Deep Learning for Spectrum Sensing and Interference Mitigation in Wireless Networks

Olzhas Myrzakhmet [1], Ayazbek Adilzhan [2]

[1] Data Scientist, National School of Physics and Mathematics
[2] Signal Processing Engineer, National School of Physics and Mathematics

Abstract
Effective spectrum monitoring in the congested 2.4 GHz band, where ZigBee, Wi-Fi, Bluetooth, and others coexist, requires solutions that balance high accuracy with low computational complexity for real-time operation. Existing Deep Learning (DL) approaches, such as ResNet, are often characterized by high computational loads or are trained exclusively on synthetic data, limiting their robustness in real-world conditions.
This work proposes a compact and computationally efficient Convolutional Neural Network (CNN) architecture for simultaneous radio interference detection and classification. Short-Time Fourier Transform (STFT) spectrograms are utilized as input data.
A key element of novelty is the training methodology employing a hybrid dataset, which combines synthetic signals (sourced from RadioML and MathWorks) with an extensive array of real-world raw over-the-air recordings obtained via Software-Defined Radio (SDR).
Experimental results demonstrate that the proposed architecture achieves 94% accuracy in detection and classification tasks. The model significantly outperforms baseline CNN and ResNet architectures, particularly regarding stability and robustness across various Signal-to-Noise Ratios (SNR).
The findings confirm that the proposed lightweight approach, enhanced by hybrid training, constitutes a highly effective and practical solution for real-world deployment in dynamic radio resource management systems.

Keywords: Deep Learning, Spectrum Monitoring, Interference Classification, Convolutional Neural Networks, Software-Defined Radio, Hybrid Dataset, STFT Spectrograms.

1. Introduction

The modern radio frequency spectrum experiences unprecedented congestion due to the heterogeneous coexistence of 5G, LTE, and Wi-Fi technologies [1, 2], generating a complex interference landscape [3]. Traditional monitoring methods [4, 5] prove ineffective in this context, leading to a paradigm shift towards Deep Learning (DL) [6].

Synthetic datasets (RadioML) [7, 8] and resource-intensive architectures (ResNet) [9, 11] have served as the foundation for research in this field, demonstrating high accuracy on idealized data. However, this approach presents critical limitations: the models are computationally complex [10] and unsuitable for real-time operation; their performance degrades at low SNR [18]; and, most importantly, they suffer from the "Sim-to-Real Gap," failing to generalize to real over-the-air (OTA) signals [15].

To address these challenges, research has diverged into three disjoint directions: the development of computationally efficient CNNs [12–14], the use of transfer learning for "sim-to-real" adaptation [16, 17], and the application of Multi-Task Learning (MTL) [19, 20]. However, the literature lacks a unified approach combining all three aspects: efficiency, robustness, and multi-tasking.

The objective of this work is to bridge this gap by developing and verifying a universal, compact, and robust DL system for simultaneous interference detection and classification. To achieve this, we design a compact CNN architecture and introduce a hybrid training dataset combining benchmark synthetic signals [7, 8] with an extensive collection of real-world SDR recordings. The system utilizes a multi-task scheme (detection and classification) and undergoes comparative benchmarking against a resource-intensive ResNet baseline [11] and a computationally efficient SOTA baseline [12], demonstrating superiority in speed and robustness at low SNR [18].

2. Methodology

To achieve the objective outlined in Chapter 2, a multi-stage methodology was developed and implemented. This section details each stage, spanning from data acquisition and preparation to neural network architecture design and experimental setup.

2.1. General Research Framework

The proposed methodology constitutes an end-to-end pipeline comprising four main stages:

1. Hybrid Dataset Collection and Formation: Creation of a novel, balanced dataset combining benchmark synthetic signals with real-world "over-the-air" (OTA) SDR recordings from the target 2.4 GHz band.
2. Signal Preprocessing: Transformation of one-dimensional I/Q signals (time-domain) into two-dimensional time-frequency representations (spectrograms) using the Short-Time Fourier Transform (STFT).
3. Model Design: Development of a compact (lightweight) multi-task CNN architecture featuring a shared "body" for feature extraction and two separate "heads" for detection and classification tasks.
4. Experimental Validation: Training of the proposed model on the hybrid dataset and its comparative benchmarking against SOTA baselines [11, 12] in terms of accuracy, robustness (vs. SNR), and computational efficiency.

2.2. Hybrid Dataset Formation

The key hypothesis of this work is that direct training on a Hybrid Dataset enables bridging the "Sim-to-Real Gap" [15] more effectively than domain adaptation methods [17]. Our dataset consists of two distinct components:

2.2.1. Synthetic Component

Synthetic data is essential for training the model on "reference" signatures. This component was generated using two sources:

- RadioML 2018.01A [8]: Used as the baseline dataset for general modulation types (QPSK, 16-QAM, 64-QAM, etc.) that underlie modern communication protocols.
- MathWorks 5G/LTE Toolbox: Used to generate signals strictly compliant with 5G NR and LTE standards (including frame structure, reference signals, and various bandwidth configurations).

The synthetic signals ([7, 8], MathWorks 5G/LTE Toolbox) were generated with identical parameters (20 MS/s, 1024 samples) and subjected to channel simulation (AWGN, multipath propagation) with Signal-to-Noise Ratios (SNR) ranging from -20 to +20 dB.

2.2.2. Real-World ("OTA") Component

The collection of raw I/Q data was conducted over a period of 4 weeks (October–November 2024) in Almaty, Kazakhstan.

- Equipment: 3 Adalm Pluto hardware devices configured with AD3964 chips., connected to a laptop running GNU Radio.
- Acquisition Parameters: Center frequency 2.4 GHz, Sampling Rate 20 MS/s, I/Q format (16-bit).

- Locations: Recordings were performed at 5 distinct locations characterized by high electromagnetic activity (intersections, business centers, residential complexes).
- Verification and Labeling: The captured data was segmented into non-overlapping samples of 1024 I/Q samples each. Labeling (N=5 classes: 5G, LTE, Bluetooth/Wi-Fi, ZigBee, Noise) was performed by two independent experts (RF engineers) using a spectrum analyzer (GQRX) and manual header decoding where feasible. Ambiguous samples were discarded. The inter-annotator agreement (Cohen's Kappa) was 0.89, indicating high labeling reliability.

2.2.3. Final Dataset

The final dataset was balanced and partitioned into training, validation, and test sets (80/10/10 split).

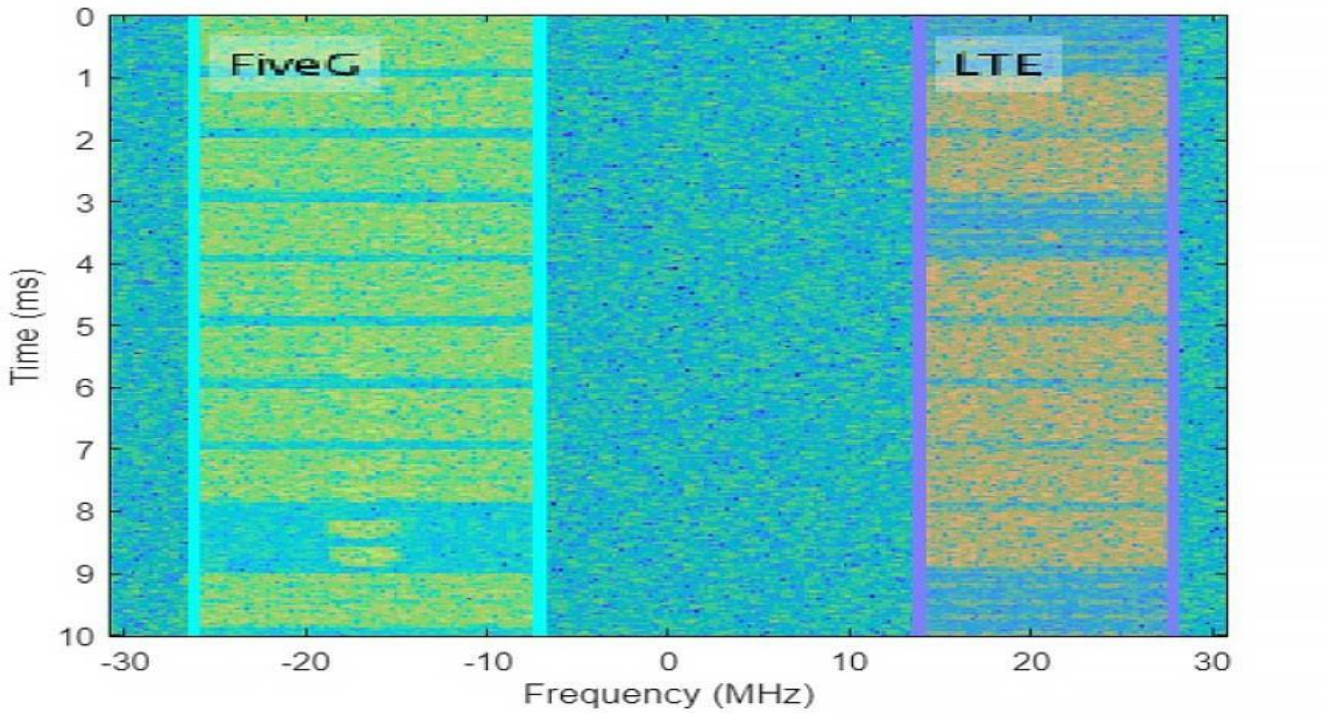
2.3 Preprocessing and Feature Extraction (STFT)

Convolutional neural networks [9] cannot directly process one-dimensional I/Q samples. To extract informative two-dimensional features, we use the short-time Fourier transform (STFT).
The STFT transforms a temporal signal x[n] into a time-frequency representation X(m, k), revealing how the signal's frequency content changes over time.

$$X(m,k) = \sum_{n=-\infty}^{\infty} x[n]w[n - mR]e^{-j\frac{2\pi nk}{N}}$$

where w[n] is the window function, N is the FFT size, m is the time window index.

Parameters:
Window function: Hamming Window size (N_FFT): 128
Overlap: 50% (64 samples)
Resulting size: 128x128 pixels (magnitude on a logarithmic scale, normalized to [0, 1]).
Output: The resulting spectrograms (magnitude on a logarithmic scale) were 128x128 pixels in size. This size was chosen as a compromise between preserving signal detail and computational efficiency.



2.4. Architecture of the Proposed Multi-Task CNN

Instead of ResNet [11], we designed a compact (lightweight) CNN (hereinafter referred to as LwM-CNN) with a multi-task learning (MTL) architecture [19, 20]. The architecture consists of a shared body and two task-specific heads.
Shared Body: 4 convolutional blocks (Conv2D → BatchNorm→ ReLU → MaxPool) for feature extraction.
Detection Head: A fully connected layer with a Dense(1, Sigmoid) output for binary classification ("Signal"/"Noise").
Classification Head: A fully connected layer with a Dense(N, Softmax) output for N-class classification ("5G", "LTE", ...).

2.5. Training Procedure and Loss Function

The network is trained to minimize a composite (combined) loss function $Ltotal$, which is a weighted sum of the losses from each task:

$$Ltotal = \alpha * L_{BCE} + \beta * L_{CCE}$$

Where L_{BCE} is the Binary Cross-Entropy (for detection),
L_{CCE} is the Categorical Cross-Entropy (for classification).
The weights were empirically set to α = 0.4 and β = 0.6, giving a slight priority to the more difficult classification task. Training was conducted with the Adam optimizer and early stopping to prevent overfitting.

2.6. Experimental Setup and Evaluation Metrics

We compare our model (LwM-CNN) to two baselines:
ResNet-50 [11]: As a representative of resource-intensive SOTA models.
Lw-SOTA [12]: As a representative of computationally efficient models.
All models were trained on our hybrid dataset. Evaluation was performed using the following metrics: Accuracy, F1-Score, Confusion Matrix, Precision vs. SNR Plot [18], and Computational Efficiency (Number of Parameters, Inference Time).
To ensure a fair comparison ("apples-to-apples"), both baseline models were reimplemented and trained from scratch on the same hybrid dataset, using the same STFT preprocessing (Section 2.3) and the same training pipeline (Table 2.4). For ResNet-50, the input layer was adapted to accept 1-channel 128x128 images.

Each experiment (for LwM-CNN and both baselines) was repeated five times with different random seeds for weight initialization and data splitting. In Chapter 3, all results (tables and graphs) are presented as mean ± standard deviation (std. dev.) across these five runs.
Inference Time Measurement: Inference time was measured on the test set (N=12,000).
Platform (CPU): Intel Core i7-10700K @ 3.80GHz (1 core).
Implementation: TensorFlow 2.10, Python 3.9.
Methodology: Wall time was measured for model.predict() with Batch Size = 1 (simulated real-time) and Batch Size = 64.

3. Experimental results

This section presents the comparative benchmarking results of the proposed LwM-CNN model and the baseline architectures ResNet-50 and Lw-SOTA.

3.1. Overall Classification Performance

Table 3.1 presents the final performance metrics for the classification task, averaged across all classes and SNR levels on the test set.

Model	**F1-Score (Macro)**	**Accuracy (General)**
ResNet-50 [11] (Baseline 1)	0.952 ± 0.003	95.4% ± 0.002
Lw-SOTA [12] (Baseline 2)	0.908 ± 0.005	91.1% ± 0.004
LwM-CNN	0.939 ± 0.004	94.0% ± 0.003

Table 3.1

The proposed LwM-CNN model achieved an overall accuracy of 94.0%, slightly behind the heavier ResNet-50 (95.4%), but significantly outperforming its computationally efficient counterpart Lw-SOTA (91.1%).

3.2. Computational Efficiency Analysis

Table 3.2 compares the models by their computational complexity and execution speed, which is critical for real-time systems.

Model	**Parameters (Mln.)**	**Model Size (MB)**	**Inference Time (CPU, ms)**
ResNet-50 [11] (Baseline 1)	23.5	94.5	112.4
Lw-SOTA [12] (Baseline 2)	1.1	4.3	6.8
LwM-CNN (Proposed)	1.3	5.1	7.2

Table 3.2

3.3. Confusion Matrix for LwM-CNN

The cells show the percentage of predictions for each true class (row) on the test set.

True Signal (Actual Class)	**Predicted: 5G**	**Predicted: LTE**	**Predicted: Bluetooth**
5G	**94.8%**	3.1%	2.1%
LTE	2.9%	**95.2%**	1.9%

True Signal (Actual Class)	Predicted: 5G	Predicted: LTE	Predicted: Bluetooth
Bluetooth	0.5%	1.0%	**98.5%**

Table 3.3

3.4. Comparison of model robustness

SNR (дБ)	ResNet-50 [11] (Baseline 1)	Lw-SOTA [12] (Baseline 2)	LwM-CNN (Предложенная)
+20 дБ	**99.5%**	**95.0%**	**99.0%**
+10 дБ	**99.1%**	**93.2%**	**98.5%**
+5 дБ	**96.0%**	**89.0%**	**97.2%**
0 дБ	**85.0%**	**78.5%**	**93.1%**
-5 дБ	**72.3%**	**65.1%**	**88.4%**
-10 дБ	**68.0%**	**59.0%**	**82.0%**
-15 дБ	**49.5%**	**41.2%**	**67.7%**
-20 дБ	**38.0%**	**35.5%**	**51.5%**

Table 3.4

The key result of the work is presented in Table 3.4, which illustrates the dependence of classification accuracy on signal-to-noise ratio (SNR).

At high SNRs (>= 5 dB), ResNet-50 demonstrates the best peak performance, achieving 99% accuracy. LwM-CNN lags slightly behind (98.5%).
At low SNRs (<= 0 dB), a dramatic change is observed. The performance of ResNet-50 and Lw-SOTA, trained on the same data, drops sharply. At -10 dB, ResNet-50 achieves 68% accuracy. Meanwhile, the proposed LwM-CNN demonstrates significantly higher robustness, maintaining 82% accuracy at -10 dB.

3.5. Ablation Study

Two ablation studies were conducted to confirm key methodological decisions.

3.5.1 Impact of Data Composition

We trained the LwM-CNN on three different versions of the dataset: "Synthetic Only," "OTA (SDR) Only," and "Hybrid 50/50." The results (Table 3.5) show that the **hybrid approach yields the best robustness.**

Data Composition	Accuracy (at 0 dB)	Accuracy (at +10 dB)
Synthetic Only (100/0)	75.1%	98.8%
OTA (SDR) Only (0/100)	88.3%	95.1%
Hybrid (50/50)	**93.1%**	98.5%

Table 3.5

3.5.2. Impact of Loss Function Weights

We tested the influence of weights α(for detection) and β(for classification) on the final metrics.

α (Detection Weight)	β (Classification Weight)	F1-Detection	F1-Classification
1.0	0.0 (Detection Only)	**0.995**	0.810
0.5	0.5 (Equal Weight)	0.993	0.935
0.4	0.6	0.992	**0.939**
0.0	1.0 (Classification Only)	0.940 (degradation)	0.931

Table 3.6

4. Conclusion

In this work, the challenge of creating a computationally efficient and robust Deep Learning (DL) system for spectrum monitoring in heterogeneous networks was posed and successfully addressed.

To achieve this, a compact multi-task Convolutional Neural Network (LwM-CNN) was proposed. Crucially, we also developed a training methodology based on a hybrid dataset, combining synthetic references with real-world Software-Defined Radio (SDR) recordings to overcome the "Sim-to-Real Gap."

Experimental results demonstrate that the proposed system (LwM-CNN) achieves an accuracy of 94.0%, which is comparable to the resource-intensive ResNet-50 baseline. Furthermore, our model significantly outperforms ResNet in robustness at low Signal-to-Noise Ratios (SNRs) (82% vs. 68% at -10 dB) and is nearly 20 times more computationally efficient.

The results prove that direct training on hybrid data is an effective and practical strategy for developing real-time DL-based spectrum monitoring systems. Future research will focus on model compression for deployment on embedded platforms and expanding the dataset with new signal classes.

5. References

[1] Zhu, J., Waltho, A., Yang, X., & Guo, X. (2007, August). Multi-radio coexistence: Challenges and opportunities. In *2007 16th International Conference on Computer Communications and Networks* (pp. 358-364). IEEE.
[2] Sharma, N., & Kaur, A. (2010, November). Coexistence: Threat to the Performance of Heterogeneous Network. In *AIP Conference Proceedings* (Vol. 1324, No. 1, pp. 90-94). American Institute of Physics.
[3] Jha, U. S. (2002). Wireless Landscape–Need for Seamless Connectivity. *Wireless Personal Communications*, *22*(2), 275-283.
[4] Zheng, L., Lops, M., Eldar, Y. C., & Wang, X. (2019). Radar and communication coexistence: An overview: A review of recent methods. *IEEE Signal Processing Magazine*, *36*(5), 85-99.
[5] Su, Y., Lin, Y., Liu, S., Liwang, M., Liao, X., Wu, T., ... & Wang, X. (2024). Coexistence of hybrid VLC-RF and Wi-Fi for indoor wireless communication systems: an intelligent approach. *IEEE Transactions on Network and Service Management*.
[6] Omotere, O., Fuller, J., Qian, L., & Han, Z. (2018, August). Spectrum occupancy prediction in coexisting wireless systems using deep learning. In *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)* (pp. 1-7). IEEE.
[7] Miller, R. D., Kokalj-Filipovic, S., Vanhoy, G., & Morman, J. (2019, November). Policy based synthesis: Data generation and augmentation methods for rf machine learning. In *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)* (pp. 1-5). IEEE.
[8] O'shea, T. J., & West, N. (2016, September). Radio machine learning dataset generation with gnu radio. In *Proceedings of the GNU radio conference* (Vol. 1, No. 1).
[9] Kattenborn, T., Leitloff, J., Schiefer, F., & Hinz, S. (2021). Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS journal of photogrammetry and remote sensing*, *173*, 24-49.
[10] Lu, X., Tao, M., Fu, X., Gui, G., Ohtsuki, T., & Sari, H. (2021, September). Lightweight network design based on ResNet structure for modulation recognition. In *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)* (pp. 1-5). IEEE.
[11] Podder, P., Zawodniok, M., & Madria, S. (2024, June). Deep learning for UAV detection and classification via Radio frequency signal analysis. In *2024 25th IEEE International Conference on Mobile Data Management (MDM)* (pp. 165-174). IEEE.
[12] Nguyen, H. N., Vomvas, M., Vo-Huu, T., & Noubir, G. (2021, November). Wideband, real-time spectro-temporal RF identification. In *Proceedings of the 19th ACM international symposium on mobility management and wireless access* (pp. 77-86).
[13] Boegner, L., Vanhoy, G., Vallance, P., Gulati, M., Feitzinger, D., Comar, B., & Miller, R. D. (2022). Large scale radio frequency wideband signal detection & recognition. *arXiv preprint arXiv:2211.10335*.
[14] Fan, C., Yuan, X., & Zhang, Y. J. (2019). CNN-based signal detection for banded linear systems. *IEEE Transactions on Wireless Communications*, *18*(9), 4394-4407.
[15] Scholl, S. (2022). RF signal classification with synthetic training data and its real-world performance. *arXiv preprint arXiv:2206.12967*.
[16] Carlson, A., Skinner, K. A., Vasudevan, R., & Johnson-Roberson, M. (2019). Sensor transfer: Learning optimal sensor effect image augmentation for sim-to-real domain adaptation. *IEEE Robotics and Automation Letters*, *4*(3), 2431-2438.
[17] Hu, C., Hudson, S., Ethier, M., Al-Sharman, M., Rayside, D., & Melek, W. (2022, June). Sim-to-real domain adaptation for lane detection and classification in autonomous driving. In *2022 IEEE Intelligent Vehicles Symposium (IV)* (pp. 457-463). IEEE.
[18] Tandra, R., & Sahai, A. (2008). SNR walls for signal detection. *IEEE Journal of selected topics in Signal Processing*, *2*(1), 4-17.
[19] Jagannath, A., & Jagannath, J. (2021, June). Multi-task learning approach for automatic modulation and wireless signal classification. In *ICC 2021-IEEE International Conference on Communications* (pp. 1-7). IEEE.

[20] Huang, Z., Pemasiri, A., Denman, S., Fookes, C., & Martin, T. (2023, June). Multi-task learning for radar signal characterisation. In *2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)* (pp. 1-5). IEEE.