Article

Intelligent Fault Diagnosis of Robotic Strain Wave Gear Reducer Using Area-Metric-Based Sampling

Yeong Rim Noh 1, Salman Khalid 1, Heung Soo Kim 1,* and Seung-Kyum Choi 2,*

1 Department of Mechanical, Robotics and Energy Engineering, Dongguk University-Seoul, 30 Pil-dong 1 Gil, Jung-gu, Seoul 04620, Republic of Korea; yrnoh_381@dgu.ac.kr (Y.R.N.); salmankhalid@dgu.ac.kr (S.K.)
2 George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, North Ave NW, Atlanta, GA 30332, USA
* Correspondence: heungsoo@dgu.edu (H.S.K.); schoi@me.gatech.edu (S.-K.C.); Tel.: +82-2-2260-8577 (H.S.K.); +1-404-894-9218 (S.-K.C.); Fax: +82-2-2263-9379 (H.S.K.); +1-404-894-9342 (S.-K.C.)

Abstract: The main challenge with rotating machine fault diagnosis is the condition monitoring of machines undergoing nonstationary operations. One possible way of efficiently handling this situation is to use the deep learning (DL) method. However, most DL methods have difficulties when the issue of imbalanced datasets occurs. This paper proposes a novel framework to mitigate this issue by developing an area-metric-based sampling method. In the proposed process, the new sampling scheme can identify which locations of the datasets can potentially have a high degree of surprise. The basic idea of the proposed method is whenever significant deviations from the area metrics are observed to populate more sample points. In addition, to improve the training accuracy of the DL method, the obtained sampled datasets are transformed into a continuous wavelet transform (CWT)-based scalogram representing the time–frequency component. The dilated convolutional neural network (CNN) is also introduced as a classification process with the altered images. The efficacy of the proposed method is demonstrated with fault diagnosis problems for welding robots. The obtained results are also compared with existing methods.

Keywords: industrial robot; area metric; explainable artificial intelligence (XAI); fault diagnosis; data imbalance

MSC: 68T01

1. Introduction

In intelligent manufacturing industrial systems, robots are widely used to autonomously carry out tasks. However, their long-term and continuous operation cause malfunctions in subcomponents, such as gears and bearings. If maintenance is not promptly carried out, these malfunctions can create various failures in the system. These failures can affect productivity and increase the probability of a shutdown. At this point, preventive maintenance of industrial machinery is a common procedure to improve the reliability and safety of the equipment. This eliminates unnecessary maintenance procedures, significantly reducing operating costs, while revitalizing research on “condition-based maintenance”. Recently, vibration signal-based condition monitoring (CM) has been commonly used for system health monitoring [1]. However, vibration signal-based CM needs additional sensors for complicated data acquisition systems and is only suitable for large electric machines [2].

In modern industrial systems, the motor current signal can be easily acquired with a current transformer that is part of the protection system, and current-based monitoring can be performed remotely from the motor control center. However, robots operate under various operating and load conditions in the industry, resulting in nonstationary current signals. Several CM technologies have been developed to detect faults in the nonstationary
system. Fault diagnosis methods typically include model-based methods and data-driven methods. Data-driven models are more suitable for diagnosing systems with complex signals, because unlike model-based methods, they do not require signal patterns and error mechanisms [3]. Machine learning (ML) algorithms, such as artificial neural networks (ANNs), support vector machines (SVMs), principal component analysis (PCA), and K-nearest neighbor (KNN), have been applied to data-driven methods [4–6]. However, ML algorithms rely on human-engineered features to train, which can lead to inaccurate classification results [7]. Research on deep learning (DL)-based methods is actively being conducted to achieve much better accuracy under various operating conditions. Because DL is based on distributed feature learning, it can automatically learn correlations and characteristics among data, eliminating the need for feature engineering processes. However, autonomous learning of these features requires massive amounts of data.

Recently, many studies have been actively carried out to overcome the problem of data imbalance and data shortage through data augmentation. In real-world image augmentation techniques, image classification through DL includes image enlargement and reduction, rotation, and flipping [8–10]. For the time series, data augmentation techniques include window warping, magnitude warping, cropping, generative adversarial networks (GANs), and seasonal and trend decomposition using loess (STL) [11–14]. However, in the data augmentation methods mentioned above, the augmented data highly depend on specific input data and may not reflect the physical meaning of the original data. In previous studies, deep learning has provided excellent prediction and classification performance. However, it is impossible to interpret the model’s internal parameters for the results. If model analysis and explanation are not possible, it is difficult to convince engineers and other stakeholders, because it is challenging to present a rationale for the preventive and management schedule through system condition monitoring [15]. In this study, the challenges of data imbalance and shortage are addressed by enhancing the model’s reliability using a minimal number of samples through the area-metric-based sampling method. This approach involves a systematic sampling technique based on geometric characteristics within the gear mechanism. This method ensures that the augmented data maintain the physical essence of the original data, allowing for a more accurate representation. Subsequently, a specialized fault diagnosis approach using convolutional neural networks (CNNs) tailored for nonstationary systems is applied. Notably, this strategy not only improves the model’s interpretability but also facilitates a clear explanation of the results using explainable artificial intelligence (XAI).

The remainder of this paper is structured as follows: Section 2 introduces prior studies on ML- and DL-based fault diagnosis, data imbalance and data shortage problems, and XAI. Section 3 introduces the proposed intelligent fault diagnosis of the robotic strain wave gear reducer, while Section 4 describes the experiment in detail. Section 5 discusses robotic strain wave gear reducer fault diagnosis to verify the proposed method and interpret the results as XAI, and Section 6 then concludes this work.

2. Related Work

This section will provide an in-depth exploration of the existing research that closely aligns with our proposed study. It focuses on the utilization of ML and DL techniques for fault detection. Additionally, it will address challenges posed by data imbalance and scarcity, and shed light on the concept of explainable artificial intelligence.

2.1. Fault Diagnosis Using ML and DL

Traditional signal feature analysis methods (e.g., time domain feature analysis, frequency domain feature analysis, and time–frequency domain feature analysis) are unsuitable for diagnosing systems with multiple failure modes and nonstationary operations. As shown in Table 1, fault detection and diagnosis (FDD) based on ML and DL algorithms have recently been widely studied to solve the problems of traditional signal feature analysis methods.
Table 1. Summary of FDD methods using ML models and DL models.

<table>
<thead>
<tr>
<th>Input</th>
<th>Model</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vibration signal</td>
<td>ANN</td>
<td>Kankar et al., 2011 [4]</td>
</tr>
<tr>
<td>Statistical feature</td>
<td>SVM</td>
<td>Kankar et al., 2011 [4]</td>
</tr>
<tr>
<td>Vibration signal</td>
<td>ANN</td>
<td>Kankar et al., 2011 [4]</td>
</tr>
<tr>
<td>Statistical feature</td>
<td>SVM</td>
<td>Kankar et al., 2011 [4]</td>
</tr>
<tr>
<td>Voltage and current signal</td>
<td>KNN</td>
<td>Samanta et al., 2016 [5]</td>
</tr>
<tr>
<td>Statistical features (time and frequency domains) and select features using PCA/ICA</td>
<td>SVM/RVM</td>
<td>Widodo et al., 2019 [6]</td>
</tr>
<tr>
<td>RGB multisignal image</td>
<td>CNN</td>
<td>Xie et al., 2021 [16]</td>
</tr>
<tr>
<td>Gray scale current signal image</td>
<td>CNN</td>
<td>Hoang et al., 2020 [17]</td>
</tr>
<tr>
<td>Frequency spectrum of the vibration data</td>
<td>CNN</td>
<td>Janssens et al., 2016 [18]</td>
</tr>
<tr>
<td>Light emission images</td>
<td>CAE</td>
<td>Park et al., 2011 [19]</td>
</tr>
</tbody>
</table>

Xie et al. [16] proposed three-channel red–green–blue (RGB), based on the principal component analysis of multisignal data, and conversion into an RGB image. After data preprocessing, a classification model based on a CNN was adopted to determine the fault diagnosis of bearings. Hoang et al. [17] proposed a fault diagnosis technique based on current motor data using deep learning and information fusion (IF) applicable to the external bearings of rotating mechanical systems. Janssens et al. [18] studied the classical feature engineering-based method and a CNN-based autonomous feature learning method through bearing vibration signal analysis to discriminate multiple bearing fault states, where the latter showed 6.36% higher accuracy than the former. Park et al. [19] conducted a study on detecting defects in composite delamination by acquiring light emission images using mechanoluminescent (ML) sensors and comparing the reconstruction error between normal and malfunctioning ones using a convolutional autoencoder (CAE). Saufi et al. [20] processed limited gearbox sample data using a deep learning model based on a stacked sparse autoencoder (SSAE). The proposed methodology improved the accuracy of gearbox fault diagnosis by 10–20% compared to the standard SSAE model. Table 1 summarizes the system condition monitoring method using the ML and DL models.

2.2. Data Imbalance and Shortage

Deep learning-based system fault diagnosis requires a large amount of training data. Zhong et al. [21] performed troubleshooting of the air handling unit (AHU) subsystem by extracting 3000 samples for each health condition. Kumar et al. [22] used data from 18,000 images for bearing fault detection. However, at actual industrial sites, it is hard to secure sufficient quality data due to noise. In addition, the standard signal of the system can be sufficiently secured. Still, there is a limit to ensuring a sufficient amount of fault data, because the system is stopped when the system fails, or a fault prognosis appears. Yang et al. [23] suggested that there are difficulties in obtaining a sufficient number of data samples in industrial systems. In this respect, many studies are being actively conducted to overcome the data imbalance and shortage problem.

Table 2 summarizes the solutions to data imbalance and shortage problems in previous studies. As shown in Table 2, Lu Qian et al. [24] augmented the data by developing a dynamic model that describes the bearing system by adjusting the bearing parameters (fault size, load, and speed). Khan et al. [8] proposed the deep learning-based fault diagnosis of 42 classes of rotating machines through synthetic data augmentation using virtual sensors. Through data augmentation, 15,608 data samples were used for training. Hu et al. [25] introduced a data augmentation algorithm that simulates data at various workloads and rotational speeds through resampling technology and proposed a self-adaptive convolutional neural network for fault diagnosis. Liu et al. [26] proposed a method to detect faults in wind turbines with small samples by generating synthetic fault data through GANs. Liu et al. [27] proposed high-quality data augmentation through a wavelet capsule.
generative adversarial network (WCGAN) for rolling bearing fault diagnosis. Li et al. [28] used an improved auxiliary classifier generative adversarial network (ACGAN) model to generate high-quality samples. Wang et al. [29] attempted data augmentation using a conditional variational autoencoder generative adversarial network (CVAE–GAN). They generated fault samples that were similar to real fault data by acquiring the distribution of fault samples from the encoder network.

Table 2. Summary of approaches to data imbalances and shortages in previous studies.

<table>
<thead>
<tr>
<th>Input</th>
<th>Method</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time series data</td>
<td>Window warping</td>
<td>Le Guennec et al., 2016 [11]</td>
</tr>
<tr>
<td>Accelerometer and gyroscope sensor data</td>
<td>Cropping</td>
<td>Li et al., 2019 [13]</td>
</tr>
<tr>
<td>Wearable sensor data</td>
<td>Magnitude-warping</td>
<td>Um et al., 2017 [12]</td>
</tr>
<tr>
<td>Computed tomography (CT) image</td>
<td>GAN</td>
<td>Frid-Adar et al., 2018 [14]</td>
</tr>
<tr>
<td>Vibration signal</td>
<td>Model-based data augmentation</td>
<td>Lu Qian et al., 2022 [24]</td>
</tr>
<tr>
<td>Vibration signal</td>
<td>Synthetic data augmentation</td>
<td>Khan et al., 2021 [8]</td>
</tr>
<tr>
<td>Bearing dataset</td>
<td>Simulates data at various workloads and rotational speeds</td>
<td>Hu et al., 2020 [25]</td>
</tr>
<tr>
<td>Wind turbine conditions based on supervisory control and data acquisition (SCADA) data</td>
<td>GAN</td>
<td>Liu et al. [26]</td>
</tr>
<tr>
<td>Vibration signal</td>
<td>WCGAN</td>
<td>Liu et al. [26]</td>
</tr>
<tr>
<td>Frequency spectrum</td>
<td>ACGAN</td>
<td>Li et al., 2019 [28]</td>
</tr>
<tr>
<td>Frequency spectrum</td>
<td>CGAN</td>
<td>Wang et al., 2020 [29]</td>
</tr>
</tbody>
</table>

2.3. Explainable Artificial Intelligence

Deep learning-based system diagnosis has shown good performance in recent studies. However, deep learning models do not provide evidence for system diagnosis results and reasons for improving model performance, which leads to a decrease in model reliability. To solve these problems, research to explain deep learning models through XAI is being actively conducted.

Sairam et al. [30] used an irradiance-based three-diode model (IB3DM) and extreme gradient boosting (XGBoost) to build an XAI initial fault detection model of photovoltaic (PV) panels. Srinivasan et al. [31] proposed trust building for cooler fault detection models via XAI based on the locally interpretable model-agnostic description (LIME), which is suitable for interpreting ML models with many predictors. Morichetta et al. [32] proposed LIME to provide evidence for unsupervised clustering results for image-quality classification. Shrikumar et al. [33] proposed deep learning important features (DeepLIFT), which is neural network output prediction decomposition for a specific input that compares the difference between the activation of neurons in the network and the reference activation, and provides a contribution score. Class activation mapping (CAM) and Grad-CAM are some of the CNN-based XAI techniques. They are among the most successful methods for generating training attention region saliency maps of models [34,35]. Chen et al. [15] performed fault diagnosis using a bearing vibration signal through the application of CNN-based XAI using Grad-CAM and explained the diagnosis result.

As represented in the previous studies, many methods have been developed for the fault diagnosis of mechanical systems. However, in actual industries, it is difficult to obtain sufficient quality data which, in turn, causes data imbalance and shortage problems. The existing data augmentation method to solve data imbalance and shortage problems cannot fully reflect the physical meaning of the original data, so a new method is needed. In addition, deep learning-based fault diagnosis of mechanical systems has difficulties in actual industrial application, because it is difficult to present a basis for the judgment result. The main contributions of this study that solve the problems of existing studies are summarized as follows: (1) Area-metric-based data scanning, which intensively scans sample data that
deviate significantly from the overall data trend. This additional step enables the good detection of anomalies in the data, even with a small number of samples. The extracted samples are converted into time–frequency images. (2) We propose an improved CNN model: dilated CNN. The dilated convolution kernel is introduced to extract features of time–frequency images and perform classification. (3) The proposed method is verified using performing robotic strain wave gear reducer failure diagnosis. (4) Grad-CAM is used to present the basis for judgment in each convolutional layer. Section 3 presents the details of the proposed framework.

3. Proposed Methodology

In this section, the theoretical background and explanation of the proposed method are described. Figure 1 shows the flow of the system fault diagnosis method using area-metric-based sampling. First, area-metric-based data sampling identifies locations in datasets that have a potentially high degree of surprise. Second, the sampled data are converted into a continuous wavelet transform (CWT)-based scalogram image containing time–frequency information. Finally, the converted image is trained on an improved CNN model to proceed with classification. The basis for the judgment in each model layer is expressed in a heat map format using XAI.

Figure 1. Overall process of the proposed fault diagnosis method.

The choice of the CWT for the data conversion arose from its effectiveness in analyzing signals in both the time and frequency domains simultaneously. Unlike traditional Fourier-based approaches, CWT provides a time-scale representation, making it suitable for the analysis of nonstationary signals. This representation enables the capture of localized features at various scales, a critical aspect for detecting faults in dynamic systems. By
utilizing CWT, essential time–frequency characteristics of the signal are preserved, thereby enhancing the representational power for subsequent analysis.

CNNs were employed for fault diagnosis because of their exceptional ability to automatically learn and extract hierarchical features from input data. CNNs excel in pattern recognition tasks, making them apt for our fault diagnosis objectives. The hierarchical feature learning in CNNs, through multiple convolutional layers, enables the model to identify intricate patterns and representations in the transformed data obtained through CWT. CNNs have proven to be highly successful in various signal processing and image recognition tasks, making them a suitable choice for our fault diagnosis application. The integration of CWT for data transformation and CNN for fault diagnosis represents a synergistic approach that leverages the strengths of each method. CWT improves the data representation by extracting meaningful features in the time–frequency domains, and CNN efficiently learns and processes these features to classify and diagnose faults. This combination ensures that the diagnostic model can effectively recognize fault patterns in the transformed data, contributing to accurate and reliable fault diagnosis in nonstationary systems.

3.1. Area Metric

The area metric is a validation metric that performs quantitative comparisons mainly by calculating the area difference between the cumulative distribution function (CDF) graph of simulation prediction, and the CDF of the actual experimental results [36]. As the graph area difference is small, it can be expected that the simulation prediction result is similar to the actual data [36]. The CDF calculation for the comparison target is required to perform the area metric. The CDF of the random variable \(X\) is defined as follows:

\[
F_X(x) = P(X \leq x), \text{ for all } x \in \mathbb{R}
\] (1)

As shown in Figure 2, the area metric \(M\) is represented as the area between the experimental results CDF \(F_a(x)\) and the simulation prediction results \(F_s(x)\) [37].

\[
M = \frac{1}{\overline{P}^d} \int_{-\infty}^{\infty} |F_s(x) - F_a(x)|dx
\] (2)

where \(\overline{P}^d\) is the absolute mean of the actual data, and \(M\) is normalized by \(\overline{P}^d\). In this study, an area-metric-based approach was employed for data scanning and analysis. The computed area-based metrics provide a quantifiable measure of the dissimilarity between the experimental and simulated CDFs, aiding in assessing the accuracy of the simulation predictions. A smaller area metric indicates a closer match between the experimental and simulated data distributions, signifying a more reliable simulation. Conversely, a larger area metric suggests discrepancies between the experimental and simulated data distributions, indicating areas for potential improvement in the simulation model.

Area-Metric-Based Sampling Method

**Step 1. Data resampling**

Data acquisition (DAQ) collects motor current data for each axis of the robot operating under various operating conditions. Since the data length is different for each operating condition, the data length is set to be the same through resampling. The resampled data \(X_{RE}\), measured through the experiment, is defined as follows:

\[
X_{RE} = \{x_{ij} : x_{ij} \in \mathbb{R}^{n \times m}\}
\] (3)

The data can be represented by an \(n \times m\) matrix \((i = 1, 2, \ldots, n, j = 1, 2, \ldots, m)\). Each row \(x_j \in \mathbb{R}^{n \times 1}\) is the resampled data from the DAQ, and each column \(x_j \in \mathbb{R}^{1 \times m}\) is the number of different operating conditions.
Figure 2. Illustration depicting the computation of the area-based metrics.

**Step 2. Data segmentation and area metric calculation**

In this study, the empirical cumulative distributive function (ECDF) is derived for the area metric calculation. Based on the resampled data, a reference ECDF, \( F_{RE}(x) \), which can indicate the trend of the entire data, is derived from the entire scenario data for each operating condition. The reference ECDF is defined as follows, where \( I \) represents an indicator function:

\[
F_{RE}(x) = \frac{1}{m} \sum_{i=1}^{m} I[x_i \leq x]
\]  

(4)

As shown in Figure 3a, the data are segmented into a specific window size \((k)\) to find the abnormal part of the data. Thus, the segmented data \((X_{SE})\) are defined as:

\[
X_{SE} = \{x_{ij} : x_{ij} \in R^{k \times l}\}, i = 1, 2, \ldots, k, j = 1, 2, \ldots, l
\]  

(5)

Figure 3b shows the representation of the reference ECDF and the segmented data ECDF, offering a graphical understanding of the comparison crucial for the subsequent analysis and decision making.

Where, \( X_{SE} \) is represented by a \( k \times l \) matrix. Each row, \( x_i \in R^{k \times 1} \), denotes each segmented data divided into \( k \) lengths, while each column, \( x_j \in R^{1 \times l} \), denotes the number of segmented data.
Each segmented data are converted into an ECDF. The ECDF of the segmented data, $F_{SE}(x)$, is defined as:

$$F_{SE}(x) = \frac{1}{l} \sum_{k=1}^{l} I[x_k \leq x]$$

(6)

The area metrics ($M_{RE-SE}$) can be determined by calculating the area difference between the ECDFs of the reference and the segmentation data:

$$M_{RE-SE} = \frac{1}{x_{SE}} \int_{-\infty}^{\infty} |F_{RE}(x) - F_{SE}(x)| \, dx$$

(7)

**Step 3. Data sampling**

The data are sequentially scanned for each operating condition, and the area metrics of the reference ECDF and the segmented data ECDF are obtained. The difference in the area metrics between the reference ECDF and the segmented data ECDF is calculated to identify the segments of data that significantly deviate from the overall data trend. Figure 4 illustrates a schematic of the area-metric-based sampling.

![Area-metric based sampling](image)

Figure 4. The area-metric-based sampling method.

**Step 4. Time–frequency imaging**

Unlike other time–frequency analysis methods, a wavelet transform provides multiresolution to facilitate dynamic analysis in the time–frequency domain. The wavelet transform is a method of decomposing a signal into different frequency bands through a wavelet function in the form of a wave [38]. The wavelet transform is performed using the dot product of the original signal and a set of wavelets. The set of wavelets, called analyzing wavelet or continuous wavelet, is a wavelet family realized through translating and scaling the basic wavelet, $\psi(t)$. The continuous wavelet can be described as:

$$\psi_{a,b}(t) = |a|^{-1/2} \psi(\frac{t-b}{a})$$

(8)

where $a$ is the scale parameter, and $b$ is the orientation parameter.

The wavelet transform has high-frequency and low-time resolution in the low-frequency band [39]. Still, it has high-time resolution and low-frequency resolution in the high-frequency band [39]. The CWT of the signal $x(t)$ can be conducted through a convolution operation between $x(t)$ and the complex conjugate of the wavelet family. If $a$ is contin-
uous variables, the CWT of the signal with finite energy ($W_x(a,b)$) is mathematically defined as [40]:

$$W_x(a,b) = \langle x(t), \psi_{a,b}(t) \rangle = |a|^{-1/2} \int_{-\infty}^{\infty} x(t) \psi^*\left(\frac{t - b}{a}\right) dt$$  \hspace{1cm} (9)

where $\psi^*$ is the complex conjugate of $\psi$. The absolute value of the CWT of a signal is the scalogram. The CWT of the signal can be represented as an image via a scalogram. The energy of $x(t)$ at scale $a$ ($S(a)$) is as follows:

$$S(a) = \|W_x(a,b)\| = \left(\int_{-\infty}^{\infty} |W_x(a,b)|^2 \, db\right)^{1/2}$$  \hspace{1cm} (10)

Figure 5 shows the conversion of a sample extracted by area metric data sampling to a CWT-scalogram:

![CWT-scalogram](image)

3.2. Dilated CNN

To solve complex problems and improve accuracy, the CNN structure is deepened by stacking layers, and the number of hyperparameters and parameters increases accordingly. Sometimes, this feature causes fatal drawbacks. As the network layer increases, backpropagation of the gradient suffers from vanishing gradient problems [41]. This may result in a sharp drop or saturation in the network performance. The issues mentioned above can be mitigated through dilated CNN. As shown in Figure 6, the dilated CNN is performed using dilated convolution kernels, unlike the traditional CNN, which uses traditional convolutional kernels.

Let $P_0, P_1, \ldots, P_{n-1} : \mathbb{Z}^2 \to \mathbb{R}$ be discrete functions, and $\omega_0, \omega_1, \ldots, \omega_{n-2} : \Omega \to \mathbb{R}$ be discrete filter with size $3 \times 3$. For the application of filters with exponentially increasing dilation, this can be expressed as:

$$P_{k+1} = P_k * 2 \omega_0 \text{ for } k = 1, 2, \ldots, n - 2$$  \hspace{1cm} (11)
Figure 6. Deep learning model framework: (a) CNN model; (b) dilated-CNN.

The receptive field of an element $\beta$, a component of $P_{k+1}$, is defined as a set of elements existing in $P_0$ that can estimate the value of $P_{k+1}(\beta)$. The size of the receptive field of each component in $P_{k+1}$ is $(2^{k+2} - 1) \left(2^{k+2} - 1\right)$. Figure 7 shows the convolutional kernel according to the dilation rate. In Figure 7a, $P_1$ is made from $P_0$ by a one-dilated convolution in which every element in $P_1$ has a receptive field of size $3 \times 3$. (Figure 7b) $P_2$ is made from $P_1$ by a two-dilated convolution, in which every element in $P_2$ has a receptive field of size $7 \times 7$. (Figure 7c) $P_3$ is made from $P_2$ by a two-dilated convolution in which every element in $P_2$ has a receptive field of size $15 \times 15$. As the dilation rate increases, the receptive field becomes more expansive, but the number of parameters remains the same. More information can be accommodated through this dilated convolution kernel without increasing the amount of computation.

Figure 7. (a) Receptive field change according to the dilation rate (b) Two-dilated convolution with a receptive field of size $7 \times 7$ (c) Two-dilated convolution with a receptive field of size $15 \times 15$ [42].

3.3. Grad-CAM

CAM is applicable only to CNN models based on global average pooling (GAP) architecture, but Grad-CAM is a model that can be applied to CNN models of other structures. Grad-CAM obtains a class activation map by calculating the gradient of the
class score ($y^c$), for class, with respect to feature maps ($A^k$) of the convolutional layer. Then, a global average pooling on the gradient flowing back can be performed:

$$a^c_k = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A^k_{ij}}$$

(12)

where $i, j$ represent each pixel of the feature map, $A^k$. $H$ and $W$ are rows and columns, and $Z$ represents the number of pixels of the $A^k$. Weight $a^c_k$ captures the importance of feature map $A^k$. Based on Equation (13), the rectified linear unit (ReLU) can be derived by removing negative values [43]:

$$L^c_{Grad-CAM} = ReLU \left( \sum_k a^c_k A^k \right)$$

(13)

4. Case Study

4.1. Data Acquisition System

This study introduces the methodologies introduced in Section 3 for robotic strain wave gear reducer fault diagnosis (the Robostar R004 welding robot with 6 degrees of freedom). Each axis of the robot is controlled by repeating acceleration/deceleration through a servomotor. As shown in Figure 8, the robotic strain wave gear reducer has two states: normal and faulty. The fault type is the bearing inner ring defect, and the failure is applied to axis 3. Table 3 shows detailed information on the robot:

![Target robot Robostar R004 and fault types.](image)

**Figure 8.** Target robot Robostar R004 and fault types.

**Table 3.** Description of the robot.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robot type</td>
<td>Robostar R004 welding robot</td>
</tr>
<tr>
<td>Sensitivity of sensor</td>
<td>64 mV/A</td>
</tr>
<tr>
<td>Health state</td>
<td>Normal, fault</td>
</tr>
<tr>
<td>Fault type</td>
<td>Strain Wave gear fault (bearing inner ring fault) in axis 3</td>
</tr>
<tr>
<td>Data type</td>
<td>1st phase of electrical current</td>
</tr>
<tr>
<td>Sampling frequency</td>
<td>2048 Hz</td>
</tr>
</tbody>
</table>

4.2. Experimental Set-Up

In this study, we analyze normal and fault motor current data in the 3rd axis to which faults are applied. The experiment was performed, as shown in Figure 9. The current data collection system is as follows. The robot operation cycle was divided using the programmable logic controller (PLC). The PLC provides trigger signals at the start and end of the robot motion, and stores them in the PC. All axes of the robot are equipped
with WCS2720 sensors, which are based on the Hall effect. The sensor signals are collected through the DAQ module, and the sampling rate was $f_s = 2048$ Hz.

![Image](image.png)

Figure 9. Experimental set-up of target robot: (a) experimental set-up; (b) car body welding; (c) welding points of car door.

Target robots are widely used for body welding. Table 4 shows detailed information on the robot’s operating conditions. A total of six working conditions (Operation 1_Normal, Operation 1_Fault, Operation 2_Normal, Operation 2_Fault, Operation 3_Normal, and Operation 3_Fault) used for car body welding were set. The robot operates under the set operating conditions. The robot repeats the welding motion in the sequence of car door points A–B–C–D, as shown in Figure 9c. After welding each point (A, B, C, and D), the robot moves to the next point (e.g., A → B, B → C, C → D). First, the robot completes the A-point welding of the car door in Operation 1. After that, it moves to point B at 40% of the maximum operating speed. In Operation 2, after the welding is completed at point B, it moves to point C at a rate of 20% of the maximum operating speed. In Operation 3, when it is completed at point C, it moves to point D at the maximum operating speed of 60%, and completes welding at point D. Figure 10 shows the motor current signal according to the robot’s operating conditions.

Table 4. Operating condition specifications.

<table>
<thead>
<tr>
<th>Motion</th>
<th>Properties</th>
<th>Data Size</th>
</tr>
</thead>
</table>
| Operation 1 (A → B) | 1. A-point welding  
   2. Moving from A to B  
   (40% of the maximum operating speed) | (27,839, 10) |
| Operation 2 (B → C) | 1. B-point welding  
   2. Moving from B to C  
   (20% of the maximum operating speed) | (37,799, 10) |
| Operation 3 (C → D) | 1. C-point welding  
   2. Moving from C to D  
   (60% of the maximum operating speed) | (25,634, 10) |

In this study, 500 samples of length (28 × 28) were extracted using the previously proposed area metric sampling method under each of six operating and health conditions, as summarized in Table 5. Afterward, the extracted samples were converted into CWT-based scalogram images, and classified through a deep learning model. Table 6 shows examples sampled of converted CWT-scalogram images in the experimental dataset using the proposed method.
Figure 10. Various operation conditions: (a) normal and fault states of Operation 1; (b) normal and fault state of Operation condition 2; (c) normal and fault states of Operation 3.

Table 5. Description of the health condition in the experimental dataset.

<table>
<thead>
<tr>
<th>Type</th>
<th>Number of Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operation 1_Normal</td>
<td>500</td>
</tr>
<tr>
<td>Operation 1_Fault</td>
<td>500</td>
</tr>
<tr>
<td>Operation 2_Normal</td>
<td>500</td>
</tr>
<tr>
<td>Operation 2_Fault</td>
<td>500</td>
</tr>
<tr>
<td>Operation 3_Normal</td>
<td>500</td>
</tr>
<tr>
<td>Operation 3_Fault</td>
<td>500</td>
</tr>
</tbody>
</table>

Table 6. Examples of the converted CWT-scalogram images in the experimental dataset.

<table>
<thead>
<tr>
<th>Type</th>
<th>Normal</th>
<th>Fault</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operation 1</td>
<td><img src="image1" alt="Example" /></td>
<td><img src="image2" alt="Example" /></td>
</tr>
<tr>
<td>Operation 2</td>
<td><img src="image3" alt="Example" /></td>
<td><img src="image4" alt="Example" /></td>
</tr>
<tr>
<td>Operation 3</td>
<td><img src="image5" alt="Example" /></td>
<td><img src="image6" alt="Example" /></td>
</tr>
</tbody>
</table>
5. Results and Discussion

5.1. Evaluation of the Proposed Method

The proposed methodology is verified by comparison with previous studies. Table 7 is a previous study to be compared and includes raw signal input, grayscale image input, and STFT-spectrogram input.

**Table 7.** Description of the comparison target.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Input Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area-metric-based sampling</td>
<td>CWT-scalogram (proposed method)</td>
</tr>
<tr>
<td></td>
<td>STFT-spectrogram</td>
</tr>
<tr>
<td></td>
<td>Grayscale image</td>
</tr>
<tr>
<td></td>
<td>1D raw signal</td>
</tr>
<tr>
<td>Random sampling</td>
<td>CWT-scalogram</td>
</tr>
<tr>
<td></td>
<td>STFT-spectrogram</td>
</tr>
<tr>
<td></td>
<td>Grayscale image</td>
</tr>
<tr>
<td></td>
<td>1D raw signal</td>
</tr>
</tbody>
</table>

The performance of the proposed method was evaluated using accuracy among the four evaluation indicators (precision, recall, F1-score, and accuracy) shown in Table 8. Precision is the proportion of actual true data among the data judged to be true by the model. Conversely, recall is the proportion of data judged to be true by the model out of the actual true data. The F1-score is the harmonic mean value of the precision and recall. Accuracy is the proportion of data that the model correctly predicted out of the total data.

**Table 8.** Evaluation indicators.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>( \frac{TP}{TP+FP} )</td>
</tr>
<tr>
<td>Recall</td>
<td>( \frac{TP}{TP+FN} )</td>
</tr>
<tr>
<td>F1-score</td>
<td>( \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} )</td>
</tr>
<tr>
<td>Accuracy</td>
<td>( \frac{TP+TN}{TP+TN+FP+FN} )</td>
</tr>
</tbody>
</table>

In Table 8, TP denotes the true positive, FP is the false positive, TN represents the true negative, and FN represents the false negative. Dilated CNN and traditional CNN models have an adaptive moment algorithm optimizer with a batch size of 32 set as the optimization algorithm for fast convergence. The execution rate was set to 0.001. For both models, a total of 100 training epochs were performed. Of the total data, the training data and test data were divided into 70 and 30%. Figure 11 shows a structural diagram of the CNN model. The proposed dilated CNN and the traditional CNN model had the same structure, but, in the case of the proposed dilated CNN model, a dilation rate = 2 was applied in the convolutional layer.

Table 9 summarizes the performance comparisons of the various methods. In the traditional CNN model, the raw signal input shows a classification accuracy of 48.89% when sampled with random sampling and 83.44% with area-metric-based sampling. The grayscale image input shows a classification accuracy of 58.78% when sampled with random sampling and 87.67% with area-metric-based sampling. The STFT-spectrogram input shows a classification accuracy of 68.89% when sampling with random sampling and 91.78% classification accuracy when sampling with area-metric-based sampling. The CWT-scalogram input shows an accuracy of 87.78% when sampling with random sampling. When sampling with area-metric-based sampling, the classification accuracy is 92.11%. When training with dilated-CNN, the classification accuracy for CWT-scalogram input is 88.33% for random sampling and 95.44% for area-metric sampling, confirming that the classification accuracy improved to 0.55 and 3.33%, respectively. The proposed method shows the highest accuracy among the comparison group. This means that the classification performance is the best.
Figure 11. Structure of the CNN model.

Figure 12 shows the confusion matrix of the dilated-CNN using the CWT-scalogram results for area-metric-based sampling and random sampling. Figure 13 shows the confusion matrix of the CNN using the CWT-scalogram results for the area-metric-based sampling and random sampling. Figure 14 shows the confusion matrix of the CNN using the STFT-spectrogram results for the area-metric-based sampling and random sampling. Figure 15 shows the confusion matrix of the CNN using the grayscale image results for area-metric-based sampling and random sampling. Figure 16 shows the confusion matrix of the CNN using 1D raw signal results for area-metric-based sampling and random sampling.
Table 9. The results of the various methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>Area-Metric-Based Sampling</th>
<th>Random Sampling</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Precision (%)</td>
<td>Recall (%)</td>
</tr>
<tr>
<td>Dilated-CNN using CWT-scalogram</td>
<td>95.57</td>
<td>95.46</td>
</tr>
<tr>
<td>CNN using CWT-scalogram</td>
<td>92.53</td>
<td>92.20</td>
</tr>
<tr>
<td>CNN using STFT-spectrogram</td>
<td>91.76</td>
<td>91.76</td>
</tr>
<tr>
<td>CNN using Grayscale image</td>
<td>87.80</td>
<td>87.81</td>
</tr>
<tr>
<td>CNN using 1D raw signal</td>
<td>83.33</td>
<td>83.52</td>
</tr>
</tbody>
</table>

Figure 12. Testing the confusion matrix of the dilated-CNN using CWT-scalogram: (a) area-metric-based sampling method confusion matrix (95.44% accuracy); (b) random sampling method confusion matrix (88.33% accuracy).

Figure 13. Testing the confusion matrix of the CNN using CWT-scalogram: (a) area-metric-based sampling method confusion matrix (92.11% accuracy); (b) random sampling method confusion matrix (87.78% accuracy).
Figure 14. Testing confusion matrix of the CNN using STFT-spectrogram: (a) area-metric-based sampling method confusion matrix (91.76% accuracy); (b) random sampling method confusion matrix (68.89% accuracy).

Figure 15. Testing confusion matrix of the CNN using grayscale image: (a) area-metric-based sampling method confusion matrix (87.67% accuracy); (b) random sampling method confusion matrix (58.78% accuracy).

Figure 16. Testing confusion matrix of the CNN using grayscale image: (a) area-metric-based sampling method confusion matrix (83.44% accuracy); (b) random sampling method confusion matrix (48.89% accuracy).
5.2. Presenting the Basis for Judgment: Based on Explainable Artificial Intelligence

Table 10 shows the results of the attention map for each state of the proposed method as a heat map based on Grad-CAM. Since the robot is controlled by repeating acceleration and deceleration instantaneously, the frequency changes over time. A CWT-based scalogram representing the change in frequency over time of nonstationary data is used for model training. The model is trained by weighting the time-varying frequency components that have physical meaning in the input image and deriving a result.

Table 10. Attention map of the proposed method for each condition.

<table>
<thead>
<tr>
<th></th>
<th>Input</th>
<th>Attention Map</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operation 1 Normal</td>
<td><img src="image1" alt="Input Image" /></td>
<td><img src="image2" alt="Attention Map" /></td>
</tr>
<tr>
<td>Operation 1 Fault</td>
<td><img src="image3" alt="Input Image" /></td>
<td><img src="image4" alt="Attention Map" /></td>
</tr>
<tr>
<td>Operation 2 Normal</td>
<td><img src="image5" alt="Input Image" /></td>
<td><img src="image6" alt="Attention Map" /></td>
</tr>
<tr>
<td>Operation 2 Fault</td>
<td><img src="image7" alt="Input Image" /></td>
<td><img src="image8" alt="Attention Map" /></td>
</tr>
<tr>
<td>Operation 3 Normal</td>
<td><img src="image9" alt="Input Image" /></td>
<td><img src="image10" alt="Attention Map" /></td>
</tr>
<tr>
<td>Operation 3 Fault</td>
<td><img src="image11" alt="Input Image" /></td>
<td><img src="image12" alt="Attention Map" /></td>
</tr>
</tbody>
</table>
Compared to the traditional CNN model, the dilated-CNN derives a classification accuracy improvement of up to 3.33%. Using Grad-CAM, the basis for judgment in each layer is visualized through a heat map to present the basis for improving the classification accuracy. The receptive field becomes more expansive in the convolutional layer of the dilated CNN because dilated convolution kernels are used. In the heat map using Grad-CAM, it can be observed that the dilated-CNN model had more active parts than the traditional CNN model in each convolutional layer. In addition, in a grayscale image in which signals are arranged, a judgment is performed on each layer without any special criteria. On the other hand, in the STFT-spectrogram and CWT-scalogram, which are time–frequency images, each layer performs a judgment by focusing on the frequency component of the image.

Unlike a STFT-spectrogram, a CWT-scalogram provides multiresolution for the time–frequency domains. Since the CWT-scalogram contains various time–frequency information compared to the STFT-spectrogram, it provides improved classification accuracy, because there are many areas where judgment weights are placed in each layer. Table 11 is the result of visualizing the weights in each convolutional layer as a heat map using the Grad-CAM.

**Table 11.** The basis for judgment in each layer visualized through a heat map.

<table>
<thead>
<tr>
<th>Input</th>
<th>Convolution Layer 1</th>
<th>Convolution Layer 2</th>
<th>Convolution Layer 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dilated-CNN using CWT-scalogram</td>
<td><img src="image1" alt="Heat Map" /></td>
<td><img src="image2" alt="Heat Map" /></td>
<td><img src="image3" alt="Heat Map" /></td>
</tr>
<tr>
<td>CNN using CWT-scalogram</td>
<td><img src="image4" alt="Heat Map" /></td>
<td><img src="image5" alt="Heat Map" /></td>
<td><img src="image6" alt="Heat Map" /></td>
</tr>
<tr>
<td>CNN using STFT-spectrogram</td>
<td><img src="image7" alt="Heat Map" /></td>
<td><img src="image8" alt="Heat Map" /></td>
<td><img src="image9" alt="Heat Map" /></td>
</tr>
<tr>
<td>CNN using Grayscale image</td>
<td><img src="image10" alt="Heat Map" /></td>
<td><img src="image11" alt="Heat Map" /></td>
<td><img src="image12" alt="Heat Map" /></td>
</tr>
</tbody>
</table>

6. Conclusions

In this research, a new data sampling method was developed to solve the problems of data imbalance and shortage. In addition, we introduced the judgment basis for deep learning-based defect diagnosis using XAI. The proposed method was verified using data
of nonstationary current signals from industrial robots operating under various operating conditions. The verification results are summarized as follow.

First, the data were extracted using area-metric-based sampling. The data were scanned sequentially for each operating condition, and the area metrics of the reference ECDF and segmented data ECDF were obtained. Second, the sampled data were transformed into CWT–scalogram images containing time–frequency information. A dilated CNN using dilated convolution kernels was introduced to increase the receptive field, which made it possible to accommodate a lot of information without increasing the number of computations and number of hyperparameters. Third, the proposed method was verified by performing robotic strain wave gear reducer fault diagnosis. The test classification accuracy, precision, recall, and F1-score were used as evaluation indices. As a result, the proposed method showed the highest accuracy among the comparison groups at 95.44%, which means that the classification performance is the best. For all comparison groups, area-metric-based sampling showed a higher test classification accuracy by at least 4.33% and up to 34.55% than the conventional random sampling. This indicates that the proposed area-based sampling can overcome the limitations of traditional sampling, and solve the problem of data imbalance and shortage. In addition, the CWT–scalogram showed a minimum of 0.33% and a maximum of 38.89% higher test classification accuracy compared to other input data under the same comparison criteria. This indicates that the CWT-scalogram has the least data loss compared to other data analysis methods. When a dilated CNN was used instead of the traditional CNN as a classification model, the test classification accuracy improved by at least 0.55% and, at most, 3.33%. Through Grad-CAM, the basis for judgment in the convolutional layer of the two models was presented using a heat map. Compared to the traditional CNN, the dilated CNN had more active parts, indicating that the dilated CNN was based on more feature information than the traditional CNN. The proposed method can derive high diagnostic accuracy even with a small amount of data, so it can be applied to industries where it is difficult to secure sufficient quality data. In addition, by presenting the basis for the diagnosis result, the reliability of the diagnosis model can be increased, and it can be widely applied to actual industrial sites. Future work will involve further exploring the adaptability and robustness of the proposed method across different industrial domains and expanding the dataset to encompass a broader range of operating conditions for a comprehensive validation of the approach. Additionally, investigating ensemble techniques and exploring real-time implementation possibilities will be essential steps to enhance the method’s practical applicability.

Author Contributions: Conceptualization, S.-K.C.; content preparation, Y.R.N. and S.K.; writing—original draft preparation, Y.R.N. and H.S.K.; writing—review and editing, S.K., H.S.K. and S.-K.C.; supervision, H.S.K. and S.-K.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the MOTIE (Ministry of Trade, Industry, and Energy) in Korea, under the Fostering Global Talents for Innovative Growth Program (P0017307) supervised by the Korea Institute for Advancement of Technology (KIAT).

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References


24. Qian, L.; Pan, Q.; Lv, Y.; Zhao, X. Fault Detection of Bearing by Resnet Classifier with Model-Based Data Augmentation. Machines 2022, 10, 521. [CrossRef]


**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.