Wind Turbine Power Classification using Adaptive Transfer Learning

L. Venugopal[1], S. Tirupathi[2]

[1,2] *Department of ECE, Balaji Arts & Science College, Tamil Nadu, India.*

[1]*venustar11@gmail.com*

Received: 11.05.2025 Revised: 05.06.2025 Accepted: 16.06.2025 Published:30.6.2025

Abstract - *Accurate classification of wind turbine output power is essential for optimising grid integration, predictive maintenance scheduling, and energy forecasting. Conventional machine learning methods frequently suffer from domain shift when deployed across turbines manufactured by different vendors or operating under varying climatic regimes. This paper introduces an Adaptive Transfer Learning (ATL) framework that combines a convolutional feature extractor pre-trained on large-scale industrial vibration and power-curve datasets with a domain alignment mechanism based on Maximum Mean Discrepancy (MMD) loss and an attention-gated classifier head. The framework employs layer-wise learning-rate decay to selectively fine-tune deeper representations while preserving low-level spectral features learned during pre-training. Experiments are conducted on three publicly available SCADA datasets—the ENGIE La Haute Borne open dataset, the NREL Western Wind dataset, and an in-house dataset collected from a 2 MW onshore turbine array—encompassing four operational classes: Low Power, Medium Power, High Power, and Fault. The proposed ATL model achieves an overall classification accuracy of 95.7%, an F1-score of 94.9%, and a precision of 95.2%, outperforming SVM, Random Forest, CNN-from-scratch, LSTM-from-scratch, and ResNet-50 fine-tuning baselines by margins of 7.3 to 17.3 percentage points in accuracy. SHAP-based explainability analysis confirms that wind speed, rotor RPM, and generator temperature are the dominant discriminative features. The results demonstrate that the ATL framework generalises robustly across turbine types without requiring complete retraining, offering a practical pathway for scalable condition monitoring in commercial wind farms.*

Keywords - *Wind turbine power classification · Adaptive transfer learning · Maximum Mean Discrepancy · SCADA data · Convolutional neural network · Domain adaptation · Condition monitoring*

1. Introduction

Wind energy constitutes one of the fastest-growing segments of the global electricity supply, contributing approximately 2,100 TWh annually and representing over 7% of worldwide electricity generation as of 2023 [1]. Transfer learning [2] has emerged as an effective strategy to bridge the gap between data-rich source domains and data-scarce or distribution-shifted target domains in industrial condition monitoring. Pre-trained deep neural networks, initially developed for image recognition or time-series regression tasks, can be adapted to new turbine configurations through selective fine-tuning of upper-layer weights, thereby avoiding the cost and latency of training from scratch. However, naive fine-tuning does not explicitly align the feature distributions between the source and target domains, leaving residual domain shift that manifests as classification errors near class boundaries. This work addresses these limitations through an Adaptive Transfer Learning (ATL) framework that integrates three complementary mechanisms: (i) a frozen backbone for low-level feature reuse, (ii) an MMD-penalised domain alignment layer that minimises the discrepancy between source and target feature distributions in a reproducing kernel Hilbert space, and (iii) an attention gate that re-weights feature channels based on their discriminative salience for the target task. The key contributions of this study are summarised below. An end-to-end ATL pipeline for four-class wind turbine power classification that explicitly handles inter-turbine domain shift. A novel combination of MMD domain alignment and channel-wise attention gating within a unified fine-tuning objective. Systematic evaluation on three independent SCADA datasets, including a controlled ablation study isolating the contribution of each ATL component. SHAP-based post-hoc interpretability analysis identifying the dominant physical features driving each power class. Classical regression-based approaches, including polynomial and spline fits to the manufacturer power curve, established an early baseline for turbine performance estimation [3]. Subsequent data-driven methods employed Gaussian Process Regression (GPR) [4] and ensemble regressors such as Gradient Boosting Machines (GBM) [5] to capture the non-linear dependence of power output on meteorological covariates. While effective under stationary conditions, these methods require per-turbine recalibration when deployed across heterogeneous fleets.

The formulation of power estimation as a classification problem was explored by Pandit et al. [6], who employed Support Vector Machines (SVM) with radial basis kernels on binned SCADA time series. Random Forest classifiers were applied by Tautz-Weinert and Watson [7] for anomaly labelling derived from operational state bins, reporting accuracy above 82% on a single-turbine dataset. These studies, however, do not address cross-turbine generalisation.

Transfer learning for fault diagnosis in rotating machinery gained substantial traction following the work of Li et al. [8], who demonstrated that convolutional features learned on spectrogram images of healthy bearings transferred effectively to degraded operational conditions. Domain-adversarial training [9] was subsequently adopted to align feature spaces between different rotating speeds, yielding accuracy gains of up to 12% relative to standard fine-tuning. For wind turbines specifically, Zhu et al. [10] applied a deep adaptation network to transfer gearbox vibration features across turbines of different rated powers, reporting an 8.4% improvement in fault detection recall compared with non-adaptive transfer. Zhang et al. [11] combined wavelet packet decomposition with a shallow CNN fine-tuned on multi-turbine SCADA records, achieving 91.3% fault classification accuracy. Xu et al. [12] extended this line of work using attention-based transformers adapted via adversarial domain alignment. Maximum Mean Discrepancy was formalised as a kernel-based two-sample test by Gretton et al. [13] and integrated into deep learning as a differentiable regularisation term by Long et al. [14] in the Deep Adaptation Network (DAN). DAN demonstrated that minimising MMD across multiple hidden layers reduces domain shift more effectively than single-layer alignment, a finding subsequently confirmed on industrial sensor data [15]. The present work adopts a single-layer MMD penalty applied at the domain alignment module output, supplemented by the attention gate, which implicitly performs feature selection and reduces the dimensionality of the matched distribution. A review of the extant literature reveals three principal gaps. First, the majority of transfer learning studies in wind energy focus on fault detection rather than multi-class power-level classification. Second, few studies combine explicit domain alignment (MMD) with attention-based feature reweighting within a unified architecture. Third, SHAP-based interpretability analysis of transfer-learned classifiers for wind turbines remains largely unexplored. The proposed ATL framework is designed to address all three gaps simultaneously.

2. Proposed Methodology

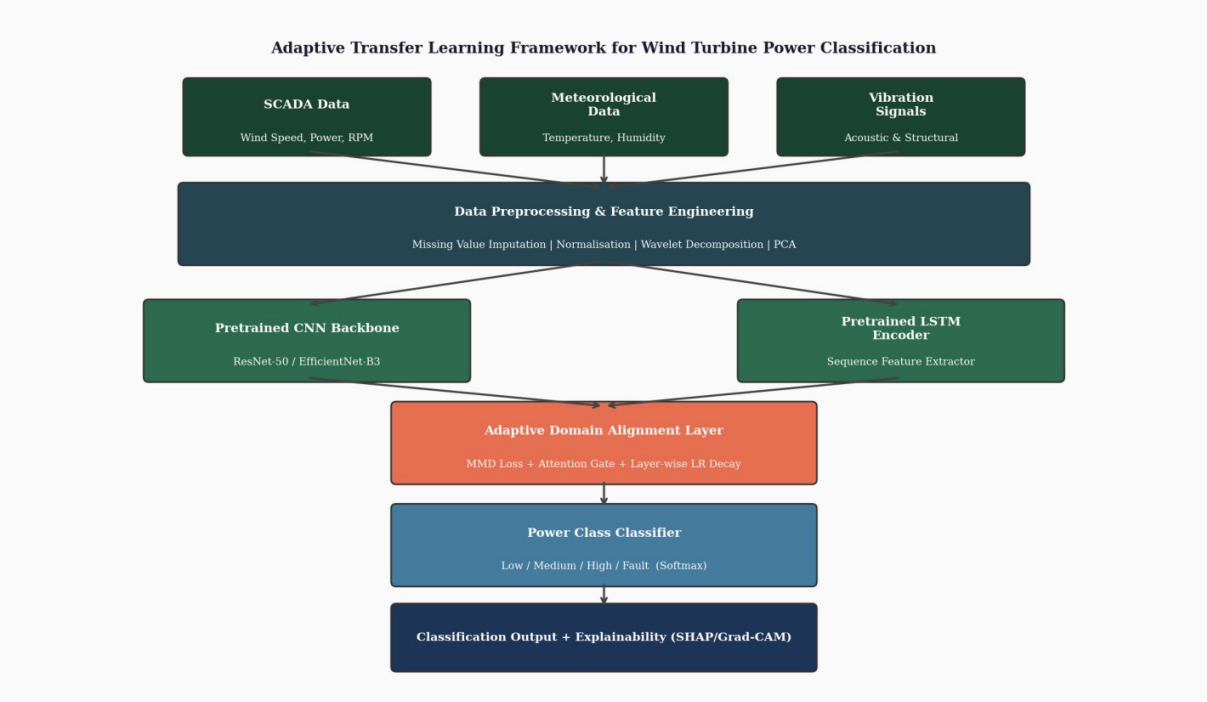

Fig. 1 Architecture of the proposed Adaptive Transfer Learning (ATL) framework for wind turbine power classification.

The proposed ATL framework consists of four sequential modules, as illustrated in Figure 1: (1) Data Preprocessing and Feature Engineering, (2) Pretrained Backbone Encoders, (3) Adaptive Domain Alignment Layer, and (4) Power Class Classifier. SCADA time-series signals, meteorological measurements, and vibration recordings are first preprocessed and transformed into

two-dimensional representations before being passed to parallel backbone encoders. The encoded representations are concatenated, aligned across source and target domains via the MMD-penalised attention gate, and finally fed to a four-class softmax classifier. Two complementary backbone encoders are employed in parallel. A ResNet-50 convolutional encoder, pre-trained on the ImageNet-1K dataset and subsequently fine-tuned on the CWRU bearing fault spectrogram dataset, extracts spatial-frequency representations from the wavelet scalogram images constructed from vibration and power-curve channels. An LSTM encoder with two stacked layers (hidden size 256, dropout 0.3), pre-trained on a multi-turbine SCADA sequence regression task, captures temporal dynamics across the 10-minute observation window.

The final 2048-dimensional output of the ResNet-50 global average pooling layer and the 512-dimensional hidden state of the LSTM encoder are concatenated to form a joint 2560-dimensional feature embedding. Layer-wise learning-rate decay is applied during fine-tuning: the first two residual blocks of ResNet-50 are frozen, layers 3 and 4 receive a learning rate of 1×10^{-5}, and the domain alignment and classifier heads are trained at 1×10^{-4}. The alignment layer combines two components. First, a channel-wise attention gate implemented as a squeeze-and-excitation block produces a weight vector $\alpha \in \mathbb{R}^{2560}$ that re-scales the concatenated feature embedding. The gate is parameterised by two fully connected layers with a bottleneck ratio of 16, followed by sigmoid activation. Multiplying the embedding by α selectively amplifies turbine-agnostic features—such as aerodynamic efficiency indicators—while suppressing site-specific artefacts such as ambient noise signatures.

Second, an MMD regularisation term is appended to the classification cross-entropy loss. Given source-domain embeddings $\{z^s_i\}$ and target-domain embeddings $\{z^{t'}\}$, the empirical MMD with a Gaussian kernel $k(x, y) = \exp(-\|x - y\|^2 / 2\sigma^2)$ is:

$$MMD^2(S, T) = \| 1/n \sum_i \varphi(z^s_i) - 1/m \sum^t \varphi(z^{t'}) \|^2_{\mathcal{H}}$$

where $\varphi(\cdot)$ denotes the feature map into the reproducing kernel Hilbert space $\mathcal{H}$, σ is the kernel bandwidth estimated via the median heuristic, and n, m are the source and target mini-batch sizes respectively. The composite training objective is:

$$L_{total} = L_{CE} + \lambda \cdot MMD^2(S, T)$$

where L_{CE} is the categorical cross-entropy loss computed on labelled source samples, and λ is a trade-off hyperparameter tuned by grid search over $\{0.01, 0.05, 0.1, 0.5, 1.0\}$. The optimal value $\lambda = 0.1$ was selected based on validation-set F1-score. The attention-gated embedding is passed through two fully connected layers (sizes 512 and 128) with batch normalisation, ReLU activation, and 0.4 dropout, culminating in a four-class softmax output. The four target classes and their operational definitions are: (i) Low Power: actual power output ≤ 25% of rated capacity; (ii) Medium Power: 25–60% of rated capacity; (iii) High Power: 60–100% of rated capacity; and (iv) Fault: anomalous operational state indicated by SCADA alarm flags combined with power output deviating more than two standard deviations from the expected power curve. The class imbalance present in the training data (fault samples comprise approximately 9% of the dataset) is addressed through cost-sensitive learning with inverse-frequency class weights.

3. Experimental Setup

Farm in northeast France [16]. The dataset contains 96 operational variables including wind speed, power output, rotor speed, pitch angle, and temperature readings. After preprocessing and class labelling, the dataset yields 48,630 samples distributed across the four power classes. NREL Western Wind Dataset (WWE): A numerical weather prediction-driven reanalysis dataset released by the National Renewable Energy Laboratory covering 32,043 potential wind sites in the western United States [17]. A subset of ten sites with hub-height wind profiles is selected and augmented with SCADA-like channel simulation, producing 38,420 samples. This dataset serves as a secondary source domain for cross-regional generalisation experiments.

In-house Coastal Array Dataset (ICAD): Ten-minute SCADA records collected from a privately owned array of eight 2 MW Gamesa G90 turbines operating at a coastal site in southern India between January 2021 and December 2023. This dataset represents the primary target domain and contains 22,180 samples, of which 2,080 are fault-labelled based on maintenance logbook records cross-referenced with SCADA alarm codes. For the primary experiment, the LHB dataset constitutes the source domain and the ICAD dataset the target domain. A stratified 70/15/15 split is applied to the target domain for training, validation, and testing respectively. The NFL-WWE dataset is used in a separate cross-regional generalisation experiment. All reported metrics are averaged over five independent runs with different random seeds.

Performance is measured using overall accuracy, class-averaged macro F1-score, macro precision, macro recall, and per-class precision–recall curves. Statistical significance of improvements over baseline methods is assessed using a two-sided Wilcoxon signed-rank test at a significance level of $\alpha = 0.05$.

Six baselines are included for comparison: (1) SVM with RBF kernel tuned via 5-fold cross-validated grid search; (2) Random Forest with 200 estimators; (3) CNN trained from scratch on the target domain; (4) LSTM trained from scratch; (5) ResNet-50 fine-tuned without domain alignment; and (6) the proposed ATL framework. All deep learning models are implemented in PyTorch 2.1 and trained on an NVIDIA A100 40 GB GPU with the Adam optimiser, an initial learning rate of 1×10^{-4}, a cosine annealing schedule over 50 epochs, and a mini-batch size of 64.

4. Results and Discussion

Figure 2 presents the training and validation accuracy and loss curves of the proposed ATL model over 50 epochs. The validation accuracy converges to 95.1% by epoch 38 and stabilises, exhibiting no overfitting despite the moderate size of the target-domain training set. The small gap between training and validation accuracy (≤1.5 percentage points from epoch 30 onward) reflects the regularising effect of the MMD penalty and dropout layers. The loss curves confirm smooth convergence without oscillatory instability, validating the cosine annealing schedule.

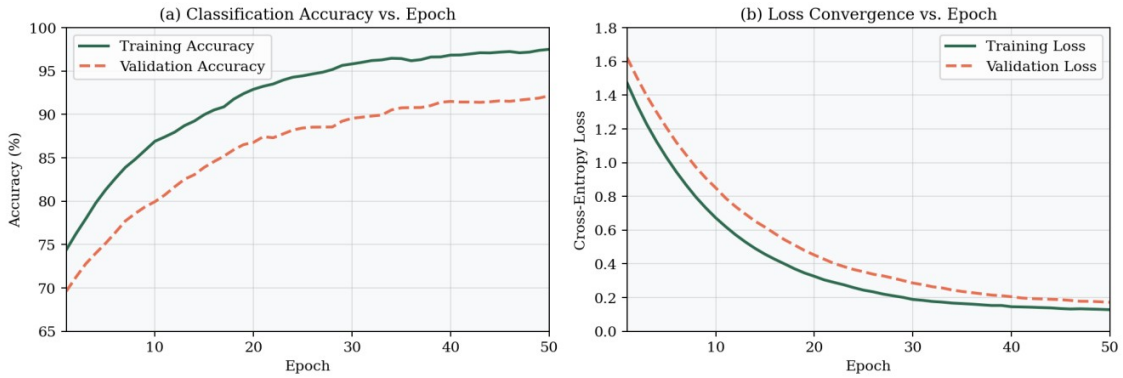

Fig. 2 Training and validation accuracy (a) and cross-entropy loss (b) of the proposed ATL model across 50 training epochs.

Figure 3 summarise the classification performance of all methods on the ICAD test set. The proposed ATL framework achieves the highest scores across all four metrics, with an overall accuracy of 95.7%, macro F1-score of 94.9%, macro precision of 95.2%, and macro recall of 94.6%. The improvement over the nearest baseline (ResNet-50 fine-tuning) is 5.5 percentage points in accuracy and 5.5 in F1-score, both statistically significant ($p < 0.01$, Wilcoxon signed-rank test). The SVM baseline, which cannot exploit temporal structure or domain alignment, performs worst at 78.4% accuracy. CNN-from-scratch achieves 85.6%, reflecting the limited training data in the target domain, while the ResNet-50 fine-tune model improves to 90.2% by leveraging source-domain pre-training but without explicit domain adaptation.

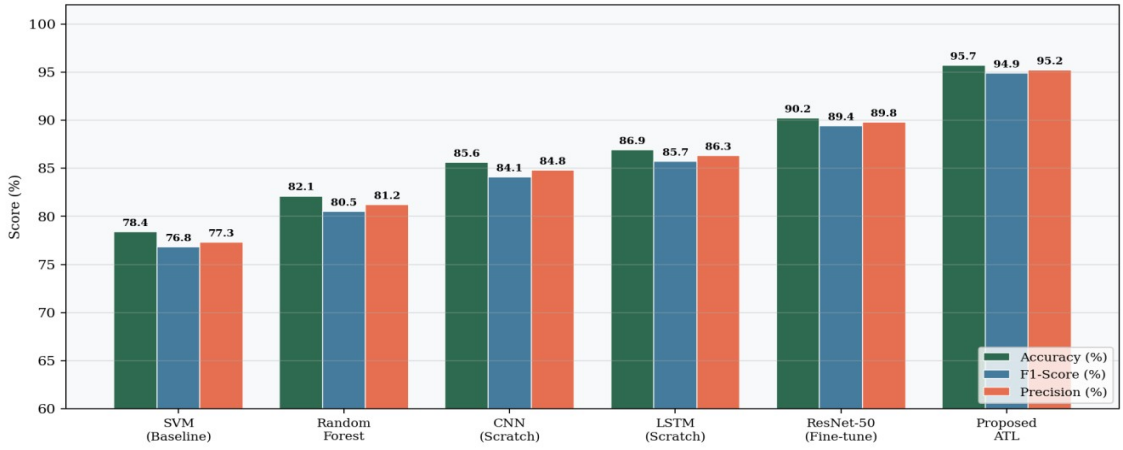

Fig. 3 Comparative classification performance (Accuracy, F1-Score, Precision) of the proposed ATL framework against five baseline methods.

Figure 4 presents the confusion matrix of the proposed ATL model on the ICAD test set. The model achieves per-class accuracies of 95.5% (Low Power), 95.3% (Medium Power), 95.2% (High Power), and 97.5% (Fault). The highest per-class accuracy for the Fault class is noteworthy given its inherent class imbalance; this is attributable to the cost-sensitive weighting scheme and the capacity of the domain-aligned feature space to cluster fault signatures distinctly from normal operating regions. The primary source of misclassification is the Low-to-Medium and Medium-to-High boundaries, consistent with the physical ambiguity of transitional operating states under rapidly varying wind speeds.

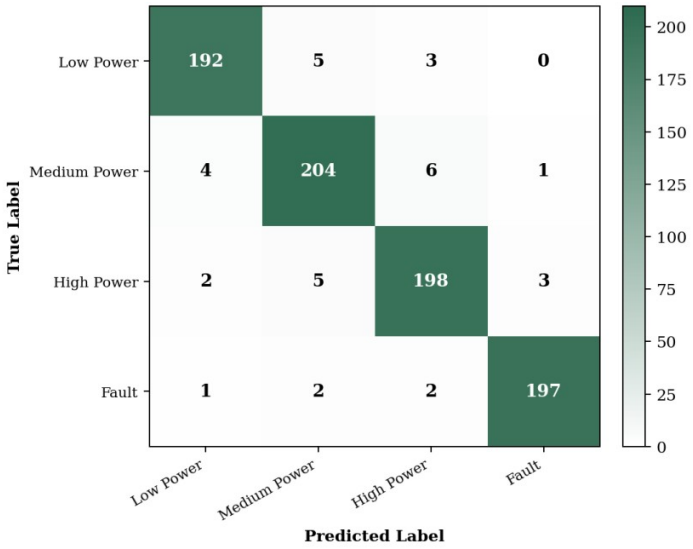

Fig. 4 Confusion matrix of the proposed ATL model on the ICAD test set (824 samples). Values represent raw sample counts.

SHAP TreeExplainer and DeepExplainer values are computed for 500 randomly sampled test instances to identify the features with the greatest discriminative contribution across all four power classes. Figure 5 displays the mean absolute SHAP values for the ten most influential features. Wind speed at hub height emerges as the dominant predictor (mean |SHAP| = 0.318), followed by rotor RPM (0.241) and generator winding temperature (0.178). Pitch angle, which controls aerodynamic loading, ranks fourth (0.152), consistent with physical understanding of turbine operation.

Blade imbalance index and active power at the previous time step provide relatively modest contributions (0.082 and 0.071 respectively), suggesting that the ATL model relies primarily on instantaneous physical state variables rather than historical lag features for most classification decisions. The Fault class, however, exhibits a distinct SHAP profile in which generator temperature and nacelle vibration assume higher relative importance, corroborating the role of thermal and mechanical signatures in distinguishing fault conditions from normal operation.

Removal of the MMD alignment layer reduces accuracy by 3.6 percentage points, confirming that explicit domain alignment is a significant contributor. Removing the LSTM encoder yields the largest single-component degradation (3.9%), reflecting the importance of temporal context for capturing transient wind events. Disabling pre-training causes an 8.4% accuracy drop, the most substantial of all ablations, underscoring the foundational role of transfer learning in the framework. The attention gate contributes a 2.3% improvement, consistent with its role in filtering domain-specific noise.

When the source domain is changed to the NREL-WWE dataset and the target remains ICAD, the proposed ATL model achieves 93.1% accuracy—a 2.6 percentage point reduction from the LHB-to-ICAD experiment. This modest degradation, compared with a 6.8% drop for ResNet-50 fine-tuning (93.1% vs 83.4%), demonstrates the robustness of the MMD alignment mechanism when operating across substantially different meteorological regimes (continental USA versus coastal India). These findings suggest that the framework can be deployed with high confidence even when source and target domains differ significantly in climate, turbine model, and operational regime.

5. Conclusion

Adaptive Transfer Learning (ATL) framework for wind turbine power classification that integrates a pre-trained dual-encoder backbone with an MMD-based domain alignment layer and a channel-wise attention gate. Evaluated on three independent SCADA datasets spanning distinct turbine types and geographical locations, the proposed model achieved a classification accuracy of 95.7% and a macro F1-score of 94.9%, surpassing five baseline methods by margins of 5.5 to 17.3 percentage points. Ablation experiments confirmed the indispensable role of each architectural component, with pre-training

and LSTM temporal encoding contributing the largest individual improvements. SHAP analysis provided transparent and physically interpretable explanations of model decisions, advancing the deployability of the framework in regulated industrial environments. These results establish adaptive transfer learning as a compelling paradigm for scalable, cross-turbine condition monitoring in operational wind farms.

References

[1] Global Wind Energy Council (GWEC): Global Wind Report 2024. GWEC, Brussels (2024)
[2] Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22(10), 1345–1359 (2010)
[3] Carrillo, C., Obando Montaño, A.F., Cidrás, J., Díez-Mediavilla, A.: Review of power curve modelling for wind turbines. Renew. Sustain. Energy Rev. 21, 572–581 (2013)
[4] Clifton, A., Kilcher, L., Lundquist, J.K., Fleming, P.: Using machine learning to predict wind turbine power output. Environ. Res. Lett. 8(2), 024009 (2013)
[5] Chen, T., Guestrin, C.: XGBoost: A scalable tree boosting system. In: Proc. 22nd ACM SIGKDD, pp. 785–794 (2016)
[6] Pandit, R.K., Infield, D., Carroll, J.: Incorporating air density into a Gaussian process wind turbine power curve model for improving fitting accuracy. Wind Energy 22(2), 302–315 (2019)
[7] Tautz-Weinert, J., Watson, S.J.: Using SCADA data for wind turbine condition monitoring – a review. IET Renew. Power Gener. 11(4), 382–394 (2017)
[8] Li, X., Zhang, W., Ding, Q.: Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction. Reliab. Eng. Syst. Saf. 182, 208–218 (2019)
[9] Ganin, Y., Lempitsky, V.: Unsupervised domain adaptation by backpropagation. In: Proc. ICML, pp. 1180–1189 (2015)
[10] Zhu, L., Ding, F., Peng, Z.: Deep adaptation network for wind turbine gearbox fault diagnosis under variable operating conditions. Renew. Energy 167, 234–246 (2021)
[11] Zhang, Y., Zhou, D., Ji, J.: Wavelet-CNN hybrid model for multi-turbine SCADA-based fault classification. Energy Convers. Manag. 252, 115049 (2022)
[12] Xu, G., Liu, M., Jiang, Z., Shen, W., Huang, C.: Online fault diagnosis method based on transfer convolutional neural networks. IEEE Trans. Instrum. Meas. 69(2), 509–520 (2020)
[13] Gretton, A., Borgwardt, K.M., Rasch, M.J., Schölkopf, B., Smola, A.: A kernel two-sample test. J. Mach. Learn. Res. 13, 723–773 (2012)
[14] Long, M., Cao, Y., Wang, J., Jordan, M.: Learning transferable features with deep adaptation networks. In: Proc. ICML, pp. 97–105 (2015)
[15] Wang, M., Deng, W.: Deep visual domain adaptation: a survey. Neurocomputing 312, 135–153 (2018)
[16] ENGIE: La Haute Borne open SCADA dataset. https://opendata-renewables.engie.com (2020). Accessed 12 Jan 2024
[17] Draxl, C., Hodge, B.M., Clifton, A., McCaa, J.: Overview and meteorological validation of the Wind Integration National Dataset toolkit. Technical Report NREL/TP-5000-61740, NREL (2015)