IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Journal Européen des Systèmes Automatisés
Vol. 59, No. 10, October, 2026, pp. 2927-2932
Journal homepage: http://iieta.org/journals/jesa

Comparative Benchmarking of Solar Power Forecasting and 17-Class Power Quality Disturbance Recognition with Grad-CAM Analysis in Renewable-Integrated Grids

Manoj M. Kharade*, Umesh T. Kute

School of Engineering and Technology, Pimpri Chinchwad University, Pune 412106, India

Corresponding Author Email: manoj.kharade23@pcu.edu.in

Copyright: ©2026 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/).

https://doi.org/10.18280/jesa.591014

Received: 2 August 2026
Revised: 16 September 2026
Accepted: 20 October 2026
Available online: 31 October 2026

Keywords:
AI, gradient-weighted class activation mapping, power quality disturbance recognition, Random Forest, solar power forecasting, time-series forecasting, transformer

ABSTRACT

Renewable-rich power systems require both anticipatory generation awareness and event-level power-quality diagnosis. This study benchmarks solar power forecasting and 17-class disturbance recognition as complementary information layers for renewable-grid monitoring. Plant-level solar active power is formed from 22 inverters at 15-min resolution and evaluated using a leakage-controlled 70/15/15 chronological split with a 48-step lookback and 12-step (3-h) direct forecast horizon. The corrected benchmark compares persistence, moving average, linear regression, recurrent, convolutional and Transformer-family models. Standard Transformer achieved the lowest pooled root mean square error (2,813.53 kW) and highest R^2 (0.8672), whereas PatchTST achieved the lowest mean absolute error (1,600.51 kW). For 17-class power-quality recognition, Random Forest achieved 67.84% accuracy and 67.89% macro-F1, while the 1D convolutional neural network (1D-CNN) was the strongest deep classifier. Random Forest used 14 handcrafted statistical, waveform and spectral features, whereas deep models used raw 100-point waveforms. A deterministically reproduced 1D-CNN enabled authentic gradient-weighted class activation mapping (Grad-CAM) analysis on five held-out examples, revealing localized attribution for transient events and broader attribution for envelope and harmonic disturbances. The results support monitoring-oriented integration without claiming closed-loop control.

1. INTRODUCTION

Renewable generation changes the operating envelope of distribution networks through short-term variability and converter-mediated waveform effects. Solar and wind output depend on weather and operating conditions, while renewable interfaces can coexist with harmonics, voltage events, transients and compound power-quality disturbances [1-4]. Forecasting and disturbance recognition therefore answer different but complementary questions: the first provides anticipatory context before an operating interval, while the second diagnoses waveform events when they occur.

In the revised system view, the two modules feed a common monitoring and supervisory decision-support layer rather than forming a validated closed-loop controller. A three-hour solar forecast can inform expected renewable operating context, while the disturbance classifier provides event-level diagnostic information for logging, prioritisation and potential downstream control decisions. Figure 1 summarizes this information flow. Hardware actuation, communication delay and closed-loop control are outside the experimental scope.

Power-quality recognition depends on disturbance construction, signal representation and validation protocol [5-8]. Explainability research in smart grids similarly emphasizes fidelity, stability and deployment context rather than heatmap appearance alone [9, 10]. Renewable forecasting performance is sensitive to horizon, temporal resolution, feature construction and baseline choice [11-18]. Long short-term memory (LSTM), gated recurrent unit (GRU), temporal convolutional network (TCN) and Transformer-family models provide different mechanisms for sequential dependence [19-25].

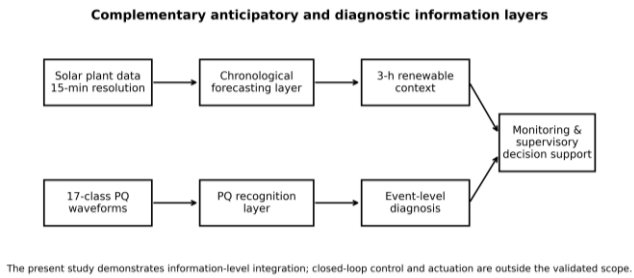


Figure 1. Complementary forecasting and power quality (PQ) diagnosis layers within a renewable-grid monitoring architecture

This work makes four contributions. First, it reports a corrected, leakage-controlled multi-horizon solar forecasting benchmark using the final processing and evaluation protocol. Second, it evaluates classical and deep classifiers on a common 17-class power-quality test split and explicitly

distinguishes handcrafted-feature and raw-waveform representations. Third, it evaluates authentic gradient-weighted class activation mapping (Grad-CAM) explanations on five held-out examples from a reproduced 17-class 1D convolutional neural network (1D-CNN). Fourth, it frames forecasting and diagnosis as complementary information layers for renewable-grid monitoring while preserving the boundaries of the software-only evidence.

2. RELATED WORK

Power-quality disturbance recognition has progressed from transform-based handcrafted features to direct and hybrid learned representations. Reviews identify dataset realism, validation design, and deployment evidence as persistent limitations [1-4]. Recent photovoltaic-integrated and deep-learning studies report strong within-study accuracy, but their class generators, signal representations, noise protocols, and validation procedures differ [5-8]. Smart-grid explainability research similarly emphasizes fidelity, stability and deployment context rather than heatmap appearance alone [9, 10].

Renewable forecasting reviews reach a similar conclusion: there is no universally best architecture across resources and horizons [11-18]. Informer, Temporal Fusion Transformer (TFT), PatchTST and temporal convolutional networks address sequence dependence differently [19-22], while LSTM, GRU and the standard Transformer remain important references [23-25]. The present revision therefore emphasizes common splits, transparent baselines and corrected test evaluation rather than architecture-first claims.

3. DATA AND METHODOLOGY

3.1 Renewable and power-quality data

The solar branch uses Plant 1 from the public Solar Power Generation Data collection, a 34-day two-plant dataset also used in recent photovoltaic forecasting work [26]. The raw generation file contains 68,778 inverter-level records over 34 days at 15-min resolution. For the corrected forecasting task, active power is aggregated across 22 physical inverters at each timestamp, producing a strictly ordered plant-level series of 3,158 timestamps from 15 May to 17 June 2020. The single forecasting target is total plant alternating-current active power in kilowatts. Wind supervisory control and data acquisition (SCADA) data contain 50,530 ten-minute observations and are retained only as a brief contextual example of a second renewable variability regime; no numerical wind forecast is claimed.

The 17-class power-quality benchmark evaluates 1,700 waveforms, 100 per class, from a larger synthetic benchmark collection. The classes include individual and compound disturbances such as sag, swell, harmonics, flicker, interruption, notch, oscillatory transient and their compound variants. Each waveform contains 100 points sampled at 5 kHz (20 ms) and is independently scaled to [-1,1]. The original 70/15/15 random split contains 1,190 training, 255 validation and 255 test samples. The split is fixed by seed 42 and is not stratified; exact per-class counts are reported in Table 1. No dataset-level normalization statistics are fitted, so no train-test leakage is introduced through scaling.

Table 1. Per-class power-quality distribution in the fixed seed-42 split (100 waveforms per class)

Class	Train	Val.	Test
Flicker	75	16	9
Flicker + Sag	68	15	17
Flicker + Swell	64	15	21
Harmonics	68	15	17
Harmonics + Sag	72	10	18
Harmonics + Swell	72	16	12
Interruption	74	14	12
Notch	69	17	14
Osc. transient	71	15	14
Pure sinusoid	73	10	17
Sag	65	12	23
Sag + Harmonics	72	16	12
Sag + Osc. transient	67	20	13
Swell	70	14	16
Swell + Harmonics	71	14	15
Swell + Osc. transient	66	22	12
Transient	73	14	13
Total	1190	255	255

3.2 Corrected forecasting protocol

The corrected solar experiment first aggregates inverter output by timestamp, then performs chronological separation before window construction. The lookback length is 48 observations (12 h), the direct forecast horizon is 12 observations (3 h), and stride is one. A StandardScaler is fitted only on the training subset (training mean 6,873.30 kW; standard deviation 8,767.14 kW) and then applied unchanged to validation and test data. Prediction metrics are computed after inverse transformation to kilowatts. Neural forecasting models use AdamW-based optimization with validation-loss scheduling and model-specific checkpoints; the revision experiment preserves the exact configuration archive and separates each checkpoint to prevent cross-model overwrite. The chronological solar data split and sample windows are summarized in Table 2, while the complete experimental hyperparameter settings and model-specific configurations are detailed in Tables 3(a) and (b), respectively.

Table 2. Chronological solar data split used in the corrected experiment

Subset	Date Range	Raw	Windows
Train	15 May 00:00 - 8 Jun 02:15	2,210	2,151
Validation	8 Jun 02:30 - 13 Jun 00:45	474	415
Test	13 Jun 01:00 - 17 Jun 23:45	474	415

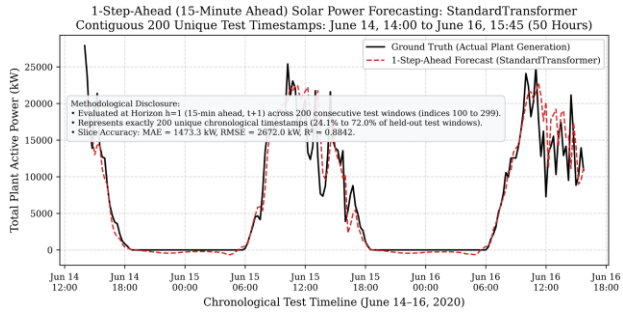


Figure 2. Corrected 1-step-ahead (15-min-ahead) Standard Transformer forecast over 200 unique chronological test timestamps (windows 100-299; 14 June 2020 14:00 to 16 June 2020 15:45)

The principal forecasting table reports pooled multi-horizon performance over 415 test windows and 12 lead times (4,980 evaluated predictions). The corrected 200-point visualization is instead defined at the t + 1 lead time and covers 200 unique consecutive timestamps from 14 June 2020 14:00 to 16 June 2020 15:45, corresponding to test windows 100-299. For this visualization slice, MAE is 1,473.3 kW, RMSE is 2,672.0 kW and R^2 is 0.8842; these slice metrics are not used to rank the multi-horizon models. As illustrated in Figure 2, the 1-step-ahead Standard Transformer forecast tracks diurnal generation closely across this contiguous 50-hour test sequence.

Table 3. Verified forecasting configuration used in the corrected revision experiment

(a) Common settings

Setting	Verified Value
Target	Plant AC active power (kW)
Temporal	L = 48 (12 h); H = 12 (3 h); stride = 1
Input/seed	1/42
Scaling	Train-only StandardScaler
Neural training	AdamW; lr = .001; batch = 32; MSE
Control	Max 30 epochs; patience = 7
Baselines	Persistence; moving avg.; Ridge-SVD
Source	Revision experiment configuration

(b) Model-specific settings

Model	Epochs	Hid./layers	Params
LSTM	12/30	32/2	13,324
BiLSTM	24/30	32/2	34,828
GRU	24/30	32/2	10,092
TCN	30/30	16/1	1,052
Std. Transformer	14/30	32/2	25,868
Informer	16/30	32/1	4,684
PatchTST	27/30	32/2	26,348
TFT	21/30	32/1	11,020

Note: long short-term memory (LSTM); BiLSTM = bidirectional long short-term memory; gated recurrent unit (GRU), temporal convolutional network (TCN); TFT = temporal fusion transformer.

3.3 Power-quality classifiers and explainability

Classical models, including Random Forest [27], use a 14-dimensional handcrafted feature vector comprising mean, standard deviation, variance, skewness, kurtosis, root mean square, peak, crest factor, shape factor, impulse factor, energy, spectral energy, spectral centroid and spectral entropy. Deep classifiers, including 1D-CNN, ResNet1D [28], CNN-LSTM, CNN-GRU, TemporalCNN, Transformer and hybrid CNN-Transformer, receive the raw 100-point waveform. All models use the same held-out test indices. Consequently, the Random Forest comparison is valid at the task level but representation-different rather than a pure architecture-only contest.

The original deep models were trained for three epochs under a constrained benchmark budget. For explainability, the 1D-CNN protocol was deterministically reproduced using the same 1,700-sample subset, seed, split, architecture, AdamW optimizer (learning rate 0.001, weight decay 0.0001), batch size 32, class-weighted cross-entropy loss and three epochs. The reproduced model achieved 48.63% accuracy and macro-F1 of 0.4487, closely matching the reported 48.24% and 0.4478 values. Grad-CAM [29] uses the final convolutional layer (conv3) and is evaluated on the first correctly classified held-out sample for each requested class.

3.4 Evaluation measures

Forecasting is summarized by mean absolute error (MAE), root mean square error (RMSE) and coefficient of determination (R^2). Percentage error is not used as a primary ranking metric because solar output includes zero and near-zero periods. Power-quality classification is summarized by accuracy, balanced accuracy, macro-precision, macro-recall, macro-F1, Matthews correlation coefficient and Cohen's kappa. Class-balanced measures are emphasized because the non-stratified test split has varying class counts.

4. RESULTS

4.1 Corrected solar forecasting benchmark

The corrected benchmark materially changes the interpretation of the solar experiment. Standard Transformer achieved the lowest pooled RMSE (2,813.53 kW) and highest pooled R^2 (0.8672), whereas PatchTST achieved the lowest MAE (1,600.51 kW). TFT, BiLSTM, GRU and LSTM also produced valid positive R^2 values. These corrected results replace the earlier forecasting values because the final experiment uses a consistent plant-level target, chronological split, training-only scaling and model-specific checkpoints. The comparative performance metrics across all evaluated solar forecasting models are presented in Table 4.

Table 4. Corrected pooled multi-horizon solar forecasting benchmark

Model	MAE (kW)	RMSE (kW)	R^2
Standard Transformer	1,618.60	2,813.53	0.8672
PatchTST	1,600.51	2,880.79	0.8607
TFT	1,757.83	2,893.30	0.8595
BiLSTM	1,647.25	2,915.93	0.8573
GRU	1,785.30	3,076.05	0.8412
LSTM	1,978.01	3,191.28	0.8291
Linear regression	2,750.54	3,828.59	0.7540
TCN	3,092.37	3,987.58	0.7332
Persistence	3,440.10	5,715.17	0.4519
Informer	5,545.46	6,880.21	0.2057
Moving average	5,160.32	7,591.84	0.0328

Notes: TCN = temporal convolutional network; TFT = temporal fusion transformer; BiLSTM = bidirectional long short-term memory; GRU = gated recurrent unit; LSTM = long short-term memory.

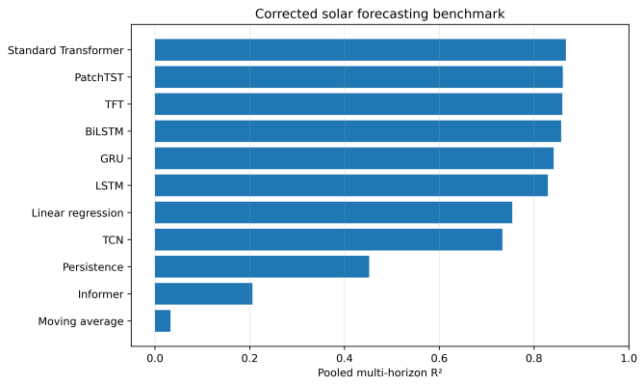


Figure 3. Corrected pooled multi-horizon solar forecasting benchmark ranked by coefficient of determination (R^2)

Horizon-specific decomposition further shows why the multi-horizon formulation matters. At t + 1 (15 min),

persistence remains competitive (R^2 = 0.9038), but at t + 12 (3 h) it falls to R^2 = -0.1628. Standard Transformer retains R^2 = 0.8696 at t + 12, while PatchTST retains R^2 = 0.8594. This study therefore reports the stronger long-horizon performance of the Transformer-based models in this experiment without claiming universal architectural superiority. The ranking of all evaluated forecasting models by their pooled multi-horizon R^2 is illustrated in Figure 3.

4.2 Power-quality disturbance recognition

Random Forest remains the strongest overall classifier with 67.84% accuracy, 68.65% balanced accuracy and 67.89% macro-F1. The strongest deep classifier is 1D-CNN at 48.24% accuracy and 44.78% macro-F1. This gap is interpreted in the context of representation and training budget: Random Forest receives 14 engineered statistical, waveform and spectral descriptors, whereas the deep models must learn directly from 100-point waveforms and were limited to three epochs. The result is therefore specific to the evaluated representation and compute budget rather than evidence that Random Forest is universally superior to deep learning. The quantitative results for all ten disturbance classifiers are presented in Table 5. In addition, the macro-F1 comparison across models is shown in Figure 4, and the normalized confusion matrix for the 1D-CNN classifier is displayed in Figure 5.

Table 5. Power-quality disturbance classifier benchmark

Model	Acc.	Bal. Acc.	Macro-F1
Random Forest	0.678	0.687	0.679
1D-CNN	0.482	0.481	0.448
ResNet1D	0.420	0.432	0.411
Logistic regression	0.427	0.450	0.407
CNN-LSTM	0.357	0.354	0.299
CNN-GRU	0.365	0.346	0.276
SVM	0.216	0.237	0.156
Hybrid CNN-Transformer	0.216	0.243	0.126
Transformer	0.149	0.174	0.088
Temporal CNN	0.114	0.132	0.050

Notes: CNN = convolutional neural network; LSTM = long short-term memory; GRU = gated recurrent unit; SVM = support vector machine.

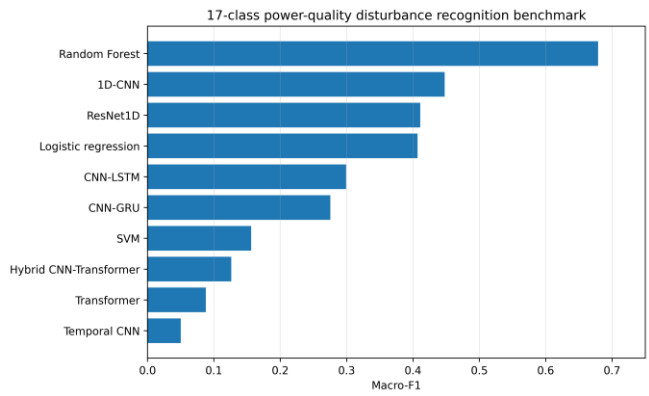


Figure 4. Macro-F1 comparison for the 17-class power-quality (PQ) disturbance benchmark

4.3 Authentic multi-sample explainability evidence

Authentic explainability evidence was generated from the deterministically reproduced 17-class 1D-CNN checkpoint. Five held-out, correctly classified examples were selected by a deterministic rule: the first correctly classified test sample in test-index order for Sag, Harmonics, Transient, Sag with Harmonics and Swell. The reproduced checkpoint classified 124 of 255 test samples correctly, compared with 123 of 255 in the original benchmark; macro-F1 differed by only 0.0009. The authentic 1D Grad-CAM saliency overlays and attribution distributions for these five representative disturbance categories are illustrated in Figure 6.

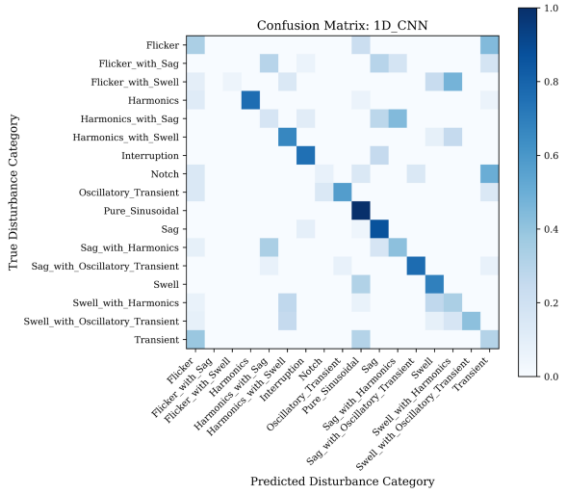


Figure 5. Normalized confusion matrix for the 1D convolutional neural network (1D-CNN) across 17 disturbance classes

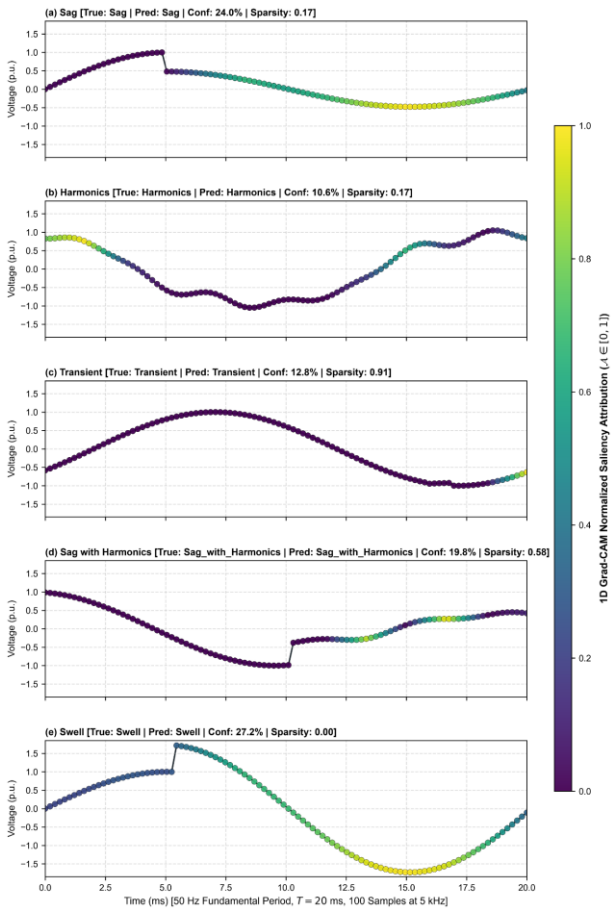


Figure 6. Authentic 1D gradient-weighted class activation mapping (Grad-CAM) saliency overlays for five correctly classified held-out disturbances from the reproduced 1D convolutional neural network (1D-CNN); color encodes normalized attribution

Attribution is disturbance-dependent. The Transient example concentrates saliency within a localized waveform interval, Sag with Harmonics shows a narrower mid-to-late emphasis, and Sag, Harmonics and Swell show broader cycle-distributed attribution. These maps indicate where the neural network assigns importance; they do not establish electrical causality or prove that the highlighted region is physically sufficient to define the disturbance.

5. SYSTEMS AND DEPLOYMENT DISCUSSION

The revised system interpretation separates anticipatory context from event diagnosis. At 15-min resolution, the forecasting branch provides a direct 12-step, three-hour view of expected solar power. The power-quality branch operates on a 20-ms waveform window and produces a disturbance label for event-level monitoring. Their outputs can be consumed by a supervisory layer for context-aware logging, alarm prioritisation and future control decisions, but no field communication or actuation loop is claimed.

Authentic Grad-CAM explanations required approximately 2.64-3.01 ms per selected sample in the recorded software setup. This value is explanation runtime, not end-to-end classifier latency. The latter, together with communication delay, hardware latency and actuator response, was not independently benchmarked. The shallow 1D-CNN and compact waveform length are compatible with online software processing in principle, but real-time deployment remains a future validation task.

The wind data are deliberately compressed to contextual evidence only. Their 50,530 ten-minute observations illustrate a second renewable variability regime, but the study does not introduce an unverified wind forecasting score. This keeps the quantitative claims aligned with the experiments actually evaluated.

6. LIMITATIONS

The solar benchmark is based on one public plant and one corrected chronological split; repeated-run confidence intervals were not used to claim statistical superiority. The power-quality experiment evaluates 1,700 of the available synthetic signals and uses a non-stratified split. Classical and deep classifiers receive different representations, so their comparison is task-level rather than architecture-pure. The deep classifiers also use a limited three-epoch training budget. Grad-CAM is evaluated on authentic test samples but remains post-hoc model attribution, not a causal electrical explanation. Finally, end-to-end real-time deployment, communication reliability, and closed-loop control are outside the validated scope.

7. CONCLUSIONS

This revised study benchmarks solar forecasting and 17-class power-quality disturbance recognition as complementary anticipatory and diagnostic layers for renewable-integrated grid monitoring. After correcting the forecasting pipeline, Standard Transformer achieved the lowest pooled RMSE (2,813.53 kW) and highest pooled R^2 (0.8672), while PatchTST achieved the lowest MAE (1,600.51 kW). Random Forest remained the strongest overall power-quality classifier, but its advantage is interpreted in the context of a 14-feature handcrafted representation and the three-epoch raw-waveform deep-learning budget. A reproduced 1D-CNN supported authentic Grad-CAM analysis across five disturbances: transient attribution was strongly localized, compound Sag with Harmonics was partially localized, and envelope/harmonic cases were more distributed. The two analytical layers can jointly support monitoring and supervisory decision-making, while closed-loop control and field deployment remain future work.

DATA AVAILABILITY

The solar and wind data are public benchmark datasets. The 17-class power-quality benchmark used in this study is a synthetic disturbance collection evaluated through the study workflow. Corrected experiment tables, model checkpoints and figure outputs are retained in the author revision archive and can be supplied if requested by the editor.

REFERENCES

[1] Chawda, G.S., Shaik, A.G., Shaik, M., et al. (2020). Comprehensive review on detection and classification of power quality disturbances in utility grid with renewable energy penetration. IEEE Access, 8: 146807-146830. https://doi.org/10.1109/ACCESS.2020.3014732

[2] Dehaghani, M.N., Korotko, T., Rosin, A. (2025). AI applications for power quality issues in distribution systems: A systematic review. IEEE Access, 13: 18346-18365. https://doi.org/10.1109/ACCESS.2025.3533702

[3] Khetarpal, P., Tripathi, M.M. (2020). A critical and comprehensive review on power quality disturbance detection and classification. Sustainable Computing: Informatics and Systems, 28: 100417. https://doi.org/10.1016/j.suscom.2020.100417

[4] Yan, Y.J., Chen, K., Geng, H., Fan, W.Q., Zhou, X.R. (2023). A review on intelligent detection and classification of power quality disturbances: Trends, methodologies, and prospects. Computer Modeling in Engineering & Sciences, 137(2): 1345-1379. https://doi.org/10.32604/cmes.2023.027252

[5] Mishra, S., Mallick, R.K., Nayak, P., Ramasamy, T.N., Panda, G. (2024). Dynamic power quality disturbance classification in grid-integrated PV systems: Leveraging Clark transformed modal voltage and subspace weighted KNN. IEEE Access, 12: 116572-116586. https://doi.org/10.1109/ACCESS.2024.3443963

[6] Al-Mashdali, M.F., Islam, A., Hassan, A., Shafiullah, M., Al-Dhaifallah, M., Al Fuwail, K. (2025). Signal processing-free intelligent model for power quality disturbances identification. IEEE Access, 13: 9910-9922. https://doi.org/10.1109/ACCESS.2025.3528296

[7] Chen, L., Chen, S., Xu, J.J., Zhou, C. (2024). Power quality disturbances identification based on deep neural network model of time-frequency feature fusion. Electric Power Systems Research, 231: 110283. https://doi.org/10.1016/j.epsr.2024.110283

[8] Li, D., Channa, I.A., Chen, X., Dou, L., Khokhar, S., Azar, N.A. (2024). A new deep learning method for classification of power quality disturbances using DWT-

MRA in utility smart grid. Computers and Electrical Engineering, 117: 109290. https://doi.org/10.1016/j.compeleceng.2024.109290
[9] Alsaigh, R., Mehmood, R., Katib, I. (2023). AI explainability and governance in smart energy systems: A review. Frontiers in Energy Research, 11: 1071291. https://doi.org/10.3389/fenrg.2023.1071291
[10] Kandula, R., Dandamudi, J.T., Amar Raj, R.D., Yanamala, R.M.R., Prakasha, K.K. (2026). Explainable AI for smart grid intelligence: A comprehensive review of techniques, applications and future directions. Energy and AI, 25: 100820. https://doi.org/10.1016/j.egyai.2026.100820
[11] Ying, C.T., Wang, W.Q., Yu, J., Li, Q., Yu, D.H., Liu, J.H. (2023). Deep learning for renewable energy forecasting: A taxonomy and systematic literature review. Journal of Cleaner Production, 384: 135414. https://doi.org/10.1016/j.jclepro.2022.135414
[12] Yang, Y., Lou, H., Wu, J.R., Zhang, S.T., Gao, S.C. (2024). A survey on wind power forecasting with machine learning approaches. Neural Computing and Applications, 36: 12753-12773. https://doi.org/10.1007/s00521-024-09923-4
[13] Liu, H., Zhang, Z.J. (2024). Development and trending of deep learning methods for wind power predictions. Artificial Intelligence Review, 57: 112. https://doi.org/10.1007/s10462-024-10728-z
[14] Barhmi, K., Heynen, C., Golroodbari, S., van Sark, W. (2024). A review of solar forecasting techniques and the role of artificial intelligence. Solar, 4(1): 99-135. https://doi.org/10.3390/solar4010005
[15] Hussan, U., Wang, H.Z., Peng, J.C., Jiang, H., Rasheed, H. (2026). Transformer-based renewable energy forecasting: A comprehensive review. Renewable and Sustainable Energy Reviews, 226: 116356. https://doi.org/10.1016/j.rser.2025.116356
[16] Liu, J.R., Fu, Y.C. (2023). Renewable energy forecasting: A self-supervised learning-based transformer variant. Energy, 284: 128730. https://doi.org/10.1016/j.energy.2023.128730
[17] Sobri, S., Koohi-Kamali, S., Rahim, N.A. (2018). Solar photovoltaic generation forecasting methods: A review. Energy Conversion and Management, 156: 459-497. https://doi.org/10.1016/j.enconman.2017.11.019
[18] Antonanzas, J., Osorio, N., Escobar, R., Urraca, R., Martinez-de-Pison, F.J., Antonanzas-Torres, F. (2016). Review of photovoltaic power forecasting. Solar Energy, 136: 78-111. https://doi.org/10.1016/j.solener.2016.06.069
[19] Zhou, H.Y., Zhang, S.H., Peng, J.Q., et al. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. Proceedings of the AAAI Conference on Artificial Intelligence, 35(12): 11106-11115. https://doi.org/10.1609/aaai.v35i12.17325
[20] Lim, B., Arik, S.O., Loeff, N., Pfister, T. (2021). Temporal Fusion Transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting, 37: 1748-1764. https://doi.org/10.1016/j.ijforecast.2021.03.012
[21] Nie, Y.Q., Nguyen, N.H., Sinthong, P., Kalagnanam, J. (2023). A time series is worth 64 words: Long-term forecasting with transformers. arXiv preprint arXiv:2211.14730. https://doi.org/10.48550/arXiv.2211.14730
[22] Bai, S.J., Kolter, J.Z., Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv preprint arXiv:1803.01271. https://doi.org/10.48550/arXiv.1803.01271
[23] Hochreiter, S., Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8): 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735
[24] Cho, K., van Merriënboer, B., Gulcehre, C., et al. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, pp. 1724-1734. https://doi.org/10.3115/v1/D14-1179
[25] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.
[26] Nahar, A., Rudro, R.A.M., Al Sohan, M.F.A., Uddin, M.H., Kumar, L. (2025). Forecasting solar photovoltaic power generation: A machine learning time series model approach. International Journal of Energy Research, 2025(1): 4092367. https://doi.org/10.1155/er/4092367
[27] Breiman, L. (2001). Random forests. Machine Learning, 45: 5-32. https://doi.org/10.1023/A:1010933404324
[28] He, K.M., Zhang, X.Y., Ren, S.Q., Sun, J. (2016). Deep residual learning for image recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 770-778. https://doi.org/10.1109/CVPR.2016.90
[29] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, pp. 618-626. https://doi.org/10.1109/ICCV.2017.74

NOMENCLATURE

AI	artificial intelligence
C	number of disturbance classes, dimensionless
CNN	convolutional neural network
Grad-CAM	gradient-weighted class activation mapping
GRU	gated recurrent unit
LSTM	long short-term memory
MAE	mean absolute error
N	number of evaluated observations, dimensionless
PQ	power quality
R^2	coefficient of determination, dimensionless
RMSE	root mean square error
SCADA	supervisory control and data acquisition
TCN	temporal convolutional network
TFT	temporal fusion transformer
y_i	observed value, kW for forecasting
$\hat{y}_i$	predicted value, kW for forecasting