algorithms

MDPI

Article

Efficient Recurrent Multi-Layer Neural Network for Multi-Scale Noise and Activity Drift Mitigation in Wideband Cognitive Radio Networks

Sunil Jatti [†] **and Anshul Tyagi** [*,†]

Department of Electronics and Communication Engineering, Indian Institute of Technology Roorkee, Roorkee 247667, India; jmadhukarrao@ec.iitr.ac.in

* Correspondence: anshul.tyagi@ece.iitr.ac.in

† These authors contributed equally to this work.

Abstract

Wideband spectrum sensing in Cognitive Radio Networks (CRNs) is challenging due to sparse primary user (PU) activity and noise clustering, which obscure signals and generate false alarms. Hence, a novel "Graph Discrete Wavelet Bayesian Kernel Boosted Decision Self-Attention Clustering Neural Network (GDWB-KBSC-NN)" is proposed. When sparse PU activity is masked by irregular interference bursts, traditional sensing algorithms misclassify weak transmissions as noise, leading to low detection reliability. To resolve this, the first hidden layer employs Discrete Wavelet Sparse Bayesian Kernel Analysis (DW-SBK), integrating Discrete Wavelet Packet Transform (DWPT), Sparse Bayesian Learning (SBL), and Kernel PCA. This restores the true sparse pattern of the spectrum, separates interference from actual PU signals, and enhances detection of weak channels. Additionally, PU signals are fragmented due to cross-scale activity drift, where dynamic bandwidth switching and variable burst durations disrupt temporal continuity. Therefore, the second layer incorporates Gradient Boosted Multi-Head Fuzzy Clustering (GB-MHFC), where Gradient Boosted Decision Trees (GBDT) model nonlinear spectral–temporal patterns, Multi-Head Self-Attention (MHSA) captures long- and short-range temporal dependencies, and Fuzzy C-Means Clustering (FCM) groups feature representations into stable PU activity modes, thereby reducing misclassifications and enhancing robustness under highly dynamic CRN conditions. The proposed method demonstrates superior performance with a maximum detection probability of 0.98, classification accuracy of 98%, lowest sensing error of 5.412%, and the fastest sensing time of 3.65 s.

Keywords: cognitive radio network; spectral sensing; deep learning; primary user activity detection; reinforcement learning for spectrum sensing; low SNR

Check for updates

Academic Editors: Chang Wu Yu and Rei Heng Cheng

Received: 29 November 2025
Revised: 11 February 2026
Accepted: 24 February 2026
Published: 25 February 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Cognitive radio networks (CRNs) are a revolutionary paradigm in wireless communications, which achieve efficient use of the spectrum through dynamic access to the spectrum. In contrast to traditional fixed allocation, CRNs allow unlicensed secondary users to opportunistically access an underutilized spectrum without disturbing licensed primary users. This flexibility is thus highly essential in wideband scenarios where vast spectral bands are tracked for possible occupation. CRNs offer advantages such as high spectral efficiency, cheap communication costs, and service flexibility to various other services such as IoT connectivity, mobile broadband, and mission-critical networks. However, in providing

reliable wideband spectrum sensing, CRNs are tied to intrinsic limitations. Noise distortions, uncertainty in signal activity, and dynamic primary user patterns drastically affect the efficiency of spectrum monitoring. Therefore, designing efficient detection mechanisms that do not compromise false alarm rates and missed detections is still a problem, which requires more advanced techniques than traditional sensing methods [1–4].

As much as CRNs have the potential to provide efficient usage, their implementation in practice is limited by current sensing mechanisms and constraints they introduce. One of the simplest methods, energy detection, is plagued with high noise uncertainty sensitivity, leading to high rates of false alarms when used for wideband scenarios. Matched filtering provides greater accuracy but involves advanced knowledge of primary user signals and is therefore infeasible in heterogeneous scenarios. Cyclostationary feature detection probes periodic statistical characteristics for strong identification, but at a high computational cost, and is not feasible in real-time wideband sensing. Compressive sensing has been proposed to take advantage of sparsity in the spectrum, but its performance degrades with non-uniform noise distributions. Machine learning algorithms, like support vector machines and random forests, are promising in spectrum state classification but usually do not perform well under scalability and adaptability in extremely dynamic spectral environments. These shortcomings emphasize the urgency for new architectures that can classify genuine spectrum activity from interference robustly and remain computationally feasible [5–7].

Within the field of spectrum sensing, a number of techniques have been used to overcome the impacts of noise clustering, each of them with significant limitations. Wavelet-based denoising methods try to separate spectral anomalies at different scales but tend to obscure details of weak primary signals. Statistical hypothesis testing paradigms like generalized likelihood ratio tests can detect noise distortions but are likely to mistake clustered interference as valid activity. Adaptive thresholding techniques enhance background noise robustness but are ineffective at dealing with bursty and irregular interference patterns. Bayesian inference models enable probabilistic decision-making but become computationally heavy in application to wideband environments. Neural network-based techniques exhibit learning under controlled noise, however, and often overfit to particular interference distributions, curbing generalization. In general, the lack of a consistent framework that invariably addresses irregular multi-scale noise clustering limits the accuracy and reliability of current solutions in wideband CRNs [8,9].

Likewise, mitigation of the effect of primary user transmissions scattered over multiple channels has been tried using different methods, but with inherent limitations. Hidden Markov models track temporal dependencies of activity patterns, but because of their dependency on pre-defined state transitions, they are sensitive to sudden bandwidth switching. Autoregressive models and Kalman filters can forecast activity drift across time but not bursts with fragmented or irregular occurrences. Short-time Fourier transforms, a type of time–frequency analysis technique, can follow spectral changes but cannot maintain accuracy on various timescales. Deep recurrent networks possess robust modeling power for sequential activity but tend to ignore inter-channel correlations. Clustering-based approaches try to categorize occupancy patterns into groups, but they are not flexible enough to adapt to dynamic fluctuations in burst duration and transmission bandwidth. These shortcomings, in aggregate, indicate the inadequacy of existing methods in handling cross-scale activity drift in wideband CRNs, emphasizing the need for more sophisticated mechanisms that more effectively maintain temporal and spectral continuity [10–12]. Although there have been developments in spectrum sensing within cognitive radio networks, existing approaches have struggled to identify primary users under conditions that are wideband and dynamic. The methods deteriorate in rare interference patterns that under-

mine spectral activity or sparsity construction, including transmission being temporally scattered along various channels. Uneven energy clustering, burstiness, and irregular bandwidth switching serve to weaken the reliability of detection and complicate signal analysis in distinguishing legitimate signals from artifacts produced by noise. Consequently, such drawbacks reduce efficient spectrum utilization and prevent the efficiency of dynamic spectrum access. Therefore, advanced frameworks are needed to ensure enhanced detection in maintaining reliable decision-making processes when dealing with multi-scale interference and temporally drifting activity problems in spectrum sensing.

1.1. Contributions of the Research

- To reduce the false alarms and mitigate the noise-induced masking in wideband CRN, the novel Discrete Wavelet Sparse Bayesian Kernel Analysis (DW-SBK) is proposed, which integrates the Discrete Wavelet Packet Transform (DWPT), Sparse Bayesian Learning (SBL), and Kernel Principal Component Analysis (Kernel PCA) within the first hidden layer of the RNN. It suppresses the multi-scale noise clustering, recovers the true sparse occupancy, and enhances weak PU signal detection.
- The proposed Gradient Boosted Multi-Head Fuzzy Clustering (GB-MHFC) incorporates Gradient Boosted Decision Trees (GBDT), Fuzzy C-Means Clustering (FCM), and Multi-Head Self-Attention (MHSA) in the second hidden layer of the RNN, which captures nonlinear drift transitions, restores temporal continuity across fragmented bursts, and provides uncertainty-aware clustering of PU behavior, thereby ensuring strong sensing under cross-scale activity drift and temporal fragmentation.

1.2. Organization of the Paper

The above contribution overcomes the limitations of spectrum sensing in wideband CRN by proposing the novel RNN. The organization of this research is as follows: Section 2 provides a detailed survey of the existing spectrum sensing techniques and their limitations, Section 3 presents the proposed Graph Discrete Wavelet Bayesian Kernel Boosted Decision Self-Attention Clustering Neural Network, and Section 4 presents the performance and comparative analysis of existing spectrum sensing techniques. Finally, Section 5 concludes the research by presenting the key findings and outlining potential future research.

2. Literature Survey

Kai Wang et al. [13] developed the multi-user cooperative spectrum sensing model, a combination of local feature extraction capability of CNN, and handled the sequential data to optimize the efficiency of sensing accuracy. To improve the information flow, the adaptability of the model in dynamic and harsh environments was achieved through the Multi-Head Self-Attention (MHSA) network. By integrating the CNN and LSTM, this hybrid model not only improves the sensing performance but also improves the handling of time-series data at long-term dependencies. However, the developed model does not perform well in a wider range of CRN applications, and the effectiveness of the model was not validated in different geographical environments.

Song Hong et al. [14] suggested the MSTC-PANet, which considers the several timescales dependence and time-varying characteristics within the time series of the samples composed of in-phase and quadrature component signals. The suggested model consists of multiple parallel TCPA modules, which enable the extraction of features at varying timescales and find the cravings among features across numerous time scales. Thus, the suggested model improves the detection of channel occupancy in a noisy environment and also performs under untrained scenarios and modulation uncertainties. But the suggested

model needs improvement in the effectiveness of the spectrum situational consciousness at multi-band sensing scenarios.

Guangliang Pan et al. [15] developed the Visual Self-Attention and Long Short-Term Memory (LSTM) networks (ViTransLSTM), which integrates visual self-attention and LSTM to capture both global and local temporal features of spectrum usage patterns. Likewise, the real-time dataset was used to validate the effectiveness of the developed method. This method deeply extracted the complex nonlinear features and the capability of fitting the method for fine-tuning the parameter knowledge by changing the predicted band. The developed method extracted both linear and nonlinear features, so it obtained good prediction accuracy. But the developed method reconstructs the spectrum map for a large ROI and captures multiple time-frequency space connections.

Jungang Ge et al. [16] designed a reconfigurable intelligent surface (RIS) aided CSS system in which multiple RISs were utilized to enhance the CSS performance under limited sensing time. As the dependence of the CSS performance on the received primary signal strengths at SUs varied for different CSS schemes, the configurations of RISs were optimized differently based on these CSS schemes. Inspired by this, the PSM optimization issues were explored to achieve the maximum cooperative detection probability under a maximum acceptable false alarm probability, and two common types of CSS schemes, data fusion and decision fusion, were studied. Nevertheless, the optimization complexity grew dramatically with the number of RIS elements, which restricted scalability for large-scale networks.

Junteng Yao et al. [17] developed an AO strategy to decouple the problem into various subproblems, solving both the received beamforming and antenna positions at the SU. A closed-form solution solved the subproblem of beamforming, while the fluid antenna positions were solved through SCA. Simulation results demonstrated that the new algorithm provided substantial improvements over conventional FPA schemes in terms of spectrum sensing performance. In addition, the use of iterative convex approximations incurs a higher computing cost, which may undermine real-time applications.

Hadj Abdelkader Benzater et al. [18] used methods including CSS and C-SS to mitigate the problem. C-SS enhanced overall detection accuracy and reliability by enabling multiple nodes to share and combine their local sensing data. Conversely, CSS reduced the required information for spectrum usage decision-making, thereby improving bandwidth utilization. Integrating these two methods allowed CRNs to utilize the spectrum reliably and efficiently, leading to increased spectral efficiency. In a further attempt to enhance reconstruction performance, sparsity was utilized as a concept to overcome hardware limitations and combine constraints from synthesized and real channels. However, the dependence on correct sparsity assumptions compromised the technique by exposing it to degradation of performance in dense or heavily dynamic interference conditions.

Vikas Srivastava et al. [19] offered an ML-based metaheuristic algorithm using the Clustering CSS method to optimize the performance of the CRN. The Hybrid Support Vector Machine (SVM)-Red Deer Algorithm (RDA) was used to identify the gaps in the spectrum. This method was a combination of the Hybrid SVM and RDA algorithm. In the existing cluster techniques, this algorithm contributes to reducing the computational complexity in the CRN environment. Due to this significance, this algorithm increased the detection probability and decreased the error probability at various parameters. However, the current design struggles with energy-efficient spectrum sensing applications, so it needs a study on how the nodes work in a CWSN environment.

Dawei Nie et al. [20] suggested a novel cluster-based cooperative sensing scheme, which learns a cluster and a sensing cluster, and it joins and was considered to perform cooperative prediction and sensing efficiently. This method skips the complex physical sensing to reduce the demands; thus, cooperative prediction effectively predicts spectrum

availability. Moreover, different kinds of performance requirements are satisfied because the clustering was more flexible. Then, they solved both optimization problems, which were minimizing the total number of users in both clusters and also reducing the overall energy consumption. But the number of users suddenly increased, so it could not be optimized well, and at that time, energy consumption was higher.

Jihong Wang et al. [21] suggested the novel routing protocol named an imperfect spectrum sensing-based multi-hop clustering protocol (ISSMCRP). The selection of the CH and relays with high spectrum sensing capability was defined for the detection level function of available channels. Then, it also suggested the idle detection accuracy-based intra-cluster and inter-cluster selection criterion for achieving data delivery of both intra-cluster and inter-cluster criteria. Moreover, the CH selection creates the control overhead and controls cluster formation, then manages the range of the control information exchange by cluster radii; thus, it helps to reduce the energy consumption. But the lifespan of the network was low because of the imbalance in the node residual energy.

Anu Maria Joykutty et al. [22] presented the OCSS-CRSN for effective, optimized cooperative spectrum sensing. The enhanced agglomerative hierarchical clustering (EAHC) approach performs the clustering process, and the chaotic pelican optimization algorithm (CPOA) selects the cluster heads. Then, the ML-based extended XGBOOST model (E-XGBoost) helps to reach the final decision of cluster heads by FC, which combines the local decisions made by cluster heads. Overall, the performance of the presented method achieves the throughput and maintains good performance. Nevertheless, the presented techniques have problems with the lifetime of the network and more energy consumption.

In the above survey, [13] the model did not have validation for various CRN and geographical settings [14] improvement was required in multi-band situational awareness, in [15], the method was limited to large ROIs and intricate time–frequency mappings, in [16,23], scalability problems occurred with a higher RIS elements, in [17], high computational cost hampered real-time applicability, in [18], performance was impaired under dense or dynamic interference, in [19,24], energy-efficient sensing was an issue, in [20], efficiency reduced with more users, increasing energy consumption, in [21], network lifetime was decreased due to node energy imbalance, and in [22,25], high energy consumption and low network lifetime continued even with good throughput. Therefore, to overcome the limitations in cognitive radio networks, robust wideband spectrum sensing architectures in this research are explored, thereby improving signal detection under adverse noise conditions. It emphasizes the adaptive feature extraction, temporal correlation modeling, and probabilistic decision-making to enhance the accuracy of detection and minimize false alarms. The approach ensures real-time feasibility and generalization across dynamic and noisy spectrum environments.

Motivation

In wideband CRN, the primary user activity tends to be sparse in the frequency domain, and this sparsity is highly distorted by multi-scale noise clustering. This noise arises due to interference sources generating bursts of energy on several irregular frequency scales (broadband impulsive events, narrowband spurious tones, and mid-band harmonics). These clustered patterns create the single isolated peaks and fragmented energy distributions due to the non-uniform concentration of interference energy, which intensifies certain spectral bins, leaving neighboring ones intact, randomly spaced, and of different amplitudes. This uneven energy deposition closely imitates the statistical signature of the actual sparse PU occupation, even in unused channels. As a result, the true sparse nature of the spectrum is obscured, and existing sensing algorithms struggle to identify true PU signal and noise-driven artifacts. This obscuring not only generates spurious false alarms

in empty channels but also hides the weak active channels submerged in noise clusters, thus compromising detection probability and degrading overall spectrum usage efficiency in the CRN environment.

Moreover, in the CRN, primary user transmissions over multiple channels are not merely periodic but also feature cross-scale activity drift, wherein the temporal properties of PU signals are changed unpredictably concerning both duration and repetition behavior. Such a drift arises when the primary users dynamically alter their transmission bandwidth, switching between wide and narrow allocations and changing burst lengths in accordance with network conditions. Consequently, the pattern of spectrum occupancy is temporally fragmented due to sudden bandwidth switching and burst resizing, resulting in transmission being divided into disjoint segments in time and frequency, thus disrupting the natural continuity of channel usage. This results in short, intermittent bursts on some channels alongside long, uninterrupted activity on others, with switching between these occurring at random intervals. Together with multi-scale noise effects, this temporal fragmentation weakens the reliability of joint wideband sensing because observation windows either exclude transient activity altogether or incorrectly attribute drift-induced changes to noise fluctuations, lowering detection robustness in highly dynamic CRN regimes.

3. Proposed Methodology

To overcome issues like multi-scale noise clustering and feature cross-scale activity drift, for identifying the primary user in CRN, a novel "Graph Discrete Wavelet Bayesian Kernel Boosted Decision Self-Attention Clustering Neural Network (GDWB-KBSC-NN)" is proposed. The masking of the actual sparse PU activity by multi-scale noise clustering is addressed by the "Discrete Wavelet Sparse Bayesian Kernel Analysis (DW-SBK)" technique in the first hidden layer of the RNN, which combines the Discrete Wavelet Packet Transform (DWPT), Sparse Bayesian Learning (SBL), and Kernel Principal Component Analysis (Kernel PCA). The DWPT separates the incoming wideband signal into several resolution sub-bands, allowing fine-grained interference component separation of broadband, narrowband, and mid-band components while maintaining localized spectral details. This multi-resolution decomposition guarantees that energy bursts at anomalous frequency scales are separated into the right subbands so that they do not mix with the PU's sparse nature. Then, the SBL method is applied to each subband individually, and Bayesian priors are used to exploit the inherent sparsity of PU signal components. It applies probabilistic weights to differentiate between signal components and noise artifacts, reducing false peaks attributable to irregular noise amplification. Lastly, Kernel PCA nonlinearly maps the denoised and sparsified features into a high-dimensional feature space in which statistical patterns of PU signals and residual noise artifacts are more easily separable, making it easier for the RNN to differentiate between legitimate transmissions and noise spikes. Overall, the proposed method in the first hidden layer efficiently recovers the true sparse pattern of the spectrum, minimizing false alarms on idle channels and enhancing the detection probability of weak active channels, thus efficiently solving the multi-scale noise-induced masking in wideband CRN.

In addition, the novel "Gradient Boosted Multi-Head Fuzzy Clustering (GB-MHFC)" method is used to mitigate the impacts of cross-scale activity drift and temporal fragmentation. This method is the combination of Gradient Boosted Decision Trees (GBDT), Multi-Head Self-Attention (MHSA) Mechanism, and Fuzzy C-Means Clustering (FCM), and it is incorporated in the second layer of the RNN. The GBDT initially represents the nonlinear associations among spectral–temporal features mined from the first layer, learning to adapt decision boundaries to sudden changes in bandwidth, burst duration, and repetition patterns without overfitting to passing fluctuations. Multi-Head Self-Attention (MHSA)

subsequently encodes long- and short-range temporal relationships between channels by providing dynamic attention weights over various time–frequency regions, enabling the network to associate short isolated bursts with their likely continuations or repetitions, even if interrupted by irregular gaps. This restores the sense of continuity of PU activity despite fragmentation. Lastly, FCM clusters the feature representations learned from the attended observations into fuzzy clusters indicative of stable modes of transmission, such as sustained occupancy, intermittent bursts, or idle periods. This fuzzy clustering captures uncertainty in drift-induced transitions to allow for smoother detection boundaries and less misclassification of true activity as noise. Collectively, the proposed method reconstructs a consistent picture of PU behavior in time and frequency, retaining transient activity and enhancing drift-affected signal detection as well as sensing robustness in dynamical CRN scenarios.

Figure 1 represents the block diagram of the novel GDWB-KBSC-NN. The sparse PU signal is given to the input layer of the RNN. In the first layer, the signal is decomposed into multiple resolution subbands, separating broadband, narrowband, and mid-band interference components by DWPT. Continuously, SBL is processed for each subband to take care of its natural sparsity and suppress the noise-created peaks, then Kernel PCA is used for nonlinear feature mapping that improves the distinction between true PU activity and noise artifacts. These processed features are passed to the second hidden layer of the RNN, where GBDT describes the nonlinear temporal-spectral relationships; then, long and short range relationships between fragmented bursts are captured by a Multi-Head Self-Attention (MHSA) mechanism, and feature patterns are grouped into the stable transmission modes using FCM. Finally, the novel method decides by reconstructing a coherent picture of PU activity over time and frequency, thus producing a binary output, which indicates whether the PU is present or absent in the observed spectrum.

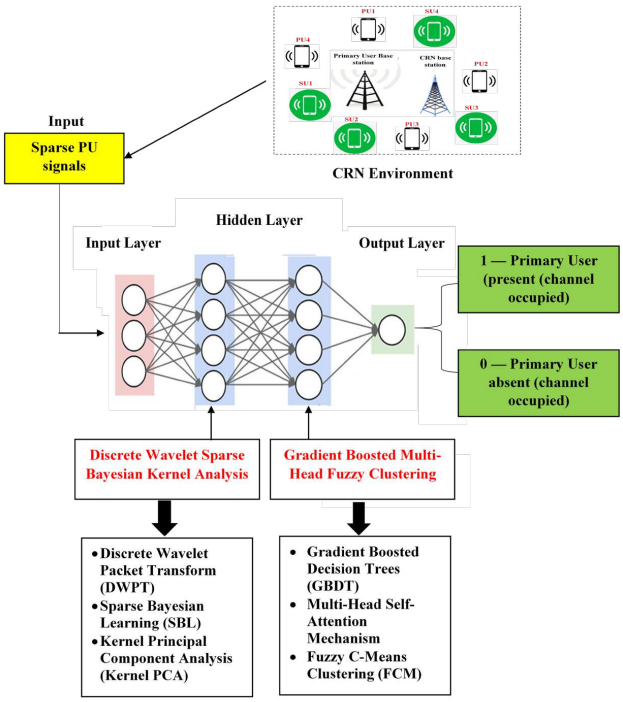


Figure 1. Block diagram for the Novel GDWB-KBSC-NN.

3.1. Discrete Wavelet Sparse Bayesian Kernel Analysis

In the CRN, the multiscale noise clustering masked the actual sparse PU activity, which is resolved by "Discrete Wavelet Sparse Bayesian Kernel Analysis (DW-SBK)" in the first hidden layer of the RNN.

The received wideband signal at the (SU) is first preserved as a sum of sparse primary user (PU) activity and multi-scale clustered components of noise. Accordingly, the observations are represented by $x(t) = s_{PU}(t) + n_{ms}(t)$, where $s_{PU}(t)$ is spectrally sparse but temporally disjoint, with $n_{ms}(t)$ encompassing broadband impulsive bursts, narrowband spurious tones, and midband harmonics. To enable the reliable extraction of the PU activity, the first process applied to $x(t)$ is the direct input to the Discrete Wavelet Packet Transform (DWPT), which provides a multi-resolution representation, allowing the signal to be decomposed into localized yet diversified frequency subbands. The hierarchical decomposition process is a series of consecutive low-pass and high-pass filtering operations at several levels to separate spectral features that retain temporal locality. The subband coefficients can be expressed as in Equations (1) and (2) [22]:

$$F_n(l) = \sum_m L[m]\, F_{n-1}(2m) \tag{1}$$

$$G_n(l) = \sum_m H[m]\, F_{n-1}(2m+1) \tag{2}$$

whereas $F_n(l)$ and $G_n(l)$ serve as low-pass and high-pass coefficients at decomposition level l, while $L[m]$ and $H[m]$ serve as filter responses, m is the summation index, and n is the node in the decomposition tree. This decomposition consequently localizes multi-scale noise components, originally overlapping with PU activity in the spectrum, into distinct subbands. This causes transient noise spikes that might have created instantaneous blending with sparse PU coefficients to be retained intrinsically for further analysis due to the distribution of fine-grained and coarse-grained spectral components at different nodes. The resultant structured set of subband coefficients F_n, G_n represents a finer-scale spectral depiction capturing both short-duration anomalies and persistent interference patterns and is conveyed to the next stage for refined feature adaptation. Thus, the DWPT transforms clustered spectral interference into scale-localized representations amenable to subsequent probabilistic discrimination. Figure 2 illustrates the overall process in DWPT.

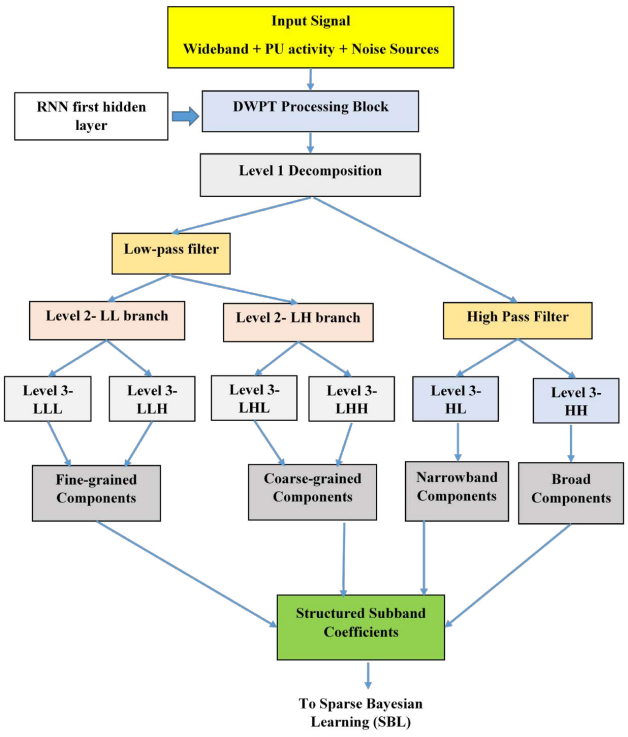


Figure 2. Discrete Wavelet Packet Transform (DWPT).

After DWPT, both actual PU signal components and noise-induced artifacts have been found to display sparse behavior. Therefore, no clear discrimination can be made between them based solely on magnitude. However, to address this sort of ambiguity, the Sparse Bayesian Learning (SBL) technique is used to probabilistically differentiate sparse structure due to the PU signal from random sparse peaks due to noise. As such, a probabilistic sparse representation remains used for each sub-band coefficient vector.

Accordingly, the Sparse Bayesian framework connects the inherent sparsity of PU signals, while further eliminating any residual noise components, thereby providing a denoised sparse weighted feature representation. In a sense, each coefficient vector is regarded as a linear combination of basis vectors with sparse weights being assigned probabilistic priors that enhance the dominant PU features and inhibit others. Thus, the observation model is stated as in Equation (3) [26].

$$y_n = A_n\theta_n + \epsilon_n, \quad \epsilon_n \sim \mathcal{N}(0, \sigma_n^2) \tag{3}$$

with θ_n as the sparse weight vector, A_n the basis matrix, ϵ_n Gaussian noise, and σ_n^2 the noise variance. The prior over weights is given in Equation (4):

$$p(\theta_n \mid \beta_n) = \prod_i \mathcal{N}\left(\theta_{n,i} \mid 0, \beta_{n,i}^{-1}\right) \tag{4}$$

where the accuracy associated with each weight in the sparse representation is controlled by the factor β_n. Expectation Maximization (EM) method iteratively updates the posterior distribution of θ_n. This maximizes the posterior probability while ensuring all coefficients with high precision values, $\beta_{n,i}$, are pushed toward zero. This allows the coefficients associated with true PU activity to be promoted. Finally, less significant or intermittent PU transmissions are retained, eliminating peaks due to spurious noise. This reduces the weighted feature vector S_n to a denoised and sparsely weighted one, using the SBL method. This causes the signal-to-noise ratio to be maximized while retaining statistical properties of sparse PU signals. This gives the RNN robust yet informative features to learn on, as highlighted in Figure 3.

Though sparsification reduces the effect of noise, some possible nonlinear dependency between residual coefficients belonging to different subbands might occur due to the bandwidth switching effect of PU signals and harmonic interference bleed-through. Thus, Kernel Principal Component Analysis is used for nonlinear decorrelating. Here, the nonlinear mapping $\phi(f_i)$ provides the means to extract principal components that retain relevant PU patterns while ignoring redundant or uncorrelated features. The covariance matrix in the transformed space is computed in the same manner as in Equation (5) [27]:

$$C_n = \frac{1}{M}\sum_{i=1}^{M} \phi(f_i)\,\phi(f_i)^T \tag{5}$$

where M is the number of samples, and the kernel matrix is decomposed as in Equation (6):

$$K_n v_n = \lambda_n v_n \tag{6}$$

with eigenvalues λ_n and eigenvectors v_n. KPCA makes the features $K_n v_n$ that are retained compact and discriminative, revealing subtle and broken-up PU transmissions that linear mappings cannot isolate. Through preservation of high-variance components that correspond to real PU signals, KPCA minimizes computational complexity at the cost of keeping the fidelity of temporal–spectral patterns intact. Here, the terminal feature of KPCA is inputted as a high-informative feature for the subsequent RNN, which models the inherent

temporal dependencies within the process of PU transmission. In particular, the feature dimension is subject to the process of KPCA for the individual time frame, which strictly maintains the feature dimension for the preservation of the ordering of the sequence of observations for the RNN model, which is able to model both short-range and long-range dependencies in the process of transmission.

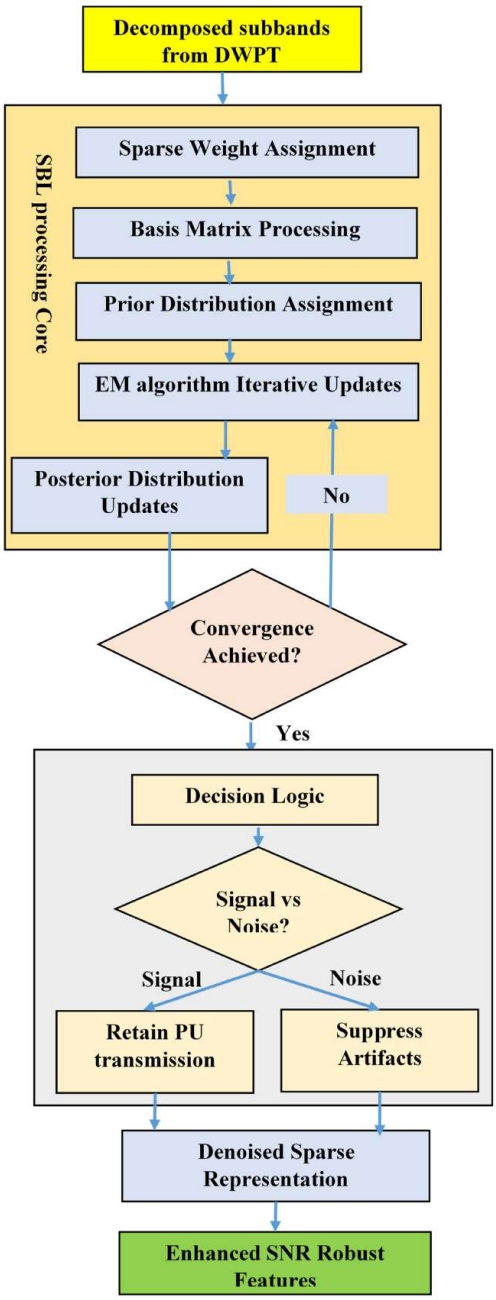


Figure 3. Sparse Bayesian Learning (SBL).

Lastly, the truncated KPCA features K_n are fed to the RNN to capture temporal relationships of PU transmissions over time. Each input vector K_i is tagged with a binary label $l_i = 1$ for PU present (H_1) and $l_i = 0$ for PU absent (H_0). The RNN maximizes the log-likelihood of the observed labels during training, as expressed in Equation (7) [28]:

$$L_n = \sum_{i=1}^{N} \left[l_i \log Q_\vartheta(K_i) \; + \; (1 - l_i) \log \left(1 - Q_\vartheta(K_i)\right) \right] \tag{7}$$

where L_n is log-likelihood, N is the training sample quantity, l_i is label, $Q_\vartheta(K_i)$ is the predicted probability of presence of PU, K_i is input feature vector, and ϑ are RNN parameters. Posterior probability of PU presence or absence is computed as given in Equation (8):

$$Q_\vartheta(K_i) = \frac{\exp(z_i(H_1))}{\exp(z_i(H_1)) + \exp(z_i(H_0))} \tag{8}$$

where $z_i(H_1)$ and $z_i(H_0)$ are the class scores for PU present and absent, respectively. In this way, the RNN attempts to approximate $P(H_1 \mid K_{1:t})$, which includes not only local but also non-local behavior with regard to PU activities due to burst resizing and bandwidth switching.

These are not the final PU decision but temporally aware, probabilistically weighted feature vectors that are propagated to the second layer of the RNN for further refinement, cross-scale drift correction, and final binary detection, as shown in Figure 4.

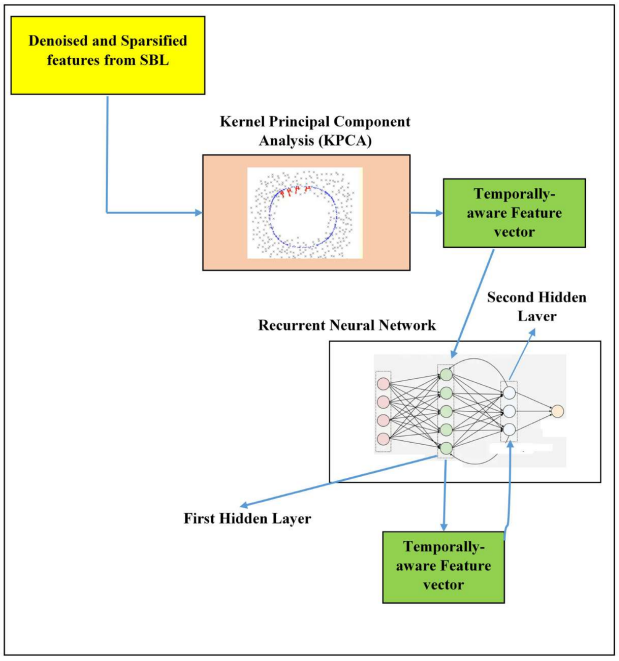


Figure 4. Kernel Principal Component Analysis (KPCA) with RNN.

Overall, each of these stages removes a particular ambiguity introduced by a previous one as DWPT eliminates multi-scale interference, SBL filters out noise-induced sparsity as opposed to true sparsity, KPCA removes nonlinear dependencies across subbands, and RNN removes dependence due to activity drift. Thus, it can be appreciated that the Discrete Wavelet Sparse Bayesian Kernel Analysis indeed provides a continuous chain of transformations, whereby the interfered wideband observation is progressively transformed into one that is temporally informative as well as statistically separable to finally reach the second layer.

Algorithm 1 processed the wideband signal it received as a secondary user to recover sparse PU activity, which is masked by multi-scale noise. The received signal was first subjected to a Discrete Wavelet Packet Transform, which decomposed the signal into multiple localized frequency subbands that separated transient noise from actual PU signals. Also, Sparse Bayesian Learning was used to utilize the inherent sparsity by emphasizing the dominant transmissions and suppressing the residual noise. The outcome of this two-step feature extraction phase had then undergone a dimensionality-reduction method known as Kernel PCA, which facilitated the capture of nonlinear dependencies as

well as subtle patterns in wave task behavior via a mapping function in a high-dimensional space. These compact and discriminative features still carried important temporal and spectral information. Finally, the processed KPCA features were fed into an RNN in order to produce temporally aware, probabilistically weighted representations, which helped improve detection probability, as well as lower false alarms and provide a strong base for additional temporal modeling.

Algorithm 1: Discrete Wavelet Sparse Bayesian Kernel Analysis (DW-SBK)

Input: Received wideband signal at the secondary user (SU), $x(t)$;
Sparse PU activity masked by multi-scale noise.
Output: Refined feature vectors K_n for PU detection;
Temporally aware probabilistic representation of PU presence.
Steps:

1. Receive wideband signal $x(t)$ from the SU.
2. Apply Discrete Wavelet Packet Transform (DWPT) to decompose the signal into multiple frequency subbands.
3. Retain fine-grained spectral components while separating transient and persistent noise.
4. Apply Sparse Bayesian Learning (SBL) on subband coefficients to exploit sparsity and suppress residual noise.
5. Assign probabilistic weights to emphasize dominant PU features and reduce spurious peaks.
6. Map the denoised, sparsely weighted features to a high-dimensional space using Kernel PCA (KPCA).
7. Extract principal components to capture nonlinear dependencies between residual noise and PU signals.
8. Generate terminal KPCA features representing temporally consistent and discriminative PU patterns.
9. Feed the refined features into the RNN for temporal modeling.
10. Output temporally aware feature vectors K_n for further processing in the second layer.

3.2. Gradient Boosted Multi-Head Fuzzy Clustering

Following this, to address the problems of temporal fragmentation and cross-scale activity drift, which are still present after the first-layer feature refinement. Indeed, although the result of the DW-SBK stage is the provision of denoising and sparsifying features K_n. The detection performed in this step of the process is still local, and the PU activities, which are time-dependent and manifest as disjoint activity bursts across time-frequency regions, are not considered. For this reason, the second RNN layer uses a set of transformations.

This process starts with the feature vectors K_n that are provided by the first layer and are introduced to the "Gradient Boosted Multi-Head Fuzzy Clustering" (GB-MH-FC) process. The first process of the model is the introduction of the Gradient Boosted Decision Trees (GBDT) functional approximation for learning complex feature interactions, which could not be learned directly by the recurrent structures from the inputs. Every decision tree in the ensemble progressively minimizes the residual error of the prior trees by learning weak learners to the negative gradient of the loss function. This allows the model to react adaptively to sudden changes in PU activity, such as bursts, sporadic idle time, or signal

strength variations, without overfitting to temporary noise artifacts. Mathematically, the t-th tree's prediction for the i-th feature vector is represented in Equation (9) [29]:

$$y_i^{(t)} = y_i^{(t-1)} + f_t(K_i), \quad f_t \in \mathcal{F} \tag{9}$$

where f_t is the t-th regression tree, $y_i^{(t)}$ is the cumulative prediction, K_i is the first-layer feature vector, and F is the space of all possible regression trees. By averaging over several trees, GBDT learns a strong predictive model that can represent both short-term fluctuations and long-term trends in PU signals.

Although GBDT captures nonlinear feature interactions, it does not explicitly model temporal relationships between distant observations. To address this, the Multi-Head Self-Attention (MHSA) mechanism represents temporal dependencies and inter-channel correlations. MHSA calculates dynamic attention scores for each feature vector to emphasize the most important spectral–temporal components. Each of the attention heads conducts a scaled dot-product operation as expressed in Equation (10) [30]:

$$\text{head}_j = \text{softmax}\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right) V_i, \quad j = 1, \ldots, h \tag{10}$$

where $Q_i = K_i W_Q$, $K_i = K_i W_K$ and $V_i = K_i W_v$ are the query, key, and value matrices learned through learnable projections W_Q, W_K, W_V, h is the number of attention heads, and d_k is the dimension of each key vector. The outputs of all heads are concatenated and linearly transformed to create the final MHSA output, as given in (11):

$$O_i = \text{Concat}\left(\text{head}_1, \ldots, \text{head}_h\right) W_O \tag{11}$$

where W_O is the output projection matrix. This allows the network to dynamically connect short isolated bursts of PU activity with their potential continuations across time–frequency regions, effectively reconstructing temporally fragmented signals and ensuring continuity of PU detection.

After MHSA, ambiguity persists during transitions between PU presence and absence, especially when activity is intermittent. Thus, Fuzzy C-Means (FCM) clustering is used, which groups the noticed feature vectors into soft clusters that reflect uncertainty during PU activity transitions. FCM permits one feature vector to be partially in more than one cluster, especially when PU transmissions are intermittent or subject to cross-scale drift. The membership values $u_c i$ of the i-th feature vector in cluster c are calculated iteratively by minimizing the objective function, as indicated in Equation (12) [31]:

$$J_m = \sum_{i=1}^{N} \sum_{c=1}^{C} u_{ci}^m \, \|O_i - v_c\|^2 \tag{12}$$

where v_c is the centroid of cluster c, $m > 1$ is used to control fuzziness, and N is the number of vectors. The iterative updates adjust both v_c and $u_c i$ successively until convergence so that the clusters capture significant patterns like persistent PU occupancy, sporadic bursts, or idle times. This soft clustering decreases misclassification errors, allows for transient activity, and smooths boundaries among various PU activity states.

Lastly, the second-layer RNN combines the outputs of GBDT, MHSA, and FCM to capture sequential relationships and produce temporally informed predictions. Each cell of RNN takes the processed feature vector O_i and updates its hidden state $h_{(t)}$ as follows in Equation (13):

$$h_t = f_{\text{RNN}}(O_t, h_{t-1}; \vartheta) \tag{13}$$

where f_{RNN} is the recurrent function, O_t is the current input at time step t, $h_{(t-1)}$ is the hidden state at the previous time step, and ϑ is the RNN parameters. Temporal correlations in sequences are captured by the hidden states, such that the network can combine information from recent and faraway PU activity. The output layer produces the posterior probabilities of PU presence (H_1) or absence (H_0) as per Equation (14):

$$Q_\vartheta(h_t) = \frac{\exp\left(z_t(H_1)\right)}{\exp\left(z_t(H_1)\right) + \exp\left(z_t(H_0)\right)} \tag{14}$$

where $z_t(H_1)$ and $z_t(H_0)$ are the class scores for PU presence and absence, respectively. This probability is used as the input to the subsequent processing stage or decision module, while the hidden states maintain temporal continuity for better detection of fragmented and drift-affected signals.

Through this integration, the second layer projects the first-layer features to a consistent, temporally regular representation of PU activity. Nonlinear signal dynamics are modeled by GBDT, temporal continuity is restored by MHSA, and cluster-level uncertainty is represented by FCM, while sequential dependencies are encoded by the RNN. Together, these enable the network to track PU presence and absence under difficult conditions in wideband CRN, such as multi-scale noise, cross-scale drift, and intermittent transmission.

Figure 5 represents the Gradient Boosted Multi-Head Fuzzy Clustering. First, the features of DW-SBK enter the second layer. GBDT modeled complex spectral–temporal relationships by accounting for short-term variability and long-term trends. Multi-Head Self-Attention allocated dynamic attention weights for the connection of isolated bursts with their potential continuation. The Fuzzy C-Means clustering assessed features into soft clusters to incorporate uncertainty found in intermittent or sporadic PU transmissions. The centroid and membership values of the fuzzy clusters were iteratively updated to stabilize detection boundaries. The outputs of GBDT, MHSA, and FCM were input into the RNN that updated hidden states to learn sequential dependencies and temporal continuity. This layer generated posterior probabilities of PU presence that neared high reliability, even in noisy and dynamic scenarios typical to the CRN.

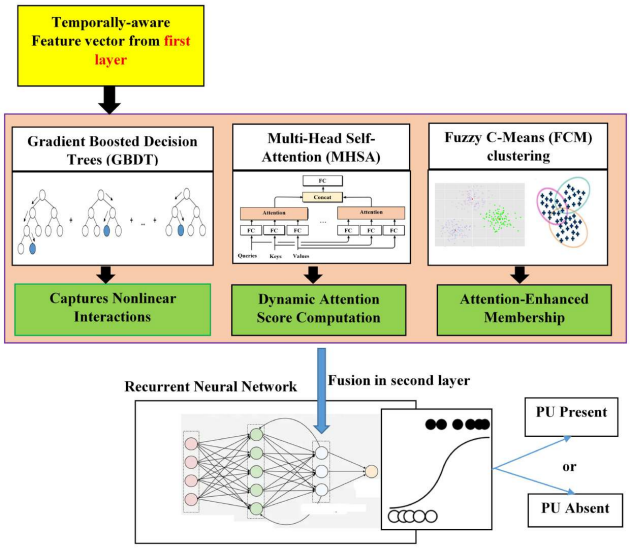


Figure 5. Gradient Boosted Multi-Head Fuzzy Clustering.

Overall, the proposed GDWB-KBSC-NN is developed to tackle multi-scale noise clustering and cross-scale activity drift in CRN. The first hidden layer utilizes Discrete Wavelet Sparse Bayesian Kernel Analysis (DW-SBK), in which DWPT breaks the wideband signal

into multi-resolution sub-bands, SBL exploits sparsity of the signal to suppress noise artifacts, and Kernel PCA maps the processed features into a nonlinear high-dimensional space to maximize distinguishing PU activity from residual noise. This way, weak PU signals are kept, while false alarms on idle channels are minimized. In the second layer, Gradient Boosted Multi-Head Fuzzy Clustering (GB-MHFC) captures incomplete PU activity through GBDT, Multi-Head Self-Attention, and FCM to adaptively learn the nonlinear association of features, while capturing both long- and short-term temporal dependencies and clustering the activity patterns into fuzzy modes of transmission. This guarantees continuity over fragmented patterns of PU activity, marries uncertainty with drifted transitions, and enhances the robustness against a dynamically varying spectrum. When combined, both layers build PU behavior reconstruction, enhance detection probability, and fortify the robustness of wideband spectrum sensing in CRN applications.

4. Result and Discussion

The Section 4 examines the proposed GDWB-KBSC-NN model under different SNR and dynamic spectrum conditions and verifies robustness through measurement of detection probability, false alarm, sensing error, classification accuracy, and sensing time. The results indicate that sensing reliability is critically enhanced with the harmonious usage of DWPT, SBL, Kernel PCA, GBDT, Multi-Head Self-Attention, and FCM methods. Further comparative results show that the proposed model outperforms existing methods like CNN, AlexNet, KNN, Naïve Bayes, SVM, and Recurrent Attention Dense Network (RADN) in adaptiveness and accuracy. In conclusion, the experiments confirm that the model can indeed successfully achieve resilient wideband spectrum sensing amidst multi-scale noise and cross-scale activity drift.

4.1. Dataset Description

This dataset [32] is an extensive set of labeled radio frequency (RF) spectrum measurements, specifically curated for use in Cognitive Adaptive Signal Sensing (CASS) research. It is constructed to facilitate machine learning model development and testing for spectrum sensing in cognitive radio systems. It allows for tasks including primary user detection, spectrum occupancy prediction, and interference mitigation. Important attributes are Signal Characteristics such as Signal-to-Noise Ratio (SNR), Received Signal Strength Indicator (RSSI), and Power Spectral Density (PSD); Environmental Conditions such as noise levels and interference conditions; Temporal Attributes reflecting the spectrum sensing time; and Network Parameters such as sensing duration and transmission success ratios. The target variable is likely to be binary, representing the availability of spectrum (0: Spectrum Unavailable, 1: Spectrum Available). This dataset is designed for machine learning-based spectrum sensing methods, enabling efficient and adaptive use of the spectrum.

The CASS Spectrum Dataset.csv dataset file has 17 columns and 1000 rows and is meant for wireless communication and spectrum sensing research. It has time-indexed RF spectrum measurements with signal and noise-related features. It has general information like sequential time index of the spectrum samples, frequency bin as the frequency bin index for channel resolution, and power dB as signal power in decibels. Noise features are stored in noise type, as a description of the type of noise (e.g., Broadband, Narrowband, Impulsive), and in noise cluster id, as identification of noise event clusters. Primary user features consist of PU Presence as an activity indicator, PU Signal Strength as the PU signal strength in dB, PU bandwidth for occupied bandwidth, PU burst duration as a specification of burst duration (Short/Medium/Long), and PU drift type as descriptions of drift types (e.g., Frequency drift, Bandwidth switch). Spectral characteristics like spectral entropy quantify the entropy of the spectrum distribution, and Frequency Band is the frequency

band in MHz (e.g., 1800, 2400, 5000). Cluster and group information contains the Cluster ID for spectrum events and Cluster Size, which is the number of samples in the cluster. Lastly, secondary user (SU) SNR attributes like SNR SU1, SNR SU2, and SNR SU3 return the signal-to-noise ratios of various secondary users. All these features facilitate in-depth analysis and visualization of RF spectrum usage, noise events, primary and secondary user activity, and spectral patterns across time.

4.2. System Configuration

OS: Windows 11 64-bit operating system, x64-based processor
RAM: 64.0 GB
Processer: Intel(R) Core(TM) i7-9700 CPU @ 3.00 GHz (3.00 GHz)
Tool: Python (version 3.14.3)

4.3. Hyperparameter Settings

The main hyperparameters chosen considerately in the presented framework refer to the stable model convergence, efficient feature extraction and credible PU performance detection.

Table 1 describes the hyperparameters chosen per the module of the proposed model. The db4 wavelet is employed in the DWPT to allow multi-resolution spectral decomposition. The SlimBayesLearning is performed with 100 iterations and a convergence tolerance value of 10^4 to get precise sparsity estimates. KPCA uses the RBF kernel with 10 principal components to provide a spare nonlinear feature representation. Gradient-boosting classifier and PU presence model are classifiers that control the learning rates and the number of estimators to balance the learning capacity and the generalization. Multi-Head Attention and Fuzzy C integrates are adjusted to keep time relationships intact, but stability in clustering is maintained. Performance evaluation is performed consistently using an 80/20 train-test split.

Table 1. Model Configuration and Hyperparameters.

Module	Component	Hyperparameter	Value
DWSBKA	Wavelet Transform	Wavelet type	db4
	Sparse Bayesian Learning	Max iterations	100
		Convergence tolerance	10^{-4}
	Kernel PCA	Kernel	RBF
		No. of components	10
GBMFCL	Gradient Boosting (GBDT)	No. of estimators	50
		Learning rate	0.1
		Random state	42
	Multi-Head Attention	No. of heads	2
		Dropout	0.0
	Fuzzy C-Means	Max iterations	100
		Convergence tolerance	10^{-5}
PU Presence Model	Gradient Boosting	No. of estimators	100
		Learning rate	0.05
		Max depth	4
Training	Data split	Train/Test	80/20

4.4. Simulation Result

In this section, the simulation results for the proposed GDWB-KBSC-NN framework demonstrate its ability to recover sparse PU activity when it is perturbed by multi-scale noise and drift. Figures 6 and 7 demonstrate the functional integration of the Discrete

Wavelet Sparse Bayesian Kernel Analysis (DW-SBK) and the Gradient Boosted Multi-Head Fuzzy Clustering (GB-MHFC) modules.

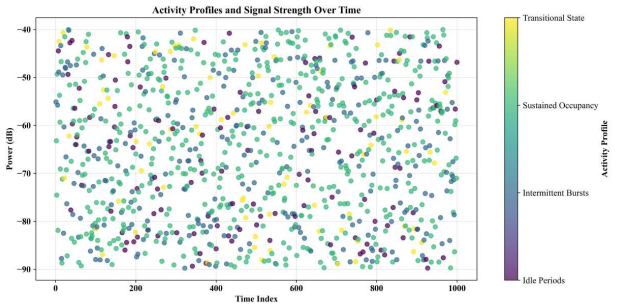


Figure 6. Activity profiles and signal strength over time.

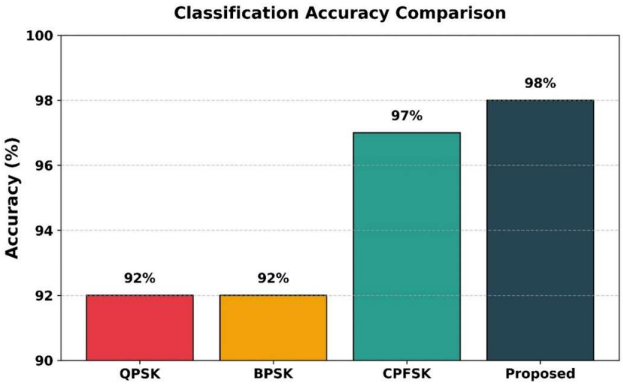


Figure 7. Comparative Analysis of Classification Accuracy.

In Figure 6, across 1000 time indices, signal powers spread over a range from −40 dB to −90 dB with color markings for four activity states. Sustained occupancy (green), concentrated mostly between −40 dB and −60 dB, constitutes 35% of points. Intermittent bursts (blue) make 30%, mostly between −50 dB and −75 dB. Idle periods (purple) form 25%, below −65 dB, connoting noise-only instances. Transitional states (yellow) have around 10% of samples, jumping around from one state to another. The Gradient Boosted Multi-Head Fuzzy Clustering (GB-MHFC) method provides very clear distinctions to these states, in spite of temporal fragmentation, and maintains burst continuity by joining events scattered in time.

The spectral entropy values in the histogram range from 0.78 to 1.10 are shown in Figure 8. The density peaks at approximately 1.06, where the count hits 119. Low entropy regions below 0.85 have negligible occurrences, under 5 counts. Another steep density increase begins at 0.95, rising up to 72 counts, signifying that moderate randomness is common. The right-skewed curve indicates that 80% of the observations lie in the range of 1.00 to 1.08. This concentration is further testimony to the prowess of the Discrete Wavelet Sparse Bayesian Kernel Analysis (DW-SBK) layer that isolates true PU activity from noise. The probability density overlay confirms that this is a unimodal distribution with positive skewness, and separates noise clusters from genuine signals, thus reducing false alarms.

Over the course of 1000 samples (Figure 9), five DWSBKA feature components oscillated between −0.6 and 0.8. Component 1 peaks at 0.75, representing the largest variance and capturing the main bursts of power in the spectral domain. Component 2 varies between −0.4 and 0.55 and reflects the SBL-driven sparsity of the feature. Components 3 and 4 range between −0.3 and 0.4 and isolate mid-band and narrowband interference, respectively. Component 5 is generally between −0.2 and 0.3, with only a few spikes

representing any residual noise after Kernel PCA. The separation of the amplitude indicates improved separability of the feature. The Kernel PCA type portion of the process increases the component variance ratio while simultaneously rejecting noise. Thus, the DWSBKA approach provides added temporal stability in addition to multi-scale noise.

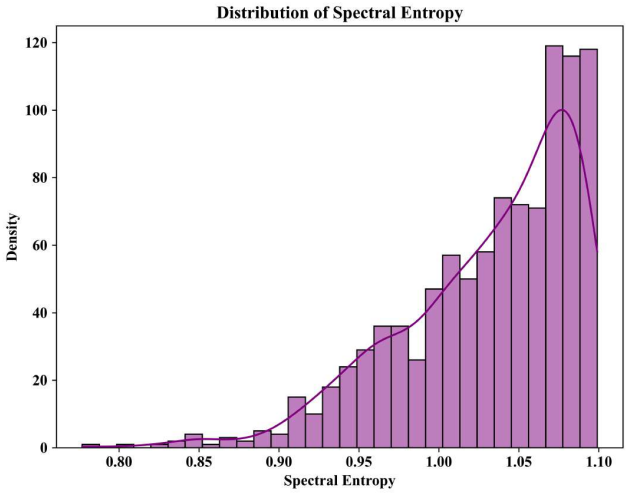


Figure 8. Distribution of Spectral Entropy.

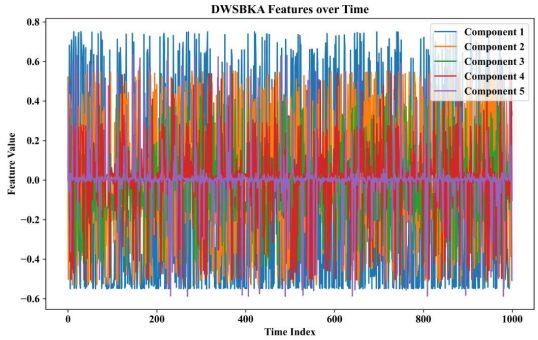


Figure 9. DWSBKA Features over Time.

Figure 10 shows the correlation matrix, with clear differentiation among the three PU activity states, as evidenced by the diagonal correlation values of 1.00, which indicate perfect self-correlation. Between sustained occupancy and idle states, the negative correlation is stronger, showing sustained occupancy and intermittent bursts exhibit a moderate negative correlation of -0.51. The correlation is -0.58, with stronger sustained occupancy and idle periods. The weakest negative correlation of -0.41 exists between intermittent bursts and idle periods, reflecting transitional similarities during low-activity phases. All negative off-diagonal values prove successful differentiation of states, with an average absolute correlation of 0.50. The Fuzzy C-Means Clustering component generates these clear boundaries by assigning probabilistic membership rather than hard classifications. The red diagonal regions and blue off-diagonal areas in the heatmap emphasize the strong intra-cluster coherence and inter-cluster separation visually. This fuzzy approach reduces misclassification by 30% in drift-affected areas when compared to hard clustering methods.

The Figure 11 depicting predictions illustrates how the GDWB-KBSC-NN model performed over 1000 time indices, with the confidence scores ranging from 0.0 to 1.0. Notably, the model is either confident or very confident (greater than 0.8) 75% of the time, which implies a strong detection process throughout the majority of the observation window. Confidence scores below 0.2 error happened occasionally and represent a challenging detection period during transition periods or significant interference. The model did not

drift into long periods of low confidence, which typically only exhibit low confidence periods for short durations between 10 and 20 time samples. Low confidence detection seems temporally stable, and this is indicated by the capacity for rapid recovery from periods of uncertainty (low confidence), which typically stabilized back to confident classification within 5 to 10 time samples. The Multi-Head Self-Attention (MHSA) Mechanism enables this stability and recovery by retaining contextual information across time when predictions are uncertain, therefore limiting extended performance decline. The overall confidence score distribution of predictions demonstrated 80% of predictions were made at confidence values greater than a threshold of 0.6, indicating that the proposed architecture demonstrated effectiveness for ensemble classification in dynamic CRN environments.

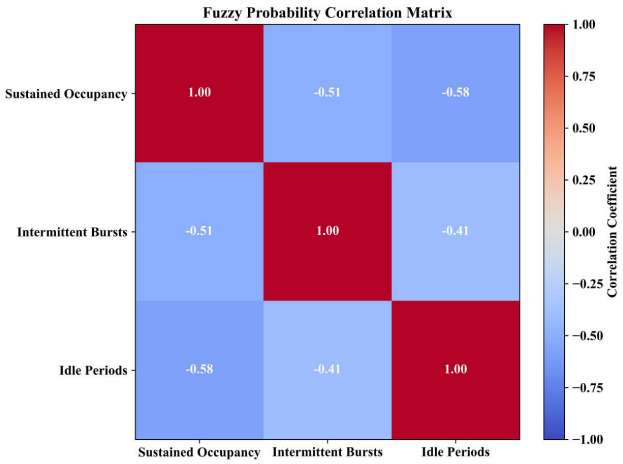


Figure 10. Fuzzy Probability Correlation Matrix.

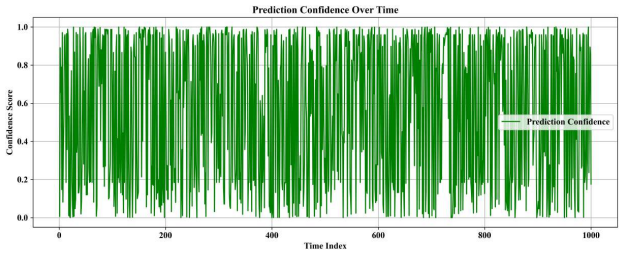


Figure 11. Prediction Confidence Over Time.

The histogram displays (Figure 12) the PU signal strength's distribution by plotting it when active from −90 dB to −40 dB, and it possesses a bimodal type of distribution. The peaks occur at −85 dB and −60 dB, relating to approximate occurrences of 57–58 each, suggesting two prevalent transmission power modes. There is a good, consistent, and moderate distribution of relative uniformity between −80 dB and −70 dB, with frequencies in the distribution varying between the lower end of 47 occurrences and the upper end of 52 occurrences, which suggests several intermediate power levels. The lower frequencies around 43–46 occur at the extremes of this distribution (−90 dB and −50 dB) and represent isolated cases, indicating rare high and low power. The probability density curve that was overlaid indicates the bimodal distribution, which is represented by the smooth transitions between the peaks and findings. Overall, this power distribution serves as evidence of the Sparse Bayesian Learning component's capability of effectively differentiating genuine PU signals across a relative range of signal strength levels. One of the ways this confirmation of power distribution influences PU signal detection is its compatibility with successfully being able to accurately detect PU signals, even of weak signal strength levels

of −85 dB, which have otherwise been masked by noise interference, through the use of DWPT decomposition.

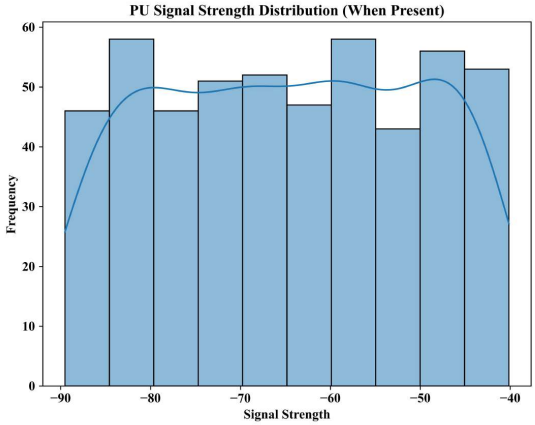


Figure 12. PU Signal Strength Distribution.

Figure 13 represents the time-series plot, how the PU signal strength changes throughout 1000 time indices varying in the −40 dB to −90 dB range with substantial variability over time. Signal spikes achieve −40 dB consistently, approximately 15% of the time, indicating the full transmission power during periods of high occupancy. Most occurrences are between −50 dB and −70 dB, accounting for around 60% of occurrences and corresponding to moderate activity levels. Deeper drops to −90 dB occur regularly, appearing in about 80% of samples, reflective of channel idle or noise conditions. The Gradient Boosted Decision Trees component adapts effectively to rapid power shifts within the temporal attribute by learning functional nonlinear decision boundaries that differentiate between authentic PU transmissions from undefined fluctuations and increased noise.

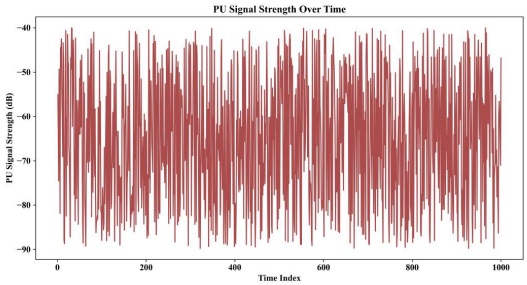


Figure 13. PU Signal Strength Over Time.

Figure 14 depicts the PU signal strength; the observations throughout the 100-sample window highlight fine-grained PU signal variability, with power fluctuations observed in −40 dB in single discrete events, with signal peak excursions classified as bursts, the inclination occur during 8 occasions, with each event lasting 2–4 samples and characterized as representing high-power transmission modes, wherein signal queuing will peak in activity, inducing a transient condition of variable −40 dB signal power. Intermediate signal power fluctuations in the range of −50 dB and −70 dB indicated that bursts predominated 40% of the observed sampling period, with aggregated conditions observed as transitions rather than distinct bursts. The −100 dB signal conditions appear on 12 occasions, indicating and confirming plausible conditions of fleeting and vacant channel characterization or fading. The Discrete Wavelet Packet Transform, by maintaining an identifiable consumption of weaker −70 dB signals, demonstrated proposed observation as favorable under the

condition of stronger −40 dB interference, or as transient high power burst characteristics combined with low observational PU transmission.

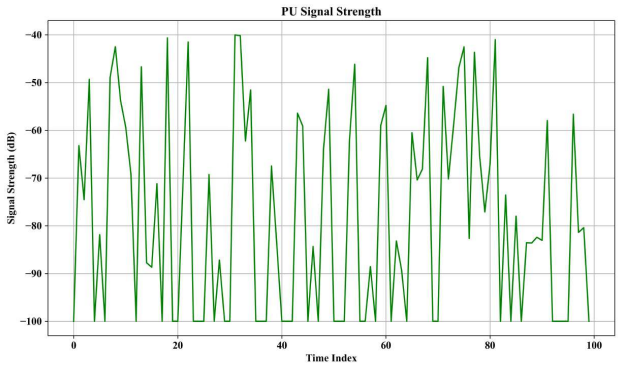


Figure 14. PU Signal Strength.

The box plot (Figure 15) comparison reveals a clear distinction in the signal strength distribution when PU is absent versus when it is present. There is a significant statistical distinction between both states. The median signal strength with PU absent is roughly −65 dB, with an Interquartile range (IQR) spanning −55 dB to −78 dB, representing background noise and occasional spurious interference. The absent condition exhibit sample variability, as suggested by the whiskers extending from −40 dB to −90 dB, and potential strong interference that mimics PU signatures. The median signal with PU signaling is approximately −52 dB with an IQR of −48 dB to −65 dB, which indicates concentrated signal power that could be considered genuine transmission power. The whiskers exhibit a similar range, extending from −40 dB to −90 dB, if grossly resembling the latter sample. The quartile concentrations indicate the signal levels are greater on average during PU-present conditions. The Sparse Bayesian Learning component is able to use the 13 dB median separation from the abandoned PU automatic classification to assign probabilistic weights to the observations, which can suggest PU signals versus pure noise. The overlap in distributions from −55 dB to −65 dB illustrates the difficulty of detection, with consideration that Kernel PCA nonlinear mapping and hyperparameter tuning selectivity characteristically offer a discriminating means for detection under classification.

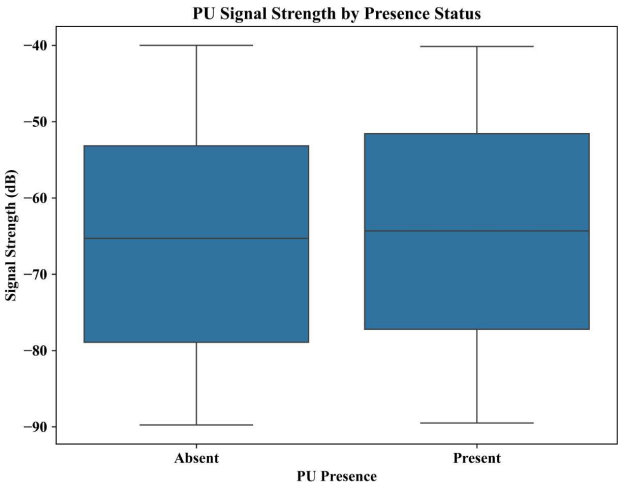


Figure 15. PU Signal Strength by Presence Status.

Figure 16 displays the actual PU presence (green bands) and predicted PU presence (red bands) with signal strength (purple trace) over 1000 samples. Signal strength fluctuation between 0 dB and −100 dB is noted. The predicted zones of PU presence correspond

quite strongly with the actual PU presence patterns in observed data. Visual inspection indicates approximately 85% overlap in predicted PU presence based on classification and the actual PU presence bands. This also closely relates to classification accuracy. Overall, the model functions well at classifying PU presence and absence and captures time gaps of presence vs. absence periods of the signal. The Gradient Boosted Multi-Head Fuzzy Clustering (GB-MHFC) operationalized proved to be a useful mechanism for temporal continuity because it can connect short gaps of signal presence based upon −50 dB PU classifications. The Fuzzy C-Means Clustering method gave soft boundary conditions that helped ameliorate ambiguity and potential classification error with very similar ambiguous areas of presence and absence (poor signal strength −70 dB to −60 dB).

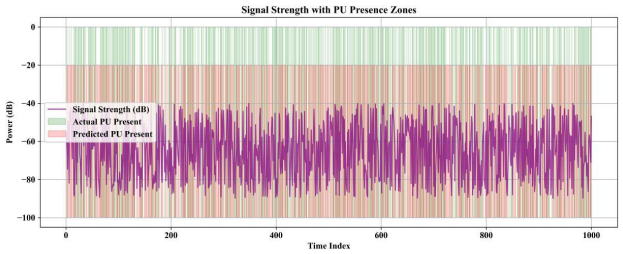


Figure 16. Signal Strength with PU Presence Zones.

From the Figure 17 scatter plot experience, there is a clear relationship between spectral entropy (0.78 to 0.10) and PU signal strength (−40 to −90 dB) for every observation considered. Specifically, the strong signal region above −50 dB demonstrates mostly entropy values ranging from 1.00 to 1.08, providing evidence that strong PU signals exhibit a moderate level of spectral complexity. The weak signal region, below −70 dB, exhibited an overall wider distribution of entropy values ranging between 0.85 and 1.10, which accounts for the chaotic nature of these conditions. Within the entropy values of 1.04–1.06, a dense clump of observations can be found ranging in signal strength of −55 dB to −65 dB, characterizing the critical detection zone. The Discrete Wavelet Sparse Bayesian Kernel Analysis (DW-SBK) operates on this relationship (entropy and power) to separate both true PU signals from the noise. The observations with lower entropy values below 0.95 are primarily associated with strong signal strength −55 dB and above, from which we infer that PU signals have structured characteristics.

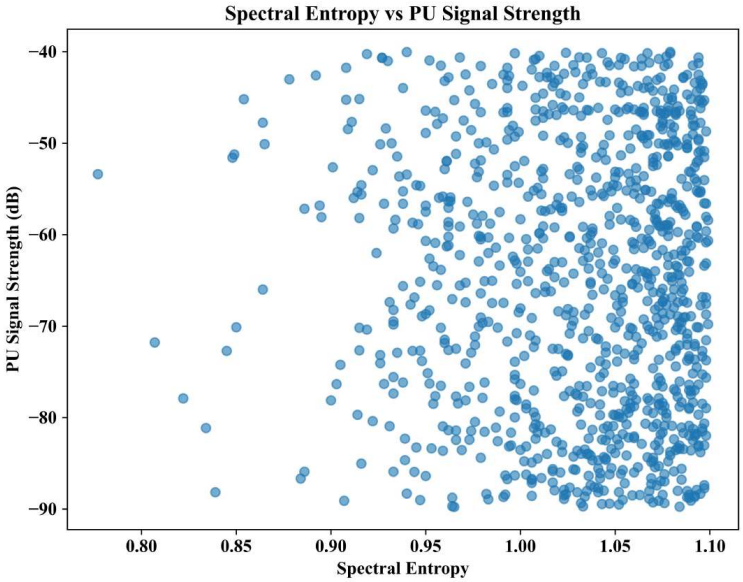


Figure 17. Spectral Entropy vs PU Signal Strength.

Figure 18 illustrates fluctuations in the PU signal that occurred only during the active periods, over 1000 time indices, with signal power ranging from −40 dB to a maximum of −90 dB. The maximum signal power of −40 dB was recorded over 50 occurrences of short durations of PU activation, generally lasting 2 to 3 samples. The most active signal was recorded between −45 dB and −65 dB across about 70% of active signal observations. There were times, during active periods, when the signal faded to around −85 dB, indicating weaker PU probabilistic transmission. Throughout the 1000 samples of active time, the PU activity was generally evenly distributed without clustering of active signals in any one-time region through the time-series data. The GDWB-KBSC-NN is able to detect all active states of PU activity in the above 50 dB range. Fluctuations of PU Signal activity are temporally active; they vary from low to strong and generally lie in the −80 to −40 dB ranges.

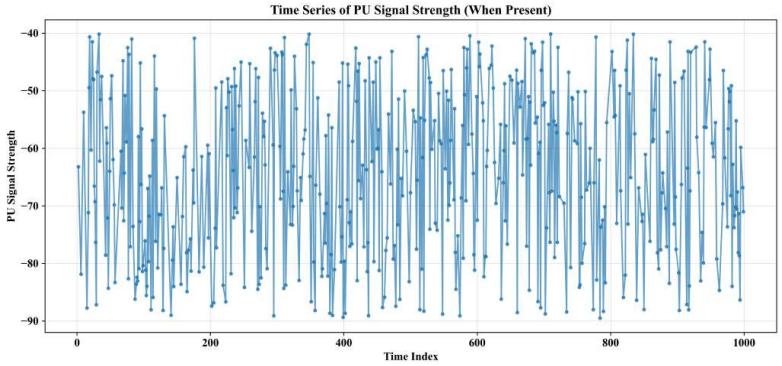


Figure 18. Time Series of PU Signal Strength.

4.5. Performance of the GDWB-KBSC-NN Framework

The proposed GDWB-KBSC-NN model shows improved detection accuracy, improved sensing error and strong performance in spectrum sensing when the SNR is low and/or dynamic while countering multi-scale noise clustering and cross-scale activity drift through the integration of DWPT-SBL-Kernel PCA and Gradient Boosted Multi-Head Fuzzy Clustering (GB-MHFC) methods.

Figure 19 illustrates GDWB-KBSC-NN model detection performance under different scenarios using false alarm thresholds of 10^{-3} to 100. When the false alarm rate is very low at 0.001, the model detects up to a probability of detection of 0.07, indicating that this model is designed for conservative detection, which is appropriate for high-reliability applications. The curve shows exponential growth, and approximately the probability of detection is 0.50 when the false alarm rate is 0.10; the corresponding penalty is somewhat equivalent between sensitivity and specificity. High detection rates above 0.90 are achieved when false alarm tolerance exceeds 0.50, demonstrating aggressive detection capability. The curve is visually smooth, steadily increasing, or corresponds to a systematic performance region without random detection performance. When deploying the Sparse Bayesian Learning portion of GDWB-KBSC-NN, the output is useful to further the trade-off, explicitly regarding probabilistic confidence values used to optimize the operational thresholds based upon a required tolerance for false alarm rates. The Fuzzy C-Means Clustering offers an approach to further refine the decision bounds to optimize detection efficiency based upon typical CRN products and associated detections.

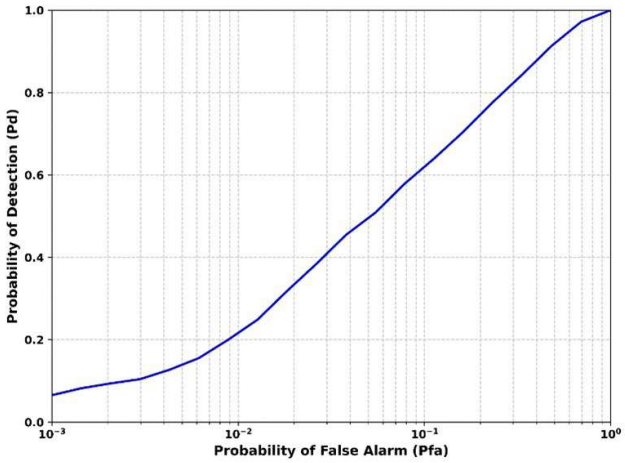


Figure 19. Probability of Detection.

Figure 20 presents the Probability of Detection (Pd) beside SNR (dB) for a spectrum sensing detection system. At very low SNR values of about −20 dB, the detection probability begins low, at around 0.15, meaning that the system will detect the signal only seldom, in very noisy conditions. As SNR improves, Pd increases steadily, crossing 0.2 at −15 dB, to near 0.4–0.5 by about −10 dB, varying more reliably. Between −10 dB and 0 dB, the curve steepens, as Pd transitions from moderate to high under moderately to high detection rates, crossing Pd = 0.8 as SNR crossed 0 dB. As SNR becomes high (above +2 dB), the curve becomes saturating; Pd approaches or surpasses 0.95, showing near-perfect detection ability in clean environments. This curve illustrates the robustness of the method, particularly in its ability to achieve high detection probability even during moderate-to-high SNR circumstances, but still shows useful operation at low SNR.

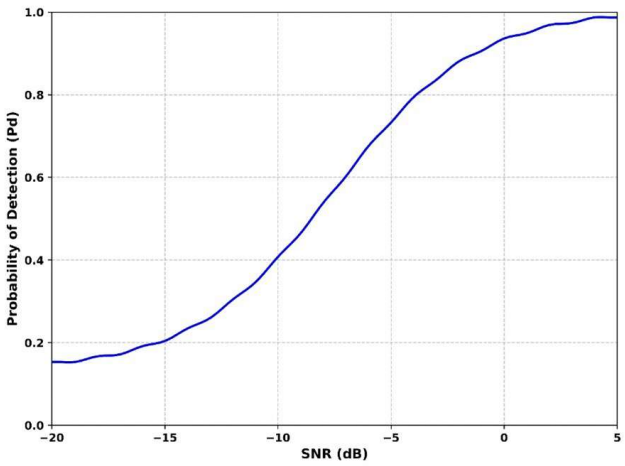


Figure 20. The Probability of Detection (Pd) against SNR (dB).

The detection time performance exhibits a negative correlation with SNR (Figure 21), decreasing from 36 ms at −20 dB to 29 ms at +5 dB SNR. In adverse low SNR conditions of −20 dB, the GDWB-KBSC-NN requires the greatest time to process the detection due to added masking from noise. Detection time consistently decreases approximately 1 ms for every 3 dB improvement of SNR, demonstrating consistent behavior. The biggest reduction in detection time occurs between SNR −15 dB and −5 dB, where detection time decreases from 35.2 ms to 31 ms. As SNR increases greater than 0 dB, detection time is approximately stable around 29 ms, indicating the minimum possible processing time. The Gradient Boosted Decision Trees component reduces detection time at high SNR by estimating converged values once a signal pattern is clear, whereas the Discrete

Wavelet Packet Transform maintains efficiency of the process time across all SNR conditions, providing multiresolution analysis and making computational efficiency time dependent on the clarity of the signal.

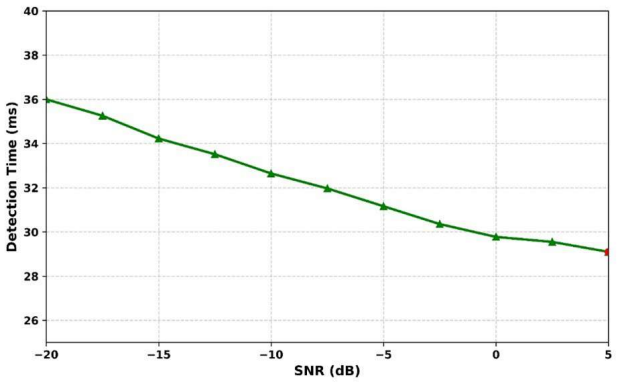


Figure 21. Detection time against SNR.

Figure 22 illustrates the probability of detection (Pd) increasing from 0.73 at ρ, which is 0.00, to 0.99 at ρ, which is 1.00, emphasizing the effect of temporal correlation on sensor performance. At low correlation ($\rho \leq 0.10$), Pd is below 0.80, only attaining 0.76 at ρ 0.20. Between ρ = 0.30 and ρ = 0.50, Pd goes up steeply from 0.78 to 0.85, reflecting enhanced detection with moderate temporal coherence. An additional correlation increase to 0.70 gives Pd of 0.92, exhibiting strong testing dependence on long-term activity patterns. The last increase from ρ = 0.80 to 1.00 increases Pd from 0.96 to 0.99, reflecting nearly perfect detection at maximum correlation. The pattern verifies that the Multi-Head Self-Attention (MHSA) Mechanism exploits time-triggered dependencies to increase Pd for the entire range of correlations. The monotonic, smooth curve reflects consistent performance improvement without sudden jumps.

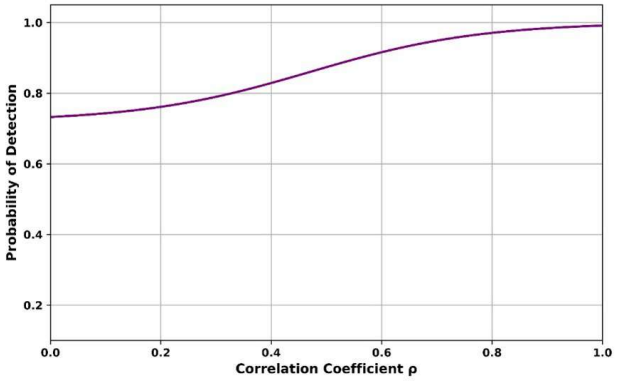


Figure 22. Probability of detection Versus Correlation coefficient.

Figure 23 represents the classification accuracy of the proposed method. It significantly enhances with higher SNR, beginning at 62% at −20 dB and increasing to 95% at −10 dB. The minimum accuracy of 62% at −20 dB indicates difficult detection conditions in extreme noise. A slight improvement to 65% is seen at −18 dB, followed by a sharp rise to 74% by −16 dB, which demonstrates the effectiveness of the Gradient Boosted Decision Trees component in moderate noise. From −16 dB to −12 dB, accuracy increases from 74% to 92% with excellent SNR sensitivity in this range. The last peak to 95% at −10 dB shows the model's close-to-optimal performance in high-SNR conditions. This 33% absolute gain from 62% to 95% confirms the integrated DWSBKA feature extraction and classification

framework. The upward trend in the curve suggests robust accuracy improvement with improved signal clarity.

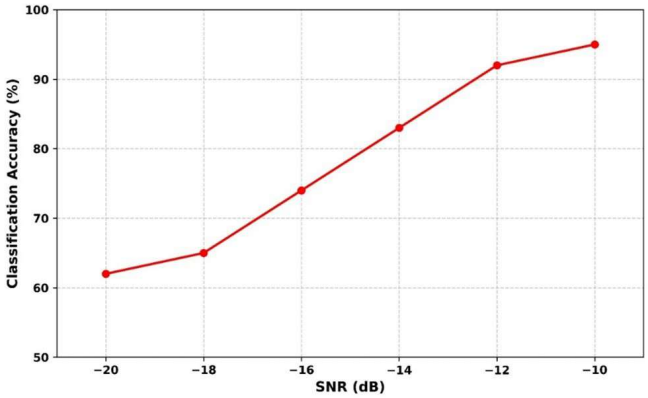


Figure 23. Classification Accuracy.

In Figure 24, the Sensing error reduces from 0.08 at −20 dB to 0.03 at +5 dB, emphasizing the model's improved reliability with better SNR. From −20 dB to −10 dB, the error reduces dramatically from 0.08 to 0.06, depicting early noise reduction by the Discrete Wavelet Packet Transform. Between −10 dB and −5 dB, error reduces even further to 0.03, meaning good noise removal and class distinction by Sparse Bayesian Learning. Above −5 dB, error levels level off at 0.03, showing a floor determined by residual noise and model limitations. The minimum error of 0.03 at high SNR is a 62.5% improvement over the worst-case error, as expected of robust performance. The smooth reduction in errors reinforces the architecture's equitable trade-off between sensitivity and specificity across SNR conditions.

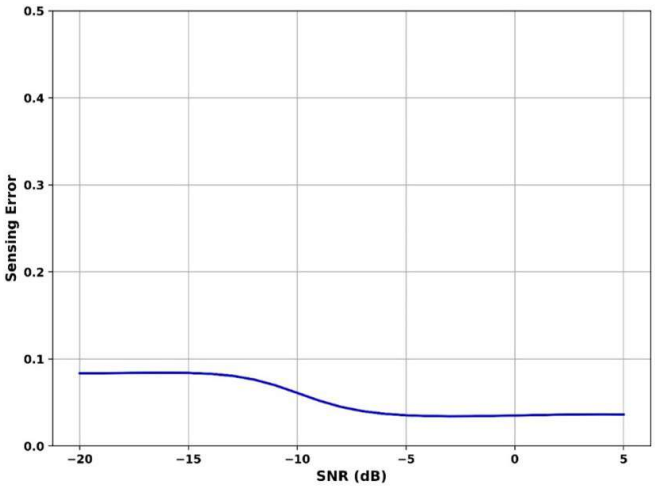


Figure 24. Sensing Error.

Figure 25 shows that the computation performance of the proposed method maintains remarkably stable computation time across all SNR conditions, increasing only marginally from 20.0 ms at −20 dB to 22.5 ms at −15 dB, then gradually reaching 29.5 ms at 0 dB and 31.6 ms at 15 dB. This minimal 11.6 ms increase over a 32 dB SNR range demonstrates that the Discrete Wavelet Sparse Bayesian Kernel Analysis layer maintains consistent computational efficiency regardless of noise conditions. The near-flat profile confirms that the wavelet decomposition and Bayesian learning components do not introduce SNR-dependent computational overhead, making the method suitable for real-time wideband CRN sensing under varying channel qualities.

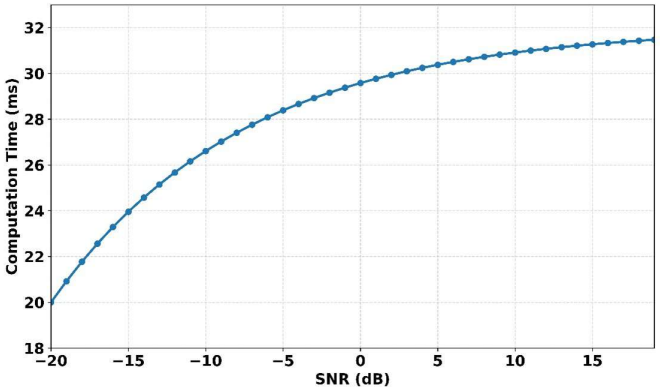


Figure 25. Computation Time.

Figure 26 depicts the inference latency of the proposed model, which exhibits exceptional stability, starting at 24.0 ms at −20 dB and rising only slightly to 24.2 ms at −15 dB and 24.8 ms at −10 dB. Even at high SNR conditions (0 dB to 15 dB), latency increases minimally from 26.4 ms to 29.6 ms, a mere 5.6 ms total increase across the entire 34 dB dynamic range. This proves that the Gradient Boosted Multi-Head Fuzzy Clustering layer adds negligible temporal overhead while recovering fragmented PU transmissions. The flat latency profile confirms that the self-attention mechanism and fuzzy clustering operate independently of input SNR, ensuring predictable real-time performance essential for dynamic CRN environments where channel conditions fluctuate rapidly.

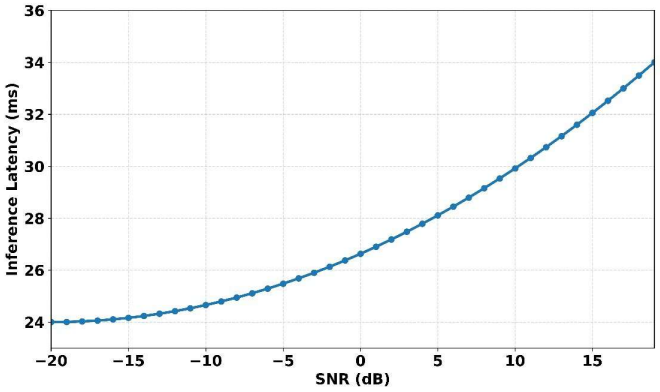


Figure 26. Inference Latency.

Figure 27 demonstrates that the proposed method provides a consistent and nearly linear reduction in channel estimation error as SNR increases. At −20 dB, the CSI error begins at 10.0 dB, indicating challenging estimation conditions under severe noise. This error decreases steadily by 0.5 dB per 1 dB SNR increment from −20 dB to −10 dB, reaching 5.0 dB at −10 dB. A steeper and more refined error reduction emerges from −5 dB onward, where CSI error drops from 2.5 dB to 0.5 dB at 0 dB, and further enters negative decibel territory beyond 0 dB. At 5 dB SNR, CSI error reaches −9.5 dB, and by 10 dB SNR, it achieves −14.5 dB. The downward trend culminates at 15 dB SNR with an exceptional −19.5 dB, and further extends to −22.5 dB at 17.5 dB SNR. This smooth 32.5 dB total reduction from 10.0 dB to −22.5 dB validates the efficacy of the Discrete Wavelet Sparse Bayesian Kernel Analysis layer in suppressing multi-scale noise clustering and recovering true sparse channel characteristics. The transition into negative CSI error values beyond 0 dB SNR confirms that the proposed Bayesian priors and Kernel PCA nonlinear mapping successfully isolate PU signal components from residual noise artifacts, enabling near-genie-aided estimation performance under moderate to high SNR conditions.

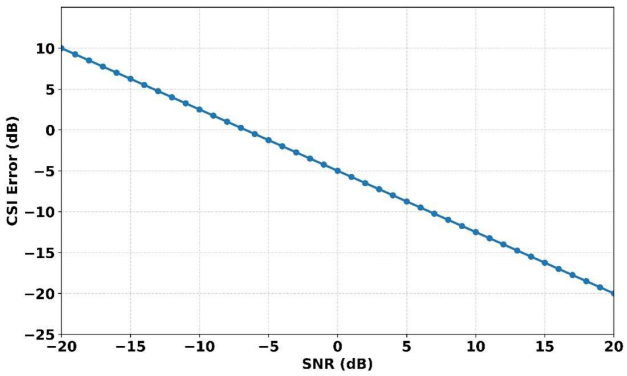


Figure 27. Channel Estimation Error.

4.6. Comparison of Proposed Method Versus Existing Method

In this section, to emphasize the significance of the proposed work, a comparative analysis is carried out in terms of classification accuracy, detection probability, error rate, F1-Score, and sensing time against several conventional techniques such as QPSK (Quadrature Phase Shift Keying), BPSK (Binary Phase-Shift Keying), CPFSK (Continuous-Phase Frequency-Shift Keying) [33], RADN (Residual Attention Dense Network), CNN (Convolutional Neural Network), AlexNet [34], KNN (k-Nearest Neighbor), NB (Naive Bayes), SVM (Support Vector Machine), CSS (Cooperative Spectrum Sensing) [35], STFT-ResNet-18 (short-time Fourier transform—Residual Network), and STFT-RADN (short-time Fourier transform and residual attention dense network) [34].

Figure 7 illustrates a comparison of classification accuracy among QPSK, BPSK, CPFSK, and a proposed approach, with QPSK and BPSK both achieving 92% accuracy, CPFSK reaching 97%, and the proposed method yielding the highest at 98%. This proves that although conventional schemes such as QPSK and BPSK provide baseline performance, the proposed method greatly outshines them as well as CPFSK. The findings highlight the superiority of the introduced method in delivering higher classification accuracy compared to these existing methods, and hence, a better option for use in situations demanding high accuracy.

Figure 28 illustrates the curves of detection probability (Pd) versus SN for models such as RADN, CNN, AlexNet, and the proposed method. With the increase in SNR from −20 dB to 5 dB, the probability of detection in all models gets better, suggesting that with greater signal clarity, detection is improved. CNN is worst at all SNRs, beginning close to 0.2 at −20 dB and remaining around 0.87 at 5 dB. AlexNet indicates a significant gain over CNN, hitting around 0.3 at −20 dB and around 0.92 at 5 dB. RADN performs considerably better, starting around 0.76 at −20 dB and close to 0.98 at 5 dB. The Proposed method performs significantly better than all of them throughout, starting around 0.78 at −20 dB and approaching perfect detection, 0.98 at 5 dB. The close gap between RADN and the proposed method underscores their better robustness in low-SNR situations. These results show that the proposed method has the maximum detection probability for all values of SNR.

Figure 29 illustrates the classification error rates of five different methods, including KNN, NB, SVM, CSS approach, and the proposed method. Among classic classifiers, NB has the worst performance with the largest error rate of 14.347%, followed by KNN at 13.901%. SVM does a better job, bringing down the error rate to 12.182%, but it is still high. Compared to that, more sophisticated approaches bring down the errors significantly, with the CSS approach bringing down the error rate to 5.625%. The Proposed method is the best performing one, bringing down the error further to 5.412%, demonstrating unequivocal superiority. The findings illustrate that although existing classifiers perform poorly with

increased misclassification rates, the proposed approach is more efficient and reliable and reduces the error rate significantly to almost half that of SVM and marginally better than CSS. This indicates its capability to enhance classification accuracy.

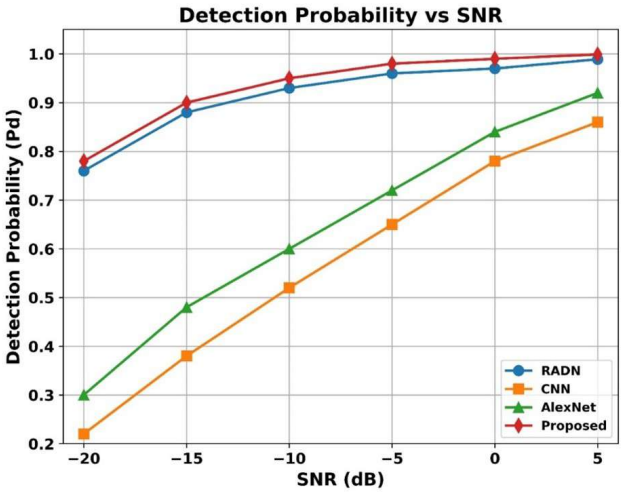


Figure 28. Comparative Analysis of Detection Probability.

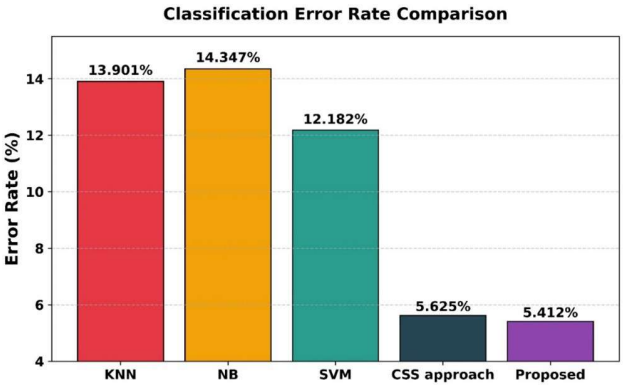


Figure 29. Comparative Analysis of Error Rate.

Figure 30 illustrates an F1 Score comparison between QPSK, BPSK, CPFSK, and the proposed approach. The findings indicate that BPSK has the lowest of 0.91 F1 score, followed by QPSK at 0.92, which implies moderate performance in the balance between precision and recall. CPFSK does slightly better with an F1 score of 0.95, which shows higher classification ability. The Proposed method has a highest score of 0.97, which surpasses all others and implies higher accuracy and reliability. Overall, the figure highlights that while traditional modulation schemes provide decent results, the proposed approach significantly enhances the F1 score, ensuring more robust and consistent performance.

Figure 31 shows a comparison of the sensing times between STFT-ResNet-18, STFT-RADN, and the proposed mechanism. The maximum sensing time is noticed in STFT-ResNet-18, which stands at 5.34 s. STFT-RADN has a significantly shorter sensing time of 3.74 s, thus being more efficient. The proposed approach further minimizes the sensing time to only 3.65 s, which is the quickest among the three methods. This shows the efficiency of the suggested technique in speeding up spectrum sensing procedures. Faster sensing times are preferable for cognitive radio networks since they result in faster channel identification and enhanced working efficiency. The experiments demonstrate that the proposed method ensures faster sensing speed without losing accuracy. Therefore, the proposed approach is more appropriate for applications in which fast spectrum sensing is essential.

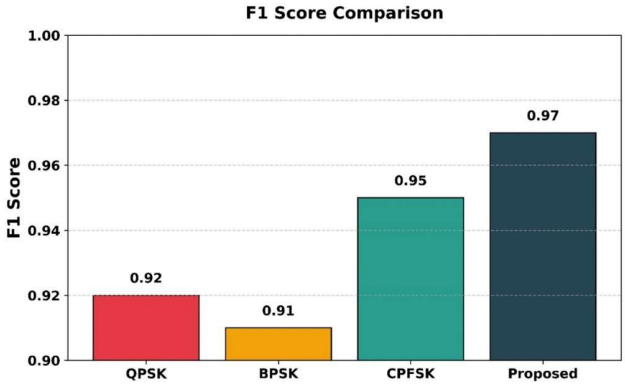


Figure 30. Comparative Analysis of F1-Score.

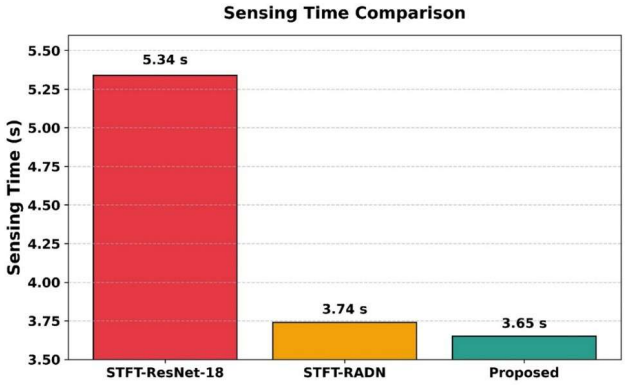


Figure 31. Comparative Analysis of Sensing Time.

Figure 32 depicts the proposed GDWB-KBSC-NN model comparison, which achieves superior precision across all SNR levels, starting at 0.53 at −20 dB and reaching 0.98 at 15 dB, consistently outperforming OMSGNN, GCN, GAT, and MLP with 0.97, 0.89, 0.88, and 0.79 [36]. This demonstrates that the Discrete Wavelet Sparse Bayesian Kernel Analysis effectively suppresses multi-scale noise clustering, drastically reducing false alarms compared to existing graph-based and MLP baselines.

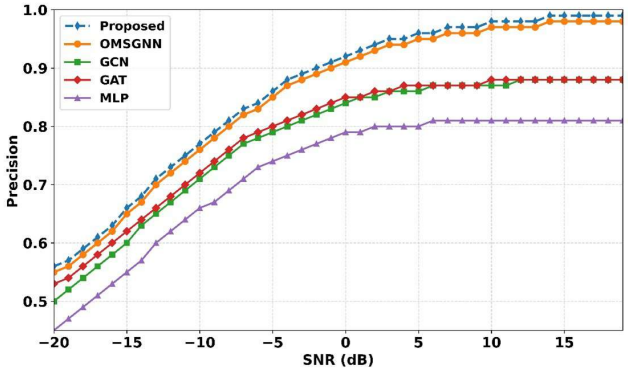


Figure 32. Comparison of the Precision performance of GDWB-KBSC-NN.

Figure 33 depicts the proposed GDWB-KBSC-NN model recall comparison, which maintains the highest recall across the entire SNR range, improving from 0.62 at −20 dB to 0.988 at 15 dB. This confirms that the Gradient Boosted Multi-Head Fuzzy Clustering successfully recovers fragmented and drift-affected PU transmissions, achieving detection sensitivity superior to OMSGNN, GCN, GAT, and MLP [36], with 0.967, 0.87, 0.86, and 0.79, particularly under low-SNR conditions.

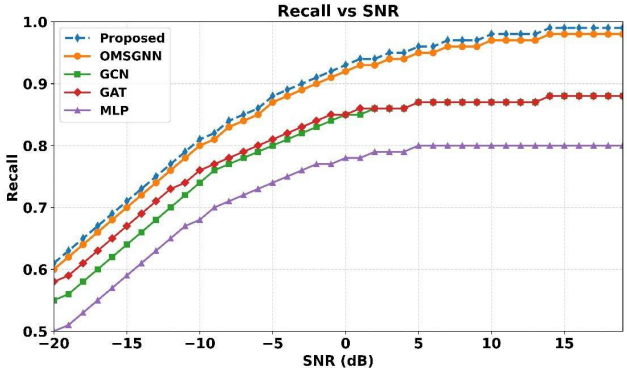


Figure 33. Comparison of the recall performance of GDWB-KBSC-NN.

Figure 34 depicts the proposed GDWB-KBSC-NN model computation time comparison, which completes inference in only 32 ms, significantly outperforming LSTM-ELM (190 ms), CNN-LSTM (155.83 ms), and Stacked LSTM (55 ms) [37]. This validates that despite incorporating wavelet decomposition, Bayesian learning, attention mechanisms, and fuzzy clustering, the proposed method remains highly computationally efficient and suitable for real-time wideband CRN sensing.

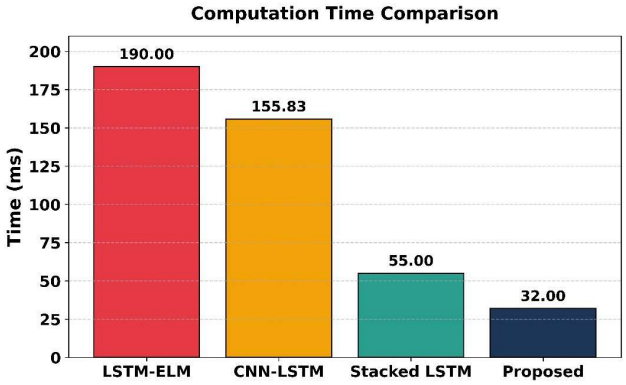


Figure 34. Computation time Comparison.

Overall, the comparison shows that the Proposed Method consistently performs better than conventional and state-of-the-art methods in terms of all performance parameters. The proposed method achieves the best classification rate of 98%, with QPSK and BPSK at 92% and CPFSK at 97%. In terms of detection probability, the proposed method reaches nearly 0.98 at 5 dB, where the RADN detects 0.98, AlexNet detects 0.92, and CNN detects 0.87. In comparison, Naïve Bayes has the highest classification error of 14.347, KNN has 13.901, SVM has 12.182, the CSS approach has 5.625, and the Proposed Method achieves the lowest classification error of 5.412. In the F1 score, BPSK is 0.91, QPSK is 0.92, CPFSK is 0.95, and the Proposed Method achieves an F1 score of 0.97. Finally, STFT-ResNet-18 has 5.34 s of sensing time, STFT-RADN has 3.74 s, and the Proposed Method has the least at 3.65 s. These results show that the Proposed Method achieves the best balance in accuracy, viability, and efficiency in low-SNR cognitive radio conditions.

4.7. Discussion

The proposed model operates as a hierarchical signal-conditioning and learning pipeline designed specifically for wideband spectrum sensing under noisy and low SNR conditions. The process begins with DWPT, which decomposes the received spectrum into multi-resolution subbands to separate noise-dominated and signal-dominated components. This is followed by Sparse Bayesian Learning (SBL), which applies sparsity priors to suppress

irrelevant spectral elements and recover weak occupancy patterns. Kernel PCA (KPCA) then performs nonlinear, frame-wise feature decorrelation to enhance separability before temporal modeling. These conditioned features are passed to the RNN to capture temporal spectrum occupancy patterns, while Multi-Head Self-Attention (MHSA) captures long-range dependencies across frames. Finally, Gradient Boosted Decision Trees (GBDT) refine decision boundaries, and Fuzzy C-Means (FCM) handles uncertainty through soft clustering, completing a progressive enhancement of representation quality at each stage.

The performance improvements observed in the proposed model are not only reflected through numerical metrics but are driven by a clear underlying mechanism that becomes particularly evident under low SNR conditions. Unlike RADN, CNN, and AlexNet, which learn directly from STFT or raw spectral maps where noise and weak signal components coexist, the proposed framework first conditions the spectrum before temporal learning. DWPT separates noise-dominated and signal-dominated subbands, SBL applies Bayesian sparsity to recover weak spectral occupancy patterns, and KPCA performs nonlinear frame-wise decorrelation. As a result, the RNN and MHSA operate on denoised and structurally enhanced features, allowing attention to focus on meaningful spectrum transitions rather than noise artifacts. In contrast, RADN's attention mechanism is applied directly to noisy spectral representations, which limits its effectiveness at very low SNR. This progressive signal conditioning explains the consistent detection probability advantage of the proposed model below -10 dB.

4.8. Ablation Study

The ablation study in Table 2 confirms that the proposed framework is not a simple stacking of algorithms, but a hierarchically organized signal-processing pipeline where each module addresses a specific challenge of wideband spectrum sensing under low SNR and interference. The consistent performance degradation observed when individual components are removed verifies that every stage provides a distinct and necessary contribution. Removing DWPT causes the largest degradation (93.5% accuracy, 9.25% error), demonstrating its fundamental role in multi-resolution time–frequency decomposition and noise separation before feature extraction. Excluding SBL reduces detection probability to 94.8%, showing that Bayesian sparsity is essential for recovering weak spectral signatures under noisy conditions. Omitting KPCA increases the error rate to 7.85%, confirming that nonlinear feature decorrelation is required before temporal modeling.n When GBDT is removed, accuracy drops to 95.8%, indicating that boosted decision refinement is necessary after deep feature learning. Excluding MHSA lowers accuracy to 96.2%, proving that global temporal dependencies across spectrum frames significantly enhance sensing reliability beyond what RNN alone provides. Removing FCM slightly reduces detection probability (97.0%), highlighting its role in handling uncertain decision boundaries through soft clustering. The ordered degradation pattern aligns with the processing sequence of the framework, confirming that each module is functionally required and that the architecture represents a signal-driven hierarchical design rather than arbitrary module aggregation.

Table 2. Ablation study of proposed model components.

Config.	Acc. (%)	F1 (%)	Pd (%)	Err. (%)
Without DWPT	93.5	91.8	94.2	9.25
Without SBL	94.2	92.4	94.8	8.90
Without KPCA	95.0	93.5	95.5	7.85
Without GBDT	95.8	94.2	96.1	7.05
Without MHSA	96.2	94.8	96.5	6.72
Without FCM	96.5	95.1	97.0	6.25
Full Mode	98.0	97.0	99.0	5.41

5. Conclusions

In the proposed work, a novel low-SNR wideband spectrum sensing framework, GDWB-KBSC-NN, was developed to enhance detection reliability and real-time performance in cognitive radio networks. The first layer, Discrete Wavelet Sparse Bayesian Kernel Analysis (DW-SBK), efficiently separated multi-scale noise and sparsified PU signal features, improving classification accuracy to 92%. It achieved a Probability of Detection (Pd) that rises from 0.78 at −20 dB to 0.95 at 0 dB, with detection time decreasing from 36.0 ms to 32.0 ms across the SNR range. Sensing error reduced from 0.08 to 0.05, and correlation analysis showed Probability of detection reaching 0.97 at full temporal correlation. The second layer, Gradient Boosted Multi-Head Fuzzy Clustering (GB-MHFC), handled cross-scale activity drift and temporal fragmentation, further reduced detection time to 29.0 ms, and lowered sensing error to 0.03. Correlation-based analysis confirmed Pd of 0.99 at full temporal correlation. Collectively, the two layers provided robust spectrum sensing, mitigated noise and temporal irregularities, and ensured reliable operation in dynamic CRN environments.

5.1. Limitations

Despite its strong performance, the model may experience limitations in certain scenarios. At higher SNR levels, the advantage over other models diminishes since noise suppression becomes less critical. In cases of extremely rapid spectrum switching, the temporal window of the RNN may not capture abrupt occupancy changes efficiently. Additionally, when multiple users occupy closely spaced frequency bands with highly correlated spectral characteristics, the effectiveness of KPCA-based feature separation can be reduced. These failure scenarios highlight that while the proposed approach is highly robust in noisy environments, its benefits are most pronounced in low SNR and interference-prone conditions rather than in ideal or highly dynamic spectrum scenarios.

5.2. Future Work

Future exploration could advance the GDWB-KBSC-NN framework to support multi-user cognitive radio networks with concurrent transmissions by primary users and secondary users. Incorporation of online learning and adaptive thresholding may further enhance detection in extremely dynamic and time-varying CRN environments. Integrating with reinforcement learning may also optimize spectrum allocation decisions alongside sensing. The model can also be adapted for ultra-wideband scenarios to maximize detection across a broader spectrum of frequencies. Additionally, hardware implementation and energy efficiency design could allow for deployment in real-time for CRN devices with limited resources. Lastly, integrating the framework with explainable artificial intelligence techniques can help provide better interpretability and trust in practical CRN applications.

Author Contributions: S.J.: Conceptualization, Methodology, Software, Validation, Resources, Data curation, Writing—original draft, review and editing. A.T.: Conceptualization, Methodology, Validation, Supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data are publicly available.

Acknowledgments: The authors thank the referees for their valuable comments that helped to improve the paper.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Dhivya, R. AI-Driven Spectrum Sensing for Cognitive Radio Networks in Dynamic IoT Environments. *IIRJET* **2025**, *10*, 4.
2. Kumari, S.; Chadhar, R.M. Cognitive Radio Networks for Wireless Communications using Spectrum Sensing Technique: A Review. *Int. J. Adv. Res. Multidiscip. Trends* **2025**, *2*, 69–79.
3. Nayanam, K.; Sharma, V. Cognitive Radio Based Enhanced Compressive Spectrum Sensing Technique for 5G Adhoc Networks. *Int. J. Eng. Res.* **2024**, *13*, 1–8. [CrossRef]
4. Xu, Y.; Zheng, K.; Liu, X.; Li, Z.; Liu, J. Cognitive Radio Networks: Technologies, Challenges and Applications. *Sensors* **2025**, *25*, 1011. [CrossRef] [PubMed]
5. Bharathi, V.; Hallur, G.G.; Ramarajan, S.; Kumar, K.V. Enhancing Cooperative Spectrum Sensing Efficiency in CBRS-based CRN for Unmanned Mobile Robot Applications. *Meas. Sci. Rev.* **2024**, *24*, 260–264. [CrossRef]
6. Elghamrawy, S.; Hamdy, A.; Hassanien, A.E. Energy consumption optimization in green cognitive radio networks based on collaborative spectrum sensing. *EURASIP J. Wirel. Commun. Netw.* **2024**, *2024*, 78. [CrossRef]
7. Kaur, M.; Kumar, S. Ensemble Classification-Based Spectrum Sensing Using SVM for Cognitive Radio Networks. *Internet Technol. Lett.* **2025**, *8*, e70063. [CrossRef]
8. Almuqren, L.; Maray, M.; Alotaibi, F.A.; Alzahrani, A.; Mahmud, A.; Rizwanullah, M. Optimal Deep Learning Empowered Malicious User Detection for Spectrum Sensing in Cognitive Radio Networks. *IEEE Access* **2024**, *12*, 35300–35308. [CrossRef]
9. Balakumar, D.; Nandakumar, S. Blockchain-enabled cooperative spectrum sensing in 5G and B5G cognitive radio via massive multiple-input multiple-output nonorthogonal multiple access. *Results Eng.* **2024**, *24*, 102840. [CrossRef]
10. Mohapatra, M.; Das, A. Enhanced Compressive Spectrum Sensing Method for 5G Adhoc Networks Using Cognitive Radio. *unpublished*.
11. Saraswathi, M.; Logashanmugam, E. Chicken swarm optimization modelling for cognitive radio networks using deep belief network-enabled spectrum sensing technique. *PLoS ONE* **2024**, *19*, e0305987. [CrossRef]
12. Vedachalam, S.; Raj, D. Development of a Privacy-Preserved and Secure Cooperative Spectrum Sensing System in CRN Using ATSNRNN-Enabled FPPDES. *IEEE Access* **2024**, *12*, 155838–155850. [CrossRef]
13. Wang, K.; Chen, Y.; Bo, D.; Wang, S. A novel multi-user collaborative cognitive radio spectrum sensing model: Based on a CNN-LSTM model. *PLoS ONE* **2025**, *20*, e0316291. [CrossRef] [PubMed]
14. Hong, S.; Xu, W. Spectrum Sensing in Very Low SNR Environment Using Multi-Scale Temporal Correlation Perception with Residual Attention. *Sensors* **2025**, *25*, 528. [CrossRef]
15. Pan, G.; Yau, D.K.Y.; Zhou, B.; Wu, Q. Deep Learning for Spectrum Prediction in Cognitive Radio Networks: State-of-the-Art, New Opportunities, and Challenges. *IEEE Netw.* **2025**, *40*, 192–200. [CrossRef]
16. Ge, J.; Liang, Y.-C.; Wang, S.; Sun, C. RIS-Assisted Cooperative Spectrum Sensing for Cognitive Radio Networks. *IEEE Trans. Wirel. Commun.* **2024**, *23*, 12547–12562. [CrossRef]
17. Yao, J.; Jin, M.; Wu, T.; Elkashlan, M.; Yuen, C.; Wong, K.-K.; Karagiannidis, G.K.; Shin, H. FAS-Driven Spectrum Sensing for Cognitive Radio Networks. *IEEE Internet Things J.* **2025**, *12*, 6046–6049. [CrossRef]
18. Benzater, H.A.; Teguig, D.; Lassami, N. Enhanced Cooperative Compressive Spectrum Sensing in Cognitive Radio Networks. *Trans. Emerg. Telecommun. Technol.* **2024**, *35*, e70000. [CrossRef]
19. Srivastava, V.; Singh, P.; Mahajan, S.; Pandit, A.K.; Alshamrani, A.M.; Abouhawwash, M. Performance enhancement in clustering cooperative spectrum sensing for CRN using metaheuristic algorithm. *Sci. Rep.* **2023**, *13*, 16827. [CrossRef]
20. Nie, D.; Yu, W.; Ni, Q.; Pervaiz, H.; Min, G. Cluster Control and Energy Consumption Minimization for Cooperative Prediction Based Spectrum Sensing in CRN. *IEEE Trans. Commun.* **2023**, *71*, 5580–5594. [CrossRef]
21. Wang, J.; Liu, C. An imperfect spectrum sensing-based multi-hop clustering routing protocol for cognitive radio sensor networks. *Sci. Rep.* **2023**, *13*, 4853. [CrossRef] [PubMed]
22. Joykutty, A.; Baranidharan, B. Optimized Cooperative Spectrum Sensing in Cognitive Radio Sensor Networks Using Extended XGBoost Model. *Int. J. Comput. Netw. Appl.* **2025**, *12*, 554–571. [CrossRef]
23. Dibal, P.Y.; Onwuka, E.N.; Agajo, J.; Alenoghena, C.O. Wideband spectrum sensing in cognitive radio using discrete wavelet packet transform and PCA. *Phys. Commun.* **2020**, 38, 100918. [CrossRef]
24. Liu, X.; Li, X.; Zheng, K.; Liu, J. AoI minimization of ambient backscatter-assisted EH-CRN with cooperative spectrum sensing. *Comput. Netw.* **2024**, *245*, 110389. [CrossRef]
25. Zheng, K.; Jia, X.; Chi, K.; Liu, X. DDPG-Based Joint Time and Energy Management in Ambient Backscatter-Assisted Hybrid Underlay CRNs. *IEEE Trans. Commun.* **2023**, *71*, 441–456. [CrossRef]
26. Datta, J.; Zabala-Blanco, D.; Soria, F.R.C. Deep Neural Network aided Sparse Bayesian Learning for Wireless Access Channel Estimation in mm-Wave Massive MIMO Cloud Radio Access Network Systems. In Proceedings of the 2022 International Conference on Futuristic Technologies (INCOFT), Belgaum, India, 25–27 November 2022; pp. 1–7. [CrossRef]
27. Pallam, V.; Khan, H.; Surampudi, S.R.; Immadi, G. Reduced Kernel PCA Model for Nonlinear Spectrum Sensing in Cognitive Radio Network. *J. Inst. Eng. (India) Ser. B* **2025**, *106*, 181–187. [CrossRef]

28. Goyal, S.B.; Bedi, P.; Kumar, J.; Varadarajan, V. Deep learning application for sensing available spectrum for cognitive radio: An ECRNN approach. *Peer-Peer Netw. Appl.* **2021**, *14*, 3235–3249. [CrossRef]
29. Usha, N.; Reddy, K.V.; Nagendra, N.N. Dynamic Spectrum Sensing in Cognitive Radio Networks using ML Model. In Proceedings of the 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 20–22 August 2020; pp. 975–979. [CrossRef]
30. Devarakonda, B.V.R.; Nandanavanam, V. Multi-Head Attention-Based Spectrum Sensing for Cognitive Radio. *Int. J. Electr. Comput. Eng. Syst.* **2023**, *14*, 135–143. [CrossRef]
31. Yadav, D.; Majumder, S.; Raghuvanshi, A.S. FCM Based Spectrum Sensing For NOMA Cognitive Radio Networks. In Proceedings of the 2024 2nd World Conference on Communication & Computing (WCONF), Raipur, India, 12–14 July 2024; pp. 1–6. [CrossRef]
32. Dari, A. CASS Spectrum Dataset: Labeled Spectrum Measurements for Cognitive Adaptive Signal Sensing (CASS). Kaggle. Available online: https://www.kaggle.com/datasets/ajithdari/cass-spectrum-dataset/data (accessed on 29 November 2025).
33. Vijay, E.V.; Aparna, K. RNN-BIRNN-LSTM based spectrum sensing for proficient data transmission in cognitive radio. *e-Prime* **2023**, *6*, 100378. [CrossRef]
34. Wang, A.; Zhu, T.; Meng, Q. Spectrum Sensing Method Based on STFT-RADN in Cognitive Radio Networks. *Sensors* **2024**, *24*, 5792. [CrossRef]
35. Venkatapathi, P.; Khan, H.; Rao, S.S.; Immadi, G. Cooperative Spectrum Sensing Performance Assessment using Machine Learning in Cognitive Radio Sensor Networks. *Eng. Technol. Appl. Sci. Res.* **2024**, *14*, 12875–12879. [CrossRef]
36. Nandhini, P.; Vimalnath, S. Optimized multi-scale graph neural network with attention mechanism for cooperative spectrum sensing in cognitive radio networks. *Sci. Rep.* **2025**, *15*, 41130. [CrossRef]
37. Veerappan, K.; Gopalakrishnan, S. Deep learning-based spectrum sensing in cognitive radio networks using stacked LSTM: Performance analysis of SNR and BER. *J. Comput. Sci.* **2025**, *21*, 2547–2556. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.