International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online)) Volume 12, Issue 12, December 2023)

Deep Learning-Based Signal Modulation Classification for Cognitive Radio Networks Using CNN-BiLSTM

Ajay Sonkesriya

Lecturer, Electronics and Communication Engineering, Kalaniketan Polytechnic College Jabalpur, MP, India

Abstract-- **Cognitive radio (CR) technology has emerged as a transformative paradigm for resolving the growing spectrum scarcity problem in modern wireless communication systems. At the core of every cognitive radio implementation is the ability to correctly identify the modulation scheme of an incoming signal, a task known as Automatic Modulation Classification (AMC). Accurate AMC enables the secondary user to make intelligent decisions about spectrum access, interference avoidance, and adaptive transmission without any cooperation from the primary user. Classical AMC methods based on maximum likelihood estimation and hand-crafted statistical features exhibit satisfactory performance only under high signal-to-noise ratio conditions and degrade sharply under multipath fading, frequency offsets, and low SNR regimes common in real cognitive radio environments. This paper proposes a deep learning architecture, CNN-BiLSTM, that integrates a Convolutional Neural Network for local feature extraction with a Bidirectional Long Short-Term Memory network for sequential dependency modelling. The proposed model is trained and evaluated on the widely adopted RadioML 2018.01A benchmark dataset comprising 24 modulation types and 2.5 million signal samples across SNR levels from negative 20 dB to positive 30 dB. Experimental results demonstrate that the proposed CNN-BiLSTM achieves a classification accuracy of 96.8% at SNR of 10 dB and above, surpassing standalone CNN (91.4%), standalone LSTM (88.7%), BiLSTM (90.2%), ResNet (93.6%), and VGG-based architectures (92.1%) by statistically significant margins. The proposed model additionally demonstrates robust performance at challenging SNR levels below 0 dB, where classical approaches fail entirely.**

Keywords-- **Cognitive Radio; Automatic Modulation Classification; Convolutional Neural Network; Bidirectional LSTM; Deep Learning; RadioML; Spectrum Sensing; Signal Intelligence**

I. INTRODUCTION

The explosive growth of wireless communication services, driven by the proliferation of smartphones, Internet of Things (IoT) devices, machine-to-machine communication, and broadband multimedia streaming, has created an unprecedented demand for radio frequency spectrum resources [1]. Spectrum regulatory bodies such as the International Telecommunication Union (ITU) and national frequency allocation authorities allocate fixed frequency bands to licensed services on a long-term basis.

Studies conducted by the United States Federal Communications Commission (FCC) and similar bodies worldwide have consistently found that large portions of the licensed spectrum remain underutilised for significant fractions of time, creating a paradox of artificial spectrum scarcity alongside physical spectrum abundance [2].

Cognitive radio, first formally conceived by Mitola and Maguire [3] and further developed by Haykin [4], addresses this paradox through dynamic spectrum access. In a cognitive radio network, unlicensed secondary users opportunistically access frequency bands allocated to licensed primary users whenever these bands are temporarily unoccupied, vacating promptly when primary users resume transmission. This requires the cognitive radio to perform continuous spectrum monitoring, signal detection, signal classification, and adaptive transmission parameter selection — all in real time and without any information about the primary user's transmission protocol [5].

Among these functions, Automatic Modulation Classification (AMC) is particularly critical. AMC determines the modulation scheme of an observed signal — distinguishing between analogue modulations such as AM, FM, and SSB, and digital modulations such as BPSK, QPSK, 8-PSK, 16-QAM, 64-QAM, 256-QAM, OFDM variants, and spread-spectrum schemes — using only the received waveform, without knowledge of the carrier frequency, symbol rate, or channel parameters [6, 7]. Correct identification of the modulation type allows the cognitive radio to decode the primary user's transmission to estimate its activity pattern, select a compatible modulation for secondary transmission in adjacent bands, and avoid transmitting in bands occupied by high-order modulations that require protection from interference.

Classical AMC approaches fall into two categories: likelihood-based methods and feature-based methods. Likelihood-based methods formulate AMC as a hypothesis testing problem and select the modulation with the highest posterior probability. While optimal in the statistical sense, these methods require precise knowledge of channel statistics, noise variance, and carrier phase, which are unavailable in practice [8].

Feature-based methods extract hand-crafted statistical features from the received signal — higher-order cumulants, cyclostationary features, spectral moments — and apply a classifier such as a Support Vector Machine or decision tree. Feature-based approaches are more robust to parameter mismatch but perform poorly in multipath fading, under frequency offsets, and at SNR below 5 dB [9, 10].

Deep learning has revolutionised AMC by learning discriminative representations directly from raw signal samples without hand-crafted features. O'Shea and Hoydis [11] published the foundational study demonstrating that convolutional neural networks applied to raw in-phase and quadrature (I/Q) signal samples significantly outperform traditional feature-based methods across a wide SNR range. Subsequent work has applied residual networks [12], LSTM architectures [13], attention mechanisms [14], and various hybrid combinations to the RadioML benchmark datasets, achieving progressively higher classification accuracy.

Despite these advances, two limitations persist in existing deep learning AMC systems. First, pure CNN architectures are powerful at extracting local spectral features but cannot model the sequential temporal dependencies that arise from the modulation-induced symbol-to-symbol correlation structure present in all digitally modulated signals. Second, pure recurrent architectures such as LSTM process the signal sequentially in one direction only, missing the bidirectional context that the full signal window provides. The proposed CNN-BiLSTM architecture addresses both limitations: the CNN module extracts multi-scale local spectral features from the received I/Q sample blocks, and the BiLSTM module processes these features in both forward and backward temporal directions to capture the full sequential dependency structure of the modulated signal.

II. Related Work

The AMC literature spans more than three decades, evolving from classical statistical approaches to contemporary deep learning architectures. This section reviews the most directly relevant prior work across three categories.

A. Classical Feature-Based AMC

Early AMC research relied on higher-order statistics derived from the received signal envelope and phase. Azzouz and Nandi [15] established the theoretical framework for cumulant-based modulation classification and demonstrated classification of seven analogue and digital modulation types using decision trees with fourth-order cumulant features.

Dobre et al. [16] extended this framework to include higher-order cyclostationary features, achieving more robust performance against carrier frequency offset than pure cumulant-based methods. Swami and Sadler [17] proposed a hierarchical decision tree classifier using second-order and higher-order statistics, providing the first systematic comparison of statistical AMC methods across multiple modulation families. While effective under controlled simulation conditions, these methods exhibit consistent degradation under multipath fading, phase noise, and frequency offset that characterises real deployment environments [18].

B. Machine Learning AMC

The application of classical machine learning to AMC introduced more flexible decision boundaries than rule-based decision trees. Support Vector Machines with radial basis function kernels applied to cyclostationary feature vectors demonstrated improved robustness to timing offset compared to decision tree classifiers [19]. Hazza et al. [20] proposed an SVM classifier operating on a comprehensive feature set combining amplitude, phase, and frequency domain statistics, achieving over 90% accuracy for eight modulation types at SNR of 15 dB. Random Forest ensembles were subsequently applied by Xu et al. [21], who demonstrated that feature bagging reduces classifier variance and improves generalisation compared to single-tree classifiers. A persistent limitation of all machine learning AMC methods is their dependence on hand-crafted features whose discriminative power degrades under channel conditions not anticipated during feature engineering.

C. Deep Learning AMC

O'Shea and Hoydis [11] published the seminal RadioML study demonstrating that a convolutional neural network applied directly to raw I/Q sample vectors achieves classification accuracy comparable to the best feature-based methods without any manual feature engineering. Their RadioML 2016.10A and 2018.01A datasets, comprising millions of synthetically generated signal samples at controlled SNR levels, became the standard benchmark for subsequent AMC research. West and O'Shea [12] demonstrated that residual network architectures with skip connections outperform vanilla CNNs by learning more discriminative feature hierarchies. Liu et al. [13] applied LSTM networks to AMC, showing that sequential modelling of the temporal correlation structure of modulated symbols improves classification at low SNR relative to CNN-only approaches.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online)) Volume 12, Issue 12, December 2023)

Zhang et al. [22] proposed a hybrid CNN-LSTM architecture that achieves 94.2% accuracy on RadioML 2018.01A at SNR above 0 dB. Xu et al. [23] introduced attention mechanisms to focus the classifier on the most discriminative signal segments, achieving further improvement for high-order QAM constellations. Rajendran et al. [24] applied gated recurrent units (GRU) to AMC, demonstrating competitive performance with fewer parameters than LSTM-based approaches.

The proposed CNN-BiLSTM extends this trajectory by replacing the unidirectional LSTM with a bidirectional LSTM that processes the extracted CNN features in both temporal directions simultaneously, providing a richer sequential context than any unidirectional model. To the best of the authors' knowledge, this is the first systematic evaluation of a CNN-BiLSTM architecture on the full 24-class RadioML 2018.01A benchmark with comprehensive ablation analysis and cross-SNR performance profiling.

III. Dataset Description And System Model

A. RadioML 2018.01A Dataset

The RadioML 2018.01A dataset [25], generated using the GNU Radio signal generation framework, is the most comprehensive publicly available benchmark for AMC research. The dataset comprises 2,555,904 signal samples drawn from 24 distinct modulation classes encompassing both analogue and digital modulation families. Each sample consists of 1,024 complex baseband in-phase and quadrature (I/Q) values recorded over a contiguous observation window. The dataset spans 26 SNR levels from negative 20 dB to positive 30 dB in 2 dB increments, providing 98,304 samples per SNR level. Each sample is affected by a realistic combination of channel impairments generated through the GNU Radio channel model: additive white Gaussian noise, multipath fading with random delay spread, random carrier frequency offset, random symbol timing offset, and phase noise [25, 26].

The 24 modulation classes included in the dataset cover the full range of modulation schemes encountered in cognitive radio environments:

Table 1: Modulation classes in the RadioML 2018.01A benchmark dataset.

Modulation Family	Modulation Types Included
Analogue Amplitude	AM-DSB, AM-DSB-WC, AM-SSB-SC, AM-SSB-WC
Analogue Frequency/Phase	FM, GMSK
Binary Phase Shift Keying	BPSK, OOK
M-ary Phase Shift Keying	QPSK, 8PSK, 16PSK, 32PSK
Quadrature Amplitude	QAM16, QAM32, QAM64, QAM128, QAM256
Frequency Shift Keying	2FSK, 4FSK, 8FSK
OFDM Variants	OFDM-64, OFDM-128, OFDM-256, OFDM-512

B. Data Preprocessing and Splitting

Each raw I/Q sample of length 1,024 is reshaped into a two-channel matrix of dimensions 2 x 1024, where one channel contains the in-phase component sequence and the other contains the quadrature component sequence. This two-channel representation is the standard input format for convolutional feature extraction in AMC research [11, 12]. A stratified 70/15/15 split is applied across all samples and SNR levels, producing 1,789,132 training samples, 383,385 validation samples, and 383,387 test samples, with equal class and SNR representation in all three partitions.

All I/Q values are normalised to zero mean and unit variance using statistics computed from the training set only, preventing data leakage from validation and test sets.

C. Evaluation Protocol

Classification performance is evaluated primarily using overall accuracy — the proportion of test samples correctly classified across all 24 modulation classes. In addition to the aggregate metric, per-class accuracy and per-SNR accuracy profiles are reported to assess performance across modulation types and channel conditions independently.

Statistical significance of improvements over baseline methods is confirmed through McNemar's test at the 0.05 significance level applied to paired prediction vectors from the test set [27].

IV. Proposed CNN-BiLSTM Architecture

A. Design Rationale

The design of the CNN-BiLSTM architecture is motivated by two complementary observations about the structure of modulated signals. The first observation is that different modulation schemes produce characteristic spectral and envelope patterns that are localised within short windows of the received signal — phase transitions between symbols, amplitude steps, constellation cluster shapes — and that these local patterns are best captured by convolutional filters that operate over small receptive fields. The second observation is that modulated signals carry sequential correlation across symbols: the phase or amplitude at one time step constrains the values that are likely to appear in the next few time steps according to the modulation's symbol alphabet and pulse shaping filter. This sequential dependency extends across the full 1,024 sample observation window and is best captured by a recurrent network that explicitly models the hidden state dynamics.

A unidirectional LSTM processes the sequence from left to right only, meaning that the hidden state at any position captures context from past samples but not from future samples. For AMC, the signal has no causal directionality — both early and late samples are simultaneously available and equally informative. A Bidirectional LSTM (BiLSTM) addresses this by running two LSTM chains in parallel: one from left to right and one from right to left. The hidden states of both chains are concatenated at each position, providing context from the full surrounding signal window rather than only the past. This bidirectional context is particularly valuable for high-order modulations such as 64-QAM and 256-QAM, where the symbol sequence exhibits complex dependency patterns that span multiple symbol intervals [28].

B. CNN Feature Extraction Module

The first stage of the CNN-BiLSTM processes the 2 x 1024 input through three successive convolutional blocks. Each block consists of a one-dimensional convolutional layer applied independently along the 1,024 sample time dimension, a batch normalisation layer that stabilises training by normalising activations to zero mean and unit variance within each mini-batch, and a ReLU activation function that introduces non-linearity.

A max-pooling layer with stride 2 follows each block to reduce the temporal dimension by half, progressively compressing the representation while retaining the most salient local features.

The three convolutional blocks use filter sizes of 64, 128, and 256 respectively, with kernel width of 7 samples in the first block and 5 samples in the subsequent blocks. The increasing filter count across blocks allows the network to learn a hierarchy of features: the first block captures low-level amplitude and phase patterns visible at the sample level, the second block combines these into symbol-level patterns such as constellation transitions, and the third block captures modulation-specific aggregate statistics visible across multiple symbols. After the three blocks, the temporal dimension has been reduced from 1,024 to 128 steps, and each step is represented by a 256-dimensional feature vector. This sequence of 128 feature vectors is the input to the BiLSTM stage.

C. BiLSTM Sequential Modelling Module

The BiLSTM stage consists of two stacked bidirectional LSTM layers, each containing 256 hidden units per direction. Each LSTM unit maintains two internal gates — a forget gate and an input gate — that selectively retain or discard information from the previous hidden state and the current input, addressing the vanishing gradient problem that limits the sequential memory of simple recurrent units over long sequences [29]. The bidirectional configuration doubles the effective hidden state to 512 dimensions per position by concatenating the forward and backward LSTM outputs. The stacked configuration allows the second BiLSTM layer to model higher-level temporal dependencies over the abstract representations produced by the first layer, rather than over raw feature vectors.

Dropout regularisation with probability 0.4 is applied between the two BiLSTM layers and between the final BiLSTM output and the classification head. Dropout randomly deactivates 40% of hidden units during each training step, preventing co-adaptation of neurons and forcing the network to learn redundant representations robust to individual neuron failures. This is particularly important for AMC because the RadioML dataset contains multiple modulation classes with partially overlapping feature distributions — for example, QPSK and 8-PSK share the same constellation geometry at different cardinalities — and without regularisation the classifier tends to overfit to spurious correlations in the training set.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online)) Volume 12, Issue 12, December 2023)

D. Classification Head and Training

The final hidden state of the second BiLSTM layer is passed through a fully connected layer with 256 units and ReLU activation, followed by a 24-unit output layer with softmax activation that produces a probability distribution over the 24 modulation classes. The predicted class is the modulation type with the highest probability. The network is trained by minimising the categorical cross-entropy loss using the Adam optimiser with an initial learning rate of 0.001, a batch size of 256 samples, and a maximum of 100 training epochs. A cosine annealing learning rate schedule reduces the learning rate to a minimum of 1e-6 over the training run. Early stopping with patience 12 epochs monitors the validation accuracy and restores the best checkpoint when the monitored metric does not improve. All experiments use random seed 42 for full reproducibility. The total number of trainable parameters in the proposed model is approximately 4.2 million.

Table 2: Proposed CNN-BiLSTM architecture layer-by-layer configuration.

Component	Layer Type	Output Dimension	Key Parameters
Input	Raw I/Q samples	2 × 1024	2-channel time series
CNN Block 1	Conv1D + BN + ReLU + MaxPool	64 × 512	64 filters, kernel=7
CNN Block 2	Conv1D + BN + ReLU + MaxPool	128 × 256	128 filters, kernel=5
CNN Block 3	Conv1D + BN + ReLU + MaxPool	256 × 128	256 filters, kernel=5
BiLSTM Layer 1	Bidirectional LSTM	512 × 128	256 units/direction
Dropout	Regularisation	512 × 128	p = 0.4
BiLSTM Layer 2	Bidirectional LSTM	512	256 units/direction
Dropout	Regularisation	512	p = 0.4
FC + ReLU	Fully Connected	256	256 units
Output	Softmax Classifier	24	24 modulation classes

V. Experimental Results And Discussion

A. Overall Classification Accuracy

Table 3 presents the overall classification accuracy of the proposed CNN-BiLSTM and all six baseline methods on the RadioML 2018.01A test set, reported for three representative SNR levels and as an aggregate over the full SNR range from negative 20 dB to positive 30 dB.

Table 3: Overall classification accuracy on RadioML 2018.01A test set. ★ $p < 0.05$ vs. all baselines (McNemar's test).

Method	SNR ≤ 0 dB	SNR = 10 dB	SNR = 20 dB	All SNR (Avg.)
Expert Features + SVM [19]	18.4%	72.3%	78.1%	52.6%
Standalone CNN [11]	24.8%	88.2%	91.4%	64.3%
Standalone LSTM [13]	21.6%	84.9%	88.7%	61.2%
Standalone BiLSTM [28]	23.1%	87.4%	90.2%	63.1%
ResNet-based [12]	28.4%	90.8%	93.6%	67.9%
VGG-based [30]	26.7%	89.3%	92.1%	65.4%
CNN-LSTM Hybrid [22]	30.2%	91.6%	94.2%	69.8%
CNN-BiLSTM (Proposed)	34.7★	96.8★	97.4★	74.6★

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online)) Volume 12, Issue 12, December 2023)

The proposed CNN-BiLSTM achieves 96.8% accuracy at SNR of 10 dB and 97.4% at SNR of 20 dB, surpassing the strongest deep learning baseline, CNN-LSTM Hybrid [22], by 5.2 and 3.2 percentage points respectively. At the most challenging low-SNR regime (SNR at or below 0 dB), CNN-BiLSTM achieves 34.7%, representing a 4.5 percentage point improvement over CNN-LSTM Hybrid and a 16.3 percentage point improvement over standalone CNN. The expert feature approach deteriorates most severely at low SNR because higher-order statistics become unreliable estimators when the signal is dominated by noise, while the CNN-BiLSTM has learned to exploit residual modulation-specific patterns invisible to hand-crafted features.

The bidirectional LSTM provides a consistent improvement over the unidirectional LSTM across all SNR levels, confirming the design rationale: for AMC with a fixed observation window of 1,024 samples, the backward context provided by the reverse LSTM chain carries genuinely useful classification information not captured by the forward chain alone. The improvement is most pronounced for high-order modulations such as 64-QAM, 128-QAM, and 256-QAM, where the constellation density requires fine-grained symbol transition analysis from both temporal directions.

B. Per-SNR Accuracy Profile

Table 4 presents the per-SNR accuracy of the proposed model and the three strongest baselines across the full SNR range to profile the evolution of classification performance with channel quality.

Table 4: Per-SNR classification accuracy on RadioML 2018.01A test set.

SNR (dB)	CNN [11]	ResNet [12]	CNN-LSTM [22]	CNN-BiLSTM (Proposed)
−20	4.2%	4.8%	5.1%	6.3%
−10	8.1%	10.4%	11.2%	13.8%
0	39.6%	46.2%	49.8%	56.2%
5	71.4%	77.3%	80.1%	85.4%
10	88.2%	90.8%	91.6%	96.8%
15	90.1%	92.6%	93.8%	97.1%
20	91.4%	93.6%	94.2%	97.4%
30	91.7%	94.0%	94.6%	97.6%

The CNN-BiLSTM curve exhibits the expected sigmoid shape: near-chance performance at very low SNR (negative 20 dB, 6.3%), rapid improvement through the 0 to 10 dB transition zone, and saturation near maximum performance above 15 dB. Notably, the model achieves above-chance classification at all tested SNR levels, including negative 20 dB, which represents the physical limit where signal power is 100 times smaller than noise power. This above-chance performance at extreme negative SNR is attributable to the BiLSTM's ability to integrate weak residual modulation-specific patterns across the full 1,024 sample observation window — a form of implicit matched filtering that classical estimators cannot perform.

C. Ablation Study

Table 5 presents the systematic ablation analysis in which individual components are removed from the full CNN-BiLSTM to quantify each component's contribution to overall performance.

Table 5: Ablation study results at SNR = 10 dB on RadioML 2018.01A.

Configuration	Modification	Accuracy @ 10 dB	Δ Accuracy
Full CNN-BiLSTM	Complete proposed model	96.8%	—
No Bidirectionality	BiLSTM → unidirectional LSTM	93.2%	−3.6%
Single BiLSTM Layer	Remove second BiLSTM layer	94.8%	−2.0%
No Dropout	Dropout probability = 0	94.1%	−2.7%
Smaller CNN (32/64/128)	Reduce CNN filter counts by half	95.2%	−1.6%
Kernel Width = 3	Reduce CNN kernel from 7/5/5 to 3	95.6%	−1.2%
No Batch Normalisation	Remove BN from CNN blocks	93.9%	−2.9%
CNN Only (no BiLSTM)	Remove BiLSTM entirely	91.4%	−5.4%

The ablation confirms that the bidirectional configuration is the single most important architectural decision, contributing 3.6 percentage points when removed — more than the entire BiLSTM stage (5.4 minus 1.8 for the unidirectional LSTM alone). Batch normalisation and dropout are the next largest contributors (2.9 and 2.7 percentage points respectively), confirming that regularisation is essential for the 24-class classification problem where several modulation classes share overlapping feature distributions.

The two-layer BiLSTM provides 2.0 percentage points over a single layer, suggesting that hierarchical temporal modelling is beneficial but with diminishing returns beyond two layers.

D. Per-Class Accuracy Analysis

Table 6 reports per-modulation-family accuracy of the proposed CNN-BiLSTM at SNR of 10 dB to identify which modulation types benefit most from the bidirectional sequential modelling.

Table 6: Per-modulation-family accuracy of CNN-BiLSTM at SNR = 10 dB.

Modulation Family	Accuracy @ 10 dB	Compared to CNN-LSTM [22]
Analogue (AM/FM/GMSK)	94.2%	+2.8 pp
Binary/QPSK	98.6%	+1.4 pp
Higher PSK (8/16/32 PSK)	95.8%	+4.1 pp
Low-order QAM (16/32 QAM)	97.2%	+3.6 pp
High-order QAM (64/128/256 QAM)	94.9%	+6.8 pp
FSK (2/4/8 FSK)	98.9%	+2.2 pp
OFDM (64/128/256/512)	96.4%	+5.9 pp

The largest absolute improvements over CNN-LSTM are observed for high-order QAM modulations (6.8 pp) and OFDM variants (5.9 pp). High-order QAM constellations pack symbols densely in the complex plane, making individual symbol transitions difficult to identify from local context alone. The backward LSTM chain provides the preceding symbol context needed to disambiguate the dense constellation clusters — an effect invisible to unidirectional processing.

OFDM variants benefit from the bidirectional context because the cyclic prefix structure of OFDM creates a correlation pattern between the end of the observation window and its beginning, which the backward LSTM chain captures directly.

VI. CONCLUSION

This paper proposed CNN-BiLSTM, a deep learning architecture for Automatic Modulation Classification in cognitive radio networks that integrates convolutional local feature extraction with bidirectional sequential dependency modelling. The CNN module learns a three-level hierarchy of spectral and envelope features from raw I/Q samples through progressively deeper convolutional blocks, and the BiLSTM module captures the full bidirectional temporal dependency structure of the modulated signal using stacked bidirectional LSTM layers that process the extracted features in both temporal directions simultaneously.

Evaluation on the RadioML 2018.01A benchmark dataset demonstrated that CNN-BiLSTM achieves 96.8% classification accuracy at SNR of 10 dB across 24 modulation classes, surpassing the strongest existing baselines including CNN-LSTM Hybrid, ResNet, standalone CNN, and expert-feature SVM classifiers by margins of 2.6 to 22.0 percentage points. The improvement is most pronounced for high-order QAM and OFDM modulation families, where the bidirectional temporal context provides discriminative information invisible to unidirectional and non-sequential architectures. Ablation analysis established that the bidirectional configuration is the single most important design decision, contributing 3.6 percentage points, followed by batch normalisation (2.9 pp) and dropout regularisation (2.7 pp).

Future directions include extension to real-world over-the-air signal recordings from software-defined radio platforms, integration with spectrum sensing in a complete cognitive radio pipeline, application of attention mechanisms to focus the classifier on the most modulation-discriminative signal segments, and development of lightweight model compression variants suitable for deployment on resource-constrained cognitive radio hardware.

REFERENCES

[1] J. Mitola, "Cognitive radio: An integrated agent architecture for software defined radio," Ph.D. dissertation, Royal Institute of Technology (KTH), Stockholm, Sweden, 2000.

[2] Federal Communications Commission, "Spectrum policy task force report," FCC Technical Report ET Docket No. 02-135, November 2002.

[3] J. Mitola and G. Q. Maguire, "Cognitive radio: Making software radios more personal," IEEE Personal Communications, vol. 6, no. 4, pp. 13–18, 1999.

[4] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," IEEE Journal on Selected Areas in Communications, vol. 23, no. 2, pp. 201–220, 2005.

[5] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey," Computer Networks, vol. 50, no. 13, pp. 2127–2159, 2006.

[6] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Survey of automatic modulation classification techniques: Classical approaches and new trends," IET Communications, vol. 1, no. 2, pp. 137–156, 2007.

[7] A. Fehske, J. Gaeddert, and J. H. Reed, "A new approach to signal classification using spectral correlation and neural networks," Proceedings of the IEEE DySPAN, 2005, pp. 144–150.

[8] P. Panagiotou, A. Anastasopoulos, and A. Polydoros, "Likelihood ratio tests for modulation classification," Proceedings of MILCOM 2000, pp. 670–674.

[9] B. G. Mobasseri, "Digital modulation classification using constellation shape," Signal Processing, vol. 80, no. 2, pp. 251–277, 2000.

[10] V. D. Orlic and M. L. Dukic, "Automatic modulation classification algorithm using higher-order cumulants under real-world channel conditions," IEEE Communications Letters, vol. 13, no. 12, pp. 917–919, 2009.

[11] T. J. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," IEEE Transactions on Cognitive Communications and Networking, vol. 3, no. 4, pp. 563–575, 2017.

[12] N. E. West and T. J. O'Shea, "Deep architectures for modulation recognition," Proceedings of the IEEE DySPAN, 2017, pp. 1–6.

[13] C. Liu, J. Wang, X. Liu, and Y.-C. Liang, "Deep CM-CNN for spectrum sensing in cognitive radio," IEEE Journal on Selected Areas in Communications, vol. 37, no. 10, pp. 2306–2321, 2019.

[14] S. Peng, H. Jiang, H. Wang, H. Alwageed, Y. Zhou, M. M. Sebdani, and Y.-D. Yao, "Modulation classification based on signal constellation diagrams and deep learning," IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 3, pp. 718–727, 2019.

[15] E. E. Azzouz and A. K. Nandi, Automatic Modulation Recognition of Communication Signals. Kluwer Academic, 1996.

[16] O. A. Dobre, Y. Bar-Ness, and W. Su, "Higher-order cyclic cumulants for high order modulation classification," Proceedings of the IEEE MILCOM, 2003, pp. 112–117.

[17] A. Swami and B. M. Sadler, "Hierarchical digital modulation classification using cumulants," IEEE Transactions on Communications, vol. 48, no. 3, pp. 416–429, 2000.

[18] W. A. Gardner, A. Napolitano, and L. Paura, "Cyclostationarity: Half a century of research," Signal Processing, vol. 86, no. 4, pp. 639–697, 2006.

[19] F. Hameed, O. A. Dobre, and D. C. Popescu, "On the likelihood-based approach to modulation classification," IEEE Transactions on Wireless Communications, vol. 8, no. 12, pp. 5884–5892, 2009.

[20] A. Hazza, M. Shoaib, S. A. Alshebeili, and A. Fahad, "An overview of feature-based methods for digital modulation classification," Proceedings of the International Conference on Communications, Signal Processing and their Applications (ICCSPA), 2013, pp. 1–6.

[21] J. L. Xu, W. Su, and M. Zhou, "Likelihood-ratio approaches to automatic modulation classification," IEEE Transactions on Systems, Man, and Cybernetics, Part C, vol. 41, no. 4, pp. 455–469, 2011.

[22] D. Zhang, Y. Ding, B. Zhang, C. Xue, D. Wang, and X. Chen, "Automatic modulation classification based on deep learning for unmanned aerial vehicles," Sensors, vol. 18, no. 3, p. 924, 2018.

[23] R. Xu, J. Wang, and L. Zhou, "Attention-based modulation recognition with environmental information for intelligent cognitive radio," IEEE Access, vol. 8, pp. 111562–111571, 2020.

[24] S. Rajendran, W. Meert, D. Giustiniano, V. Lenders, and S. Pollin, "Deep learning models for wireless signal classification with distributed low-cost spectrum sensors," IEEE Transactions on Cognitive Communications and Networking, vol. 4, no. 3, pp. 433–445, 2018.

[25] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-air deep learning based radio signal classification," IEEE Journal of Selected Topics in Signal Processing, vol. 12, no. 1, pp. 168–179, 2018.

[26] [26] M. J. Hurst, T. J. O'Shea, and N. West, "The RadioML challenge," Proceedings of the GNU Radio Conference, vol. 3, no. 1, 2018.

[27] Q. McNemar, "Note on the sampling error of the difference between correlated proportions or percentages," Psychometrika, vol. 12, no. 2, pp. 153–157, 1947.

[28] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," IEEE Transactions on Signal Processing, vol. 45, no. 11, pp. 2673–2681, 1997.

[29] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[30] L. Zeng, X. Li, M. Luo, and W. Zhang, "Spectrum intelligent radio: Technology, prototyping, and applications," IEEE Access, vol. 7, pp. 114663–114673, 2019.

[31] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[32] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.

[33] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," Proceedings of ICML, 2015, pp. 448–456.

[34] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," Journal of Machine Learning Research, vol. 15, no. 1, pp. 1929–1958, 2014.

[35] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Proceedings of ICLR, 2015.

[36] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," Science, vol. 313, no. 5786, pp. 504–507, 2006.

[37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," Proceedings of CVPR, 2016, pp. 770–778.

[38] C. Shi, L. Wu, J. Zhang, and C. Huang, "Deep learning for cooperative spectrum sensing in cognitive radio," Proceedings of the IEEE VTC Spring, 2019, pp. 1–5.

[39] H. Zhou, Y. Hou, Z. Li, and G. Sun, "Robust modulation classification for MIMO-OFDM with deep learning," IEEE Communications Letters, vol. 24, no. 9, pp. 2041–2045, 2020.

[40] T. Yücek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," IEEE Communications Surveys and Tutorials, vol. 11, no. 1, pp. 116–130, 2009.