

# Transfer Learning Strategy for Fault Identification in Wind Turbine High-Speed Shaft Bearing with Limited Samples

SM Gbashi<sup>a</sup>, OO Olatunji<sup>a</sup>, PA. Adedeji<sup>a</sup>, N Madushele<sup>a,b</sup>

Received 1 January 2024, in revised form 22 April 2024 and accepted 22 April 2024

**Abstract:** *The application of deep learning algorithms for fault identification in wind turbine components is contingent on extensive data. Such data is often scarce, especially in the faulty category. While adversarial data augmentation helps, biases from the original data persist, and larger datasets strain computational resources. As a solution, experts are turning to transfer learning. Leveraging insights from related domains, transfer learning enables machine learning models to circumvent the exigency of training from scratch with extensive data. This study proposed a transfer learning strategy for fault identification in high-speed wind turbine shaft bearings. Two-dimensional matrices extracted from vibration signals sampled from the turbine bearings are employed to train ResNet50 and VGG16 convolutional neural network models with frozen weights based on transfer learning. While both models performed well on the normal test samples, they showed differing robustness when evaluated with noise-induced test samples. Contrarily, the ResNet50 had an accuracy, F-score, and training time of 82.21 %, 78.34 %, and 26.3 s, respectively, while the VGG16 model had an accuracy and F-score of 95.55 % and 95.35 %, respectively, but trained for 46 s. The ResNet50 may have converged quickly due to “skip connection” in its architecture, typical of residual learning models. While the VGG16 is computationally intensive, its superior performance and resilience to noise make it suited for vibration-based defect detection in the high-speed shaft bearing, where severe background noise is prevalent.*

**Additional keywords:** Convolutional Neural Network; Fault Identification; High-speed Shaft Bearing; Transfer Learning, Wind Turbine Gearbox.

## 1 Introduction

The failure of the gearbox high-speed shaft bearing (HSSB) accounts for 20 % of wind turbine downtime [1]. In recent years, due to limitations in the use of conventional models for fault identification in the HSSB, there has been a growing trend in the use of deep learning techniques [2]. The effi-

<sup>a</sup> Department of Mechanical Engineering Science, University of Johannesburg, Cnr Kingsway & University Roads, Auckland Park, Johannesburg, South Africa. E-mail: samgbash247@gmail.com

<sup>b</sup> Department of Mechanical Engineering, Durban University of Technology, 70 Steve Biko Road, Berea, Durban, South Africa.

cacy of deep learning algorithms stems from their capacity for end-to-end learning, obviating the need for costly feature engineering. The application of deep learning models, however, depends on the availability of extensive training data. Due to various reasons, including access restrictions and privacy concerns, condition monitoring datasets (especially in the faulty category) are often limited [3]. The latter presents practical application limitations for deep learning models. Some authors have attempted to address the difficulties of training deep neural networks in the face of limited data [4, 5]. The core of these efforts has centred on using generative models, particularly generative adversarial networks (GANs), for data augmentation. Zhou et al. [4] employed GANs to augment data for improved wind turbine power forecasting. Liang et al. [5] employed GANs for data augmentation in single and compound fault diagnosis of wind turbine gearboxes. While adversarial data augmentation helps, biases from the original data persist, and larger datasets strain computational resources [6]. As a solution, experts are turning to transfer learning (TL). Leveraging insights from related domains, TL enables machine learning models to circumvent the exigency of training from scratch with extensive data, a possibility that this study seeks to explore in the wind turbine space.

Transfer learning has become increasingly popular in the wind turbine space. However, there are still research gaps. The application of transfer learning in the wind turbine space has been limited to a few algorithms, including convolutional autoencoder, LSTM, MobileNetV1-YOLOv4, and a few others [7]. Even fewer studies have applied TL to the gearbox HSSB. These studies employed deep convolutional neural networks and recurrent neural networks [8]. In light of the emergence of pre-trained neural networks in recent years, this study area remains fertile for further investigation.

Recently, the use of convolutional neural network (CNN) pre-trained variants, including the Residual Network – 50 (ResNet50) and the Visual Geometry Group - 16 (VGG16), for transfer learning with limited samples has gained significant interest in many domains. Among others, these algorithms have been applied for improved feature classification with significant success rates in agriculture, cybersecurity, machinery health monitoring, and medicine [9, 10]. The ResNet50 is a residual model proposed by He et al. [11] in 2015. The depth of the ResNet50 makes it excel in image recognition, segmentation, and classification. The model extracts features from image data through its convolutional layers, comprising 3x3 and 1x1 filters. The Visual Geometry Group 16 (VGG16) is a variant of CNN developed by the

Visual Geometry Group at the University of Oxford in 2014 [12]. The VGG16 primarily employs 3x3 convolutional filters. The small filter sizes allow the network to learn complex spatial features in image data. While the VGG16 and ResNet50 models have demonstrated outstanding performance in image classification tasks in other domains, they have not been implemented for fault recognition in the HSSB. To bridge this gap, this study proposes a transfer learning strategy based on the ResNet50 and VGG16 models for fault identification in HSSB.

## 2 Study Methodology

### 2.1 Dataset information

In this work, we used the National Renewable Energy Laboratory benchmarking dataset as our case study. The data consists of 10-minute vibration signals sampled from the high-speed shaft bearing (SKF 32222-J2 tapered roller bearing) of a 750 kW stall-controlled wind turbine. The vibration signals were acquired from accelerometers mounted in the gearbox at a sample rate of 40 kHz using the National Instrument’s PXI-4472B data acquisition system.

### 2.2 Model development

#### 2.2.1 Transfer learning

The theoretical basis of transfer learning is founded on the concepts of domain and task, denoted by  $D$  and  $T$  respectively [13]. A domain comprises two parts, a marginal probability distribution and a feature space, denoted by  $P(x)$  and  $\chi$ , in order. If  $x \in \chi$ , then the domain may be defined mathematically as:

$$D = \{\chi, P(x)\} \tag{1}$$

In the same vein, if a decision function  $f(x)$  and a label space  $\Upsilon$  are components of a task, then the task may be defined as:

$$T = \{\Upsilon, f(x)\} \tag{2}$$

The expression above may be viewed as a conditional probability distribution  $P(y|x)$  in which  $y \in \Upsilon$ . In practice, a domain with a substantial quantity of sample data accompanied by label information is called a source domain. The term “target domain” refers to the domain with a knowledge deficit. Transfer learning seeks to leverage the knowledge acquired from a source domain and effectively apply it to a target domain [14] to accomplish the desired task in the target domain (Figure 1). In this study, we employ two convolutional neural network variants, ResNet50 and VGG16, for transfer learning. Both models have been pre-trained on the ImageNet dataset.

#### 2.2.2 Transfer learning with ResNet50 and VGG16

A CNN is a machine learning model suited for learning features in data arranged in grid patterns, such as image data. In CNNs, spatial features are learned from image data through

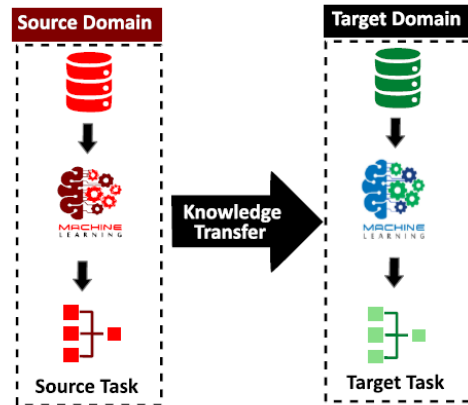


Figure 1 Knowledge transfer from source domain to target domain

convolution operations. In this study, we employ two variants of the CNN, the ResNet50 (Figure 2) and VGG16 (Figure 3) for transfer learning. To achieve transfer learning, the final layers of the VGG16 and ResNet50 were removed and replaced with two dense layers and a sigmoid activation function (Figure 4). Whereas the first dense layer had 512 neurons, the last dense layer had a single neuron. The sigmoid function in the updated models converts real-valued inputs to a probability range of [0, 1], allowing for binary classification of the gearbox HSSB states. Subsequently, the weights of the new models were frozen while the fully connected layers remained trainable. This enabled knowledge transfer from the pre-trained CNN to our modified models. Table 1 presents the training hyperparameters of the modified models.

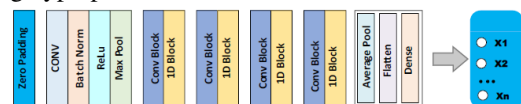


Figure 2 Architecture of the ResNet50 CNN model

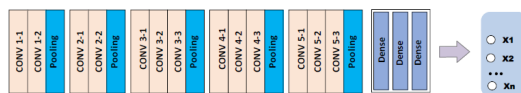


Figure 3 Architecture of the VGG16 CNN model

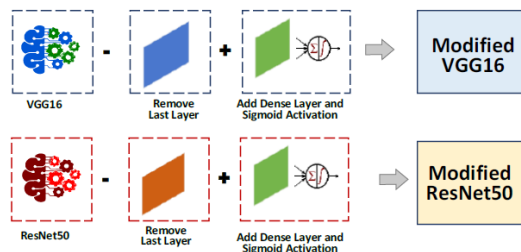


Figure 4 Modification of pre-trained VGG16 and ResNet50 models for transfer learning

### 2.3 Gaussian noise addition

To evaluate the robustness of the VGG16 and ResNet50 models, the models were tested with vibration signals corrupted with Gaussian white noise at a signal-to-noise ratio of 8dB

Table 1 The VGG16 and ResNet50 training hyperparameters

Parameter Name	Value
Learning rate	0.01
Loss function	Binary cross entropy
Metrics	Accuracy
Batch Size	200
Epochs	5

[15]. The signal-to-noise ratio (SNR) of a vibration signal is expressed as the ratio of the average power of the signal ( $P_{signal\ avr}$ ) to the average power of noise ( $P_{noise\ avr}$ ):

$$SNR = \frac{P_{signal\ avr}}{P_{noise\ avr}} \quad (3)$$

### 2.4 Performance evaluation of the models

Four performance metrics, defined in the equations below, were used to evaluate the performance of the study models. They include accuracy, recall (sensitivity), precision (specificity) and F1-score based on the number of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) results reported by the model under investigation.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$F1\text{-score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (7)$$

## 3 Results and Discussion

The scripts for the models were written in Python and installed on a laptop PC with a 12<sup>th</sup> generation Intel Core i7 multi-core microprocessor with 32 GB of RAM. The study employed 70 % of the dataset for training the models, 10 % for validation, and 20 % for testing.

### 3.1 Comparison of the ResNet50 and VGG16 models

Table 2 compares the performance and computational time of the ResNet50 and VGG16 models on the normal and corrupted test data. The analysis of the findings presented in Table 2 indicates that there is no significant difference in the performance of the study models when evaluated on normal vibration data. Both models had peak performance scores for accuracy, precision, recall, and F1-score. The models' excellent results align with those found by Yoo et al. [16], who used transfer learning with a ResNet50 and a VGG16

model to achieve defect classification accuracies of 99.74 % and 99.88 %, respectively, in ball bearing multi-defect identification.

Table 2 Performance scores and computational time of the ResNet50 and VGG16 models on the normal and noise-induced test data

Performance Measure	Normal		Noise-induced	
	ResNet50	VGG16	ResNet50	VGG16
Accuracy (%)	99.96	99.97	82.21	95.55
Precision (%)	100.0	99.99	88.00	98.85
Recall (%)	99.99	100.0	64.39	91.24
F1-Score (%)	99.94	99.96	78.34	95.35
Computation Time (s)	21.5	43.1	26.3	46.3

Table 2 also compares the performance and computational time of the models on the noise-induced test data. The table shows that the ResNet50 and a VGG16 have differing resilience to noise when evaluated on the corrupted vibration signals. The VGG16 is seen to outperform the ResNet50 with an average performance margin of 15.35 % across all the metrics evaluated. The related confusion matrices and area under the curves (AUCs) of the models are depicted in Figure 5. The superior AUC of the VGG16 demonstrates the model's diagnostic prowess over its counterpart. The confusion matrices of the two models provide an intuitive confirmation of this observation. While the ResNet50 exhibited multiple misclassified observations, the VGG16 had only a few misclassifications. The high performance of the VGG16 network is a result of its compact filter sizes, which enable it to learn complex spatial features in image data. The low performance of the ResNet50 compared to the VGG16 may be linked to its extreme depth, which may have caused it to overfit the training data. The last column in Table 2 compares the computation times of the ResNet50 and VGG16 models. It is seen from these results that although the VGG16 outperformed the ResNet50, this outstanding performance came at the expense of computation speed. While the small filter sizes enabled the VGG16 to learn complex patterns in the vibration signal, convolutional operations required much processing time. The ResNet50 may have converged quickly due to "skip connection" in its architecture, typical of residual learning models. While the VGG16 is computationally intensive, its superior performance and resilience to noise make it suited for vibration-based defect detection in the wind turbine gearbox HSSB, where severe background noise is prevalent.

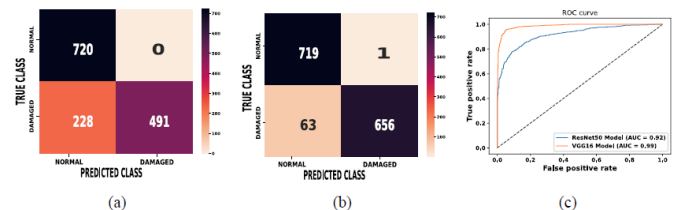


Figure 5 (a) Confusion matrix of the ResNet50 model, (b) confusion matrix of VGG16 model, (c) AUC of the ResNet50 and the VGG16 model

In the bar chart shown in Figure 6, the accuracies and F1-scores of the ResNet50 and VGG16 models are compared with those of a vanilla CNN model employed in a compara-

ble study for fault identification in the HSSB [17]. In contrast to this study's models, the reference study's CNN model was trained from scratch. The bar chart compares the robustness of the models to noise induced in the test data, as evidenced by the performance scores of the respective models. The chart shows that the ResNet50 and VGG16 models demonstrate superior resilience to noisy vibration signals compared to the vanilla CNN model. This result underscores the prowess of pre-trained models. By leveraging knowledge from other domains, transfer learning, while requiring limited samples, obtains comparable (and sometimes superior) performance to models trained from scratch on similar tasks.

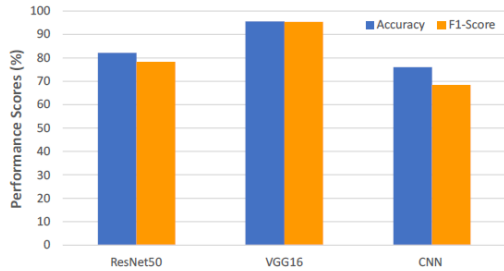


Figure 6 Comparison of accuracy and F1-score of ResNet50, VGG16 and a vanilla CNN model in a related study [17]

## 4 Conclusion

This study proposed a novel transfer learning technique for fault identification in the high-speed shaft bearing of the wind turbine gearbox. The study employed pre-trained ResNet50 and VGG16 CNN models to learn fault features from the vibration signals of the HSSB. Analysis of the results showed that the VGG16 outperformed its counterpart, albeit with a trade-off in computational efficiency. The VGG16 model represents a potential resource for data analysts seeking efficient fault diagnosis with limited samples in the wind turbine space. The proposed strategy is a feasible alternative to employing synthetic data generated through adversarial learning. Future studies could employ the proposed method for fault identification in other wind turbine components, like the main bearing, where data limitations are a persistent concern.

## References

- [1] Chengjia Bao, Tianyi Zhang, Zhixi Hu, Wangjing Feng, and Ru Liu. Wind turbine condition monitoring based on improved active learning strategy and KNN algorithm. *IEEE Access*, 11:13545–13553, 2023.
- [2] Hamida Maatallah and Kais Ouni. Health assessment of wind turbine bearings progressive degradation based on unsupervised machine learning. *Wind Engineering*, 46(6):1888–1900, 2022.
- [3] Stefan Petscharnig, Mathias Lux, and Savvas Chatzichristofis. Dimensionality reduction for image features using deep learning and autoencoders. In *Proceedings of the 15th international workshop on content-based multimedia indexing*, pages 1–6, 2017.
- [4] Bin Zhou, Haoran Duan, Qiuwei Wu, Huaizhi Wang, Siu Wing Or, Ka Wing Chan, and Yunfan Meng. Short-term prediction of wind power and its ramp events based on semi-supervised generative adversarial network. *International Journal of Electrical Power & Energy Systems*, 125:106411, 2021.
- [5] Pengfei Liang, Chao Deng, Xiaoming Yuan, and Lijie Zhang. A deep capsule neural network with data augmentation generative adversarial networks for single and simultaneous fault diagnosis of wind turbine gearbox. *ISA transactions*, 135:462–475, 2023.
- [6] Niharika Jain, Alberto Olmo, Sailik Sengupta, Lydia Manikonda, and Subbarao Kambhampati. Imperfect ImaGANation: Implications of GANs exacerbating biases on facial data augmentation and Snapchat face lenses. *Artificial Intelligence*, 304:103652, 2022.
- [7] Yanting Li, Wenbo Jiang, Guangyao Zhang, and Lianjie Shu. Wind turbine fault diagnosis based on transfer learning and convolutional autoencoder with small-scale data. *Renewable Energy*, 171:103–115, 2021.
- [8] Joyjit Chatterjee and Nina Dethlefs. Deep learning with knowledge transfer for explainable anomaly prediction in wind turbines. *Wind Energy*, 23(8):1693–1710, 2020.
- [9] Haoyan Yang, Jiangong Ni, Jiyue Gao, Zhongzhi Han, and Tao Luan. A novel method for peanut variety identification and classification by improved VGG16. *Scientific Reports*, 11(1):15756, 2021.
- [10] Edmar Rezende, Guilherme Ruppert, Tiago Carvalho, Antonio Theophilo, Fabio Ramos, and Paulo de Geus. Malicious software classification using VGG16 deep neural network's bottleneck features. In *Information Technology-New Generations: 15th International Conference on Information Technology*, pages 51–59. Springer, 2018.
- [11] Mangalam Sankupellay and Dmitry Konovalov. Bird call recognition using deep convolutional neural network, ResNet-50. In *Proc. Acoustics*, volume 7, pages 1–8, 2018.
- [12] Mohammad Yaseliani, Ali Zeinal Hamadani, Abtin Ijadi Maghsoodi, and Amir Mosavi. Pneumonia detection proposing a hybrid deep convolutional neural network based on two parallel visual geometry group architectures and machine learning classifiers. *IEEE access*, 10:62110–62128, 2022.
- [13] Silvio Simani, Saverio Farsoni, and Paolo Castaldi. Retracted: Supervisory control and data acquisition for fault diagnosis of wind turbines via deep transfer learning. *Energies*, 16(9):3644, 2023.
- [14] Xiaohang Jin, Hengtuo Pan, Chengzuo Ying, Ziqian Kong, Zhengguo Xu, and Bin Zhang. Condition monitoring of wind turbine generator based on transfer learning and one-class classifier. *IEEE Sensors Journal*, 22(24):24130–24139, 2022.

- [15] Hongchun Sun, Xu Cao, Changdong Wang, and Sheng Gao. An interpretable anti-noise network for rolling bearing fault diagnosis based on fswt. *Measurement*, 190:110698, 2022.
- [16] Youngjun Yoo and Seongcheol Jeong. Vibration analysis process based on spectrogram using gradient class activation map with selection process of cnn model and feature layer. *Displays*, 73:102233, 2022.
- [17] Samuel M Gbashi, Obafemi O Olatunji, Paul A Adedeji, and Nkosinathi Madushele. Hyperparameter optimization on cnn using hyperband for fault identification in wind turbine high-speed shaft gearbox bearing. In *2023 International Conference on Electrical, Computer and Energy Technologies (ICECET)*, pages 1–7. IEEE, 2023.