

Deep-learning-enabled Predictive Maintenance of Synchrotron Beamline Ion Pumps: A Real-time Anomaly Detection Framework

Cheng-Ying Chung,^{1*} Yu-Shan Huang,² and Longlife Lee²

¹Department of Mechanical and Electro-Mechanical Engineering, National Ilan University,
No. 1, Sec. 1, Shennong Rd., Yilan City, Yilan County 260007, Taiwan

²National Synchrotron Radiation Research Center,
101 Hsin-Ann Road, Hsinchu Science Park, Hsinchu 300092, Taiwan

(Received January 6, 2026; accepted August 31, 2026)

Keywords: deep learning, ion pumps (IPs), predictive maintenance (PdM), synchrotron beamlines

Machine health monitoring is used to forecast remaining useful life in predictive maintenance (PdM) to optimize production efficiency. Given the substantial cost and critical operational role of ion pumps (IPs) in synchrotron beamlines, in this study, we developed a deep learning framework for PdM to obtain failure predictions, allowing for maintenance scheduling to be optimized. We extracted historical data over a one-year period before an IP experienced failure. The developed framework uses sensor data from the previous 5 min to predict the target machine's health status (normal/abnormal) in the next minute. A prediction of abnormality indicates that the machine may fail within a maximum planning window of five months, which provides engineers sufficient time to schedule maintenance. The following four models were trained for use in the proposed framework and then evaluated: long short-term memory (LSTM), gated recurrent unit, independent recurrent neural network, and one-dimensional convolutional neural network. The results indicated that among these models, the LSTM model achieved the lowest false negative rate of 5%. This trained LSTM model performed well when applied to seven additional IPs on an experimental beamline. Furthermore, a proposed iterative mechanism for interpreting results considerably reduced the false positive and false negative rates of the LSTM model from 19 to 0.69% and from 5 to 0.01%, respectively. Thus, the proposed real-time anomaly detection framework is suitable for practical industrial application in beamline maintenance. Future enhancements will focus on integrating data from additional sensors to improve predictive accuracy.

1. Introduction

Making use of rapid advances in technology as part of Industry 4.0, companies have begun leveraging big data, cloud computing, and edge computing to enhance production efficiency and product quality.^(1,2) To achieve customized manufacturing, which is

*Corresponding author: e-mail: cychung@niu.edu.tw
<https://doi.org/10.18494/SAM6162>

characterized by low volume and wide variety, factories are integrating sensors, robots, and IoT technologies to enable rapid adjustments to their manufacturing processes. Rapid developments in digital twin technology, cyber-physical systems, and artificial intelligence have accelerated the advancement of smart factories. A key part of Industry 4.0 is the integration of production processes and machinery into a collaborative workspace for collecting, receiving, transmitting, and analyzing data. Such integration is aimed at predicting future production trends and identifying optimal solutions for potential problems.⁽³⁾ Methods based on advanced models, frameworks, and technologies have been developed to improve production process planning, supply chain management, transportation and logistics, and product design.⁽⁴⁾ Among these methods, the predictive maintenance (PdM) of equipment is critical for achieving efficient and high-quality production.^(5,6)

Machine maintenance can come in one of five types: corrective (reactive) maintenance (CM), preventive maintenance (PM), condition-based maintenance (CBM), PdM, and prescriptive maintenance.^(7–10) In CM, machinery is repaired after a failure has occurred. This approach is commonly used for low-cost components, which do not affect overall system operation when they fail. However, this method is costly to implement for expensive equipment or complex, large-scale systems (such as vehicle control systems). The unpredictable failure times of machines give rise to a large maintenance burden. Unexpected machine failures can also pose safety risks. Moreover, the entire production line must be halted for maintenance, at the cost of production efficiency and sometimes for a long period. In PM, machinery is regularly inspected and replaced on the basis of its specifications to prevent failure. The replacement cycle of a machine is equivalent to its specified lifespan scaled by a certain safety factor and is used to standardize the maintenance schedule for engineers. Although PM can reduce unplanned downtime, it may lead to premature replacement and thus resource wastage. Furthermore, machinery might unexpectedly fail before the scheduled maintenance period when subject to harsh operating conditions. In CBM, various sensors are used to monitor the real-time status of machinery, and alerts are triggered when the sensed parameters exceed predefined thresholds. To some extent, this approach enables maintenance to be performed prior to equipment failure. However, setting appropriate threshold values is challenging. If these values are set very low, false alarms may occur; however, if they are set very high, abnormal conditions prior to failure might go undetected, leading to missed alerts. To solve this problem, PdM is used to predict the health of machinery or equipment through statistical or machine learning methods.^(7, 8) In PdM, historical data are used to determine the health of machinery and predict its potential failure time or remaining useful life (RUL). This approach allows engineers to perform maintenance at the optimal time. Finally, in prescriptive maintenance, predicted failure types are used to propose constructive improvement measures, potentially allowing engineers to extend the RUL of a machine. In this study, we focused on PdM, developing a deep-learning-based framework for this task.

You *et al.*⁽¹¹⁾ proposed a statistical PdM method in which the long-term planning benefits of lifetime distribution analysis are combined with PdM-based precise RUL predictions for individual equipment to enhance average system availability. Lei *et al.*⁽¹²⁾ analyzed vibration data by using the weighted minimum quantization error to obtain a health indicator for bearings. They employed maximum likelihood estimation and particle filtering algorithms to predict the RUL of bearings. Qiu *et al.*⁽¹³⁾ proposed an improved robust approach for the PdM of rolling bearings. In this method, weak signal enhancement based on wavelet filtering is integrated with performance degradation assessment based on self-organizing maps to address the challenge of noise obscuring early defects and inconsistent machine degradation patterns. Hu and Chen⁽¹⁴⁾ proposed a strategy for identifying systems with a risk of “hard failure”. This strategy allows the estimation of RUL and the development of optimal maintenance plans when no fixed degradation threshold is available. The strategy is based on a Weibull proportional hazards model that incorporates degradation data simulated through a random-effects Wiener process as time-varying covariates; thus, it captures the role of degradation in increasing the risk of hard failure.

Cutting-edge PdM methods use AI to analyze large sets of data gathered from IoT sensors. Garcia *et al.*⁽¹⁵⁾ used monitoring parameters such as gearbox bearing temperature, generated power, nacelle temperature, and cooler fan status as inputs to a multilayer perceptron neural network model to predict the health status of wind turbine gearboxes. Moreover, they employed fuzzy logic to diagnose the types of fault shown by these gearboxes. Liao *et al.*⁽¹⁶⁾ proposed an enhanced restricted Boltzmann machine and introduced a regularization technique for automatically generating suitable features and machine degradation trends, thereby improving the accuracy of RUL predictions. Li *et al.*⁽¹⁷⁾ proposed an ensemble prediction method that considers the effect of system degradation on the accuracy of RUL predictions. Their ensemble framework included models such as recurrent neural networks (RNNs), support vector machines (SVMs), and Bayesian linear regression. They estimated the RUL of machinery by calculating the weighted sum of the outputs of these models, and they validated their approach by conducting case studies on aircraft bearings and aircraft engines. Cheng *et al.*⁽¹⁸⁾ adopted building information modeling, IoT technology, an artificial neural network model, and an SVM model to predict the state of electrical and plumbing components, thus facilitating the planning of appropriate maintenance schedules. Ruiz-Sarmiento *et al.*⁽¹⁹⁾ utilized Bayesian filters to predict the gradual degradation of machinery during a hot-rolling production process. Roberts and Ahmed⁽²⁰⁾ used random forest and XGBoost to predict potential faults from real-time data, providing early warnings to operators and preventing production line stoppages. In addition to industrial applications, deep learning has also been successfully applied to multiple tasks. For instance, a multi-view imaging framework was employed to detect distracted driving behaviors, demonstrating the versatility of deep learning in extracting complex features and improving recognition accuracy.⁽²¹⁾ In the field of biometric recognition, a hierarchical finger-vein feature extraction method was proposed for intelligent

vehicles,⁽²²⁾ and an improved finger-vein pattern method using wavelet-based analysis was developed for real-time personal identification systems,⁽²³⁾ both underscoring the effectiveness of advanced feature extraction techniques in biometric applications. Furthermore, depth-image-based rendering techniques have been improved through adaptive compensation methods on autostereoscopic 3-D displays for Kinect sensors, illustrating the broad applicability of deep learning and advanced imaging approaches in enhancing visual recognition and rendering tasks.⁽²⁴⁾

The Taiwan Photon Source (TPS), a 3 GeV synchrotron radiation facility operated by the National Synchrotron Radiation Research Center (NSRRC) in Taiwan, has 44 beamlines. To ensure a high-quality and safe experimental environment, the NSRRC prioritizes robust maintenance, including in vacuum systems, to ensure safety and operational integrity. In this study, we address vacuum system maintenance, explicitly focusing on the ion pumps (IPs) of the TPS 44A beamline. Traditional maintenance approaches predominantly rely on PM and CBM. However, to enhance operational efficiency, we developed a PdM system that uses real-time sensor data obtained through CBM for online machine health prediction. This system can enable proactive maintenance scheduling by beamline personnel.

The IP in the TPS is a critical and costly component for maintaining ultrahigh-vacuum environments. Its stable operation requires advanced maintenance technologies. Although numerous studies have explored PdM, few have specifically addressed deep-learning-based PdM frameworks tailored for IPs in synchrotron radiation facilities. In this study, we aim to develop a PdM system targeting this essential component. Given the rapid advancement of deep learning and its capability for automatic feature extraction, we focus on constructing a deep learning (DL) framework for PdM implementation.

Owing to the limited availability of high-quality historical data—particularly failure data—regression models are insufficient for accurately predicting the RUL of IPs. Therefore, in this research, we propose a classification-based approach to detect whether the IP has entered an abnormal state and to estimate the maximum possible RUL after such a transition.

In synchrotron beamlines, IPs are indispensable for maintaining ultrahigh vacuum, yet they exhibit distinct failure modes that complicate maintenance. Their finite titanium plate capacity leads to gradual degradation, making PM schedules inaccurate and often resulting in premature replacement. In addition, the localized saturation of the titanium plate can cause transient outgassing events, which may inadvertently trigger the vacuum interlock system. These phenomena are difficult to capture through CBM, as overly sensitive thresholds generate false alarms, whereas overly lenient thresholds risk missing genuine anomalies. Consequently, PdM is particularly critical in synchrotron facilities, as it enables the early detection of abnormal states, provides engineers with sufficient lead time to plan interventions, and ensures stable and safe beamline operation.

2. Methods

In this study, we attempted to develop and evaluate a deep learning framework for the PdM of synchrotron IPs. The section includes an overview of the IP systems under study, the architecture of the DL framework, the data processing workflow, and the selected models. In particular, the framework aims to achieve reliable abnormal-state detection⁽²⁵⁾ as a precursor to RUL estimation, thereby providing engineers with sufficient lead time for maintenance planning.

2.1 Ion pump systems

The examined ultrahigh-vacuum system comprises several critical components: IPs, IP controllers, ion gauges, and ion gauge controllers. Among these components, the IP controllers are the most prone to malfunction. IP controller malfunction primarily occurs because of damage to the electronic circuitry of the connected channel, which is usually resolved by switching to a backup channel. The working principle of an ion gauge's sensing head is similar to that of a traditional tungsten filament bulb with two filaments. If one filament burns out, the other filament can function in its place, thus allowing the user's experiment to proceed uninterrupted; the burnt-out filament can then be replaced during the system's downtime, which occurs once every six months and lasts for one month each time. The ion gauge controllers are relatively unlikely to fail. Finally, the IPs are the most complex of the aforementioned components and are the subject of this research.

IPs operate by releasing electrons under high voltage. Before reaching the anode, these electrons are acted on by a magnetic field. This magnetic field causes the electrons to move in a helical trajectory, which makes them more likely to collide with gas molecules. After the gas molecules are dissociated into ions through electron collisions, they are adsorbed onto a titanium cathode plate within the vacuum chamber, which reduces the number of gas molecules in this chamber, thereby facilitating evacuation.

Because IPs function as a closed vacuum system, they do not experience problems related to pressure increase caused by power outages or mechanical failures. Moreover, unlike mechanical pumps, they produce no noise and are not subject to vibration problems. IPs are thus able to meet the design requirements for beamlines and are installed in each vacuum section, with different pumping rates selected on the basis of the size of the vacuum chamber. Practical experience indicates that a pressure of 10^{-10} Torr (ultrahigh vacuum) can be achieved when the outgassing and pumping rates reach equilibrium.

To maintain high absorption capacity and extend the lifespan of an IP, a turbo pump is used to evacuate the IP and the connected vacuum chamber before use. These two components are then heated to accelerate the detachment of gas molecules adhered to the vacuum chamber's walls or to the IP, thus enhancing the outgassing rate. The released gas molecules are subsequently expelled from the chamber through the turbo pump. The IP is only activated after the establishment of an ultrahigh-vacuum environment, which reduces the load on the IP, thereby extending its lifespan. Moreover, the heating process results in unstable gas molecules

being expelled from the surface of the titanium plate, which increases the cleanliness of the chamber walls and ensures that high absorption efficiency is maintained.

In this study, we focused on the IPs used in the TPS 44A beamline of the NSRRC [TiTan-# (where # represents different pumping speeds in liters per second), Gamma Vacuum, Shakopee, Minnesota, USA]. Figures 1(a)–1(c) show the appearance,⁽²⁶⁾ internal structure, and controller of these IPs,⁽²⁷⁾ respectively. The ion gauges and ion gauge controllers at the TPS are manufactured by Agilent (Santa Clara, California, USA), and a schematic of an ion gauge is shown in Fig. 1(d).

2.2 DL framework for the IP system

The proposed system is shown in Fig. 2. A deep-learning-based framework was developed to conduct PdM for eight IPs on the TPS 44A beamline of the NSRRC. Sensor signals, including voltage, current, and pressure, were preprocessed (details in Sect. 2.3) and used to train multiple candidate models for PdM (see Sect. 2.4). The model achieving the highest performance was subsequently deployed on the beamline server. The system utilizes real-time sensor streams supplemented by short-term historical records to predict the health status of IPs, thereby enabling engineers to plan maintenance tasks in advance and reduce the risk of unexpected beamline interruptions.

2.3 Data processing workflow

The data were collected, cleaned, resampled, normalized, visualized, and finally labeled. Subsequently, lag features were selected for model construction, and machine status was encoded. Ninety percent and 10% of the collected data were used for model training and testing, respectively. Figure 3 depicts the workflow of the aforementioned procedure.

The monitoring data for the examined IPs are stored in a beamline database. We used Python to extract the following information from this database: voltage (V), current (I), and pressure (P) values determined by the IP controllers, as well as the pressure measured by the ion gauges (G).

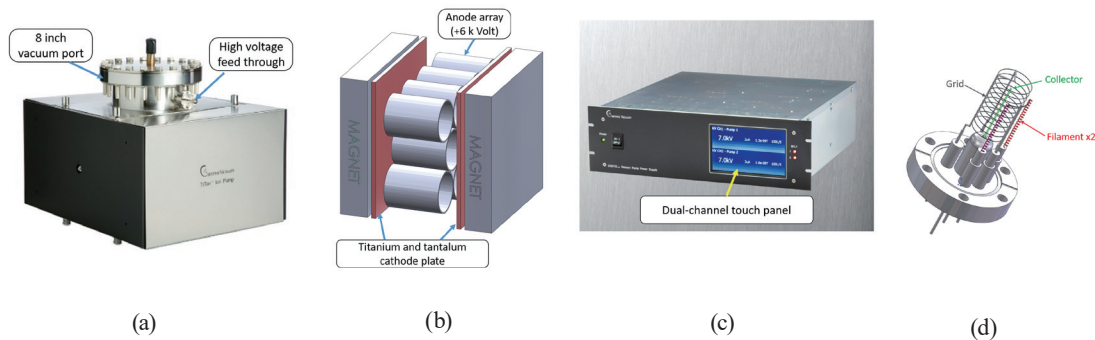


Fig. 1. (Color online) (a) Appearance, (b) internal structure, and (c) controller of an IP used in the TPS 44A beamline of the NSRRC. (d) Schematic of an ion gauge head used in this beamline.

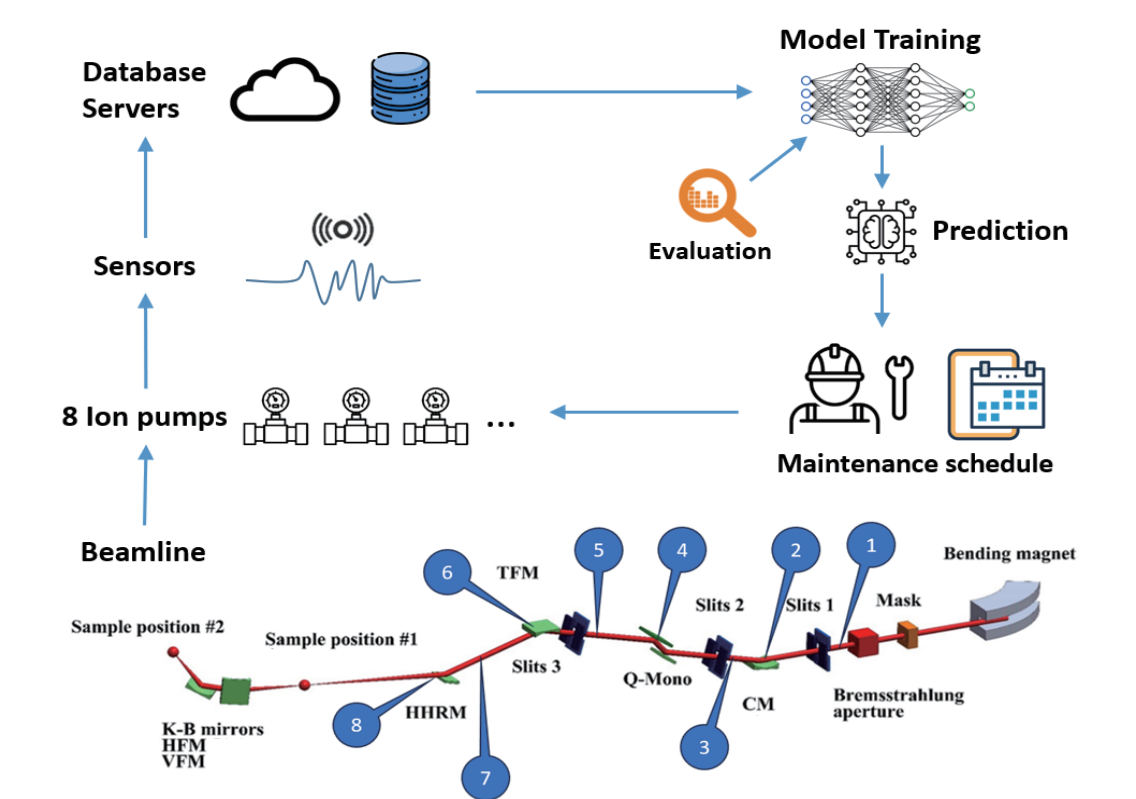


Fig. 2. (Color online) System architecture. Data recorded on eight IPs in the beamline server database were used to train and evaluate multiple models to identify the optimal model. The optimal model was then used to predict the health statuses of the IPs on the basis of current and past data. Maintenance personnel can use the predictions of the optimal model to plan PdM tasks.

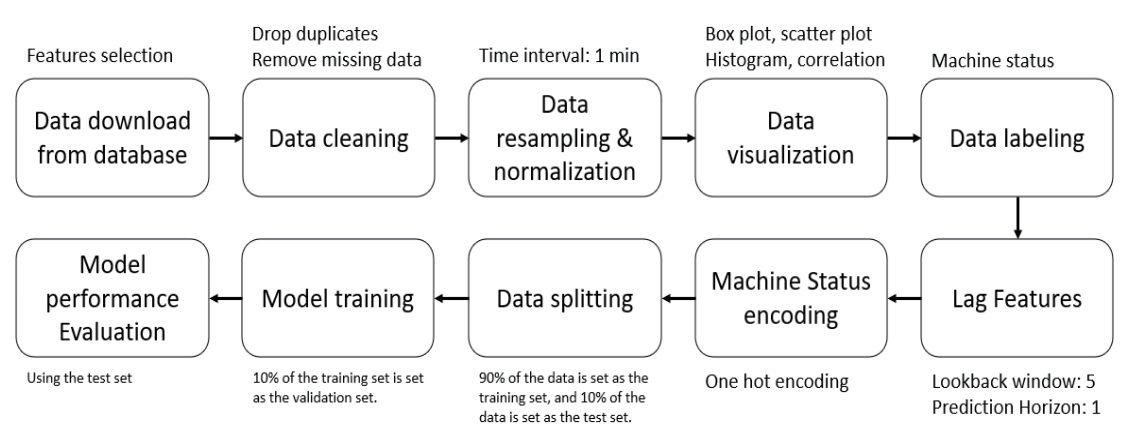


Fig. 3. Data processing workflow.

In early November 2021, IP4 experienced a failure; therefore, data from approximately 1 year before this failure event (from November 1, 2020 to December 21, 2021) were gathered and analyzed. Table 1 presents the following information on the extracted data: data categories, data types, time ranges, original numbers of samples, and final numbers of samples.

The data were cleaned to eliminate duplicate and missing values. The numbers of records obtained for the four considered features (V, I, P, and G) varied because of the adoption of different recording strategies (fixed-time or value-change-triggered sampling). Resampling was conducted to balance the data set for model training. The sampling interval was set as 1 min, and interpolation or extrapolation was performed to fill in any missing data. After this step was conducted, the total number of data entries was 520952, and the cutoff date was set as October 28, 2021 (i.e., before the failure). This date was selected because it was the last date in October for which data were available on all four considered features, namely, V, I, P, and G. The values of the four features varied considerably because they had different units and were thus subject to normalization. In this study, we used the scikit-learn min/max scaling method⁽²⁸⁾ to scale the feature values to a range of 0–1 to increase stability and reduce training time.

We used box plots to visualize the median, quartiles, skewness, and potential outliers of different features and to facilitate comparisons of their data distributions. The correlations among and trends of the features were visually analyzed using scatter plots of feature pairs. These correlations were also quantified through a Pearson correlation analysis.⁽²⁹⁾ Histograms were used to understand the distribution and modality of the data. We conducted these analyses to understand the data on which the model was trained.

The developed framework outputs the health status of an IP. A regression model can be developed to predict the RUL of IP4 on the basis of its failure time and monitoring data from the previous year. However, for the reasons explained in the discussion section, we classified model outputs into two general classes: a normal state and an abnormal state. An abnormal state indicates that a failure is expected to occur within a certain maximum period. The maximum period is determined using the k-means algorithm^(30,31) and other methods (details are provided in the data labeling section). The data are then divided into two clusters to find the time point that separates the two classes. The data collected before and after this time point are labeled 0 and 1, which indicate a normal state and an abnormal state, respectively.

Time-series lag features are created by shifting the time series backward by a certain number of time steps to capture historical values in the series. These features allow models to incorporate

Table 1
Historical data retrieved for IP4.

	Time range	Counts	Mod. counts	Type
V	2020-11-01 00:01:49 2021-12-21 23:54:15	61276	520952	Float 6400.0
I	2020-11-01 00:00:03 2021-12-21 23:59:34	1442129	520952	Float 4.000e−06
P	2020-11-01 00:00:08 2021-12-21 23:58:55	2068406	520952	Float 6.500e−10
G	2020-11-01 00:00:00 2021-12-21 23:59:55	26362795	520952	Float 1.213e−09

past information to predict future values. The developed framework uses feature data for the past 5 min to predict the health status 1 min into the future. To ensure that model outputs do not differ in order or magnitude, the statuses 0 and 1 were converted into the tuples (0, 1) and (1, 0), respectively, through one-hot encoding.

Both the training and test sets contained appropriate proportions of data on healthy and abnormal states. During the training process, 10% of the training data was used as a validation set to check for overfitting. Furthermore, imbalance in the proportion of healthy and abnormal data can cause bias in the model training results. In this study, the ratio between healthy and abnormal data was 1:0.7. Therefore, during model training, different weight values were assigned to the two classes to address the class imbalance.

2.4 Selected models

PdM is defined as the use of historical and real-time machine statuses to forecast potential failure time points. Consequently, model selection must accommodate the characteristics of time-series data. RNNs known for their memory capabilities are commonly used for such tasks; however, they suffer from vanishing and exploding gradient issues. To address these limitations, we adopted two improved variants: long short-term memory (LSTM) and gated recurrent unit (GRU). Additionally, the independently recurrent neural network (IndRNN) is considered owing to its superior ability to stack deeper layers, handle longer sequences, offer greater interpretability, and achieve a higher computational efficiency than LSTM and GRU.

While 2D convolutional neural networks (CNNs) are widely applied to image data, the analogous concept of 1D CNNs can be effectively utilized for time-series analysis. Thus, a 1D CNN is included as the fourth model in this study. In addition to the four DL models, regression-based approaches (e.g., linear regression and ARIMA) were initially considered as baselines. However, these methods rely on explicit degradation trajectories and sufficient failure samples, which are limited in this study. Moreover, regression models are less effective in handling multivariate sensor data and temporal dependencies. For these reasons, regression-based approaches were excluded. The selected models aim to strike a balance among model complexity, computational efficiency, and predictive performance. For practical industrial applications that require real-time monitoring and forecasting, model lightweighting and ease of deployment are critical considerations. To ensure a fair comparison of model performance, consistent hyperparameters are applied across all four architectures. Detailed configurations are described in the following sections.

2.4.1 LSTM and GRU models

An RNN is a neural network with memory,⁽³²⁾ and its hidden state update process can be represented as

$$h_t = \sigma(W_h x_t + U_h h_{t-1} + b_h), \quad (1)$$

where h_t is the hidden state at time step t , x_t is the input at time step t , h_{t-1} is the hidden state from the previous time step, W_h and U_h are weight matrices of the input and recurrent input, respectively, b_h is the bias term, and σ is the activation function. However, an RNN is affected by gradient vanishing and explosion, and is unsuitable for long-term memory.

Compared with an RNN, an LSTM network has additional input, forget, and output gates, which control the flow, retention, and output of information, respectively.⁽³³⁾ An LSTM network contains an independent memory cell that can effectively preserve long-term details, thereby mitigating the problems of vanishing gradients and addressing the poor performance of RNNs in long-term memory. Multilayer LSTM architectures are capable of capturing temporal dependencies at different time scales, with lower layers modeling short-term variations and higher layers accommodating longer-term trends; however, higher model complexity might lead to overfitting. Therefore, in this study, we adopted a single-layer LSTM model design with 64 neurons. The LSTM layer of this model was followed by a dense layer with two neurons representing the two output states. The softmax activation function was used for classification, and the model had 17794 parameters. The loss function selected for training was categorical cross-entropy, which measures the difference between predicted and actual probability distributions.⁽³⁴⁾ Moreover, the Adam optimizer was used to minimize errors through backpropagation.⁽³⁵⁾ This commonly used optimizer adopts the momentum method⁽³⁶⁾ to maintain an exponentially decaying average of past gradients, and it employs the AdaGrad method⁽³⁷⁾ to adjust the learning rate according to the square of past gradients.

Compared with LSTM, GRU simplifies the input and forget gates into a single update gate, which controls the recording and forgetting of information. This simplification can accelerate execution and reduce memory usage.⁽³⁸⁾ The GRU model used in this study also contained a single layer with 64 neurons, and the model had 13570 parameters.

2.4.2 IndRNN model

Although LSTM and GRU use the hyperbolic tangent and sigmoid activation functions as gate functions, they may still encounter gradient vanishing problems. IndRNN⁽³⁹⁾ can avoid this problem because it features neurons that are independent of each other within the same layer, which makes the behavior of each neuron more interpretable. By regulating recurrent weights, IndRNN can learn longer-term dependencies to process longer time-series data more effectively. The hidden state at time t for IndRNN can be described as

$$h_t = \sigma(Wx_t + u \odot h_{t-1} + b) \quad (2)$$

where x_t is the input, u is the recurrent weight vector, $\odot$ is the Hadamard product, W is the weight for the current input, b is the bias, and σ is the activation function. In contrast to traditional RNN models, IndRNN can robustly apply the rectified linear unit (ReLU) activation function, enabling the construction of deeper networks. IndRNN has a simplified hidden state update process and does not require complex weight matrices to capture dependencies between neurons; thus, the single-layer IndRNN model used in this study had only 450 parameters.

2.4.3 1D CNN models

1D CNN models apply 1D convolution to time signals, which makes these models suitable for analyzing time-series data. The operation of 1D CNN models can be expressed as⁽⁴⁰⁾

$$y_k^l = \sigma \left(b_k^l + \sum_{i=1}^{N_{l-1}} \text{conv1D} \left(w_{ik}^{l-1}, x_i^{l-1} \right) \right), \quad (3)$$

where y_k^l is the output of the k th neuron at layer l , σ is the nonlinear activation function, such as the ReLU activation function, x_i^{l-1} is the output of the i th neuron at layer $l-1$, w_{ik}^{l-1} is the filter from the i th neuron at layer $l-1$ to the k th neuron at layer l , and conv1D is the 1D convolution operation. A 1D CNN model contains a pooling layer that reduces the spatial dimensions of feature maps through data flattening while retaining crucial feature information. The output of this model is produced by a dense layer with two neurons. The 1D CNN model used in this study had a parameter count of 962. Table 2 shows the architectures and parameter settings of the four models used in this study.

In summary, we aimed to develop and implement a DL framework for the real-time health prediction of IPs. The proposed approach begins with comprehensive data preprocessing, followed by the evaluation and comparison of multiple DL models. The performance and generalization capability of the selected model are then demonstrated, validating the practicality and applicability of the framework within the operational environment of synchrotron radiation beamlines.

3. Results

The results include data visualization and statistical analysis results, model training and evaluation outcomes, and comparative performance across different architectures. Particular attention is given to abnormal-state detection accuracy, false positive and false negative rates, and the robustness of the framework when applied to multiple IPs. These findings demonstrate the feasibility of real-time anomaly detection.

Table 2
Architectures of the four adopted models.

	LSTM	GRU	IndRNN	1D CNN
Input layer (neurons)	LSTM (64)	GRU (64)	IndRNN (64)	Conv1D (64)
Other layers	—	—	—	global max pooling1d
Output layer (neurons)	Dense (2)			
Parameters	17794	13570	450	962
Loss function	categorical cross entropy			
Optimizer	Adam			

3.1 Visualization and analysis

Figure 4 shows the normalized box plots of the four considered parameters (V, I, P, and G) for IP4. The minimum and maximum values in these plots are 0 and 1, respectively. The normalization process was conducted for each IP examined in this study, with IP4 exhibiting the largest data dispersion because of the failure it experienced.

To understand the interactions among the features of IP4, we created scatter plots for qualitative observations and calculated Pearson correlation coefficients for quantitative analysis (Fig. 5). The following insights were obtained through the aforementioned analyses:

- 1) Among the four examined features, the strongest correlation (coefficient: 0.84) existed between P and I, followed by between G and P (coefficient: 0.77) and between G and I (coefficient: 0.75).
- 2) The feature V had weak correlations with the other three features, being nearly constant over time.
- 3) The machine state had the strongest correlation with G (coefficient: 0.72), followed by P and I. The data for individual features were primarily concentrated between 0 and 0.4.

These three insights are elaborated in the discussion section.

3.2 Data labeling

The labeling strategy adopted in this study aims to classify the operational status of IPs as either normal or abnormal, thereby supporting the objectives of PdM. The primary goal is to define clear state boundaries for training binary classification models, rather than merely identifying structural transition points within time-series data. By leveraging historical operational records, the proposed approach enables the model to learn characteristic “anomalous patterns” that typically precede failure events. This facilitates the early detection of potential risks and enhances the system’s ability to anticipate equipment degradation before critical faults occur.

We used the *k*-means algorithm to cluster the data on the four examined features into two groups (blue points in Fig. 6), one representing a healthy state and the other representing an abnormal state (0 and 1, respectively). The machine was more likely to be in an abnormal state (1) than a normal state (0) at time points closer to the time of failure. The orange solid curve in Fig. 6 was obtained on the basis of the cumulative clustering results (blue points). This curve resembles two intersecting straight lines. Starting from the first time point, subsequent time points were iteratively added, and a regression line was drawn. The fitted line’s mean square error (MSE) was computed. When MSE exceeded the set threshold, the iteration halted, and the turning point was determined. As indicated by the green dashed line in Fig. 6, the final turning point was determined to be June 1, 2021. Therefore, data points before this point were defined as representing a healthy status and labeled as 0. By contrast, the data points after the turning point were defined as representing an abnormal state and labeled as 1. Although the clustering results (blue points) show some interleaving of 0 and 1 labels due to clustering errors and transient fluctuations, the cumulative curve (orange line) was adopted to overcome this. In the healthy

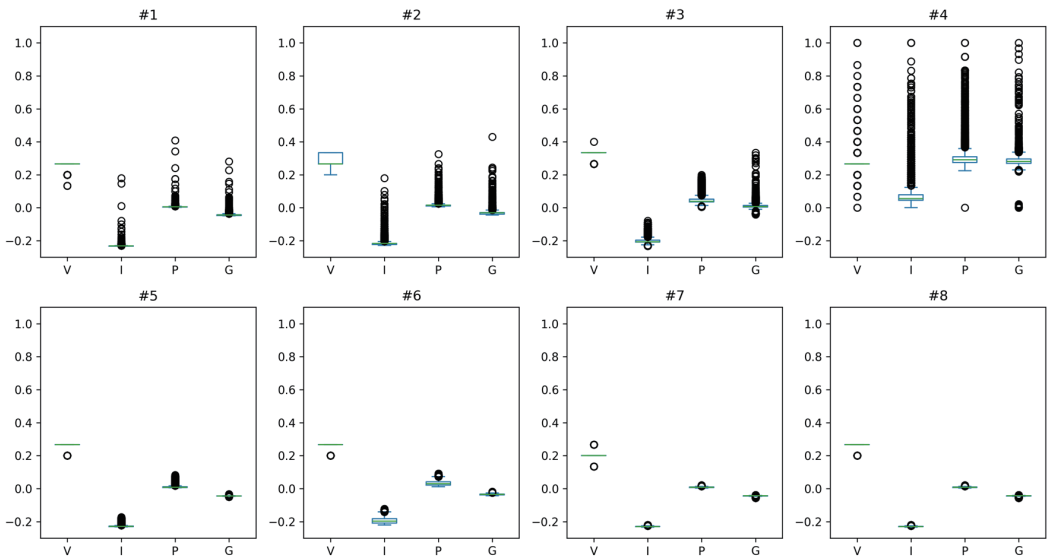


Fig. 4. (Color online) Boxplot of 1-year historical data for eight IPs on the TPS 44A beamline.

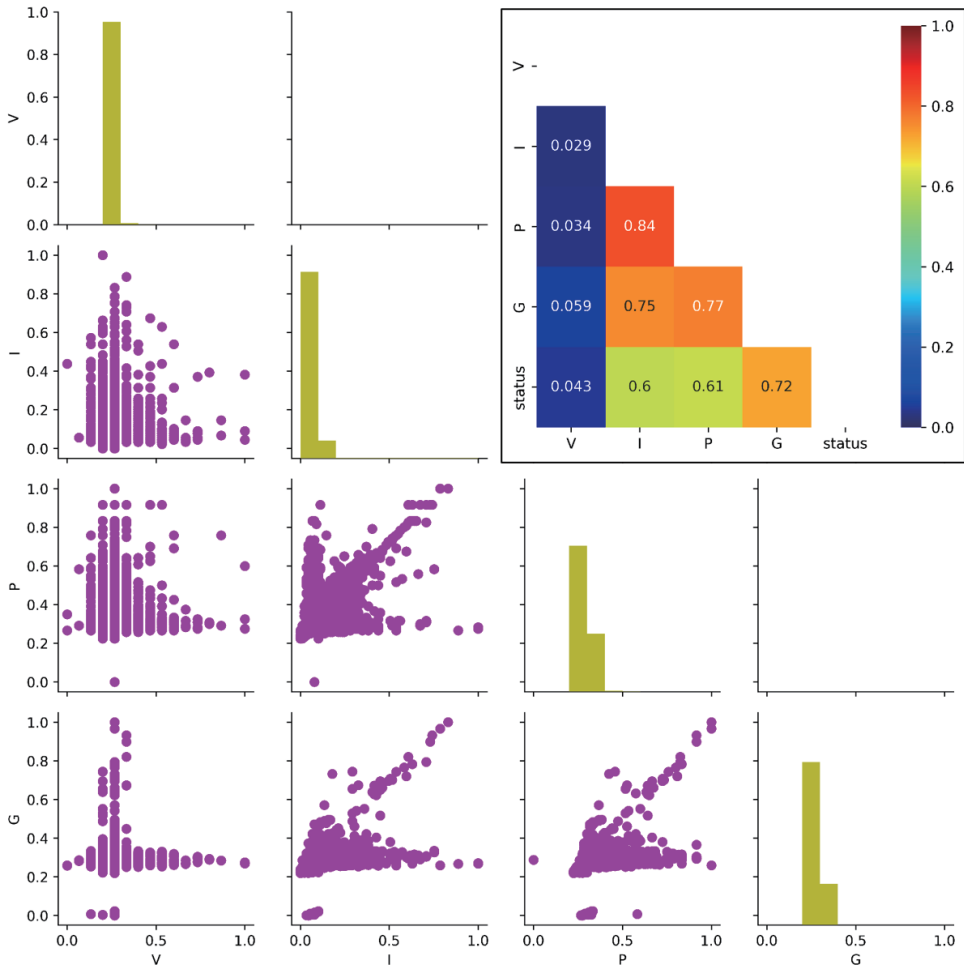


Fig. 5. (Color online) Scatter plots (lower-left triangle, magenta points) and Pearson correlation coefficients (upper-right subplot) of feature interactions for IP4 as well as feature histograms (diagonal, yellow bars) for IP4.

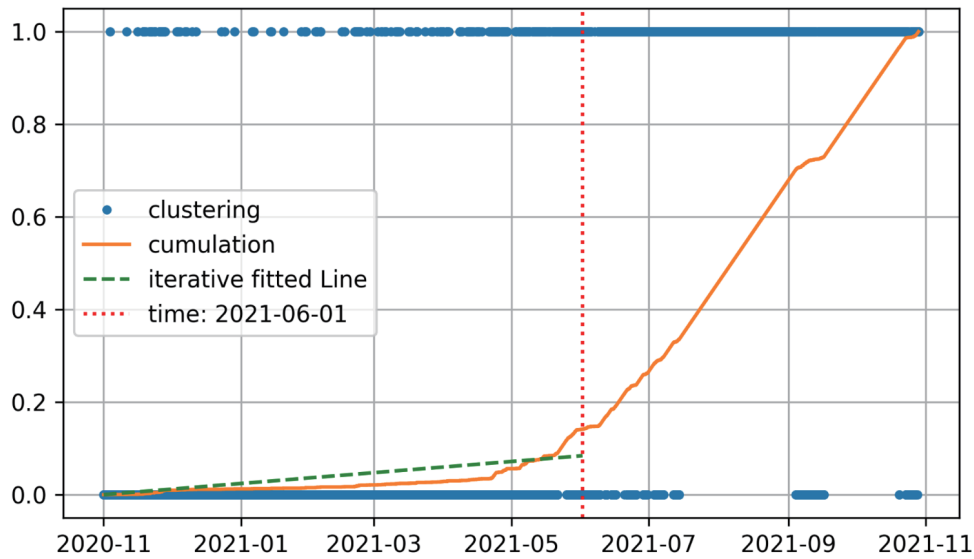


Fig. 6. (Color online) Labeling outcomes. The blue points represent the results of k-means clustering (0 and 1), the orange solid line denotes the results of cumulative clustering, the green dashed line is the iterative fitted line, and the red dotted line indicates the termination point of the iteration, where the fitted line's MSE exceeded the set threshold.

stage, the slope remains low as most points are 0, whereas in the abnormal stage, the slope increases as 1 labels accumulate. The turning point of this cumulative curve (red dotted line) was then defined as the scientific boundary between the healthy and abnormal states. A model output of 0 indicated that the machine was healthy, whereas a model output of 1 indicated that machine failure might occur within five months from June 1, 2021 (i.e., by November 1, 2021). Maintenance schedules can then be arranged for engineers according to the model outputs.

3.3 Model comparison

The data collected for IP4 over approximately 1 year before its failure were used to train the four adopted models. The collected dataset comprised 520952 records, with 81, 9, and 10% of the data used for model training, model validation, and model testing, respectively. The early stopping technique was employed in this study; this technique involves monitoring the validation loss during training to prevent model overfitting. The number of epochs was set to 30 and the batch size was set to 1024. Trained versions of each model were evaluated on the test dataset in terms of accuracy, precision, and recall.

Given the focus on detecting potential failures, the abnormal state was considered the positive class in this study for model evaluation. Therefore, the model predictions were categorized as follows.

- True positives (TP): scenarios in which the predicted and actual states were abnormal.
- True negatives (TN): scenarios in which the predicted and actual states were normal.
- False positives (FP, or false alarms): scenarios in which the predicted state was abnormal but the actual state was normal.

- False negatives (FN, or misses): scenarios in which the predicted state was normal but the actual state was abnormal.

Precision and recall were calculated to evaluate the performance of each model in identifying an abnormal state as

$$\text{Precision} = TP / (TP + FP), \quad (4)$$

$$\text{Recall} = TP / (TP + FN). \quad (5)$$

Furthermore, the false alarm and missing rates of each model in identifying an abnormal state were defined as

$$\text{False alarm rate} = 1 - \text{precision} = FP / (TP + FP), \quad (6)$$

$$\text{Missing rate} = 1 - \text{recall} = FN / (TP + FN). \quad (7)$$

The missing rate represents the proportion of actual anomalies that the model failed to detect. Table 3 presents the accuracy, precision, recall, false alarm rate, and missing rate values of the four adopted models.

Among the four models, IndRNN had the highest accuracy, but LSTM had the lowest missing rate. The primary focus of this study was the performance of the four models in detecting actual abnormal states, which was represented by their missing rates; thus, the LSTM model had the best performance in this study.

The selected model achieved a false alarm rate of 19% and a miss detection rate of 5%. Furthermore, by implementing a status confirmation mechanism—requiring the model to produce consistent predictions across three consecutive time steps (e.g., three successive normal or abnormal states)—the false alarm and miss detection rates can be further reduced to 0.69 and 0.01%, respectively. This enhancement significantly improves the model's reliability for real-time anomaly detection.

This model was subsequently used to predict the statuses of the other seven IPs, and the results are shown in Table 4. These results indicate that the LSTM model accurately predicted the health statuses of these IPs, with only one false alarm occurring for IP1 across 520952 data points.

Table 3
Performance characteristics of the four adopted models.

	LSTM	GRU	1D CNN	IndRNN
Accuracy	.89	.89	.89	.90
Precision	.81	.82	.84	.85
Recall	.95	.94	.91	.91
False alarm	.19	.18	.16	.15
Missing rate	.05	.06	.09	.09

Table 4

Results obtained with the LSTM model (optimal) for all IPs except for IP4. In the results obtained from the other seven normal IPs (each with 520,946 data entries), only one instance was misclassified as abnormal.

	Predicted anomaly	Real anomaly	Support samples
1	1	0	520,946
2	0	0	520,943
3	0	0	520,946
5	0	0	520,946
6	0	0	520,939
7	0	0	520,940
8	0	0	520,946

4. Discussion

Several insights can be obtained from the results of this study. First, voltage (V) almost remains constant before a failure. Thus, the exclusion of this variable can simplify the model training process without affecting model performance. Second, among the four considered parameters, the strongest correlation was that between the pressures (P) estimated by the IP controllers and the currents (I) measured by them. This strong correlation is attributable to the IP controllers estimating P on the basis of output current (I) and other factors (V, correction parameters, and pump size). Therefore, I can be removed to reduce the number of parameters, with P being retained. In summary, P and G are indeed the most critical parameters. Nevertheless, our preliminary experiments showed that using only these two inputs yielded a slightly lower accuracy than the four-feature model. To maximize predictive performance, we therefore adopted the four-feature setting. Moreover, regression-based approaches were also explored in the early stage to estimate the RUL, but the outcomes were unsatisfactory, likely owing to the limited availability of failure data and insufficient critical sensor variables. Consequently, a classification framework was employed to ensure reliable abnormal-state detection, while future work will focus on extending the model toward robust RUL prediction once richer datasets become available.

The false alarm and missing rates of the LSTM model for IP4 were 0.19 and 0.05, respectively. The false alarm rate was almost 0 for the other seven healthy IPs; however, the missing rate for these IPs could not be determined because no abnormal data were available for them. In this study, the ratio of normal to abnormal training data was approximately 1:0.7, and appropriate weights were used for the abnormal class to avoid data imbalance problems and enhance model performance in detecting the abnormal class. In the future, the performance of the LSTM model can be validated by determining its false alarm and missing rates for other IPs with abnormal states.

The eight IPs examined in this study were manufactured by the same company and have similar designs; however, they have different pumping capacities. Specifically, IP4 has a pumping capacity of 400 L/s; IP1, IP3, IP5, IP7, and IP8 have a pumping capacity of 300 L/s; and IP2 and IP6 have a pumping capacity of 500 L/s. Although the LSTM model was trained using

data for IP4, it accurately predicted the statuses of the other IPs, demonstrating sufficient generalizability.

Although advanced PdM approaches, such as Transformer-based models, have demonstrated promising results, they typically require large-scale datasets, richer sensor inputs, and substantial computational resources. Given the limited sensor parameters available in the early stage of this study, as well as the real-time constraints of practical deployment, in this research, we focused on developing a foundational framework that operates efficiently under existing resource conditions. Therefore, we emphasize a comparative evaluation of mainstream DL architectures under realistic constraints. Among the tested models, the LSTM-based architecture achieved exceptional performance in anomaly detection. Through iterative refinement, the model significantly reduced both false positives and false negatives, demonstrating its effectiveness and reliability in the target application context.

Finally, we developed a simple model architecture (only a single-layer structure), which may not clearly reflect the advantages of IndRNN. A more complex IndRNN model may overcome the drawbacks of LSTM and GRU models, thus outperforming these models.

5. Conclusions

In Industry 4.0, IoT-collected production data are analyzed using AI to enhance production efficiency and flexibility. PdM can increase production stability and efficiency by facilitating the prediction of equipment failures in advance. Implementing PdM in beamlines can reduce operational and maintenance costs, improve production efficiency, and reduce downtime caused by equipment failures.

We collected 1-year monitoring data on four parameters [the voltage (V), current (I), and pressure values determined by the IP controllers (P) as well as the pressure values measured by the ion gauges (G)] for eight IPs on the TPS 44A beamline of the NSRRC. The collected data were preprocessed by cleaning, resampling, normalization, and labeling. Subsequently, lag features were selected, status encoding was performed, and the data were split. LSTM, GRU, IndRNN, and 1D CNN models were trained to predict the state of the examined IP. For fairness in comparison, a simple single-layer structure was adopted for all the models, with the number of neurons set as 64. The results obtained with these models for IP4 indicate that the models had similar prediction performance characteristics, with the IndRNN and LSTM models having the highest accuracy (90%) and the lowest missing rate (5%), respectively. A minimum miss rate for an abnormal machine state is preferred in failure prediction; therefore, the LSTM model was selected for subsequent application. This model was trained using the data collected for IP4 and then used to predict the health statuses of the other seven IPs. The LSTM model accurately predicted the health statuses of these IPs across more than 520000 data points over 1 year, with only one false alarm observed. This model had high prediction reliability because it conducted weight adjustments to prevent problems resulting from an imbalance between normal and abnormal data.

After training, the developed framework can identify abnormal signals in monitoring data five months before machine failure. However, further improvements can be made to the

framework to enhance its performance. First, data on additional parameters can be input into the framework to improve its prediction capabilities. Second, an alert notification function can be incorporated into the framework to enable the scheduling of timely equipment maintenance. Finally, the framework can be periodically retrained using new monitoring data to maintain or enhance its performance. In summary, the proposed generalized classification framework for PdM can improve the operational performance of the TPS and other synchrotron research facilities, in addition to contributing to efforts at realizing Industry 4.0.

Acknowledgments

This study was supported by the National Science and Technology Council under grant numbers NSTC 113-2222-E-197-002 and NSTC 113-2112-M-029-007.

References

- 1 C. Bai, P. Dallasega, G. Orzes, and J. Sarkis: *Int. J. Prod. Econ.* **229** (2020) 107776. <https://doi.org/10.1016/j.ijpe.2020.107776>
- 2 R. Sahal, J. G. Breslin, and M. I. Ali: *J. Manuf. Syst.* **54** (2020) 138. <https://doi.org/10.1016/j.jmsy.2019.11.004>
- 3 H. S. Kang, J. Y. Lee, S. Choi, H. Kim, J. H. Park, J. Y. Son, B. H. Kim, and S. D. Noh: *Int. J. Precision Eng. Manuf.-Green Technol.* **3** (2016) 111. <https://doi.org/10.1007/s40684-016-0015-5>
- 4 Y. Liao, F. Deschamps, E. d. F. R. Loures, and L. F. P. Ramos: *Int. J. Prod. Res.* **55** (2017) 3609. <https://doi.org/10.1080/00207543.2017.1308576>
- 5 G. Medina-Oliva, A. Voisin, M. Monnin, and J.-B. Leger: *Knowledge-Based Syst.* **68** (2014) 40. <https://doi.org/10.1016/j.knosys.2013.12.020>
- 6 Y. Li and Z. Mao: *Sens. Mater.* **38** (2026) 3611. <https://doi.org/10.18494/SAM6268>
- 7 E. Sezer, D. Romero, F. Guede, M. Macchi, and C. Emmanouilidis: 2018 IEEE Int. Conf. Engineering, Technology and Innovation (ICE/ITMC) (2018) 1. <https://doi.org/10.1109/ICE.2018.8436307>
- 8 J. Lee, E. Lapira, B. Bagheri, and H.-A. Kao: *Manuf. Lett.* **1** (2013) 38. <https://doi.org/10.1016/j.mfglet.2013.09.005>
- 9 P. Nunes, J. Santos, and E. Rocha: *CIRP J. Manuf. Sci. Technol.* **40** (2023) 53. <https://doi.org/10.1016/j.cirpj.2022.11.004>
- 10 J. Lee, H.-A. Kao, and S. Yang: *Procedia CIRP* **16** (2014) 3. <https://doi.org/10.1016/j.procir.2014.02.001>
- 11 M.-Y. You, F. Liu, W. Wang, and G. Meng: *IEEE Trans. Reliab.* **59** (2010) 744. <https://doi.org/10.1109/TR.2010.2085572>
- 12 Y. Lei, N. Li, S. Gontarz, J. Lin, S. Radkowski, and J. Dybala: *IEEE Trans. Reliab.* **65** (2016) 1314. <https://doi.org/10.1109/TR.2016.2570568>
- 13 H. Qiu, J. Lee, J. Lin, and G. Yu: *Adv. Eng. Inf.* **17** (2003) 127. <https://doi.org/10.1016/j.aei.2004.08.001>
- 14 J. Hu and P. Chen: *Reliab. Eng. Syst. Safety* **196** (2020) 106707. <https://doi.org/10.1016/j.ress.2019.106707>
- 15 M. C. Garcia, M. A. Sanz-Bobi, and J. Del Pico: *Comput. Ind.* **57** (2006) 552. <https://doi.org/10.1016/j.compind.2006.02.011>
- 16 L. Liao, W. Jin, and R. Pavel: *IEEE Trans. Ind. Electron.* **63** (2016) 7076. <https://doi.org/10.1109/TIE.2016.2586442>
- 17 Z. Li, D. Wu, C. Hu, and J. Terpeny: *Reliability Eng. Syst. Safety* **184** (2019) 110. <https://doi.org/10.1016/j.ress.2017.12.016>
- 18 J. C. Cheng, W. Chen, K. Chen, and Q. Wang: *Autom. Construct.* **112** (2020) 103087. <https://doi.org/10.1016/j.autcon.2020.103087>
- 19 J.-R. Ruiz-Sarmiento, J. Monroy, F.-A. Moreno, C. Galindo, J.-M. Bonelo, and J. Gonzalez-Jimenez: *Eng. Appl. Artif. Intell.* **87** (2020) 103289. <https://doi.org/10.1016/j.engappai.2019.103289>
- 20 J. A. Roberts and U. Ahmed: 2025 7th Int. Conf. Electron. Eng. Informatics (EEI) (2025) 543. <https://doi.org/10.1109/EEI67650.2025.11334875>
- 21 C.-H. Hsia, L.-Y. Ke, S.-T. Chen, and Y.-C. He: *IEEE Access* **14** (2026) 8107. <https://doi.org/10.1109/ACCESS.2026.3652826>

- 22 C.-H. Hsia and C.-H. Liu: IEEE Sens. J. **22** (2022) 13612. <https://doi.org/10.1109/JSEN.2022.3177472>
- 23 C.-H. Hsia: J. Imaging Sci. Technol. **62** (2018) 030402. <https://doi.org/10.2352/J.ImagingSci.Technol.2018.62.3.030402>
- 24 C.-H. Hsia: IEEE Sens. J. **15** (2015) 994. <https://doi.org/10.1109/JSEN.2014.2359225>
- 25 C.-C. Chen, Y.-S. Lin, H.-Y. Chen, W.-Y. Sun, and C.-J. Kuo: Sens. Mater. **37** (2025) 5501. <https://doi.org/10.18494/SAM5746>
- 26 Products: Large Titan Ion Pumps <https://www.gammavacuum.com/detail/index/sArticle/3385/number/300LCV8SSCNN/sCategory/17668> (Jan. 4, 2025).
- 27 Products: Digitel Controllers <https://www.gammavacuum.com/products/digitel-controllers/3335/digitel-mpcq?number=MPCQ1PS1U1SSNI&c=17669> (Jan. 4, 2025).
- 28 L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, and J. Grobler: arXiv preprint arXiv:1309.0238 (2013). <https://doi.org/10.48550/arXiv.1309.0238>
- 29 J. Benesty, J. Chen, Y. Huang, and I. Cohen: Noise Reduction in Speech Processing (Springer, Berlin, 2009) pp. 1–4. https://doi.org/10.1007/978-3-642-00296-0_5
- 30 M. Ahmed, R. Seraj, and S. M. S. Islam: Electronics **9** (2020) 1295. <https://doi.org/10.3390/electronics9081295>
- 31 A. Likas, N. Vlassis, and J. J. Verbeek: Pattern Recognit. **36** (2003) 451. [https://doi.org/10.1016/S0031-3203\(02\)00060-2](https://doi.org/10.1016/S0031-3203(02)00060-2)
- 32 L. R. Medsker and L. C. Jain: Recurrent Neural Networks: Design and Applications (CRC Press, Boca Raton, 1999). <https://doi.org/10.1201/9781420049176>
- 33 A. Graves: Supervised Sequence Labelling with Recurrent Neural Networks (Springer, Berlin, 2012) pp. 37–45. https://doi.org/10.1007/978-3-642-24797-2_4
- 34 Z. Zhang and M. Sabuncu: Adv. Neural Inf. Process. Syst. **31** (2018) 8778.
- 35 D. P. Kingma and J. Ba: arXiv preprint arXiv:1412.6980 (2014). <https://doi.org/10.48550/arXiv.1412.6980>
- 36 V. V. Phansalkar and P. S. Sastry: IEEE Trans. Neural Networks **5** (1994) 505. <https://doi.org/10.1109/72.286925>
- 37 R. Ward, X. Wu, and L. Bottou: J. Mach. Learn. Res. **21** (2020) 1.
- 38 R. Dey and F. M. Salem: 2017 IEEE 60th Int. Midwest Symp. Circuits and Systems (MWSCAS) (2017) 1597. <https://doi.org/10.1109/MWSCAS.2017.8053243>
- 39 S. Li, W. Li, C. Cook, C. Zhu, and Y. Gao: Proceedings of the IEEE Conf. Computer Vision and Pattern Recognition (2018) 5457. <https://doi.org/10.1109/CVPR.2018.00572>
- 40 S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman: Mech. Syst. Signal Process. **151** (2021) 107398. <https://doi.org/10.1016/j.ymssp.2020.107398>

About the Authors



Cheng-Ying Chung received his B.S. and M.S. degrees in mechanical engineering from National Central University, Taiwan, in 2003 and 2005, respectively, and his Ph.D. degree in electrical engineering from National Cheng Kung University, Taiwan, in 2015. From 2005 to 2014, he was an engineer at the Metal Industries Research and Development Centre, Taiwan. He then served as a researcher at the R&D Center of Nidec Corporation, Taiwan, from 2014 to 2015. From 2016 to 2018, he was a chief engineer at Delta Electronics, and from 2018 to 2021, he was an R&D manager at Lite-On Technology Corporation. From 2021 to 2024, he worked as an assistant engineer at the National Synchrotron Radiation Research Center, Taiwan. Since 2024, he has been an assistant professor with the Department of Mechanical and Electro-Mechanical Engineering, National Ilan University, Taiwan. His research interests include opto-mechatronic system integration, artificial intelligence applications, and signal processing.
(cychung@niu.edu.tw)



Yu-Shan Huang received his B.S. and Ph.D. degrees in physics from National Tsing Hua University, Taiwan, in 1990 and 1999, respectively. He was a postdoctoral researcher in the Department of Physics at National Tsing Hua University from 1999 to 2000 and a visiting scholar at the Stanford Synchrotron Radiation Laboratory, USA, from 2000 to 2001. Since 2000, he has been with the National Synchrotron Radiation Research Center (NSRRC), Taiwan, where he served as an assistant research fellow until 2007 and has been a research fellow since 2008. His research interests include X-ray scattering and diffraction physics, biophysics, and X-ray optics.

(jade@nsrrc.org.tw)



Longlife Lee received his master's degree in mechanical engineering from National Chiao Tung University, Taiwan, in 2016. Since 2004, he has been serving as a mechanical engineer at the National Synchrotron Radiation Research Center (NSRRC). His work involves the design and construction of beamline mechanical systems and experimental station mechanisms, with expertise in vacuum application technology. (longlife@nsrrc.org.tw)