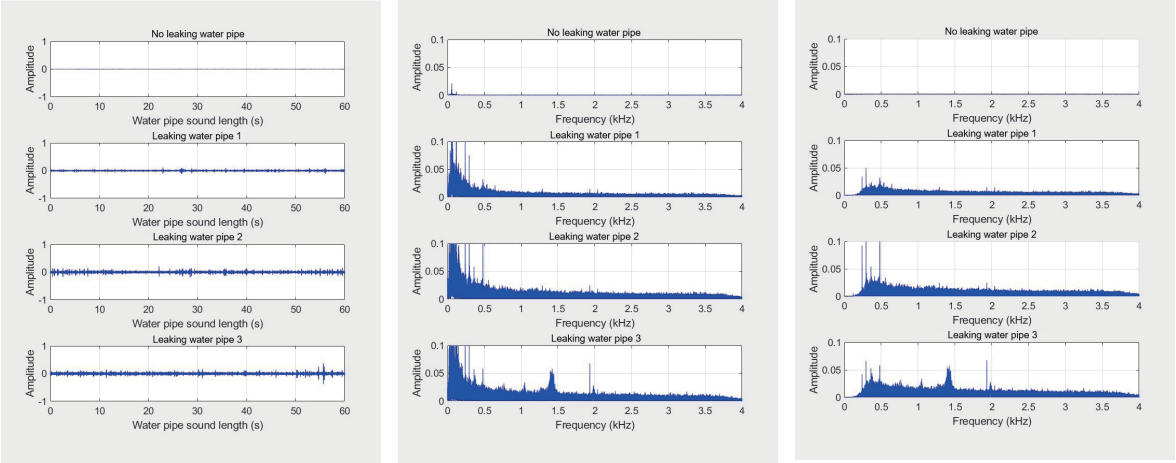
3.2.1 Analysis results of audio anomaly system

Figures 6 and 7 present the acoustic characteristics of water pipes with diameters of 16 and 20 mm under normal and ruptured conditions. Both pipes exhibit similar trends across time-domain



(a) (b) (c)

Fig. 6. (Color online) Wall sound of 16 mm water pipe in (a) time, (b) frequency, and (d) 300 Hz frequency domains.



(a) (b) (c)

Fig. 7. (Color online) Wall sound of 20 mm water pipe in (a) time, (b) frequency, and (d) 300 Hz frequency domains.

and frequency-domain analyses, confirming the consistency of rupture-induced acoustic signatures regardless of pipe size.

In the time domain [Figs. 6(a) and 7(a)], normal pipes demonstrate stable vibration amplitudes without impulsive peaks, indicating smooth water flow. In contrast, ruptured pipes display pronounced amplitude spikes and irregular fluctuations, which reflect turbulence and leakage transients generated at the damaged section. These observations suggest that time-domain analysis can provide preliminary rupture identification; however, its susceptibility to environmental noise limits reliability for practical applications.

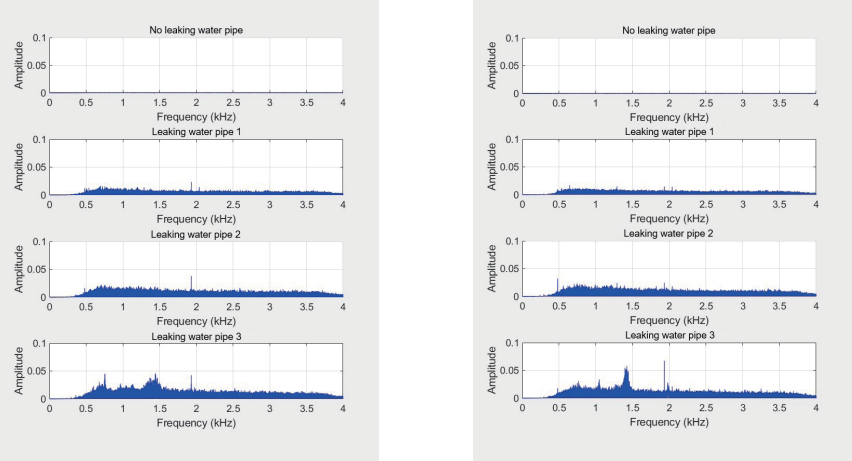
The frequency-domain representations [Figs. 6(b) and 7(b)] offer clearer discrimination between normal and damaged states. For both pipe sizes, ruptured conditions exhibit substantially higher amplitude energy across multiple frequency bands compared with normal pipes, with distinct spectral spikes appearing within approximately 0–300 and 0–600 Hz. These spikes correspond to leakage bursts triggered at rupture onset, making spectral analysis more robust against transient noise interference than time-domain signals.

To further enhance feature separability, a fifth-order high-pass filter with a cutoff frequency of 300 Hz was applied to both pipes [Figs. 6(c) and 7(c)]. This preprocessing step effectively suppresses low-frequency operational noise while preserving rupture-related high-frequency components. Combined with normalization, this approach ensures amplitude consistency across samples, producing more distinguishable spectral signatures for subsequent machine-learning-based classification and leak localization. The similarity in filtered results for both pipe diameters indicates that the proposed acoustic processing pipe is scalable and effective across various pipe sizes.

Overall, these findings confirm that rupture-induced acoustic patterns are consistent across different pipe diameters, and that frequency-domain analysis combined with high-pass filtering significantly improves anomaly detection robustness. This consistency provides a strong foundation for developing generalized machine-learning models capable of handling diverse pipe configurations in real-world water distribution systems.

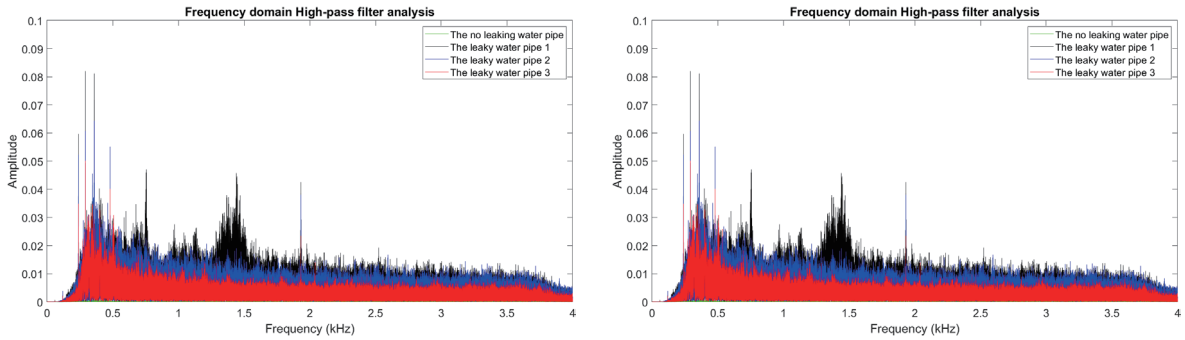
Figure 8 demonstrates that applying a high-pass filter effectively suppresses low-frequency noise below 300 and 600 Hz, thereby enhancing the spectral clarity of rupture-induced acoustic features. Compared with the unfiltered spectra in Figs. 5 and 6, the filtered results show significantly more pronounced high-frequency components for ruptured pipes, while normal pipes maintain relatively low and stable amplitudes. This confirms that high-pass filtering improves the system's sensitivity to leakage-related sound patterns and strengthens its ability to discriminate between normal operation and structural damage.

To further support rupture identification and localization, a movable pipe test configuration was implemented, as illustrated in Figs. 9 and 10. The microphone, positioned near the left water storage tank, collects vibration-induced acoustic signals under various hole locations. After high-pass filtering and FFT processing, the resulting frequency-domain signatures reveal distinct amplitude distributions corresponding to holes in different pipe sections. In the visualized spectra, green represents the normal condition, whereas red, blue, and black correspond to holes 1, 2, and 3, respectively. These color-coded plots exhibit clear stratification, indicating that each rupture location produces unique spectral energy patterns.



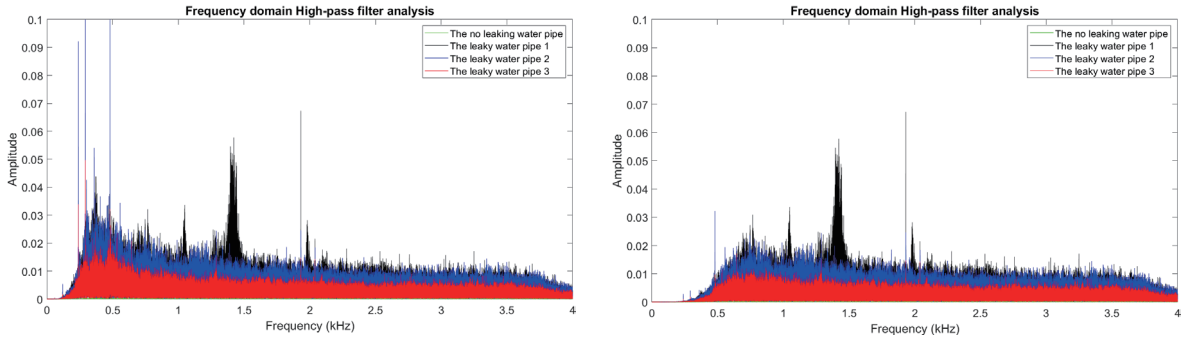
(a) (b)

Fig. 8. (Color online) Comparison of (a) 16 and (b) 20 mm water pipes with high-pass filter in 600 Hz frequency domain.



(a) (b)

Fig. 9. (Color online) Frequency responses of 16 mm water pipe with (a) 300 and (b) 600 Hz frequency domains.



(a) (b)

Fig. 10. (Color online) Frequency responses of 20 mm water pipe with (a) 300 and (b) 600 Hz frequency domains.

The clear separability among the three damage categories provides a robust foundation for machine-learning-based classification and allows accurate leakage localization along the pipe. Overall, the integration of high-pass filtering, frequency-domain transformation, and spatially varied data acquisition significantly enhances the robustness and diagnostic capability of the proposed acoustic monitoring framework.

3.2.2 Classification results of audio anomaly system

The audio anomaly detection system was evaluated under two high-pass filtering conditions (300 and 600 Hz) using five classifiers: NB, NN, RF, SVM, and KNN. Performance was assessed using Accuracy, Sensitivity, Specificity, and F1-score, as summarized in Table 1. Across both filtering conditions, RF consistently achieved the highest accuracy of 89–90%, coupled with strong sensitivity (89%) and specificity (96%). SVM ranked second, offering high sensitivity (95%) and specificity (95–96%) despite moderate overall accuracy (62–78%). NB exhibited the weakest performance under both conditions, limited by its assumption of feature independence, resulting in low accuracy (59%) and poor F1-scores (≈0.50). NN and KNN showed intermediate results but lacked robustness for multi-class classification.

This consistency across 300 and 600 Hz filtering indicates that classifier architecture plays a more critical role than filtering frequency in determining detection performance. RF's ensemble learning mechanism and SVM's kernel-based nonlinear mapping enable the superior handling of high-dimensional acoustic features, whereas NB struggles with correlated spectral data.

Figures 11 and 12 illustrate confusion matrices for all classifiers under both filtering conditions. RF demonstrates balanced recognition across all categories, achieving 98% accuracy for normal pipes and 86–88% for rupture positions (Holes 1–3). SVM maintains strong detection for normal pipes (96–97%) but exhibits reduced accuracy for rupture categories, particularly hole 2 (≈66%). NB performs poorly in distinguishing rupture positions, with recognition rates dropping to 24–57%, confirming its unsuitability for complex multi-class tasks. NN and KNN show moderate performance but fail to achieve the stability required for reliable anomaly detection.

Table 1
Comparison of various machine learning evaluation indicators at 300 and 600 Hz.

	300 Hz					600 Hz				
	NB	NN	RF	SVM	KNN	NB	NN	RF	SVM	KNN
Accuracy	0.59	0.59	0.89	0.78	0.64	0.59	0.60	0.90	0.62	0.59
Sensitivity	0.59	0.59	0.89	0.95	0.95	0.59	0.60	0.89	0.95	0.98
Specificity	0.43	0.59	0.96	0.96	0.82	0.43	0.60	0.96	0.95	0.87
F1-score	0.50	0.56	0.88	0.61	0.64	0.50	0.57	0.89	0.61	0.56

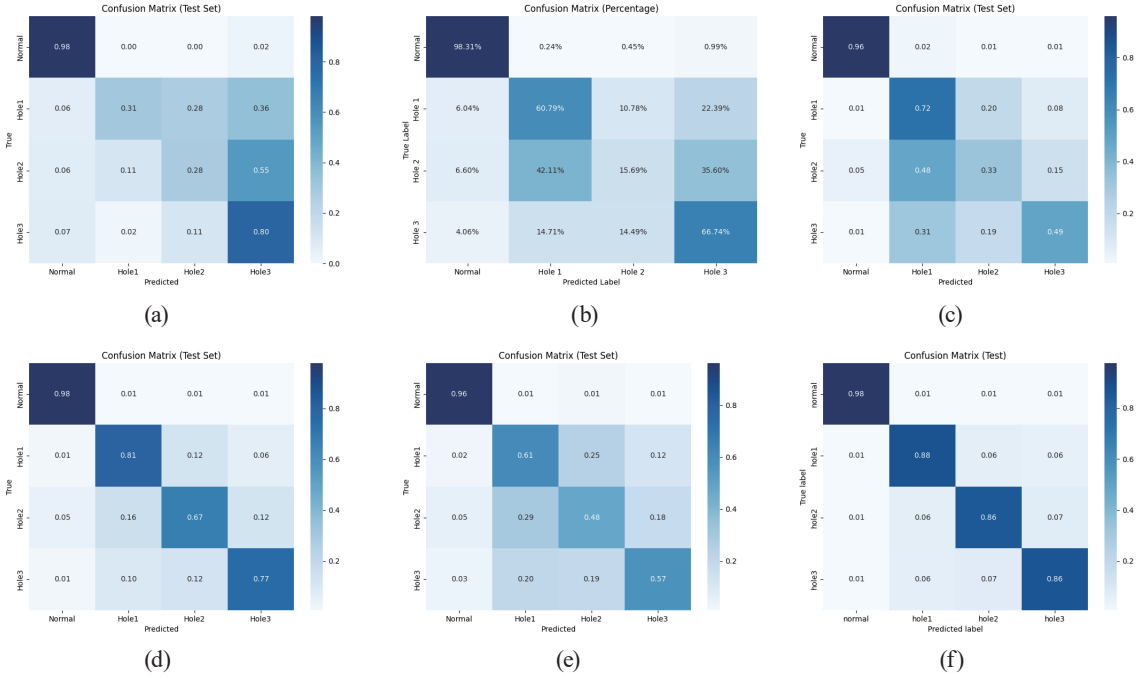


Fig. 11. (Color online) Confusion matrices for (a) KNN, (b) NN, (c) SVM with 10-point average, (d) SVM with 5-point average, (e) NB, and (f) RF with filtering condition of 300 Hz.

Overall, RF and SVM consistently outperform other classifiers under both filtering conditions, while NB exhibits substantial limitations across nearly all evaluation metrics. These findings validate the robustness of RF for high-dimensional, nonlinear acoustic anomaly classification and highlight SVM as a viable alternative when computational efficiency is prioritized.

3.3 Experimental results of image recognition for unmanned vehicles

In this study, the crack detection experiments were conducted on pipe sections in a dry, non-submerged environment. Although the pipe material is the same as that used in water supply systems, the inspection scenario simulated here does not involve underwater operation. The camera and sensing components were positioned externally and were not designed for underwater immersion.

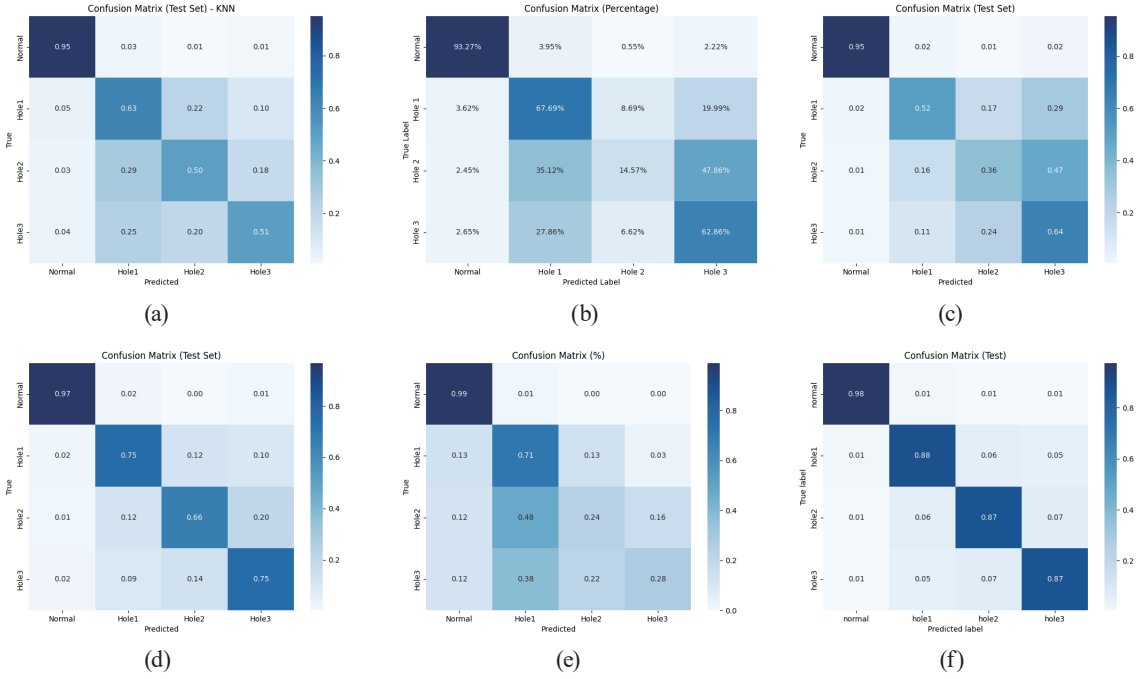


Fig. 12. (Color online) Confusion matrices for (a) KNN, (b) NN, (c) SVM with 10-point average, (d) SVM with 5-point average, (e) NB, and (f) RF with filtering condition of 600 Hz.

The unmanned-vehicle image recognition subsystem exhibited outstanding performance, as evidenced by the YOLOv9 confusion matrix (Fig. 13). All defect and background samples were correctly classified, achieving 100% accuracy during validation. This result demonstrates the robustness of the detection model and confirms its capability for real-time pipe defect identification under controlled experimental conditions. Such performance indicates that the proposed architecture effectively addresses challenges associated with visual inspection in confined environments, including variable lighting and complex surface textures.

Further evaluation using real pipe images (Fig. 14) validates the practical applicability of the YOLOv9 model. The system successfully detected multiple damaged points, with confidence scores of 0.69, 0.86, and 0.78 for three distinct defects. These results highlight two critical aspects of the proposed image recognition system. First, the detection reliability observed in both controlled and real-world scenarios demonstrates that the YOLOv9 model can consistently identify structural anomalies even under non-ideal conditions. This capability significantly reduces the likelihood of undetected cracks or holes, which is essential for maintaining the integrity of water pipe systems. Second, the variation in confidence scores across detected defects suggests that factors such as defect size, shape, and surrounding surface conditions affect detection certainty. Smaller or partially occluded defects tend to yield lower confidence values, indicating that while the model performs well overall, there is room for improvement. Enhancing dataset diversity through augmentation and incorporating advanced feature extraction techniques can further strengthen the system's robustness and ensure more uniform confidence levels across different defect types.

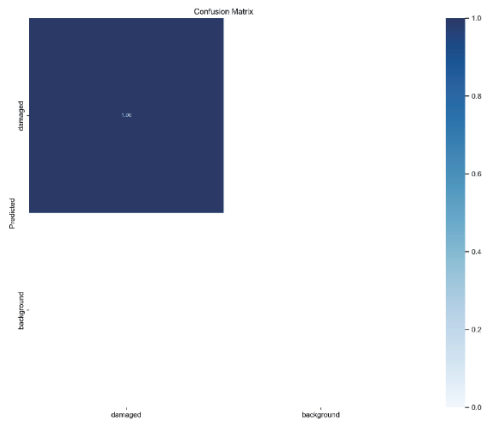


Fig. 13. (Color online) YOLOv9 confusion matrix.

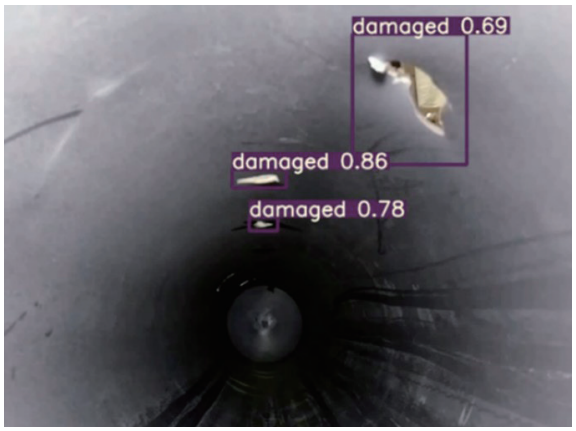


Fig. 14. (Color online) Image recognition training results.

From an operational perspective, achieving near-perfect classification in controlled tests and strong performance in real-world scenarios underscores the suitability of YOLOv9 for pipe monitoring applications. However, while the model demonstrates high precision and recall, its reliance on high-quality image acquisition implies that environmental factors—such as water turbidity, lighting, and camera positioning—remain critical for maintaining detection accuracy. Future work should explore adaptive illumination techniques and domain-specific data augmentation to further enhance robustness under diverse conditions.

4. Conclusions

We presented a cyber-physical sensing system for the spatial monitoring of water pipe conditions. By combining acoustic anomaly detection and visual inspection, the system provides robust multimodal detection capabilities. Random Forest outperforms all other classifiers in audio-based leak localization, whereas YOLOv9 achieves high performance in visual crack detection. The proposed CPS framework offers significant advantages: (1) high accuracy in detecting and locating pipe defects, (2) low-cost and flexible sensing architecture, and (3) real-time monitoring capabilities with mobile sensing agents. Future work will integrate continuous real-time sensing, advanced deep-learning-based audio classification, and deployment in operational-scale pipe environments. RF performs excellently in training accuracy, testing accuracy, sensitivity, specificity, and F1-score, with a training accuracy of 99%, a testing accuracy of 90%, a sensitivity of 89%, a specificity of 96%, and an F1-score of 89%. These indicators show that RF performs well in handling high-dimensional data and multi-class problems, with strong adaptability and stability.

Acknowledgments

This work was supported in part by the National Science and Technology Council (NSTC) of Taiwan under Grant nos. NSTC 111-2622-E-218-005 and NSTC 113-2622-E-218-002, and in part by the Center of Intelligent Healthcare, STUST, from the Higher Education Sprout Project, Ministry of Education, Taiwan.

References

1 A. Anish, R. Sharan, A. H. Malini, and T. Archana: Proc. 2nd Int. Conf. Automation, Computing and Renewable Systems (ICACRS) (IEEE, 2023) 1254–1257. https://doi.org/10.1109/ICACRS58579.2023.10404710
2 T. Özseven and M. Arpacioglu: Proc. 5th Int. Congr. Human-Computer Interaction, Optimization and Robotic Applications (HORA) (IEEE, 2023) 1–3. https://doi.org/10.1109/HORA58378.2023.10156803
3 J. Wei and Y. Li: Proc. 6th Int. Congr. Image and Signal Processing (CISP) (IEEE, 2013) 1277–1281. https://doi.org/10.1109/CISP.2013.6743869
4 A. Aradi and A. K. Varga: IEEE Int. Workshop Metrology for the Sea (MetroSea) (IEEE, 2023) 449–453. https://doi.org/10.1109/MetroSea58055.2023.10317350
5 M. I. Habibie and N. Nurda: Proc. 1st Int. Conf. Smart Technology, Applied Informatics, and Engineering (APICS) (IEEE, 2022) 48–52. https://doi.org/10.1109/APICS56469.2022.9918793
6 S. Badar, S. A. Butt, M. N. Khan, A. Mujtaba, and M. A. Ali: Proc. Int. Conf. Information Networking (2023) 433–437.
7 R. Kawada, A. Nishitani, and J. Kojima: Proc. OCEANS 2023 MTS/IEEE U. S. Gulf Coast (IEEE, 2023) 1–5. https://doi.org/10.23919/OCEANS52994.2023.10337397
8 G. Xiong, J. Wang, J. Zhang, H. Zhang, and C. Liu: IEEE/CAA J. Autom. Sinica **2** (2015) 320. https://doi.org/10.1109/JAS.2015.7152667
9 C. Semeraro, M. Caggiano, and M. Dassisti: Proc. Int. Conf. Cyber-Physical Social Intelligence (ICCSI) (IEEE, 2021) 1–6. https://doi.org/10.1109/ICCSI53130.2021.9736167
10 N. Mohd, I. Kumar, and A. A. Khurshid: Proc. Int. Conf. Communication, Security and Artificial Intelligence (ICCSAI) (IEEE, 2023) 647–651. https://doi.org/10.1109/ICCSAI59793.2023.10421085
11 M. R. Islam, S. Azam, B. Shanmugam, and D. Mathur: IEEE Access **10** (2022) 107177. https://doi.org/10.1109/ACCESS.2022.3212769
12 M. R. Islam, S. Azam, B. Shanmugam, K. Salah, and N. N. N. Abd. Rahman: IEEE Access **11** (2023) 123625. https://doi.org/10.1109/ACCESS.2023.3329467
13 M. Fishta, E. Raviola, and F. Fiori: IEEE Access **10** (2022) 108955. https://doi.org/10.1109/ACCESS.2022.3213973
14 A. Krizhevsky, I. Sutskever, and G. Hinton: Advances in Neural Information Processing Systems (Proc. NeurIPS) (2012). https://proceedings.neurips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html
15 Y. Xie, J. Hu, Q. Cui, H. Liu, X. Liu, and Y. Li: Sensors **20** (2020) 5040. https://doi.org/10.3390/s20185040
16 UNICEF: http://www.unicef.org/wash/water-scarcity (accessed Jun. 09, 2024).

About the Authors

Zhong-Wei Ye received his M.S. degree in computer science and information engineering from Southern Taiwan University of Science and Technology, Tainan, Taiwan, in 2024. His research interests include intelligent systems, IoT, and artificial intelligence. (bill89518@gmail.com)

Rong-San Lin received his Ph.D. degree in electrical engineering from National Cheng Kung University, Taiwan, in 2002. He is an associate professor in the Department of Computer Science and Information Engineering, Southern Taiwan University of Science and Technology, Tainan, Taiwan. His research interests include speech coding, audio signal processing, design of fast algorithms, and digital signal processing. (rslin@stust.edu.tw)

Kun-Yi Huang received his Ph.D. degree from the Institute of Computer Science and Information Engineering of National Cheng Kung University (NCKU), Tainan, Taiwan, in 2019. He is currently an assistant professor in the Department of Computer Science and Information Engineering of Southern Taiwan University of Science and Technology. His research interests include artificial intelligence and deep learning, speech and speaker recognition, multimedia emotion recognition, biomedical signal processing, and spoken language processing. (iamkyh77@stust.edu.tw)

Gwo-Jiun Horng received his Ph.D. degree in computer science and information engineering from National Cheng Kung University, Taiwan, in 2013. He is a full Distinguished Professor in the Department of Computer Science and Information Engineering, Southern Taiwan University of Science and Technology, Tainan, Taiwan. His research interests include mobile services, AIoT, intelligent computing, and cloud networks. (grojium@stust.edu.tw)

Yeou-Jiunn Chen received his Ph.D. degree from the Institute of Information Engineering, National Cheng Kung University, Tainan, Taiwan, in 1995 and 2000, respectively. He is currently a Distinguished Professor in the Department of Electrical Engineering, Southern Taiwan University of Science and Technology, Tainan, Taiwan. His research interests include biomedical signal processing, natural language processing, artificial intelligence, and deep learning. (chenyj@stust.edu.tw)