A Lightweight Deep Learning-Based Approach for Automatic Modulation Classification Using Multi Modal Data

by

Istiaque Ahmed
21301589

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
Brac University
June 2025.

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Istiaque Ahmed
21301589

Approval

The thesis titled "**A Lightweight Deep Learning-Based Approach for Automatic Modulation Classification Using Multi-Modal Data**" submitted by **Istiaque Ahmed (21301589)** of Spring, 2025 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on June 20, 2025.

Examining Committee:

Supervisor:
(Member)

Md. Ashraful Alam, PhD

Associate Professor
Department of Computer Science & Engineering
Brac University

Thesis Coordinator:
(Member)

Dr. Md. Golam Rabiul Alam

Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD

Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

Automatic Modulation Classification (AMC) is a vital task in modern communication systems, enabling efficient spectrum utilization and reliable signal demodulation. This thesis proposes a lightweight deep learning-based approach for AMC using multi-modal signal data. The methodology involves generating radio signals with eight distinct modulation types (six digital and two analog) using MATLAB's Communications Toolbox and capturing them with ADALM-PLUTO software-defined radios. To enhance signal diversity and realism, channel impairments such as AWGN, Rician fading, and clock offsets are applied. Each signal is transformed into time-frequency representations using the Continuous Wavelet Transform (CWT), separating amplitude and phase components into grayscale scalograms. These dual-modal images are then classified using a custom dual-stream Convolutional Neural Network (CNN), designed to be computationally efficient while maintaining high accuracy. Experimental results demonstrate that the proposed model achieves 98% classification accuracy with significantly fewer parameters compared to baseline models like DenseNet, EfficientNet, and SqueezeNet. This balance of accuracy and efficiency makes the model well-suited for deployment on resource-constrained edge devices. The findings affirm the effectiveness of wavelet-based image representation and lightweight CNN design in advancing AMC performance.

Keywords: Modulation classification, deep learning, lightweight architecture

Acknowledgement

Firstly, all praise to the Great Allah for whom our thesis have been completed without any major interruption.
Secondly, to our advisor Dr. Md. Ashraful Alam sir for his kind support and advice in our work. He helped us whenever we needed help.
And finally to our parents without their throughout support it may not be possible. With their kind support and prayer we are now on the verge of our graduation.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

In modern communication systems, accurate classification of modulation techniques is critical for ensuring reliable data transmission and reception. Modulation involves altering a carrier signal to encode information, and each modulation scheme introduces distinct variations to the signal in terms of amplitude, phase, or frequency. A receiver must correctly identify the modulation scheme used by the transmitter to apply the appropriate demodulation algorithm and recover the original information. However, the misclassification of modulation schemes can lead to incorrect demodulation, resulting in data corruption or loss. Recently, deep learning architectures have been widely used in this field for their ability to extract features without manual intervention. These methods cannot take signals directly as inputs for classification so they are represented by features, images, sequences, or a combination of such modalities. Image-based approaches have gained much popularity as they can represent features that are harder to extract using other forms of representations. The drawback in this approach is that images are usually fed to Convolutional Neural Networks(CNN) and regular CNNs are computationally very demanding. In this research, we propose a novel method for image-based automatic modulation classification, a lightweight dual-stream CNN that takes the signals amplitude and phase scalograms generated using continuous wavelet transform(CWT). The multimodal method helps our model learn the signals features much faster and efficiently, making the model computationally light while still being accurate. This opens new doors for resource-constraint devices that take advantage of the technology of modulation classification in the signal processing domain.

1.2 Motivation

Conventional deep learning techniques for automatic modulation classification often require high computation time and memory space. In order to meet this challenge, researchers are compelled to delve into light-weight architectures that entail high classification rates and low resource consumption. Recently Image-based approaches have gained popularity in this domain for a number of reasons. First, it allows visual representation of the signals that capture modulation specific patterns that help better determine the characteristics of the signal. Secondly, these approaches can benefit from Transfer learning as the models can be pre-trained and enable fine-tuning. Thirdly, CNN models can be easily modified through adding or removing different filters, layers and activations. These models are also less sensitive to noise and maintain robust performance in various levels of SNRs. Wavelet scalograms are incredibly useful for deep learning models such as the CNNs as they provide 2D representations of the signals, preserving temporal and spectral features. Continuous Wavelet Transform (CWT) serves as a great way to generate these scalograms. CWT has many advantages over FFT and STFT as it gives both time and frequency information simultaneously. It also analyzes transient features like, sudden spikes, frequency shifts, modulations and pulses. In addition, there are options for choosing different mother wavelets, which is very helpful when working on signals with specific characteristics. The amplitude and phase are key features that are used in the signal modulation domain to determine a modulation scheme. In this research, we take advantage of these features through CWT and feeding the resultant scalograms to our multi-input CNN for faster learning, reducing computational overhead and providing better accuracy.

1.3 Problem Statement

The classification of modulation techniques is particularly challenging due to the presence of multiple schemes with similar characteristics. For example, modulation schemes within the same family, such as 16-QAM and 64-QAM, exhibit overlapping features such as bandwidth, symbol rate, and phase continuity, making it difficult for traditional classification methods to distinguish between them. This issue becomes even more prominent in heterogeneous networks, where multiple communication standards coexist, requiring devices to accurately classify and adapt to different modulation schemes for seamless communication. Modern day modulation classifications are handled by deep learning architectures due to their superior performance over traditional methods. However, the trade-off between accuracy and computational overhead has been a major concern for researchers for years. Iot Devices, Drones/UAVs, Military and Tactical Radios, Wireless Sensor Networks(WSNs), Autonomous vehicles etc are examples of edge devices that take advantage of this technology. Optimizing the deep learning architectures would greatly help in making the system deployable on such devices.

1.4 Objective

In many cases where there is a scarcity of resources and infrastructure, deep learning algorithms that use images are hard to deploy because of the memory and computational requirements. For example, if we consider military implementations, hardware-based FPGA boards will not be able to host heavyweight deep learning-based classification schemes. Our objective is to overcome this limitation by developing a fast, robust solution for resource-scarce hardware. For this we use a custom dataset consisting of amplitude and phase scalograms generated from signals so that our proposed model can utilize the multi-modal input to learn faster and provide accurate classification results while being computationally lighter than traditional state-of-the-art models.

1.5 Scopes and Challenges

The lightweight structure of the proposed architecture allows real-time deployment on edge devices. Using multi-modal data (Amplitude and Phase) also increases performance under noisy conditions. The proposed architecture also sets a foundation for future optimizations or fusion-based classification methods. Using transfer learning, the model can serve as a base model for domain adaptation. However, the research comes with challenges. Under various SNRs modulation patterns can vary significantly and may cause inconsistent classification. CNN models also cause a black-box problem, making it difficult to understand why a particular modulation is classified in a certain way. Since the data generated and classified under a virtual simulation, the model may perform differently in real-world scenarios. The research demonstrates the lightweight approach with 8 classes, but the efficiency factor may be compromised when other modulation schemes are introduced as we try to retain accuracy.

Chapter 2

Literature Review

2.1 Preliminaries

Signal modulation is a fundamental topic in wireless communications, cognitive radio, electronic warfare, and spectrum monitoring. It is the process of modifying a carrier signal to encode data. It is typically done by altering the signals amplitude, frequency, phase, or multiple properties combined. Figure 2.1 demonstrates how an input carrier signal is modulated. Common modulation types include Amplitude Modulation (AM), Frequency Modulation (FM), Phase Shift Keying (PSK), and Quadrature Amplitude Modulation (QAM), among others. Each modulation scheme affects the carrier in a unique way to represent digital or analog data. In practical systems, identifying the modulation type is crucial for correctly demodulating and decoding the received signal. Modulation classification algorithms, whether based on traditional likelihood methods or modern machine learning approaches, analyze the characteristics of the received signal, such as its constellation, power spectrum, or time-frequency representation, to determine the most probable modulation scheme. This automatic classification enables adaptive communication systems to optimize performance, maintain interoperability, and detect potential intrusions or interference in real time.

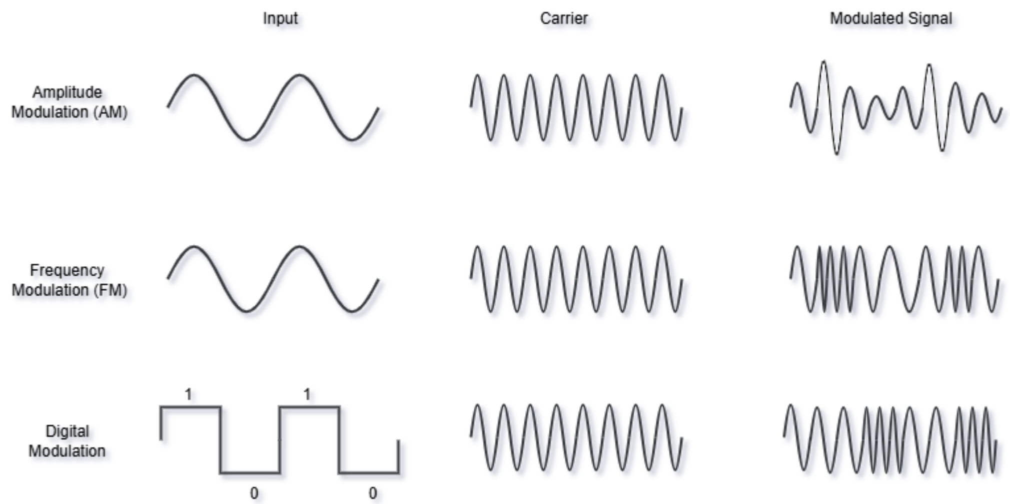

Figure 2.1: Signal Modulation

Continuous Wavelet Transform (CWT) is a powerful signal processing technique used to analyze non-stationary signals those whose frequency characteristics change over time. Unlike the Fourier Transform, which represents a signal in terms of its global frequency content, CWT provides both time and frequency localization, making it ideal for analyzing signals that exhibit transient behaviors or time-varying structures, such as communication signals affected by noise, fading, or interference. At its core, CWT works by decomposing a signal into wavelets, which are small, wave-like functions that are localized in both time and frequency. These wavelets are generated by scaling (stretching or compressing) and translating (shifting) a base function called the mother wavelet. The CWT computes the similarity between the signal and these wavelets at various scales and positions, resulting in a two-dimensional representation called a scalogram. This scalogram visually captures how the frequency content of a signal evolves over time. Mathematically, the CWT of a signal x(t) is defined as:

$$W_x(a,b) = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{|a|}} \psi^* \left(\frac{t-b}{a} \right) dt \tag{2.1}$$

Where, a is the scale parameter (analogous to frequency), b is the time shift parameter and $\psi^*(t)$ is the complex conjugate of the wavelet function.
In practical applications like modulation classification, CWT is used to convert 1D radio signals into 2D time-frequency images, separating phase and amplitude components. These images, known as scalograms, highlight unique patterns and temporal variations in different modulation schemes, making them suitable inputs for deep learning models. The ability of CWT to preserve both time and frequency details makes it especially useful in analyzing complex, real-world communication signals that are affected by channel impairments or environmental noise.

Convolutional Neural Networks (CNNs) are a specialized class of deep learning models designed primarily for processing data with a grid-like structure, such as images, audio spectrograms, or time-frequency representations like scalograms. CNNs are particularly effective at automatically learning spatial hierarchies of features, ranging from simple edges and textures to more complex patterns by using a series of convolutional layers. These layers apply learnable filters (or kernels) that slide over the input data to detect local features, preserving the spatial relationship between pixels. This is followed by pooling layers that reduce the dimensionality, helping to generalize the learned features and reduce computational complexity. Unlike traditional fully connected neural networks, CNNs significantly reduce the number of parameters, making them more efficient and scalable for high-dimensional inputs. In tasks like image classification, object detection, or modulation classification, CNNs excel due to their ability to extract robust and translation-invariant features. For example, in modulation classification, CNNs can be trained on 2D scalograms generated from radio signals to automatically distinguish between different modulation types by learning their unique time-frequency patterns. Their strong feature extraction capability, combined with end-to-end training, makes CNNs one of the most widely used and powerful architectures in modern machine learning. Below is a figure 2.2 of a simple CNN architecture

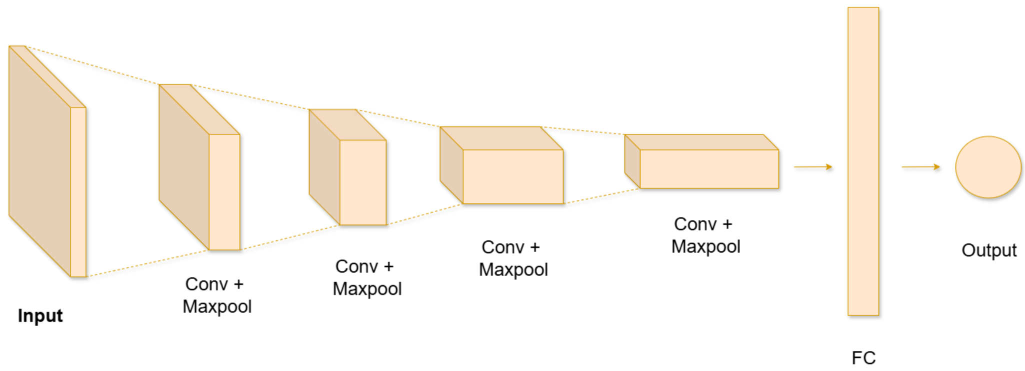

Figure 2.2: Structure of Convolutional Neural Networks

2.2 Review of Existing Research

The traditional methodologies for AMC include the Maximum Likelihood method and the Feature-Based method. According to the available research, likelihood-based approaches assist SDR systems to accomplish reliable and effective transmission through various channels by means of adaptive modulation. These techniques are also useful in environments that are noncoherent in nature where the carrier phase or any other signal characteristics are not known about and hence are useful in the scenarios where communication is not cooperative in nature. The likelihood-based algorithms also offer a general framework for incorporating different degrees of prior information on the data and thus can be applied to many modulation formats [2].
On the other hand, the FB methods provide near-optimal performance compared to the maximum likelihood methods at a much lower complexity cost [4]. These methods can be implemented in conjunction with different classification models such as ANN, SVM and DT hence they are flexible to use in different applications. The properly designed FB algorithms are also comparatively efficient in the classification of the modulation schemes which is helpful in applications such as military intelligence and adaptive modulation systems.
The studies also show that FB methods are easier to apply and require fewer computations as opposed to LB methods. Nevertheless, the LB approaches and especially the ML-based ones, are known to provide the best results. The FB methods are relatively less affected by model errors and other unknown factors, which is advantageous in some cases. However, the selection of the most appropriate method depends on the extent of the trade-off between the described features of complexity, performance, and robustness [1].
These traditional methods have been discovered to pose some serious disadvantages compared to the current methods under Deep Learning [26]. The Likelihood-Based approaches require complex calculation procedures to approximate the likelihood function and model parameters and are not feasible in actual applications. The Feature-Based approaches have limited ability to learn from data due to the weak discriminative power of handcrafted features and the constraints of traditional classification algorithms. Both approaches struggle with the increasing number of modulation formats and the discrimination of higher-order digital modulations. These approaches also have drawbacks in handling strong channel distortions that impact signal quality and the classification accuracy. On the contrary, Deep learning meth-

ods excel in automatic feature extraction, learning underlying characteristics of radio signals for modulation pattern recognition. It has higher learning capacity, which improves classification performance, especially under channel distortions.
Deep learning algorithms can also autonomously extract features avoiding the difficult task of manual feature selection. With these advantages in mind, researchers started utilizing these algorithms in communication systems, specially Convolutional Neural Networks(CNN)[11]. CNNs have layers with constrained connectivity patterns that reduce the number of parameters and allows larger input dimensionality. The filters used in convolutional operations are applied across the entire input which leads to fewer parameters and improved efficiency. Pooling layers in CNNs help the model become invariant to small translations, enhancing the ability to recognize patterns regardless of their position in the input. CNNs can also learn a hierarchy of features, from simple to complex, which is beneficial for tasks like image classification where different levels of abstraction are useful. In comparison with traditional and SVM-based algorithms, the CNN-based model performs well across different SNR regions, matching the accuracy of cumulant-based models in high SNR and SVM-based models in low SNR scenarios without the need for manual feature selection.
Recent research has focused on using convolutional neural networks (CNNs) for modulation classification in wireless communication systems, combining CNNs with various signal processing techniques to improve classification accuracy[36]. The proposed method applies the Radon Transform(RT) to constellation diagrams before feeding them into pre-trained CNNs. The authors tested the data on AlexNet, VGG-16, and VGG-19 models for classification task and demonstrated improved performance in noisy channels.
Abdulkarem et al. (2022) proposed a two-staged method combining CNN with continuous wavelet transform(CWT)[32]. The method extracted time-frequency information from the signals using CWT to generate 2D pictures which were then fed to the Convolutional neural networks as inputs. The experiment included six modulation types: ASK, PSK, FSK, QASK, QPSK and QFSK, all tested under 0-25dB SNR and showed great results.
Daldal et al. (2020) also utilized time-frequency representations and generated 2D images employing short-time Fourier transform. These images served as direct inputs to CNNs for classifying multiple modulation types. for experiment, four popular CNN models have been used (AlexNet, VGG-16, VGG-19, GoogLeNet, ResNet-101) and all of them achieved high classification accuracy (over 99% of accuracy).In fact, the VGG-16 and ResNet-101 provided a classification accuracy of 100%[21].
Al-Makhlasawy et al. (2020) investigated modulation classification in the presence of adjacent channel interference using constellation diagrams as CNN inputs[16]. The studies compared different CNN architectures, including AlexNet and VGG variants, and evaluated performance across various signal-to-noise ratios and fading channel conditions. These research demonstrate the effectiveness of CNN-based approaches for modulation classification.
Hany et al. (2023) proposed three CNN-based AMC models, each using a different classification layer: Mean absolute error-based, Sum of squared errors-based and Cross-entropy-based[37]. The models were tested with three different optimizers at two signal-to-noise ratio (SNR) levels. Results show that the models can achieve up to 100% classification accuracy, depending on the choice of optimizer and loss

function.
Thien et al. (2021) proposed a CNN-based approach for modulation classification with skip-connections and reusable-feature depth-concatenation[25]. The Reusable-Feature Convolutional Neural Network (RefNet) uses skip-connection to prevent vanishing gradients and reusable-feature depth-concatenation to optimize feature utilization. RefNet achieves high accuracy of 85.10% at +10 dB SNR on the DeepSig: RadioML dataset, outperforming traditional methods.
Van-Sang et al. (2020) proposed a deep CNN model can accurately classify modulation formats by learning the constellation map[17]. This method uses gray-scale constellation images of 8 classes to feed the CNN architecture that is specified by multiple processing blocks. The proposed network achieves approximately 87% classification accuracy at 0 dB signal-to-noise ratio (SNR) under multipath Rayleigh fading conditions, outperforming other state-of-the-art models.
Jiacheng et al. (2021) developed a CNN-based method using short-time-Fourier-transformation (STFT) spectrograms.[29] The radio signals are converted into gray-scale STFT spectrograms and then compressed to a fixed size using bi-cubic interpolation. Using a dataset of 16 modulation types, the results show a major improvement over conventional methods—boosting accuracy from 74.3% to 99.0%. Additionally, the model can distinguish between different signals of the same modulation type.
Sun et al. (2022) proposed two machine learning models that eliminate the need for phase and ferquency lock[35]. The models used Constellation Images (CI) and Graphic Representation of Features (GRF), transforming statistical signal properties into spider graphs. The authors used transfer learning with pre-trained CNNs such as Inception-v3[9], GoogleNet[5] and SqueezeNet[7]. The models have been tested with real SDR data and simulated data, and the outcomes demonstrate that while CI performs well on high SNRs, GRF performs better on Low SNRs.
Zhao et al. (2024) introduced a transfer-learning-based approach using spectrum and constellation diagrams for modulation classification. The authors use Efficient-Net[13] and ViT[23] to demonstrate the power of transfer learning capabilities of computer vision models, achieving 8.97% improved performance over traditional methods[44].
Rehman et al, (2025) used eye diagrams and deep convolutional networks for modulation classification[45]. The authors used ResNet-18, ResNet-50[34] and MobileNet-V2[39] which achieved robust performance on low SNRs(-20 to 30dB). The eye diagram approach captures temporal distortions better than constellation diagrams. But, the eye diagrams often fail to capture certain dynamic channel behaviors that could arise in fast varying environments.
Wu et al. (2025) presented a visition transformer model enhanced with multi-scale feature fusion strategies[46] that include vertical shortcut connections and multi-head subspace self-attention (MHSSA). The techniques strengthen the model's ability to learn across different signal characteristics and scales. The MLP classifiers are also replaced by a Kolmogorov-Arnold Network to enhance nonlinear representation and interpretability. The proposed method outperforms conventional CNNs, LSTMs, and other transformer-based models, especially in low SNR environments.
Zhang et al.(2025) combines the temporal and image-based modalities in a novel approach[47]. Through this method, 1D signal inputs (such as IQ, amplitude, and phase) with 2D Markov Transition Field (MTF) images the authors leverage both

local and global characteristics of the signals. The modalities are processed using a LSTM and a focal modulation network (FMN). Experimental results on the RadioML2016.10A dataset demonstrate an accuracy of 92.30% at 0 dB which is superior compared to baseline models like CNN, ResNet, LSTM, CLDNN, and Transformer networks. However, the increase in model parameters (up to 28.64M) may impact inference speed and deployment on resource-constrained systems.

Waqas et al. introduced a novel approach for Automatic Modulation Classification using spectrogram-based deep learning[41]. The I/Q radio signal data are converted into spectrogram images by applying resolution transformations. This sigficantly reduces computational load, upto 99.61% accelerate data processing speed by 8x. Several CNN architectures (e.g., SqueezeNet[7], ResNet-50[34], InceptionResNet-V2[12]) have been used to evaluate the method and achieved 91.2% accuracy under noisy conditions.

Liu et al. challenged the conventional downsampling strategies that are responsible for degrading spatial resolution and hindering classification of closely related modulation schemes[31]. The authors propose a novel Multiscale Dilated Pyramid Module (MDPM) which preserves high-resolution features by employing dilated convolutions and multiscale feature aggregation. This approach achieves classification robustness in noisy environments. The model was adapted with group convolution and channel shuffle to reduce computational overhead and allow deployment in resource constrained platforms.

In recent years, the focus on developing lightweight automatic modulation classification (AMC) techniques has grown due to the increasing demand for efficient and deployable models in resource-constrained environments. Various approaches have been proposed to balance the trade off between model complexity and classification performance. For example, LightAMC combines deep learning with compressive sensing to create a lightweight neural network for AMC [22]. The method involves training a convolutional neural network (CNN) on mixed datasets with varying signal-to-noise ratios (SNRs) and applying neuron pruning via L_1 regularization to remove redundant neurons. However, The training process involves multiple stages, including pruning and fine tuning, which can be complex and time-consuming.

A lightweight AMC must also be cost-efficient and an architecture specifically designed to achieve that is MCNet [18]. The model is designed with specific convolutional blocks that use asymmetric convolutional kernels, reducing the number of parameters while effectively learning spatiotemporal signal correlations. Authors claim that it requires approximately 40% and 45% fewer trainable parameters than ResNet and VGG respectively while having similar inference time with CNN-AMC which is a significantly larger model than MCNet. The inclusion of skip connections helps preserve residual information across multi-scale feature maps and prevents the vanishing gradient problem during training. MCNet achieves an overall classification rate of 93% at 20dB SNR on DeepSig dataset. However, the model is still not light enough to be deployed in resource-constrained environments and it is subject to a trade-off between network capacity and computational complexity.

The MCNet received further improvements as a better CNN was proposed based on its designs [19]. The newer model achieved higher classification accuracy in SNR range from 4 dB to 20 dB, with improvements of 5.52% and 5.92% at 0 dB and 10 dB of SNRs, respectively. The model further reduces the computational complexity by saving 67% of trainable parameters compared to MCNet. It also cuts down the

inference time by 54.4% compared to MCNet, making it more efficient for resource-constrained environments.
The Ultralight Convolutional Neural Network (ULCNN) is designed for UAV systems, balancing accuracy and efficiency with a small model size and fast inference speed [43]. The architecture utilized data augmentation to increase the training sample size and enhance generalization. It employs complex-valued convolution to fuse in-phase and quadrature components of signal samples, capturing correlations. ULCNN integrates separable convolution, channel shuffle and channel attention mechanisms to extract in-depth features while reducing feature dimensionality. It also fuses features from different layers and dynamically adjusts the learning rate during training to align with progress, aiding in convergence and performance. The ULCNN achieves 62.47% accuracy on the RML2016.10a dataset with only 9751 parameters, demonstrating its relevance in the resource-constrained environment like UAV systems.
Stressing the importance of AMC in military and civilian contexts, a model was proposed applying pruning techniques to decrease computational processing and storage burdens and enable the method for EDGE devices with limited capacities[20]. This is achieved through filter-level pruning which is implemented by utilizing an activation maximization (AM) method which helps remove filters that are of minor significance and thus promotes the architecture's accuracy when using the same compression rate as compared to the traditional ones. In this research study the evaluation of the network pruning with respect to the accuracy and computational complexity is done following the VT-CNN2 architecture with adam optimizer and ReLUs. However, pruning can still be disadvantageous in certain contexts since assessing which convolutional filter to prune is more challenging if the deviation of the filter norm is too low.
Another lightweight and efficient architecture LWAMCNet also introduces two strategies to alleviate computational overhead [28]. First, it suggests a residual architecture with depthwise separable convolution (DSC) that prevents the vanishing gradient problem and reduces the computational complexity during feature extraction. Then for feature reconstruction, Global depthwise convolution (GDWConv) is applied after the last convolutional layer to reduce model parameters and inference time. The experiments show that the proposed architecture reduces model parameters by 70-98% and inference time by 30-99% for RadioML2018. 01A and RadioML2016.10A datasets while maintaining high accuracy.
Underlining the performance limitations of the centralized learning AMC method, DecentAMC is an architecture that introduces decentralized learning methods to AMC using model aggregation and lightweight design [24]. It uses a separable convolution neural network (S-CNN) that replaces standard convolution layers, reducing space complexity by about 94% and time complexity by about 96% compared to traditional CNN, with minimal impact on classification accuracy. Multiple edge devices (EDs) are used for joint training, improving training efficiency compared to S-CNN based centralized learning (CentAMC). DecentAMC also shares model weights of S-CNN instead of large datasets or complex CNN model weights which also contributes to reducing communication overhead. However, the training efficiency of this model is not as high and heavily dependent on the number of edge devices. While the model fails to outperform S-CNN based CentAMC in terms of classification accuracy, it significantly reduces space and time complexity making

the tradeoff worth it.

MCMBNN is another lightweight decentralized-learning-based architecture that is suitable for edge devices with limited computing and memory resources[33]. This DecentAMC architecture uses three exclusive blocks (phase estimator and transformer block, spatial feature extraction block and temporal feature extraction and Softmax block) to extract features to achieve robust classification accuracy. Complimentary features of varying channels are also extracted using multichannel input. MCMBNN achieved better accuracy in classification with low model complexity for the classification of all the datasets. It was superior to other state-of-the-art networks particularly in the low SNR region. The method also showed good adaptability to edge devices with less communication overhead. As for the findings in this study, the results confirmed that the proposed MCMBNN and DecentAMC method are capable of realizing decentralized learning with high classification accuracy and efficient low-cost communication.

PDARTS-AMC, a Progressive Differentiable Architecture Search for AMC uses gradient based search strategy to jointly learn the network weights and architecture parameters for the purpose of finding efficient neural network architectures for modulation classification [42]. It performs the architecture design of neural networks without requiring the user's expertise. Moreover, it cuts down the computational complexity and enhances the search effectiveness by employing the approximations of search space and imposing regularization. But, the process of searching for the optimal architecture can still be complex and time-consuming.

The LSTM Auto-Encoder is a novel framework that extracts features from noisy signals and classifies modulation schemes efficiently [27]. The method employs a small model size that can be implemented on low-cost computational platforms like Raspberry Pi and yet obtain high classification accuracy. The auto-encoder has two parts: an encoder that encodes data and produces a code and decoder that decodes the code and produces the original data. The input signals are L2-normalized and some of them are partially masked by setting some of the samples to zero during training. This allows the model to learn and recover signals and also extract stable features. The last hidden state vector of LSTM is fed into the fully connected layers and softmax function to classify the modulation type of the input signal. In comparison with other state-of-the-art AMC models like CNN, CLDNN and LSTM across different SNRs the framework achieves higher top-1 classification accuracy while requiring fewer FLOPs, less memory space and the smallest number of trainable parameters. As for the real-world situations, the framework's performance could be different, since the synthetic and real over-the-air radio data used in the experiments may not be highly representative of real-world signal environments.

The CVResNet is an advanced, yet compact neural network framework that is specifically designed for Automatic Modulation Classification (AMC)[40]. This new architecture combines the complex convolutions with the residual networks to enhance the performance. Due to the use of depthwise separable convolution, it has a simplified structure and reduces the number of parameters and FLOPs by approximately 83. 34% and 83. 77% respectively, and at the same time, maintaining good performance. In addition, it incorporates a new hybrid data augmentation strategy that helps in increasing the classification performance without the shortcomings that are often linked to a more simple design. The CVResNet's shows its potential application in resource-constrained environments due to its ability to reduce computational complexity and parameter count but the experiments does not explore the effect of noise on feature extraction, indicating that it might not be fully robust in practical scenarios.

ShuffleAMC is a CNN-based architecture designed for IoT systems and OFDM that is based on ShuffleNet, a lightweight neural network [30]. The method employs the Fast Fourier Transform (FFT) to preprocess the received OFDM signals for improving classification performance at low SNRs. It uses a Light Convolutional Neural Network (CNN) to reduce the number of model parameters and computational costs. Also, L2 regularization is used in the training process and, as seen from the results, if used with an appropriate regularization parameter, it helps to avoid overfitting, improves the classification performance, and reduces the training time. In terms of classification accuracy, the ShuffleAMC performs very similar to CNN-based AMC method with less than 1% performance gap. But in terms of complexity, the ShuffleAMC uses only 1.8% of the model parameters and 2.4% of the model size than that of CNN based AMC.

To improve both accuracy and efficiency, Lightweight AMC architectures utilize residual learning and Squeeze-and-Excitation (SE) blocks [38]. Research has shown that adding residual learning in the training phases can help to solve the issues arising from deep networks. This is done through shortcut connections that allow layers in the network to learn directly from the identity mapping of their inputs hence increasing classification accuracy. SE blocks also enhance the network's functionality by dynamically fine-tuning channel-wise feature responses. This is done by capturing the dependencies between the channels in a dynamic way, thus enhancing the capacity of the network to represent the topology. The architecture's design focuses on reduction of the number of parameters, and overall lighter architecture, even though it does not affect the performance. Interestingly, it outperforms conventional approaches, especially in conditions characterized by high Signal-to-Noise Ratio (SNR).

2.3 Summary of Key Findings

To put in perspective the effectiveness of the proposed approach, Table 2.1 shows a comparative summary of the most current image-based modulation classification methods. Each paper is classified according to its image representation, network architecture, the strength of modulation types it can compensate for, and performance. During this comparison, it also indicates the distinctiveness of our approach, which is designed to use CWT-based dual-stream CNNs to integrate the amplitude and phase information and thus, reach high recognition accuracy over a broad SNR region, surpassing many state-of-the-art methods in both digital and analog modulation recognition.

Study / Paper	Image Representation	Architecture	Modulation Types	Performance
Abdulkarem et al., 2022[32]	CWT	Custom deep CNN	6 modulations	99% accuracy at 0 dB to 25 dB SNRs
Daldal et al. (2020)[21]	Short-time Fourier transform	AlexNet, VGG-16, VGG-19, GoogLeNet, ResNet-101	5 modulations	over 99% accuracy on most models
Al-Makhlasaw et al., 2020[16]	Constellation diagrams	AlexNet, VGG-16, VGG-19	5 modulations	95% to 98% accuracy across models
Hany et al., 2023[37]	Intermediate frequency (IF) time-domain signals.	Three CNN-based AMC models	11 modulations	up to 100% accuracy on various SNRs
Thien et al. 2021[25]	I/Q time-domain samples	RefNet	24 modulations	85.10% overall classification accuracy
Van-Sang et al. 2020[17]	Gray-scale constellation images	Deep CNN model	8 modulations	87% accuracy at 0 dB
Jiacheng et al. 2021[29]	Gray-scale STFT spectrograms	Custom deep CNN model	16 modulations	99% accuracy
Sun & Ball, 2022[35]	Constellation Images (CI) & Spider Graphs	CNN & ML classifiers	8 digital modulations	86% accuracy at 10 dB SNR
Zhao & Luo, 2024[44]	Spectrum & Constellation Diagrams	Transfer Learning (EfficientNet, ViT)	11 modulations (RadioML)	Accuracy ↑ by 8.97% over baselines
Rehman et al., 2025[45]	Eye Diagrams	ResNet-18/50, MobileNetV2	11 modulations	93.6% accuracy at -20dB SNR
Wu et al., 2025[46]	Time–Frequency Maps	Vision Transformer (MSFF-ViT) & Kolmogorov–Arnold Net	24 modulations (RadioML2018.01A)	94% accuracy at 8 dB SNR, peaks at 97.3% accuracy at 16 dB
Zhang et al., 2025[47]	Multimodal (1D&2D features)	Multi-task CNN & GRU	11 modulations	92.3% accuracy at 0 dB SNR
Waqas et al., 2023[41]	Resolution Transformed Spectrograms	CNN (SqueezeNet, ResNet, DenseNet)	11 modulations	91.2% accuracy in AWGN
Liu et al., 2024[31]	Time–Frequency Images	CNN & Multiscale Dilated Pyramid Module	10 modulations	90% accuracy, 1/8 parameter count compared to traditional models

Table 2.1: Comparison of Image-Based Modulation Classification Techniques

In the recent literatures, increasing application of deep learning in particular, Convolutional Neural Networks (CNNs) in AMC is observed due to their capability of feature learning as well as processing data in a large scale independently. CNNs achieve superior performance against counterpart traditional methods and SVMs at different SNR with no human intervention to extract features. Different researchers have improved the CNN-based AMC by integrating some signal transformation methods, such as the Radon Transform, STFT[9], and CWT[10], which used the signal transformation to transform the one-dimensional (1D) signal into a joint two-dimensional (2D)image. Especially, the CWT and STFT images based models have achieved excellent classification accuracies (often exceeding 99%) on various modulation categories and for different channel conditions. Furthermore, advanced architectural decisions, such as skip connections, feature reuse (RefNet), and the usage of optimal classification layers, have improved model results (LinkNet). Despite these improvements, the state-of-the-art models are mostly large and computationally expensive, making it difficult to be deployed in real-time or on the edge applications. This makes it possible for lightweight CNN architectures, such as the dual-stream architecture proposed here, to find a better trade-off between accuracy and efficiency.

Chapter 3

Signal Generation, Transmission and Data Credibility

This section outlines the rigorous process used to generate, transmit, and collect the signal data used for automatic modulation classification in this study. The methodology combines precise mathematical signal synthesis using MATLAB with realistic wireless transmission via Software Defined Radio (SDR). The resulting dataset closely emulates the behavior of real-world wireless communication systems, thereby ensuring both experimental control and environmental validity.

3.1 Digital Baseband Signal Synthesis

The process begins with the generation of a digital baseband signal using MATLAB. In wireless communications, the baseband signal is a complex-valued time-domain representation of a modulated waveform, expressed as:

$$s[n] = I[n] + jQ[n] \tag{3.1}$$

where $I[n]$ and $Q[n]$ are the in-phase and quadrature-phase components, respectively. These components represent the amplitude and phase variations used to encode binary information.

To construct the signal, a stream of binary bits is grouped into symbols based on the chosen modulation scheme. For instance, QPSK maps two bits per symbol while 64-QAM maps six bits. Each symbol is mapped to a point in the complex constellation according to standardized rules such as Gray coding. This results in a discrete sequence of complex-valued symbols.

To convert this sequence into a continuous waveform suitable for analog transmission, we apply pulse shaping using a Root Raised Cosine (RRC) filter. The baseband signal is obtained by convolving the symbol sequence with the pulse shaping filter:

$$x(t) = \sum_{n=-\infty}^{\infty} s[n] \cdot p(t - nT) \tag{3.2}$$

where $p(t)$ is the impulse response of the RRC filter and T is the symbol period. In MATLAB, this is implemented using:

```
symbols = qammod(bits, M, 'UnitAveragePower', true);
```

```
rrc = rcosdesign(0.35, span, sps);
txWaveform = upfirdn(symbols, rrc, sps);
```

The resulting $x(t)$ is a smooth, complex-valued waveform that accurately represents the modulated baseband signal.
To ensure a comprehensive and challenging dataset, eight modulation schemes are employed in this study: six digital and two analog. The digital modulation schemes are defined as follows:
Binary Phase Shift Keying (BPSK) is the simplest phase modulation scheme where each bit modulates the phase of the carrier. The signal is given by:

$$s(t) = \sqrt{2P}\cos(2\pi f_c t + \pi b), \quad b \in \{0, 1\} \tag{3.3}$$

A binary '1' induces a 180° phase shift compared to '0'.
Quadrature Phase Shift Keying (QPSK) encodes two bits per symbol using four phase shifts:

$$s(t) = \sqrt{2P}\cos(2\pi f_c t + \phi_k), \quad \phi_k \in \left\{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\right\} \tag{3.4}$$

This improves bandwidth efficiency over BPSK by doubling the bit rate.
64-Quadrature Amplitude Modulation (64-QAM) is a high-order scheme that modulates both amplitude and phase:

$$s(t) = I_k \cos(2\pi f_c t) - Q_k \sin(2\pi f_c t) \tag{3.5}$$

where I_k and Q_k are selected from a 64-point constellation, allowing the transmission of 6 bits per symbol.
4-Pulse Amplitude Modulation (PAM4) uses four discrete amplitude levels to represent two bits per symbol:

$$s(t) = A_k p(t - kT), \quad A_k \in \{-3A, -A, A, 3A\} \tag{3.6}$$

It is bandwidth-efficient but sensitive to amplitude distortion and noise.
Gaussian Frequency Shift Keying (GFSK) is a frequency modulation scheme where the carrier frequency varies with the data bits, smoothed by a Gaussian filter. Its instantaneous phase is given by:

$$\phi(t) = 2\pi \int_0^t f(\tau)d\tau = 2\pi \int_0^t f_c + k \cdot m(\tau)\, d\tau \tag{3.7}$$

This smoothing reduces spectral width, making it useful in Bluetooth and low-power applications.
Continuous Phase Frequency Shift Keying (CPFSK) ensures that the phase of the modulated signal changes smoothly, avoiding discontinuities:

$$s(t) = \sqrt{2P}\cos\left(2\pi f_c t + 2\pi h \sum_n b_n q(t - nT)\right) \tag{3.8}$$

where h is the modulation index and $q(t)$ is a continuous-phase shaping function. It is more power-efficient than abrupt phase modulation schemes.

The analog modulation schemes used are:
Broadcast Frequency Modulation (B-FM) varies the frequency of the carrier in proportion to the input message signal:

$$s(t) = A_c \cos\left(2\pi f_c t + 2\pi k_f \int_0^t m(\tau) d\tau\right) \tag{3.9}$$

This provides high noise immunity and is commonly used in FM radio broadcasting.
Single Sideband Amplitude Modulation (SSB-AM) transmits only one sideband of an AM signal. The upper sideband signal is expressed as:

$$s_{\text{USB}}(t) = A_c m(t) \cos(2\pi f_c t) - \hat{m}(t) \sin(2\pi f_c t) \tag{3.10}$$

where $\hat{m}(t)$ is the Hilbert transform of $m(t)$. This reduces bandwidth and is widely used in long-distance communication systems.

3.2 Channel Modeling and Signal Impairments

To simulate realistic channel conditions, we introduce impairments including Additive White Gaussian Noise (AWGN), multipath fading, and clock offset. AWGN models the background thermal noise present in all communication systems:

$$y(t) = x(t) + n(t), \quad n(t) \sim \mathcal{N}(0, \sigma^2) \tag{3.11}$$

To model multipath propagation, we use a Rician fading channel, where the channel response $h(t)$ is a combination of line-of-sight (LOS) and non-line-of-sight (NLOS) components:

$$h(t) = \sqrt{\frac{K}{K+1}} h_{\text{LOS}} + \sqrt{\frac{1}{K+1}} h_{\text{NLOS}} \tag{3.12}$$

where K is the Rician factor.
Clock offset and timing errors are introduced by resampling the signal with a small perturbation to the sampling rate:

$$x_{\text{offset}}(t) = x(t \cdot (1 + \delta)) \tag{3.13}$$

These impairments are either simulated in MATLAB or naturally incurred during real hardware transmission.

3.3 Over-the-Air Transmission via SDR

To enhance realism beyond simulation, the generated waveform is transmitted over-the-air using an ADALM-PLUTO SDR. This device performs digital-to-analog conversion and RF upconversion, emitting the signal through an antenna.
The complex baseband signal is modulated onto a radio frequency carrier f_c using quadrature modulation:

$$x_{\text{RF}}(t) = \Re\left\{x(t) \cdot e^{j2\pi f_c t}\right\} = I(t) \cos(2\pi f_c t) - Q(t) \sin(2\pi f_c t) \tag{3.14}$$

The signal is transmitted at a carrier frequency of 902 MHz. It is subject to environmental effects, hardware non-linearities, and quantization noise-factors that cannot be fully modeled in software alone.
Reception is handled by another SDR (or the same device in loopback mode), which downconverts the signal back to baseband:

$$x_{\mathrm{BB}}(t) = x_{\mathrm{RF}}(t) \cdot e^{-j2\pi f_c t} \tag{3.15}$$

The resulting complex samples are digitized by an ADC and stored for further processing.

Complete Received Signal Model

Combining the modulation, channel effects, and impairments, the final received signal at the receiver can be modeled as:

$$r(t) = \left[s(t) \cdot e^{j(2\pi\Delta f t+\theta)} * h(t)\right] + n(t) \tag{3.16}$$

Here, $s(t)$ is the transmitted baseband signal, Δf is the frequency offset, θ is the random phase offset, $h(t)$ is the multipath channel impulse response, and $n(t) \sim \mathcal{N}(0, \sigma^2)$ represents additive white Gaussian noise. The operator $*$ denotes convolution. This model captures the key real-world distortions introduced by oscillator mismatch, multipath fading, and thermal noise.

3.4 Data Credibility and Validity Justification

Firstly, the underlying signals are generated according to strict mathematical and engineering principles that are foundational to modern wireless communication. The use of standardized modulation schemes such as BPSK, QPSK, and 64-QAM ensures that the generated signals conform to those used in practical systems like LTE, Wi-Fi, and satellite communication [14], [15].
Secondly, the use of MATLAB enables full control over signal characteristics, including modulation depth, symbol rate, pulse shape, and channel impairments. These parameters can be adjusted and documented with high precision, ensuring reproducibility and consistency across experiments.
Thirdly, the introduction of real-world impairments through simulation makes the signals significantly realistic [10]. The simulation introduces controlled impairments such as AWGN or Rician fading, unavoidable imperfections such as amplifier non-linearities, frequency drift, and phase noise. These distortions enrich the dataset with natural variability that mirrors real deployment scenarios.
Lastly, the pipeline is fully deterministic and reproducible. Every waveform, filter, channel condition, and SDR configuration is generated using parameterized scripts. This ensures that the dataset is not only credible in theory, but also practical for validation.
Taken together, these elements make the dataset used in this study both theoretically valid and empirically grounded. It is ideally suited for developing and evaluating lightweight deep learning models for automatic modulation classification in constrained wireless environments.

Chapter 4

Proposed Methodology

4.1 Methodology Overview

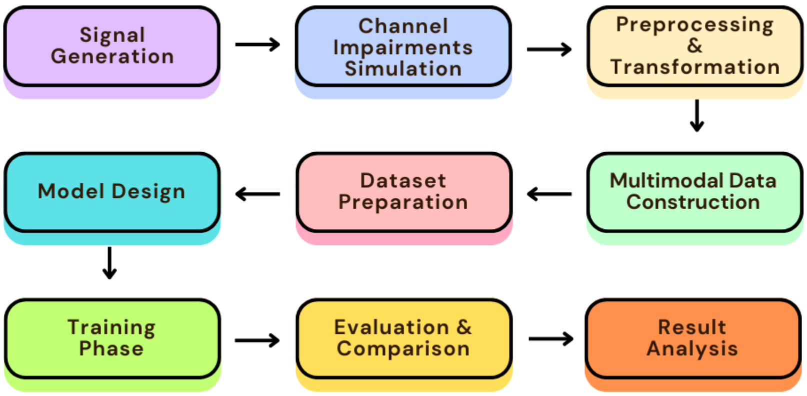

Figure 4.1: Top down view of the proposed methodology

Modulation classification using images from signals is a fascinating and innovative approach in the field of signal processing and machine learning. This method involves converting radio frequency (RF) signals into visual representations, such as spectrograms, scalograms, constellation diagrams, or other image formats, and then using image processing and deep learning techniques to classify the modulation schemes of these signals. This research presents an innovative approach to modulation classification, using a CWT to analyze time-frequency data and a custom deep learning architecture to classify signals based on wavelet scalograms. The signals are generated using the Communication Toolbox from MATLAB and captured I/Q components from the signals, which are later used to extract amplitude and phase information from each signal and converted into wavelet scalograms using CWT. Finally, the scalograms are fed into a custom dual-stream CNN for classification. Baseline models have also been adjusted similarly, modified to take single-channel (grayscale) input in a dual-stream architecture, placing them into each individual streams for comparison. After classification of the data, we observe each model's performance alongside their computational efficiency.

4.2 Dataset

As stated before, we have selected 8 modulation types (6 digital and 2 analog). 1000 frames have been generated for each modulation type. The signals are sampled at 200 kHz, with each frame containing 1024 I/Q samples. The modulation types include:

Digital Modulation Types:

- Binary Phase Shift Keying (BPSK)
- Quadrature Phase Shift Keying (QPSK)
- 64-ary Quadrature Amplitude Modulation (64-QAM)
- 4-ary Pulse Amplitude Modulation (PAM4)
- Gaussian Frequency Shift Keying (GFSK)
- Continuous Phase Frequency Shift Keying (CPFSK)

Analog Modulation Types

- Broadcast FM (B-FM)
- Single Sideband Amplitude Modulation (SSB-AM)

4.2.1 Data generation

The Radio signals have been collected using the **Communications Toolbox Support Package** form **MATLAB**. We used two **ADALM-PLUTO** radios to transmit the signals and capture the channel-impaired signals using the SDR configured for signal reception.

4.2.2 Channel Impairment

The generated frames pass through the following channel impairments:

- **Additive White Gaussian Noise (AWGN):** The channel adds AWGN with an SNR of 30 dB using the awgn function from the Communications Toolbox.
- **Rician Multipath Fading:** The signals are passed through a Rician multipath fading channel using the comm.RicianChannel System object from the Communications Toolbox. The channel is modeled with:
  - Delays of [0 1.8 3.4] samples.
  - Corresponding average path gains of [0 -2 -10] dB.
  - A K-factor of 4, and
  - A maximum Doppler shift of 4 Hz, simulating a walking speed at 902MHz.

- **Clock Offset:** Clock offset is caused by inaccuracies in internal clock sources of transmitters and receivers. This results in:
  - Frequency Offset: The frequency offset is modeled by scaling the clock offset factor C with the center frequency.
  - Sampling Rate Offset: A sampling rate offset is applied using the interp1 function to resample the frame at a new rate of C $\times$ fs.

These channel impairments are added to the signals to simulate realism and bring variety to the signals.

4.2.3 Data Transform

To analyze a complex signal in terms of its instantaneous amplitude and phase, we use its in-phase (I) and quadrature (Q) components, often obtained from I/Q demodulation. From the I/Q samples, we can compute the amplitude and phase as follows:

$$\text{Amplitude } A[n] = \sqrt{I[n]^2 + Q[n]^2}$$
$$\text{Phase } \phi[n] = \tan^{-1}\left(\frac{Q[n]}{I[n]}\right)$$

To analyze the time-frequency characteristics of the signal, the Continuous Wavelet Transform (CWT) was performed using the Morlet wavelet. We used the pywt.cwt function for this purpose. We also defined a range of scales to cover different frequency components of the signal. The CWT of a signal x(t) is defined as:

$$W_x(a, b) = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{|a|}} \psi^* \left(\frac{t-b}{a}\right) dt \tag{4.1}$$

where a is the scale parameter (analogous to frequency), b is the time shift parameter and $\psi^*(t)$ is the complex conjugate of the wavelet function.
The complex Morlet wavelet (which we use) is:

$$\psi(t) = \frac{1}{\sqrt{\pi B}} e^{-t^2/B} e^{j2\pi Ct} \tag{4.2}$$

where B is the bandwidth and C is the center frequency. We apply complex Morlet Wavelet (cmor1.5-0.5) over a set of log-spaced scales(from $10^{0.5}$ to 10^2) with 200 steps. This provides a time-frequency representation of the signal. To normalize the CWT coefficients, we apply MinMax Scaling (0 to 1 range). The output scalograms are resized to 224$\times$224 pixels using Lanczos interpolation, making them suitable for our deep learning model. We stack the amplitude and phase scalograms as two channels in a NumPy array and save them. The following Figure 4.2 shows the amplitude(on the left) and phase scalograms(on the right) for each modualtion scheme.

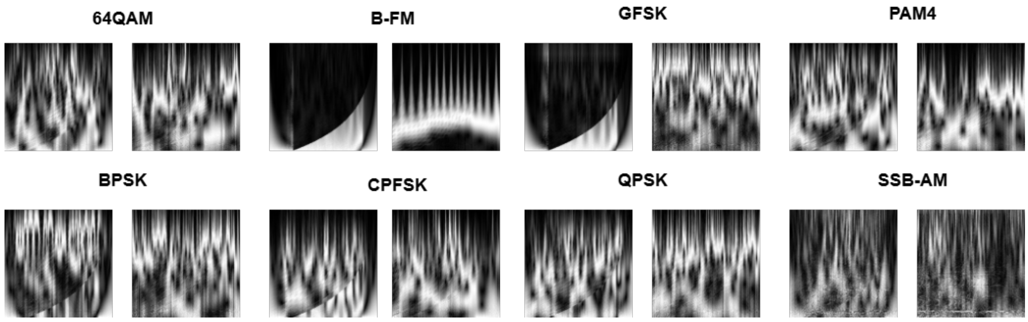

Figure 4.2: Sample scalograms for each class

4.3 Model architecture

Our proposed model is a dual-stream CNN architecture. We use two streams to process the modalities of the signal separately which are amplitude and phase scalograms. After passing through the streams the features are concatinated and passed through another feature extraction block from combined feature extraction. After that, the data is forwarded through a classification block.

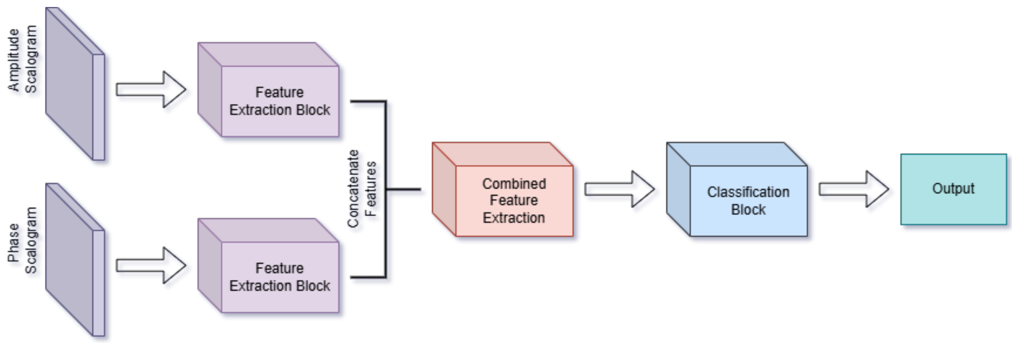

Figure 4.3: Model architecture

DualStreamCNN architecture uses two parallel feature extraction streams each based on a reduced version of AlexNet[3] to independently process two input channels (i.e., amplitude and phase scalograms). Hierarchical spatial features are extracted using convolutional, batch normalization, ReLU, and pooling layers in each stream. Finally, outputs from both streams are concatenated along the channel dimension to get a complete representation. The combined feature map is subsequently processed through additional convolutional layers in the combined feature extraction block, enhancing cross-channel exploitation. Lastly, we use a dropout-regularized classification block that outputs the final class predictions from the spatially pooled and flattened features. The major components of the model are broken down in Figure 4.4.

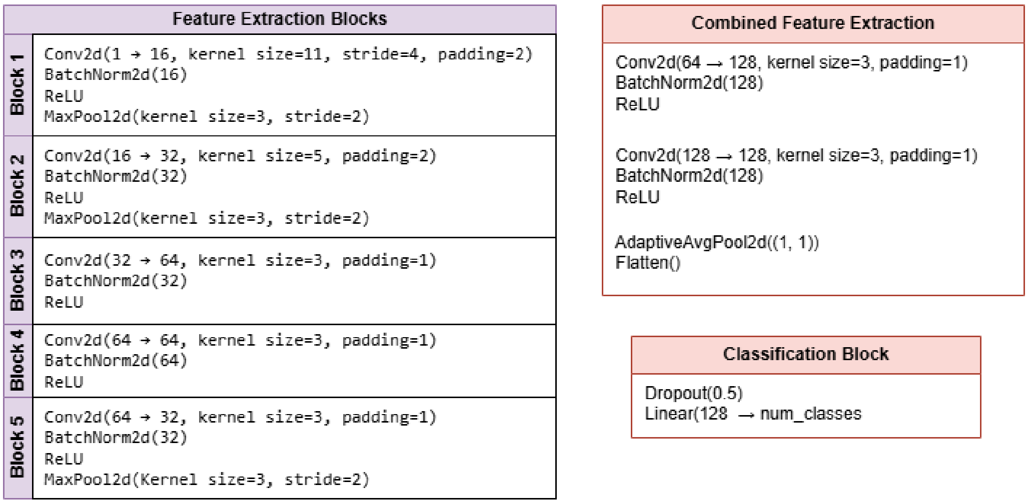

Figure 4.4: Model breakdown

Chapter 5

Result Analysis

5.1 Experimental Setup

5.1.1 Hardware & Software Specifications

We conducted the experiments using an Intel Core i7-14700K CPU with 64GB DDR5 memory and an Nvidia 4080 Super 16GB GPU for simulation. For signal generation, we used MATLAB R2021a with the Communication Toolbox to generate modulated signals (e.g., BPSK, QAM, FSK, etc.) and apply channel impairments like AWGN and Rayleigh fading. The Signal Processing Toolbox was used for filtering, transformations, noise modeling, and spectral analysis. As our deep learning framework, we used PyTorch 2.3.1 and imported the necessary libraries using Python.

5.1.2 Dataset Preparation

For training, validation, and testing, we split the dataset such that 80% of the data is used for training (800 signals), 10% for validation, and the remaining 10% for testing from each class. We used shuffling to ensure unbiased splitting. Input scalograms were resized to 224x224 pixels and normalized using standard deviation. For data augmentation, we applied random horizontal flips to increase variability and prevent overfitting.

The loss was computed using the `CrossEntropyLoss` function, and the AdamW optimizer with weight decay was used for stable weight updates. The learning rate was initialized at 0.001. A `ReduceLROnPlateau` scheduler was used to reduce the learning rate if validation accuracy plateaued. Training was performed over 50 epochs with a batch size of 64 and early stopping enabled to prevent overfitting.

5.2 Baseline Models

5.2.1 SqueezeNet

SqueezeNet is a lightweight deep neural network for image classification released in 2016 by researchers at DeepScale, UC Berkeley, and Stanford University. The authors claim that the model achieves similar accuracy to AlexNet on ImageNet, but is 50 times smaller [8]. This makes SqueezeNet an ideal baseline model. It uses Fire modules to reduce computational cost and parameter count without significantly sacrificing accuracy, making it suitable for edge device deployment. In this study, the SqueezeNet architecture was adapted to accept grayscale scalograms as input for modulation classification.

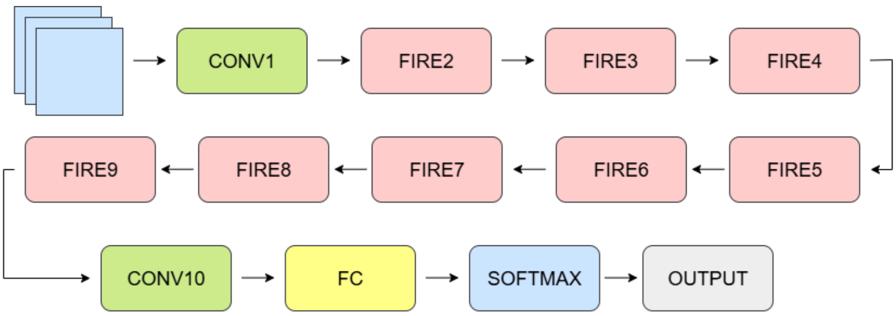

Figure 5.1: SqueezeNet Architecture [8]

5.2.2 EfficientNet

EfficientNet, introduced by Google Research, scales network dimensions to optimize both accuracy and efficiency. It uses a novel compound scaling method that uniformly scales depth, width, and resolution. Built on Mobile Inverted Bottleneck Convolutions (MBConv) from MobileNetV2 [13], EfficientNet models range from B0 to B7. EfficientNet-Lite0 was selected for comparison due to its low parameter count (4.7 million) and grayscale image support.

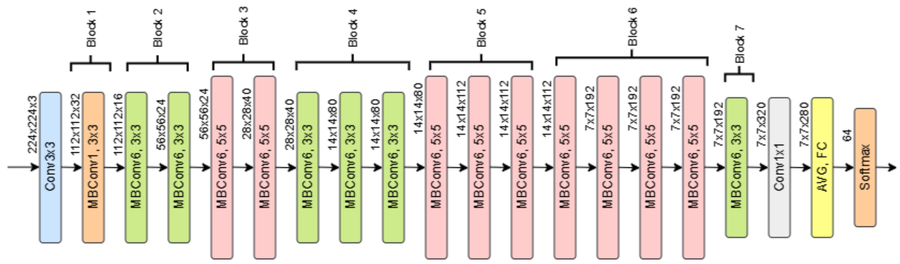

Figure 5.2: EfficientNet-Lite0 Architecture [13]

5.2.3 DenseNet

DenseNet-121 is a densely connected convolutional network known for its efficient feature reuse and improved gradient flow during backpropagation [6]. Each layer receives the feature maps of all previous layers as input, which improves feature propagation and reduces vanishing gradient issues. Transition layers (1x1 convolutions and pooling) between dense blocks control the size of the model. A global average pooling layer at the end minimizes the parameter count. The classifier was replaced with a fully connected layer for our 8-class modulation classification task.

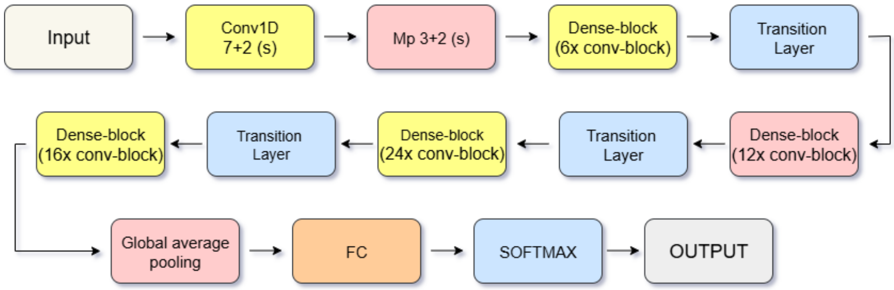

Figure 5.3: DenseNet-121 Architecture [6]

5.3 Performance Evaluation Metrics

The proposed and baseline models are evaluated using test accuracy, and macro-averaged precision, recall, and F1-score. A confusion matrix is used to visualize prediction results. We also compare model parameter counts to highlight the lightweight efficiency of our proposed architecture.

5.4 Performance of the Proposed Model

5.4.1 Training and Validation Loss Curves

The steady decrease in training loss indicates that the model is effectively learning. Early fluctuations in validation loss suggest some instability or overfitting during initial epochs. However, after epoch 20, validation loss begins to decrease, indicating improved generalization and convergence.

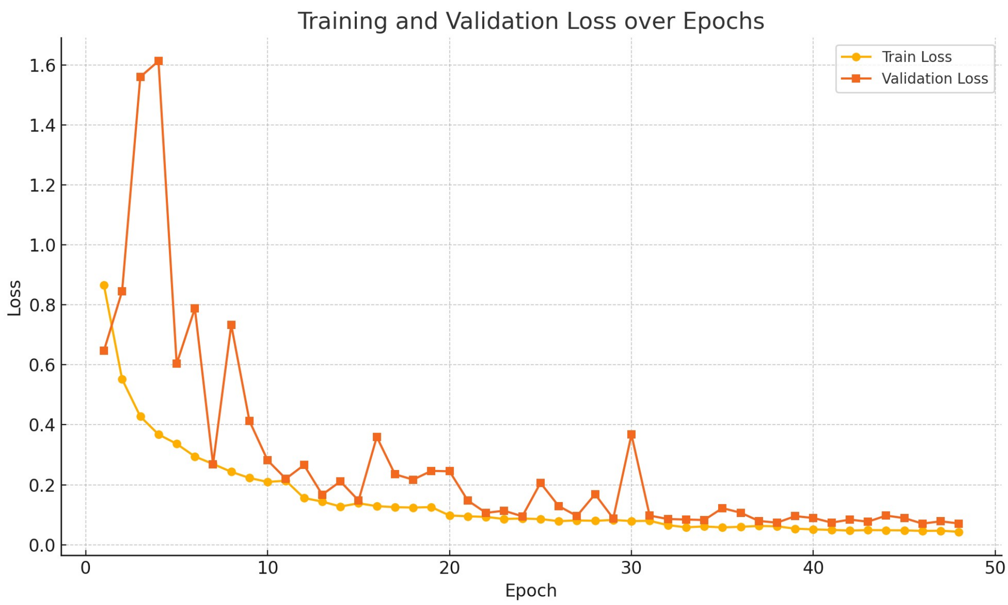

Figure 5.4: Training and Validation Loss Curve

5.4.2 Accuracy and Confusion Matrix

The classification report for the proposed dual-stream CNN model, in Table 5.1, shows that the above advantages of adding multi-modal learning can effectively classify different modulation schemes. The classwise values of macro, weighted average, precision, recall and F1-score as **0.98** suggests that the model performance remains consistent across all the classes. The model scores perfect precision, recall, and F1-score for a set of modulation types such as **GFSK**, **CPFSK**, and **SSB-AM**, noting that the network architecture is capable of capturing discriminative features for both analog and digital modulation types.

Performance for 64QAM, PAM4 and QPSK is marginally worse, the F1-scores ranging from 0.92 to 0.96, but still indicative of strong predictive power (and acceptable, given spectral overlap between some of these modulation schemes). The relatively worse performance of 64QAM and PAM4 can be explained by the complexity and the high symbol density of these higher order modulations.

Table 5.1: Classification report for modulation classes

Class	**Precision**	**Recall**	**F1-Score**	**Support**
64QAM	0.92	0.92	0.92	100
B-FM	1.00	0.99	0.99	100
BPSK	1.00	0.99	0.99	100
CPFSK	0.99	1.00	1.00	100
GFSK	1.00	1.00	1.00	100
PAM4	0.92	0.97	0.95	100
QPSK	0.98	0.94	0.96	100
SSB-AM	1.00	1.00	1.00	100
Accuracy			0.98	800
Macro Avg	0.98	0.98	0.98	800
Weighted Avg	0.98	0.98	0.98	800

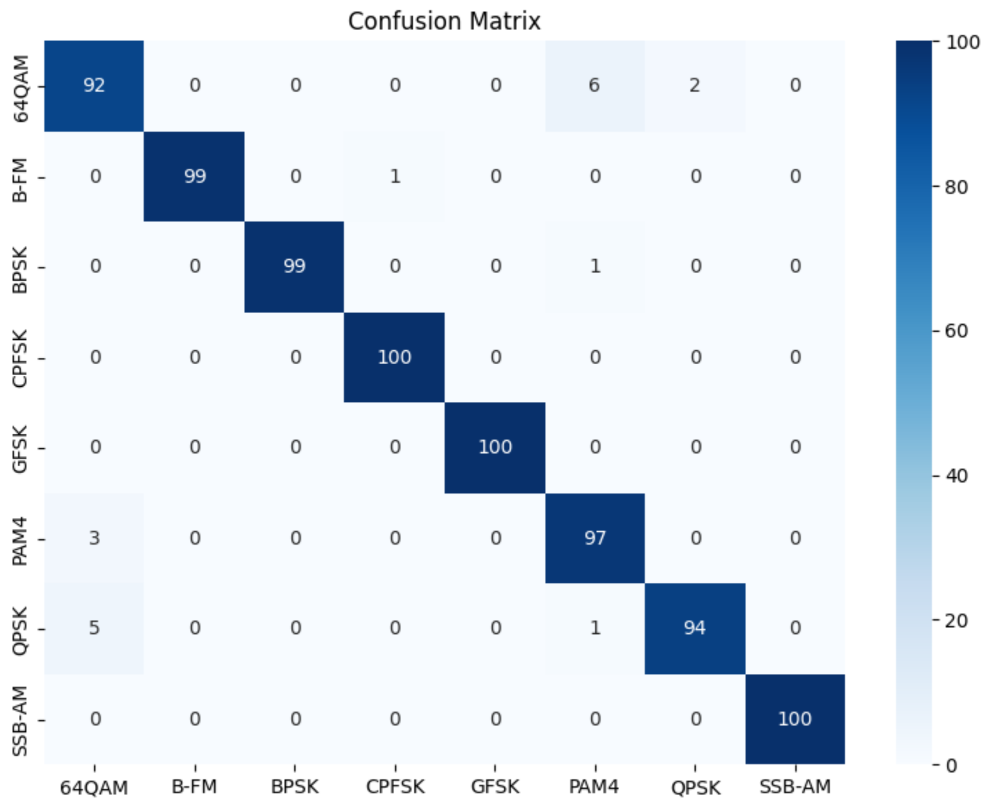

Figure 5.5: Confusion Matrix

In summary, the results confirm that our wavelet-based dual-stream architecture efficiently models amplitude and phase indications for robust classification under channel impairments and for different modulation types.

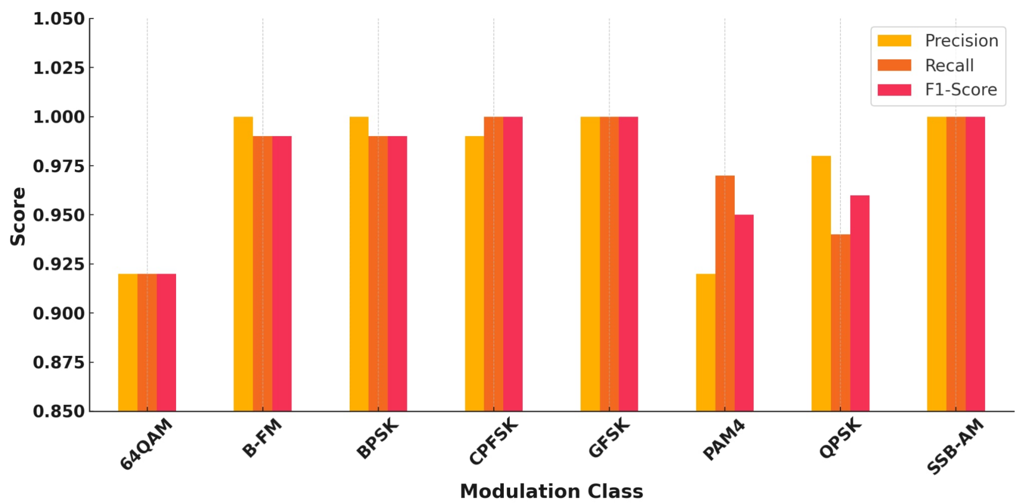

Figure 5.6: Classification Metrics per Modulation Class

5.4.3 Lightweight Design Efficiency

Our proposed model contains only 401,160 parameters, making it over 7× smaller than SqueezeNet and 34× smaller than DenseNet-121. Despite the smaller size, it achieves competitive accuracy, demonstrating strong efficiency and suitability for edge deployment.

5.5 Comparison with Baseline Models

The dual-stream CNN exhibits a competitive classification performance compared with the standard baseline CNN and at a considerably lower level of model complexity. Although DenseNet-121 and EfficientNet-Lite0 give slightly better accuracy and F1-scores, they also have a much larger parameter count, more than 13 million and 6.7 million respectively. SqueezeNet being substantially more compact, yet it contains almost eight times more parameters than our model. With 401160 parameters, the current model achieves the same SqueezeNet accuracy (98%) and is more than seven times smaller than it, thus making edge deployment possible with a trade-off of a decreased size.

Table 5.2: Performance Metrics on Test Dataset

Model	Precision	Recall	F1-Score	Accuracy	Parameters
DenseNet-121	0.989	0.976	0.990	0.990	13,911,560
EfficientNet-Lite0	0.984	0.979	0.982	0.985	6,761,352
SqueezeNet	0.978	0.983	0.980	0.980	3,150,984
Proposed	0.976	0.981	0.979	0.980	401,160

This figure shows the precision, recall, and F1-score for each modulation class. Most classes such as B-FM, BPSK, GFSK, CPFSK, and SSB-AM have perfect or near-perfect scores. These results confirm that the model correctly identifies these modulations with high reliability. Slight drops are observed in PAM4 and 64QAM due to their similar signal patterns, which can confuse the classifier. QPSK also shows minor recall reduction. Overall, the model performs consistently across both analog and digital modulation types, ensuring robustness in classification.

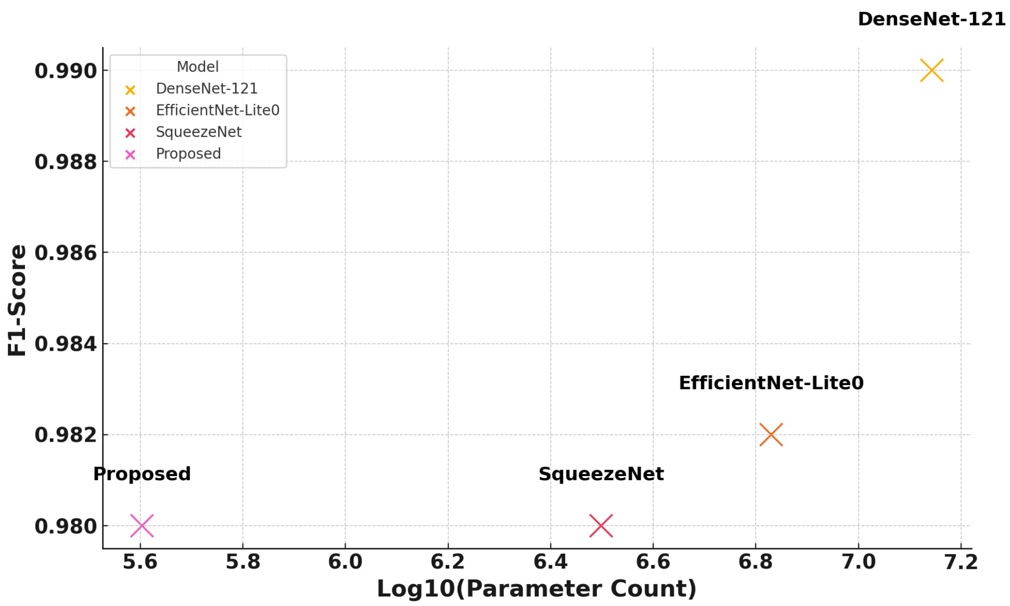

Figure 5.7: F1-Score vs. Log(Parameter Count)

This plot compares the performance and size of different models. DenseNet-121 has the highest F1-score but also the largest number of parameters. The proposed model achieves the same F1-score as SqueezeNet with far fewer parameters. It also closely matches EfficientNet-Lite0 in accuracy while using fewer resources. This balance between high accuracy and low complexity shows that the proposed model is both efficient and effective. It is especially suitable for devices with limited processing power.

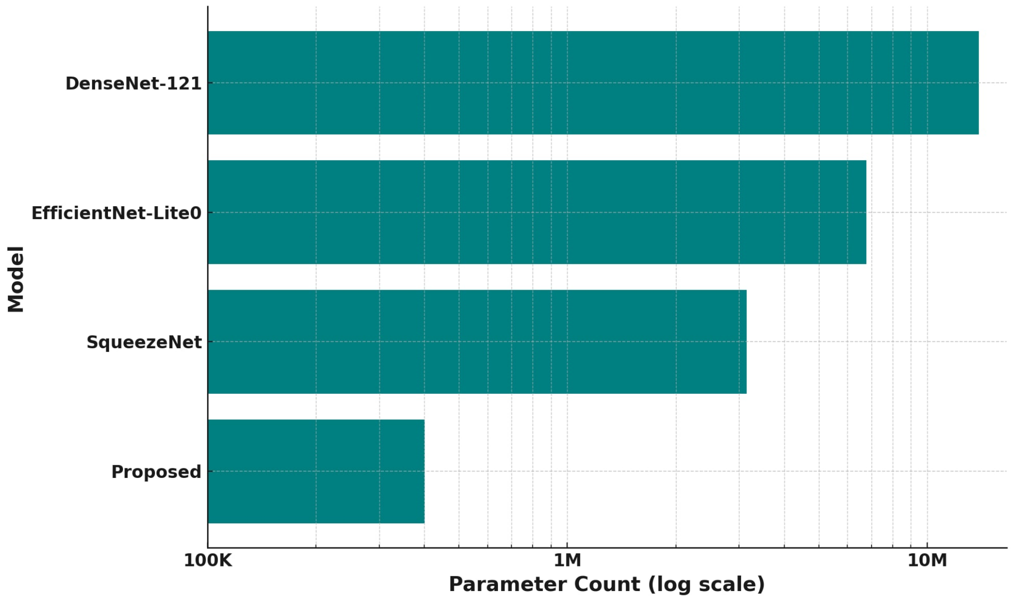

Figure 5.8: Parameter Count (Log Scale)

This bar chart shows the number of parameters for each model on a logarithmic scale. The proposed model has the smallest size, with just over 400 thousand parameters. In contrast, DenseNet-121 and EfficientNet-Lite0 have millions of parameters. SqueezeNet is smaller than these two but still much larger than the proposed model. This figure highlights how compact the proposed model is. It proves that a very small model can still achieve high classification performance, making it ideal for real-time and embedded applications.

5.6 Discussions

The results indicate that the proposed model achieves comparable accuracy and F1-score to state-of-the-art lightweight architectures while maintaining a significantly smaller model size. This makes it well-suited for resource-constrained environments like edge devices. However, while the architecture is efficient, it may still benefit from further optimization for specific hardware. Additionally, scalability to more complex signal scenarios and robustness to adversarial channel conditions remain areas for future work.

Chapter 6

Conclusion

6.1 Summary of Findings

Our research explores a new branch in image-based approaches for automatic modulation classification. We demonstrate an innovative way to represent signal data for lightweight deep learning models. Our dual-stream custom CNN model shows exceptional performance on the data with 98% accuracy. Other state of the art models like DenseNet-121 and lightweight models SqueezeNet and EfficientNet-Lite0 have shown similar performance on our dataset but at the cost of much higher parameters. Our proposed model is 7x smaller than SqueezeNet, 16x smaller than EffcientNet-Lite0 and 34x smaller than DenseNet121. This implies that our model is well suited for edge deployment, as the two light baseline models are already great candidates for resource-scarce environments. The accuracy and performance of these deep learning methods also show the robustness of the data representation technique.

6.2 Contributions to the Field

Our contributions to this research are mentioned as follows:

- In our research, we propose an innovative approach to AMC and introduce the lightweight deep learning model, particularly beneficial because of its size and deployability in the computation-scarce scenarios. Therefore, by integrating the architecture into our work, we envision expanding the effectiveness of AMC systems in challenging environments where computational power is a constraint.

- Our work involves generating signals using MATLAB and converting them into wavelet scalograms. These scalograms provide a visual representation of signal characteristics across time and frequency, enabling more effective feature extraction for classification. This facilitates further image-based AMC implementations, particularly using wavelet transforms.

- To enhance classification performance, we incorporate a multi-modal approach. This approach lets deep learning models to effectively extract features and have better results compared to using single modalities. This demonstrates that learning from multiple signal representations simultaneously allow deep learning models to work with fewer parameters to output same results.

Overall, our research aims to advance AMC techniques by combining lightweight architectures and signal processing with continuous wavelet transform. These contributions have the potential to benefit various applications, including wireless communication systems and cognitive radios.

6.3 Recommendations for Future Work

To improve the performance and generality of the model, it can be adopted with attention layers or transformer-based modules for more effective learning of the global signal correlation information. Generalization could be improved by augmenting the dataset with additional modulation types and a wide range of channel conditions, low SNR scenarios, and sampling the models on a wider range of channel conditions. Furthermore, the deployment and verification of the model on a RT embedded systems or edge devices can assist in quantifying the practical efficiency. Investigating self-supervised or few-shot learning methods will also alleviate the requirement of large annotated datasets in real applications.

Bibliography

[1] 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.

[2] J. L. Xu, W. Su, and M. Zhou, "Likelihood-ratio approaches to automatic modulation classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 41, no. 4, pp. 455–469, 2011. DOI: 10.1109/TSMCC.2010.2076347.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds., vol. 25, Curran Associates, Inc., 2012. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

[4] A. Hazza, M. Shoaib, S. A. Alshebeili, and A. Fahad, "An overview of feature-based methods for digital modulation classification," in *2013 1st International Conference on Communications, Signal Processing, and their Applications (ICCSPA)*, 2013, pp. 1–6. DOI: 10.1109/ICCSPA.2013.6487244.

[5] C. Szegedy, W. Liu, Y. Jia, *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[6] G. Huang, Z. Liu, and K. Q. Weinberger, "Densely connected convolutional networks," *CoRR*, vol. abs/1608.06993, 2016. arXiv: 1608.06993. [Online]. Available: http://arxiv.org/abs/1608.06993.

[7] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size," *arXiv preprint arXiv:1602.07360*, 2016.

[8] F. N. Iandola, M. W. Moskewicz, K. Ashraf, S. Han, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and ¡1mb model size," *CoRR*, vol. abs/1602.07360, 2016. arXiv: 1602.07360. [Online]. Available: http://arxiv.org/abs/1602.07360.

[9] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[10] T. 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. DOI: 10.1109/TCCN.2017.2758370.

[11] S. Peng, H. Jiang, H. Wang, H. Alwageed, and Y.-D. Yao, "Modulation classification using convolutional neural network based deep learning model," in *2017 26th Wireless and Optical Communication Conference (WOCC)*, IEEE, 2017, pp. 1–5.

[12] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 31, 2017.

[13] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," *CoRR*, vol. abs/1905.11946, 2019. arXiv: 1905.11946. [Online]. Available: http://arxiv.org/abs/1905.11946.

[14] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation (Release 16)," 3rd Generation Partnership Project (3GPP), Technical Specification TS 36.211, version 16.0.0, Jul. 2020, https://www.3gpp.org/DynaReport/36211.htm.

[15] 3GPP, "Study on Channel Model for Frequencies from 0.5 to 100 GHz (Release 16)," 3rd Generation Partnership Project (3GPP), Technical Report TR 38.901, version 16.1.0, Jan. 2020, https://www.3gpp.org/DynaReport/38901.htm.

[16] N. Daldal, Z. Cömert, and K. Polat, "Automatic determination of digital modulation types with different noises using convolutional neural network based on time–frequency information," *Applied Soft Computing*, vol. 86, p. 105 834, 2020.

[17] V.-S. Doan, T. Huynh-The, C.-H. Hua, Q.-V. Pham, and D.-S. Kim, "Learning constellation map with deep cnn for accurate modulation recognition," in *GLOBECOM 2020-2020 IEEE Global Communications Conference*, IEEE, 2020, pp. 1–6.

[18] T. Huynh-The, C.-H. Hua, Q.-V. Pham, and D.-S. Kim, "Mcnet: An efficient cnn architecture for robust automatic modulation classification," *IEEE Communications Letters*, vol. 24, no. 4, pp. 811–815, 2020.

[19] S.-H. Kim, J.-W. Kim, V.-S. Doan, and D.-S. Kim, "Lightweight deep learning model for automatic modulation classification in cognitive radio networks," *IEEE Access*, vol. 8, pp. 197 532–197 541, 2020.

[20] Y. Lin, Y. Tu, and Z. Dou, "An improved neural network pruning technology for automatic modulation classification in edge devices," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 5, pp. 5703–5706, 2020.

[21] R. M. Al-Makhlasawy, A. A. Hefnawy, M. M. Abd Elnaby, and F. E. Abd El-Samie, "Modulation classification in the presence of adjacent channel interference using convolutional neural networks," *International Journal of Communication Systems*, vol. 33, no. 13, e4295, 2020.

[22] Y. Wang, J. Yang, M. Liu, and G. Gui, "Lightamc: Lightweight automatic modulation classification via deep learning and compressive sensing," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 3, pp. 3491–3495, 2020.

[23] A. Dosovitskiy, L. Beyer, A. Kolesnikov, *et al.*, *An image is worth 16x16 words: Transformers for image recognition at scale*, 2021. arXiv: 2010.11929 [cs.CV]. [Online]. Available: https://arxiv.org/abs/2010.11929.

[24] X. Fu, G. Gui, Y. Wang, *et al.*, "Lightweight automatic modulation classification based on decentralized learning," *IEEE Transactions on Cognitive Communications and Networking*, vol. 8, no. 1, pp. 57–70, 2021.

[25] T. Huynh-The, C.-H. Hua, V.-S. Doan, and D.-S. Kim, "Accurate modulation classification with reusable-feature convolutional neural network," in *2020 IEEE Eighth International Conference on Communications and Electronics (ICCE)*, IEEE, 2021, pp. 12–17.

[26] T. Huynh-The, Q.-V. Pham, T.-V. Nguyen, *et al.*, "Automatic modulation classification: A deep architecture survey," *IEEE Access*, vol. 9, pp. 142 950–142 971, 2021.

[27] Z. Ke and H. Vikalo, "Real-time radio modulation classification with an lstm auto-encoder," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2021, pp. 4935–4939.

[28] Z. Wang, D. Sun, K. Gong, W. Wang, and P. Sun, "A lightweight cnn architecture for automatic modulation classification," *Electronics*, vol. 10, no. 21, p. 2679, 2021.

[29] J. Wu, Y. Zhong, and A. Chen, "Radio modulation classification using stft spectrogram and cnn," in *2021 7th International Conference on Computer and Communications (ICCC)*, 2021, pp. 178–182. DOI: 10.1109/ICCC54389.2021.9674714.

[30] J. Yin, L. Guo, W. Jiang, S. Hong, and J. Yang, "Shufflenet-inspired lightweight neural network design for automatic modulation classification methods in ubiquitous iot cyber–physical systems," *Computer Communications*, vol. 176, pp. 249–257, 2021, ISSN: 0140-3664. DOI: https://doi.org/10.1016/j.comcom.2021.05.005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0140366421001894.

[31] F. Zhang, C. Luo, J. Xu, and Y. Luo, "An efficient deep learning model for automatic modulation recognition based on parameter estimation and transformation," *IEEE Communications Letters*, vol. 25, no. 10, pp. 3287–3290, 2021.

[32] A. M. Abdulkarem, F. Abedi, H. M. Ghanimi, *et al.*, "Robust automatic modulation classification using convolutional deep neural network based on scalogram information," *Computers*, vol. 11, no. 11, p. 162, 2022.

[33] B. Dong, Y. Liu, G. Gui, *et al.*, "A lightweight decentralized-learning-based automatic modulation classification method for resource-constrained edge devices," *IEEE Internet of Things Journal*, vol. 9, no. 24, pp. 24 708–24 720, 2022.

[34] M. Shafiq and Z. Gu, "Deep residual learning for image recognition: A survey," *Applied sciences*, vol. 12, no. 18, p. 8972, 2022.

[35] Y. Sun and E. A. Ball, "Automatic modulation classification using techniques from image classification," *IET Communications*, vol. 16, no. 11, pp. 1303–1314, 2022.

[36] H. S. Ghanem, R. M. Al-Makhlasawy, W. El-Shafai, *et al.*, "Wireless modulation classification based on radon transform and convolutional neural networks," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 5, pp. 6263–6272, 2023.

[37] H. S. Hussein, M. H. E. Ali, M. Ismeil, M. N. Shaaban, M. L. Mohamed, and H. A. Atallah, "Automatic modulation classification: Convolutional deep learning neural networks approaches," *IEEE Access*, 2023.

[38] M. Z. Nisar, M. S. Ibrahim, M. Usman, and J.-A. Lee, "A lightweight deep learning model for automatic modulation classification using residual learning and squeeze–excitation blocks," *Applied Sciences*, vol. 13, no. 8, p. 5145, 2023.

[39] R. Raj, P. Tiwari, M. K. Gourisaria, and H. Das, "Efficient object detection and labeling in retail environments using mobilenetv2 with inverted residuals," in *2023 OITS International Conference on Information Technology (OCIT)*, IEEE, 2023, pp. 634–639.

[40] F. Wang, T. Shang, C. Hu, and Q. Liu, "Automatic modulation classification using hybrid data augmentation and lightweight neural network," *Sensors*, vol. 23, no. 9, p. 4187, 2023.

[41] M. Waqas, M. Ashraf, and M. Zakwan, "Modulation classification through deep learning using resolution transformed spectrograms," *arXiv preprint arXiv:2306.04655*, 2023.

[42] X. Zhang, X. Chen, Y. Wang, *et al.*, "Lightweight automatic modulation classification via progressive differentiable architecture search," *IEEE Transactions on Cognitive Communications and Networking*, 2023.

[43] L. Guo, Y. Wang, Y. Liu, Y. Lin, H. Zhao, and G. Gui, "Ultra lite convolutional neural network for automatic modulation classification in internet of unmanned aerial vehicles," *IEEE Internet of Things Journal*, 2024.

[44] W. Zhao and Q. Luo, "Leveraging computer vision for automatic modulation classification: Insights from spectrum and constellation diagram analysis," in *International Conference on Pattern Recognition*, Springer, 2024, pp. 284–294.

[45] F. U. Rehman, Q. Abbas, and M. K. Shehzad, "Dl-amc: Deep learning for automatic modulation classification," *arXiv preprint arXiv:2504.08011*, 2025.

[46] T. Wu, W. Zheng, and L. Liu, "Automatic modulation recognition using vision transformers with cross-layer feature fusion and kolmogorov–arnold representation," *Wireless Networks*, pp. 1–12, 2025.

[47] H. Zhang, Y. Kuang, R. Huang, S. Lin, Y. Dong, and M. Zhang, "Modulation recognition method based on multimodal features," *Frontiers in Communications and Networks*, vol. 6, p. 1 453 125, 2025.