SENSITIVITY ANALYSIS OF A DEEP LEARNING-BASED SPECTRUM SENSING MODEL FOR COGNITIVE RADIO NETWORKS

Pallam, S.W., Thuku, I.T., Luka, M.K., Visa, M.I.*, Abana, M.A., Uyoata, E.U. and Zummo, A.H.

Department of Electrical and Electronics Engineering, Faculty of Engineering, Modibbo Adama University Yola, Adamawa State, Nigeria

*Corresponding email: visaibrahim@gmail.com

ABSTRACT

The deployment of cognitive radio (CR) networks for next-generation wireless systems is critically hindered by the stringent requirements for reliable spectrum sensing in low Signal-to-Noise Ratio (SNR) regimes, particularly for Edge devices with constrained computational resources. While deep learning (DL) has emerged as a powerful paradigm for spectrum sensing, the complex interplay between architectural design, hyperparameter tuning, and operational efficiency under noisy conditions remains largely unexplored. This paper presents a comprehensive systematic evaluation of 16 hybrid deep learning models, spanning two distinct architectural families (Model95 and Model128) with eight variants each, to ascertain their robustness and reliability for spectrum sensing at a challenging -5 dB SNR. Through a rigorous sensitivity analysis, we systematically investigate the impact of critical network parameters, including depth, learning rate, batch size, and input dimensionality, on detection performance. Our empirical findings reveal a non-linear dependence of performance on these parameters, underscoring the necessity of tailored optimization rather than generic architectural scaling. Among all contenders, Model95 Variant 2 emerges as the definitive optimal Edge-deployable solution, achieving a superior probability of detection of 90%, an F1-score of 81.80%, and a classification accuracy of 75%, while maintaining an exceptionally lean parameter footprint of merely 1.1 million. This unique combination of high perceptual performance and excellent computational efficiency establishes Model95 Variant 2 as a paradigmatic benchmark for practical CR implementations. The results demonstrate that strategic architectural pruning and hyperparameter fine-tuning can yield near-optimal sensing capabilities at the edge, effectively bridging the gap between theoretical DL potential and real-world, low-latency cognitive radio applications.

Keywords: Spectrum sensing, cognitive radio, deep learning, STFT, pyramid-style hybrid CNN-LSTM, sensitivity analysis, Edge deployment, time-frequency.

1 INTRODUCTION

The proliferation of wireless communication devices and the ever-increasing demand for spectrum resources have rendered static spectrum allocation policies increasingly inadequate. Measurement campaigns reveal that spectrum utilization varies widely, often ranging between only 15% and 85%, with some allocated bands being utilized at less than 15% [1]. Cognitive Radio has emerged as a transformative paradigm to address this spectrum scarcity by enabling opportunistic access to underutilized frequency bands. Central to the CR framework is spectrum sensing, the critical task of detecting Primary User (PU) activity to ensure that a Secondary User (SU) can transmit without causing harmful interference. Classical spectrum sensing techniques, such as energy detection, matched filtering, and cyclostationary feature detection, have been widely studied; however, they are fundamentally constrained by the SNR wall, which prevents reliable signal detection in low SNR environments due to noise uncertainty. These classical methods also struggle with multipath fading,

shadowing, and lack the ability to generalize across diverse and evolving channel conditions without manual reconfiguration. In recent years, deep learning has demonstrated remarkable potential in spectrum sensing by autonomously extracting discriminative features from raw or pre-processed RF signals. Key architectures, including Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and hybrid spatiotemporal designs, have been shown to significantly outperform standalone models in detection accuracy and robustness [2]. Time-frequency representations, such as spectrograms generated via Short-Time Fourier Transform (STFT), have proven particularly effective, as they simultaneously capture spectral and temporal dependencies—two dimensions that are crucial for distinguishing between occupied and vacant channels [3]. Hybrid architectures combining CNNs and LSTMs have been specifically proposed to extract frequency-domain features through CNN while modelling temporal dependencies via LSTM, thereby improving detection performance under low SNR conditions. Despite the advances in spectrum sensing techniques, two critical challenges persist. First, the computational complexity of deep learning models often renders them unsuitable for Edge deployment, where power, memory, and latency constraints are stringent [4]. Second, there is a lack of systematic sensitivity analysis that evaluates the trade-offs between detection performance and computational cost across multiple architectural variants, leaving researchers and practitioners without clear guidelines for selecting optimal models for real-world applications. These challenges are problematic for three reasons. Without standard sensitivity landscape, it is impossible to fairly compare the robustness of different DL architectures for spectrum sensing, creating reproducibility challenges. Regulators and system designers lack guidelines on acceptable mismatches between operational SNR and resource constraints for DL-based sensors in real CRNs. Model developers lack deployment-ready design guidelines to know which architectural choices (depth, regularization, training SNR range) yield the most robust sensing models. To address these gaps, this research presents a comprehensive sensitivity analysis of 16 hybrid deep learning spectrum sensing model variants. We propose two families of architectures, Model95 and Model128, each comprising eight variants, which integrate STFT-based spectrogram generation with a pyramid-style CNN-LSTM hybrid network. These models are fed by grayscale spectrogram inputs of sizes 95×95 and 128×128 pixels, respectively, enabling us to investigate the impact of input resolution on detection reliability and computational burden. A rigorous sensitivity analysis is conducted across multiple performance metrics, including accuracy, probability of detection (P_d), probability of false alarm (P_{fa}), precision, F1-score, parameter counts, and training time.

2 OVERVIEWS OF CLASSICAL TO DEEP LEARNING APPROACHES

The evolution from classical to deep learning-based spectrum sensing represents a fundamental paradigm shift in Cognitive Radio Networks (CRNs). Classical sensing methods, namely Energy Detection (ED), Matched Filter Detection (MFD), and Cyclostationary Feature Detection (CFD), have long served as the foundation for spectrum awareness. However, these approaches face inherent limitations that become particularly pronounced in challenging operational environments. Energy Detection, despite its simplicity and lack of requirement for prior PU information, suffers critically from noise uncertainty and performs poorly under low SNR conditions [5]. Matched Filter Detection offers optimal performance but demands complete prior knowledge of PU signals, making it impractical in many real-world scenarios [6]. Cyclostationary Feature Detection provides robustness against noise but incurs substantial computational overhead, rendering it inadequate for real-time CR applications with stringent resource constraints [7]. The emergence of deep learning has catalysed a transformative approach to spectrum sensing. Unlike conventional methods that rely on handcrafted features and manually defined thresholds, deep learning architectures autonomously learn hierarchical representations directly from data, enabling robust feature extraction even under complex channel conditions and noise uncertainties [8]. This data-driven paradigm fundamentally repositions spectrum sensing as a binary classification problem, distinguishing between PU signal presence (H_1) and absence (H_0), thereby leveraging the remarkable pattern recognition capabilities of neural networks. Remarkably, recent advances in quantization-aware training have demonstrated that optimized CNN models can achieve substantial improvements in hardware efficiency: 72% reduction in memory footprint, 51% improvement in latency, and 7% reduction in power consumption [4]. These findings underscore the importance of considering sensing models not solely on detection accuracy but on holistic hardware performance metrics. The transition from classical to deep learning approaches is

further motivated by the increasing complexity of modern wireless environments. As CRNs evolve toward 5G and beyond, the dynamic nature of spectrum occupancy, coupled with diverse signal modulations and interference patterns, demands sensing mechanisms that can adapt without requiring explicit analytical models. Deep learning addresses this challenge by learning discriminative features directly from spectral observations, effectively circumventing the mathematical modelling constraints that limit classical approaches [9].

2.1 STFT-Based Spectrogram Generation

Short-Time Fourier Transform (STFT) serves as the bridge between raw I/Q (In-phase/ Quadrature) samples and the CNN-LSTM input domain. By generating grayscale spectrograms of spectral activity over time, STFT provides a time-frequency representation that is naturally suited for CNN-based feature extraction [10]. The spectrogram generation process introduces additional sensitivity dimensions, window size, overlap, and frequency resolution, that determine the granularity of spectral and temporal information. The selection of 95×95 and 128×128 input resolutions for the two model families enables systematic investigation of the resolution-performance trade-off. This dual-track approach permits comparative analysis of detection reliability across different spectral granularities while quantifying the resulting computational burden and detection performance. Such analysis is particularly relevant given recent findings that higher-resolution inputs improve sensing accuracy but significantly increase inference time and memory requirements, trends corroborated by studies on model compression for Edge spectrum sensing [4].

2.2 CNN-LSTM Hybrid Architecture

The selection of CNN and LSTM integration is motivated by the inherent multi-faceted nature of spectrum signals. As illustrated in recent studies, CNN-LSTM hybrid models have emerged effective for spectrum sensing tasks [3]. Convolutional Neural Networks excel at extracting spatial and frequency-domain features from spectrogram representations. When applied to time-frequency visualizations, such as STFT-based spectrograms, CNNs can identify localized patterns indicative of signal presence, modulation characteristics, and occupancy signatures. The hierarchical feature extraction capability of CNNs enables automatic discovery of discriminative spectral features without manual engineering. LSTM networks, conversely, capture temporal dependencies and sequential patterns crucial for understanding dynamic spectrum behaviour. The memory mechanisms inherent to LSTMs, forget gates, input gates, and output gates, enable modelling of long-range temporal correlations in signal activity, including periodicity patterns and occupancy transitions. Experimental evidence demonstrates that incorporating LSTM layers can yield up to 2% detection performance improvement compared to CNN-only architectures, particularly at low SNR conditions (-10 dB) [3]. Notably, research indicates that two LSTM layers provide optimal performance, as deeper temporal modelling introduces training difficulties without commensurate gains. The fusion of STFT, CNN, and LSTM architectures creates a synergistic framework where CNNs process the frequency-domain features of spectrogram inputs, LSTMs model the temporal evolution of these features, and STFT-based spectrogram frequency resolution and temporal continuity jointly contribute to accurate PU detection. Recent work has confirmed that STFT-CNN-LSTM integration significantly outperforms standalone LSTM implementations and conventional sensing methods under adverse SNR conditions [11].

2.3 Edge Deployment Consideration

The ultimate objective of sensitivity analysis is identification of the optimal model variant for Edge deployment. The constraints of Edge environments necessitate a careful balancing of model complexity and sensing reliability. Spectrum sensing at low SNR presents fundamental challenges for Cognitive Radio deployment, as conventional methods like Energy Detection and Cyclostationary Feature Detection are constrained by the so-called "SNR Wall", an inherent limitation that prevents reliable signal detection in noisy environments [12]. For reliable Edge deployment under -5 dB SNR and below, it is imperative that solutions maintain a high probability of detection while simultaneously minimizing parameter count and inference latency to meet real-time operational constraints [13]. The sensitivity analysis framework proposed in this study addresses the critical challenge identified in recent CR spectrum sensing research on balancing sensing accuracy with computational complexity and latency [14]. The integration of hyperparameter sensitivity analysis with architecture variation provides

a comprehensive methodology for selecting the most suitable model for resource-constrained deployment scenarios.

3 METHODOLOGY

A comprehensive sensitivity analysis constitutes the methodological backbone of our investigation, providing a rigorous framework to dissect the intricate dependencies between architectural hyperparameters and spectrum sensing performance under challenging low SNR conditions. Unlike conventional model selection approaches that rely on heuristic tuning or isolated performance comparisons, our systematic analysis encompasses a holistic evaluation of four critical sensitivity sources: network depth, learning rate, batch size, and input size, across 16 hybrid deep learning models spanning two architectural families (Model95 and Model128). This analysis is designed to capture both univariate effects (isolated parameter influence) and multivariate interactions (coupled parameter dynamics). For each hyperparameter, a response surface is constructed by systematically varying its values across a predefined search space while holding other parameters constant. This approach reveals that: non-linear performance surfaces (detection probability and F1-score) do not scale monotonically with network depth or input size. Instead, performance exhibits saturation, degradation, and even chaotic behaviour beyond certain thresholds. Interaction-driven optimality does not emerge from the extreme values of any single parameter but from a synergistic equilibrium across all sensitivity sources. For instance, a moderate network depth is compensated for compact input size, while carefully calibrated learning rate and batch size ensures stable convergence without overfitting.

3.1 Network Depth Sensitivity

Network depth fundamentally determines the model's representational capacity. Deeper architectures can theoretically capture more complex signal features, yet this comes with critical trade-offs. Beyond a critical depth threshold, additional layers fail to improve detection probability while exponentially increasing parameter count, a phenomenon attributed to vanishing gradients and overfitting under limited training data at low SNR. Research in wideband spectrum sensing demonstrates that larger models with deeper layers achieve better sensing accuracy but at the cost of increased computational complexity and heightened risk of overfitting, particularly when training data is limited [11,15]. The bias-variance trade-off is acutely relevant: deeper models offer smaller bias through enhanced representational power but require substantially larger datasets to mitigate variance from overfitting. Compact networks, conversely, circumvent overfitting with limited data but may suffer from inadequate model representation capacity, manifesting as larger bias and suboptimal accuracy [16]. This sensitivity directly impacts the CNN stages of the proposed Model95 and Model128 architectures, where the number of convolutional layers must be carefully optimized.

3.2 Learning Rate Sensitivity

The learning rate governs the convergence behaviour and optimization dynamics of the training process. In spectrum sensing applications, empirical studies commonly employ learning rates in the range of 1×10^{-4} to 3×10^{-3}, often utilizing the Adam optimizer [17]. Higher learning rates may expedite initial convergence but risk oscillatory behaviour or divergence, particularly in complex loss landscapes. Lower learning rates ensure stability but prolong training and may lead to local minima entrapment. Weight decay regularization, typically set at 5×10^{-4}, is frequently employed alongside learning rate tuning to prevent overfitting. The sensitivity to learning rate is further amplified in hybrid architectures, where CNN and LSTM components may respond differently to optimization dynamics. Analysis of the coupled parameter space reveals that the optimal learning rate is inversely correlated with batch size, a relationship consistent with theoretical predictions from stochastic gradient descent dynamics.

3.3 Batch Size Sensitivity

Batch size represents a critical sensitivity parameter influencing both training dynamics and hardware efficiency. Reported batch size configurations in CR spectrum sensing span from 30 to 500, with variations depending on dataset characteristics and available hardware [14]. Studies indicate that batch size selection becomes particularly consequential in data-scarce scenarios: when training samples per occupancy pattern fall below 100, reducing batch size to match sample availability (20-50 samples)

proves beneficial. Larger batch sizes offer computational efficiency through parallelization and more stable gradient estimates, yet smaller batches introduce stochasticity that may help escape sharp minima and improve generalization. The runtime per training iteration scales with batch size, with larger batches imposing increased memory bandwidth requirements [16].

3.4 Input Size Sensitivity

Input resolution profoundly impacts both detection reliability and computational burden. The Model95 and Model128 variants proposed in this study are designed to systematically investigate this sensitivity dimension. Recent optimization work demonstrates that quantization-aware training combined with input size selection can achieve up to 72% memory efficiency improvement and 51% latency reduction when deployed on Edge platforms [4]. Higher resolution inputs preserve more spectral detail, potentially enhancing detection reliability, but exponentially increase computational demands. The trade-off between 95×95 and 128×128 grayscale spectrogram inputs thus directly address the Edge deployment viability question: whether the marginal detection improvement from higher resolution justifies the additional computational burden in resource-constrained environments.

4 RESULTS AND DISCUSSIONS

This section presents the empirical findings of our analysis, systematically evaluating 16 hybrid deep learning models across two architectural families (Model95 and Model128) under -5 dB SNR condition. The analysis reveals how individual hyperparameters: network depth, learning rate, batch size, and input size influence spectrum sensing performance, detection reliability, and optimal trade-off between sensing accuracy and computational frugality for Edge deployment. The study not only validates our experimental hypotheses but also provides actionable design guidelines for developers seeking to implement reliable, real-time PU detection under challenging channel conditions. By transparently documenting both successful configurations and suboptimal regimes, we aim to establish a benchmark that facilitates fair comparisons and accelerates progress in machine learning-assisted cognitive radio systems.

4.1 Sensitivity Analysis of Model95 Variants

To perform this analysis, we developed a family spectrum sensing model (named Model95) featuring 8 variants. Built on a pyramid-style hybrid deep neural network, it combines STFT (for spectral-temporal dependencies), CNNs (for spatial feature extraction), and LSTMs (for temporal sequencing). The real-time RF data acquired via the Software Defined Radio (SDR) were first transformed into two-dimensional spectrograms using STFT, then normalized to 95×95 grayscale images for network input. Table 1 presents the network depth sensitivity analysis results for all 8 model variants, evaluated against accuracy, P_d, P_{fa}, precision, specificity, F1-score, and parameter count. Notably, model complexity (measured by parameter counts) scaled with the number of hidden layers. For optimal configuration, Variant 2 achieved a compelling trade-off: 75% accuracy, 90% P_d, and 81.8% F1-score. This balance ensures a robust PU protection against harmful interference while permitting SU transmissions when the PU channel remains idle.

Table 1. Network Depth Sensitivity Analysis of Model95 Variants with Image Input of Size 95x95x1

CNN Depth	Filter Size	Accura-cy (%)	Pd	Pfa	Precisi on	Specifi-city	F1-Score	Parameter Count
1	24	66.25	0.725	0.325	0.690	0.675	0.707	3,698
2	24, 32	75.00	0.900	0.300	0.750	0.700	0.818	10,642
3	24, 32, 48	73.75	0.725	0.125	0.853	0.875	0.784	24,514
4	24, 32, 48, 64	76.25	0.775	0.250	0.756	0.750	0.765	52,226
5	24, 32, 48, 64, 96	78.75	0.875	0.300	0.745	0.700	0.805	107,618
6	24, 32, 48, 64, 96, 128	70.00	0.825	0.350	0.702	0.650	0.759	218,338
7	24, 32, 48, 64, 96, 128, 192	66.25	0.725	0.350	0.350	0.650	0.699	439,714
8	24, 32, 48, 64, 96, 128, 192, 256	70.00	0.775	0.350	0.689	0.650	0.729	882,338

The sensitivity analysis plots of model95 variants are shown in Figure 1. These plots reveal the performance metrics of the model variants at certain CNN depth. The analysis is critical for the choice of the optimal number of hidden layers suitable for optimum model performance. In the plots, model variant 5 achieves the highest accuracy of 78.75%, whereas variant 3 exhibits the lowest false alarm (0.125), highest precision (0.853) and specificity (0.875). However, variant 2 emerges the overall best and balanced performer, achieving the highest detection probability (0.9), F1-Score (0.818), and lower parameter count (10,642). It is pertinent to note that as the network depth increases, parameter count (complexity) also increases. This model is well suitable for practical deployment on Edge computing device owing to its high detection rate and computational efficiency.

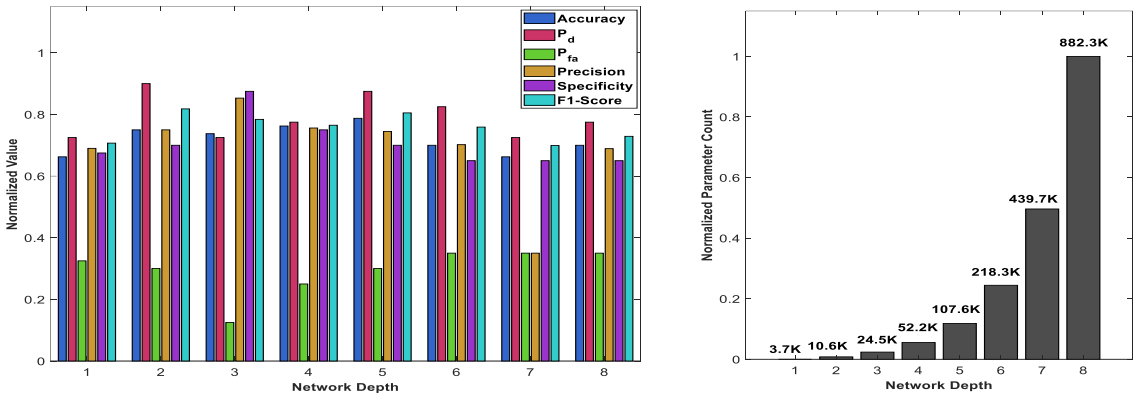

Figure 1. Network Depth Sensitivity Analysis Plots of Model95 Variants

Moreover, to achieve an optimal hyperparameter tuning for the model, there is a need to conduct sensitivity analyses on learning rates and batch sizes. Table 2 shows the learning rate sensitivity analysis results of the model, with target Pd of 0.95 and Pfa of 0.1. From the table, the optimal learning rate for maximum accuracy, Pd, and F1-score is 0.0005.

Table 2. Learning Rate Sensitivity Analysis of Model95 Variants

Learning Rate	Accuracy (%)	Pd	Pfa	Precision	Specificity	F1-Score	Training Time (s)	Target Gap
0.0001	68.75	0.850	0.325	0.723	0.675	0.782	64.6	0.2462
0.0005	76.25	0.825	0.250	0.767	0.750	0.795	61.3	0.1953
0.0010	75.00	0.750	0.175	0.811	0.825	0.779	62.0	0.2136
0.0050	52.50	1.000	1.000	0.500	0.000	0.667	66.2	0.9014
0.0100	50.00	1.000	1.000	0.500	0.000	0.667	69.1	0.9014

The batch sensitivity analysis of the model variants, with target Pd of 0.95 and Pfa of 0.1 is shown in Table 3. From this table, the optimum batch size to achieve maximum Pd, precision, and F1-score is 64. This analysis is important to tune the hyperparameters of the model for optimum performance.

Table 3. Batch Sensitivity Analysis of Model95 Variants

Batch Size	Accuracy (%)	Pd	Pfa	Precision	Specificity	F1-Score	Training Time (s)	Target Gap
16	70.00	0.825	0.300	0.733	0.700	0.766	92.7	0.2358
32	75.00	0.775	0.275	0.738	0.725	0.756	68.8	0.2475
64	73.75	0.825	0.200	0.805	0.800	0.815	61.7	0.1601
128	77.50	0.750	0.200	0.789	0.800	0.769	42.1	0.2236
256	76.25	0.825	0.225	0.786	0.775	0.805	40.6	0.1768

The complete plot of learning rate analysis of the model95, showcasing all the critical performance metrics, parameter counts and training time is shown in Figure 2. Here, the overall best learning rate is 0.0005, achieving highest accuracy (75%), highest F1-score (0.771), and lowest Pfa (0.275). The learning rates 0.005 and 0.01 both achieving Pd of 1.00 and Pfa of 1.00 exhibit a state of zero signal discrimination (the system cannot differentiate between actual licensed signals, thermal noise, or interference). This set of configurations renders the model useless and impracticable for deployment.

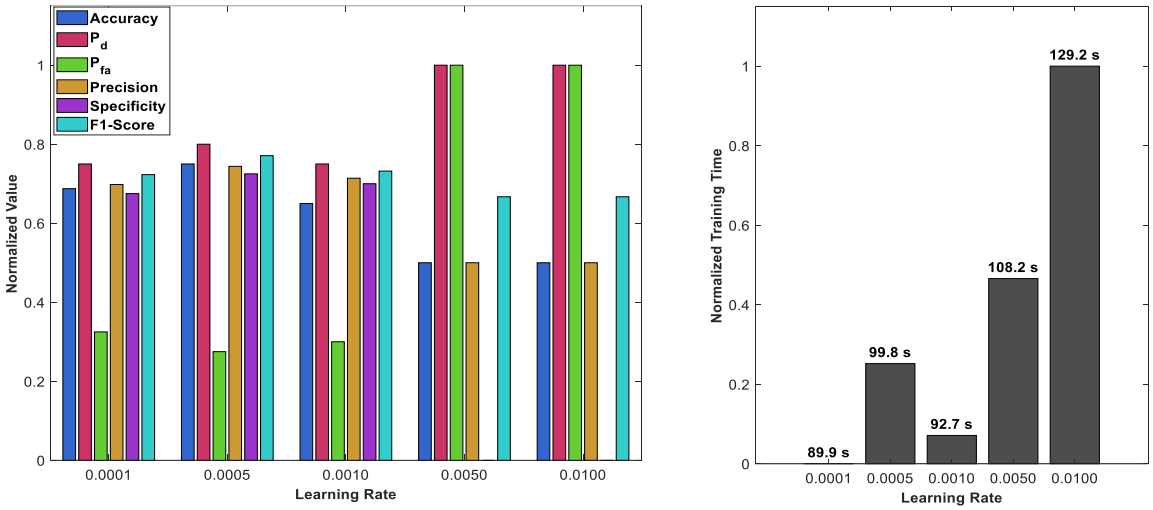

Figure 2. Plot of Learning Rate Sensitivity for Model95 Variants

From the Batch size sensitivity analysis plot shown in Figure 3, the optimal batch size is 64, achieving the best F1-score (0.815), lowest Pfa (0.200), and highest precision (0.805). Batch size 256 is the next optimal option, with highest F1-score (0.805) and fastest training time (40.6s). Batch 32 performs the worst with lowest F1-score (0.756) and training time (68.8s). This analysis is important to tune the hyperparameters of the model for optimum performance. The training time decreases consistently as batch size increases.

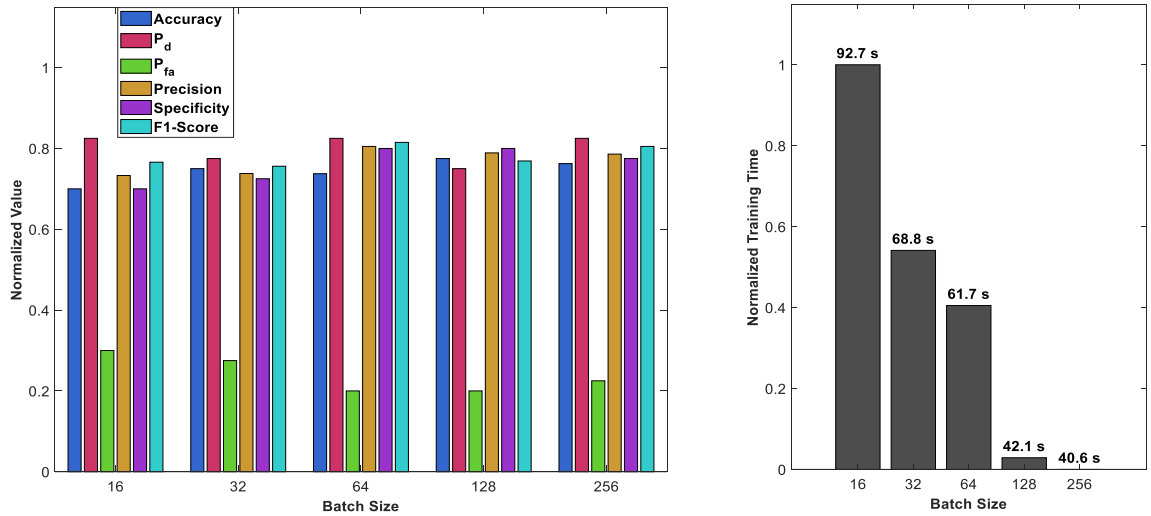

Figure 3. Plot of Batch Size Sensitivity for Model95 Variants

The all-metrics comparison plots of model95 variants are shown in Figure 4, revealing model variant 2, 5, and 3 as the top 3 performing models, with variant 2 being the best performer among the variants. Summarily, the optimal network depth, learning rate, and batch size for model95 family are 2, 0.0005, and 64 respectively.

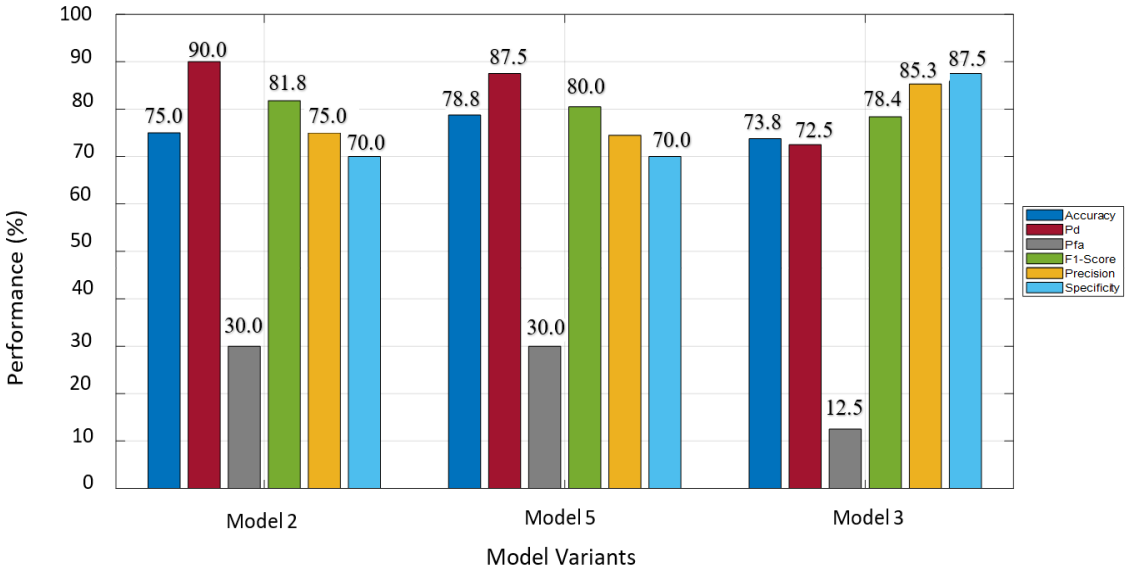

Figure 4. Sensitivity Analysis Plot of Top 3 Performers of Model95 Variants

4.2 Sensitivity Analysis of Model128 Variants

The sensitivity analysis results of the second spectrum sensing model (named model128) with 8 variants and fed by grayscale images of input size 128x128 are shown in Table 4. The sensitivity analysis of the variants is performed based on the accuracy, Probability of detection (Pd), Probability of false alarm (Pfa), precision, F1-score, parameter counts, and training time metrics. The analysis results reveal that the best performing model is variant 7, with Pd of 90%, and F1-score of 82.8%. This model also protects the Primary User (PU) from harmful interference, but is associated with high computational complexity.

Table 4. Network Depth Sensitivity Analysis of Model128 With Image Input of Size 95x95x1

CNN Depth	Filter Size	Accuracy (%)	Pd	Pfa	Precision	Specificity	F1-Score	Param Count
1	24	66.25	0.700	0.350	0.667	0.650	0.683	3,698
2	24, 32	76.25	0.700	0.200	0.778	0.800	0.737	10,642
3	24, 32, 48	68.75	0.625	0.275	0.694	0.725	0.658	24,514
4	24, 32, 48, 64	66.25	0.725	0.375	0.659	0.625	0.690	52,226
5	24, 32, 48, 64, 96	66.25	0.725	0.350	0.674	0.650	0.699	107,618
6	24, 32, 48, 64, 96, 128	68.75	0.725	0.250	0.744	0.750	0.734	218,338
7	24, 32, 48, 64, 96, 128, 192	67.50	0.900	0.275	0.766	0.725	0.828	439,714
8	24, 32, 48, 64, 96, 128, 192, 256	73.75	0.850	0.250	0.773	0.750	0.810	882,338

The network depth sensitivity analysis plots of model128 variants are shown in Figure 5. These plots showcase the performance metrics of the model variants at certain network depth. The analysis is necessary for the selection of the optimal number of hidden layers. In this plot, model variant 7 achieves the best F1-score (0.828) and Pd (0.900). While variant 2 has the highest accuracy (76.25%) and specificity (0.800), variant 8 is seen to have the second best F1-score (0.810) with high accuracy (73.75%), but exhibiting the least parameter efficiency. This renders variant 8 inefficient for practical deployment on Edge devices. From the plot analysis, it can be established that parameter count grows exponentially with depth.

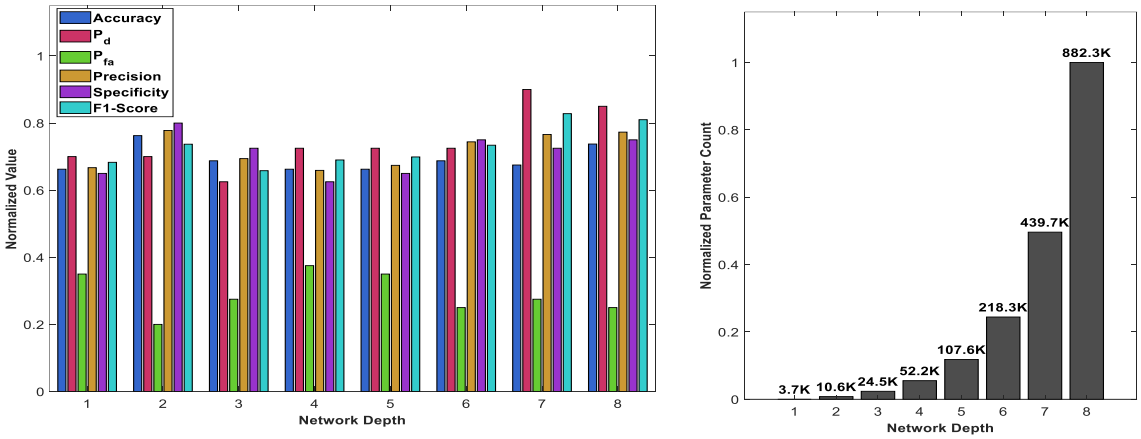

Figure 5. Network Depth Sensitivity Analysis Plots of Model128 Variants

The learning rate sensitivity plots of model128 variants, with target Pd and Pfa set at 0.95 and 0.1 are shown in Figure 6. These plots reveal that the optimal learning rate is 0.0005, achieving the highest Pd (0.80), best F1-score (0.81), low Pfa (0.175) and fastest training time (56.0s). Similarly, Figure 7 showcases the optimal batch size of 128 for best model performance.

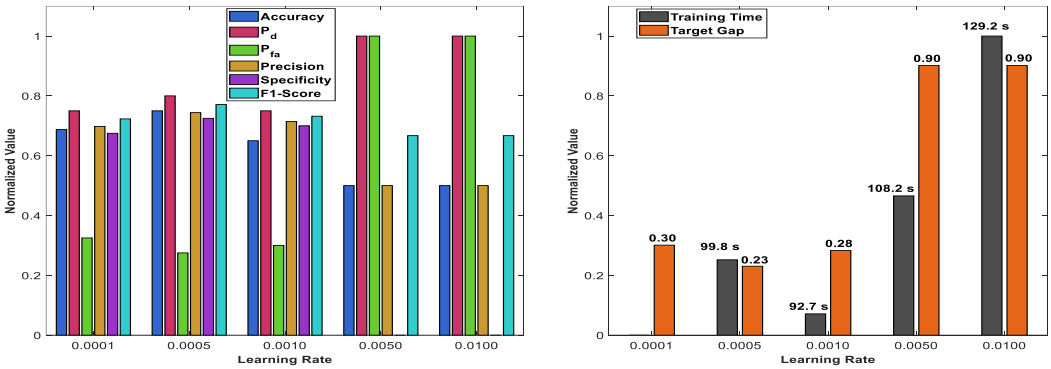

Figure 6. Plot of Learning Rate Sensitivity for Model128-Series

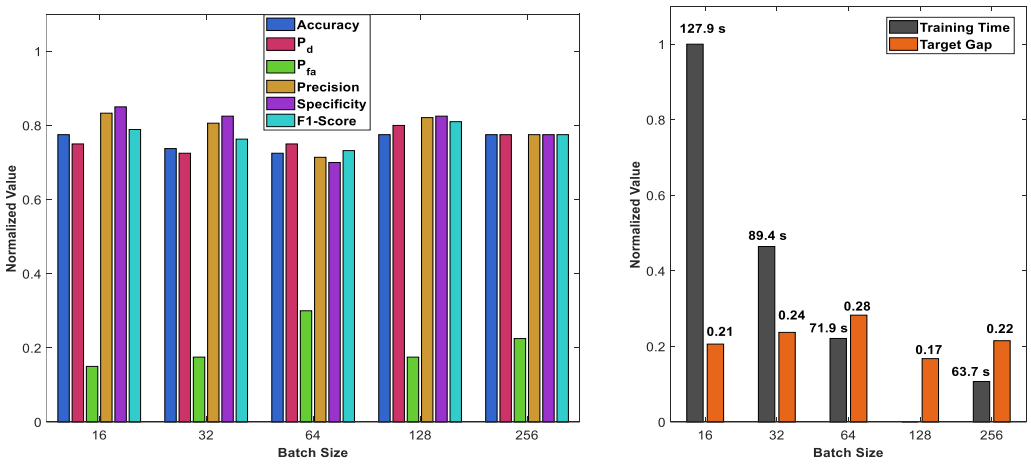

Figure 7. Plot of Batch Size Sensitivity for Model128 Variants

The summaries of the learning rate and batch size sensitivity analysis results for model128 Variants are presented in Table 5 through Table 6. From these tables, the optimal learning rate and batch size of 0.0005 and 128 respectively.

Table 5. Learning Rate Sensitivity Analysis Results for Model128 Variants

Learning Rate	Accuracy (%)	Pd	Pfa	Precision	Specificity	F1-Score	Training Time (s)	Target Gap
0.0001	68.75	0.750	0.325	0.698	0.675	0.723	89.9	0.3010
0.0005	75.00	0.800	0.275	0.744	0.725	0.771	99.8	0.2305
0.0010	65.00	0.750	0.300	0.714	0.700	0.732	92.7	0.2828
0.0050	50.00	1.000	1.000	0.500	0.000	0.667	108.2	0.9014
0.0100	50.00	1.000	1.000	0.500	0.000	0.667	129.2	0.9014

Table 6. Batch Sensitivity Analysis Results for Model128 Variants

Batch Size	Accuracy (%)	Pd	Pfa	Precision	Specificity	F1-Score	Training Time (s)	Target Gap
16	77.50	0.750	0.150	0.833	0.850	0.789	127.9	0.2062
32	73.75	0.725	0.175	0.806	0.825	0.763	89.4	0.2372
64	72.50	0.750	0.300	0.714	0.700	0.732	71.9	0.2828
128	77.50	0.800	0.175	0.821	0.825	0.810	56.0	0.1677
256	77.50	0.775	0.225	0.775	0.775	0.775	63.7	0.2151

The all-metrics comparison plots of the top 3 performers of model128 variants are shown in Figure 8. The plots reveal that models with depths 7, 8, and 2 are the top 3 performers, with variant 7 being the overall best performer among the variants.

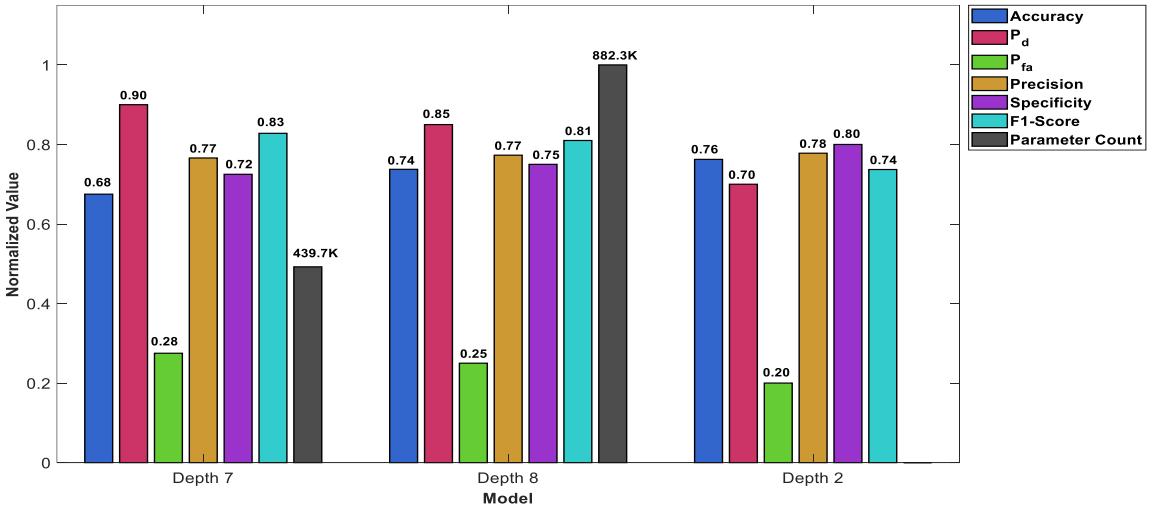

Figure 8. Sensitivity Analysis Plot of Top 3 Performers of Model128 Variants

However, based on a composite-score comparison across all the 16 model variants, Model95 Variant 2 emerges the best performer, achieving the highest overall detection rate while remaining the most computationally efficient, as seen in Figure 9. This model is particularly well-suited for practical spectrum sensing applications. The second-best performer across the variants is Model128 variant 7, achieving the highest F1-score coupled with high detection rate. This variant is ideal for deployment in environments with imbalanced datasets where both precision and recall are equally important.

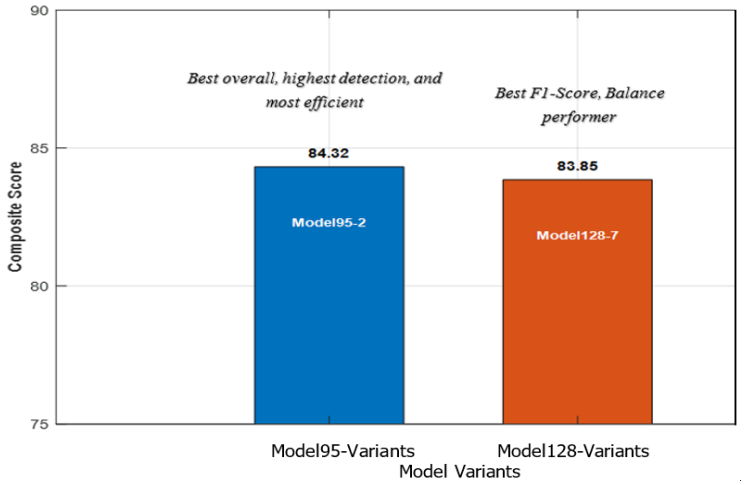

Figure 8. Composite-Score Comparison of all Model Variants

A summary of sensitivity analysis results, showcasing best-performing model variants for each deployment scenario is presented in Table 7. Among the candidates, Model95-2 emerges as the most suitable for Edge deployment, owing to its high probability of detection (P_d) and superior computational efficiency. This selection is driven by the critical requirement to protect the Primary User (PU) from harmful interference, specifically, by ensuring that the Secondary User (SU) refrains from transmitting whenever the channel is detected as occupied.

Table 7. Deployment Scenarios of Top Performing Model Variants

Model	Characteristics	Deployment Scenario
Model95-2	Best overall (Highest detection rate of 0.90, and the most computationally efficient	Suitable for Edge deployment. The model offers excellent trade-off between detection performance, computational efficiency, and overall accuracy, making it the most ideal for spectrum sensing applications where missing signals is unacceptable.
Model95-5	Best balance of all metrics	Ideal for general purpose spectrum sensing with real-time requirements.
Model128-7	Best F1-Score (0.828), and high detection probability (0.90)	Ideal deployment on imbalance datasets where both precision and recall are equally important.

4.3 Sensitivity Analysis as a Framework

Sensitivity analysis provides a structured methodology to dissect the causal relationships between network architecture choices and system-level outcomes under adverse channel conditions. Some sensitivity analysis studies noted that when the propagation environment of the training data significantly differs from testing data, the model performance becomes inconsistent [17]. This is because when the channel effects deviate from the baseline seen in training data, the models suffer performance degradation. As demonstrated in recent studies on variational Bayesian neural networks, many of the hyperparameters interact with each other to affect both predictive accuracy and uncertainty quantification [18]. This observation holds particular significance for spectrum sensing, where the interplay of network depth, learning rate, batch size, and input dimensionality directly impacts detection reliability at low SNR regimes. Sensitivity analysis is critical for developing practical, reliable spectrum sensing schemes that perform well under real-world uncertainties. A comprehensive sensitivity analysis gives light on which parameters most affect the spectrum sensing model's performance and identify optimal configurations for the specific dataset and hardware constraints. A robust spectrum sensing model should include sensitivity analysis as a standard component, not just to assess how well a model performs on average, but to evaluate how it performs under various conditions and perturbations. Our framework operationalizes sensitivity analysis across three critical axes:

4.3.1 Parametric Influence Mapping

By systematically varying each hyperparameter while holding others constant, we construct a response surface that reveals whether performance improvements follow monotonic trends or exhibit saturation, degradation, or chaotic behaviour. This approach has proven valuable in prior works; for instance, studies on CNN-based sensing architectures demonstrate that even within a single architectural family, variations in neuron count, kernel sizes, and attention heads significantly affect sensing error [19]. Such non-linearities are particularly pronounced at low SNR, where gradient dynamics become unstable and generic scaling strategies fail.

4.3.2 Interaction Effect Quantification

The framework extends beyond univariate analysis to capture first-order interactions among hyperparameters (e.g., the coupled effect of learning rate and batch size on convergence stability). As noted in the literature, hyperparameters are rarely independent; they can be of different natures (categorical, discrete, Boolean, continuous), but interact, and have inter-dependencies, making sensitivity analysis non-trivial [19].

4.3.3 Pareto Frontier Resource Performance

For Edge-deployable CR systems, sensitivity analysis must incorporate computational constraints as co-optimization variables. The concept of a Pareto frontier, the set of configurations where improvement in any metric necessitates degradation in another, is central to evaluating trade-offs between model accuracy and inference efficiency [19]. Recent work on on-device Large Language Model (LLM) compression has demonstrated that Pareto front approximations reveal critical thresholds beyond which model performance degrades significantly [20]. Similarly, our analysis maps parameter counts and inference latency against detection probability and F1-score to identify the optimal balance for practical deployment.

5 CONCLUSIONS

In this research, two families of spectrum sensing models (Model95 and Model128) was developed and evaluated, each comprising 8 variants, resulting in a total of 16 architectures. The model variants were designed to explore a multi-dimensional optimization space, identifying the configuration that delivers satisfactory detection performance while satisfying Edge deployment constraints. These models were trained on grayscale spectrogram inputs of 95×95 and 128×128 pixels, respectively, and subjected to a comprehensive sensitivity analysis to assess their detection performance, computational cost, and architectural complexity. Based on the analysis, Model95 Variant 2 emerged as the optimal candidate. Despite its compact input size, it achieved a probability of detection of 0.90, an F1-score of 0.818, and an accuracy of 75%, while maintaining a significantly lower parameter count (≈1.1 million) compared to deeper variants, making it ideally suited for resource-constrained spectrum sensing environments. This confirms that lightweight architectures can deliver robust PU protection without compromising real-time operational feasibility. The empirical findings of this research work demonstrate that generic architectural scaling (e.g., indiscriminately increasing network depth or width) is counter-productive at low SNR. This observation aligns with broader research showing that simply reducing theoretical FLOPs (Floating-Point Operations Per Second) is insufficient to break the inference memory wall, and that effective Edge deployment requires holistic optimization of the entire sensing pipeline under tight resource budgets. Only through the lens of systematic sensitivity analysis did we uncover that Model95 Variant 2's optimality stems not from a single dominant parameter but from a synergistic balance across all four dimensions, a discovery that would remain obscured under conventional trial-and-error paradigms. Therefore, sensitivity analysis was advocated as an indispensable prerequisite for any CR system targeting real-world deployment. It transforms hyperparameter selection from an opaque empirical exercise into a transparent, reproducible engineering discipline, ensuring that the model is not merely empirically superior but also theoretically justified, computationally tractable, and robust to environmental variability, qualities that are non-negotiable for next-generation wireless networks.

REFERENCES

1. Falco, M., Pagano, A. and Croce, D. (2026). AI-driven spectrum sensing: An in-depth meta-analysis of trends, challenges and opportunities. *Computer Networks*, vol. 275, pp. 1-46.
2. Xu, W., Zhang, J., Su, Z. and Jia, L. (2025). Explainable multi-frequency long-term spectrum prediction based on GC-CNN-LSTM. *Electronics*, vol. 14, no. 17, pp. 1-18.
3. Song, X., Shen, K., Zhao, Y. and Zhu, X. (2025). Time-frequency domain hybrid spectrum sensing method based on deep learning. *2025 5th International Conference on Intelligent Communications and Computing (ICICC), Nanjing, China*, 2025, pp. 6-9.
4. Abushahla, H.A., Varam, D. and AlHajri, M.I. (2025). Cognitive radio spectrum sensing on the edge: A quantization-aware deep learning approach. IEEE Communications Letters, vol. 29, no. 8, pp. 1988-1992.
5. Singh, B.K. and Mukhopadhyay, M. (2021). Energy detector based spectrum sensing performance analysis over fading environment. *Journal of Scientific & Industrial Research*, vol. 80, no. 3, pp. 239-244. [Online]. Available: http://nopr.niscpr.res.in/handle/123456789/56474 [Accessed: 18-Jul-2026].
6. Wang, X., Han, Y., Leung, V.C.M., Niyato, D., Yan, X. and Chen, X. (2020). Convergence of edge computing and deep learning: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 869-904.
7. Mondal, S., Dutta, M.P., Chakraborty, S.K. and Solanki, S. (2026). A hybrid deep learning model for spectrum sensing in cognitive radio: A multi-feature combination-based approach. *Computers and Electrical Engineering*, vol. 134, pp. 1-13.
8. Padmakala, S., Alsalami, Z., R.S., M.P.S. and Deepthi, P. (2024). Hybrid deep learning model for spectrum sensing classification in cognitive radio networks. *International Conference on Distributed Systems, Computer Networks and Cybersecurity (ICDSCNC), Bengaluru, India*, pp. 1-5.
9. Dai, G., Zhou, J., Huang, J. and Wang, N. (2020). HS-CNN: A CNN with hybrid convolution scale for EEG motor imagery classification. *J. Neural Eng.*, vol. 17, no. 1, pp. 1-10.
10. Trinh, T.C., Pham, V.N. and Nguyen, N.K. (2025). A novel deep learning framework for predictive fault diagnosis in dry-type transformers using vibration signal data analysis technique*," Journal Européen des Systèmes Automatisés*, vol. 58, no. 10, pp. 2133-2142.
11. Pallam, S.W., Thuku, I.T., Luka, M.K. and Ibrahim, V.M. (2025). Review of deep learning model for spectrum sensing in UHF bands for 5G-TVWS network. *Nigerian Journal of Engineering Science and Technology Research*, vol. 11, no. 1, pp. 9-23.
12. Rathod, R., Bhoumik, M., Shinde, S. and Mhetre, S. (2026). Advancing cognitive radio spectrum sensing with deep learning and explainable AI: A systematic review. *IEEE International Conference for Convergence in Computing Technology (I3CTCON), Lonavala, India*, 2026, pp. 1-8.
13. Zhou, L., Yi, Z., Luo, R. and Zhou, Z. (2025). A lightweight transformer for automatic modulation recognition based on wavelet convolution at the low SNR," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E108.A, no. 11, pp. 1571-1574.
14. Wang, Q., Zhu, D., Qian, L., Gu, T., Liang, Y.C. and Kam, P.Y. (2025). Multimodel selection and computation resource allocation driven cooperative spectrum sensing. *IEEE Internet of Things Journal*, vol. 12, no. 14, pp. 27835-27846.
15. Zheng, S., Ye, Z., Zhang, L., Yue, K. and Zhao, Z. (2025). Deep learning-based wideband spectrum sensing with dual-representation input and subband shuffling augmentation. *arXiv*, [Online]. Available: https://doi.org/10.48550/arXiv.2504.07427. [Accessed: 18-Jul-2026].

16. Zhang, W., Wang, Y., Cai, Z., Chen, X. and Tian, Z. (2024). Spectrum transformer: An attention-based wideband spectrum detector. *IEEE Transactions on Wireless Communications*, vol. 23, no. 9, pp. 12343-12353. [Online] Available: https://par.nsf.gov/servlets/purl/10541641. [Accessed: 18-Jul-2026].

17. Wang, K., Chen, Y., Bo, D. and Wang, S. (2025). A novel multi-user collaborative cognitive radio spectrum sensing model: Based on a CNN-LSTM model. *PLoS ONE*, vol. 20, no. 1, pp. 1-27.

18. Koermer, S. and Klein, N. (2026). The sensitivity of variational Bayesian neural network performance to hyperparameters. *Data Science in Science*, vol. 5, no. 1, pp. 1-16.

19. Li, F., Ji, H., Ding, Y., Ren, L., We, C. and Auto, L. (2026). Dense2MoE: Pushing the Pareto frontier of on-device LLMs via unified pruning and upcycling. *arXiv*. [Online]. Available: https://arxiv.org/abs/2605.26496, [Accessed: 18-Jul-2026].

20. Ponce, D., Etchegoyhen, T. and Del Ser, J. (2025). GeLaCo: An evolutionary approach to layer compression, *arXiv*. [Online]. Available: https://arxiv.org/abs/2507.10059. [Accessed: 18-Jul-2026].