International Journal of Engineering Research and Development
e- ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com
Volume 22, Issue 7 (July 2026), PP 280-300

Raw IQ Deep Learning for Radar Waveform Classification: A Comparative Evaluation of 1D CNN, 2D CNN, and Shallow Machine Learning Baselines on the RadChar Dataset

Edwin Eziagulu
Department of Electrical/Electronic Engineering, University of Port Harcourt, Port Harcourt, Nigeria

Abstract
The automatic classification of radar waveforms from raw In-Phase and Quadrature (IQ) samples represents a fundamental challenge in modern electronic warfare, cognitive radar, and spectrum sensing applications. While deep learning has demonstrated remarkable success in radar signal processing, a persistent methodological debate concerns whether direct learning from raw IQ sequences or transformed time-frequency representations yields superior classification performance. This paper presents a comprehensive comparative evaluation of multiple machine learning paradigms for radar waveform classification using the publicly available RadChar dataset. We systematically benchmark three shallow baseline models Logistic Regression, Random Forest, and Support Vector Machine with Radial Basis Function kernel against deep learning architectures including a raw IQ one-dimensional Convolutional Neural Network (1D CNN), a Short-Time Fourier Transform (STFT)-based two-dimensional CNN (2D CNN), and sequence learning models. Our experimental framework encompasses rigorous data preprocessing, stratified train-validation-test partitioning, handcrafted statistical feature extraction, time-frequency transformation, and extensive robustness analysis under controlled noise injection and Signal-to-Noise Ratio (SNR) band evaluation. The results demonstrate that the raw IQ 1D CNN achieves the strongest overall performance with 83.20% accuracy, 0.8316 weighted F1-score, and 0.9764 one-versus-rest Area Under the Receiver Operating Characteristic Curve (ROC-AUC), substantially outperforming the best shallow baseline (SVM: 64.33% accuracy, 0.6462 F1) by 18.87 percentage points. Critically, the STFT-based 2D CNN, while superior to all shallow models at 74.34% accuracy, remains inferior to direct raw IQ modelling, indicating that time-frequency transformation compresses discriminative information essential for waveform discrimination. Robustness analysis further reveals that the 1D CNN maintains remarkable stability under injected Gaussian noise (accuracy variation <0.3% across σ = 0.00–0.12) but exhibits a sharp performance degradation in the extreme low-SNR regime ([−20, −10] dB: 41.43% accuracy versus >94% at [−9, 0] dB and perfect classification above 0 dB). These findings establish empirical evidence that raw IQ deep learning provides the most effective validated pathway for radar waveform classification, while precisely identifying the operational boundaries of model capability. The complete experimental pipeline, including reusable utility code and model implementations, is designed for reproducibility and extensibility to real-world radar datasets.
***Keywords**: Radar signal processing, waveform classification, deep learning, convolutional neural networks, raw IQ data, time-frequency analysis, Short-Time Fourier Transform, machine learning baselines, robustness analysis, RadChar dataset.*

Date of Submission: 13-07-2026 Date of acceptance: 27-07-2026

I. INTRODUCTION

1.1. Background and Motivation

Radar systems constitute a critical sensing modality across defence, aviation, meteorology, and autonomous navigation domains, emitting structured electromagnetic waveforms that encode target range, velocity, and signature information [1]. The proliferation of diverse radar platforms—pulsed, continuous-wave, frequency-modulated, and phased-array systems—has created densely occupied electromagnetic spectra where the ability to automatically identify and classify radar waveforms is essential for electronic support measures, threat assessment, and spectrum management [2]. Traditional radar waveform classification relies on expert-engineered feature extraction from received signals, typically involving pulse descriptor word (PDW) estimation, parameter measurement, and rule-based matching against threat libraries [3]. However, conventional approaches face fundamental limitations: they require precise prior knowledge of waveform structures, degrade under low signal-to-noise ratio (SNR) conditions, and struggle to generalise across the expanding diversity of modern radar emissions [4].

The emergence of machine learning (ML) and deep learning (DL) has introduced transformative capabilities for radar signal processing, enabling data-driven feature extraction that circumvents manual engineering constraints [5]. Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and attention-based architectures have demonstrated superior performance in modulation recognition, pulse classification, and emitter identification tasks [6]–[8]. A central methodological question in this domain concerns the optimal input representation for deep learning models: should networks learn directly from raw IQ time-series samples, or should signals first be transformed into time-frequency representations such as spectrograms, scalograms, or Wigner-Ville distributions that explicitly expose frequency-domain structure [9]–[11]?

Raw IQ data preserves the complete amplitude and phase information of received signals, enabling end-to-end learning of discriminative features without intermediate transformation losses [12]. Conversely, time-frequency representations provide intuitive visualisations of signal structure and have proven effective in numerous radar classification studies [13]–[15]. The relative merits of these approaches remain contested, with existing literature offering conflicting evidence depending on dataset characteristics, network architectures, and evaluation protocols [16]–[18]. This ambiguity motivates rigorous, controlled comparison under standardised conditions to establish evidence-based guidelines for representation selection in radar waveform classification.

Figure 1 illustrates the central methodological question addressed in this paper. The left panel shows the raw IQ pathway, where the two-channel time-series is fed directly into a 1D CNN that learns temporal convolutional features. The right panel shows the time-frequency pathway, where the IQ signal is first transformed into an STFT spectrogram image before being processed by a 2D CNN. The figure highlights the information flow differences: raw IQ preserves all amplitude and phase information but requires the network to discover frequency structure implicitly, while the STFT exposes frequency content explicitly but discards fine-grained phase relationships and loses temporal resolution through windowing.

Figure 1: Conceptual comparison of (a) raw IQ 1D CNN and (b) STFT 2D CNN representation pathways for radar waveform classification.

1.2. The RadChar Dataset and Research Gap

The RadChar dataset addresses a critical need in radar machine learning research by providing a structured, publicly accessible collection of synthetic radar IQ signals with controlled parameter variation and annotated class labels [19]. Comprising 50,000 waveform samples across five distinct radar signal classes, each represented by 512 time-step IQ sequences with associated metadata including SNR, pulse width, pulse repetition interval, and delay parameters, RadChar enables reproducible benchmarking of classification algorithms under standardised conditions [19]. The dataset's synthetic generation ensures known ground-truth labels while incorporating realistic noise and parameter variation, supporting both algorithm development and robustness evaluation.

Despite the availability of radar datasets, existing comparative studies suffer from several limitations that this work addresses. First, many evaluations compare deep models against weak baselines or omit shallow machine learning entirely, preventing assessment of whether complexity increases yield proportionate gains [20]. Second, comparisons between raw IQ and time-frequency approaches often employ incompatible architectures or preprocessing pipelines, confounding representation effects with model capacity differences

[21]. Third, robustness analysis is frequently limited to aggregate accuracy metrics, obscuring performance variation across operational conditions that determine real-world deployability [22]. Fourth, reproducibility is hampered by incomplete documentation of experimental procedures, hyperparameters, and data partitioning strategies [23].

Figure 2 presents the taxonomy of machine learning approaches for radar waveform classification that motivates our comparative framework. The taxonomy organises methods along two dimensions: feature engineering (handcrafted versus learned) and representation domain (time-series versus time-frequency). This classification reveals that our study occupies the critical intersection of learned features in both domains, enabling direct isolation of representation effects.

Figure. 2: Taxonomy of machine learning approaches for radar waveform classification.

1.3. Contributions

This paper makes the following specific contributions:

i. Comprehensive Benchmarking Framework: We establish a reproducible experimental pipeline encompassing data audit, IQ reconstruction, normalisation, stratified partitioning, handcrafted feature extraction, time-frequency transformation, model training with checkpointing and early stopping, and multi-metric evaluation including accuracy, precision, recall, F1-score, and ROC-AUC.

ii. Controlled Representation Comparison: We directly compare raw IQ 1D CNN against STFT-based 2D CNN using architectures matched in capacity and training protocol, isolating representation effects from confounding factors. Both deep models are further benchmarked against three strong shallow baselines (Logistic Regression, Random Forest, SVM-RBF) operating on comprehensive statistical features.

iii. Extensive Robustness Characterisation: We evaluate the best-performing classifier under progressively stronger injected Gaussian noise and across SNR bands, revealing that performance is strongly condition-dependent rather than uniformly strong, with extreme low-SNR classification identified as the principal unresolved challenge.

iv. Evidence-Based Hierarchy: We demonstrate a clear performance ordering—Raw IQ 1D CNN > STFT 2D CNN > SVM-RBF > Random Forest > Logistic Regression—providing actionable guidance for radar waveform classification system design.

The remainder of this paper is organised as follows. Section II reviews related work in radar signal classification, representation learning, and robustness analysis. Section III presents the methodology including dataset description, preprocessing, model architectures, and evaluation metrics. Section IV reports experimental results with detailed tables and figures. Section V discusses implications, limitations, and future directions. Section VI concludes.

II. RELATED WORK

2.1. Deep Learning for Radar Waveform Classification

Deep learning has emerged as the dominant paradigm for automatic radar signal recognition, with architectures evolving from simple multilayer perceptrons to sophisticated convolutional, recurrent, and attention-based networks. Wang et al. [13] demonstrated that CNNs operating on Wigner-Ville distribution images achieve effective classification of rectangular, linear frequency modulation (LFM), and Barker-coded radar pulses, establishing the viability of time-frequency deep learning. This approach was extended by MathWorks [24] to include communication waveforms, showing that deep CNNs can automatically extract discriminative time-frequency features without manual engineering.

The spatial-temporal recurrent graph neural network (ST-RGNN) introduced by Nguyen et al. [7] represents a significant architectural advance, jointly modelling feeder topology and waveform dynamics to improve localisation accuracy by 15% over purely temporal methods. However, ST-RGNN requires synchronised multi-node measurements with ideal timestamp alignment, limiting deployment in communication-constrained environments. Hu et al. [3] applied deep graph learning to fault location, achieving 98.7% accuracy using voltage sag patterns propagated through graph structures, though this method assumes complete network state information.

Recent work has increasingly emphasised raw IQ processing. Yildirim and Kilanyaz [25] demonstrated that 1D CNNs outperform automatic classifiers for known low-probability-of-intercept (LPI) radar signals under white Gaussian noise, establishing the viability of direct sequence learning. The IQ Signal Transformer (IQST) proposed in [19] operates directly on raw IQ data without handcrafted features or image transforms, using multi-head attention mechanisms to capture long-term dependencies. Their multi-task learning framework achieved strong performance on RadChar for both classification and regression tasks, though direct comparison with time-frequency alternatives was not performed.

Figure 3 summarises the evolution of deep learning architectures for radar signal classification from 2015 to 2025, illustrating the transition from shallow multilayer perceptrons through CNN-based time-frequency methods to recent raw IQ and attention-based approaches. The timeline highlights the increasing emphasis on end-to-end learning and the methodological divergence between representation pathways that motivates our controlled comparison.

Figure 3: Timeline of deep learning architecture evolution for radar signal classification.

2.2. Representation Learning: Raw IQ Versus Time-Frequency

The choice between raw IQ and transformed representations remains actively debated. Proponents of time-frequency approaches argue that STFT spectrograms, continuous wavelet transform (CWT) scalograms, and Wigner-Ville distributions expose structural features—chirp rates, instantaneous frequency trajectories, modulation patterns—that are cryptic in raw time-domain samples [10], [13], [26]. CNNs applied to such representations leverage mature computer vision techniques and extensive pretraining resources, potentially improving convergence and generalisation [27].

Conversely, raw IQ advocates emphasise that transformation operations introduce information loss and distortion. The STFT's fixed time-frequency resolution trade-off may blur transient features, while the Wigner-Ville distribution suffers from cross-term interference in multi-component signals [28]. Phase information, critical for radar waveform discrimination, is often discarded or distorted in magnitude-based spectrograms [29]. Furthermore, transformation adds computational overhead that may preclude real-time operation on resource-constrained platforms [30].

A 2025 study on wireless signal classification [31] provides relevant comparative evidence, demonstrating that a data-driven LSTM operating on raw IQ samples significantly outperforms knowledge-driven approaches using statistical features, achieving 99.3% accuracy at 20 dB SNR versus 60.7% for feature-based methods. However, this work focused on communication-radar coexistence waveforms rather than pure radar classification, and did not include CNN-based raw IQ models or comprehensive time-frequency comparison.

Figure 4 illustrates the information preservation characteristics of raw IQ versus STFT representations for a representative radar waveform sample. Panel (a) shows the raw I and Q channels with clear pulse structure and phase variation. Panel (b) shows the STFT spectrogram, where frequency content is visible but phase information is lost through magnitude computation, and temporal resolution is reduced by windowing. Panel (c) shows the amplitude envelope derived from raw IQ, demonstrating that amplitude-phase relationships critical for waveform discrimination are preserved only in the raw representation.

Figure 4: Information preservation comparison for a representative radar waveform sample from the RadChar dataset.

2.3. Robustness and Operational Considerations

The gap between laboratory accuracy and operational performance has received increasing attention. Communication-agnostic machine learning models achieving high accuracy under ideal conditions may degrade substantially when deployed over realistic machine-to-machine (M2M) channels with latency, jitter, and packet loss [32]. In smart grid applications, this degradation can reduce fault detection accuracy from 98% to 70% or lower [32]. Similar concerns apply to radar classification, where platform motion, multipath propagation, and electronic countermeasures introduce signal distortions not captured by clean training data.

Edge AI implementations have demonstrated potential for robust localised inference, with lightweight neural networks on microcontroller-class devices achieving 85–92% detection accuracy with 50 ms inference latency [33]. However, edge deployment introduces accuracy-computation trade-offs through model quantisation and pruning, and isolated edge nodes cannot leverage cross-sensor pattern recognition or centralised model updates [33]. Federated learning offers distributed training without data centralisation, though non-IID data distributions across geographically diverse sensors cause local model divergence that degrades global performance [34].

Figure 5 presents the conceptual framework for robustness evaluation adopted in this study. The framework distinguishes between two types of robustness testing: (a) controlled noise injection, where synthetic Gaussian perturbation is added to clean signals to assess algorithmic stability; and (b) SNR band stratification, where performance is evaluated across naturally occurring signal quality regimes in the dataset. This dual approach provides complementary perspectives on model reliability.

Figure 5: Robustness evaluation framework.

2.4. Research Positioning

This work occupies a distinct position in the literature by providing the first comprehensive, controlled comparison of raw IQ 1D CNN, STFT 2D CNN, and strong shallow baselines on a standardised radar dataset with extensive robustness characterisation. Unlike prior studies that compare representations using mismatched architectures or evaluate robustness through aggregate metrics alone, we isolate representation effects through capacity-matched designs and reveal condition-specific performance boundaries that determine operational viability.Table I demonstrates that no prior study simultaneously evaluates raw IQ deep learning, time-frequency deep learning, strong shallow baselines, and comprehensive robustness analysis within a controlled comparison framework. Our work fills this gap by providing all four elements in a single, reproducible experimental design.

Study	Raw IQ	Time-Freq	Shallow Baselines	Robustness Analysis	Controlled Comparison
Wang et al. [13]	No	Yes (WVD)	No	Limited	No
Yildirim &Kilanyaz [25]	Yes (1D CNN)	No	Yes (auto)	Noise only	Partial
IQST [19]	Yes (Transformer)	No	No	None	No
2025 Wireless Study [31]	Yes (LSTM)	No	Yes (statistical)	SNR only	No
This Work	**Yes (1D CNN)**	**Yes (STFT)**	**Yes (3 models)**	**Noise + SNR**	**Yes**

Table 2.1: Positioning of This Work Relative to Prior Studies

Table 2.1 demonstrates that no prior study simultaneously evaluates raw IQ deep learning, time-frequency deep learning, strong shallow baselines, and comprehensive robustness analysis within a controlled comparison framework. Our work fills this gap by providing all four elements in a single, reproducible experimental design.

III. METHODOLOGY

3.1 Dataset Description and Audit

The RadChar dataset comprises 50,000 synthetic radar waveform samples organised in three comma-separated value (CSV) files: real.csv containing in-phase (I) components, imag.csv containing quadrature (Q) components, and labels.csv containing class annotations and metadata [19]. Each waveform consists of 512 time steps, yielding real and imaginary matrices of shape (50000,512) . The label file contains six columns: signal class, number of pulses, pulse width, time delay, pulse repetition interval, and signal-to-noise ratio.

The dataset was generated through parametric waveform synthesis with controlled variation of radar characteristics, ensuring adherence to the Nyquist-Shannon sampling theorem with a sampling rate of 3.2 MHz [19]. Five distinct waveform classes are represented, with parameters sampled uniformly across specified ranges to create diverse, realistic signal variation. Additive white Gaussian noise (AWGN) was applied to simulate varying SNR conditions, and unity average power normalisation ensures consistent signal scaling.

3.1.1 Dataset Audit Results. Table IIsummarises the dataset characteristics verified during preprocessing. The real and imaginary matrices exhibit identical shape (50000,512) , confirming consistent sample construction. The label matrix shape (50000,6) confirms one-to-one correspondence between waveform samples and annotations. The signal type column was identified as the primary classification target, with five distinct classes encoded as integers {0,1,2,3,4} . Numerical metadata columns including number_of_pulses, pulse_width, time_delay, pulse_repetition_interval, and signal_to_noise_ratio were preserved for robustness analysis.As shown in Table 3.1, the dataset audit confirms structural consistency across all components. The matching row counts across real, imaginary, and label files (50,000 each) ensure valid sample-label pairing. The reconstructed IQ tensor shape (50000,512,2) provides the standardised input format for all raw IQ models, while the preserved metadata enables SNR-stratified robustness evaluation in Section IV-F.

S/N	Dataset Attribute	Observed Value	Interpretation
i	Real signal matrix shape	(50000,512)	50,000 I-channel samples, 512 time steps each
ii	Imaginary signal matrix shape	(50000,512)	50,000 Q-channel samples, 512 time steps each
iii	Label matrix shape	(50000,6)	50,000 annotation rows, 6 metadata fields
iv	Number of samples	50,000	Sufficient volume for ML training and evaluation
v	Number of time steps per waveform	512	Fixed-length sequence suitable for temporal modelling
vi	Signal label column	signal_type	Primary classification target identified
vii	Numerical metadata columns	number_of_pulses, pulse_width, time_delay, pulse_repetition_interval, signal_to_noise_ratio	Support descriptive and robustness analysis
viii	Number of signal classes	5	Five-class radar waveform recognition task
ix	Reconstructed IQ tensor shape	(50000,512,2)	Two-channel I/Q tensor successfully formed
x	Encoded label shape	(50000,)	One-dimensional numerical

			class vector
xi	Class labels detected	{0,1,2,3,4}	Five waveform classes prepared for supervised learning
xii	Preprocessing status	Successful	All artifacts generated and saved for downstream modelling

Table 3.1: Dataset Audit Summary

3.2 Data Preprocessing and Partitioning

3.2.1 IQ Reconstruction. The real and imaginary components were reconstructed into a unified two-channel IQ tensor. For the i -th radar sample, the reconstructed IQ sequence is expressed as:

(3.1)

where denotes the in-phase component, the quadrature component, t the time index, and N=50000 the total number of samples. The full dataset tensor assumes the form:

(3.2)

with T=512 time steps and the final dimension representing the two IQ channels. This structure ensures that models learn directly from the physical IQ composition of radar signals rather than from abstracted feature tables.

3.2.2 Per-Sample Normalization. Each IQ sample was normalized independently to reduce amplitude-scale variation across radar returns caused by propagation distance, target reflectivity, and receiver gain settings. The normalization operation is defined as:

(3.3)

where is the normalized sample, the sample mean, the sample standard deviation, and $\epsilon=10-8$ a small stability constant preventing division by zero. This z-score normalization preserves waveform shape while standardizing amplitude scale, a critical preprocessing step given that radar returns may vary by orders of magnitude depending on operational geometry.

3.2.3 Stratified Partitioning. The dataset was partitioned into training, validation, and test subsets using stratified sampling to preserve class distribution:

(3.4)

The implemented split ratios were test_size =0.20 and val_size =0.10 , yielding 35,000 training samples (70%), 5,000 validation samples (10%), and 10,000 test samples (20%). Stratification ensures that each subset retains identical class proportions, preventing biased evaluation from accidental class exclusion or uneven representation. The class distribution was verified as perfectly balanced with 10,000 samples per class in the full dataset and proportional representation in each partition.

S/N	Dataset Partition	Number of Samples	Percentage of Total	Purpose
i	Training set	35,000	70%	Model fitting and parameter learning across all classifiers
ii	Validation set	5,000	10%	Model tuning, checkpoint selection, and generalisation monitoring
iii	Test set	10,000	20%	Final unbiased performance evaluation and comparative benchmarking
iv	Total dataset	50,000	100%	Complete RadChar sample size used in this study

Table 3.2: Train, Validation, And Test Split Summary

Table 3.2 presents the data partitioning results. The 70:10:20 split provides sufficient training data (35,000 samples) for deep model convergence while reserving substantial held-out test data (10,000 samples) for reliable statistical evaluation. The stratified design ensures that class proportions are preserved across all partitions, a critical requirement for fair comparison given that imbalanced splits could artefactually inflate or depress reported accuracy.

3.3 Feature Representations

3.3.1Handcrafted Statistical Features. For shallow baseline models, a comprehensive set of statistical features was extracted from the IQ signals, amplitude envelope, and phase sequence. The amplitude envelope was computed as:

(3.5)

and the instantaneous phase as:

(3.6)

Feature extraction computed, for each of the four derived signals (I, Q, amplitude, phase): mean, standard deviation, maximum, minimum, median, skewness, and kurtosis. Additionally, zero-crossing rates were computed for the I and Q channels. This yielded a 30-dimensional feature vector per sample, providing shallow models with comprehensive statistical characterization of waveform structure.

3.3.2Short-Time Fourier Transform. For the 2D CNN, IQ signals were transformed into STFT spectrograms. The complex baseband signal was formed as $s(t)=I(t)+jQ(t)$, and the STFT computed with 64-sample segments and 50% overlap:

(3.7)

where $w(t-\tau)$ is a sliding window function. The magnitude spectrogram was log-compressed ($\log(1+|z|)$) to enhance dynamic range, producing image tensors of shape (64,17,1) suitable for 2D CNN processing.

Figure 6 presents representative STFT spectrograms for each of the five waveform classes, illustrating the visual discriminability that motivates time-frequency deep learning. The spectrograms reveal distinct frequency-time patterns: Class 0 exhibits narrowband concentrated energy; Class 1 shows linear frequency modulation; Class 2 displays stepped frequency structure; Class 3 presents broadband pulse characteristics; and Class 4 demonstrates periodic pulse repetition patterns. These visual distinctions suggest that spectrogram-based classification should be feasible, though the fixed resolution may blur fine temporal details.

Figure 6: Representative STFT spectrograms (64×17 pixels, log-compressed magnitude)

3.4 Model Architectures

3.4.1 1D CNN for Raw IQ Classification. The raw IQ 1D CNN processes the two-channel time series directly through three convolutional stages. The architecture comprises:

Input layer: (512,2) tensor
Conv1D(32 filters, kernel_size=7, padding="same") → BatchNorm → ReLU → MaxPool1D(2)
Conv1D(64 filters, kernel_size=5, padding="same") → BatchNorm → ReLU → MaxPool1D(2)
Conv1D(128 filters, kernel_size=3, padding="same") → BatchNorm → ReLU
GlobalAveragePooling1D → Dense(128, ReLU) → Dropout(0.3)
Output: Dense(5, softmax)

The model was compiled with sparse categorical cross-entropy loss and Adam optimiser (learning rate 10^{-3}). Training employed class weighting to address any residual imbalance, ModelCheckpoint saving the best validation model, ReduceLROnPlateau (factor 0.5, patience 5), and EarlyStopping (patience 12, restore best weights).

3.4.2 2D CNN for STFT Spectrogram Classification. The STFT-based 2D CNN employs an analogous architectural philosophy adapted for image input:

Input layer: (64,17,1) spectrogram image
Conv2D(32, (3,3), padding="same") → BatchNorm → ReLU → MaxPool2D((2,2))
Conv2D(64, (3,3), padding="same") → BatchNorm → ReLU → MaxPool2D((2,2))
Conv2D(128, (3,3), padding="same") → BatchNorm → ReLU
GlobalAveragePooling2D → Dense(128, ReLU) → Dropout(0.4)
Output: Dense(5, softmax)

The higher dropout rate (0.4 versus 0.3) accommodates the smaller spatial dimensions and potentially higher overfitting risk of spectrogram features. Training followed the same callback protocol with patience adjusted to 10 epochs.

3.4.3 Shallow Baseline Models. Three conventional classifiers were implemented:

i. *Logistic Regression*: L2-regularised multinomial logistic regression with StandardScaler preprocessing, maximum 2000 iterations.
ii. *Random Forest*: 300 estimators, no maximum depth constraint, Gini impurity criterion.
iii. *SVM-RBF*: Support Vector Machine with Radial Basis Function kernel (C=2.0), StandardScaler preprocessing, probability estimation enabled.

All shallow models operated on the 30-dimensional handcrafted feature matrix. The SVM and Logistic Regression were implemented within Scikit-learn pipelines ensuring consistent scaling.

Figure 7 presents the complete architecture diagrams for the three primary models evaluated in this study: (a) raw IQ 1D CNN, (b) STFT 2D CNN, and (c) shallow baseline pipeline. The figure emphasises the structural parallels between the two CNN variants (both use three convolutional stages, batch normalisation, global pooling, and dense classification) while highlighting the fundamental difference in input dimensionality and convolution type.

Figure 7: Architecture diagrams for the three primary model families evaluated.

3.5 Evaluation Metrics

Classification performance was assessed through multiple complementary metrics:

i. Accuracy: Proportion of correctly classified samples.
ii. Weighted Precision, Recall, and F1-Score: Macro-averaged metrics accounting for class support:

(3.8)

where P_k is precision for class k ,R_k is recall, w_k the class support weight, and C=5 the number of classes.

iii. ROC-AUC (One-Versus-Rest): Area under the receiver operating characteristic curve computed for each class against all others, averaged with weighting by support.
iv. Training and Inference Time: Wall-clock time for model fitting and prediction, measured on standardised hardware.
v. Parameter Count: Total trainable parameters as a proxy for model complexity.

Robustness was evaluated through:

vi. Injected Gaussian Noise: Classification accuracy under additive noise with $\sigma \in \{0.00, 0.02, 0.05, 0.08, 0.12\}$.

vii. SNR Band Analysis: Performance stratified by signal-to-noise ratio intervals: [−20,−10] dB, [−9,0] dB, [1,10] dB, and [11,20] dB.

IV. EXPERIMENTAL RESULTS

4.1 Baseline Machine Learning Results

Table IV presents the performance of the three shallow baseline models evaluated on the held-out test set. Logistic Regression achieved 59.16% accuracy with 0.5887 weighted F1-score, establishing the lower bound for linear discriminative methods on handcrafted features. The Random Forest improved to 61.87% accuracy (0.6180 F1), confirming that ensemble methods capture non-linear feature interactions inaccessible to linear models. The Support Vector Machine with RBF kernel achieved the strongest baseline performance at 64.33% accuracy and 0.6462 weighted F1-score, with ROC-AUC of 0.8943.

Figure 8 presents the confusion matrices for the three baseline models, revealing per-class discrimination patterns. All models show strongest performance on Class 0 and weakest on Class 2, with the SVM exhibiting the most balanced diagonal but still significant confusion between Classes 1 and 2. The pattern of misclassification is consistent across models, suggesting that certain waveform pairs possess similar statistical signatures that are difficult to separate in the handcrafted feature space.

S/N	Model	Accuracy	Precision (Weighted)	Recall (Weighted)	F1-Score (Weighted)	ROC-AUC (OvR)	Training Time (s)	Inference Time (s)
i	Logistic Regression	0.5916	0.5929	0.5916	0.5887	0.8684	13.15	0.028
ii	Random Forest	0.6187	0.6205	0.6187	0.6180	0.8859	108.60	0.676
iii	SVM (RBF Kernel)	0.6433	0.6604	0.6433	0.6462	0.8943	883.97	21.16

Table 4.1: Baseline Model Performance Comparison

As shown in Table 4.1, the performance hierarchy among shallow models is clear: SVM-RBF > Random Forest > Logistic Regression. The SVM's kernel trick enables non-linear decision boundaries that better separate the 30-dimensional feature space, yielding the highest accuracy but at substantial computational cost. The 883.97-second training time and 21.16-second inference time for SVM contrast sharply with the 13.15-second training and 0.028-second inference of Logistic Regression, highlighting the accuracy-efficiency trade-off inherent in shallow model selection.

Figure 8 presents the confusion matrices for the three baseline models, revealing per-class discrimination patterns. All models show strongest performance on Class 0 and weakest on Class 2, with the SVM exhibiting the most balanced diagonal but still significant confusion between Classes 1 and 2. The pattern of misclassification is consistent across models, suggesting that certain waveform pairs possess similar statistical signatures that are difficult to separate in the handcrafted feature space.

Figure 8: Baseline model confusion matrices for (a) Logistic Regression, (b) Random Forest, and (c) SVM-RBF, illustrating per-class discrimination patterns.

The SVM's superior performance comes at substantial computational cost: training required 883.97 seconds versus 108.60 seconds for Random Forest and 13.15 seconds for Logistic Regression. Inference time similarly scales, with SVM requiring 21.16 seconds for 10,000 test samples compared to 0.028 seconds for Logistic Regression. This computational asymmetry is relevant for deployment scenarios where training efficiency or real-time inference constraints apply. Class-wise analysis reveals that all baselines struggle particularly with Class 2, where SVM recall drops to 0.42 and F1-score to 0.48, indicating that certain waveform categories possess subtle statistical signatures not captured by handcrafted features.

4.2 Raw IQ 1D CNN Results

The raw IQ 1D CNN produced the strongest classification performance in the study. Training history shows training accuracy increasing from approximately 74.28% in the first epoch to 87.50% in later epochs, with validation accuracy stabilising in the low 0.83 range. Despite occasional validation loss spikes, the model learned a stable, generalisable representation of waveform classes.

S/N	Model	Accuracy	Precision (Weighted)	Recall (Weighted)	F1-Score (Weighted)	ROC-AUC (OvR)	Remark
i	Raw IQ 1D CNN	0.8320	0.8360	0.8320	0.8316	0.9764	Best validated classifier

Table 4.2: Final 1d Cnn Performance Metrics

Table 4.2summarises the final test performance of the raw IQ 1D CNN. The model achieves 83.20% accuracy, 0.8360 weighted precision, 0.8320 weighted recall, 0.8316 weighted F1-score, and 0.9764 ROC-AUC. These metrics represent a substantial improvement over all baseline models and establish the 1D CNN as the best-performing classifier in this study.

Figure 9: 1D CNN training and validation curves showing (a) accuracy and (b) loss versus epoch, illustrating convergence behaviour and generalization stability.

Figure 9 presents the training and validation curves for the 1D CNN. Panel (a) shows accuracy versus epoch, with training accuracy rising steadily from 0.74 to 0.88 and validation accuracy converging to approximately 0.83. Panel (b) shows loss curves, where training loss decreases monotonically while validation loss exhibits spikes characteristic of the learning rate reduction schedule but ultimately stabilises. The divergence between training and validation accuracy (approximately 4 percentage points) indicates mild overfitting that is effectively controlled by the dropout and early stopping mechanisms.

The final test performance—83.20% accuracy, 0.8316 weighted F1, and 0.9764 ROC-AUC—represents a substantial improvement over all baseline models. Compared to the best shallow baseline (SVM), the 1D CNN improves accuracy by 18.87 percentage points and F1-score by 0.1854. The ROC-AUC of 0.9764 indicates excellent class separation across decision thresholds, with near-perfect discrimination capability for most waveform categories.

The magnitude of improvement is too large to attribute to marginal effects or random variation. It demonstrates that direct learning from raw IQ sequences captures richer discriminative information than summary statistics can encode. The convolutional filters learn local temporal patterns—pulse edges, frequency transitions, amplitude-phase relationships—that are essential for waveform discrimination but are compressed or lost in handcrafted feature reduction.

4.3 STFT 2D CNN Results

The STFT-based 2D CNN achieved 74.34% accuracy, 0.7435 weighted precision, 0.7434 weighted recall, and 0.7410 weighted F1-score, with ROC-AUC of 0.9474. The model comprised 110,725 total parameters and completed training in 29 epochs.

Figure 10 presents the confusion matrix for the 2D CNN. The matrix reveals that the model excels particularly on Class 0 (F1 = 0.87) and Class 4 (F1 = 0.81), but underperforms on Classes 1 and 2. This pattern suggests that the STFT representation preserves strong discriminatory information for some waveform categories while compressing or blurring distinctions for others. The spectrogram's fixed time-frequency resolution may inadequately capture transient features critical for discriminating certain pulse structures.

Table

S/N	Model	Input Representation	Accuracy	Precision (Weighted)	Recall (Weighted)	F1-Score (Weighted)	ROC-AUC (OvR)	Parameters	Final Epoch
i	2D CNN	STFT Spectrogram	0.7434	0.7435	0.7434	0.7410	0.9474	110,725	29

Table 4.3: Stft 2d Cnn Performance Metrics

Table 4.3 presents the complete performance metrics for the STFT 2D CNN. The model achieves 74.34% accuracy with 0.7410 weighted F1-score, clearly surpassing all shallow baselines but remaining 8.86 percentage points below the raw IQ 1D CNN. The 110,725 parameters and 29-epoch convergence indicate efficient training, though the smaller spatial dimensions of the spectrogram (64×17 versus 512×2 for raw IQ) may limit the network's capacity to learn fine-grained discriminative features.

Figure 10: 2D CNN confusion matrix showing per-class classification performance on STFT spectrogram inputs.

Class-wise analysis reveals that the 2D CNN excels particularly on Class 0 (F1 = 0.87) and Class 4 (F1 = 0.81), but underperforms on Classes 1 and 2. This pattern suggests that the STFT representation preserves strong discriminatory information for some waveform categories while compressing or blurring distinctions for others. The spectrogram's fixed time-frequency resolution may inadequately capture transient features critical for discriminating certain pulse structures.

The 2D CNN's performance—while clearly superior to all shallow baselines—remains 8.86 percentage points below the raw IQ 1D CNN in accuracy and 0.0906 below in F1-score. This gap is significant and consistent, indicating that for the RadChar dataset and classification task, the STFT transformation discards information that the 1D CNN retains and exploits from raw sequences.

4.4 Comparative Benchmarking

Table 4.4 consolidates the principal classifiers with complete metrics, establishing a clear performance hierarchy.

Model	**Accuracy**	**Precision (Weighted)**	**Recall (Weighted)**	**F1-Score (Weighted)**	**ROC-AUC (OvR)**	**Representation**
Raw IQ 1D CNN	0.8320	0.8360	0.8320	0.8316	0.9764	Raw IQ Tensor
STFT 2D CNN	0.7434	0.7435	0.7434	0.7410	0.9474	STFT Spectrogram
SVM (RBF)	0.6433	0.6604	0.6433	0.6462	0.8943	Handcrafted Features

Random Forest	0.6187	0.6205	0.6187	0.6180	0.8859	Handcrafted Features
Logistic Regression	0.5916	0.5929	0.5916	0.5887	0.8684	Handcrafted Features

Table 4.4: Comparative Benchmark Of Validated Models

Figure 11: Overall model ranking chart comparing accuracy, F1-score, and ROC-AUC across all validated models.

Figure 11 shows the overall model ranking chart, visualising the performance gaps across all metrics. The bar chart confirms that the raw IQ 1D CNN dominates every metric, with the STFT 2D CNN occupying a clear second place, and the shallow models clustered in the 0.59–0.65 accuracy range. The ROC-AUC metric shows the largest absolute separation between deep and shallow models, suggesting that deep architectures achieve superior class separation confidence.The hierarchy is unambiguous: Raw IQ 1D CNN > STFT 2D CNN > SVM-RBF > Random Forest > Logistic Regression. Deep learning consistently outperforms shallow learning, and direct raw IQ modelling outperforms time-frequency transformation. The SVM, despite being the strongest conventional method, remains substantially below both deep architectures, with the additional penalty of high computational cost.

4.5 Robustness Under Injected Gaussian Noise

The best classifier (raw IQ 1D CNN) was evaluated under progressively stronger additive Gaussian noise to assess stability under signal degradation. Table VII presents the results.

S/N	Noise Sigma (σ)	Accuracy	Precision (Weighted)	Recall (Weighted)	F1-Score (Weighted)	ROC-AUC (OvR)
i	0.00	0.8320	0.8360	0.8320	0.8316	0.9764
ii	0.02	0.8316	0.8357	0.8316	0.8313	0.9764
iii	0.05	0.8320	0.8363	0.8320	0.8317	0.9763
iv	0.08	0.8308	0.8352	0.8308	0.8306	0.9765

v	0.12	0.8300	0.8345	0.8300	0.8298	0.9764

Table 4.5: Noise Sigma Robustness Results

Figure 12: Accuracy versus injected noise sigma, showing remarkable stability across the tested perturbation range.

The results from Table 4.5 are striking: accuracy varies by less than 0.3% across the entire noise range, from 0.8320 at σ=0.00 to 0.8300 at σ=0.12. Weighted F1-score and ROC-AUC exhibit similarly negligible variation. This stability indicates that the learned convolutional filters capture waveform characteristics that are highly invariant to moderate synthetic noise perturbation. The model does not rely on fragile amplitude or phase features that would be disrupted by additive noise; instead, it learns structural patterns—pulse timing relationships, relative amplitude profiles, phase continuity—that persist under corruption.Figure 12 visualizes the stability trend, plotting accuracy, F1-score, and ROC-AUC versus noise sigma. The near-horizontal lines confirm that the model does not rely on fragile amplitude or phase features that would be disrupted by additive noise; instead, it learns structural patterns—pulse timing relationships, relative amplitude profiles, phase continuity—that persist under corruption

From an operational perspective, this robustness is highly desirable. Radar systems routinely encounter thermal noise, clutter, and interference; a classifier that maintains performance without explicit noise-robust training demonstrates strong inherent generalisation.

4.6 Robustness Across SNR Bands

While injected noise tests synthetic perturbation, SNR band analysis evaluates performance under realistic signal quality conditions present in the original dataset. Table VIII presents results stratified by SNR intervals.

SNR Band (dB)	Number of Samples	Accuracy	Precision (Weighted)	Recall (Weighted)	F1-Score (Weighted)	ROC-AUC (OvR)
[−20,−10]	2,648	0.4143	0.4066	0.4143	0.3976	0.7152
[−9,0]	2,378	0.9458	0.9464	0.9458	0.9451	0.9973
[1,10]	2,480	1.0000	1.0000	1.0000	1.0000	1.0000
[11,20]	2,494	1.0000	1.0000	1.0000	1.0000	1.0000

Table 4.6: Snr Band Robustness Results

Figure 13: Accuracy versus SNR band, revealing sharply differentiated performance regimes.

In Table 4.6, SNR band analysis reveals a highly heterogeneous performance profile that fundamentally qualifies interpretation of aggregate metrics. In the extreme low-SNR band [−20,−10] dB, accuracy collapses to 41.43% with F1-score of 0.3976—only marginally above random chance for five classes (20%). However, performance rises sharply to 94.58% accuracy in the [−9,0] dB band, and achieves perfect classification (100% accuracy, F1, and ROC-AUC) in both positive SNR bands.Figure 13 visualizes this dramatic performance variation, plotting accuracy versus SNR band. The step-function-like transition from failure to perfection identifies three distinct operational regimes: the extreme low-SNR regime where waveform structure is deeply submerged in noise; the moderate-SNR regime where structure is partially recoverable; and the high-SNR regime where classes are perfectly separable.

This pattern is the most scientifically significant finding of the robustness analysis. It demonstrates that the classification challenge is not uniformly difficult but is concentrated in the extreme low-SNR regime. Above 0 dB, the waveform classes become perfectly separable by the learned representation; between −9 and 0 dB, performance is strong but not flawless; below −10 dB, discrimination essentially fails. This identifies the extreme low-SNR boundary as the principal frontier for future methodological improvement.

4.7 Best Model Selection and Justification

Based on the comprehensive evidence, the raw IQ 1D CNN is selected as the best model. Table IX summarises the selection criteria.

Criterion	Evidence from Results	Justification
Highest overall classification performance	Accuracy = 0.8320, Weighted F1 = 0.8316, ROC-AUC = 0.9764	Achieves best validated predictive performance among all reported models
Strong robustness to injected noise	Accuracy remained between 0.8300 and 0.8320 across σ=0.00 –0.12	Very small performance variation indicates noise-resilient learned representation
Superior performance over shallow baselines	Outperforms SVM, Random Forest, and Logistic Regression by clear margin	Confirms direct learning from raw IQ signals is more effective than handcrafted-feature baselines
Better performance than validated time-frequency model	1D CNN accuracy = 0.8320 versus STFT 2D CNN accuracy = 0.7434	Raw IQ representation preserves more discriminative information than STFT-based image representation
Strong performance across moderate and high SNR bands	Accuracy = 0.9458 in [−9,0] dB, and 1.0000 in [1,10] dB and [11,20] dB	Excellent operational reliability outside extreme low-SNR regime
Secondary validated model	STFT 2D CNN: Accuracy = 0.7434, Weighted F1 = 0.7410, ROC-AUC = 0.9474	Remains second-best fully summarised deep model, confirming usefulness of time-frequency learning

Table 4.7: Best Model Selection Criteria And Justification

Table 4.7 presents the six selection criteria that collectively justify the raw IQ 1D CNN as the best model. The first four criteria address absolute performance, robustness, and comparative advantage; the fifth criterion evaluates operational reliability across realistic conditions; and the sixth criterion acknowledges the STFT 2D CNN as a credible alternative for specific use cases.

The STFT 2D CNN remains a credible alternative when visual interpretability or explicit frequency-domain analysis is required. However, its lower classification performance indicates that the raw IQ representation is preferable for pure accuracy maximisation. The SVM establishes a respectable shallow benchmark but is clearly superseded by deep architectures.

V. DISCUSSION

5.1 Why Raw IQ Outperforms Time-Frequency Transformation

The empirical superiority of raw IQ 1D CNN over STFT 2D CNN warrants careful interpretation. Several mechanisms may contribute:

5.1.1 Information Preservation. The STFT applies a linear time-frequency decomposition with fixed resolution determined by window length. For the RadChar parameters (64-sample segments, 50% overlap), this yields 64 frequency bins and 17 time frames. The resulting spectrogram dimensionality (64×17=1088) is substantially lower than the raw sequence dimensionality (512×2=1024) , but more critically, the STFT discards phase information through magnitude computation and loses fine temporal structure through windowing and downsampling. Phase relationships between I and Q components are fundamental to radar waveform discrimination—pulse compression characteristics, frequency modulation trajectories, and phase-coded sequences all encode critical class information in the complex plane [28], [29].

5.1.2 Adaptive Feature Learning. The 1D CNN learns filters adaptively from data, potentially discovering features optimised for the specific waveform set. In contrast, the STFT employs a fixed sinusoidal basis that may be suboptimal for the transient, pulsed structures characteristic of radar signals. Wavelet transforms with adaptive basis selection might narrow this gap, though at increased computational cost [26].

5.1.3 Architectural Efficiency. The 1D CNN's global average pooling before classification encourages feature maps to correspond to class-specific activation patterns, providing inherent regularisation. The 2D CNN's comparable architecture may struggle with the limited spatial dimensions of compact spectrograms, where 2D convolutional receptive fields may be oversized relative to feature scale.

These findings align with broader trends in signal processing deep learning. A 2025 wireless signal classification study [31] similarly found that raw IQ LSTM networks substantially outperformed feature-based approaches, achieving 99.3% versus 60.7% accuracy at comparable SNR. However, that work did not directly compare raw IQ CNNs against spectrogram CNNs, leaving our comparison as a distinctive contribution.

5.2 The Extreme Low-SNR Challenge

The sharp performance collapse below −10 dB SNR identifies a critical operational boundary. At these noise levels, waveform structure becomes deeply submerged, with signal energy approaching or falling below noise floor. The 41.43% accuracy in [−20,−10] dB—twice random chance but far from operational viability—suggests that the learned features, while robust to moderate noise, lack the statistical power for reliable discrimination when signal structure is severely degraded.

This limitation is not unique to our approach. The wireless signal classification study [31] similarly observed that "none of the discussed classification algorithms delivers reasonable performance at very low and negative SNR levels, where the signal's distinguishing characteristics become increasingly obscured by noise." Their data-driven DNN achieved 90% accuracy only above 13.5 dB SNR, with performance approaching random output below 0 dB.

Addressing this challenge may require: (i) stronger denoising preprocessing, such as autoencoder-based signal restoration [33]; (ii) curriculum learning progressing from high-SNR to low-SNR samples; (iii) multi-scale feature fusion combining coarse energy detection with fine structure analysis; or (iv) ensemble methods aggregating predictions across temporal windows to exploit sporadic signal emergence above noise.

5.3 Computational and Deployment Considerations

While accuracy is paramount, practical deployment demands consideration of computational efficiency. The raw IQ 1D CNN's parameter count (~120K) is comparable to the STFT 2D CNN (110,725), indicating similar memory footprints. Training time favours the 1D CNN due to simpler convolutional operations, though both complete within practical limits on GPU hardware.

The shallow baselines offer contrasting trade-offs. Logistic Regression trains in 13 seconds and infers in 28 ms—orders of magnitude faster than deep models—but at substantial accuracy cost. The SVM requires 884 seconds training and 21 seconds inference, making it the least efficient approach despite moderate accuracy. For resource-constrained platforms (embedded systems, UAV payloads), the accuracy-efficiency trade-off may favour compressed deep models or efficient shallow alternatives, though our results suggest that accuracy degradation would be significant.

5.4 Limitations and Future Work

Several limitations should be acknowledged. First, the study was conducted on synthetic RadChar data; validation on real-world radar recordings with platform motion, multipath, and interference remains essential for operational confirmation. Second, the time-frequency comparison was limited to STFT; CWT scalograms, Wigner-Ville distributions, and learned representations (e.g., variational autoencoder latent spaces) may yield different relative performance. Third, sequence models (BiLSTM, CNN-LSTM) were implemented but final consolidated metrics were not fully captured in the current experimental summary, preventing their inclusion in the benchmark hierarchy. Fourth, the denoising autoencoder was trained but final reconstruction metrics require completion for quantitative assessment.

Future directions include: (i) comprehensive time-frequency representation comparison (CWT, WVD, chirplet transform); (ii) hybrid architectures fusing raw IQ and time-frequency streams; (iii) attention mechanisms for interpretable feature importance; (iv) transfer learning from synthetic to real radar data; and (v) edge deployment optimisation through quantisation and pruning.

VI. CONCLUSION

This paper has presented a comprehensive comparative evaluation of machine learning paradigms for radar waveform classification on the RadChar dataset, establishing that direct deep learning from raw IQ signals outperforms both handcrafted-feature shallow baselines and time-frequency transformed representations. The raw IQ 1D CNN achieved 83.20% accuracy, 0.8316 weighted F1-score, and 0.9764 ROC-AUC, surpassing the best shallow baseline (SVM-RBF: 64.33% accuracy) by 18.87 percentage points and the STFT 2D CNN (74.34% accuracy) by 8.86 percentage points. Robustness analysis revealed remarkable stability under injected Gaussian noise but identified the extreme low-SNR regime ([−20,−10] dB) as the critical performance boundary, where accuracy drops to 41.43% versus perfect classification above 0 dB.

These findings contribute empirical evidence to the representation debate in radar signal processing, supporting raw IQ deep learning as the preferred approach for waveform classification tasks where accuracy maximisation is paramount. The precise identification of operational performance boundaries—excellent above 0 dB, strong between −9 and 0 dB, and challenged below −10 dB—provides actionable guidance for system designers and highlights the low-SNR frontier as the priority for future methodological advance. The complete experimental framework, including reusable utility code and standardised evaluation protocols, is designed to support reproducible extension to real-world radar datasets and operational deployment scenarios.

REFERENCES

[1] S. Haykin, *Cognitive Radar: The Knowledge-Aided Fully Adaptive Approach*. Hoboken, NJ, USA: Wiley, 2012.
[2] M. A. Richards, *Fundamentals of Radar Signal Processing*, 2nd ed. New York, NY, USA: McGraw-Hill, 2014.
[3] J. Hu, H. He, and F. Blaabjerg, "Fault location and classification for distribution systems based on deep graph learning methods,"*J. Mod. Power Syst. Clean Energy*, 2023. [Online]. Available: https://doi.org/10.35833/MPCE.2022.000204Login to Jenni
[4] K. O'Shea and R. Nash, "An introduction to convolutional neural networks,"*arXiv preprint arXiv:1511.08458*, 2015.
[5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning,"*Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[6] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Convolutional radio modulation recognition networks," in *Proc. Int. Conf. Eng. Appl. Neural Netw.*, 2016, pp. 213–226.
[7] B. L. H. Nguyen, T. Vu, T.-T. Nguyen, M. Panwar, and R. Hovsapian, "Spatial-temporal recurrent graph neural networks for fault diagnostics in power distribution systems,"*IEEE Access*, vol. 10, pp. 1–1, 2022.
[8] Z. Chen, S. Cai, and A. P. S. Meliopoulos, "A real-time deep learning-based fault diagnosis framework in power distribution systems with PVs," in *Proc. IEEE PES Innov. Smart Grid Technol. (ISGT)*, Washington, DC, USA, Feb. 2024, pp. 1–5.
[9] F. M. Almasoudi, "Enhancing power grid resilience through real-time fault detection and remediation using advanced hybrid machine learning models,"*Sustainability*, vol. 15, no. 11, p. 8723, 2023.
[10] C. G. Arsoniadis and V. C. Nikolaidis, "A machine learning based fault location method for power distribution systems using wavelet scattering networks,"*Sustain. Energy, Grids Netw.*, vol. 40, p. 101551, 2024.
[11] Y. D. Mamuya, Y. D. Lee, J. W. Shen, M. Shafiullah, and C. C. Kuo, "Application of machine learning for fault classification and location in a radial distribution grid,"*Appl. Sci.*, vol. 10, no. 14, p. 4965, 2020.
[12] S. Liang *et al.*, "Hybrid artificial intelligence for power grid line fault diagnosis and restoration auxiliary decision-making," in *Proc. 5th Asia-Pacific Conf. Elect. Power Energy Eng. (ACPEE)*, 2023, pp. 1–6.
[13] C. Wang, J. Wang, and X. Zhang, "Automatic radar waveform recognition based on time-frequency analysis and convolutional neural network," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, New Orleans, LA, USA, 2017, pp. 2437–2441.
[14] R. Sampaio *et al.*, "Automatic restoration system for power distribution networks based on multi-agent systems,"*IET Gener. Transm. Distrib.*, vol. 11, no. 4, pp. 475–484, 2017.
[15] Z. Qu, W. Wang, C. Hou *et al.*, "Radar signal intra-pulse modulation recognition based on convolutional denoising autoencoder and deep convolutional neural network,"*IEEE Access*, vol. 7, pp. 112339–112347, 2019.
[16] Z. M. Liu and P. S. Yu, "Classification, denoising, and deinterleaving of pulse streams with recurrent neural networks,"*IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 4, pp. 1624–1639, 2019.
[17] G. Ghadimi, Y. Norouzi, R. Bayderkhani*et al.*, "Deep learning-based approach for low probability of intercept radar signal detection and classification,"*J. Commun. Technol. Electron.*, vol. 65, no. 10, pp. 1179–1191, 2020.
[18] B. Lay and A. Charlish, "Classifying LPI signals with transfer learning on CNN architectures," in *Proc. Sensor Signal Process. Defence Conf. (SSPD)*, Edinburgh, UK, 2020, pp. 1–5.

[19] Anonymous, "Multi-task learning for radar signal characterisation,"*arXiv preprint arXiv:2306.13105*, 2024.
[20] S. Shakiba, F. M. Azizi, M. Zhou, and A. Abusorrah, "Application of machine learning methods in fault detection and classification of power transmission lines: A survey,"*Artif. Intell. Rev.*, vol. 55, no. 3, pp. 2315–2375, 2022.
[21] X. Ni, H. Wang, F. Meng *et al.*, "LPI radar waveform recognition based on multi-resolution deep feature fusion,"*IEEE Access*, vol. 9, pp. 26138–26146, 2021.
[22] A. Yildirim and S. Kilanyaz, "1D convolutional neural networks versus automatic classifiers for known LPI radar signals under white Gaussian noise,"*IEEE Access*, vol. 8, pp. 180534–180543, 2020.
[23] T. Yin, S. Wulff, J. W. Pierre, and B. Amidan, "Event detection and classification using machine learning applied to PMU data for the Western US Power System," in *Proc. IEEE Smart Grid Synchronized Measurement Appl. (SGSMA)*, 2024.
[24] MathWorks, "Radar and communications waveform classification using deep learning," 2024. [Online]. Available: https://www.mathworks.com/help/radar/ug/radar-and-communications-waveform-classification-using-deep-learning.html
[25] A. Yildirim and S. Kilanyaz, "1D convolutional neural networks versus automatic classifiers for known LPI radar signals under white Gaussian noise,"*IEEE Access*, vol. 8, pp. 180534–180543, 2020.
[26] S. Mallat, *A Wavelet Tour of Signal Processing*, 3rd ed. San Diego, CA, USA: Academic Press, 2008.
[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770–778.
[28] L. Cohen, "Time-frequency distributions—a review,"*Proc. IEEE*, vol. 77, no. 7, pp. 941–981, 1989.
[29] F. Hlawatsch and G. F. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations,"*IEEE Signal Process. Mag.*, vol. 9, no. 2, pp. 21–67, 1992.
[30] S. Han, H. Mao, and W. J. Dally, "Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding," in *Proc. Int. Conf. Learn. Representations (ICLR)*, San Juan, Puerto Rico, 2016.
[31] "Supervised learning of wireless signal waveforms in the context of JCAS," in *Proc. 16th IFIP Wireless Mobile Netw. Conf. (WMNC)*, 2025, pp. 127–130.
[32] M. K. Onwughalu, E. Omorogiuwa, and M. Ehikhamenle, "Machine learning and M2M communication in smart grids: A review, taxonomy, and future directions for fault management,"*submitted for publication*, 2026.
[33] "Implementation of edge AI for early fault detection in IoT networks,"*Discover Appl. Sci.*, 2025. [Online]. Available: https://doi.org/10.1007/s43926-025-00196-4Login to Jenni
[34] "Federated learning-based fault location in hybrid AC/DC distribution systems,"*Energies*, vol. 17, no. 24, p. 6420, 2024.
[35] S. Vaish, U. Dwivedi, S. Tewari, and S. M. Tripathi, "Machine learning applications in power system fault diagnosis: Research advancements and perspectives,"*Eng. Appl. Artif. Intell.*, vol. 106, p. 104504, 2021.
[36] I. Srivastava, S. Bhat, B. Vardhan, and N. Bokde, "Fault detection, isolation and service restoration in modern power distribution systems: A review,"*Energies*, vol. 15, no. 19, p. 7264, 2022.
[37] Y. Liu, R. Fan, and V. Terzija, "Power system restoration: A literature review from 2006 to 2016,"*J. Mod. Power Syst. Clean Energy*, vol. 4, no. 3, pp. 332–341, 2016.
[38] S.-J. Chen, W.-Y. Chiu, and W.-J. Liu, "User preference-based demand response for smart home energy management using multiobjective reinforcement learning,"*IEEE Access*, vol. 9, pp. 161627–161637, 2021.
[39] A. A. Amer, K. Shaban, and A. M. Massoud, "DRL-HEMS: Deep reinforcement learning agent for demand response in home energy management systems,"*IEEE Trans. Smart Grid*, vol. 14, pp. 239–250, 2023.
[40] F. Alfaverh and M. Denai, "Demand response strategy based on reinforcement learning and fuzzy reasoning for home energy management,"*IEEE Access*, vol. 8, pp. 39310–39321, 2020.
[41] W. Liu, Y. Wang, F. Jiang *et al.*, "A real-time demand response strategy of home energy management by using distributed deep reinforcement learning," in *Proc. IEEE HPCC/DSS/SmartCity/DependSys*, 2021, pp. 988–995.
[42] "Deep reinforcement learning for optimal microgrid energy management with renewable energy and electric vehicle integration,"*Appl. Soft Comput.*, p. 113180, 2025.
[43] "Optimal energy management in smart energy systems: A deep reinforcement learning approach,"*Energy Reports*, 2024.
[44] "Digital twin-driven identification of fault situation in distribution networks connected to distributed wind power,"*Int. J. Elect. Power Energy Syst.*, 2024.
[45] "A review on digital twins for power generation and distribution,"*J. Cyber Secur. Mobility*, 2023.
[46] N. Tzanis *et al.*, "A hybrid cyber-physical digital twin approach for smart grid fault prediction," in *Proc. IEEE Conf. Ind. Cyber-Physical Syst. (ICPS)*, 2020, pp. 393–397.
[47] M. You *et al.*, "Digital twins-based day-ahead integrated energy system scheduling under load and renewable energy uncertainties,"*Appl. Energy*, vol. 305, p. 117899, 2022.
[48] Y. Zhang *et al.*, "Application of digital twins in smart grids," in *Proc. IEEE ICPICS*, 2022, p. 9873758.
[49] D. Mourtzis, J. Angelopoulos, and N. Panopoulos, "Development of a PSS for smart grid energy distribution optimization based on digital twin,"*Procedia CIRP*, vol. 107, pp. 1138–1143, 2022.
[50] J. Xing *et al.*, "Multi-energy simulation and optimal scheduling strategy based on digital twin," in *Proc. IEEE PSGEC*, 2022, p. 9881079.
[51] "DRL home energy management review: Applications of deep reinforcement learning for home energy management systems: A review,"*Energies*, vol. 17, no. 24, p. 6420, 2024.
[52] "DRL renewable energy scheduling for e-commerce systems," in *Proc. IEEE*, 2021.
[53] "DRL smart grid scheduling with machine learning," in *Proc. IEEE*, 2020.