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. 2855-2864
Journal homepage: http://iieta.org/journals/jesa

Robustness-Cost Trade-Offs in Power Quality Disturbance Classification: A Compact Noise-Resilient Convolutional Neural Network Versus Classical Learner

Muna Hassan Hussien[1*], Azam Isam Aladwani[1], Mohammed Omar Mafoodh[2], Osamah Basheer[1]

[1] Technical College of Engineering/Mosul, Northern Technical University, Mosul 41001, Iraq
[2] Electrical Engineering Department, College of Engineering, University of Mosul, Mosul 41002, Iraq

Corresponding Author Email: mti.lec306.Muna@ntu.edu.iq

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.591008

Received: 7 August 2026
Revised: 8 October 2026
Accepted: 20 October 2026
Available online: 31 October 2026

Keywords:
power quality disturbances, convolutional neural network, noise robustness, signal-to-noise ratio, computational cost, classifier comparison, IEEE 1159

ABSTRACT

Automatic recognition of power quality disturbances (PQDs) is a core function of monitoring terminals and protection devices, where noise resilience, latency, and model size govern deployment. This study benchmarks a compact raw-waveform one-dimensional convolutional neural network (1-D CNN) against feature-based support vector machine (SVM), random forest (RF), and k-nearest neighbours (k-NN) classifiers for 17 synthetic IEEE 1159 disturbance classes. The protocol combines validation-only baseline tuning, five independent runs, architecture ablation, unseen-severity testing, repeated CPU timing, representation isolation, and stress tests under impulsive, harmonic, and narrowband electromagnetic interference. At 20 dB AWGN, the CNN achieves 95.0 ± 1.3% accuracy, compared with 91.3 ± 1.2% for SVM, 84.1 ± 5.5% for RF, and 80.1 ± 2.5% for k-NN, while requiring 185.6 kB and 2.956 ± 0.534 ms/sample on the CPU. The CNN shows the strongest AWGN and impulsive-noise robustness among the evaluated models, but severe structured harmonic and narrowband interference favours the classical classifiers. SVM reaches 96.27% at 20 dB when driven by CNN features, indicating that the main robustness gain lies in the learned representation.

1. INTRODUCTION

Power quality disturbances (PQDs), including sags, swells, interruptions, harmonics, flicker, transients, notches, and their combinations, degrade sensitive loads and are increasingly common in distribution networks with high penetration of power-electronic interfaces and distributed generation. Automatic recognition of these disturbances underpins mitigation, source identification and compliance monitoring [1, 2]. From an engineering-automation standpoint, PQD classifiers are not stand-alone algorithms but embedded functions inside power-monitoring terminals, digital protection relays and edge-based real-time diagnosis units. Such devices operate on noisy field measurements, run on modest embedded processors without a GPU, and must return a decision within a fraction of a cycle [3, 4]. Two properties therefore matter as much as raw accuracy: the classifier must remain reliable as measurement noise rises, and it must have a small memory and latency footprint so that it can be deployed on constrained hardware [5]. A method that is marginally more accurate on clean laboratory signals but collapses under noise, or that cannot fit on a protection terminal, is of limited practical use. Despite this, two limitations recur in the literature [2, 6]. First, reported accuracy is usually obtained on clean or lightly perturbed signals; behaviour across a range of signal-to-noise ratio (SNR) is rarely quantified. Second, the computational cost of deep models-training time, inference latency and memory is seldom reported, so the practical price of their added complexity is unclear. Because synthetic parametric generation of PQD signals is widely used in this field, a fully controlled benchmark is possible in which the clean waveform is known exactly and calibrated noise can be injected to measure degradation directly [1, 7]. This paper exploits that possibility to study the joint question of robustness and cost. A compact one-dimensional convolutional neural network (1-D CNN) architecture operating on the raw waveform is compared against support vector machine (SVM), random forest (RF), and k-nearest neighbours (k-NN) baselines that use a 24-dimensional hand-crafted feature vector, under one unified protocol. The contributions are:

1. A seventeen-class PQD dataset (nine single, eight composite) generated from the IEEE 1159 models, with a reproducible noise-augmentation and evaluation protocol.
2. A degradation-aware evaluation reporting accuracy as a function of SNR (clean to 20 dB) for all classifiers, averaged over five independent runs with standard deviations to demonstrate training stability.
3. A hardware-neutral computational-cost analysis measuring per-sample inference latency (including feature extraction), training time, and model size on a common CPU.
4. An architecture ablation that justifies the chosen network depth from a signal-characteristics perspective, an out-of-range generalisation test on unseen disturbance severities, and a representation-versus-classifier study that isolates the source of the observed noise robustness.
5. A validation-only search that documents fair operating parameters for SVM, RF, and k-NN, together with non-

Gaussian stress tests that distinguish robustness to impulsive corruption from sensitivity to structured harmonic and narrowband electromagnetic interference.

The paper is organised as follows. Section 2 reviews related work; Section 3 describes the signal models and dataset; Section 4 details the classifiers and protocol; Section 5 gives the experimental setup; Section 6 presents and discusses the results; and Section 7 concludes.

2. RELATED WORK

Classical PQD recognition combines a signal-processing transform for feature extraction with a statistical classifier. Wavelet- and Stockwell-transform features feeding decision trees, SVMs or shallow networks remain accurate on synthetic data [8-11]. Recent feature-based methods also combine multiple transforms, feature selection and ensemble classifiers to improve discrimination while controlling feature dimensionality [12, 13]. Feature-based classical and dense neural models have also been evaluated across Gaussian-noise levels and software platforms [14]. More recent work replaces hand-crafted features with learned representations: convolutional and transformer architectures operating on raw or image-reshaped waveforms achieve very high accuracy [15-18]. Direct one-dimensional architectures have also been compared for waveform-level multi-label recognition and noise tolerance [19], while hybrid deep models have been applied to fault and disturbance classification in power networks [20, 21]. Multiple and composite disturbances have been addressed using transform–deep-learning pipelines, feature-driven ensembles and deep-feature/classical-classifier combinations [22-24]. Two gaps motivate this study. First, although noise-aware evaluation is receiving increasing attention, protocols remain heterogeneous. Cross-attention models have reported performance at selected high-noise operating points [25], a hierarchical CNN states that no systematic SNR evaluation was performed [15], and a vision-transformer study reports a marked reduction below 30 dB [16]. Second, studies that jointly report noise robustness, complete computational cost, tuned classical baselines, and representation effects under identical conditions remain scarce. The novelty claimed here is not a new convolution operator; it is the controlled quantification of degradation patterns, robustness-cost trade-offs, and representation effects within one reproducible, deployment-oriented protocol.

The present study therefore focuses on the combined experimental design: a paired SNR sweep, CPU cost accounting that includes feature extraction, architecture ablation, out-of-range severity testing, and reuse of learned features with the same classical classifiers to isolate the source of robustness.

3. POWER QUALITY DISTURBANCES SIGNAL MODELING AND DATASET

3.1 Parametric disturbance models

Each observation is a single-phase voltage waveform sampled at f_s = 6.4 kHz over ten cycles of the f_0 = 50 Hz fundamental, giving N = 1280 samples per signal. The base waveform is $(v(t))$= sin $(2\pi f_0 t)$ in per unit. Nine single disturbances are produced by applying the IEEE 1159 parametric modifiers summarized in Table 1; eight composite classes are formed by superposing two modifiers with independently randomized parameters.

Disturbance onset, duration, depth, and spectral content are drawn uniformly from the ranges in Table 1 so that classes overlap realistically at their boundaries. Representative clean waveforms of all seventeen classes are shown in Figure 1.

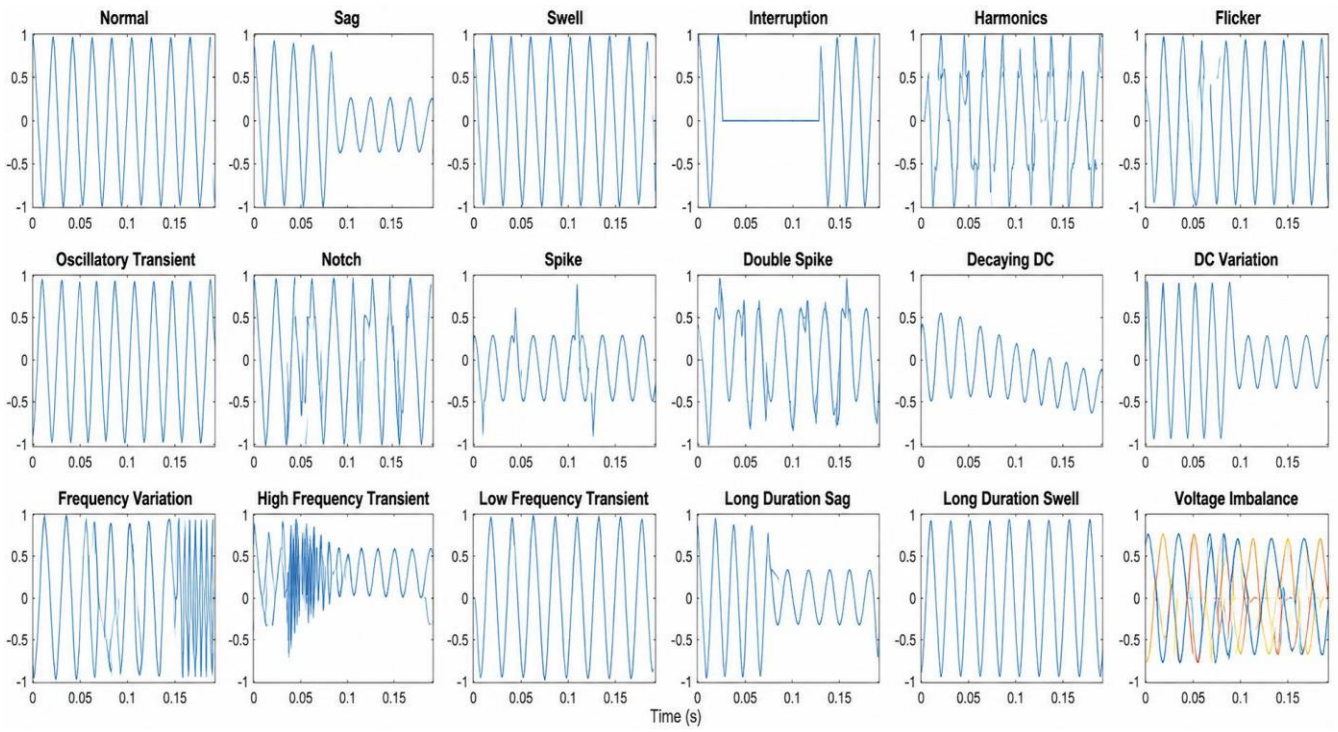


Figure 1. Representative clean waveforms of the seventeen PQD classes generated from the IEEE 1159 parametric models

Table 1. PQD classes and their parametric models (per unit, $\omega = 2\pi f_0$)

Class	Model/Modifier	Parameters
Normal	$\sin(\omega t)$	—
Sag	$(1 - \alpha)$ during window	$\alpha \in [0.1,0.9]$
Swell	$(1 + \alpha)$ during window	$\alpha \in [0.1,0.8]$
Interruption	$(1 - \alpha)$ during window	$\alpha \in [0.9,1.0]$
Harmonics	$+ a_3 \sin 3\omega t + a_5 \sin 5\omega t + a_7 \sin 7\omega t$	$a_i \in [0.05,0.15]$
Flicker	$(1 + \alpha f \sin 2\pi f_f t)$	$\alpha f \in [0.1,0.2]$, $\epsilon f_f \epsilon [5,20]$ Hz
Osc. Transient	$+\beta e^{-(t-t1)/\tau} \sin 2\pi f_n n(t-t1)$	$f_n \in [300,900]$ Hz, $\tau \in [8,40]$ ms
Notch	Periodic notches, 4/cycle	depth $\in [0.1,0.4]$
Spike	2–5 impulsive spikes	amp. $\in [0.5,2.0]$
Composite classes (8): Sag + Harm., Swell + Harm., Interruption + Harm., Flicker + Harm., Sag + Osc. transient, Swell + Osc. transient, Sag + Flicker, Notch + Harm.		

3.2 Dataset and noise protocol

For each class, 400 signals are generated, yielding 6800 signals in total. The data are split, with class-wise stratification, into 70% training (4760), 15% validation (1020) and 15% test (1020). To emulate field conditions and to train noise-tolerant models, each training and validation signal is corrupted with additive white Gaussian noise at an SNR drawn uniformly from [20,50] dB, with 15% of signals kept clean so that the noise-free condition remains in-distribution. For evaluation, the same clean test waveforms are corrupted at each level of a fixed SNR grid {∞,50,40,30,20} dB, isolating the effect of noise from that of the underlying signals. The reference operating point for per-class analysis is 30 dB.

4. METHODOLOGY

4.1 Deep model: Raw-waveform 1-D CNN

The deep classifier consumes the raw single-channel waveform directly. Its architecture (Table 2) has four temporal convolution blocks of increasing width followed by global average pooling, dropout, and two fully connected layers [26]. The waveform is standardised using training-set statistics only, so no test information leaks into the model.

Table 2. Compact 1-D CNN architecture (N = 1280 input, 17 outputs)

Stage	Layer	Output
Input	raw waveform (z-scored)	1280 × 1
Conv-1	conv(9) × 16, BN, ReLU, pool/2	640 × 16
Conv-2	conv(7) × 32, BN, ReLU, pool/2	320 × 32
Conv-3	conv(5) × 64, BN, ReLU	160 × 64
Conv-4	conv(3) × 128, BN, ReLU	160 × 128
Pooling	global average pooling	128
Head	dropout(0.4), FC(64), ReLU	64
	dropout(0.3), FC(17), softmax	17

The design follows the physical structure of PQD waveforms rather than being chosen arbitrarily. At 6.4 kHz, the first-layer kernel of width nine spans about 1.4 ms, i.e. a fraction of one 50 Hz cycle, so it captures localised waveform events—notch edges, transient onsets and spikes—at their native time scale. Successive kernels shrink (7, 5, 3) while the number of filters grows (16, 32, 64, 128) and pooling halves the temporal resolution, so deeper layers aggregate the local cues into cycle-level and multi-cycle patterns such as sags, flicker envelopes and harmonic distortion. The number of blocks was selected through the architecture ablation study. Dropout rates of 0.4 after pooling and 0.3 before the output layer were used as moderate regularisation to reduce overfitting.

4.2 Classical baselines and features

The three baselines use a 24-dimensional feature vector: eleven time-domain descriptors, seven frequency-domain descriptors, and six relative db4 wavelet sub-band energies. Features are z-score normalised using training statistics only. Model selection uses the validation partition only. The search covers SVM box constraint and kernel scale in {0.1, 1, 10}; RF size in {60, 120, 240} trees and minimum leaf size in {1, 3, 5}, with four randomly selected candidate features considered at each split (mtry = 4); and k-NN with k in {1, 3, 5, 7, 9} and Euclidean, city-block, or cosine distance. The selected settings are an RBF SVM with C = 10 and kernel scale 10, an RF with 120 trees and minimum leaf size 1, and k-NN with k = 1 and city-block distance. The untouched test set is used only after selection. All baselines use the same noise-augmented training signals as the CNN [27].

4.3 Evaluation protocol and cost measurement

Performance is reported as accuracy at each SNR and per-class precision, recall, and F1 at 30 dB. The full pipeline is repeated for five random seeds, and Section 6.1 reports mean and standard deviation. Per-class, confusion-matrix, cost, and deep-feature results use the representative seed-42 run. The principal SNR sweep and the Round-2 noise-family test use the same trained seed-42 models but independently seeded noise realisations; their single-realisation AWGN values therefore need not be identical. CPU latency includes feature extraction for classical models and is measured in 20 independent runs of 300 samples after warm-up. Training time is measured once for the representative run.

5. EXPERIMENTAL SETUP

All experiments ran in MATLAB R2024a with the Deep Learning, Statistics and Machine Learning, and Wavelet Toolboxes on a Windows 11 laptop with an Intel Core i7-1165G7 CPU (4 cores, 8 threads, 2.8 GHz) and 16 GB RAM. No discrete GPU was used. The CNN used Adam [28], an initial learning rate of 10^{-3} halved every 20 epochs, mini-batches of 64, L2 regularisation of 10^{-4}, and at most 45 epochs with early stopping on validation loss (patience 8). Seeds 11, 22, 33, 44, and 55 were used for stability, while seed 42 defines

the representative detailed run. Latencies are hardware-dependent and are interpreted comparatively on identical hardware.

6. RESULTS AND DISCUSSION

6.1 Accuracy and noise robustness

Table 3 reports five-run accuracy across the SNR grid and macro-F1 at 30 dB. At 20 dB, the CNN retains 95.0 ± 1.3%, compared with 91.3 ± 1.2% for SVM, 84.1 ± 5.5% for RF, and 80.1 ± 2.5% for k-NN. Relative to clean signals, the mean reductions are 3.1, 5.8, 13.6, and 13.2 percentage points, respectively. After validation-only tuning strengthens the classical baselines, the CNN remains the most accurate primary model at the most demanding AWGN condition. Figure 2 shows the corresponding accuracy trends across the SNR grid.

6.2 Training stability

At 20 dB, the CNN standard deviation is 1.3 points, close to SVM (1.2) and below k-NN (2.5) and RF (5.5). The CNN therefore combines the highest mean accuracy with controlled dispersion, whereas RF is the most sensitive to run-to-run variation under severe AWGN. In the representative trajectory, validation accuracy reaches 98.92%; the minimum-validation-loss checkpoint is retained with early stopping on validation loss (patience 8). Figure 3 shows the representative training and validation trajectories.

Table 3. Test accuracy (%) versus SNR, mean ± standard deviation over five runs; macro-F1 at 30 dB

Model	Clean	50 dB	40 dB	30 dB	20 dB	M-F1
CNN (raw)	98.1 ± 0.7	98.1 ± 0.7	98.0 ± 0.7	97.6 ± 0.9	95.0 ± 1.3	97.6 ± 0.9
SVM	97.1 ± 0.6	97.3 ± 0.6	97.3 ± 0.7	96.4 ± 0.7	91.3 ± 1.2	96.4 ± 0.7
Random forest	97.7 ± 0.3	97.7 ± 0.3	97.8 ± 0.3	97.6 ± 0.5	84.1 ± 5.5	97.6 ± 0.5
k-NN	93.3 ± 0.6	93.6 ± 0.3	93.6 ± 0.7	90.7 ± 1.4	80.1 ± 2.5	90.5 ± 1.5

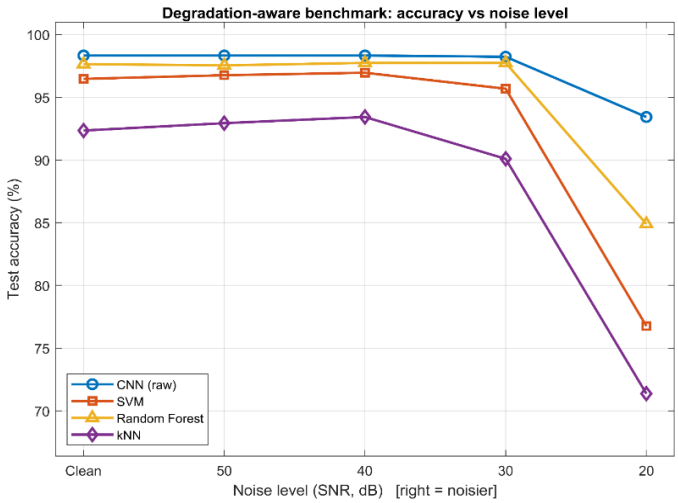


Figure 2. Accuracy versus noise level, showing that the CNN remains nearly stable across the SNR grid while the classical models degrade sharply at 20 dB

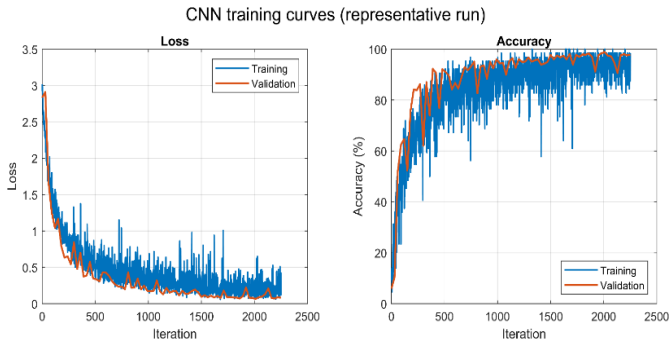


Figure 3. Representative CNN training and validation loss and accuracy curves showing close agreement and stable training without evident overfitting

6.3 Per-class behavior

Table 4. Per-class performance of the CNN at 30 dB (%)

Class	Precision	Recall	F1
Normal	96.77	100.00	98.36
Sag	91.94	95.00	93.44
Swell	95.24	100.00	97.56
Interruption	93.75	100.00	96.77
Harmonics	100.00	100.00	100.00
Flicker	100.00	100.00	100.00
Osc. transient	100.00	98.33	99.16
Notch	100.00	100.00	100.00
Spike	100.00	98.33	99.16
Sag + Harm.	100.00	95.00	97.44
Swell +Harm.	100.00	100.00	100.00
Interrupt + Harm.	95.24	100.00	97.56
Flicker + Harm.	100.00	100.00	100.00
Sag + Osc. trans.	100.00	91.67	95.65
Swell + Osc. trans.	100.00	95.00	97.44
Sag + Flicker	100.00	98.33	99.16
Notch + Harm.	100.00	100.00	100.00
Macro average	98.41	98.33	98.34

Table 4 reports the CNN per-class metrics at 30 dB (macro-F1 = 98.34%). Sag has the lowest classwise F1 (93.44%) because its principal cue is an amplitude reduction whose random depth and duration can approach the boundaries of Normal, Interruption, and composite sag classes. Sag + Flicker reaches 99.16% because periodic low-frequency modulation supplies a second discriminative time-frequency cue. Classes with distinctive signatures, including Harmonics, Flicker, Notch, Swell + Harmonics, and Notch + Harmonics, reach 100% F1. The corresponding confusion matrix and per-class summary are shown in Figures 4 and 5.

6.4 Computational cost

Table 5 shows that the CNN has the smallest footprint (185.6 kB, versus 796.4–4,253.4 kB). Across 20 timing runs, its latency is 2.956 ± 0.534 ms/sample, only 0.25 ms slower than k-NN and approximately 8.8 and 32 times faster than SVM and RF, respectively. Classical feature extraction alone requires 1.307 ms/sample. CNN training takes 1,469.28 s, but this is an offline cost; the deployment path remains compact and fast. RF has the largest timing variation, so both the mean and standard deviation are reported. Figure 6 visualises the per-sample latency and robustness-cost trade-off.

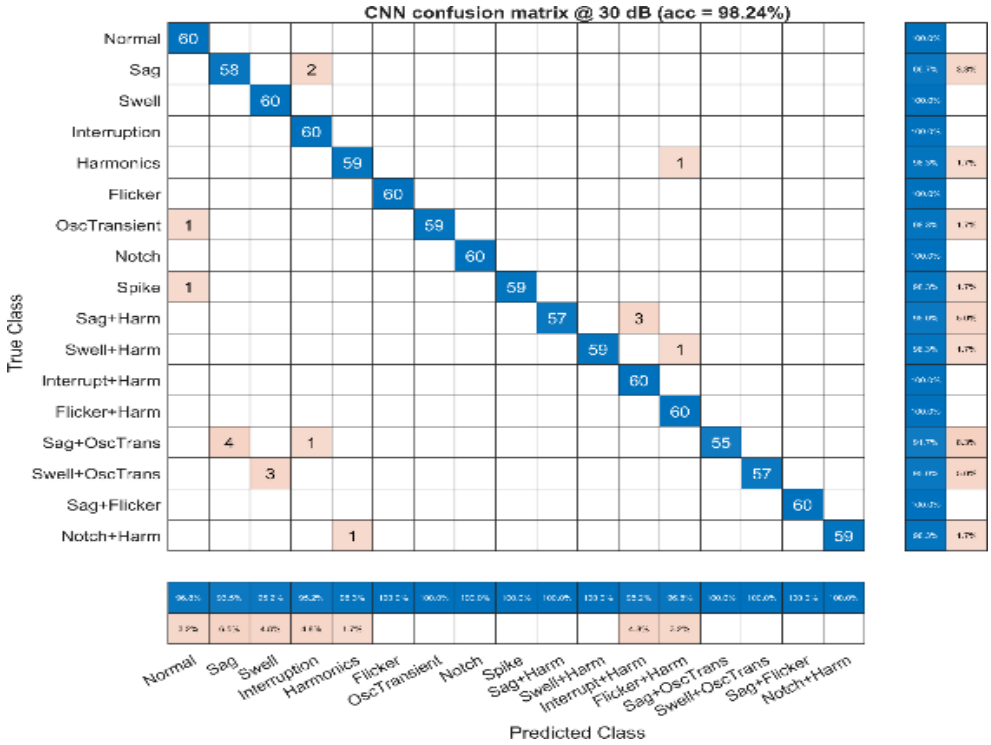


Figure 4. Confusion matrix of the CNN at 30 dB

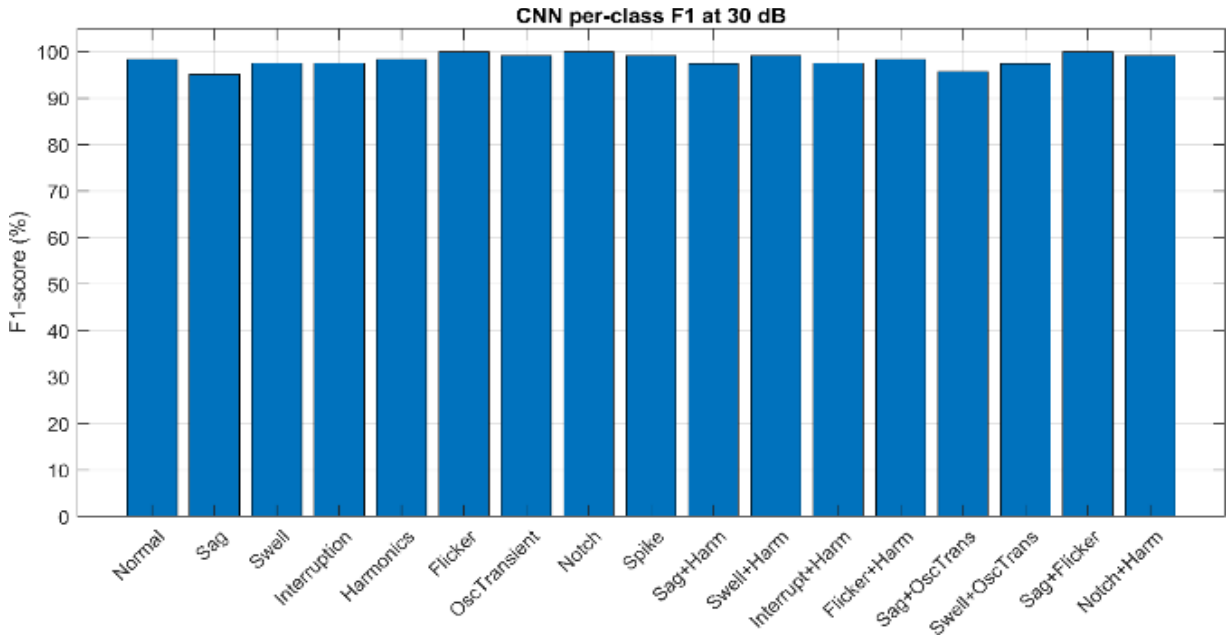


Figure 5. Per-class performance of the CNN at the reference SNR

Table 5. Per-sample inference cost on the central processing unit (CPU), training time, and model size

Model	Feat. (ms)	Pred. (ms)	Total (ms)	SD (ms)	Train (s)	Size (kB)
CNN (raw)	0.000	2.956	2.956	0.534	1,469.28	185.6
SVM	1.307	24.739	26.046	2.927	1.26	1,673.5
Random forest	1.307	93.320	94.628	43.969	2.51	4,253.4
k-NN	1.307	1.398	2.705	0.347	0.01	796.4

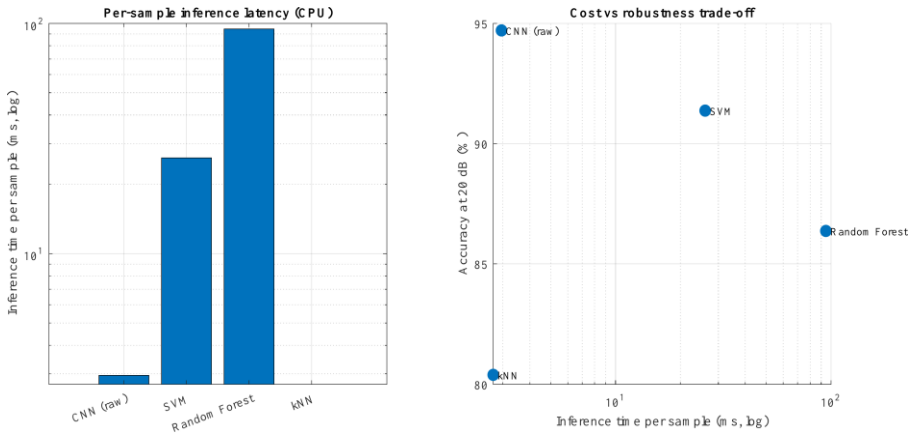


Figure 6. Left: per-sample inference latency (log scale). Right: the robustness–cost trade-off (accuracy at 20 dB versus latency). The CNN combines high robustness with low latency and the smallest footprint

6.5 Architecture ablation

Table 6 varies the network from two to five convolution blocks under the same representative stability seed. Four blocks attain 94.12% at 20 dB, compared with 79.51%, 79.22%, and 92.84% for two, three, and five blocks. Although five blocks reach the highest clean accuracy (98.14%), they use 155,889 parameters, 3.21 times the four-block count, and are less accurate at 20 dB. Four blocks therefore maximise high-noise accuracy in this ablation while preserving the intended embedded footprint. Figure 7 summarises the architecture-ablation results.

Table 6. Architecture ablation over the number of convolution blocks

Conv Blocks	Clean (%)	20 dB (%)	Parameters
2	84.31	79.51	7,089
3	92.45	79.22	19,569
4 (chosen)	96.37	94.12	48,625
5	98.14	92.84	155,889

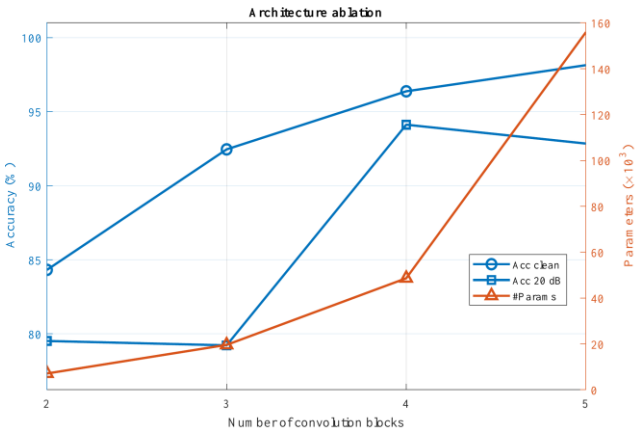


Figure 7. Architecture ablation: accuracy (clean and 20 dB) and parameter count versus the number of convolution blocks. Accuracy saturates at four blocks

6.6 Out-of-range generalisation

To test transfer beyond a random split, all models are trained on mild ranges and evaluated on non-overlapping severe ranges at 30 dB. The mild/severe intervals are: sag 0.10-0.50/0.50-0.90; swell 0.10-0.45/0.45-0.80; interruption 0.90-0.95/0.95-1.00; harmonics 0.05-0.10/0.10-0.15; flicker 0.10-0.15/0.15-0.20; oscillatory transient 0.10-0.45/0.45-0.80; notch 0.10-0.25/0.25-0.40; and spike 0.50-1.25/1.25-2.00. The CNN retains the highest severe-range accuracy (81.96%) and the smallest drop (17.35 points), followed by SVM at 80.20%. RF and k-NN reach 74.80% and 73.63%, showing that training-range coverage remains important. The numerical results and corresponding comparison are presented in Table 7 and Figure 8.

Table 7. Out-of-range generalisation: trained on mild severities, tested on unseen severe severities (accuracy at 30 dB, %)

Model	In-range	Severe (Unseen)	Drop
CNN (raw)	99.31	81.96	17.35
SVM	98.43	80.20	18.24
Random forest	99.02	74.80	24.22
k-NN	93.14	73.63	19.51

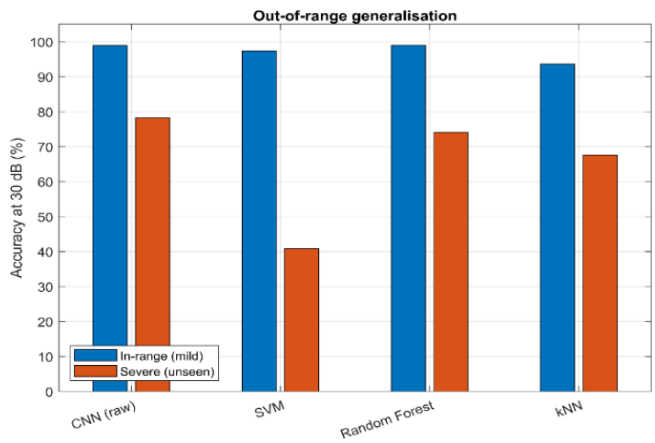


Figure 8. In-range and out-of-range (severe, unseen) accuracy at 30 dB, showing the strongest generalisation for the CNN and RF and the largest degradation for SVM

6.7 Representation versus classifier

Because the CNN uses the raw waveform while the classical models use hand-crafted features, their performance gap could stem from either the representation or the classifier. To separate the two, the 128-dimensional activations of the CNN's global-average-pooling layer were extracted as deep features and fed to the same SVM, RF and k-NN classifiers

(Table 8). With deep features the classical classifiers become far more noise-robust: at 20 dB, SVM improves from 91.37% to 96.27% (+4.90 points), k-NN from 80.39% to 94.31% (+13.92 points), and RF from 86.37% to 93.63% (+7.26 points). Remarkably, the auxiliary SVM trained on the CNN's deep representation—rather than used as an independent baseline—even exceeds the end-to-end CNN at 20 dB. This shows that the robustness originates chiefly from the learned representation rather than the classifier, and directly explains why the deep model is worth its training cost. Figure 9 visualises the effect of the two feature representations across the SNR grid.

Table 8. Accuracy (%) using hand-crafted and CNN deep features for the representative seed-42 run under the principal-sweep AWGN realisations

Classifier/Features	Clean	50	40	30	20
SVM - hand-crafted	97.84	97.84	97.94	97.35	91.37
SVM – deep	99.61	99.61	99.51	99.41	96.27
RF - hand-crafted	97.75	97.84	97.84	97.45	86.37
RF – deep	99.51	99.51	99.51	99.41	93.63
k-NN - hand-crafted	94.12	93.92	93.53	92.75	80.39
k-NN – deep	99.41	99.41	99.12	99.31	94.31
CNN (raw, end-to-end)	98.33	98.33	98.33	98.33	94.71

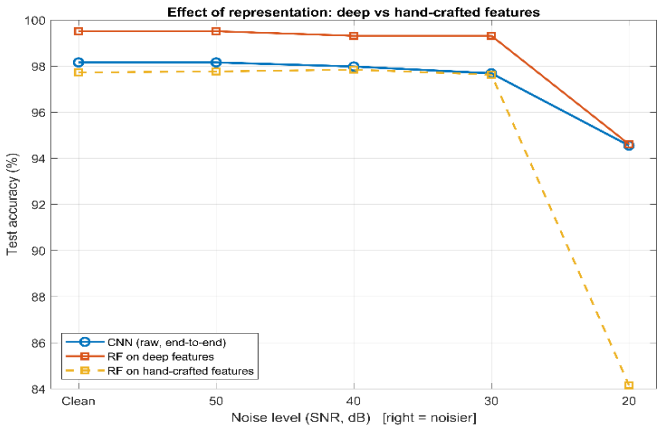


Figure 9. Effect of hand-crafted versus CNN-learned representations

6.8 Validation-based hyperparameter selection

Table 9 reports the complete validation-only search. Settings maximise validation accuracy, with macro-F1 retained as a class-balanced check; the test set is not consulted. This prevents arbitrary baseline settings or test-set tuning.

Table 9. Validation-only selection of classical-classifier parameters

Model	Search Grid	Selected	Val. Acc.	Val. M-F1
SVM	C, scale ∈ {0.1, 1, 10}	C = 10; scale = 10	96.37	96.36
RF	Trees ∈ {60,120,240}; leaf ∈ {1,3,5}	120 trees; leaf = 1	97.25	97.25
k-NN	k ∈ {1,3,5,7,9}; 3 distances	K = 1; city-block	91.47	91.37

6.9 Robustness beyond AWGN

Table 10 evaluates the seed-42 models under three unseen contamination families in addition to an independently seeded AWGN control. At 20 dB impulsive noise, the CNN retains 92.65% accuracy and 91.73% macro-F1; the next-highest accuracy is 68.33% for RF. At 30 dB harmonic interference the CNN leads at 98.43%, whereas SVM leads the 30 dB narrowband-EMI condition at 97.84%. Under severe 20 dB harmonic and EMI conditions, k-NN reaches 74.31% and 70.39%, compared with 54.61% and 55.29% for the CNN. Thus, AWGN augmentation transfers strongly to sparse impulses but does not fully cover coherent spectral interference. Figures 10 and 11 summarise the corresponding accuracy and macro-F1 results.

The dominant 20 dB errors reflect overlapping cues: Normal is mapped to Oscillatory Transient in 50.0% of the AWGN cases, Normal to Spike in 78.3% under impulsive noise, Spike to Swell + Oscillatory Transient in 73.3% under harmonic interference, and Harmonics to Notch + Harmonics in 68.3% under EMI. These stringent tests support interference-aware augmentation or front-end suppression as the next engineering improvement.

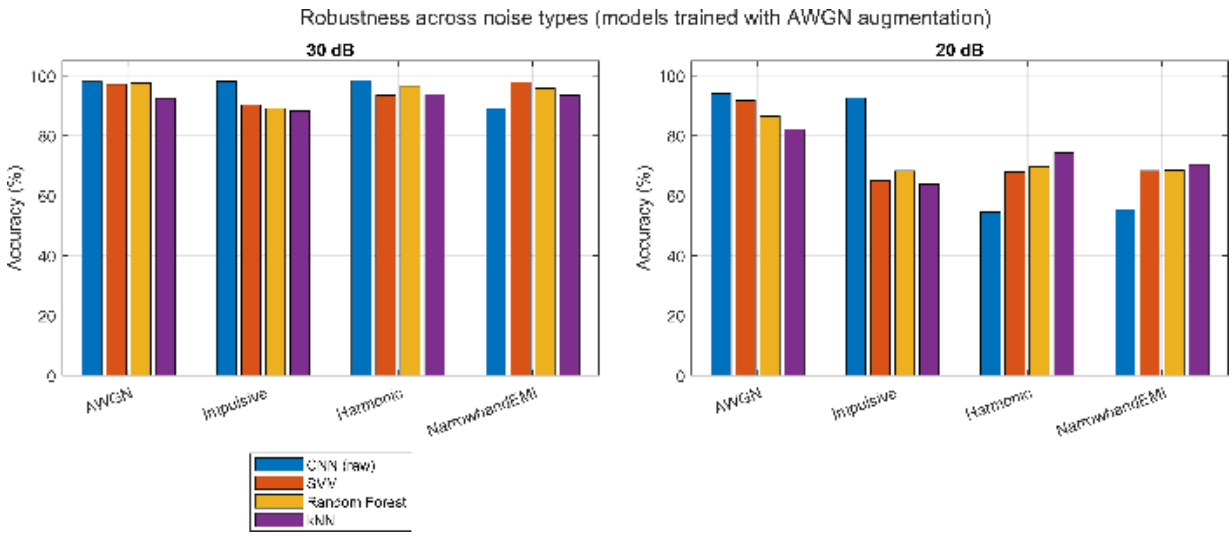


Figure 10. Accuracy under AWGN, impulsive, harmonic, and narrowband-EMI conditions at 30 and 20 dB

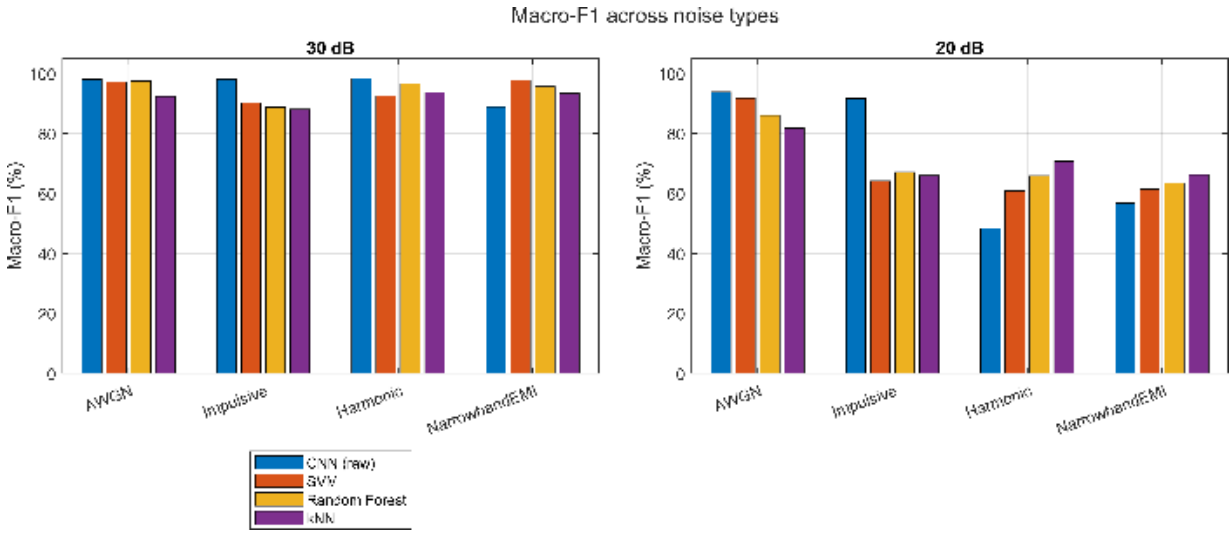


Figure 11. Macro-F1 under AWGN, impulsive, harmonic, and narrowband-EMI conditions

Table 10. Accuracy/macro-F1 (%) under unseen noise and interference

Level	Condition	CNN	SVM	RF	k-NN
30 dB	AWGN	98.04/98.04	97.16/97.14	97.55/97.54	92.45/92.40
30 dB	Impulsive	98.14/98.14	90.29/90.16	89.12/88.82	88.24/88.11
30 dB	Harmonic	98.43/98.43	93.43/92.51	96.57/96.57	93.73/93.68
30 dB	Narrowband EMI	89.02/88.93	97.84/97.84	95.78/95.67	93.43/93.36
20 dB	AWGN	94.02/93.84	91.77/91.82	86.47/86.04	82.16/81.87
20 dB	Impulsive	92.65/91.73	64.90/64.20	68.33/67.13	63.92/66.00
20 dB	Harmonic	54.61/48.45	67.84/60.90	69.71/65.93	74.31/70.83
20 dB	Narrowband EMI	55.29/56.76	68.33/61.37	68.43/63.50	70.39/66.26

6.10 Comparison with prior work

Tables 11 and 12 position the study against recent PQD and event-classification work. Numerical results are not directly comparable because class sets, signal models, hardware, and noise conditions differ. The distinction here is the joint evaluation of raw-waveform learning, tuned classical baselines, per-sample CPU latency, depth ablation, unseen-severity transfer, representation isolation, and three non-Gaussian contamination families under one protocol.

For scale, the Vision Transformer reported by Akkaya and Dümen contains approximately 86 million parameters, whereas the selected CNN contains 48,625; the present network therefore uses about 1,770 times fewer parameters. This comparison concerns footprint and does not imply direct accuracy comparability between different datasets or protocols.

Table 11. Numerical comparison with recent PQD/event classification studies (values verified from the full texts; not directly comparable due to differing conditions)

Study	Year	Method	# Classes	SNR Tested	Reported Result	Cost/Hardware
Akkaya and Dümen [16]	2025	Vision Transformer	21	20–50 dB (curve)	Acc.: 99.95/93.57	3.58 ms (P100); 86M
Juarez et al. [15]	2026	Hierarchical 1D-CNN	29	none (a)	F1: 98.86/—	23.8 ms/signal (Apple M2)
Tunio et al. [20]	2025	DWT + TCN/BiLSTM/GRU	12	noisy (level NR)	Acc.: 99.60/-- (b)	25 ms (platform NR); 221k
Singh et al. [8]	2023	ST + HT + DWT + tree	18	20 dB (1 level)	Acc.: 98.58/97.62	80.564 ms (platform NR)
Veizaga et al. [21]	2023	STFT + Fortescue + DTW	7 causes	10–40 dB (curve)	Acc.: 100; field F1: >99	not reported
Yılmaz et al. [9]	2022	UWT + SVM	9	25–40 dB (3 lvl)	Acc.: 99.85/99.32	12.33 ms (2-level UWT only) (c)
Palomares-Salas et al. [14]	2025	Hand-crafted features + ML/DNN	10	1–40 dB (curve)	Acc.: 98.25/84.07 (d)	18.7-21.5 s/test set (e)
This work (CNN)	2026	Compact 1-D CNN (raw)	17	Clean-20 dB	Acc.: 98.1/95.0	2.96 ms (CPU)
This work (RF)	2026	Features + RF	17	Clean-20 dB	Acc.: 97.7/84.1	94.63 ms (CPU)

(a) Authors state no systematic SNR robustness evaluation was performed. NR: not reported.
(b) The reported TCN accuracy varies across the abstract and Table 8 of Tunio et al. [20]; the table value is used here.
(c) Yılmaz et al. [9] timed only the UWT stage; the complete classification latency and timing platform are not reported.
(d) Python DNN results at 40 and 1 dB; no clean-signal result was reported.
(e) Total Python DNN test-set time including preprocessing and feature extraction; not per-sample latency.

Table 12. Coverage of evaluation dimensions (✓ addressed, ~ partial, × not addressed)

Study	Raw	Classical	SNR	Cost	Per-sample	Arch.	Unseen	Representation
	Waveform	baselines	sweep	Analysis	CPU latency	Ablation	severity	isolation
Akkaya & Dümen [16]	~	×	✓	✓	×	×	×	×
Juarez et al. [15]	✓	×	×	✓	×	✓	×	×
Tunio et al. [20]	×	×	×	✓	×	×	×	×
Singh et al. [8]	×	~	×	~	×	×	×	×
Veizaga et al. [21]	×	✓	✓	×	×	×	×	×
Yılmaz et al. [9]	×	✓	~	~	×	×	×	~
Palomares-Salas et al. [14]	×	✓	✓	✓	×	×	×	×
This work	✓	✓	✓	✓	✓	✓	✓	✓

Note: Akkaya & Dümen reshape raw samples into a 32 × 32 image; Yılmaz et al. [9] control UWT versus DWT features, not learned versus hand-crafted features.

6.11 Discussion

The combined evidence defines a deployment-relevant robustness-cost trade-off. Under AWGN, the CNN reaches the highest 20 dB mean accuracy (95.0%) with a 185.6 kB footprint and 2.956 ± 0.534 ms CPU latency. It also retains 92.65% under unseen 20 dB impulsive contamination. Validation-only tuning makes the comparison meaningful against strengthened classical baselines, while the deep-feature experiment identifies the learned representation, rather than the softmax head, as the principal source of robustness. The operating boundary is equally clear: the CNN does not dominate severe coherent harmonic or narrowband EMI, and its 81.96% unseen-severity accuracy shows that parameter-range coverage still matters. For monitoring terminals, protection devices, and edge-diagnosis units, training can occur offline, whereas memory and inference latency are continuous deployment costs. The compact CNN is therefore attractive when AWGN-like and impulsive contamination dominate, with front-end suppression or expanded training needed for structured interference.

7. CONCLUSIONS

This paper presented a controlled benchmark of a compact four-block raw-waveform 1-D CNN against validation-tuned SVM, RF, and k-NN baselines for 17 PQD classes. Across five runs, the CNN achieved 95.0 ± 1.3% at 20 dB AWGN with a 185.6 kB footprint and 2.956 ± 0.534 ms/sample CPU latency. It retained 92.65% under unseen 20 dB impulsive contamination, and the four-block configuration gave the best high-noise ablation result. Representation isolation, unseen-severity testing, and repeated timing show that the main benefit is a compact learned representation robust to AWGN-like and impulsive corruption. At 20 dB harmonic interference and narrowband EMI, however, the best classical accuracies were 74.31% and 70.39%, while the CNN reached 54.61% and 55.29%. The conclusions are therefore limited to the evaluated synthetic single-phase IEEE 1159 conditions and do not imply universal field robustness. Measured multi-site, nonlinear, and unbalanced three-phase validation with interference-aware training remains necessary before operational deployment.

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] Caicedo, J.E., Agudelo-Martínez, D., Rivas-Trujillo, E., Meyer, J. (2023). A systematic review of real-time detection and classification of power quality disturbances. Protection and Control of Modern Power Systems, 8(1): 1-37. https://doi.org/10.1186/s41601-023-00277-y

[3] Igual, R., Medrano, C. (2020). Research challenges in real-time classification of power quality disturbances applicable to microgrids: A systematic review. Renewable and Sustainable Energy Reviews, 132: 110050. https://doi.org/10.1016/j.rser.2020.110050

[4] Zlatkova, A., Kokolanski, Z., Taskovski, D. (2024). Real-time system for detection and classification of power quality disturbances. In 2024 IEEE International Conference on Environment and Electrical Engineering and 2024 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe), Rome, Italy, pp. 1-6. https://doi.org/10.1109/EEEIC/ICPSEurope61470.2024.10751665

[5] Hou, Z., Wang, B., Liu, J., He, Y., Yao, Y. (2025). Physics-inspired time-frequency feature extraction and lightweight neural network for power quality disturbance classification. Frontiers in Physics, 13: 1616367. https://doi.org/10.3389/fphy.2025.1616367

[6] Khokhar, S., Zin, A.A.B.M., Mokhtar, A.S.B., Pesaran, M. (2015). A comprehensive overview on signal processing and artificial intelligence techniques applications in classification of power quality disturbances. Renewable and Sustainable Energy Reviews, 51: 1650-1663. https://doi.org/10.1016/j.rser.2015.07.068

[7] IEEE Recommended Practice for Monitoring Electric Power Quality, IEEE Std 1159-2019, pp. 1-98. https://doi.org/10.1109/IEEESTD.2019.8796486

[8] Singh, S., Sharma, A., Garg, A.R., et al. (2023). Power quality detection and categorization algorithm actuated by multiple signal processing techniques and rule-based decision tree. Sustainability, 15(5): 4317. https://doi.org/10.3390/su15054317

[9] Yılmaz, A., Küçüker, A., Bayrak, G., Ertekin, D., Shafie-Khah, M., Guerrero, J.M. (2022). An improved automated PQD classification method for distributed generators with hybrid SVM-based approach using un-decimated wavelet transform. International Journal of Electrical Power & Energy Systems, 136: 107763. https://doi.org/10.1016/j.ijepes.2021.107763

[10] Li, J., Teng, Z., Tang, Q., Song, J. (2016). Detection and classification of power quality disturbances using double resolution S-transform and DAG-SVMs. IEEE Transactions on Instrumentation and Measurement,

65(10): 2302-2312. https://doi.org/10.1109/TIM.2016.2578518
[11] Khetarpal, P., Nagpal, N., Alhelou, H.H., Siano, P., Al-Numay, M. (2024). Noisy and non-stationary power quality disturbance classification based on adaptive segmentation empirical wavelet transform and support vector machine. Computers and Electrical Engineering, 118: 109346. https://doi.org/10.1016/j.compeleceng.2024.109346
[12] Li, P., Ma, T., Shi, J., Jia, Q. (2024). Multi-dimensional feature multi-classifier synergetic classification method for power quality disturbances. Computers and Electrical Engineering, 120: 109720. https://doi.org/10.1016/j.compeleceng.2024.109720
[13] Ismael, M.R., Abd, H.J., Homod, R.Z. (2024). An algorithm for the classification of power quality disturbance signals using a tunable-Q-factor wavelet transform and ensemble learning methodology. Electrical Engineering, 106(4): 5253-5265. https://doi.org/10.1007/s00202-024-02294-y
[14] Palomares-Salas, J.C., Aguado-González, S., Sierra-Fernández, J.M. (2025). Robustness of machine learning and deep learning models for power quality disturbance classification: A cross-platform analysis. Applied Sciences, 15(19): 10602. https://doi.org/10.3390/app151910602
[15] Juarez, M.G., Cerda, J., Zamora-Mendez, A., Ortiz-Bejar, J., Silva-Chavez, J.C. (2026). Multi-stage hierarchical CNN model for power quality disturbance detection and classification. AI, 7(6): 220. https://doi.org/10.3390/ai7060220
[16] Akkaya, S., Dümen, S. (2025). A novel vision transformer-based power quality disturbance classification method. Ain Shams Engineering Journal, 16(11): 103718. https://doi.org/10.1016/j.asej.2025.103718
[17] Khetarpal, P., Nagpal, N., Al-Numay, M.S., Siano, P., Arya, Y., Kassarwani, N. (2023). Power quality disturbances detection and classification based on deep convolution auto-encoder networks. IEEE Access, 11: 46026-46038. https://doi.org/10.1109/ACCESS.2023.3274732
[18] Wang, G., Zhang, H., Gao, M., Ding, W., Qian, Y. (2025). Identification and classification of power quality disturbances using CNN-transformer. Journal of Electrical Engineering & Technology, 20(5): 2993-3007. https://doi.org/10.1007/s42835-025-02213-6
[19] Shao, H., Henriques, R., Morais, H., Tedeschi, E. (2024). Comparing deep neural networks for 1D multi-label power quality disturbance classification: ResNet, MobileNet, and DenseNet. In IECON 2024 - 50th Annual Conference of the IEEE Industrial Electronics Society, Chicago, IL, USA, pp. 1-6. https://doi.org/10.1109/IECON55916.2024.10905211
[20] Tunio, N.A., Tunio, M.A., Raza, M.A., Faheem, M., Hashmani, A.A., Nadeem, R. (2025). Performance comparison between deep learning models for fault classification in transmission lines using time series data. Energy Science & Engineering, 13(5): 2330-2351. https://doi.org/10.1002/ese3.70033
[21] Veizaga, M., Delpha, C., Diallo, D., Bercu, S., Bertin, L. (2023). Classification of voltage sags causes in industrial power networks using multivariate time-series. IET Generation, Transmission & Distribution, 17(7): 1568-1584. https://doi.org/10.1049/gtd2.12765
[22] Cui, C., Duan, Y., Hu, H., Wang, L., Liu, Q. (2022). Detection and classification of multiple power quality disturbances using Stockwell transform and deep learning. IEEE Transactions on Instrumentation and Measurement, 71: 1-12. https://doi.org/10.1109/TIM.2022.3214284
[23] Liu, Y., Yuan, D., Fan, H., Jin, T., Mohamed, M.A. (2023). A multidimensional feature-driven ensemble model for accurate classification of complex power quality disturbance. IEEE Transactions on Instrumentation and Measurement, 72: 1-13. https://doi.org/10.1109/TIM.2023.3265756
[24] Soni, P., Mishra, P., Mondal, D. (2025). A novel amalgamation of pre-processing technique and CNN model for accurate classification of power quality disturbances. Electrical Engineering, 107(4): 5187-5206. https://doi.org/10.1007/s00202-024-02818-6
[25] Liao, T., Wang, W., Xing, Y. (2024). A method for disturbance identification in power quality based on cross-attention fusion of temporal and spatial features. Electric Power Systems Research, 234: 110560. https://doi.org/10.1016/j.epsr.2024.110560
[26] LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. Nature, 521(7553): 436-444. https://doi.org/10.1038/nature14539
[27] Breiman, L. (2001). Random forests. Machine Learning, 45(1): 5-32. https://doi.org/10.1023/A:1010933404324
[28] Kingma, D.P., Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980. https://doi.org/10.48550/arXiv.1412.6980