IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 1 June 2023, accepted 15 June 2023, date of publication 20 June 2023, date of current version 12 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3287879

RESEARCH ARTICLE

Automatic Optimization of One-Dimensional CNN Architecture for Fault Diagnosis of a Hydraulic Piston Pump Using Genetic Algorithm

OYBEK ERALIEV MARIPJON UGLI[1], KWANG-HEE LEE[2], AND CHUL-HEE LEE[2], (Member, IEEE)

[1]Department of Future Vehicle Engineering, Inha University, Incheon 22212, South Korea
[2]Department of Mechanical Engineering, Inha University, Incheon 22212, South Korea

Corresponding author: Chul-Hee Lee (chulhee@inha.ac.kr)

This paper was supported by Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government [Ministry of Trade, Industry and Energy (South Korea) (MOTIE)] (20014615).

ABSTRACT A hydraulic piston pump is an essential component of a hydraulic transmission system and is extensively used in contemporary industrial settings. Therefore, fault diagnosis of piston pumps is a crucial topic in the engineering field. The convolutional neural network (CNN) is currently the most popular deep neural network model and has been successfully employed for fault detection and other tasks. The design and hyperparameter settings of CNNs significantly affect the overall diagnosis performance. In this study, a genetic method is proposed that can quickly investigate a specific set of potentially viable one-dimensional CNN (1D-CNN) architectures while also optimizing their hyperparameters for a fault detection task of an axial hydraulic piston pump. The proposed model is automatically designed based on a direct connect 1D-CNN block, which is another contribution of this study. The proposed method is evaluated on the raw sound signal dataset of an axial hydraulic piston pump without any signal pre-processing techniques. The experimental results demonstrate that the proposed method outperforms several well-known deep learning (DL) models in terms of fault diagnosis performance. Additionally, the suggested method uses significantly less computational power to determine the best 1D-CNN structures than most peer rivals.

INDEX TERMS Fault diagnosis, hydraulic piston pump, convolutional neural network, genetic algorithm, hyperparameter optimization.

I. INTRODUCTION

A hydraulic piston pump is considered one of the most important components of a hydraulic transmission system and it has been using in several engineering fields such as mechanical engineering, aerospace engineering, ship industries and heavy construction machinery industries [1], [2]. The pump's malfunction might cause downtime or perhaps paralyze the entire manufacturing line. From the standpoint of one's own safety, it could result in terrible mishaps. Hence, the fault detection of a hydraulic pump can play a key role as an avoidance tool in worker safety and secure manufacturing, and it has been stood at the center of the scholarly interest [3], [4], [5]. For example, complexity and hiddenness are considered two main characteristics for the failure can be seen in real world applications. In order to maintain the operation of the entire hydraulic system, effective state monitoring and accurate fault detection of a hydraulic piston pump are main aspects which cannot be ignored.

Traditional fault diagnosis of hydraulic piston pumps typically involves a combination of manual inspection, expert knowledge, and the use of sensor-based measurements. This approach aims to identify and analyze potential faults or abnormalities in the pump system to ensure its reliable operation [6]. One commonly used method in traditional fault diagnosis is visual inspection, where operators visually

The associate editor coordinating the review of this manuscript and approving it for publication was Guillermo Valencia-Palomo.

examine the pump components for any signs of wear, damage, or leakage. This approach relies on the experience and expertise of the inspector to identify potential faults based on visual cues. Additionally, sensor-based measurements play a crucial role in fault diagnosis. Various sensors, such as vibration sensors, pressure sensors, temperature sensors, and acoustic sensors, are installed on the pump system to monitor its performance. These sensors capture data related to operating conditions, performance parameters, and potential fault signatures. The collected sensor data is then analyzed using traditional signal processing techniques [7]. This may involve time domain analysis, frequency domain analysis, or statistical analysis to extract relevant features that indicate potential faults. Feature extraction methods, such as Fourier analysis, wavelet analysis, or statistical moments, are commonly employed to identify specific fault patterns or anomalies in the sensor data. Once the features are extracted, they are compared against pre-defined thresholds or reference values to determine the presence of faults [5]. Rule-based algorithms or expert systems may be used to interpret the extracted features and make diagnostic decisions. These algorithms are typically built based on expert knowledge and experience, incorporating a set of rules or decision criteria to identify specific fault conditions. Overall, traditional fault diagnosis of hydraulic piston pumps relies on manual inspection, sensor-based measurements, and signal processing techniques to detect and diagnose faults. While effective to some extent, this approach often requires a high level of expertise, is time-consuming, and may not be capable of detecting subtle or complex faults. As a result, there is a growing interest in integrating advanced techniques such as machine learning and deep learning to enhance the accuracy and efficiency of fault diagnosis in hydraulic systems.

Machine learning (ML) algorithms have recently become frequently used in engineering due to the expansion of data from mechanical systems and the advancement of artificial intelligence [8], [9]. However, for feature extraction, ML algorithms heavily rely on the expertise and experience of an engineer. As a result, they are inappropriate for scenarios with substantially nonrepresentational characteristics. Meanwhile, deep learning (DL) models rely less on past information and have stronger representational capabilities than ML models. Therefore, DL-based fault diagnosis systems that have the advantage of automatic feature extraction have been employed for condition monitoring of machineries [10], [11]. There are several DL models that have been proposed for fault diagnosis. For example, a unique autoencoder has been suggested to model both local and global geometries of the input by developing various cost functions because current DL approaches ignore the geometry of input samples [12]. A stacked sparse autoencoder (SSAE) is suggested for limited data samples [13]. A deep neural network (DNN) is successfully applied to a fault diagnosis of a wind turbine and produced promising results [14]. A bearing fault diagnosis system based on a generated adversarial network (GAN) is proposed and the model has outstanding diagnostic performance and can address the issue of zero-shot in novel conditions [15]. Among the DL models, the most popular deep neural network model at the moment is the convolutional neural network (CNN), which has been utilized successfully for fault detection and other tasks. Several optimization techniques are applied for hyperparameter tuning [16], [17]. For instance, Tang Sh. et. al. has proposed an adaptive CNN model for fault detection of an axial hydraulic piston pump using acoustic signal [18]. They have used a continuous wave transform (CWT) signal processing technique for converting row acoustic signals to time-frequency domain images, and Bayesian optimization technique is utilized for hyperparameter optimization. Although the study has shown promising results, signal processing technique requires adequate knowledge and the skills of an expert. These authors have also improved CNN model by the use of adaptable learning rate [19]. They have used signal data from vibration sensor, pressure sensor and acoustic sensor and this method can be financially expensive for fault diagnosis. Furthermore, a vibration signal-based fault diagnosis is commonly investigated by researchers [20], [21], [22], [23], [24]. Although the vibration-based method has reliable diagnosis performance, the installation of vibration sensor on applications might be challenging in real world conditions. To address this issue, acoustic sensors can be used. However, acoustic sensors also have disadvantages. For instance, acoustic sensors are very sensitive to environmental noise, and it might cause a decrease of overall diagnosis performance.

CNN models can extract enriched hierarchical features from the input data by increasing the number of layers of it [25]. These features are essential for completing the tasks that have been set as targets. Although the depth of architecture plays a big role, there are a finite number of layers that can be added. The fundamental reason is that training entire models thoroughly using back-propagation techniques is exceedingly challenging for deep architectures because of issues with gradient information flow [26]. Additionally, because there are so many trainable parameters, deeper networks are more susceptible to overfitting issues. In the research disciplines of image recognition and speech recognition, various studies have been undertaken to improve the information flow inside deep neural network designs by adding additional short connections between layers [27], [28], [29]. Such factors must be taken into account when developing hydraulic piston pump system DL-based defect diagnostic algorithms. By taking these concerns into account, richer and appropriate characteristics for high diagnosis performance can be obtained by effective deep neural network model training.

To tackle aforementioned issues, a novel fault diagnosis system is proposed in this research. This system evolves DL model architecture which is based on 1D-CNN blocks and hyperparameter tuning of the model using genetic algorithm (GA). The CNN block is constructed based on the direct

connection-based CNN. In the CNN design, the gradient information flow can be maximized and effectively train deep networks by directly connecting the feature maps of the various layers. Simultaneously, dimension reductions in both the width and height as well as the depth-wise direction are intended to address the issues that can arise from the increased number of parameters as a result of direct connections. Additionally, the suggested method's performance is checked by visualization of the results of the learnt features using t-distributed stochastic neighbor embedding (t-SNE) method. The suggested method also exhibits consistent and reliable diagnosis performance in the presence of noisy data, which is a problem that is commonly observed in real world environments.

The following is a summary of this paper's significant contributions:

1) A novel automatically optimization of 1D-CNN architecture for fault diagnosis of a hydraulic piston pump is proposed. The network architecture and hyperparameters are optimized simultaneously by genetic algorithm (GA). Therefore, the suggested algorithm does not even require users to have a working background of GAs, CNNs, or the examined topic.
2) A direct connection-based CNN block is built. It can improve the gradient information flow within the block's layers. Consequently, enhanced hierarchical features of the input data can be extracted, and the CNN architectures can be trained effectively. To address the potential issues brought on by the increasing number of parameters because of direct connections, dimension redaction module has been built in the CNN block.
3) The proposed diagnosis system uses row acoustic signal. Therefore, it does not require any signal processing techniques, feature extraction methods, and an experienced worker neither on signal processing nor CNN architectures.
4) Experimental findings show that the suggested strategy can perform more effectively than a number of other standard ways. The effectiveness of the suggested strategy is thoroughly examined utilizing various analyses.

The rest of the research is constructed as follows. The fundamental theory of 1D-CNN and GA is described in Section II. Section III talks about the proposed fault diagnosis system, which is based on automatically optimization 1D-CNN architecture using GA. Experimental setup and data acquisition progress is described in Section IV. Section V talks about analysis and visualization of the results. Section VI concludes the paper.

II. BRIEF THEORY

A. ONE-DIMENSIONAL CONVOLUTIONAL NEURAL NETWORK (1D-CNN)

A Convolutional neural network (CNN) is one of the best DL models and it is frequently used to analyze two-dimensional data, like photos and movies. Due to its local connections, weight sharing, and down-sampling, CNN differs from other DL models [30], [31]. CNN excels in extracting both local and global properties from data. Because there are fewer parameters in CNN than in traditional feed-forward neural networks, training is simpler. Generally, CNN architectures consist of three layers including input, hidden and output and the hidden layer also includes several layers, such as a convolution layer, a pooling layer, a fully connected layer. The simple structure of an 1D-CNN is illustrated in Fig.1. A 1D-CNN architecture is particularly suited for processing sequential data, such as time series signals. The input data, in this case, is a one-dimensional signal, such as acoustic data collected from a hydraulic piston pump. As the signal flows through the layers of the 1D-CNN, several operations take place that result in changes in the data dimensions. The initial layers of the 1D-CNN consist of convolutional layers, which apply filters to extract local features from the input signal. The function of the convolution layers is to extract high-level features from the input row vector. Kernels, also known as filters, are a convolutional layer's parameters, and the input row vector is convolved by each filter throughout the feed-forward operation. After computing the dot product between the filter and the input vector, a 1D activation vector of the filter is produced. The size of the filter can directly affect how many hidden layers are present. These convolutional layers produce feature maps that preserve the spatial information but reduce the signal's temporal dimension.

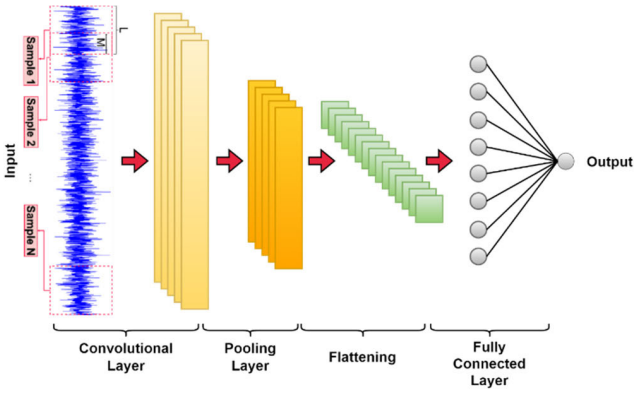

FIGURE 1. Structure of one dimensional convolutional neural network (1D-CNN).

The convolution layer can be expressed as follows:

$$X_v^l = F(\sum_{u \epsilon M_v} X_u^{l-1} \cdot K_v^l + B_v^l), \quad (1)$$

where X denotes the input vector, X_v^l is the updated feature maps produced by the convolution layer, l stands for l^{th} layer of the network, X_u^{l-1} represents u^{th} feature maps produced by the $(l-1)^{th}$ layer, K_v presents a convolutional filter, and B_v^l denotes the bias in the convolution operation.

The network's ability to model nonlinear representations is provided by the activation function. Saturated and non-saturated functions come in two varieties. Tanh is a

saturated function, as is the sigmoid function. The rectified linear unit (ReLU) and its variations are typically favored since they are non-saturated. ReLUs are frequently used in CNNs because they are quick and, thanks to their linear and unsaturated properties, can overcome the gradient vanishing problem [32].

Subsequently, pooling layers are often employed in the architecture. These pooling layers downsample the feature maps by aggregating neighboring values, which further reduces the spatial dimensions of the data. Max pooling is a common type of pooling operation used in 1D-CNNs, where the maximum value within a pooling window is selected as the representative value. And it can help the network to avoid overfitting problem. A mathematical expression of the pooling operation is as follows:

$$a_{v-s}^{l} = f(W_v^l down\left(M_v^{l-1}\right) + B_v^l) \tag{2}$$

where W_v^l represents the weight vector, $down\,(\cdot)$ denotes the pooling operation and M_v presents the feature maps produced by pooling operations.

Following the pooling layers, the feature maps are usually flattened into a one-dimensional vector, which essentially collapses the spatial dimensions into a single dimension. This prepares the data for the fully connected layers of the 1D-CNN. The fully connected layers consist of densely connected neurons that learn complex relationships between the extracted features. These layers may introduce additional dimensionality changes, depending on the specific architecture design and the number of neurons in each layer.

Following that, the classifier receives a 1D vector as the output of the fully connected layer and uses it in conjunction with a softmax logistic regression model to make the final prediction and multi-classification.

B. GENETIC ALGORITHM (GA)

GA is frequently used to solve optimization issues. It is especially helpful when the task at hand has several local optima and/or a large number of factors. In GAs, the set of parameters that the suggested approach aims to solve is referred to as a chromosome. Until the desired chromosomes are formed, GA first generates random chromosomes. There are four key operations that make up the algorithm. Each chromosome's fitness value is calculated first. The selection operator is then used to select strong chromosomes from the population based on how they are. Third, the crossover operator is used to divide the existing chromosomes into new ones. Finally, these chromosomes undergo random mutation to produce new chromosomes. The fitness of these newly created chromosomes is then calculated in the subsequent cycle. Up until the intended outcomes are achieved, the process is repeated.

However, in a deep learning environment, training a GA model might be computationally expensive. The fitness of chromosomes is determined after each iteration. Therefore, getting fit chromosomes in fewer iterations is essential. By enhancing and optimizing GAs, it is achieved. The idea of changing GAs is not new in and of itself. A number of scholars have already suggested certain improvements for specific jobs and applications. In order to achieve better final results, we attempt to optimize common GA operators in this work to fit our framework.

There have been advancements in the following factors:

1) Not each altered chromosome is better than unmutated ones, according to the selection operator. As a result, we choose the best chromosomes from the most recent and previous generations.
2) Crossover operator: rather than using a fixed value, the adaptive crossover probability is employed.
3) Reduced danger of losing beneficial genes due to mutation operators.

1) SELECTION

The selection operator selects the strongest chromosomes and discards the weaker ones after determining the fitness of the existing chromosomes. According to this definition, the likelihood that chromosome ch will be chosen is:

$$P_{select} = \frac{f(ch)}{\sum_1^{N_{ch}} f(ch)} \tag{3}$$

where $f(ch)$ denotes the fitness of the chromosome and N_{ch} denotes the total number of chromosomes in the population. Here, the only chromosomes that were produced during this iteration are eligible to compete. Not all the present chromosomes, nevertheless, are an improvement over the past ones. To put it another way, there is no guarantee that chromosomes created at time t will be more fit than those created at time $(t-1)$. To choose the best chromosome from the population, the following procedures are used:

1) Calculate each chromosome's fitness and retain the strong S chromosomes. Throw away the remaining $N_{ch} - S$ chromosomes, where N_{ch} is the population's total number of chromosomes.
2) If the present chromosomes are more fit than the previous ones, keep them all. Otherwise, preserve the most powerful chromosomes from the most recent generation and swap out comparatively weak chromosomes for those from the most recent iteration.
3) To continue evolution, apply mutation operators to the remainder chromosomes.

By doing this, we can stop crossover and mutation operators from wiping out the strongest individuals.

2) CROSSOVER

By changing segments of the matched father chromosomes, a crossover operator creates two new individuals. A one-point crossover is used in this work.

$$\begin{cases} ch_1^u(t) = r \cdot ch_2^u(t-1) + (1-r) \cdot ch_1^u(t-1) \\ ch_2^u(t) = (r-1) \cdot ch_2^u(t-1) + r \cdot ch_1^u(t-1) \end{cases} \tag{4}$$

where ch_1^u and ch_2^u display the chromosomes, the point of crossover is denoted by u, while the uniform random real

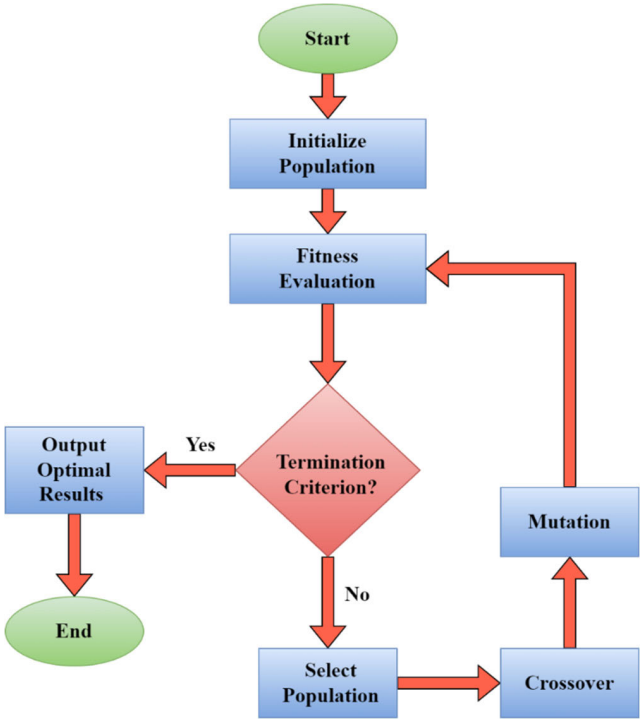

FIGURE 2. Overall flowchart of genetic algorithm.

number in (0, 1) is denoted by r. Eq. (4) demonstrates the pairing operations of crossover parent chromosomes $ch_i(t-1)$ to create new child chromosomes $ch_i(t)$. To prevent losing healthy chromosomes, the crossover rate should be decreased as chromosomes evolve. To do this, decrease the likelihood of crossing. Therefore, as demonstrated in Eq. (5), the adaptive crossover probability is used.

$$\begin{cases} P_{cross}(t) = \dfrac{P_{cross}(t-1)}{1+(f_{parent}-f_{current})}, & if\ f_{parent} > f_{current} \\ P_{cross}(t) = P_{cross}(t-1), & otherwise \end{cases} \tag{5}$$

where $f_{current}$ denotes the average fitness value of the current generation and f_{parent} denotes the average fitness value of the parent chromosome. Therefore, if the fitness value drops after a few iterations, the crossing chance will also drop. P_{cross} doesn't change anything else.

3) MUTATION

Another method of creating new chromosomes is by mutation, which involves modifying one or more genes in an existing chromosome. The algorithm can avoid becoming caught in a local minimum by using mutation. However, if convergence proceeds too slowly, random gene mutations could damage healthy chromosomes. The steps listed below are used to solve this issue:

1) Take the crossover operator's S chromosomes and leave them all unaltered.
2) Up until N_{ch} generations have passed, duplicate these chromosomes and use mutation to create new ones.
3) Keep the probability of mutation (1%) as small as possible to avoid chaotic behavior.

These little enhancements to the selection, crossover, and mutation operators produce better outcomes and hasten chromosome convergence.

When the best answer is identified, a genetic algorithm comes to an end. After a given number of repetitions, if a desired chromosome is not created, the algorithm terminates, and the chromosome with the best fitness in the most recent generation is considered an output of GA. Fig. 2 depicts a GA model's overall flowchart.

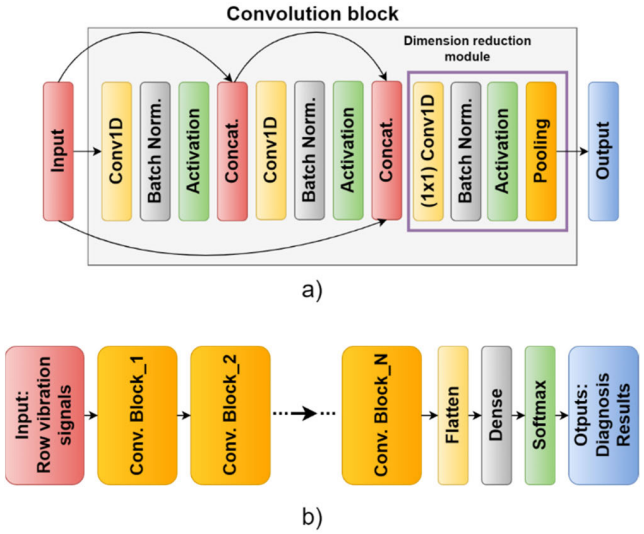

FIGURE 3. a) 1D-CNN block, b) overall network architecture.

III. AUTOMATICALLY OPTIMIZED A CNN ARCHITECTURE-BASED FAULT DIAGNOSIS SYSTEM

The suggested method for fault diagnosis of a hydraulic piston pump is presented in this section. First, direct connection based 1D-CNN block is explained, following that optimization process of 1D-CNN architecture and overall flowchart of the suggested system are described.

A. 1D-CNN BLOCK

Deep CNN architectures can be used to obtain more hierarchical features, which are essential for achieving good fault detection accuracy. Nevertheless, as network architectures become more complex, they become much more difficult to train perfectly because of issues with a gradient information flow that occur when utilizing back-propagation methods for training [33], [34], [35]. In order to solve this issue, the direct connection based 1D-CNN block is proposed for hydraulic piston pump fault diagnostics in this work. This block can help not only to improve the gradient information following and decrease the number of trainable parameters but also reduce the optimization time of network architecture. The proposed 1D-CNN block and entire network architecture are illustrated in Fig. 3.

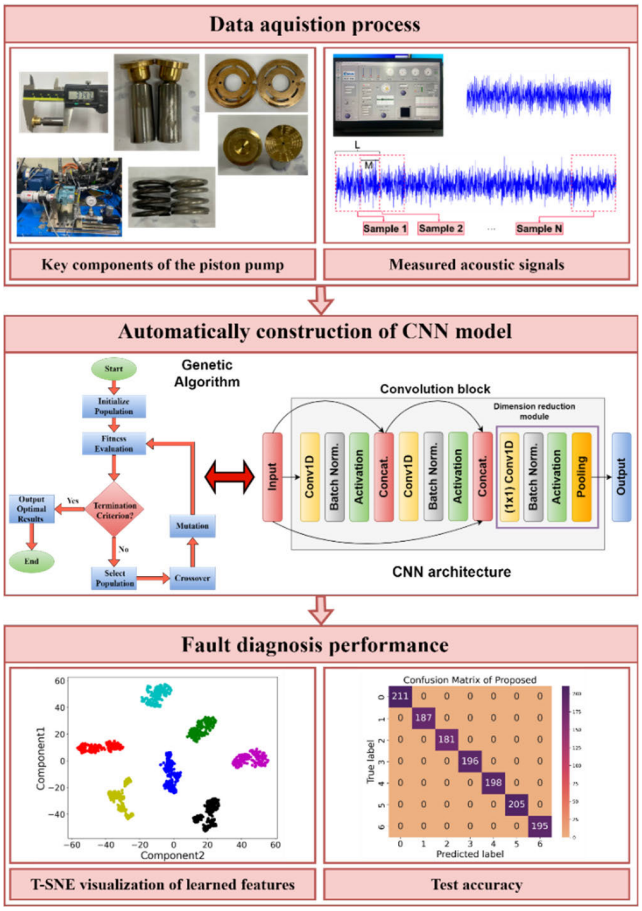

FIGURE 4. Overall framework of the represented approach.

The 1D-CNN block is one of the fundamental parts of the entire CNN architecture. The main goal of this block is to increase the gradient information flow by strengthening the connections between different CNN layers and reduce the number of trainable parameters. In this block, a 1-dimensional Conv layer extracts features of an input data, and a batch normalization (BN) layer is applied. Because BN is especially effective for deep networks and has produced positive results in deep learning [36]. The output of the BN layer passes an activation layer, and the output is concatenated with the initial input data. This data is passed the same three layers (Conv, BN, and activation) once again, and this data is concatenated with the initial input data and the previous concatenated data as shown in Fig. 3 (a). Following this, a dimension reduction module is applied. This module consists of four layers including Conv layer with 1×1 filter size for depth-wise reducing an input data, BN, activation and a pooling layer for widthwise reducing a data.

B. OVERALL PROCEDURE OF THE PROPOSED DIAGNOSIS SYSTEM

Fig. 4 shows the general framework for using the suggested automatically optimized 1D-CNN architecture for fault diagnosis of a hydraulic piston pump. The proposed diagnosis approach for fault detection of a hydraulic piston pump consists of three main parts, as illustrated in the overall framework. First, sensors are utilized to gather information about important parts of a piston pump, such as sound signals under normal and failure modes. The collected signals are separated into several samples. These samples are divided into train and test samples. The model is directly fed to the raw sound data. There is no requirement for manually created signal processing features like skewness, kurtosis, etc. Second, a neural network architecture based the 1D-CNN blocks and its hyperparameters are optimized simultaneously using GA. The hyperparameters of the neural network, their ranges and optimal results are listed in Table 1. The objective function of GA is the classification accuracy which is expressed in Eq. 6 on the test dataset.

TABLE 1. The best hyperparameters obtained by GA.

No.	Hyperparameters	Ranges	Optimal results
1	Number of blocks	[1, 7]	2
2	Number of filters	[8, 256]	32
3	Filter size	[3, 7]	7
4	Number of neurons	[25, 200]	100
5	Learning rate	[0.00001, 0.01]	0.00092
6	Activation functions	[relu, tanh, sigmoid, LeakyReLU, selu, elu]	relu
7	Optimizers	[SGD, Adam, RMSprop, Adadelta, Adagard, Adamax, Nadam]	Adam
8	Number of filters for dimension reduction module 1	[8, 256]	16
9	Number of filters for dimension reduction module 2	[8, 256]	8
10	Batch size	[4, 256]	8
11	Number of epochs	[20, 200]	80

$$Test\ error = \frac{true_{value} - predicted_{value}}{total_{predictions}} \tag{6}$$

As described above, automatically constructing a 1D-CNN architecture with the ideal number of CNN layers and other hyperparameters for fault diagnosis is the major goal of this paper. The following parameters are used in the GA: the maximum number of generations is 20 and the probabilities of crossover and mutation are 0.4 and 0.2, respectively. A categorical cross entropy function is used as a loss function of the proposed neural network. The prediction and classification are finished using the softmax regression tool. These terms are expressed as follows:

Categorical cross entropy:

$$L_{CE} = -\sum_{i=1}^{N} t_i \cdot \log(y_i), \tag{7}$$

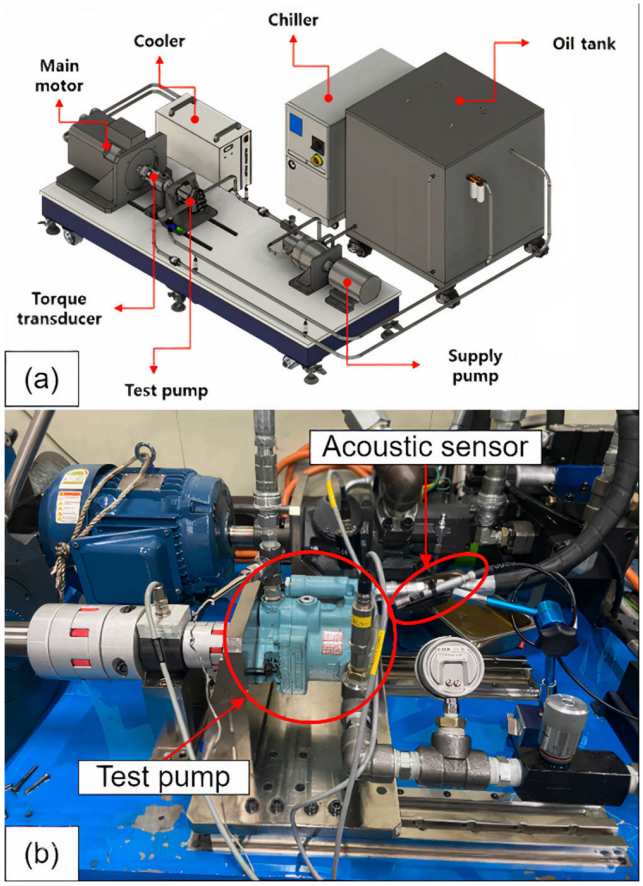

FIGURE 5. (a) overall experimental set up, (b) tested pump.

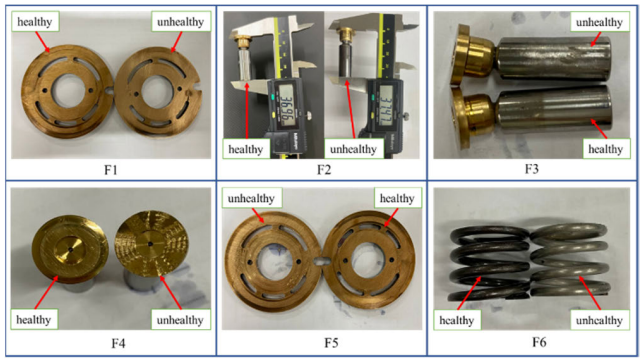

FIGURE 6. Damaged components of the hydraulic piston pump.

Softmax:

$$f(y)_i = \frac{e^{y_i}}{\sum_j^N e^{y_j}}, \tag{8}$$

where N denotes a number of classes, t_i is the truth label, y_i represents the softmax probability for i^{th} class, y_j is the score inferred by the network for each class in N. Finally, fault classification is performed using the automatically optimized 1D-CNN model and feature visualization is performed using the t-SNE method.

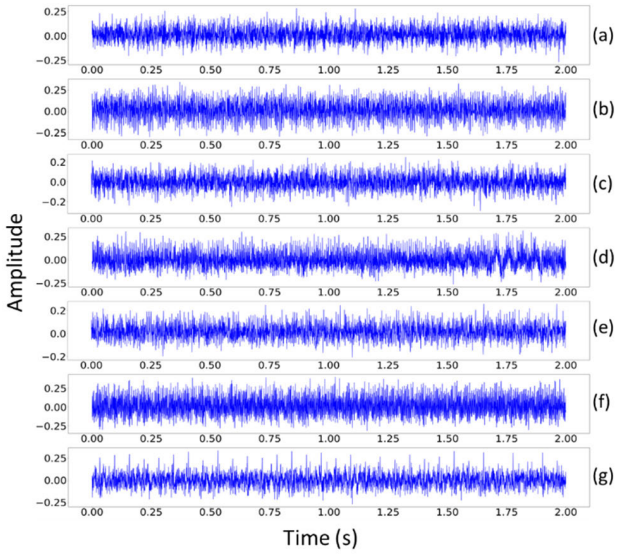

FIGURE 7. Row acoustic signals, (a) normal, (b) cavitation erosion on port plate, (c) loose slipper, (d) worn piston, (e) worn slipper, (f) worn port plate, (g) central spring wear.

IV. EXPERIMENTAL SET UP AND DATA COLLECTION

Experimental testing performed in the mechanical engineering department's lab at Inha University served as the basis for the research effort. Table 2 lists the pump's specifications. The sound signals of the piston pump are acquired by an acoustic sensor while working conditions. Pumps are not used in the same conditions in real environments, they can be operated at different rotation speeds or under different pressure. Therefore, the test pump is operated in several conditions which are listed in Table 3 and the sound signals are collected. By this, enough datasets can be collected for feeding the proposed DL model and more complex and different enriched datasets can be established. The tested hydraulic piston pump (a) and overall experimental set up (b) are represented in Fig. 5. If the sensor is not enough positioned to the sound signal source, the desired sound data might be masked by environmental noise or mechanical transmission sound. In order to obtain a signal with a greater signal-to-noise ratio, the near sound field measurement approach is used. The acoustic sensor (Model: Bruel & Kjaer Type 2671) is mounted on a stationary base near to the hydraulic piston pump as shown in Fig. 5. The distance between the sensor and the body of the pump is 0.15 *meters*. The measured signals while operating is collected on a laptop via DAQ (NI cDAQ-9174) module. The signals are recorded at 20 *kHz* sampling rate during 10 *seconds* for each condition.

Acoustic signals from the pump in various health conditions are collected for fault diagnosis. Four common faults of hydraulic piston pumps including worn port plate, cavitation erosion, worn slipper and damaged cylinder block are investigated in the reference [37]. The references [18], [19] also focus four types of failure modes. In this study, six very

TABLE 2. Specification of the tested hydraulic piston pump.

Name	Specification
Model	PV-0B-80-30
Volume/rev	8.0 (cm^3/rev)
Pressure adjustment range	30.6 to 214 *MPa*
Permitted peak pressure	25 *MPa*
Rotating speed	500-2000 *1/min*
Mass	7.7 *kg*

TABLE 3. Operating conditions of the hydraulic piston pump for data acquisition.

Angular Velocity	Swash Angle	Pressure (*bar*)	
		100	200
1000 *rev/min*	12.8 *deg*	✓	✓
1500 *rev/min*	12.8 *deg*	✓	✓
2000 *rev/min*	12.8 *deg*	✓	✓

TABLE 4. The health conditions of a piston pump.

Operating modes	Description	Code	Label index
Healthy condition	Normal	H	0
Failure conditions	Cavitation erosion on port plate	F1	1
	Loose slipper	F2	2
	Worn piston	F3	3
	Worn slipper	F4	4
	Worn port plate	F5	5
	Central spring wear	F6	6

common failure modes which are listed in Table 4 of a piston pump are analyzed. Damaged components of the pump are illustrated in Fig. 6.

As mentioned above, the proposed diagnosis system receives 1D time series data. Therefore, obtained acoustic signals can be fed directly to 1D-CNN model. The row acoustic signals of 2 *minutes* time duration for each mode of the pump are shown in Fig. 7.

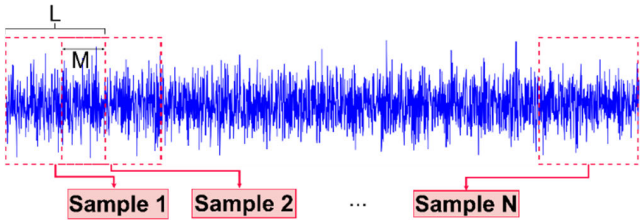

FIGURE 8. Sampling process of the row sound signal.

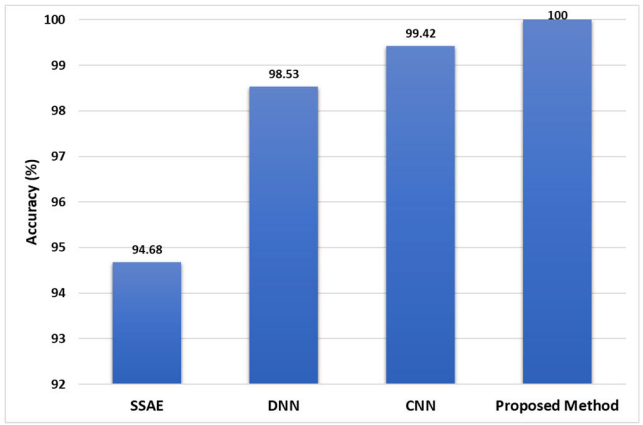

FIGURE 9. Diagnosis performances of selected DL models and the proposed model.

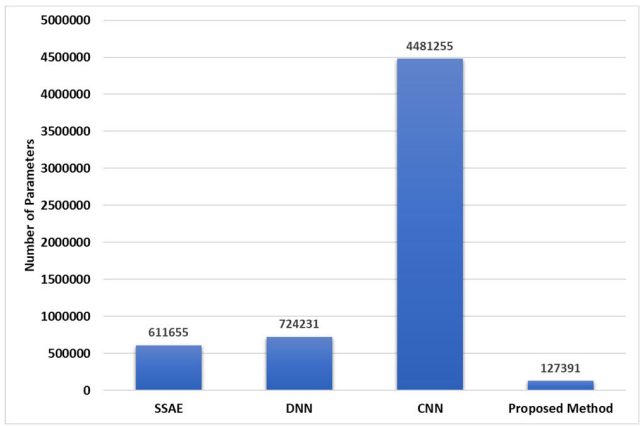

FIGURE 10. Number of trainable parameters of the models.

V. RESULTS AND DISCUSSION

A. INPUT DATA DESCRIBPTION

The collected sound signals are divided into the same length of pieces. In other words, one sample of the dataset makes up 512 data points (L) of the row sound signal. And the next data sample makes up another 512 data points by moving next to 256 data points (M) as shown in Fig. 8.

Acoustic signals are pieced 6860 data samples overall, 980 samples for each of the seven modes. The dataset is arbitrarily divided into a training dataset and a test dataset with an 8:2 sample ratio, giving the training dataset 5488 samples and the test dataset 1800 samples. 840 training samples and 1372 test samples are utilized to train and test the model, respectively, for each mode. Please take note that no training was done using the test data. To extract more helpful features, the training data is randomly flipped horizontally.

B. ANALYSIS OF THE EXPERIMENTAL RESULTS

Aforementioned above, DNN, SSAE and CNN DL models are commonly employed on a fault diagnosis and health monitoring. Therefore, these DL models have been selected for comparative study. The models are implemented using one of DL toolkits called Tensorflow 2.0 version. Hardware settings used in this study are as follows. Central Processing Unit (CPU) is an AMD Ryzen 9 5900X 12-Core Processor

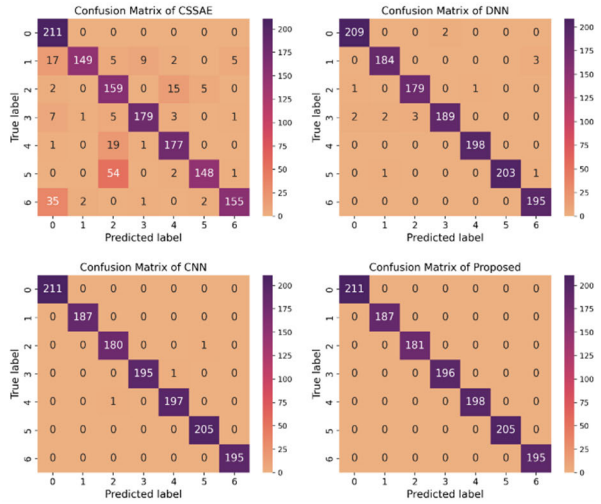

FIGURE 11. Confusion matrix of DL and proposed models.

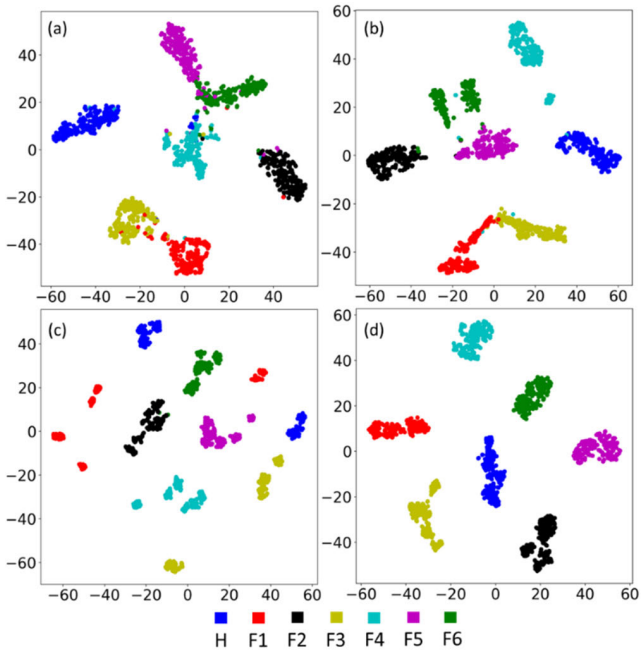

FIGURE 12. Feature representations of the last layer of models via t-SNE. (a) SSAE, (b) DNN, (c) CNN and (d) proposed model.

3.70 GHz (RAM: 64 GB), Graphics Processing Unite (GPU) is NVIDIA GeForce RTX 3060 (12 GB memory). The result of comparative study is represented in Fig. 9. The accuracy of the models in this figure is an average value of ten times repetition of training. Fig. 9 shows that the automatically optimized CNN architecture has performed the highest accuracy among the other compared models. If it is looked the graph, SSAE and DNN has shown relatively poor diagnosis performance. Therefore, it can be said that convolutional layers can help solve domain knowledge dependency issues.

Dimension reduction module is used in 1D-CNN block to lower the number of parameters required to build effective DL models. The results of the suggested fault diagnosis method are evaluated in this research, and the number of trainable parameters for DL-based methods is examined. The best diagnostic performance with the fewest parameters is achieved using our proposed method, as shown in Fig. 10. Although the proposed model is the deepest model, the number of trainable parameters is fewer than that of the rest models. The dimension reduction module in the 1D-CNN block is responsible for this benefit, which can be verified by counting the parameters in both the standard and reduced-size versions. Numerous parameters might be eliminated if only the dimension reduction module is used in the 1D-CNN block. Thus, the benefits of 1D-CNN blocks-based methods allow for the efficient construction of diagnosis models with deep architectures, which in turn improves the diagnostic performance.

Classification accuracy is a parameter that can only be used to measure the algorithm's overall performance and not to draw attention to the most pressing problems with data classification. To tackle these problems, the confusion matrix is commonly used. Accuracy in statistical categorization can be shown using the confusion matrix. The rows of the matrix represent the actual values, while the columns represent the predicted category. The confusion matrix for a classifier with N categories is a N×N square matrix.

Fig. 11 demonstrates the classifiers that enable accurate classification of the pump's state of health. The confusion matrix, in particular, enables detailed observation of which fault scenarios are appropriately categorized. Considerable things can be done based on the matrices' analysis. Looking at the figure, SSAE and DNN models fail in detection of all classes while CNN model has only failed in predicting label 2 (F2), label 4 (F4) and label 5 (F5) one time. All faults are successfully classified by the proposed method. It can be said that the proposed method is reliable.

The visualization of the results is crucial and useful to reveal high-dimensional feature representations in order to further leverage the internal CNN model process for the automatic learning of hidden features. t-distributed Stochastic Neighbor Embedding (t-SNE) method is a popular and useful technique in DL for addressing the issue of non-linear dimension reduction [38]. For SNE, a probability distribution is constructed in high-dimensional space to describe the similarity between the points by translating the Euclidean distance into conditional probability. The likelihood of choosing similar objects is higher than the likelihood of choosing objects that are not similar. The Gaussian distribution is used to establish the probability distribution in low-dimensional space. The two probability distributions mentioned above have been tuned for the best approximation. As a result, SNE frequently keeps data's regional characteristics. Symmetric SNE cost function is used in t-SNE as opposed to conventional SNE. It is clear that SNE and t-SNE differ from one other in two ways. One is that the joint probability distribution in high-dimensional space takes the role of the conditional probability distribution, and the gradient calculation procedure

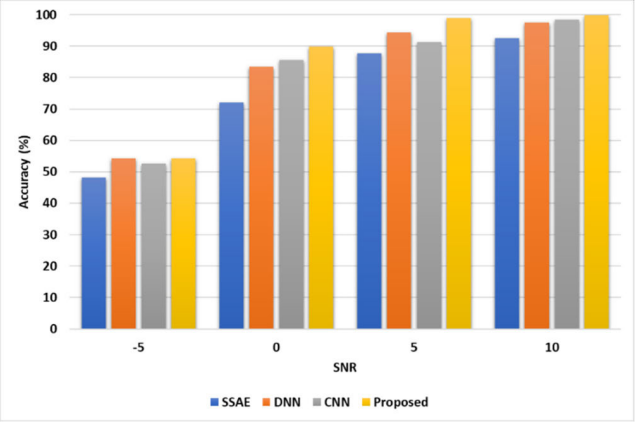

FIGURE 13. Diagnostic performances of DL-based approaches when subjected to varying degrees of additive noise.

is made simpler. The second is that t-distribution, which emphasizes a long-tail distribution, is used in place of Gaussian distribution. This avoids the issues of optimization and crowding by allowing the medium and low distances in the high dimension to display a bigger distance after mapping. Therefore, t-SNE is better for acquiring overall properties. For instance, the features in one layer of a neural network are reduced to 2D or 3D using a principal component analysis (PCA), and reduced features can then be transferred to 2D, or 3D space for display.

It is possible to map the extracted feature data points to probability distributions. Both 2D and 3D matching distributions are possible. In this study, the 2D visualization of characteristics is chosen. The extracted features in the last layer of all models is shown in Fig. 12. From the figure, SSAE, DNN and CNN models have a few errors in learned features while the proposed model learns the successful features from row signal data points. Consequently, it can be said from the learnt feature visualization above that the proposed model can learn better features more quickly than other techniques.

The gathered acoustic signals in actual industrial contexts are invariably accompanied by background noises. Therefore, the effectiveness of the suggested model in a noisy environment must be confirmed. In this experiment, acoustic signals with varying signal-to-noise ratios (SNRs) in the range of -5 to 10 *dB* are combined with Gaussian white noise [39]. The SNR can be expressed a following equation:

$$SNR_{db} = 10log_{10}\left(\frac{P_{signal}}{P_{noise}}\right) = P_{signal,\ \ db} - P_{noise,\ db} \quad (9)$$

where P_{signal} denotes the power of the signal while P_{noise} denotes the power of the additive white Gaussian noise.

The comparative results under different noisy datasets are shown in Fig. 13. The result shows that adding more noise to the mix alters the distributions and properties of the test data, which leads to a decline in diagnosis performance across the board. In addition, the proposed optimized 1D-CNN architecture outperforms alternative DL-based models across the board in terms of signal-to-noise ratio (SNR). The proposed model also shows high diagnosis performance when noise is added, while the other models quickly degrade with the addition of noise.

VI. CONCLUSION

This study proposes an automatic optimization approach for fault diagnosis of hydraulic piston pumps using a 1D-CNN architecture and genetic algorithm. This approach introduces essential features and advantages that set it apart from other reviews. Firstly, our automated optimization process with the genetic algorithm eliminates manual tuning, making the fault diagnosis more efficient and reliable. Secondly, by incorporating direct connections and a dimension reduction model within the 1D-CNN block, we enhance gradient information flow and effectively manage trainable parameters, improving model performance and reducing overfitting risks. Furthermore, our use of raw acoustic signals without preprocessing or statistical feature extraction simplifies the data processing pipeline and potentially captures more relevant information for accurate fault diagnosis. Comprehensive comparisons with popular DL models demonstrate the superiority of our proposed system, achieving 99.99% accuracy with fewer parameters.

However, our study has limitations. Further validation on diverse datasets from different hydraulic piston pump systems is necessary for generalizability. Additionally, exploring unsupervised or transfer learning schemes can overcome the dependency on labeled data and enhance real-world applicability.

REFERENCES

[1] S. Nie, M. Guo, F. Yin, H. Ji, Z. Ma, Z. Hu, and X. Zhou, "Research on fluid-structure interaction for piston/cylinder tribopair of seawater hydraulic axial piston pump in deep-sea environment," *Ocean Eng.*, vol. 219, no. 100, pp. 108–222, 2021.

[2] J. M. Bergada, S. Kumar, and J. Watton, "Axial piston pumps, new trends and development," in *Fluid Dynamics, Mechanical Applications and Role in Engineering*. New York, NY, USA: Nova, 2012.

[3] J. Zhao, Y. Fu, J. Ma, J. Fu, Q. Chao, and Y. Wang, "Review of cylinder block/valve plate interface in axial piston pumps: Theoretical models, experimental investigations, and optimal design," *Chin. J. Aeronaut.*, vol. 34, no. 1, pp. 111–134, Jan. 2021.

[4] S. Ye, J. Zhang, B. Xu, L. Hou, J. Xiang, and H. Tang, "A theoretical dynamic model to study the vibration response characteristics of an axial piston pump," *Mech. Syst. Signal Process.*, vol. 150, Mar. 2021, Art. no. 107237.

[5] S. Guo, J. Chen, Y. Lu, Y. Wang, and H. Dong, "Hydraulic piston pump in civil aircraft: Current status, future directions and critical technologies," *Chin. J. Aeronaut.*, vol. 33, no. 1, pp. 16–30, Jan. 2020.

[6] L. Liu, J. Liu, Q. Zhou, and D. Huang, "Machine learning algorithm selection for windage alteration fault diagnosis of mine ventilation system," *Adv. Eng. Informat.*, vol. 53, Aug. 2022, Art. no. 101666.

[7] O. Eraliev, K.-H. Lee, and C.-H. Lee, "Vibration-based loosening detection of a multi-bolt structure using machine learning algorithms," *Sensors*, vol. 22, no. 3, p. 1210, Feb. 2022.

[8] C. Wang, C. Xin, and Z. Xu, "A novel deep metric learning model for imbalanced fault diagnosis and toward open-set classification," *Knowl.-Based Syst.*, vol. 220, May 2021, Art. no. 106925.

[9] H. Zhiyi, S. Haidong, Z. Xiang, Y. Yu, and C. Junsheng, "An intelligent fault diagnosis method for rotor-bearing system using small labeled infrared thermal images and enhanced CNN transferred from CAE," *Adv. Eng. Informat.*, vol. 46, Oct. 2020, Art. no. 101150.

[10] Y. Li, C. K. L. Lekamalage, T. Liu, P. Chen, and G. Huang, "Learning representations with local and global geometries preserved for machine fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 67, no. 3, pp. 2360–2370, Mar. 2020.
[11] S. R. Saufi, Z. A. B. Ahmad, M. S. Leong, and M. H. Lim, "Gearbox fault diagnosis using a deep learning model with limited data sample," *IEEE Trans. Ind. Informat.*, vol. 16, no. 10, pp. 6263–6271, Oct. 2020.
[12] X. Wen and Z. Xu, "Wind turbine fault diagnosis based on ReliefF-PCA and DNN," *Expert Syst. Appl.*, vol. 178, Sep. 2021, Art. no. 115016.
[13] K. Xu, X. Kong, Q. Wang, S. Yang, N. Huang, and J. Wang, "A bearing fault diagnosis method without fault data in new working condition combined dynamic model with deep learning," *Adv. Eng. Informat.*, vol. 54, Oct. 2022, Art. no. 101795.
[14] S. Tang, Y. Zhu, and S. Yuan, "A novel adaptive convolutional neural network for fault diagnosis of hydraulic piston pump with acoustic images," *Adv. Eng. Informat.*, vol. 52, Apr. 2022, Art. no. 101554.
[15] S. Tang, Y. Zhu, and S. Yuan, "An improved convolutional neural network with an adaptable learning rate towards multi-signal fault diagnosis of hydraulic piston pump," *Adv. Eng. Informat.*, vol. 50, Oct. 2021, Art. no. 101406.
[16] X. Liu, Q. Zhou, J. Zhao, H. Shen, and X. Xiong, "Fault diagnosis of rotating machinery under noisy environment conditions based on a 1-D convolutional autoencoder and 1-D convolutional neural network," *Sensors*, vol. 19, no. 4, p. 972, Feb. 2019.
[17] S. R. Saufi, Z. A. B. Ahmad, M. S. Leong, and M. H. Lim, "Low-speed bearing fault diagnosis based on ArSSAE model using acoustic emission and vibration signals," *IEEE Access*, vol. 7, pp. 46885–46897, 2019.
[18] Z. Chen, A. Mauricio, W. Li, and K. Gryllias, "A deep learning method for bearing fault diagnosis based on cyclic spectral coherence and convolutional neural networks," *Mech. Syst. Signal Process.*, vol. 140, Jun. 2020, Art. no. 106683.
[19] L. S. Dhamande and M. B. Chaudhari, "Compound gear-bearing fault feature extraction using statistical features based on time-frequency method," *Measurement*, vol. 125, pp. 63–77, Sep. 2018.
[20] P. B. Mallikarjuna, M. Sreenatha, S. Manjunath, and N. C. Kundur, "Aircraft gearbox fault diagnosis system: An approach based on deep learning techniques," *J. Intell. Syst.*, vol. 30, no. 1, pp. 258–272, Aug. 2020.
[21] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[22] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *J. Mach. Learn. Res.*, vol. 9, pp. 249–256, May 2010.
[23] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.
[24] K. Greff, "Training very deep networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 1–9.
[25] Y. Zhang, G. Chen, D. Yu, K. Yao, S. Khudanpur, and J. Glass, "Highway long short-term memory RNNS for distant speech recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 5755–5759.
[26] S. Tang, S. Yuan, and Y. Zhu, "Convolutional neural network in intelligent fault diagnosis toward rotatory machinery," *IEEE Access*, vol. 8, pp. 86510–86519, 2020.
[27] T. Shanthi and R. S. Sabeenian, "Modified Alexnet architecture for classification of diabetic retinopathy images," *Comput. Electr. Eng.*, vol. 76, pp. 56–64, Jun. 2019.
[28] Y. Han, B. Tang, and L. Deng, "An enhanced convolutional neural network with enlarged receptive fields for fault diagnosis of planetary gearboxes," *Comput. Ind.*, vol. 107, pp. 50–58, May 2019.
[29] M. Kim, J. H. Jung, J. U. Ko, H. B. Kong, J. Lee, and B. D. Youn, "Direct connection-based convolutional neural network (DC-CNN) for fault diagnosis of rotor systems," *IEEE Access*, vol. 8, pp. 172043–172056, 2020.
[30] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.
[31] T. Tan, Y. Qian, H. Hu, Y. Zhou, W. Ding, and K. Yu, "Adaptive very deep convolutional residual network for noise robust speech recognition," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 8, pp. 1393–1405, Aug. 2018.
[32] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn.*, 2015, vol. 37, no. 6, pp. 730–743.
[33] P. Casoli, M. Pastori, F. Scolari, and M. Rundo, "A vibration signal-based method for fault identification and classification in hydraulic axial piston pumps," *Energies*, vol. 12, no. 5, p. 953, Mar. 2019.
[34] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 1, pp. 2579–2605, 2008.
[35] Y. Zhang, L. Gao, X. Wen, and H. Wang, "Intelligent fault diagnosis of machine under noisy environment using ensemble orthogonal contractive auto-encoder," *Expert Syst. Appl.*, vol. 203, Oct. 2022, Art. no. 117408.
[36] V. Thakkar, S. Tewary, and C. Chakraborty, "Batch normalization in convolutional neural networks—A comparative study with CIFAR-10 data," in *Proc. 5th Int. Conf. Emerg. Appl. Inf. Technol. (EAIT)* , 2018, pp. 1–5.
[37] P. Casoli, M. Pastori, F. Scolari, and M. Rundo, "A vibration signal-based method for fault identification and classification in hydraulic axial piston pumps," *Energies*, vol. 12, no. 5, 2019.
[38] A. Gisbrecht, A. Schulz, and B. Hammer, "Parametric nonlinear dimensionality reduction using kernel t-SNE," *Neurocomputing*, vol. 147, no. 1, pp. 71–82, 2015.

OYBEK ERALIEV MARIPJON UGLI received the B.S. and M.S. degrees in electrical engineering from the Fergana Polytechnic Institute, Fergana, Uzbekistan, in 2012 and 2014, respectively. He is currently pursuing the Ph.D. degree with Inha University, Incheon, South Korea. From 2014 to 2019, he was a Lecturer with the Fergana Polytechnic Institute. His research interests include machine learning, deep learning, and fault diagnosis.

KWANG-HEE LEE is currently pursuing the Ph.D. degree with Inha University, Incheon, South Korea. His research interests include machine learning, fault diagnosis, and tribology.

CHUL-HEE LEE (Member, IEEE) received the Doctor of Philosophy degree in mechanical and industrial engineering from the University of Illinois at Urbana–Champaign, in 2006. He was a Research Engineer with Hyundai Motor Company, from 1996 to 2002, and a Senior Research and Development Engineer with Caterpillar Inc., USA, from 2006 to 2007. He is currently a Professor with the School of Mechanical Engineering, Inha University. His research interests include virtual product development by design optimization and FE analysis.