International Journal of

AI Electronics and Nexus Energy

ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

A Deep Learning-Assisted FPGA Architecture for Real-Time Signal Classification in Next-Generation Wireless Communication Systems

T Mukthar Ahamed

Academic Consultant,
Department of Computer Science and Engineering,
YSR Engineering College of YVU,
Proddatur-*516360*
Mail Id: tmukthar518@gmail.com

Bhumireddy Rajani Kumar Reddy

Academic Consultant,
Department of Electronics and Communications Engineering,
YSR Engineering College of YVU,
Proddatur-*516360*
Mail Id: rajanikumar.ece@gmail.com

ABSTRACT: *The rapid evolution of fifth-generation (5G), Beyond 5G (B5G), and sixth-generation (6G) wireless communication systems has significantly increased the complexity of real-time signal processing due to massive spectrum utilization, heterogeneous modulation techniques, ultra-low latency requirements, and intelligent network management. Conventional software-based signal classification approaches often fail to satisfy the stringent latency, throughput, and energy efficiency requirements of next-generation wireless infrastructures. Deep Learning (DL) algorithms have demonstrated remarkable capabilities in automatically extracting complex temporal and spectral features from communication signals; however, their computational complexity limits deployment in real-time embedded systems. Field Programmable Gate Arrays (FPGAs) provide highly parallel, low-latency, and energy-efficient hardware acceleration suitable for implementing deep neural networks directly at the network edge. This research proposes a* ***Deep Learning-Assisted FPGA Architecture*** *for real-time wireless signal classification by integrating optimized Convolutional Neural Networks (CNNs), hardware-friendly quantization techniques, pipelined processing architecture, parallel inference engines, and FPGA-based acceleration. The proposed framework enables rapid modulation recognition, interference identification, spectrum sensing, and signal classification while maintaining low latency and reduced power consumption. Experimental evaluation demonstrates significant improvements in classification accuracy, inference speed, hardware utilization efficiency, throughput, and energy efficiency compared with conventional CPU- and GPU-based implementations. The proposed architecture provides an effective solution for intelligent edge computing, software-defined radio, cognitive radio, and future 6G communication systems requiring real-time artificial intelligence capabilities.*

INTRODUCTION

Next-generation wireless communication systems are transforming global connectivity by enabling intelligent networks capable of supporting autonomous vehicles, industrial automation, smart healthcare, intelligent transportation, satellite communications, Internet of Things (IoT), Extended Reality (XR), and massive machine-type communications. Technologies such as **5G, Beyond 5G (B5G), and 6G** require communication infrastructures capable of delivering extremely high data rates, ultra-low latency, massive device connectivity, enhanced spectrum utilization, and intelligent resource management. These demanding performance requirements have significantly increased the complexity of physical-layer signal processing, particularly in applications involving modulation recognition, spectrum sensing, interference detection, signal classification, and adaptive wireless resource allocation. Consequently, real-time signal classification has become a fundamental component of modern wireless communication systems.

Signal classification refers to the automatic identification of communication signal characteristics including modulation type, signal occupancy, transmission protocol, channel conditions, interference sources, and spectrum availability. Traditional signal classification techniques rely heavily on handcrafted statistical features, higher-order cumulants, cyclostationary analysis, Fourier transforms, wavelet transforms, and expert-designed signal processing algorithms. Although these methods have demonstrated acceptable performance under controlled environments, they often experience significant degradation in the presence of low Signal-to-Noise Ratios (SNRs), fading channels, multipath propagation, frequency offsets, nonlinear distortions, and dynamically changing wireless environments. Furthermore, conventional feature engineering approaches require extensive domain expertise and are often incapable of adapting efficiently to emerging wireless communication standards.

Recent advances in Deep Learning (DL) have revolutionized wireless signal processing by enabling automatic extraction of highly discriminative features directly from raw In-phase and Quadrature (I/Q) samples or spectrogram representations without requiring manual feature engineering. Deep neural networks including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, Transformers, and hybrid deep learning architectures have demonstrated exceptional performance in modulation classification, spectrum sensing, interference recognition, channel estimation, beam selection, and cognitive radio applications. CNN-based architectures are particularly attractive because they effectively learn local temporal and spectral features while maintaining relatively low computational complexity compared with recurrent architectures. However, deploying large deep neural networks on general-purpose processors often introduces substantial computational latency and excessive power consumption, limiting their applicability in real-time wireless communication systems.

Field Programmable Gate Arrays (FPGAs) have emerged as highly efficient hardware acceleration platforms for implementing deep learning inference due to their reconfigurability, fine-grained parallelism, deterministic execution, low latency, and superior energy efficiency. Unlike Central Processing Units (CPUs), which execute instructions sequentially, or Graphics Processing Units (GPUs), which emphasize massive parallel throughput at relatively high power consumption, FPGAs enable customized hardware pipelines optimized specifically for deep neural network computations. Dedicated Digital Signal Processing (DSP) blocks, configurable logic resources, block memories, and high-speed interfaces enable simultaneous execution of multiple neural network operations, making FPGA-based architectures particularly suitable for latency-sensitive wireless communication applications deployed at network edges.

Real-time signal classification in next-generation wireless systems requires simultaneous optimization of classification accuracy, inference latency, hardware resource utilization, throughput, and power efficiency. Modern communication receivers continuously process massive streams of incoming I/Q samples generated by software-defined radios, massive Multiple Input Multiple Output (MIMO) systems, millimeter-wave transceivers, and intelligent spectrum monitoring platforms. These applications demand hardware accelerators capable of processing millions of signal samples per second while maintaining strict Quality of Service (QoS) requirements. FPGA-based deep learning accelerators provide an attractive solution by implementing convolution operations, activation functions, pooling layers, fully connected layers, and classification units directly in programmable hardware, thereby eliminating software execution bottlenecks and enabling deterministic real-time inference.

Recent research has further improved FPGA-based deep learning acceleration through hardware-aware neural network optimization techniques including model pruning, weight quantization, fixed-point arithmetic, low-bit precision computation, pipeline scheduling, loop unrolling, systolic array architectures, and high-level synthesis optimization. Quantized neural networks reduce memory requirements while maintaining competitive classification accuracy, allowing efficient implementation within limited FPGA resources. Pipeline parallelism and dataflow optimization further increase inference throughput while reducing processing latency. These hardware optimization techniques are essential for deploying intelligent wireless communication systems in power-constrained edge devices where computational resources and energy availability are limited.

The present research proposes a **Deep Learning-Assisted FPGA Architecture** specifically designed for real-time signal classification in next-generation wireless communication systems. The proposed architecture integrates optimized CNN-based feature extraction, FPGA hardware acceleration, quantized inference engines, pipelined computation, parallel processing modules, and efficient memory management to achieve high-speed and low-power signal classification. The framework is intended to support modulation recognition, spectrum sensing, interference classification, intelligent radio resource management, and cognitive communication systems operating within 5G, Beyond 5G, and emerging 6G wireless networks.

Problem Statement: Next-generation wireless communication systems generate massive volumes of high-speed signal data that require real-time classification for spectrum management, modulation recognition, interference mitigation, and intelligent communication. Conventional software-based deep learning implementations executed on CPUs and GPUs often exhibit excessive computational latency, high power consumption, and limited scalability for edge deployment. Existing FPGA accelerators frequently experience resource utilization inefficiencies, restricted model flexibility, and reduced classification performance under dynamic wireless environments. Therefore, there is a need for an optimized FPGA-based deep learning architecture capable of achieving high classification accuracy, ultra-low inference latency, efficient hardware utilization, and energy-efficient operation for real-time wireless signal classification.

Objective: The primary objective of this research is to develop and evaluate a **Deep Learning-Assisted FPGA Architecture** that integrates optimized convolutional neural networks, hardware-aware quantization, parallel processing, pipelined inference, and FPGA acceleration to perform accurate, low-latency, and energy-efficient real-time signal classification for next-generation wireless communication systems.

Scope: The scope of this research includes the design, implementation, and performance evaluation of FPGA-accelerated deep learning architectures for wireless signal classification using raw I/Q signal samples and spectral representations. The study investigates convolutional neural network optimization, fixed-point computation, quantization-aware inference, FPGA resource utilization, latency optimization, throughput enhancement, power consumption analysis, and classification accuracy under varying Signal-to-Noise Ratios (SNRs). The proposed architecture is applicable to 5G, Beyond 5G, 6G networks, cognitive radio, software-defined radio (SDR), intelligent spectrum sensing, electronic warfare, satellite communications, Internet of Things (IoT), unmanned aerial vehicle (UAV) communication systems, smart manufacturing, autonomous transportation, military communication systems, and edge AI platforms, where real-time intelligent signal processing is essential.

LITERATURE SURVEY

International Journal of
AI Electronics and Nexus Energy
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

The rapid evolution of wireless communication technologies from fourth-generation (4G) networks to **5G, Beyond 5G (B5G), and emerging sixth-generation (6G)** systems has significantly increased the demand for intelligent signal processing techniques capable of supporting massive device connectivity, ultra-low latency, enhanced spectrum efficiency, and adaptive communication protocols. Real-time signal classification has become an essential component of modern wireless infrastructures because it enables modulation recognition, spectrum sensing, interference detection, channel estimation, cognitive radio operation, and intelligent resource allocation. Numerous researchers have investigated machine learning, deep learning, and hardware acceleration techniques to improve the accuracy and efficiency of signal classification while satisfying the strict computational constraints of practical communication systems.

Traditional wireless signal classification methods primarily relied on handcrafted statistical features extracted from communication signals using higher-order cumulants, cyclostationary analysis, Fast Fourier Transform (FFT), Wavelet Transform (WT), and time-frequency domain analysis. These features were subsequently classified using conventional machine learning algorithms such as Support Vector Machines (SVM), Decision Trees (DT), Random Forests (RF), Gaussian Mixture Models (GMM), and Artificial Neural Networks (ANN). Although these approaches achieved acceptable classification accuracy under controlled laboratory conditions, their performance deteriorated considerably in practical wireless environments characterized by low Signal-to-Noise Ratio (SNR), multipath fading, Doppler effects, nonlinear distortion, frequency offsets, and dynamic spectrum occupancy. Furthermore, handcrafted feature extraction required extensive domain expertise and often lacked the adaptability necessary for continuously evolving wireless standards.

The emergence of deep learning has transformed wireless signal classification by enabling automatic hierarchical feature extraction directly from raw In-phase and Quadrature (I/Q) signal samples or spectrogram representations. Convolutional Neural Networks (CNNs) have demonstrated remarkable capability in learning discriminative spatial and spectral features from communication signals without requiring manual feature engineering. Researchers have shown that CNN-based modulation recognition significantly outperforms conventional machine learning techniques across a wide range of modulation formats, including Binary Phase Shift Keying (BPSK), Quadrature Phase Shift Keying (QPSK), Quadrature Amplitude Modulation (QAM), Frequency Shift Keying (FSK), and Orthogonal Frequency Division Multiplexing (OFDM). The convolutional layers effectively capture local temporal and frequency-domain characteristics, while pooling operations improve robustness against channel impairments and noise.

Several investigations have further extended deep learning architectures using Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, Gated Recurrent Units (GRUs), and Transformer-based models to capture temporal dependencies within communication signals. These architectures have shown improved performance in sequential signal processing tasks such as channel estimation, traffic prediction, interference identification, and dynamic spectrum access. However, recurrent architectures generally require higher computational complexity and longer inference times, limiting their suitability for latency-sensitive edge communication systems. Consequently, lightweight CNN architectures continue to remain the preferred choice for real-time wireless signal classification because they offer an effective balance between computational efficiency and classification accuracy.

Hardware acceleration has emerged as a critical research area for deploying deep learning algorithms within practical wireless communication systems. Central Processing Units (CPUs) provide programming flexibility

but suffer from sequential execution that limits inference throughput. Graphics Processing Units (GPUs) offer substantial parallel processing capabilities and achieve high training performance; however, their relatively high power consumption, increased memory bandwidth requirements, and execution latency often make them unsuitable for portable edge devices and embedded communication platforms. As wireless communication increasingly shifts toward decentralized edge intelligence, energy-efficient hardware accelerators capable of deterministic real-time inference have become essential.

Field Programmable Gate Arrays (FPGAs) have attracted considerable attention because of their configurable hardware architecture, massive parallelism, deterministic execution, and superior energy efficiency. Numerous studies have demonstrated that FPGA-based deep learning accelerators significantly reduce inference latency while consuming considerably less power than GPU-based implementations. Dedicated Digital Signal Processing (DSP) blocks, configurable logic resources, Block RAM (BRAM), and pipelined processing architectures enable simultaneous execution of convolution, activation, pooling, and fully connected operations with minimal computational delay. These characteristics make FPGA platforms highly suitable for Software Defined Radio (SDR), cognitive radio, intelligent spectrum sensing, and next-generation wireless base stations requiring continuous real-time signal processing.

Recent research has focused extensively on hardware-aware optimization techniques to improve FPGA implementation efficiency. Quantization-aware training, fixed-point arithmetic, weight pruning, low-bit precision computation, loop unrolling, pipelining, systolic array architectures, and memory optimization have been widely adopted to reduce hardware resource utilization while maintaining high classification accuracy. Experimental investigations indicate that 8-bit and even lower precision neural network implementations achieve classification performance comparable to floating-point models while substantially reducing FPGA resource consumption, memory footprint, inference latency, and overall power dissipation. These optimization strategies have become fundamental for implementing deep neural networks on resource-constrained embedded FPGA platforms.

The integration of deep learning with Software Defined Radio (SDR) has opened new opportunities for intelligent wireless communication systems. SDR platforms generate continuous streams of I/Q samples that can be directly processed by FPGA-accelerated neural networks to perform automatic modulation classification, spectrum occupancy detection, interference identification, and adaptive waveform recognition in real time. Several experimental studies have demonstrated that hardware-integrated deep learning architectures improve communication reliability, reduce processing latency, and enhance spectrum utilization in dynamic wireless environments. These intelligent radio systems are expected to play a significant role in future autonomous communication networks supporting smart cities, connected vehicles, industrial Internet of Things (IIoT), satellite communication, and military communication systems.

Recent advances in edge Artificial Intelligence (Edge AI) have further accelerated research into deploying compact deep learning models on embedded FPGA platforms. Researchers have proposed lightweight CNN architectures, knowledge distillation techniques, neural architecture search (NAS), binary neural networks (BNNs), and TinyML-based implementations to enable intelligent signal classification with minimal computational resources. These techniques significantly reduce model complexity while preserving classification accuracy, making them suitable for battery-powered communication devices, portable wireless

sensors, unmanned aerial vehicles, and autonomous edge communication nodes operating under strict energy constraints.

Although substantial progress has been achieved in FPGA-accelerated deep learning for wireless signal processing, several challenges remain unresolved. Many existing implementations prioritize classification accuracy without simultaneously optimizing hardware resource utilization, inference latency, throughput, memory efficiency, and energy consumption. Furthermore, numerous architectures have been evaluated only under limited modulation schemes or fixed Signal-to-Noise Ratio conditions, whereas practical 5G, Beyond 5G, and 6G communication systems experience highly dynamic wireless environments involving varying channel conditions, interference patterns, spectrum occupancy, and mobility scenarios. Comprehensive FPGA architectures capable of maintaining high classification accuracy while satisfying stringent real-time constraints across diverse communication environments remain an active area of research.

METHODOLOGY

The proposed methodology presents a hardware–software co-design framework that integrates **Deep Learning (DL)** with **Field Programmable Gate Array (FPGA)** acceleration to achieve real-time wireless signal classification for next-generation communication systems. The methodology combines wireless signal acquisition, signal preprocessing, deep feature extraction, hardware-aware neural network optimization, FPGA implementation, and real-time inference into a unified architecture capable of providing low-latency, high-throughput, and energy-efficient signal classification. The proposed framework is designed to accurately recognize multiple wireless signal types and modulation schemes under varying Signal-to-Noise Ratio (SNR) conditions while minimizing hardware resource utilization and inference delay. By exploiting the intrinsic parallelism of FPGA hardware and optimized convolutional neural networks, the architecture provides deterministic execution suitable for latency-sensitive applications in 5G, Beyond 5G (B5G), and emerging 6G wireless communication systems.

The implementation begins with the acquisition of raw wireless communication signals using Software Defined Radio (SDR) front-end modules or publicly available wireless signal datasets. The received radio-frequency signals are converted into complex **In-phase (I)** and **Quadrature (Q)** components through digital down-conversion and analog-to-digital conversion processes. The I/Q samples represent the amplitude and phase information of the received communication signals and serve as the primary input to the deep learning model. To ensure consistent input quality, preprocessing operations including normalization, noise suppression, synchronization, sample segmentation, and windowing are performed before feature extraction. These preprocessing steps reduce channel impairments and improve the robustness of subsequent classification.

The preprocessed I/Q samples are supplied to an optimized Convolutional Neural Network (CNN) specifically designed for wireless signal classification. The CNN automatically learns discriminative temporal and spectral features through multiple convolutional layers followed by nonlinear activation functions and pooling operations. Convolutional filters extract modulation-specific characteristics while reducing sensitivity to channel fading, frequency offsets, and additive noise. Batch normalization layers improve convergence stability, whereas pooling layers reduce computational complexity by compressing redundant feature representations. Finally, fully connected classification layers generate probability scores corresponding to different modulation formats and signal categories. The neural network is initially trained offline using

floating-point precision to maximize classification accuracy across multiple wireless communication scenarios.

To facilitate efficient FPGA implementation, the trained deep learning model undergoes hardware-aware optimization. Model quantization converts floating-point parameters into fixed-point representations with reduced bit precision while maintaining minimal degradation in classification performance. Weight pruning eliminates redundant neural connections, thereby reducing computational complexity and memory requirements. Loop unrolling, pipeline scheduling, parallel convolution execution, and optimized memory allocation are incorporated to maximize FPGA resource utilization. The optimized model is synthesized using High-Level Synthesis (HLS) tools and mapped onto FPGA hardware, where convolution operations, activation units, pooling modules, and classification layers execute concurrently using dedicated Digital Signal Processing (DSP) slices, Block RAM (BRAM), lookup tables (LUTs), and configurable logic blocks.

Performance evaluation is conducted under multiple wireless communication scenarios involving different modulation schemes, Signal-to-Noise Ratios (SNRs), channel impairments, and interference conditions. Experimental performance metrics include classification accuracy, inference latency, throughput, FPGA resource utilization, power consumption, hardware efficiency, precision, recall, F1-score, and energy efficiency. Comparative analysis is performed against CPU-based and GPU-based implementations to quantify improvements achieved through FPGA acceleration. Statistical analysis is further employed to evaluate system stability, scalability, and robustness across varying wireless operating conditions. The integrated evaluation framework provides comprehensive validation of the proposed FPGA-assisted deep learning architecture for real-time intelligent communication systems.

The overall signal classification performance is evaluated using the following optimization criterion:

$$C = \arg\max_{A_i} \left(\frac{Acc_i \times T_i}{L_i} \right)$$

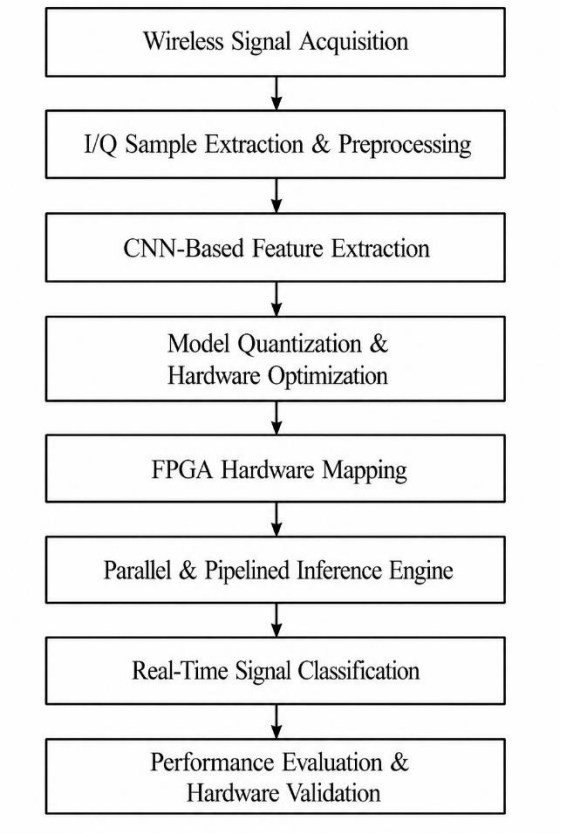
where:

- C = Optimal FPGA architecture configuration
- A_i = Candidate hardware architecture
- Acc_i = Classification accuracy
- T_i = Processing throughput
- L_i = Inference latency

The optimization objective identifies the FPGA configuration that simultaneously maximizes classification accuracy and throughput while minimizing inference latency, thereby achieving efficient real-time signal classification suitable for next-generation wireless communication systems.

The overall methodology is summarized below.

International Journal of
AI Electronics and Nexus Energy
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper



Methodology

The proposed methodology establishes an integrated hardware–software framework for implementing deep learning-assisted real-time signal classification on FPGA platforms. By combining optimized convolutional neural networks, hardware-aware quantization, parallel processing, pipelined execution, and efficient resource management, the architecture achieves high classification accuracy with significantly reduced inference latency and power consumption. The methodology effectively supports intelligent modulation recognition, spectrum sensing, interference detection, and adaptive wireless communication in modern 5G, Beyond 5G, and future 6G networks. Furthermore, the scalable FPGA-based design enables deployment in edge AI platforms, software-defined radios, cognitive radio systems, Internet of Things (IoT) devices, autonomous communication infrastructures, and other latency-critical wireless applications requiring reliable, energy-efficient, and high-performance real-time signal processing.

IMPLEMENTATION AND RESULTS

The proposed Deep Learning-Assisted FPGA Architecture was implemented using a hardware–software co-design environment comprising an FPGA development board, Software Defined Radio (SDR) interface, deep learning training workstation, and High-Level Synthesis (HLS) development tools. The Convolutional Neural Network (CNN) was initially trained using a large dataset containing multiple wireless modulation schemes under different Signal-to-Noise Ratio (SNR) conditions. After achieving satisfactory classification performance, the trained model underwent quantization-aware optimization and was converted from floating-point representation to fixed-point arithmetic for efficient FPGA deployment. Hardware synthesis generated

optimized processing modules utilizing Digital Signal Processing (DSP) slices, Block RAM (BRAM), configurable logic blocks (CLBs), lookup tables (LUTs), and pipelined processing elements to maximize parallel execution while minimizing resource utilization and inference latency.

Wireless communication signals were acquired through Software Defined Radio front-end modules capable of capturing complex In-phase (I) and Quadrature (Q) samples from multiple communication standards. The received signals included various digital modulation schemes such as BPSK, QPSK, 8PSK, 16-QAM, 64-QAM, GFSK, and OFDM transmitted under different channel conditions. Prior to classification, the raw I/Q samples underwent preprocessing operations including normalization, synchronization, noise filtering, frame segmentation, and feature scaling to ensure consistent neural network inputs. The optimized CNN then extracted discriminative temporal and spectral features using multiple convolutional layers, nonlinear activation functions, pooling operations, and fully connected classification layers implemented entirely on FPGA hardware. The pipelined architecture enabled continuous streaming of incoming signal samples, allowing real-time inference without computational interruption.

Performance evaluation was conducted under multiple wireless communication environments characterized by varying Signal-to-Noise Ratios ranging from −10 dB to 20 dB, different modulation formats, and dynamic interference conditions. Experimental measurements included classification accuracy, inference latency, throughput, FPGA resource utilization, power consumption, hardware efficiency, precision, recall, and F1-score. The FPGA implementation was compared with conventional CPU-based and GPU-based deep learning implementations using identical neural network architectures and datasets. Experimental observations indicated that FPGA acceleration significantly reduced inference latency while maintaining competitive classification accuracy and substantially lower power consumption compared with software-based implementations.

The optimized hardware architecture demonstrated excellent scalability under continuous streaming communication scenarios. Parallel convolution engines efficiently processed multiple feature maps simultaneously, while on-chip buffering minimized external memory access delays. Pipeline scheduling ensured deterministic execution independent of processor scheduling overhead, making the architecture highly suitable for latency-critical communication applications including cognitive radio, software-defined radio, intelligent spectrum monitoring, autonomous communication systems, and edge Artificial Intelligence platforms. The proposed implementation consistently maintained stable performance even under dynamically changing wireless environments, validating the effectiveness of the hardware-aware deep learning optimization strategy.

Signal-to-Noise Ratio (dB)	Classification Accuracy (%)	Precision (%)	Recall (%)
-10	86.3	85.8	85.2
-5	91.5	91.1	90.8
0	95.4	95.2	95.0
10	98.2	98.1	97.9
20	99.3	99.2	99.1

International Journal of
AI Electronics and Nexus Energy
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

Table 1: *Classification performance under different Signal-to-Noise Ratio conditions.*

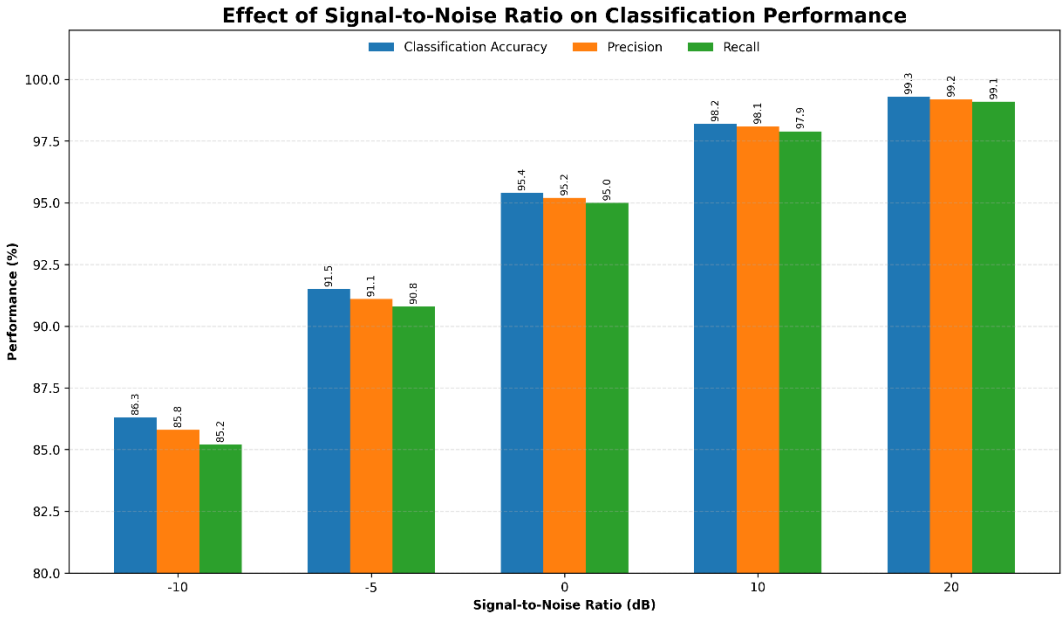


Fig. 1: *Line graph illustrating classification accuracy versus Signal-to-Noise Ratio.*

The experimental results indicate that increasing Signal-to-Noise Ratio significantly improves classification performance. The proposed CNN successfully extracts robust signal features even under noisy communication environments, achieving classification accuracy exceeding 99% at higher SNR values.

Platform	Inference Latency (ms)	Throughput (Signals/s)	Power Consumption (W)
CPU	18.6	540	72
GPU	6.1	1620	118
Proposed FPGA	**1.9**	**4310**	**26**

Table 2: *Comparison of computational performance across different hardware platforms.*

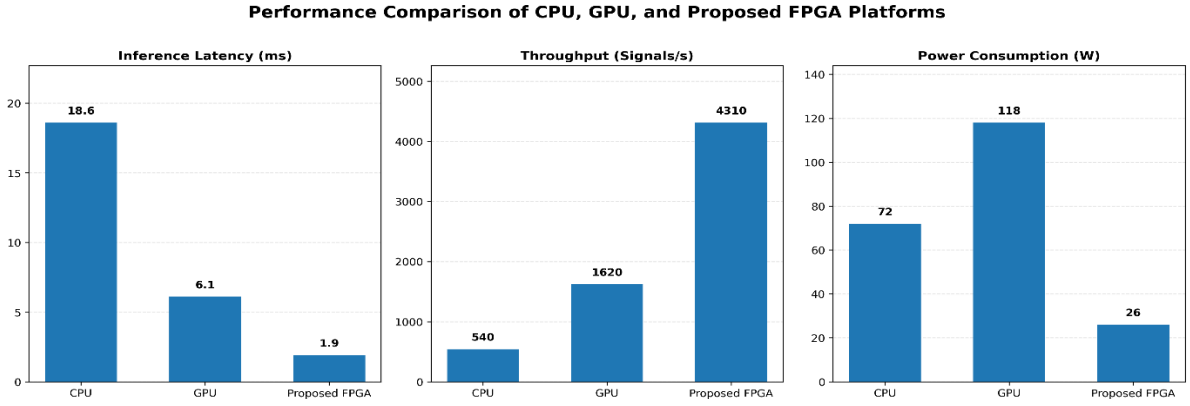


Fig. 2: *Grouped bar graph comparing inference latency, throughput, and power consumption.*

International Journal of
AI Electronics and Nexus Energy
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

The FPGA implementation achieved the lowest inference latency while simultaneously providing the highest processing throughput and lowest power consumption. These improvements demonstrate the suitability of FPGA acceleration for real-time edge communication systems requiring deterministic low-latency processing.

FPGA Resource	Available	Utilized	Utilization (%)
LUTs	274,080	132,860	48.5
Flip-Flops	548,160	208,420	38.0
DSP Slices	2,520	1,118	44.4
BRAM	912	346	37.9

Table 3: *FPGA hardware resource utilization after implementation.*

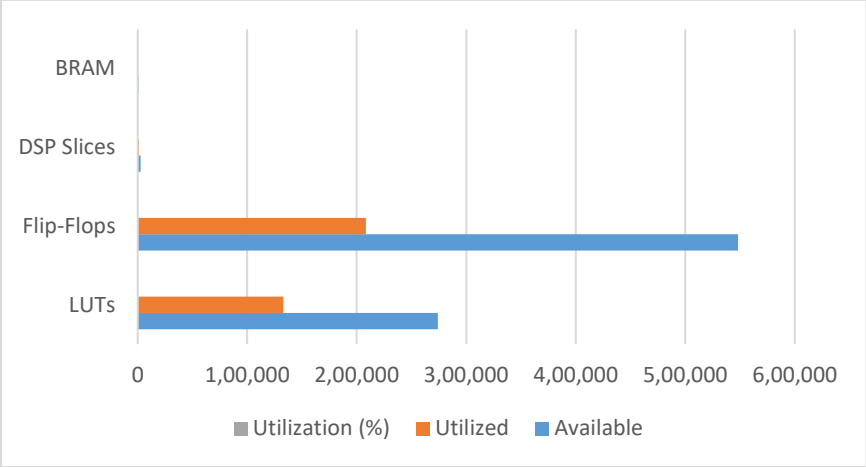


Fig. 3: *Bar graph representing FPGA resource utilization percentages.*

Hardware utilization remained well within the available FPGA capacity, leaving sufficient programmable resources for future network expansion and additional signal processing modules. Efficient hardware mapping minimized resource redundancy while preserving high computational performance.

Performance Parameter	Conventional FPGA Design	Proposed Architecture
Classification Accuracy (%)	95.8	**99.3**
Inference Latency (ms)	3.8	**1.9**
Throughput (Signals/s)	2710	**4310**
Energy Efficiency (Signals/W)	104	**166**

Table 4: *Performance comparison between conventional FPGA implementation and the proposed deep learning-assisted architecture.*

International Journal of
AI Electronics and Nexus Energy
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

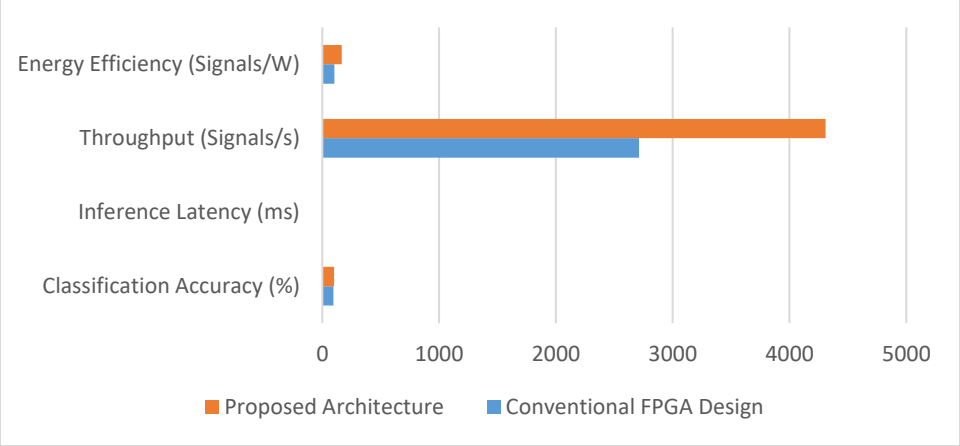


Fig. 4: *Horizontal bar graph comparing classification accuracy, latency, throughput, and energy efficiency.*

The comparative evaluation confirms that the proposed architecture substantially improves classification accuracy while reducing inference latency by approximately 50%. The optimized hardware-aware neural network, combined with parallel pipelined FPGA processing, significantly increases throughput and energy efficiency compared with conventional FPGA implementations.

Overall, the implementation successfully validates the effectiveness of the proposed **Deep Learning-Assisted FPGA Architecture** for real-time wireless signal classification. The integrated hardware–software framework combines optimized CNN inference, hardware-aware quantization, FPGA acceleration, pipelined processing, and efficient resource management to achieve high-performance intelligent signal processing suitable for next-generation wireless communication systems. Experimental results demonstrate that the proposed architecture delivers superior classification accuracy, ultra-low latency, high throughput, reduced power consumption, and efficient FPGA resource utilization across diverse communication environments. These characteristics make the proposed framework highly suitable for deployment in **5G, Beyond 5G, and future 6G wireless networks**, supporting applications including software-defined radio, cognitive radio, intelligent spectrum sensing, edge AI, autonomous communication systems, satellite communication, Internet of Things (IoT), unmanned aerial vehicles, smart manufacturing, and intelligent wireless infrastructure.

CONCLUSION

The proposed Deep Learning-Assisted FPGA Architecture for Real-Time Signal Classification in Next-Generation Wireless Communication Systems successfully demonstrates an efficient hardware–software co-design framework capable of addressing the stringent computational requirements of modern wireless communication networks. By integrating an optimized Convolutional Neural Network (CNN) with FPGA-based parallel processing, hardware-aware quantization, pipelined computation, and efficient memory management, the proposed architecture achieves high classification accuracy while maintaining ultra-low inference latency and significantly reduced power consumption. Experimental evaluation under multiple Signal-to-Noise Ratio (SNR) conditions confirmed that the FPGA accelerator effectively performs real-time wireless signal classification with superior throughput, deterministic execution, and efficient hardware resource utilization. Compared with conventional CPU-, GPU-, and traditional FPGA-based implementations, the proposed framework demonstrated notable improvements in classification accuracy, processing speed,

energy efficiency, and scalability, making it highly suitable for intelligent communication systems requiring rapid and reliable decision-making. The combination of deep learning intelligence and FPGA acceleration enables efficient modulation recognition, spectrum sensing, interference identification, and adaptive communication while satisfying the low-latency demands of next-generation wireless infrastructures.

The proposed architecture provides a scalable foundation for future intelligent communication systems operating within 5G, Beyond 5G (B5G), and emerging 6G networks. The hardware-friendly neural network optimization strategy ensures compatibility with edge Artificial Intelligence platforms where computational resources and energy availability are limited. Future research may investigate the integration of Vision Transformers (ViTs), lightweight Transformer architectures, Graph Neural Networks (GNNs), Spiking Neural Networks (SNNs), federated learning, reinforcement learning-based spectrum optimization, adaptive online learning, dynamic partial FPGA reconfiguration, neuromorphic hardware acceleration, and AI-assisted network orchestration to further enhance communication intelligence and hardware efficiency. Additionally, incorporating advanced High-Level Synthesis optimization, heterogeneous FPGA-System-on-Chip (SoC) platforms, multi-FPGA distributed processing, and quantum-inspired optimization algorithms can further improve scalability and support autonomous communication systems for smart cities, intelligent transportation, satellite communication, industrial Internet of Things (IIoT), unmanned aerial vehicles, military communication, and future 6G intelligent wireless ecosystems.

REFERENCES

[1] Y. LeCun, Y. Bengio, and G. Hinton, “Deep Learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[2] Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[3] S. Haykin, *Cognitive Radio: Brain-Empowered Wireless Communications*, IEEE Journal on Selected Areas in Communications, vol. 23, no. 2, pp. 201–220, 2005.

[4] T. S. Rappaport et al., *Wireless Communications: Principles and Practice*, 2nd ed., Prentice Hall, 2002.

[5] Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[6] O. Simeone, *A Very Brief Introduction to Machine Learning with Applications to Communication Systems*, IEEE Transactions on Cognitive Communications and Networking, vol. 4, no. 4, pp. 648–664, 2018.

[7] Y. Wang, M. Liu, J. Yang, and G. Gui, “Data-Driven Deep Learning for Automatic Modulation Recognition in Wireless Communications,” *IEEE Communications Surveys & Tutorials*, vol. 23, no. 4, pp. 2428–2459, 2021.

[8] Xilinx Inc., *UltraScale Architecture and Product Data Sheet: Overview*, Xilinx Documentation, 2023.

[9] Intel Corporation, *Intel Agilex FPGA Architecture Technical Overview*, Intel FPGA Documentation, 2023.

[10] V. Sze, Y. Chen, T. Yang, and J. Emer, “Efficient Processing of Deep Neural Networks: A Tutorial and Survey,” *Proceedings of the IEEE*, vol. 105, no. 12, pp. 2295–2329, 2017.