

Article

Transfer Learning-Based Health Monitoring of Robotic Rotate Vector Reducer Under Variable Working Conditions

Muhammad Umar Elahi ¹, Izaz Raouf ¹, Salman Khalid ¹, Faraz Ahmad ² and Heung Soo Kim ^{1,*}

¹ Department of Mechanical, Robotics and Energy Engineering, Dongguk University–Seoul, 30 Pildong-ro 1-gil, Jung-gu, Seoul 04620, Republic of Korea; umarelahi@dgu.ac.kr (M.U.E.); izazraouf@dongguk.edu (I.R.); salmankhalid@dgu.ac.kr (S.K.)

² School of Mechanical Engineering, Purdue University, 585 Purdue Mall, West Lafayette, IN 47907, USA; ahmad97@purdue.edu

* Correspondence: heungsoo@dgu.edu; Tel.: +82-2-2260-8577; Fax: +82-2-2263-9379

Abstract: Due to their precision, compact size, and high torque transfer, Rotate vector (RV) reducers are becoming more popular in industrial robots. However, repetitive operations and varying speed conditions mean that these components are prone to mechanical failure. Therefore, it is important to develop effective health monitoring (HM) strategies. Traditional approaches for HM, including those using vibration and acoustic emission sensors, encounter such challenges as noise interference, data inconsistency, and high computational costs. Deep learning-based techniques, which use current electrical data embedded within industrial robots, address these issues, offering a more efficient solution. This research provides transfer learning (TL) models for the HM of RV reducers, which eliminate the need to train models from scratch. Fine-tuning pre-trained architectures on operational data for the three different reducers of health conditions, which are healthy, faulty, and faulty aged, improves fault classification across different motion profiles and variable speed conditions. Four TL models, EfficientNet, MobileNet, GoogleNet, and ResNET50v2, are considered. The classification accuracy and generalization capabilities of the suggested models were assessed across diverse circumstances, including low speed, high speed, and speed fluctuations. Compared to the other models, the proposed EfficientNet model showed the most promising results, achieving a testing accuracy and an F1-score of 98.33% each, which makes it best suited for the HM of robotic reducers.

Keywords: transfer learning; industrial robots; health monitoring; fault detection; rotate vector reducer



Academic Editor: Dan Zhang

Received: 7 December 2024

Revised: 14 January 2025

Accepted: 15 January 2025

Published: 16 January 2025

Citation: Elahi, M.U.; Raouf, I.; Khalid, S.; Ahmad, F.; Kim, H.S. Transfer Learning-Based Health Monitoring of Robotic Rotate Vector Reducer Under Variable Working Conditions. *Machines* **2025**, *13*, 60. <https://doi.org/10.3390/machines13010060>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In contemporary society, many industries are utilizing robust robots for manufacturing, which takes advantage of the advances of existing industries such as automotive manufacture, aerospace engineering, and heavy machinery production [1,2]. The systematic incorporation of robotic systems in production has markedly increased efficiency and improved customer satisfaction. Nevertheless, the ongoing operational demands placed on industrial robots often lead to component degradation, which may disrupt production and heighten the risk of system failure, potentially resulting in significant economic loss [3]. To address these issues, researchers have devised and executed diverse HM and diagnostic methodologies, which are commonly referred to as prognosis and health management (PHM) [4,5]. Comprehensive studies have increasingly focused on the PHM of industrial robots [6]. Due to its small design, efficient speed transmission, and high torque capability,

the Rotate vector (RV) reducer is extensively employed in industrial robotic systems [7,8]. Notwithstanding these significant benefits, the demanding operational conditions of elevated torque, impact forces, and sudden load fluctuations make the RV reducer vulnerable to possible failure [9,10]. Frequent failures of RV reducers encompass tooth surface erosion, pitting of the planetary gear, cracking, elevated backlash, bearing degradation, lubrication defect, and misalignment [11,12]. These failures may result in substantial operational problems, such as diminished accuracy, impaired robotic functionality, and unforeseen interruptions. In essential manufacturing operations, such disruptions may lead to postponed production timelines, heightened maintenance expenses, and diminished product quality. Furthermore, extended failures may induce cascading harm to interconnected robotic systems, leading to economic and operational consequences.

Robots are an essential part of modern industry and the production environment, and they reshape industries with increased precision, speed, and efficiency. They are applied in a range of industries, including automotive, electronics, aerospace, medicine, and logistics [13]. The implementation of robots in various industries increases production and ultimately fulfills customer needs. The operational timescales of robots range from hours to months, which increases the likelihood of defects in their components. Continuous operation of industrial robots often leads to failure in various components, impacting productivity and increasing the risk of operational blackouts, which can result in substantial financial loss. To identify and reduce major faults, researchers have proposed various health monitoring and diagnostic techniques under the umbrella of PHM [14,15]. Therefore, HM is generally assessed through the use of three principal diagnostic methods: ferrography analysis (FA) [16], vibration analysis (VA) [17], and acoustic emission analysis (AEA) [18]. Ferrography analysis examines particles mixed with lubricating oil to identify faults. One method uses neural networks to categorize particle types to diagnose wear modes, though this technique is limited to identifying defects in the internal structure. An alternative diagnostic technique, VA, entails the installation of vibration sensors on the reducer surface for identifying defects [19].

Data-driven techniques in VA, including frequency-domain analysis, have been established for early fault detection. Nevertheless, the main challenge with VA is the early detection of flaws as it is challenging to recognize issues like fracture initiation and propagation in their initial phases [20]. Alternatively, AEA mounts sensors on the robotic system to monitor acoustic emissions from the RV reducer and facilitate fault identification. While AEA can proficiently identify problems, it encounters constraints such as acoustic emission interferences, which result from variations in speed, and fluctuations in system temperature [21]. These strategies present distinct issues for fault identification in RV reducers. The requirement of supplementary sensors in both VA and AEA complicates the monitoring configuration, rendering it more cumbersome and intricate. Sensor installation is challenging in industrial settings with constrained space [22]. Furthermore, the implementation of several sensors throughout different robotic components mandates a comprehensive data-collecting system, hence complicating the design of the PHM system and elevating both computational and financial expenditures [23].

Recently, data-driven approaches are being used as an alternative to physics-based approaches due to the complexity of creating PHM models for different applications [24]. Data-driven approaches, such as Machine Learning (ML) and deep learning (DL), are widely used in the PHM of various applications based on the availability of data [25,26]. DL is effective in addressing the challenges of large-scale data processing, automatic feature extraction, and knowledge transfer across various operational contexts, including different machine types, motion types, and conditions [27]. Due to significant requirements for feature engineering and data processing, which restrict scalability and real-time application,

DL faces computational challenges. Traditional ML is effective in classification, regression, and clustering and is finding application in autonomous vehicles, fault detection, healthcare, and manufacturing [28]. Fault identification utilizing ML employs time-series data, alongside diverse analytical methods used to extract fault features, whose methods include time, frequency, and time–frequency analysis. These techniques are computationally intensive and necessitate expert knowledge, which, therefore, restricts their real-time application [29]. In general, robots are essential for enhancing precision, efficiency, and production in such industries as automotive, electronics, and logistics. The RV reducer is a crucial element that guarantees high torque transmission and accurate motion, which renders it indispensable for robotic functionality. Nonetheless, its failures, such as in gear erosion and misalignment, highlight the necessity for efficient health monitoring and fault diagnosis.

AI-driven fault identification plays a vital role in mechanical fault diagnosis, but acquiring adequate fault samples poses significant challenges, particularly in complex systems such as RV reducers. AI-driven approaches broadly include ML and DL, which enable systems to learn and make decisions from data. They also encompass other AI methods like expert systems and symbolic reasoning, highlighting their versatility. Previously, the study presented an improved deep discriminative transfer learning network (IDDTLN), which is optimized using a dynamic whale optimization technique. The evaluation of RV reducer data finds that IDDTLN attained a classification accuracy of 99.8%, demonstrating significant potential for fault diagnosis [26]. An adaptive transfer auto-encoder (ATAE), developed to tackle the issue of limited labeled data in the identification of defects in rotating machinery, improves fault classification through the integration of a data adaptation module, which utilizes kernel Hilbert space metrics, alongside variations of particle swarm optimization and k-nearest neighbors, for health status classification. In evaluations that utilize vibration data from bearing and gearbox rigs, ATAE demonstrates superior diagnostic accuracy, compared to conventional methods, while eliminating the requirement for target signal labels [30].

While data-driven AI models are essential for mechanical fault diagnosis, the scarcity of fault samples restricts their application in systems such as RV reducers. The proposed numerical simulation-enhanced multi-task integrated learning network (NSE–MTILN) optimizes RV reducer parameters through a whale optimization algorithm, uses a noise generative network (NGN) to generate fault samples, and employs ResNet for multi-task fault classification and prediction. Experiments indicate that NSE–MTILN effectively detects faults by utilizing solely simulated fault samples [31]. However, the fault diagnosis approach for RV reducer crankshaft bearings, which operate under variable-speed conditions and employ a white shark optimizer with variational mode decomposition (WSO–VMD) for signal reconstruction, enhances signal noise reduction and diagnostic accuracy, surpassing other optimization and classification models under both variable and constant-speed conditions [32]. AI-driven methodologies frequently necessitate large amounts of labeled data, which can be challenging and expensive to acquire. Moreover, these models may exhibit difficulties in adapting to novel surroundings or operational situations, and their computational complexity could impede real-time implementation. These constraints underscore the necessity for more efficient and adaptable methodologies such as transfer learning.

Building upon the existing literature, we propose, in this paper, the application of TL to enhance PHM systems for industrial robots, specifically focusing on RV reducer fault detection. TL is particularly advantageous in PHM because it allows models, which were trained on one set of conditions, such as specific machines, environments, or operational parameters, to be adapted for new but related tasks with minimal retraining [33]. Traditional HM methods face significant challenges, including a heavy reliance on large

amounts of labeled data and limited adaptability to new operating conditions or machines. TL approaches mitigate these restrictions by limiting reliance on labeled data, which allows models to use information from similar domains to enhance performance. Transfer learning improves generalization across various datasets and situations by matching feature distributions throughout domains, which renders it a viable solution for multiple health monitoring applications. This capability addresses a critical gap in traditional data-driven methods, where, when exposed to different conditions or machinery, model performance often diminishes due to the high variance in operational contexts.

Previously, a novel deep learning framework utilizes TL to optimize convolutional neural networks (CNNs) for the fault diagnosis of running mechanical equipment, which was coupled with support vector machines (SVMs) for enhanced classification performance. The method is validated through extensive experiments on seven mechanical datasets, showcasing its superior accuracy and generalization capabilities across diverse fault scenarios [34]. Further, the TL framework with Dynamic Multi-scale Representation (DMR) is used to address feature shifts under varying working conditions in fault diagnosis. The model uses a Multi-gate Mixture-of-Experts (MMoE) to balance scale-specific representation and inter-scale correlation and a TL backend for cross-domain feature alignment. It outperforms existing methods on PHM2009 and MCP datasets, improving accuracy by 4.71% and 3.86%, respectively [35]. For limited data, a deep Convolutional Transfer Learning Network (DCTLN) was implemented to address challenges in intelligent fault diagnosis, using TL to handle data distribution discrepancies across machines. The model combines a 1D CNN for feature learning and health condition recognition with a domain adaptation module to align cross-domain features [36]. Unlike traditional ML [37] and DL [38] methods, which require significant data from the target domain, TL optimizes pre-trained models to expedite training and improve accuracy in defect identification across diverse domains [39,40]. The reuse of acquired features, especially those that embody generic patterns in time-domain data in transfer learning, reduces the need for vast labeled datasets in each new context. This versatility renders it a potent strategy for sectors with restricted access to domain-specific data or where data annotation is laborious and expensive.

This study focuses on the implementation of transfer learning to diagnose faults of the robotic RV reducer by using limited data. Current data are obtained from the robotic RV reducer to validate the proposed methodology. Moreover, the raw 1D current signals are converted into 2D scalogram images to improve the feature extraction process. This process is conducted using continuous wavelet transform (CWT) analysis, where the time series data are converted to time–frequency images. In the present study, we consider various TL models, namely, ResNET50v2, MobileNet, GoogleNet, and EfficientNet, due to their robustness. These models are utilized because of their effectiveness in autonomously extracting deep features from scalogram images, thus eliminating the need for manual feature extraction. EfficientNet is identified as the optimal model based on its superior performance across evaluation metrics, including accuracy, precision, recall, and F1-score.

The primary contributions of this study are as follows. First, motor current signature analysis (MCSA) is employed, instead of traditional techniques such as vibration analysis (VA), acoustic emission analysis (AEA), and ferrography, offering a more efficient and practical approach to fault diagnosis. Embedded sensors are utilized for data collection, which offer a simpler and more integrated solution compared to the complexity of external sensors. Second, to overcome the limitations of handcrafted feature extraction, which is time-consuming and less suitable for generalization, TL models are utilized for autonomous feature extraction in the HM of industrial robots. Third, robust features for the HM of the robotic reducer are extracted using 2D scalogram images generated from motor current signals. Finally, the integration of transfer learning has enabled the development

of a generalized HM model for the RV reducer, ensuring improved diagnostic accuracy and adaptability in varying conditions. Therefore, the pre-trained TL models, which were trained on large source data, are fine-tuned on the robotic RV reducer target data. Numerous evaluation metrics of accuracy, precision, recall, and F1-score are used to validate the efficiency of the proposed approach. This novel approach represents a significant advancement over conventional fault diagnosis methods.

In the remainder of this paper, Section 2 demonstrates the methodology employed in the experimental investigation, while Section 3 describes the TL algorithms. Section 4 presents the results and their analysis, while Section 5 concludes the paper and suggests future work.

2. Methodology

This study introduces a methodology for HM in industrial robots by analyzing electric current data from the RV reducer under three health conditions: healthy, faulty, and faulty aged. The process involves preprocessing and synchronizing the electric current signals, followed by converting one-dimensional signals into two-dimensional scalograms to enhance feature extraction. Advanced transfer learning models, including EfficientNet, MobileNet, GoogleNet, and ResNet50v2, are applied for fault classification. Figure 1 provides a detailed overview of the suggested methodology, which was carried out on an experimental test bench using an industrial robot from Hyundai Robotics Co. in a laboratory setting. In the fault diagnosis of the RV reducer, two fault types are simulated on reducers with different degradation levels, specifically on joint 4 of the robot. These conditions are categorized into three distinct states: healthy, faulty, and faulty aged. Initially, multiple sensors record voltage data through data acquisition modules, where data are subsequently transformed into electric current signals using current–voltage sensor calibration. Our technique relies on current signals from the RV reducer to detect and then classify problems. To improve the diagnostic value of these signals, the 1D time-domain current signals are first converted into 2D scalogram images. These images illustrate three distinct states of healthy, faulty, and aged faulty situations, providing a comprehensive visual representation of the characteristics of each signal. Transformation of raw signals into scalograms produces a data format that is better suited to deep learning models, which are adept at identifying intricate visual patterns. Four transfer learning models are applied to the scalogram picture data of the RV reducer for various speeds ranging (10 to 100)%, encompassing the architectures of EfficientNet, MobileNet, GoogleNet, and ResNET50v2.

2.1. Experimental Details

Figure 2 depicts the Hyundai YS080 robot used in this study to implement and analyze faults in RV reducers. This six-degree-of-freedom (6-DOF) robot has a payload capacity of 80 kgf and is designed with six independent joints, each of which is driven by an electric motor. RV reducers are coupled with each motor to enable full 360° rotation around all axes. For this study, we focus on the RV reducer located at joint 4, where faults are intentionally introduced. The RV reducer is examined under three operational conditions: healthy, faulty, and aged faulty. The healthy state denotes a new, defect-free reducer, while the Faulty state denotes a reducer with an inherently produced low-severity fault. The faulty aged state simulates progressive wear and tear, which is modeled through an accelerated degradation test.

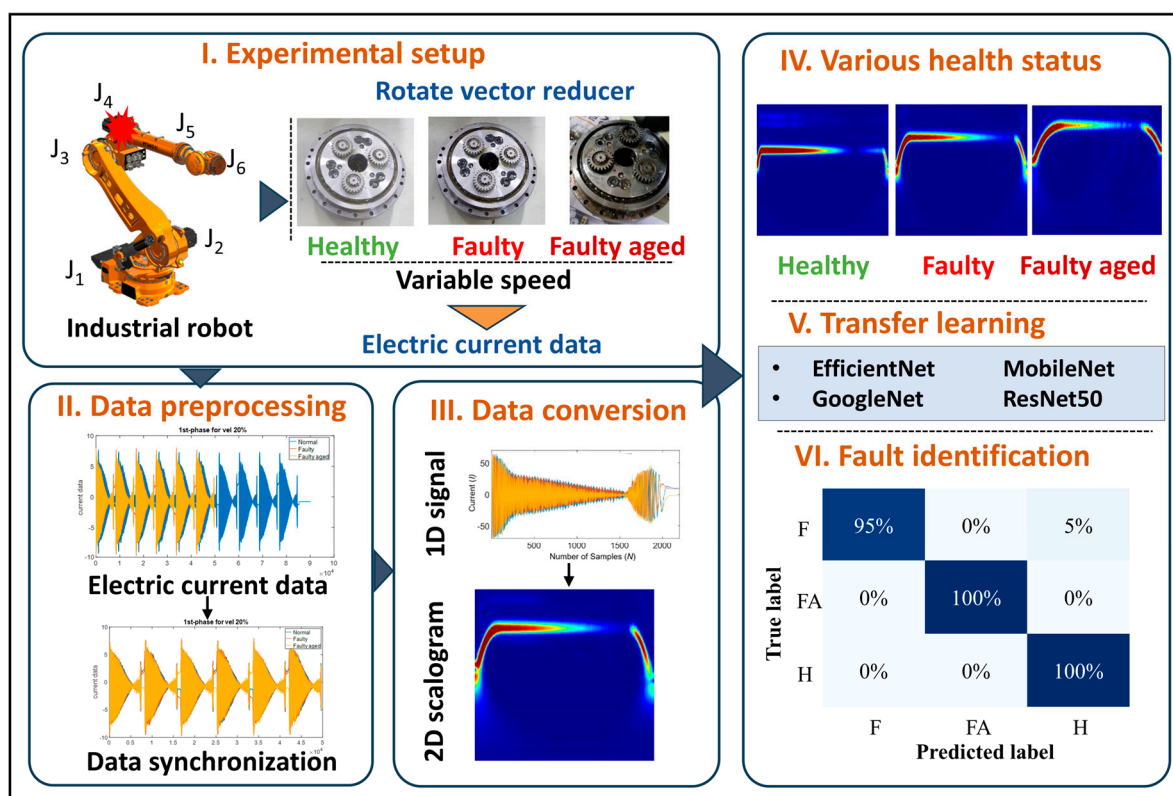


Figure 1. Proposed methodology of the identification of healthy, faulty and faulty aged state of the RV rotor robot using various TL models.

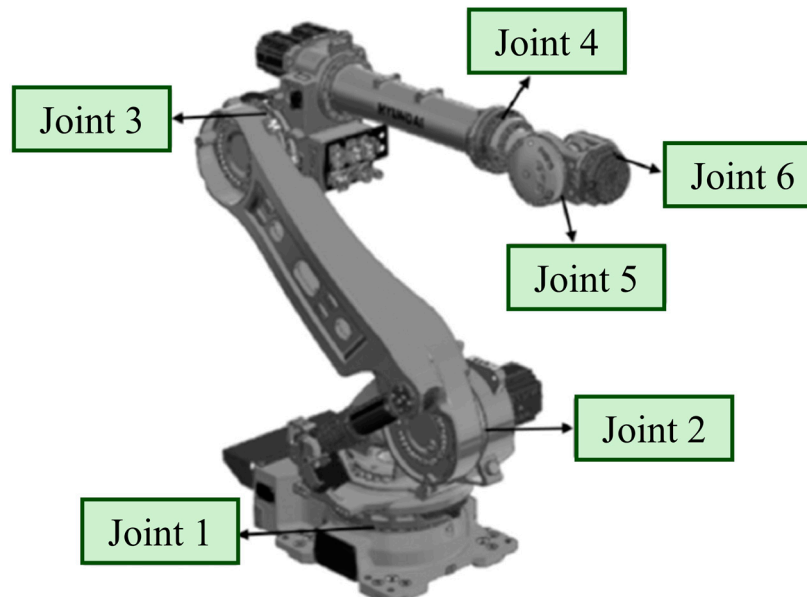


Figure 2. Illustrates the 6-degree-of-freedom industrial robot, highlighting each joint's pivotal role in achieving precise, multi-axis movements for advanced automation.

2.2. Data Collection Framework

Data collection is facilitated by equipping the robot with sensors on each of its six axes. WCS6800 Hall-effect sensors are used to record electric current data for MCSA analysis. These embedded sensors are attached to the first, second, and third arms of the robotic motors, as well as on all six axes. To comprehensively monitor the robot's performance, 18 sensors are installed to capture data across all phases of the fault induction process.

These sensors interface with NIDAQ 9230 data acquisition modules, with the acquired data being systematically stored on a PC for in-depth analysis. Figure 3 provides a schematic of data acquisition setup, which consists of a control unit, DAQ modules, a function generator for precise signal calibration, Hall-effect sensors for current measurement, and a PC for data processing. Figure 3 demonstrates the connectivity between these components that facilitates seamless data flow.

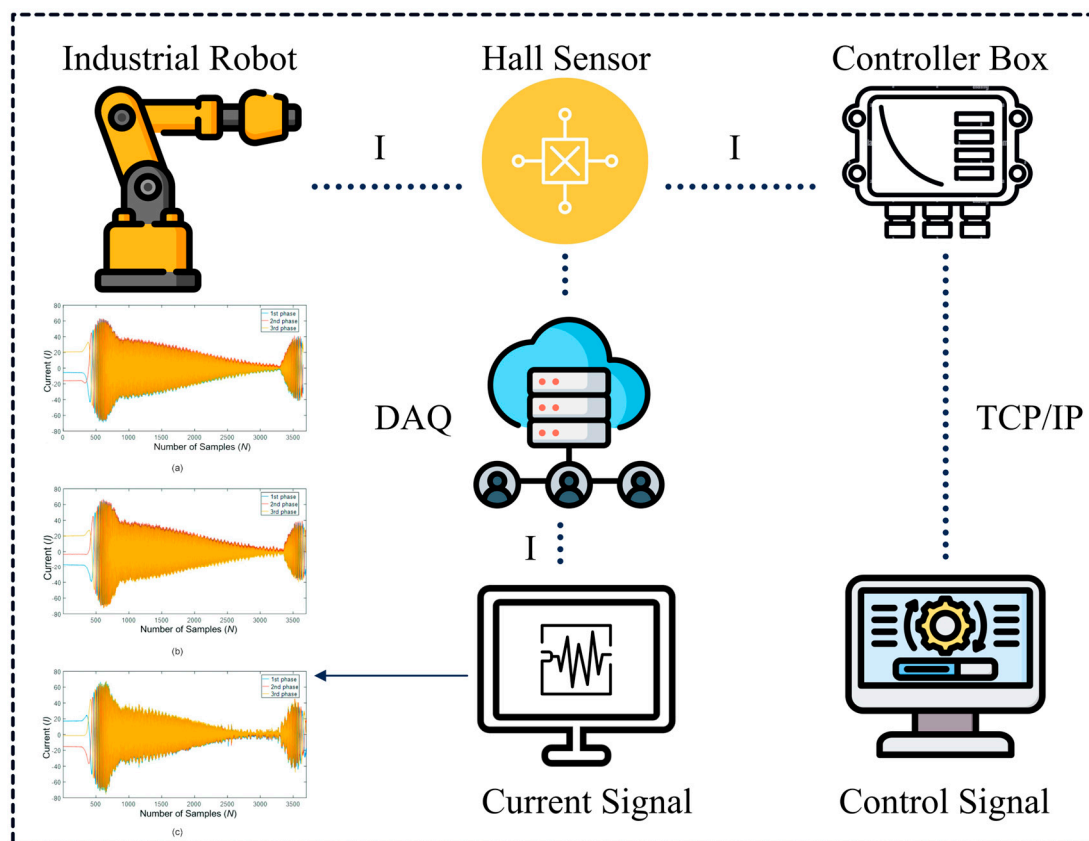


Figure 3. Schematic of the DAQ system, current signal (I) from sensor data capture through processing, storage, and output for analysis. The visual plots of the following current signals: (a) Healthy; (b) Faulty; (c) Faulty aged [41].

The data collection process involves recording signals from all 18 sensors for each reducer state (healthy, faulty, and faulty aged), while joint 4 operates in cyclic motion, as shown in Figure 3. The tests are conducted across a series of speed settings, ranging from 10 to 100%, with a maximum speed of $170^\circ/\text{s}$ in 10% increments, to capture a comprehensive dataset. Data are collected at a sampling frequency (f_s) of 12,800 Hz. Figure 1 depicts the Fault induction procedures, showing the healthy, faulty, and faulty aged reducers, respectively, where faults are induced manually by changing the RV reducers. This methodology ensures a comprehensive assessment of fault diagnostics across different fault severities and operational conditions using the four TL models.

2.3. Data Analysis

Data from all eighteen sensors were obtained. Despite the potential for mechanical coupling to induce patterns in the electrical current signals from different axes caused by a failure in one axis, data visualization and preliminary findings show that a fault in one joint does not influence the electric current patterns of the remaining axes. As a result, joint 4 data are primarily used for analysis, as each fault is investigated separately at its individual position. Data regarding joint 4 were gathered when the joint 4 motor

alone was operational while the other joints remained motionless. To maintain dataset consistency, data from the healthy, faulty, and faulty aged states were standardized. This method included all three phases of electric current (first, second, and third) and covered velocities ranging from 10 to 100%. Table 1 tabulates the properties of the collected data:

Table 1. Specifications of the recorded dataset [41].

Parameter	Description/Values
Sampling Frequency	12,800 Hz
Current phases	First phase, second phase, and third phase
Data types	Healthy, faulty, and faulty aged
Joint Motion	To and fro motion of joint 4
Speed	10 different speeds (10–100%)
No. of cycles	10 cycles

The 1D raw data of current signals are converted into scalogram images using the CWT. The CWT is a mathematical tool used to decompose a signal into time-frequency components, providing a multi-resolution analysis [42]. In this analysis, the Morlet wavelet, which is commonly used for generating scalograms, is utilized for the transformation. The CWT of a current signal $x(t)$ is defined as follows:

$$W(a, b) = \int_{-\infty}^{\infty} x(t) \psi_{a,b}^*(t) dt \quad (1)$$

where

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right) \quad (2)$$

Here, $x(t)$ is the input current signal, a is the scale parameter that determines the frequency resolution, b is the translation parameter that determines the time localization, ψ^* denotes the complex conjugate of the wavelet function ψ . The wavelet $\psi(t)$ is a localized oscillatory function with zero mean, enabling it to analyze signals with time-varying frequency content. Through this process, 1D raw signals were converted into scalogram images. Continuous Wavelet Transform (CWT) was chosen for its ability to provide superior time-frequency localization compared to other methods like Short-Time Fourier Transform (STFT) and Empirical Mode Decomposition (EMD). Unlike methods like STFT and EMD, CWT adapts to non-stationary signals and provides multi-scale analysis, capturing transient and localized features effectively. The flexibility of wavelet selection allows tailoring to domain-specific signal characteristics. This enhances model performance by offering richer feature representations, improved generalization, and reduced information loss, leading to higher accuracy and less dependency on extensive preprocessing [43,44].

3. Transfer Learning Algorithms

Transfer learning algorithms, including EfficientNet, MobileNet, GoogleNet, and ResNet50v2, are utilized in this study to address the challenges of fault classification and enhance the accuracy and efficiency of HM in industrial robots. Each model offers unique benefits in terms of classification efficiency and accuracy, specifically tailored to the challenges of defect detection. EfficientNet optimizes the equilibrium between performance and computational requirements, rendering it appropriate for diverse applications while preserving superior accuracy. MobileNet provides a lightweight, low-latency framework, which is ideal for real-time applications that require swift classification. GoogleNet employs inception modules to capture multi-scale features, enhancing its capacity to identify subtle variations in scalograms and differentiate between healthy and defective conditions. The

skip connections in ResNET50v2 mitigate the vanishing gradient problem, facilitating the more effective convergence of deeper network layers. This makes it ideal for collecting complicated signal fluctuations over time. TL has become an effective solution in deep learning for fault identification in robotic systems, especially when large datasets are scarce. By pre-training models on diverse datasets and leveraging acquired features for specific tasks, TL mitigates data limitations, making it ideal for robotics applications where comprehensive labeled data are difficult and expensive to obtain. Our study, which combines scalogram transformation and TL models, aims to achieve high fault classification accuracy, thus providing a robust and efficient way to use an RV reducer to monitor the health of industrial robotics.

This paper presents a TL methodology for the HM of RV reducers during variable-speed operations. The methodology consists of several essential steps: (a) Obtaining operational data from the experimental apparatus; (b) Converting raw current signals into scalogram images via CWT; (c) Implementing TL models of EfficientNet, MobileNet, GoogleNet, and ResNET50v2 on these scalogram images; (d) Determining the health states of the RV reducers. Initially, current signal data were collected from the RV reducers under three specific health conditions, healthy, faulty, and faulty aged, which correspond to different speed scenarios, as shown in Figure 4. The scalograms in Figure 4, derived using the CWT, show time-frequency plots with resolution of 224×224 pixels. The x -axis represents time; the y -axis indicates frequency (or scale), and color intensity reflects amplitude, with warmer colors (e.g., red) indicating higher amplitudes. In Figure 4a, the scalogram for the healthy condition shows a smooth and consistent frequency distribution, indicating no faults. In Figure 4b, the faulty condition displays noticeable disturbances in the frequency distribution caused by a fault in the RV reducer. In Figure 4c, the faulty aged condition reveals pronounced disturbances and irregularities, signifying advanced degradation.

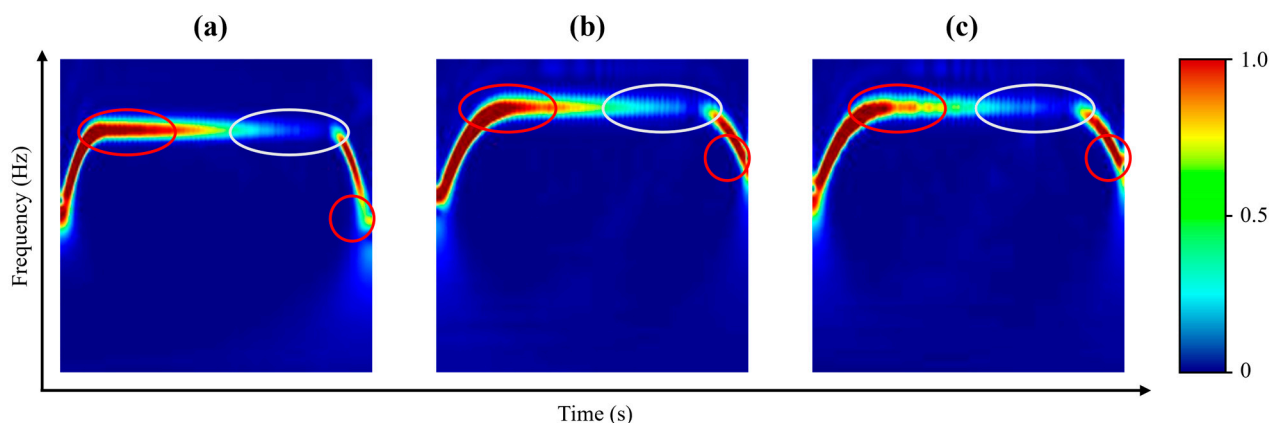


Figure 4. Scalogram images obtained after CWT analysis for the following states: (a) Healthy; (b) Faulty; (c) Faulty aged.

Several methodologies have been employed to improve the efficacy of TL in robotic fault detection. One method entails refining pre-trained models on extensive datasets, allowing these models to acquire general properties that are applicable to specific tasks with smaller datasets [45]. Furthermore, TL models gain advantages from employing architecture like convolutional neural networks (CNNs), which proficiently extract pertinent elements from operational data [46]. Automated machine learning (Auto-ML) methodologies have significantly advanced the application of TL by automating critical processes such as model selection, hyperparameter optimization, and model refinement. Through Auto-ML, pre-trained models are selected and fine-tuned based on the specific needs of a given task, reducing the need for expert intervention and ensuring efficient performance in

complex scenarios. This automation not only streamlines the workflow but also enhances the accuracy and adaptability of TL models, particularly in fault detection within robotics, where fluctuating environments and resource constraints are prevalent. The selection of models such as EfficientNet, MobileNet, ResNet50v2, and GoogleNet as transfer learning benchmarks reflects a balance between computational efficiency, model complexity, and proven performance in diverse applications [32,47–50]. EfficientNet achieves state-of-the-art accuracy with fewer parameters due to compound scaling, while MobileNet excels in resource-constrained environments with its lightweight design. Previous studies propose a deep learning-based industrial machine malfunction detection model using acoustic signals, where a weighted ensemble of EfficientNet-B0, B5, and B7 outperforms individual classifiers and other ensemble methods on the MIMII dataset [47]. Wang et al. introduce a lightweight, intelligent fault diagnosis model using multi-sensor data fusion for fault detection in roll bearings. The ultra-lightweight GoogleNet model not only fulfills the “small, light, and fast” requirements for lightweight systems but also achieves high-accuracy fault diagnosis even in noisy environments [50]. Kumar et al. proposed a TL-based method for bearing fault detection, using the pre-trained ResNetV2 model to autonomously extract features. By converting 1D data to 2D, manual feature extraction and selection are eliminated, enabling effective detection of bearing faults such as outer race, inner race, and ball defects, with performance metrics confirming the approach’s efficacy [51]. ResNet50v2, with its residual connections, improves gradient flow and enables the training of deeper networks, making it a robust choice for complex tasks. GoogleNet employs inception modules for multi-scale feature extraction and reduced computational overhead, making it efficient yet powerful [52]. Each model’s distinct strengths, including computational complexity, parameter size, and demonstrated success in case studies, justify their inclusion as benchmarks.

EfficientNet models are advancements in lightweight transfer learning, utilizing Auto-ML approaches to optimize model scaling. By adjusting the network’s depth, breadth, and resolution, they reduce parameters while enhancing performance, overcoming challenges like fragmentation and memory access issues. This makes EfficientNet models well-suited for real-time fault detection in robotics. Researchers have successfully deployed TL-based models to detect defects in robotic systems across diverse operational scenarios. Pre-trained models allow for quick adaptation to new tasks with minimal data, boosting system reliability and performance. As the field evolves, TL is poised to play an increasingly vital role in refining fault detection and enabling more intelligent, efficient robotic systems.

TL models are trained on scalogram images, enabling them to identify and respond to the unique features associated with each health state. This approach optimizes the CNNs in the architectures of EfficientNet, MobileNet, GoogleNet, and ResNet50v2 to extract hierarchical features from visual inputs. These architectures differ in design, efficiency, and suitability for various applications: EfficientNet excels in balancing accuracy and computational cost through compound scaling, making it ideal for resource-constrained environments; MobileNet is optimized for lightweight, high-performance models on mobile and edge devices; GoogleNet incorporates inception modules to capture multi-scale features, offering a balance between depth and computational efficiency; and ResNet50v2, with its residual connections, mitigates the vanishing gradient problem, allowing for deeper networks without sacrificing performance [53]. The trained TL model can be used to predict and determine the different health statuses of the rotary vector reducers, which, in our study, are limited to healthy, faulty, and faulty aged. The performance of the model is evaluated by a variety of metrics, including a confusion matrix, which provides insights into the accuracy, as well as the validity of health state predictions. This section highlights the attributes of EfficientNet, MobileNet, GoogleNet, and ResNet50v2, emphasizing their

strengths in enhancing condition health monitoring for robotic RV reducers under diverse operational conditions.

3.1. Convolutional Neural Network (CNN)

Convolutional Neural Networks (CNNs) are a class of deep learning models specifically designed to process data with grid-like topology, including images. Inspired by the human visual system, CNNs are widely utilized in image identification applications. The CNNs process starts by producing feature maps of specified areas of source data through the convolution of kernels or filters of diverse dimensioning. This convolution includes determining the dot product between a chosen segment of the source data and a parameterized weighted matrix enhanced by biases. A standard CNN design comprises multiple layers that include convolutional, pooling, flattening, dense, and classification layers [54]. The convolutional layers detect local features from input images, such as edges or textures, by applying filters. Pooling layers reduce the spatial dimensions of feature maps, decreasing computational complexity while preserving important information. The flattening layer converts the 2D feature maps into a 1D vector, enabling the output to be processed by subsequent layers. Dense layers then combine the extracted features, allowing the network to learn complex patterns. Finally, the classification layer produces the output by assigning probabilities to each possible health state based on the learned features [55]. This study emphasizes the critical role of the convolutional layer, which performs operations on scalogram regions to generate feature maps, which aggregate critical characteristics from the input data and enable the model to identify intricate patterns related to the various health statuses of the rotary vector reducers. Recent literature substantiates the efficacy of CNNs in the diverse conditions of monitoring applications, illustrating their ability to precisely classify and diagnose fault conditions across distinct machines. By optimizing the strengths of CNNs and TL, this study aims to improve the predictive maintenance capacities of rotary vector reducers at varying operational speeds [56]:

$$X_{m'}^i = \mathcal{C} \left(\sum_{m=1}^M W_{mm'}^i * X_m^{(i-1)} + B_m^i \right) \quad (3)$$

where W denotes the weight matrix, $*$ the convolution operation, B the filter bias, and $\mathcal{C}$ the activation function. Furthermore, m and m' represent the indices of the input and output feature maps, respectively.

3.2. EfficientNet Model

EfficientNet is a family of convolutional neural networks introduced by Tan et al. [57], designed to achieve a high level of accuracy while being computationally efficient. EfficientNet models scale uniformly across depth, width, and resolution using a compound scaling method, unlike traditional models that scale one dimension at a time. The scaling method optimizes each of these three dimensions to achieve a better trade-off between accuracy and efficiency. EfficientNet has become a prominent model in DL, especially for image classification applications that utilize the ImageNet dataset [57,58]. The primary aim of EfficientNet is to create more efficient models with fewer parameters while maintaining good accuracy. This is achieved by the technique known as compound scaling; this uniformly modifies the depth, width, and resolution of the network, resulting in improved performance while avoiding a proportional increase in processing requirements. The target network is developed by augmenting the baseline network through a compound scaling method that adjusts various parameters of the model according to specified resource limitations. Moreover, the architecture uses depth-wise separable convolutions, reducing processing requirements by a factor of k^2 , where k is the kernel size. The compound scaling

approach consistently adjusts the depth (d), width (ω), and resolution (r) of the model by means of a scaling coefficient ψ . This procedure is directed by constants α , β , and γ , which are refined by grid search within established resource limitations and guarantee effective use of computational resources [57]. The scaling equations are expressed as follows:

$$\text{depth} : d = \alpha^\psi \quad (4)$$

$$\omega = \beta^\psi \quad (5)$$

$$r = \gamma^\psi \quad (6)$$

$$\alpha, \beta, \gamma \geq 1 \quad (7)$$

These equations enable EfficientNet to maximize the balance between model complexity and performance, rendering this method especially appropriate for diverse applications, such as fault detection in robotic systems. The advantages of EfficientNet allow researchers to create resilient models that excel, notwithstanding the constrained computational resources, hence increasing their relevance in practical applications. EfficientNet has demonstrated efficacy as a model to identify flaws in robotic systems, utilizing its sophisticated architecture and effective scaling techniques. The constants α , β , and γ in the EfficientNet model govern the allocation of additional resources across the network's dimensions, while the coefficient ψ controls the overall scaling of the model. After establishing these parameters, the baseline network is scaled by adjusting the ψ value, resulting in larger models and allowing EfficientNet to improve its performance.

In this research, the EfficientNet model is utilized for fault detection in robots. However, the performance of EfficientNet is compared against three other transfer learning models, MobileNet, GoogleNet, and ResNet50v2, to assess the most effective architecture for this task. The EfficientNet model architecture, illustrated in Figure 5, emphasizes the arrangement of mobile inverted bottleneck convolution (MB-Conv) layers, with the output feature map dimensions indicated at each layer. At first, the layers of the EfficientNet model undergo pre-training with the ImageNet dataset, facilitating the acquisition of broad features, prior to fine-tuning for specific applications. Modification of the pre-trained model for fault detection eliminates the final classification layer and incorporates three new layers: a global average pooling (GAP) layer, a dense layer, and one additional dense layer for classification. The GAP layer condenses each feature map of the EfficientNet layers into a singular value via averaging, thereby effectively summarizing the features. The three newly added layers are trainable, enabling them to assimilate the distinct attributes of the vibration data linked to faulty conditions in robots. Simultaneously, to maintain the acquired features from the pre-training phase, the weights of the preceding layers remain static throughout training. This methodology allows the EfficientNet model to proficiently identify defects in robotic systems, even under conditions of restricted data availability. The TL approach employs pre-trained knowledge to discern small patterns suggestive of problems by fine-tuning the model on the vibration scalograms of robotic components. The lightweight design of EfficientNet facilitates its deployment in real-time applications, where it offers an effective solution for predictive maintenance and defect detection in diverse robotic systems. The optimized architecture and novel scaling methodologies of EfficientNet facilitate its ability to detect flaws in robots, rendering it a formidable instrument to improve the dependability and efficacy of robotic systems in diverse practical applications.

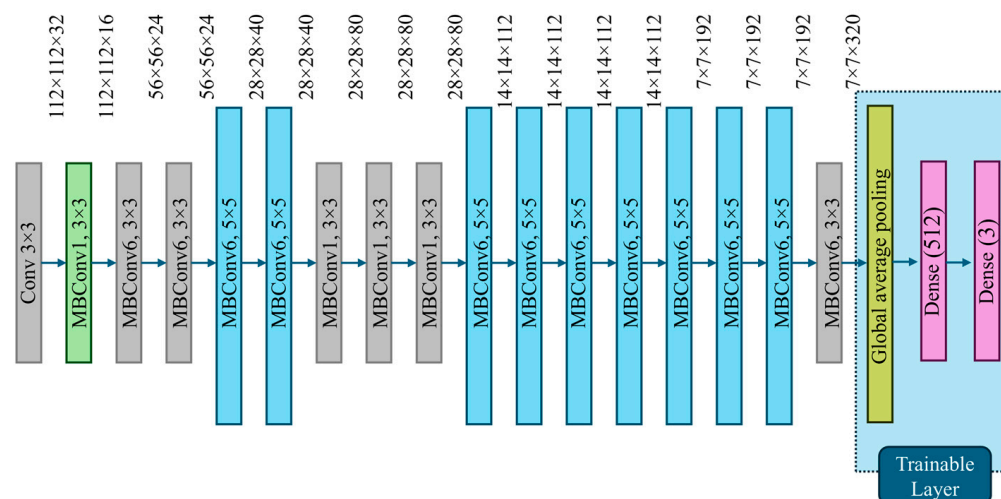


Figure 5. The architecture of the EfficientNet model.

3.3. GoogleNet

GoogleNet is a deep CNN architecture that innovatively uses the “Inception module”. Instead of using a simple stacking of convolutional layers, the Inception module applies multiple convolution operations with different kernel sizes (e.g., 1×1 , 3×3 , 5×5) in parallel, then concatenates the results. This allows the network to capture different levels of feature abstraction in parallel, leading to improved model performance. GoogleNet architecture is known for its depth, comprising 22 layers, and its use of global average pooling instead of fully connected layers to reduce the number of parameters. Convolutional, pooling, and fully connected layers are all integrated into the 22-layer deep CNN model known as GoogleNet, which, in comparison to LeNet and AlexNet, effectively lowers computing loads. Additionally, it has fewer features and is lighter than VGG, which results in quicker model training and fewer recognition failures. A significant advance in GoogleNet is its use of “inception modules”, which are intended to capture features at many scales while lowering the number of parameters. The main goal of the inception module is to expand the breadth and depth of the network without appreciably raising its computing cost. It includes a 3×3 pooling operation and four parallel convolutional layers with varying kernel sizes of 1×1 , 3×3 , and 5×5 . The outputs of these layers are concatenated. The inception module collects characteristics at multiple scales by employing different kernel sizes in parallel; meanwhile, pooling techniques lower the dimensionality of the output, as illustrated in Figure 6. These modules perform local feature extraction and multi-scale convolutions. The combination of the outputs of many convolutional layers and aggregation of the visual input at different sizes enables the fusion of features across scales. To make feature extraction at various sizes easier, dimensionality reduction is used on larger matrices. Because of their high number of parameters, added processing complexity, and overfitting risk, GoogleNet avoids the three fully linked layer design used by networks like VGG, LeNet, and AlexNet. Instead, the addition of average pooling and dropout units immediately after the inception modules lessens overfitting and decreases dimensionality.

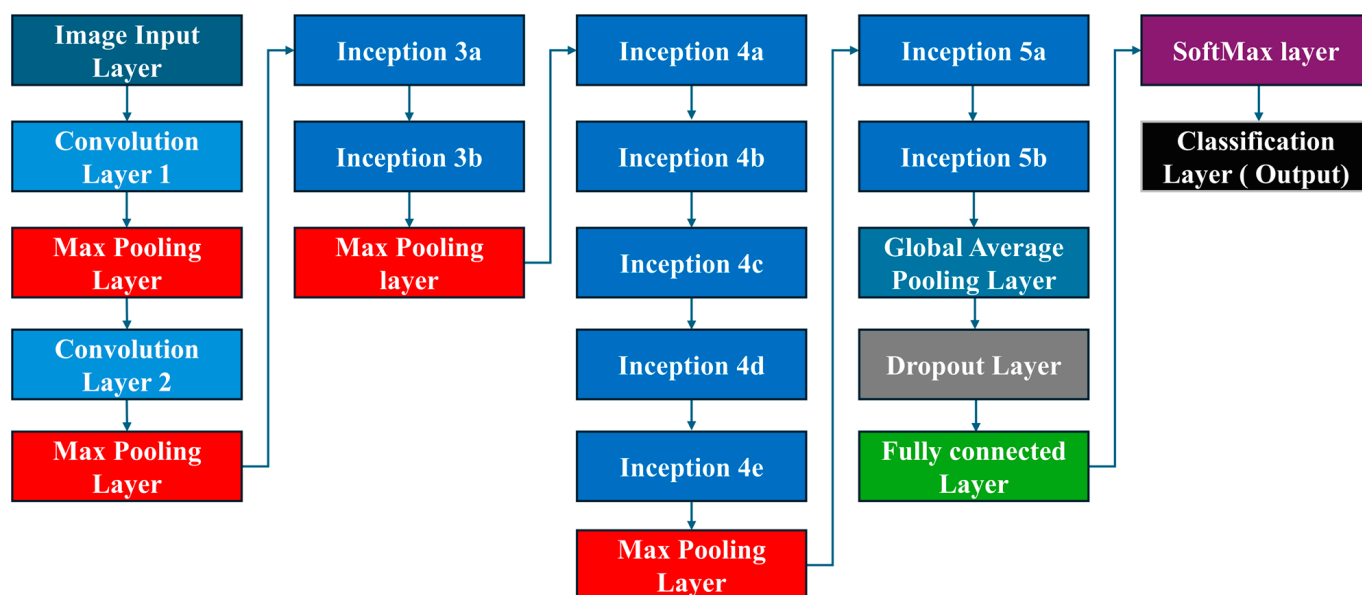


Figure 6. The architecture of the implemented GoogleNet-based transfer learning model.

3.4. MobileNet

MobileNet is a family of CNN architectures optimized for mobile and embedded vision applications. MobileNet is designed to run efficiently on devices with limited computational power, such as smartphones. The key innovation in MobileNet is the use of depth-wise separable convolutions, which split the convolution operation into two steps: a depth-wise convolution, which applies a single filter to each input channel, and a pointwise convolution, which combines the output channels. This reduces the computational cost compared to traditional convolutions. MobileNet also introduces a width multiplier and a resolution multiplier, which allows the user to trade off accuracy for computational efficiency, making it highly customizable for real-time applications [59]. MobileNet is a series of deep transfer learning (DTL) models, which have been developed for efficient use on mobile and embedded devices and are characterized by low computational and memory needs. Their architecture employs depth-wise separable convolutions (DSCs) to decrease computational demands and lower the parameter count to that needed for classification and image recognition applications. A conventional convolution in Depth-wise Separable Convolution (DSC) is divided into two phases: a depth-wise convolution independently filters each input channel, while a pointwise convolution integrates these filtered outputs through 1×1 convolutions. This decomposition reduces both the parameter count and computational cost and hence enhances the efficiency of the network.

Furthermore, MobileNet improves computational efficiency by utilizing bottleneck layers that condense input feature maps prior to their processing through convolutional layers, as shown in Figure 7. The MobileNet family of MobileNet, MobileNetV2, and MobileNetV3 features enhanced architectural innovations, including inverted residual blocks and linear bottlenecks. These innovations improve performance across diverse applications while maintaining minimal computing demands.

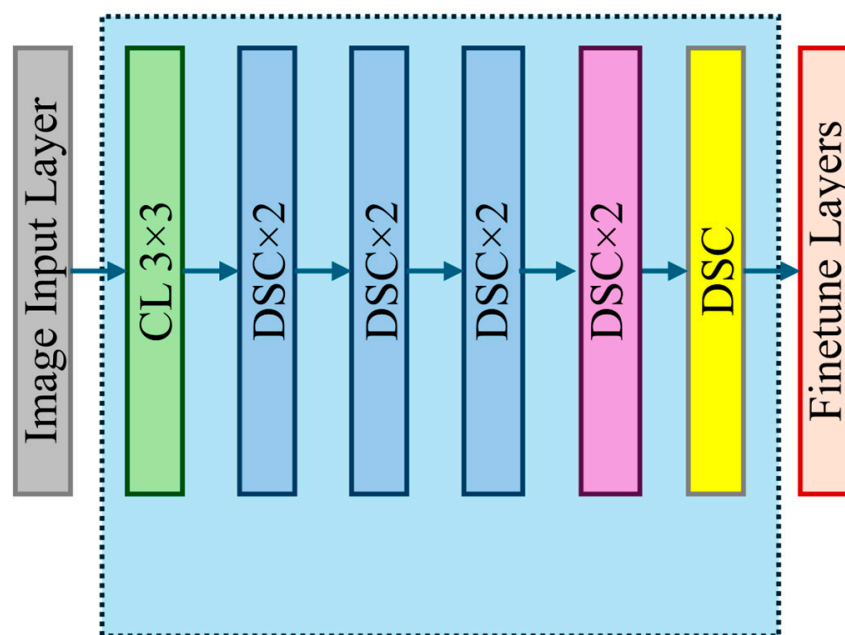


Figure 7. Schematic of the MobileNet-based DTL model. CL, Convolution layer; DSC, Depth-wise separable convolution.

3.5. ResNET50v2

ResNET50v2 is a modified iteration of the ResNet-50 model aimed at enhancing training stability and convergence through the optimization of residual block architecture. This architecture has 50 layers, and by adopting a pre-activation design, it improves the original ResNet framework by performing batch normalization and ReLU activation prior to each convolution operation. This enhancement optimizes gradient flow, so it is more efficient at addressing the vanishing gradient problem. The model employs bottleneck layers within its residual blocks to reduce the number of parameters and computational complexity, hence preserving efficiency without sacrificing performance. The design consists of two main block types: Residual Block and Identity Block. The Residual Block, referred to as Block-1, is utilized when there is a mismatch between input and output dimensions, and it is composed of three convolutional layers, 1×1 , 3×3 , and 1×1 , which are accompanied by a projection shortcut link that synchronizes the input to the output dimensions. This enables the model to obtain residual mappings while resolving dimensional differences. In contrast, the Identity Block, or Block-2, is employed when the input and output dimensions are congruent. This block exhibits a similar design of three convolutional layers, employing a simple identity shortcut that immediately adds the input to the output without requiring dimensional alterations. The blocks are arranged in layers with progressively larger filter sizes, allowing ResNET50v2 to effectively extract hierarchical features, rendering it ideal for image classification and object recognition tasks [60]. The ResNET50v2 model does not require a large amount of data. Figure 8 illustrates the whole network architecture for RN50V2:

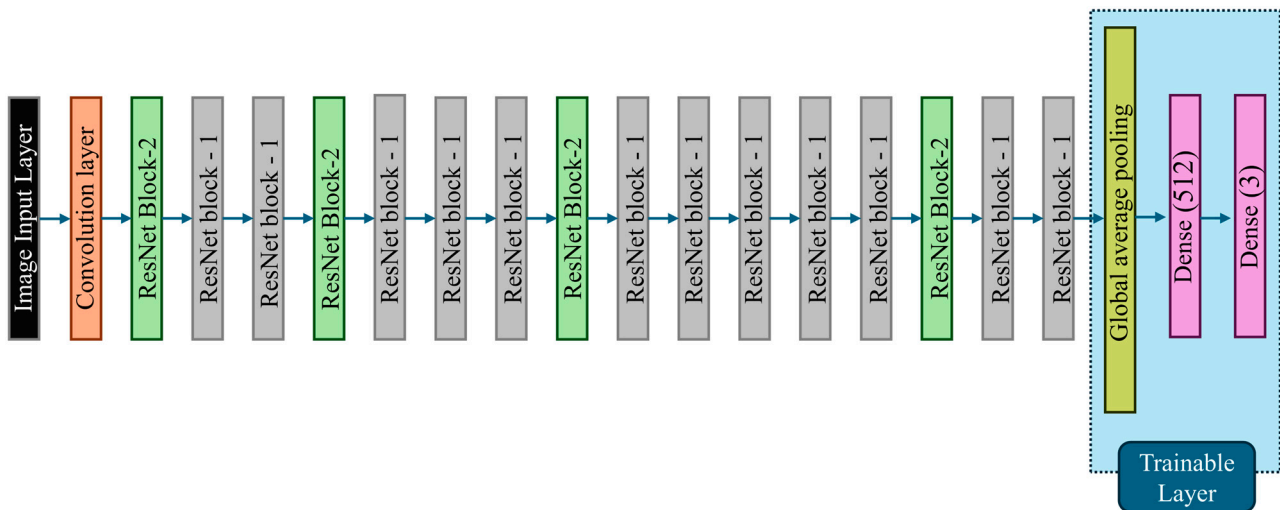


Figure 8. Schematic of the architecture of the ResNET50v2-based TL model.

3.6. Performance Metrics

Various performance criteria of the accuracy, confusion matrices, precision, recall, and F1-score were used to assess the efficiency of the TL models for fault detection. Accuracy measures the proportion of correct predictions in the dataset. Precision indicates the proportion of true positives among all predicted positives, while recall measures the proportion of actual positives correctly identified by the model. The F1-score is the harmonic mean of precision and recall, providing a balanced measure when both are in tension, especially in imbalanced datasets. Values for accuracy, precision, recall, and F1-score range from 0 to 1, where 0 represents the worst performance and 1 represents perfect performance. A value of 1 for accuracy means all predictions are correct, for precision, it indicates no false positives, and for recall, it means all actual positives are identified. The F1-score of 1 reflects a perfect balance between precision and recall. Higher values generally indicate better model performance, with 1 being the ideal [61]. These metrics offer a thorough evaluation of the capacity of the model to discern health conditions in robotic systems. The mathematical formulae to calculate these metrics, utilizing true positive (TP), true negative (TN), false positive (FP), and false negative (FN) predictions, are shown in Equations (6)–(9) [62]:

$$\text{Accuracy} = \frac{T_p + T_n}{T_p + T_n + F_n + F_p} \quad (8)$$

$$\text{Precision} = \frac{T_p}{T_p + F_p} \quad (9)$$

$$\text{Recall} = \frac{TP}{T_p + F_n} \quad (10)$$

$$F1\text{-score} = \frac{2T_p}{2T_p + F_n + F_p} \quad (11)$$

True Positive (TP) denotes the accurate identification of a healthy state, while True Negative (TN) denotes the accurate identification of an unhealthy (faulty or aged faulty) state. False Positive (FP) denotes the erroneous detection of a healthy condition that is unhealthy, while False Negative (FN) denotes the erroneous detection of an unhealthy condition that is in reality healthy. The confusion matrix offers a depiction of true and expected outputs, facilitating the assessment of the dependability and efficacy of the models in accurately identifying defects in industrial robots.

4. Results

The TL models of MobileNet, ResNET50v2, GoogleNet, and EfficientNet were implemented on the data. The dataset of scalogram images was trained for 100 epochs with the Adam optimizer at a learning rate of 1×10^{-3} . To introduce non-linearity in the models, ReLU activation was used in all dense layers except the last dense layer (classification later), which used SoftMax activation. Moreover, the multi-class problem resulted in the categorical cross-entropy loss function being utilized during model compilation. The dataset was split into three subsets: training (60%), validation (20%), and testing (20%). Each subset maintained an equal distribution across the three classes: faulty aged (FA), faulty (F), and healthy (H). The dataset was split to maintain class balance in all subsets, ensuring robust model evaluation. Cross-validation was not conducted in this study; instead, the test dataset was selected to represent diverse operating conditions of the RV reducer robot, capturing the typical variations in data. The validation dataset was used to fine-tune the model during training, while the test dataset served as an independent set to evaluate the model's generalization performance.

Figure 9 includes the training and validation accuracy curves over 100 epochs for each of the following transfer learning models: (a) MobileNet; (b) ResNet50v2; (c) GoogleNet; (d) EfficientNet. These curves provide insights into the model's performance during training, highlighting how well each model generalizes to validation data. From the accuracy curves, it is evident that the MobileNet model exhibits signs of overfitting, as the training accuracy curve rises higher than the validation accuracy curve, especially in the later epochs. In comparison, the ResNet50v2 and GoogleNet models display less overfitting than MobileNet, with their training and validation accuracy curves being more closely aligned. However, both ResNet50v2 and GoogleNet exhibit noticeable fluctuations in accuracy, indicating some instability during the training process. EfficientNet, on the other hand, shows the least overfitting among all models, with consistently smooth and closely aligned training and validation accuracy curves, indicating robust generalization and stable learning. This performance underscores EfficientNet's superior ability to handle the dataset effectively, achieving high accuracy in predictions while maintaining balanced metrics. For testing accuracy, MobileNet achieves 88.33%, indicating overfitting despite its high training accuracy. ResNet50v2 and GoogleNet perform better, with testing accuracies of 91.67% and 95.00%, respectively, showing stronger generalization. EfficientNet outperforms all models, achieving the highest testing accuracy of 98.33%, particularly after fine-tuning. As a result, EfficientNet emerges as the best-performing model in this study, demonstrating exceptional accuracy, precision, recall, and F1-score, making it an ideal choice for fault diagnosis in robotic RV reducers. The performance confirms its ability to manage the dataset with minimal classification errors, yielding results that are superior to the other models.

Figure 10 compares the training accuracy, validation accuracy, and testing accuracy of the implemented MobileNet, ResNET50v2, GoogleNet, and EfficientNet models for the fault diagnosis of the robotic RV reducer. Although MobileNet provides 100.00% training accuracy, its validation and testing accuracy are (92.98 and 88.33)%, respectively, showing overfitting results. However, it shows learning ability regarding training data, which indicates the struggle to generalize the new data. ResNET50v2 at 98.83% shows lower accuracy compared to MobileNet, with validation and test accuracy of (94.74 and 91.67)%, respectively, indicating strong generalization without overfitting or underfitting. However, GoogleNet achieves similar results of training and validation accuracy to ResNET50v2, which are (98.83 and 94.74)%, respectively, but higher test accuracy at 95.00%, showing high performance. The ablation study in Table 2 highlights the impact of fine-tuning on the performance of EfficientNet. Without fine-tuning, the model achieves an accuracy of 93.33%, with precision, recall, and F1-score values around 93%. However, after applying fine-tuning,

the model significantly improves, reaching an accuracy of 98.33%, with corresponding increases in precision, recall, and F1-score, demonstrating the substantial contribution of fine-tuning to overall system performance.

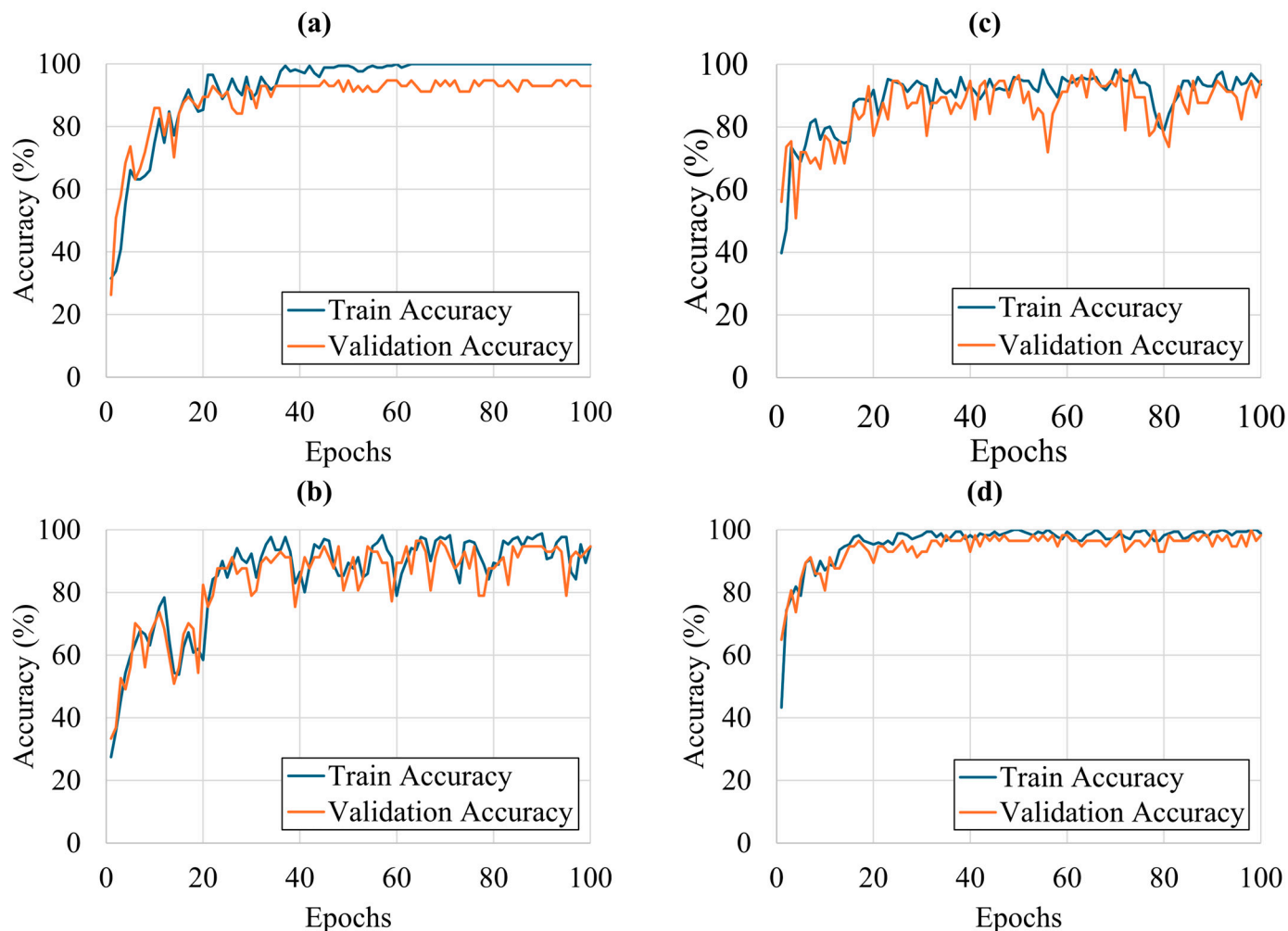


Figure 9. The training and validation accuracies of the following: (a) MobileNet; (b) ResNet50v2; (c) GoogleNet; (d) EfficientNet.

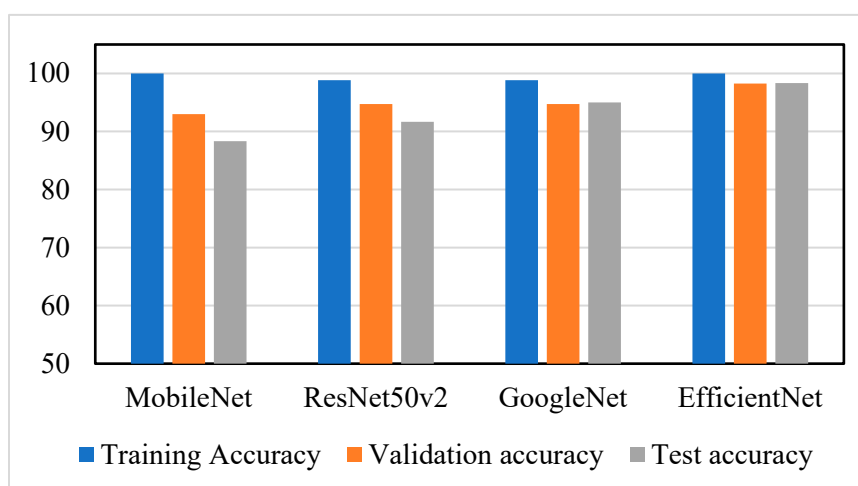


Figure 10. Training, validation, and testing accuracies of the proposed models.

Table 2. Ablation study for proposed EfficientNet model.

Model	Accuracy	Precision	Recall	F1-Score
EfficientNet without trainable layer	93.33%	93.60%	93.33%	93.32%
EfficientNet with trainable layer	98.33%	98.41%	98.33%	98.33%

EfficientNet Model shows 100.00% training accuracy and (98.25 and 98.33)% validation and test accuracy, respectively; these results are higher than those of the previous models. Due to its optimized structure, which is designed for good efficiency, the EfficientNet model maintains high accuracy, matching or slightly surpassing the other models. Therefore, the EfficientNet model is the best choice for these kinds of data in that they can reliably perform better on seen and unseen data and do so when computational efficiency is a priority. In scenarios where generalization and resource utilization are equally critical, EfficientNet emerges as a highly proficient and efficient choice.

The results highlight the performance of the four TL models MobileNet, ResNET50v2, GoogleNet, and EfficientNet for the CM of RV reducers in industrial robots. EfficientNet emerges as the most robust model, achieving the highest test accuracy of 98.33%, with an outstanding F1-score of 98.33%. The high precision, recall, and F1-scores of this model across all classes indicate a reliable classification of Healthy, Faulty, and Faulty Aged reducers. However, the slightly higher evaluation and prediction times of (2.32 and 7) s, respectively, suggest a trade-off between computational cost and superior fault-detection capabilities. GoogleNet, while faster in evaluation (0.63 s) and prediction (3 s), achieved a slightly lower accuracy of 95.00%. Its F1-score of 94.96% indicates consistent performance, though class-specific recall for Class 0 (healthy reducers) was slightly reduced (90.00%). For GoogleNet, the training time is approximately 70 min for 100 epochs, with a testing time of 0.63 s. ResNet50v2 has a training time of around 60 min (estimated based on step times), with a testing time of 0.87 s. MobileNet requires about 6 min for training and 0.63 s for testing. In contrast, EfficientNet, while achieving the highest accuracy, requires a significant computational overhead, with a training time of approximately 30 min and a testing time of 2.32 s. Although EfficientNet demonstrates the highest accuracy, its relatively long testing time presents potential challenges for real-time industrial applications. Therefore, a trade-off analysis is essential to balance accuracy and computational efficiency when selecting the most suitable model for practical use in industrial settings. This reveals a potential limitation in differentiating certain health states under varying conditions of operation.

Figure 11 shows the confusion matrices that visualize the fault diagnosis performance of the four deep transfer learning models of MobileNet, ResNET50v2, GoogleNet, and EfficientNet on robotic RV reducers across the three operational states of healthy (H), faulty (F), and faulty aged (FA), with 96 images analyzed per category. True labels are the actual ground truth classes from the dataset, while predicted labels are the classes assigned by the model. EfficientNet demonstrates the highest accuracy, correctly classifying (95.00, 100.00, and 100.00)% of faulty, faulty aged, and healthy samples, respectively, with only a small misclassification of 5.00% where faulty samples were labeled as healthy. Similarly, GoogleNet achieves good performance, correctly identifying (90.00, 100.00, and 95.00)% of faulty, faulty aged, and healthy samples, respectively, but shows minor errors with 5.00% of faulty samples classified as faulty aged and 5.00% of healthy misclassified as faulty. In comparison, MobileNet exhibits more noticeable misclassifications. While it achieves 95.00% accuracy for Faulty samples, it struggles with faulty aged and healthy states, correctly identifying only 85.00% of samples in both categories. Misclassifications include 15.00% of faulty aged samples classified as faulty and 10.00% of healthy samples misclassified as faulty. The performance of ResNET50v2 is comparable to MobileNet,

achieving (95.00, 85.00, and 85.00)% accuracy for faulty, faulty aged, and healthy samples, respectively, with similar patterns of misclassification. The higher misclassification rate for the faulty aged state is due to the overlapping features with the faulty condition, making it difficult for the models to distinguish between them. The limited training data for the “faulty aged” class also contributes to this misclassification. This combination results in reduced accuracy for the FA state. EfficientNet emerges as the most reliable model to diagnose faults in robotics RV reducers, followed by GoogleNet, while the differentiation of faulty aged states from healthy states is more challenging for MobileNet and ResNET50v2.

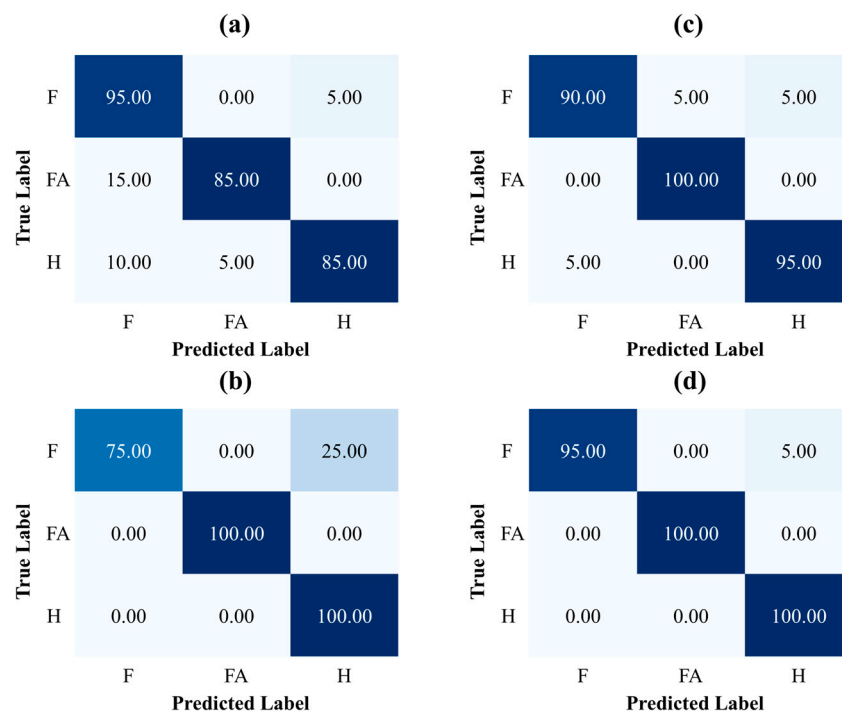


Figure 11. Confusion matrix for classification results of healthy (H), faulty (F), and faulty aged (FA) data for the following: (a) MobileNet; (b) ResNET50v2; (c) GoogleNet; (d) EfficientNet.

To better understand the implemented models, other evaluation metrics were studied. Table 3 shows the precision, recall, and F1-score of the various TL models used in this study. For precision, which is the correctness of the true prediction, the EfficientNet model correctly predicted with the highest precision value of 98.41%, while MobileNet attained an accuracy of 88.33%, representing the lowest performance among the four models evaluated. The precision, recall, and F1-score exhibited slight increases yet correspond to a comparatively lower accuracy. These results suggest that although MobileNet demonstrates satisfactory performance, it is outperformed by the other models in attaining a balanced trade-off between precision and recall, yielding an F1-score of 88.44%. ResNET50v2 shows enhanced performance, achieving an accuracy of 91.67%, as well as superior precision, recall, and F1-score, in comparison to MobileNet. This indicates that ResNET50v2 demonstrates greater reliability in data classification and maintains an effective balance between false positives and false negatives. The F1-score of 91.53% indicates consistent and comprehensive performance across essential metrics. GoogleNet also demonstrates enhanced performance, attaining an accuracy of 95.00%, along with comparable precision, recall, and F1-scores, all of approximately 95.00%. GoogleNet demonstrates high performance in both precisions by minimizing false positives and recall and by minimizing false negatives, thereby proving to be effective for classification tasks. EfficientNet achieves superior performance compared to the other models, attaining an accuracy of 98.33%, with precision, recall, and F1-scores also of 98.33%. The computational overhead of each model varies based on

its design and intended use. MobileNet is lightweight with minimal overhead, making it ideal for resource-constrained environments. ResNet50v2 offers a balanced trade-off between performance and efficiency. GoogleNet has a moderate computational load, which is suitable for tasks requiring higher complexity. ResNet101 is more computationally heavy, offering high accuracy at the cost of significant training time. InceptionV3 is demanding and optimized for accuracy but requires considerable resources. VGG16 is expensive, both in terms of parameter size and training time, which makes it resource intensive. Lastly, EfficientNet is optimal, offering top-notch accuracy with low computational overhead, striking the best balance between performance and efficiency.

Table 3. Comparison of accuracy, precision, recall, and F1-score for numerous TL models for robotic RV reduce fault detection.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Parameter Size (MB)	Training Time (s)
<i>MobileNet</i>	88.33	89.35	88.33	88.44	~12.35	99.5
<i>ResNET50v2</i>	91.67	93.33	91.53	91.53	~87.2	69
<i>GoogleNet</i>	95	94.99	95	94.96	~95.2	211.89
<i>ResNet101</i>	96.67	96.97	96.67	96.66	~175	1458.52
<i>InceptionV3</i>	90.00	90.24	90.00	89.97	~95.2	229.27
<i>VGG16</i>	91.67	93.33	91.67	91.53	~60	911.67
<i>EfficientNet</i>	98.33	98.41	98.33	98.33	~23	451

5. Conclusions

In this study, a TL-based fault diagnosis approach for RV reducers was developed and validated with the use of experimental data collected under varying operational conditions. The technology utilized CWT to transform current signal data into time–frequency scalogram images, hence improving feature representation for classification tasks. EfficientNet demonstrates superior performance with an accuracy and F1-score of 98.33% each, outperforming GoogleNet by 3.33% in accuracy and 3.37% in F1-score. Similarly, it surpasses ResNet50v2 by 6.66% in accuracy and 6.8% in F1-score and MobileNet by a significant margin of 10% in accuracy and 9.89% in F1-score, clearly establishing its advantages in fault diagnosis for robotic RV reducers. The findings indicated that EfficientNet attained superior classification performance, demonstrating few misclassifications across all health statuses. GoogleNet exhibited commendable performance; however, it showed marginally elevated misclassification rates relative to EfficientNet, particularly for the faulty aged condition. Conversely, despite their computational efficiency, MobileNet and ResNET50v2 demonstrated elevated misclassification rates, especially between the faulty aged and healthy categories, highlighting their inadequacies in differentiating nuanced variations in degrees of deterioration.

The use of TL models such as MobileNet underscores the compromise between computational efficiency and classification precision. While MobileNet offers significant benefits in terms of speed and resource consumption, its lower performance (88.33% accuracy, 88.44% F1-score) in distinguishing fault types calls for further exploration of trade-offs in industrial settings. The exceptional performance of EfficientNet arises from its compound scalability and efficient architecture, which utilizes depth-wise separable convolutions and bottleneck layers to enhance feature extraction. EfficientNet’s optimized architecture ensures high classification accuracy with minimal computational and memory requirements, making it ideal for deployment on resource-constrained industrial hardware. Moreover, its robustness

across varying operational speeds demonstrates its adaptability to real-world industrial environments. Moreover, the resilience of EfficientNet guarantees stable performance with negligible reduction in accuracy across training, validation, and test datasets, which illustrates its generalization ability. The suggested system is adept at tackling the obstacles presented by limited data and variable speed operations and demonstrates the efficacy of the models in fault diagnosis applications. The streamlined architecture of the assessed models reduces computing expenses, rendering them appropriate for implementation in industrial environments that have limited resources. However, the limitations observed in MobileNet and ResNET50v2 suggest their need for further optimization, or the integration of ensemble learning techniques, to improve their fault diagnosis capabilities.

Future work could focus on extending the framework to include additional fault types, a broader range of operational conditions, and real-world scenarios to enhance model robustness. Moreover, integrating domain adaptation techniques, ensemble methods, and real-time implementation in industrial robots would further ascertain the practical usefulness of this work. However, TL methods face limitations due to their reliance on pre-trained models trained on datasets like ImageNet, which may not align with the characteristics of industrial datasets. This mismatch can affect generalization performance under varying operational conditions. To address this, future work could explore domain adaptation techniques and hybrid approaches, combining TL with unsupervised learning to enhance robustness and adaptability to diverse datasets. Additionally, the creation of public databases for RV reducer fault diagnosis could facilitate broader applicability while improving model validation. In future research, expanding the current framework to include more fault types, varying operational conditions, and real-world scenarios should ensure its robustness and adaptability for diverse industrial applications.

Author Contributions: Conceptualization, I.R. and F.A.; Methodology, M.U.E. and I.R.; Software, M.U.E.; Validation, M.U.E., I.R. and S.K.; Formal analysis, S.K. and F.A.; Investigation, M.U.E.; Writing—original draft, I.R.; Writing—review & editing, S.K., F.A. and H.S.K.; Visualization, I.R., S.K., F.A. and H.S.K.; Supervision, H.S.K.; Project administration, H.S.K.; Funding acquisition, H.S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by a National Research Foundation of Korea (NRF) grant, graciously funded by the Korean government (MSIT) (RS-2024-00405691).

Data Availability Statement: Data will be made available upon request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study, the collection, analyses, or interpretation of data, the writing of the manuscript, or the decision to publish the results.

References

1. Shukla, A.; Karki, H. Application of Robotics in Onshore Oil and Gas Industry—A Review Part I. *Robot. Auton. Syst.* **2016**, *75*, 490–507. [\[CrossRef\]](#)
2. Liu, L.; Guo, F.; Zou, Z.; Duffy, V.G. Application, Development and Future Opportunities of Collaborative Robots (Cobots) in Manufacturing: A Literature Review. *Int. J. Hum. Comput. Interact.* **2022**, *40*, 915–932. [\[CrossRef\]](#)
3. Raouf, I.; Kumar, P.; Soo Kim, H. Deep Learning-Based Fault Diagnosis of Servo Motor Bearing Using the Attention-Guided Feature Aggregation Network. *Expert Syst. Appl.* **2024**, *258*, 125137. [\[CrossRef\]](#)
4. Zio, E. Prognostics and Health Management (PHM): Where Are We and Where Do We (Need to) Go in Theory and Practice. *Reliab. Eng. Syst. Saf.* **2022**, *218*, 108119. [\[CrossRef\]](#)
5. Hu, Y.; Miao, X.; Si, Y.; Pan, E.; Zio, E. Prognostics and Health Management: A Review from the Perspectives of Design, Development and Decision. *Reliab. Eng. Syst. Saf.* **2022**, *217*, 108063. [\[CrossRef\]](#)
6. Qiao, G.; Weiss, B.A. Advancing Measurement Science to Assess Monitoring, Diagnostics, and Prognostics for Manufacturing Robotics. *IJPHM* **2020**, *7*. [\[CrossRef\]](#)

7. E, Y.; Liu, Z.; Chen, H.; Zhang, Y.; Li, Z.; Feng, H.; Feng, Z. Dynamic Modeling and Vibration Analysis for Fault Diagnosis of Rotate Vector Reducers. *Mech. Syst. Signal Process.* **2025**, *224*, 111965. [[CrossRef](#)]
8. Pham, A.-D.; Ahn, H.-J. High Precision Reducers for Industrial Robots Driving 4th Industrial Revolution: State of Arts, Analysis, Design, Performance Evaluation and Perspective. *Int. J. Precis. Eng. Manuf.-Green Tech.* **2018**, *5*, 519–533. [[CrossRef](#)]
9. Liu, W.; Han, B.; Zheng, A.; Zheng, Z.; Chen, S.; Jia, S. Fault Diagnosis of Reducers Based on Digital Twins and Deep Learning. *Sci. Rep.* **2024**, *14*, 24406. [[CrossRef](#)] [[PubMed](#)]
10. Wei, B.; Wang, J.; Zhou, G.; Yang, R.; Zhou, H.; He, T. Mixed Lubrication Analysis of Modified Cycloidal Gear Used in the RV Reducer. *Proc. Inst. Mech. Eng. Part J J. Eng. Tribol.* **2016**, *230*, 121–134. [[CrossRef](#)]
11. Aivaliotis, P.; Arkouli, Z.; Georgoulas, K.; Makris, S. Degradation Curves Integration in Physics-Based Models: Towards the Predictive Maintenance of Industrial Robots. *Robot. Comput. Integr. Manuf.* **2021**, *71*, 102177. [[CrossRef](#)]
12. Qian, H.-M.; Li, Y.-F.; Huang, H.-Z. Time-Variant Reliability Analysis for Industrial Robot RV Reducer under Multiple Failure Modes Using Kriging Model. *Reliab. Eng. Syst. Saf.* **2020**, *199*, 106936. [[CrossRef](#)]
13. Licardo, J.T.; Domjan, M.; Orehovalčki, T. Intelligent Robotics—A Systematic Review of Emerging Technologies and Trends. *Electronics* **2024**, *13*, 542. [[CrossRef](#)]
14. Yan, Y.; Guo, Y.; Liu, X. Tooth Root Crack Detection of Planet Gear in Industrial Robot RV Reducer. *Meas. Control* **2023**, *56*, 1720–1731. [[CrossRef](#)]
15. Raouf, I.; Kumar, P.; Lee, H.; Kim, H.S. Transfer Learning-Based Intelligent Fault Detection Approach for the Industrial Robotic System. *Mathematics* **2023**, *11*, 945. [[CrossRef](#)]
16. Wang, D.; Liu, X.; Wu, X.; Liu, T.; Xu, K.; Li, J. Transient Feature Extraction of Encoder Signal Based on Envelope Demodulation for Fault Diagnosis of RV Reducer. *IEEE Sens. J.* **2024**, *24*, 21082–21092. [[CrossRef](#)]
17. Peng, P.; Wang, J. NOSCNN: A Robust Method for Fault Diagnosis of RV Reducer. *Measurement* **2019**, *138*, 652–658. [[CrossRef](#)]
18. An, H.; Liang, W.; Zhang, Y.; Li, Y.; Liang, Y.; Tan, J. Rotate Vector Reducer Crankshaft Fault Diagnosis Using Acoustic Emission Techniques. In Proceedings of the 2017 5th International Conference on Enterprise Systems (ES), Beijing, China, 22–24 September 2017; IEEE: Piscataway, NJ, USA; 2017; pp. 294–298.
19. Wu, H.; Kwok, N.M.; Liu, S.; Li, R.; Wu, T.; Peng, Z. Restoration of Defocused Ferrograph Images Using a Large Kernel Convolutional Neural Network. *Wear* **2019**, *426–427*, 1740–1747. [[CrossRef](#)]
20. Peng, P.; Wang, J. Wear Particle Classification Considering Particle Overlapping. *Wear* **2019**, *422–423*, 119–127. [[CrossRef](#)]
21. Zhang, Y.; Lu, W.; Chu, F. Planet Gear Fault Localization for Wind Turbine Gearbox Using Acoustic Emission Signals. *Renew. Energy* **2017**, *109*, 449–460. [[CrossRef](#)]
22. Xuan, L.; Hong, W.; Wan, Z.; Lin, Z. Review on Vibration Test and Fault Analysis of Rv Reducer. *J. Phys. Conf. Ser.* **2023**, *2472*, 012061. [[CrossRef](#)]
23. Rohan, A. Wavelets-Based Features Infusion for Machine Learning Classification; Towards a Holistic Rotate Vector (RV) Reducer Fault Detection and Diagnosis System for an Industrial Robot. *CO/JRA* **2020**, *1*, 1587–1589. [[CrossRef](#)]
24. Khan, A.; Ko, D.-K.; Lim, S.C.; Kim, H.S. Structural Vibration-Based Classification and Prediction of Delamination in Smart Composite Laminates Using Deep Learning Neural Network. *Compos. Part B Eng.* **2019**, *161*, 586–594. [[CrossRef](#)]
25. Lee, H.; Raouf, I.; Song, J.; Kim, H.S.; Lee, S. Prognostics and Health Management of the Robotic Servo-Motor under Variable Operating Conditions. *Mathematics* **2023**, *11*, 398. [[CrossRef](#)]
26. Zhang, S.; Gao, J.; Wang, L.; Chen, C.; Xu, S.; Wang, B. A Novel On-Line Approach for Evaluating Transmission Errors in Harmonic Drives. *Adv. Mech. Eng.* **2024**, *16*, 16878132241276666. [[CrossRef](#)]
27. Ahmed, S.F.; Alam, M.S.B.; Hassan, M.; Rozbu, M.R.; Ishtiaq, T.; Rafa, N.; Mofijur, M.; Shawkat Ali, A.B.M.; Gandomi, A.H. Deep Learning Modelling Techniques: Current Progress, Applications, Advantages, and Challenges. *Artif. Intell. Rev.* **2023**, *56*, 13521–13617. [[CrossRef](#)]
28. Tufail, S.; Riggs, H.; Tariq, M.; Sarwat, A.I. Advancements and Challenges in Machine Learning: A Comprehensive Review of Models, Libraries, Applications, and Algorithms. *Electronics* **2023**, *12*, 1789. [[CrossRef](#)]
29. Kumar, P.; Hati, A.S. Review on Machine Learning Algorithm Based Fault Detection in Induction Motors. *Arch. Comput. Methods Eng.* **2021**, *28*, 1929–1940. [[CrossRef](#)]
30. Tang, Z.; Bo, L.; Liu, X.; Wei, D. An Autoencoder with Adaptive Transfer Learning for Intelligent Fault Diagnosis of Rotating Machinery. *Meas. Sci. Technol.* **2021**, *32*, 055110. [[CrossRef](#)]
31. Wang, H.; Wang, S.; Yang, R.; Xiang, J. A Numerical Simulation Enhanced Multi-Task Integrated Learning Network for Fault Detection in Rotation Vector Reducers. *Mech. Syst. Signal Process.* **2024**, *217*, 111525. [[CrossRef](#)]
32. Qiu, G.; Nie, Y.; Peng, Y.; Huang, P.; Chen, J.; Gu, Y. A Variable-speed-condition Fault Diagnosis Method for Crankshaft Bearing in the RV Reducer with WSO-VMD and ResNet-SWIN. *Qual. Reliab. Eng.* **2024**, *40*, 2321–2347. [[CrossRef](#)]
33. Helwa, M.K.; Schoellig, A.P. Multi-Robot Transfer Learning: A Dynamical System Perspective. In Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vancouver, BC, Canada, 24–28 September 2017; IEEE: Piscataway, NJ, USA; pp. 4702–4708.

34. Zhang, W.; Zhang, T.; Cui, G.; Pan, Y. Intelligent Machine Fault Diagnosis Using Convolutional Neural Networks and Transfer Learning. *IEEE Access* **2022**, *10*, 50959–50973. [\[CrossRef\]](#)
35. Sun, X.; Wang, S.; Jing, J.; Shen, Z.; Zhang, L. Fault Diagnosis Using Transfer Learning with Dynamic Multiscale Representation. *Cogn. Robot.* **2023**, *3*, 257–264. [\[CrossRef\]](#)
36. Guo, L.; Lei, Y.; Xing, S.; Yan, T.; Li, N. Deep Convolutional Transfer Learning Network: A New Method for Intelligent Fault Diagnosis of Machines with Unlabeled Data. *IEEE Trans. Ind. Electron.* **2019**, *66*, 7316–7325. [\[CrossRef\]](#)
37. Galan-Urbe, E.; Amezcua-Sanchez, J.P.; Morales-Velazquez, L. Supervised Machine-Learning Methodology for Industrial Robot Positional Health Using Artificial Neural Networks, Discrete Wavelet Transform, and Nonlinear Indicators. *Sensors* **2023**, *23*, 3213. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Raouf, I.; Lee, H.; Noh, Y.R.; Youn, B.D.; Kim, H.S. Prognostic Health Management of the Robotic Strain Wave Gear Reducer Based on Variable Speed of Operation: A Data-Driven via Deep Learning Approach. *J. Comput. Des. Eng.* **2022**, *9*, 1775–1788. [\[CrossRef\]](#)
39. Wen, L.; Gao, L.; Li, X. A New Deep Transfer Learning Based on Sparse Auto-Encoder for Fault Diagnosis. *IEEE Trans. Syst. Man Cybern. Syst.* **2019**, *49*, 136–144. [\[CrossRef\]](#)
40. He, F.; Liu, C.; Wang, M.; Yang, E.; Liu, X. Network Lightweight Method Based on Knowledge Distillation Is Applied to RV Reducer Fault Diagnosis. *Meas. Sci. Technol.* **2023**, *34*, 095110. [\[CrossRef\]](#)
41. Raouf, I.; Lee, H.; Kim, H.S. Mechanical Fault Detection Based on Machine Learning for Robotic RV Reducer Using Electrical Current Signature Analysis: A Data-Driven Approach. *J. Comput. Des. Eng.* **2022**, *9*, 417–433. [\[CrossRef\]](#)
42. Khorrami, H.; Moavenian, M. A Comparative Study of DWT, CWT and DCT Transformations in ECG Arrhythmias Classification. *Expert Syst. Appl.* **2010**, *37*, 5751–5757. [\[CrossRef\]](#)
43. Gu, J.; Peng, Y.; Lu, H.; Chang, X.; Chen, G. A Novel Fault Diagnosis Method of Rotating Machinery via VMD, CWT and Improved CNN. *Measurement* **2022**, *200*, 111635. [\[CrossRef\]](#)
44. Kim, Y.Y.; Kim, E.-H. Effectiveness of the Continuous Wavelet Transform in the Analysis of Some Dispersive Elastic Waves. *J. Acoust. Soc. Am.* **2001**, *110*, 86–94. [\[CrossRef\]](#)
45. Tytarenko, S.; Amin, M.R. Breaking Free Transformer Models: Task-Specific Context Attribution Promises Improved Generalizability Without Fine-Tuning Pre-Trained LLMs 2024. *arXiv* **2024**, arXiv:2401.16638.
46. Lee, H.; Lee, S.; Ryu, S. Advancements and Challenges of Micromechanics-Based Homogenization for the Short Fiber Reinforced Composites. *Multiscale Sci. Eng.* **2023**, *5*, 133–146. [\[CrossRef\]](#)
47. Tama, B.A.; Vania, M.; Kim, I.; Lim, S. An EfficientNet-Based Weighted Ensemble Model for Industrial Machine Malfunction Detection Using Acoustic Signals. *IEEE Access* **2022**, *10*, 34625–34636. [\[CrossRef\]](#)
48. Zhu, F.; Liu, C.; Yang, J.; Wang, S. An Improved MobileNet Network with Wavelet Energy and Global Average Pooling for Rotating Machinery Fault Diagnosis. *Sensors* **2022**, *22*, 4427. [\[CrossRef\]](#) [\[PubMed\]](#)
49. Mishra, A.K.; Mohanty, S.K. A Deep Convolutional Neural Network Model for Nonlinear Drift Fault Diagnosis in Sensors Using Transfer Learning and Multiresolution Analysis. In Proceedings of the 2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT), Mandya, India, 26–27 December 2022; IEEE: Piscataway, NJ, USA; pp. 1–7.
50. Wang, S.; Feng, Z. Multi-Sensor Fusion Rolling Bearing Intelligent Fault Diagnosis Based on VMD and Ultra-Lightweight GoogLeNet in Industrial Environments. *Digit. Signal Process.* **2024**, *145*, 104306. [\[CrossRef\]](#)
51. Kumar, P.; Kumar, P.; Hati, A.S.; Kim, H.S. Deep Transfer Learning Framework for Bearing Fault Detection in Motors. *Mathematics* **2022**, *10*, 4683. [\[CrossRef\]](#)
52. Luo, L.; Li, P.; Yan, X. Deep Learning-Based Building Extraction from Remote Sensing Images: A Comprehensive Review. *Energies* **2021**, *14*, 7982. [\[CrossRef\]](#)
53. Deepa, S.; Zeema, J.L.; Gokila, S. Exploratory Architectures Analysis of Various Pre-Trained Image Classification Models for Deep Learning. *JAIT* **2024**, *15*, 66–78. [\[CrossRef\]](#)
54. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaria, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions. *J. Big Data* **2021**, *8*, 53. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Mohseni-Dargah, M.; Falahati, Z.; Dabirmanesh, B.; Nasrollahi, P.; Khajeh, K. Machine Learning in Surface Plasmon Resonance for Environmental Monitoring. In *Artificial Intelligence and Data Science in Environmental Sensing*; Elsevier: Amsterdam, The Netherlands, 2022; pp. 269–298, ISBN 978-0-323-90508-4.
56. Wang, T.; Zhang, L.; Wang, X. Fault Detection for Motor Drive Control System of Industrial Robots Using CNN-LSTM-Based Observers. *Trans. Electr. Mach. Syst.* **2023**, *7*, 144–152. [\[CrossRef\]](#)
57. Tan, M.; Le, Q. *Efficientnet: Rethinking Model Scaling for Convolutional Neural Networks*; PMLR: Birmingham, UK, 2019; pp. 6105–6114.

58. Kim, T.-Y.; Park, S.; Lim, C.-K.; Choi, J.-H.; An, C.-H.; Song, L.-H.; Kim, D.-N. Deep Learning-Based Detection of Defects in Wafer Buffer Zone During Semiconductor Packaging Process. *Multiscale Sci. Eng.* **2024**, *6*, 25–32. [[CrossRef](#)]
59. Sinha, D.; El-Sharkawy, M. Thin MobileNet: An Enhanced MobileNet Architecture. In Proceedings of the 2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York City, NY, USA, 10–12 October 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 0280–0285.
60. Azad, M.M.; Kumar, P.; Kim, H.S. Delamination Detection in CFRP Laminates Using Deep Transfer Learning with Limited Experimental Data. *J. Mater. Res. Technol.* **2024**, *29*, 3024–3035. [[CrossRef](#)]
61. Goodwin, M.; Halvorsen, K.T.; Jiao, L.; Knausgård, K.M.; Martin, A.H.; Moyano, M.; Oomen, R.A.; Rasmussen, J.H.; Sjørdalen, T.K.; Thorbjørnsen, S.H. Unlocking the Potential of Deep Learning for Marine Ecology: Overview, Applications, and Outlook. *ICES J. Mar. Sci.* **2022**, *79*, 319–336. [[CrossRef](#)]
62. Latif, S.; Boulila, W.; Koubaa, A.; Zou, Z.; Ahmad, J. DTL-IDS: An Optimized Intrusion Detection Framework Using Deep Transfer Learning and Genetic Algorithm. *J. Netw. Comput. Appl.* **2024**, *221*, 103784. [[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.