

Hybridrailnet: A 5g-Enabled Dual-Stream Fmcw Radar Framework For Real Time Railway Obstacle Detection

Malateshwari Aski¹, Swetha²

¹Research Scholar Dept of Electronics, Karnataka State Akkamahadevi women university, Vijayapura.

²Assistant Professor, Dept of Electronics, Karnataka State Akkamahadevi women university, Vijayapura.

Abstract

Real-time obstacle identification under low light, fog, rain, and strong rail-clutter reflections is necessary to prevent safety-critical dangers caused by pedestrian, animal, and vehicle intrusions on railway tracks. While single-stream lightweight CNNs frequently experience noisy Range–Doppler patterns, limited motion discrimination, and higher false alarms/missed detections, conventional vision/thermal systems deteriorate in low visibility, and classical radar pipelines (hand-crafted CFAR/thresholding + rule-based logic) are brittle under nonstationary clutter and multipath. This study provides a 5G-enabled IoT radar obstacle recognition system that uses HybridRailNet, a dual-stream CNN that simultaneously learns from Range–Doppler (RD) and Micro-Doppler (MD) representations, to overcome these drawbacks. After applying Min–Max normalization, MTI-based clutter suppression, and a CFAR-like adaptive mask, the pipeline uses an enhanced ResNet-18 backbone with asymmetric convolution fusion, Mish activation, and Improved CBAM (ICBAM) attention to process RD and MD, followed by late fusion and Focal Loss training. HybridRailNet with Accuracy (0.9889), Precision(1.0000), Recall(0.9789) and F1-Score(0.9893). With a low bandwidth need (0.008), it achieves 4.2 ms edge latency and 14.7 ms end-to-end alert latency for 5G IoT applicability. These findings show that HybridRailNet supports reliable 5G-integrated safety monitoring by enabling precise, low-latency, and communication-efficient railway obstacle notifications.

Keywords: Railway obstacle detection, FMCW radar, Range–Doppler, Micro-Doppler, Dual-stream CNN, Attention mechanism.

1 Introduction

With the significant advancement of railway transportation, more emphasis has been placed on the safety and dependability of rail systems [1]. However, the obstacles invading the track pose a significant threat to driving safety owing to their suddenness and unpredictability [2]. Pedestrians, cars that cross the rails, animals that wander on the tracks [3], and a few other heavy, inanimate things that have fallen onto the track from the overpass are some of the obstacles. Many railroad crossings in both industrialized and developing nations lack the necessary safety features, such as gates and lights [4]. Traditional manual inspection and monitoring techniques are no longer enough to satisfy the rising needs for accuracy, real-time detection, and rapid response as railway networks continue to develop and train speeds rise. Therefore, in order to prevent serious traffic accidents and guarantee the safety of persons and property, it is essential to identify multi-scale unforeseen obstructions with high accuracy and to provide timely warnings [5].

An autonomous railway track obstacle detection system that uses the Internet of Things (IoT) has emerged as a viable solution to this critical issue [6]. The IoT, which allows a network powered by linked sensors and devices to monitor every movement in real-time, has a significant impact on the development of the

railway management system. Radar-based systems have drawn more attention than other sensing modalities, especially in the fields of environmental monitoring, autonomous systems, smart transportation, and surveillance [7].

Compared to vision sensors, radar sensors offer unique advantages since they can function in challenging environments such as poor illumination, bad weather, and limited visibility [8]. Because of these characteristics, radar is especially suitable for handling real-time object tracking issues, which are essential in autonomous and safety-critical systems and applications [9]. However, when used in dispersed IoT systems, radar-based tracking systems have many serious problems in spite of these benefits [10]. The foundation of traditional centralized machine learning (ML) approaches is the collection and synthesis of enormous volumes of data at one location, which presents significant privacy risks, latency, and bandwidth constraints [11], [12]. Furthermore, radar data are often sparse, noisy, and high-dimensional, particularly under low signal-noise-ratio (SNR) situations. These factors make it more difficult to use established classical ML techniques to guarantee high resilience and high generalization [13].

Deep learning (DL) has made notable progress in a number of object detection tasks in recent years [14] [15]. The field of DL has led to the gradual emergence of CNN-based object detection methods [16], which can be classified into two-stage and single-stage algorithms [17]. YOLO series [15] [18] [19], SSD [20], RetinaNet [21], and other algorithms are common single-stage target detection algorithms, while R-CNN [22], Faster R-CNN [23], and other algorithms are common two-stage target detection algorithms.

This paper introduces HybridRailNet, a 5G-enabled dual-stream FMCW radar architecture for obstacle detection that facilitates edge-level observation and utilizes 5G connectivity for low-latency warning propagation and centralized monitoring. Each radar sample is depicted through complementary RD and MD perspectives, enhanced by MTI-based clutter suppression and a CFAR-like adaptive mask to highlight target-like areas. HybridRailNet concurrently processes RD and MD inputs via two parallel streams with an enhanced ResNet-18 backbone, facilitating the model's ability to simultaneously learn spatial-range cues and motion-induced signatures. To improve discriminative learning in cluttered and low-SNR environments, the network incorporates an Asymmetric Convolution Module (3×3 , 1×3 , 3×1) to detect directional radar patterns like Doppler streaks and range bands, an Improved CBAM (ICBAM) attention mechanism to diminish irrelevant clutter and highlight energy concentrations pertinent to obstacles, and Mish activation to enhance the preservation of weak radar signatures and stabilize the optimization process. The proposed framework is intended for real-time implementation on train-side edge devices, ensuring that safety-critical actions are executed locally, while utilizing 5G URLLC/QoS support to relay concise alerts and status messages to control centers and trackside systems for oversight, coordination, and documentation.

This research's principal contributions are summarized as follows:

- 1 A dual-stream FMCW radar detection system (HybridRailNet) that concurrently utilizes RD localization and MD motion signatures for binary railway obstacle identification (Clear vs. Obstacle).

- 2 Radar-centric preprocessing and input organization, encompassing MTI-based clutter mitigation and a CFAR-like adaptive mask to direct learning towards target-resembling areas.
- 3 An enhanced ResNet-18 backbone that optimizes computing efficiency while providing robust representation learning for real-time edge inference.
- 4 Integration of ICBAM attention to improve obstacle-centric feature learning and diminish sensitivity to clutter and noise.
- 5 Asymmetric convolution fusion to successfully capture anisotropic radar patterns with minimal overhead.
- 6 Deployment-focused design incorporating 5G-enabled IoT integration, facilitating low-latency alert distribution and scalable monitoring while ensuring fail-safe safety decisions at the edge.

The remaining sections are organized as follows: Section 2 reviews related work and research gaps. Section 3 provides background on the proposed models. Section 4 describes the methodology. Section 5 presents experiments, analyzes results, and compares findings. Section 6 concludes the study and outlines future directions.

2 Literature Review

IoT, sensors, image processing methods, machine learning algorithms, and other robotics components have all been added to railway track monitoring and obstacle detection systems, according to the most recent literature analysis. These technologies improve the safety, effectiveness, and dependability of railway operations by instantly identifying fractures and anomalies in the rail tracks [24]. To find and identify obstacles on railway tracks, several studies have been conducted. Zhou et al. [25] introduced a unified framework called detecting contiguous outliers in the low low-rank representation for the detection of moving objects, combining object detection with background learning. An obstacle identification method using a thermal camera and the ADA boost algorithm was proposed by Manikandan et al. [26]. The rapid growth in field of ML is also contributing in the field of railway applications. Krummenacher et al. [27] introduced two ML strategies dependent on Support Vector Machines (SVM) and neural network to recognize rail wheels. SVM were employed by Tong et al. [28] to find potential railway obstacles based on the aspect ratio and color index that they specifically created. Although the method's structure is straightforward, it has limited detection capabilities for objects with multiple scales.

In addition to obstructions on the railroad track, unfavorable weather conditions such as rain, haze, darkness, and overcast skies pose a significant threat to railway safety. The vision of the moving train at high speed is directly impacted by the varying light levels under these settings. Low visibility may often result in significant losses for railroads. In order to give enough visibility in any kind of unfavorable weather situation, the system should also be invariant to light. Mangale et al. [29] proposed a directed acyclic graph approach for object identification in thermal pictures of low light and various weather conditions. M. Kristo et al. [30] described a YOLO framework- based method for thermal object detection in bad weather. By using CNN in the feature fusion network and

designing the SSwin-Le transformer for feature extraction, He et al. [31] greatly increased the detection accuracy of railroad impediments in tunnel lines and at night. However, the model's high computational complexity makes it unsuitable for real-time applications and hinders its performance.

The achievements of the DL in the field of object recognition motivated to analyze their performance in obstacle recognition on the railway track. Guo et al. [32] redesigned the Faster R- CNN's loss function to increase the railway dropper's detection accuracy. By using the notion of acquiring candidate areas prior to detection, two-stage detectors may achieve a greater detection accuracy; nevertheless, speed is a penalty associated with all of these detectors. He, Qiu [33] et al. proposed an enhanced foreign object detection technique based on Mask R-CNN; Tao Hui-Qing [34] proposed a better railway foreign object identification model with Faster R-CNN as its foundation, with VGG-16 as the feature extraction network of the Faster R-CNN model detection accuracy, which can reach 93.6%. It could synthesize multiscale enhancement techniques to boost target detection performance. The performance of target detection, the two-stage target detection approaches stated above, and the detection accuracy are all at a high level, but they are all plagued by the issue that the model is too large and real-time deployment is often challenging. Recent lightweight and real-time railway obstacle detection models are included in Table 1, which provides a summary of their datasets, main contributions, and limitations.

Table 1 Overview of Recent Lightweight and Real-Time Railway Obstacle Detection Models

Study	Model	Dataset	Key Contributions	Limitations
Zhao et al. [5]	RH-Net: Lightweight CNN+ Improved Transformer (LEM + IFM + EFM)	Custom railway obstacle dataset (rain, fog, low-light); MS COCO	Proven real-world and embedded applicability; robust detection in fog, rain, and low-light conditions; effective multi-scale obstacle detection; 97.02% mAP, 43 FPS	Vision-based; sensitive to severe occlusion and extreme illumination
Niu et al. [35]	MSL-YOLO: YOLO-based lightweight detector with MSCM, C2f-Star-EMA, LSCD head	Custom railway foreign object intrusion dataset	Ultra-lightweight model (2.35M parameters, 6.7 GFLOPs); very high speed (277 FPS, 3.3 ms latency); strong accuracy (94.3% mAP); excellent real-time performance	Dataset-specific optimization; generalization to unseen railway scenes not fully explored

Rahman et al. [36]	MobileNetV2-based CNN with transfer learning	Custom railway obstacle dataset	High accuracy (97.0%) with lightweight MobileNetV2; efficient resource usage; suitable for low-config devices; balance between detection speed and processing	Performance may vary under highly complex environments; limited comparative analysis with YOLOv5, ResNet50, VGG16
Liu & Lin [37]	MRS-YOLO: Improved YOLO11 Channel Pruning + MAKDF SC_Detect	Railroad transmission line FOD dataset	Improved feature extraction for multi-size objects; reduced parameters and GFLOPs; mAP50 94.8%, mAP50:9586.4%; enhanced detection efficiency	Model requires careful tuning; performance on completely unseen datasets not fully validated
Song et al. [38]	LW-DETR: Lightweight Transformer+ PP-LCNet+ optimized cross-scale fusion	Modified Pascal VOC dataset	Effective on resource-constrained devices; high detection speed (135.3% FPS increase); 71.7% fewer parameters; 73.7% lower computation; maintains acceptable detection accuracy	Tested mainly on modified VOC; real-world railway-specific robustness may need further validation
Qi et al. [39]	Improved YOLOv5+ FasterNe	AARFOD dataset (Railway+ Aviation)	Reduced parameters by 25.1%, GFLOPs by 10.6%; Improve precision (+1.2%), recall (+1.0%), mAP@0.5(+0.6%); efficient foreign object detection on railway and airway tracks	Real-time performance on embedded devices not explicitly reported; limited extreme weather testing
Li et al. [40]	CDH-DETR: Lightweight Transformer	Custom railway intrusion dataset	Real-time detection (96.8 FPS) with only 14.86 MB parameters; improved mAP	Slightly lower mAP than heavier YOLO Models in some

	Context-Guided Block DySample HSFPN		By 2% vs RT-DETR; effective multi-scale feature fusion; robust in practical video validation	scenarios; requires careful parameter selection
--	---	--	--	--

Although ML, DL, and lightweight transformer-based models have made great strides in the detection of obstacles on railways, there are still certain problems to be solved. Traditional approaches encounter difficulties with multi-scale objects, fluctuating environmental circumstances, and real-time implementation, while vision-based and thermal methods are constrained in low-light, unfavorable weather, or obstructed situations. Two-stage detectors are quite accurate, but they take a lot of processing power and time, thus they can't be used in embedded applications. The detection of tiny or weak barriers may be compromised by lightweight and transformer-based models, which also often lack generality across a variety of railway contexts. Most of the systems that are already in place just use visual data and don't use multi-modal radar data enough, which makes them less reliable in settings with a lot of noise or poor signal-to-noise ratio. This study fills these gaps by presenting a hybrid dual-stream CNN with an updated ResNet18 backbone that combines range-Doppler and micro-Doppler radar features. The ICBAM, asymmetric convolutions, and Mish activation make these features even better. This method strikes a compromise between computing economy and representational capacity, improves feature attention, and offers dependable, real-time obstacle detection in a variety of railway settings.

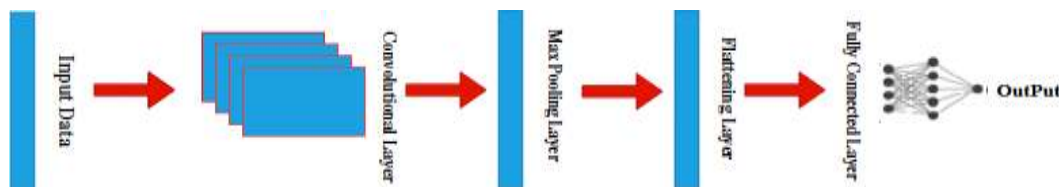
3 Background

This section presents the deep learning models and architectural components employed in this study, including CNNs, an improved ResNet-18 backbone, asymmetric convolution, attention mechanisms, and advanced activation functions. Together, these techniques enable effective extraction of discriminative spatial and motion-related features from radar representations, enhancing robustness and detection accuracy in railway obstacle detection tasks.

3.1 Convolutional Neural Networks (CNNs)

A typical CNN is composed of convolutional, pooling, flattening, and fully connected layers, each serving a specific purpose in feature extraction and classification, as described in [41] and illustrated in Figure 1.

Figure 1 Typical CNN Model



CNNs are great for analyzing radar data because range-Doppler and micro-Doppler representations seem like images, with significant objects showing up as organized energy distributions like peaks, ridges, and streaks. Early convolutional layers learn local spectral variations, intermediate layers learn motion and shape-related features, and deeper layers learn high-level semantic patterns that help tell the difference between

clear tracks and situations with obstacles.

C

m

n

$$Y(i, j) = \sum_{v=1}^n \sum_{u=1}^m \sum_{c=1}^C W_{k,c}(u, v) X_c(i + u, j + v) + b_k$$

In this formulation, X_c denotes the input feature map of the c -th channel, where C represents the total number of input channels. The convolution kernel corresponding to the k -th output feature map is denoted by W_k , and has a spatial size of $m \times n$, while b_k represents the associated bias

term. The indices (i, j) indicate the spatial location within the feature map, and Y_k denotes the

resulting k -th output feature map produced by the convolution operation.

Radar signatures can shift slightly in range or velocity due to sensor noise, rail curvature, and platform vibration; pooling improves tolerance to these shifts while preserving dominant evidence.

$$P(i, j) = \max_{(u, v) \in \Omega} Z(i + u, j + v)$$

$$Z(i + u, j + v)$$

Here, Z denotes the activated feature map, Ω represents the pooling window, and P denotes the resulting pooled feature map.

For binary detection, logits are converted into posterior probabilities; this probability is later used for decision thresholds (normal vs brake).

$$\hat{p} = \text{Softmax}(o), \hat{p}_t = \frac{e^{o_t}}{\sum_{q=1}^2 e^{o_q}}$$

$$e^{o_t}$$

$$e^{o_q}$$

$$q=1$$

Here, o represents the logits, $\hat{p}$ is the predicted probability vector, and t denotes the class index corresponding to Clear or Obstacle.

RD maps show where the item is (range bins) and how quickly it travels (Doppler bins). Micro-Doppler focuses on small changes in motion over time, including periodic components or traces that seem like vibrations. CNNs can do both because they take use of spatial localization and can be changed to work with 2D time-frequency maps.

A standard loss for classification is cross-entropy:

2

$$\mathcal{L}_{CE} = - \sum_{t=1}^2 y_t \log(\hat{p}_t)$$

$t=1$

Here, $\mathcal{L}_{CE}$ denotes the cross-entropy loss, $y_t \in \{0,1\}$ represents the one-hot encoded label, and $\hat{p}_t$

is the predicted probability for class t .

3.2 Improved ResNet-18

Kaiming He et al. [42] developed ResNet in 2016. To train deeper networks that are typically challenging to optimize, they proposed a residual learning approach. This

method reformulates network layers to learn residual functions with respect to the layer inputs. The results demonstrate that deeper networks based on residual learning achieve higher accuracy and improved optimization. The underlying mapping is given by:

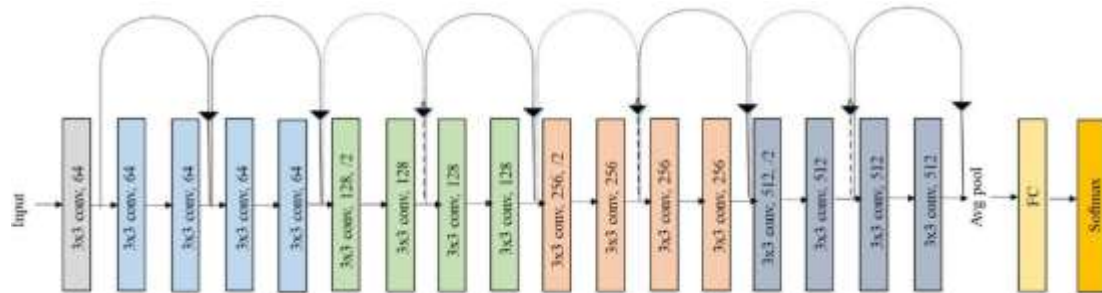
$$x(1) = F(x) + x(1)$$

Consequently, the residual function is defined as:

$$x(2) = H(x) - x(2)$$

Instead of learning the original mapping $x(1)$, the stacked layers explicitly learn the residual function $F(x)$. Figure 2 illustrates the architecture of the original ResNet-18. It consists of eighteen layers: seventeen convolutional layers, a fully connected layer, and an additional softmax layer for classification. The convolutional layers use 3×3 filters, and the network is designed such that the number of filters is maintained if the output feature map size remains unchanged. However, the number of filters is doubled when the output feature map is halved. Downsampling is performed using convolutional layers with a stride of two. Finally, the network concludes with an average pooling layer, a fully connected layer, and a softmax layer. Residual shortcut connections are consistently added between layers throughout the network.

Figure 2 Original ResNet-18 Architecture



3.2.1 Residual Learning in Improved ResNet-18

Deep convolutional neural networks can experience information loss as the network depth increases. Residual learning addresses this issue by allowing the network to learn only the residual mapping beyond the identity, thereby preserving low-level radar structures such as weak obstacle signatures near clutter bands. The output of a residual block is formulated as:

$$y = F(x; W) + x$$

where x is the block input, y is the block output, F represents the residual mapping, and W denotes the learnable weights. Residual connections also facilitate backpropagation and mitigate the vanishing gradient problem:

$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial y} \left(I + \frac{\partial F}{\partial x} \right)$$

$$\frac{\partial L}{\partial x}$$

$$\frac{\partial L}{\partial y}$$

$$\frac{\partial L}{\partial x}$$

Here, L is the loss function, and I represents the identity matrix.

3.2.2 Stem replacement

The standard ResNet-18 stem is replaced with three consecutive 3×3 convolutional layers to increase network nonlinearity while preserving receptive field coverage. This modification enhances the capture of subtle details in range-Doppler and micro-Doppler representations. The receptive field at layer i is computed as:

$$F_i = (F_{i-1} - 1) \cdot s_i + k_i$$

where s_i is the stride, k_i is the kernel size, and F_i represents the receptive field at layer i . Using strides of (2,1,1) for the three 3×3 layers preserve the effective receptive field of the original 7×7 convolution while enabling deeper feature learning.

3.2.3 Asymmetric convolution

Radar obstacle patterns often exhibit directional dominance, such as horizontal ridges or vertical streaks. Asymmetric convolution explicitly captures these anisotropic features by applying multiple directional kernels:

$$F_{\text{asym}} = (W_{3 \times 3} * X) + (W_{1 \times 3} * X) + (W_{3 \times 1} * X)$$

where X is the input feature map, $*$ denotes the convolution operation, and $W_{a \times b}$ are the directional kernels. Batch normalization and an activation function ($\cdot$) are applied to stabilize training and produce the final output feature map:

$$Y = (\text{BN}(F_{\text{asym}}))$$

Here, Y represents the output feature map after asymmetric convolution, batch normalization, and activation.

3.2.4 Mish activation

Traditional ReLU activation sets all negative inputs to zero, which can inadvertently discard faint yet meaningful radar signals especially after MTI (Moving Target Indicator) or CFAR (Constant False Alarm Rate) processing, which suppresses clutter and leaves only minimal residuals. To address this limitation, the Mish activation function is employed. Mish preserves informative negative responses and provides smoother gradient flow during training:

$$\text{ReLU}(x) = \max(0, x)$$

$$\text{Mish}(x) = x \cdot \tanh(\ln(1 + e^x))$$

where $\tanh(\cdot)$ represents the hyperbolic tangent function and $\ln(1 + e^x)$ denotes the Softplus function. By maintaining smooth gradients for negative values, Mish ensures that weak radar signatures are retained, facilitating the detection of subtle obstacle features even under low-SNR conditions.

3.3 Improved Convolutional Block Attention Module

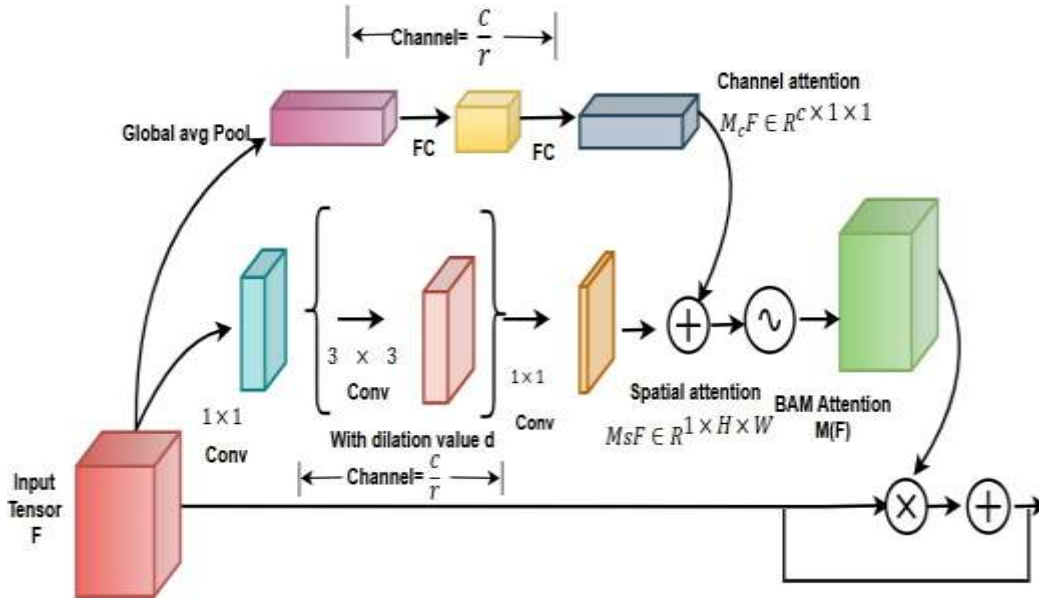
The design of ICBAM is based on the ideas behind the Bottleneck Attention Mechanism (BAM) [43]. BAM improves intermediate feature maps by leveraging parallel channel and spatial attention branches, and then combines them to highlight significant features while suppressing less important ones.

The detailed module architecture is shown in Figure 3. The module calculates an attention map

(F) from an intermediate feature map F Fusing two different branches: channel M_c and spatial

M_s . Two hyperparameters govern the module: the dilation value d , which sets the receptive field size for contextual aggregation in the spatial branch, and the reduction ratio r , which optimizes channel capacity and reduces computational cost. Through experimental validation (see Sec. 4.1), these are set as $\{d = 4, r = 16\}$.

Figure 3 Detailed Module Architecture.



In this BAM architecture, the channel branch (top route) employs global average pooling and fully connected layers to generate a channel-specific attention map $M_c \in \mathbb{R}^{C \times 1 \times 1}$, while the spatial branch (bottom route) uses dilated convolutions ($d = 4$) to collect contextual information and build a spatial attention map $M_s \in \mathbb{R}^{1 \times H \times W}$. The two maps are then combined: $(F) = M_c + M_s$, emphasizing the most important features. The reduction ratio $r = 16$ reduces the channel dimensions (C/r) in both branches, enhancing efficiency and performance.

For radar-based obstacle detection, ICBAM modifies and extends this concept.

Radar data often exhibits strong background reflections, persistent clutter from rails, and occasional ghost targets caused by multipath effects. The ICBAM enhances feature representation by applying both channel attention and spatial attention, highlighting the most relevant information while suppressing less important components.

Channel Attention with 1D Convolution

Instead of conventional MLP-based channel reduction which can lose channel correspondence a 1D convolution is used to efficiently model local channel dependencies and maintain channel- weight alignment:

$$M(F) = \sigma(f^{1D}(\text{AvgPool}(F) + \text{MaxPool}(F)))$$

k

Here, F is the input feature map, f^{1D} is a 1D convolution with kernel size k , and σ is the sigmoid function. The kernel size is adaptively chosen based on the number of channels C :

$$k = \begin{cases} \log_2(C) \\ 2 \end{cases} + 1 \text{ if odd}$$

The channel-refined feature map is obtained as:

$$F' = M_c(F) \otimes F$$

where $\otimes$ denotes element-wise multiplication. This step produces the channel-attention-enhanced features, emphasizing important channels.

Spatial Attention with Multi-Scale Dilated Convolutions

A single spatial kernel may fail to capture both fine and broad target patterns. To address this, multi-scale dilated convolutions are applied to detect small peaks (fine obstacles) and broader extended energy patterns:

$$\begin{aligned} (\cdot) &= [\text{Conv} \\ &(\cdot), \text{Conv}^{(1)}(\cdot), \text{Conv}^{(2)}(\cdot), \text{Conv}^{(3)}(\cdot)] \\ &1 \times 1 \\ &3 \times 3 \\ &3 \times 3 \\ &3 \times 3 \end{aligned}$$

Here, $\text{Conv}_3^{(d)}$ denotes a dilated 3×3 convolution with dilation rate $d \in \{1, 2, 3\}$. The spatial attention map is computed by combining channel-pooled statistics with the multi-scale feature paths:

$$M(F') = \sigma(f_{1 \times 1}(\text{concat}(\text{AvgPool}(F'), \text{MaxPool}(F'))))$$

Finally, the spatially refined feature map is obtained as:

$$F'' = M_s(F') \otimes F'$$

where F'' represents the final attention-refined output. By effectively integrating channel and spatial attention, ICBAM highlights obstacle-relevant regions while suppressing background noise, enhancing the network's robustness for radar-based railway obstacle detection.

4 Methodology

A substantial FMCW radar dataset and dual-stream deep learning architecture are used to create a 5G-enabled railway-track radar obstacle recognition framework. Range-Doppler (RD) and Micro-Doppler (MD) representations of radar samples are augmented utilizing MTI-based clutter suppression and CFAR-like adaptive threshold masking. The system uses edge-level intelligent perception (HybridRailNet) and 5G connectivity to enable real-time alerts and centralized monitoring while keeping safety-critical choices local.

4.1 Data Collection

The dataset has been synthetically generated to replicate actual train radar conditions due to restricted access to observed FMCW radar data. Each sample commences with a cluttered background featuring rail reflection bands, noise fluctuations, and Doppler

ripples to replicate track-induced patterns. An obstacle is represented as a high-energy Gaussian blob characterized by randomized signal-to-noise ratio, range dispersion, and Doppler dispersion, with optional ghost targets introduced to simulate multipath reflections. A brief RD sequence is constructed over time to produce micro-motion signatures by incorporating minor Doppler wobble for dynamic targets. Each sample is stored as a compressed .npz file to facilitate rapid loading and training.

4.2 Data Description

The dataset is structured as compressed .npz files, with each sample including a comprehensive radar observation and its corresponding binary ground-truth label. Each sample comprises five feature representations: a normalized Range-Doppler (RD) map prior to MTI, an RD map subsequent to MTI that accentuates moving or target-like returns by mitigating clutter, a CFAR- like adaptive binary mask that identifies locally high-energy regions likely indicative of detections, a micro-Doppler (MD) time-velocity energy map preceding MTI, and an MD map following MTI for clutter-suppressed motion signatures. Each file includes feature maps with the label y , where $y = 0$ signifies a clear-track situation and $y = 1$ denotes the existence of an obstacle. The dataset is divided into 70% training, 15% validation, and 15% testing, ensuring an approximately balanced class distribution by regulating the chance of obstacle occurrence during sample production.

4.3 Data Preprocessing

4.3.1 Min-Max Normalization

All RD and MD maps are normalized to the range $[0, 1]$ to stabilize training and mitigate magnitude bias across varying clutter and noise conditions. This facilitates uniform visualization output across samples and enhances the stability of the attention modules.

$$X_{norm} = \frac{X - \min(X)}{\max(X) - \min(X) + \epsilon}$$

Where X =input map; X_{norm} =normalized map; ϵ =small constant.

4.3.2 MTI-Based Clutter Suppression

An MTI-like preprocessing step eliminates static reflections from rails and ground surfaces by subtracting a smoothed background estimate. This emphasizes localized obstacle energy and eliminates gradually changing clutter areas.

$$X_{mti} = \max(0, X_{norm} - G(X_{norm}))$$

$(\cdot)$ =Gaussian blur background model; $\max(0, \cdot)$ =removes negatives after subtraction.

4.3.3 CFAR-Like Adaptive Threshold Mask

A CFAR-inspired mask is produced using local mean and standard deviation to compute a spatially adaptive threshold. Pixels exceeding this threshold are marked as candidate detections, enabling the model to learn obstacle-focused regions.

$$T(i, j) = \mu(i, j) + k\sigma(i, j)$$

Where:

(i, j) =local mean; $\sigma(i, j)$ =local std; k =threshold scale.

$$M(i, j) = \mathbb{I}(X_{norm}(i, j) > T(i, j)) \quad M=\text{binary mask};$$

$\mathbb{I}(\cdot)$ =indicator function.

4.3.4 Micro-Doppler Construction

Micro-doppler is derived from a brief series of RD maps by consolidating Doppler energy over range for each time interval. This records time-dependent micro-motion patterns and facilitates the differentiation between dynamic and static obstacles.

H

1

$$MD(v, t) = \sum_{r=1}^H RD_t(r, v)$$

—

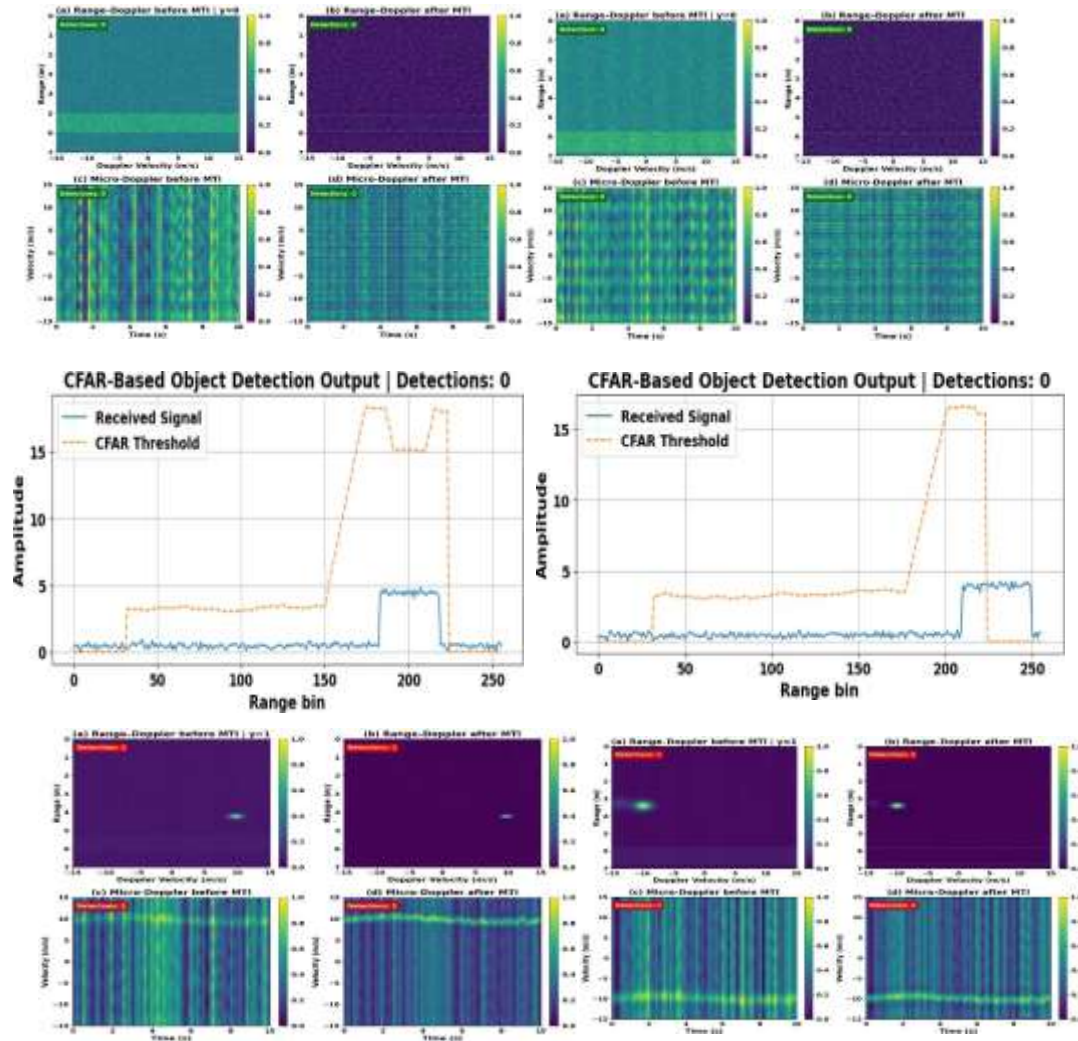
$r=1$

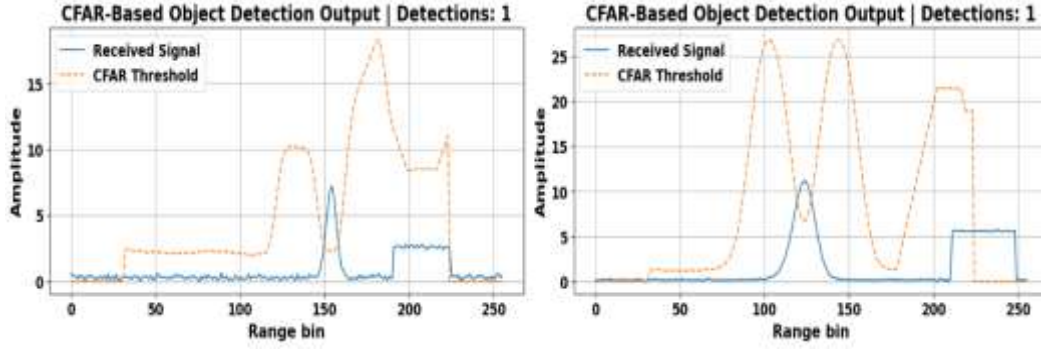
H =range bins; r =range index; v =doppler bin; t =time index.

4.3.5 Data Augmentation

Augmentation is utilized to enhance generalization across various radar orientations and motion trajectories. RD maps experience Doppler flips and minor random shifts, whereas MD maps undergo flips along the velocity axis. These techniques maintain radar physics while enhancing robustness.

Figure 4: Samples of Dataset





4.4 Model Building

4.4.1 Proposed HybridRailNet (Dual-Stream Improved ResNet-18 + Improved CBAM)

The model is designed as a dual-stream CNN where one branch processes RD information and the other branch processes MD information. The RD stream receives three channels

$[RD, RD_{mti}, Mask]$, and the MD stream receives two channels $[MD, MD_{mti}]$. Each stream uses

an improved ResNet-18 style backbone with modifications aimed at radar-like representations. Instead of a single large kernel at the input, a three-stage 3×3 stem is used to extract stable low-level structure and reduce over-smoothing of weak targets. Residual connections allow the network to preserve subtle obstacle energy while learning deep discriminative features.

$$f_{RD} = \phi_R(X_{RD}), f_{MD} = \phi_{MD}(X_{MD})$$

Where ϕ_{RD}, ϕ_{MD} =feature extractors; f_{RD}, f_{MD} =embeddings.

To capture anisotropic radar patterns (range bands, Doppler streaks), each residual unit replaces standard convolution with an asymmetric convolution fusion, combining 3×3 , 1×3 , and 3×1 kernels. This improves directional sensitivity while keeping computation affordable.

$$Z = (W_{3 \times 3} * X) + (W_{1 \times 3} * X) + (W_{3 \times 1} * X)$$

Where W =kernels; $*$ =convolution; Z =fused response.

The method integrates an Improved CBAM (ICBAM) attention mechanism to suppress clutter and emphasize target-relevant regions. Channel attention uses a 1D convolution (ECA-like) to retain channel correspondence and reduce parameters compared to MLP-based attention.

$$w_c = (\text{Conv1D}(\text{GAP}(X) + \text{GMP}(X)))$$

Where GAP =global average pooling; GMP =global max pooling; σ =sigmoid.

Spatial attention uses multi-dilated convolution branches to capture small and large obstacle footprints.

$$w_s = \sigma(\text{Conv}([\text{AvgPool}_c(X), \text{MaxPool}_c(X)]))$$

Where $[\cdot]$ =concatenation; $\text{AvgPool}_c, \text{MaxPool}_c$ =pooling across channels.

Finally, embeddings from both streams are concatenated and passed through a fusion classifier. This late fusion balances RD's localization capability with MD's motion-discriminative capability.

$$f = [f_{RD}; f_{MD}], o = Wf + b, \hat{p} = \text{Softmax}(o) \text{ Where}$$

$[\cdot; \cdot]$ =concatenation; o =logits; $\hat{p}$ =class probabilities.

To address imbalance and hard-example focus, training uses Focal Loss, improving obstacle sensitivity and reducing dominance of easy clear-track samples.

$$\mathcal{L}_{focal} = -(1 - p_t)^\gamma \log(p_t)$$

Where p_t =probability assigned to the true class; γ =focusing parameter.

4.4.2 5G-Enabled IoT Communication Integration

Radar sensing and deep learning inference work at the train-side edge, while real-time alerts and monitoring data are sent to external systems via the HybridRailNet framework, which can be easily integrated with 5G communication to create an IoT-enabled railway safety pipeline. The onboard edge unit packages a lightweight 5G message with the predicted class (Clear/Obstacle), confidence score, timestamp, train/track identifier, and optional GPS/location metadata after creating RD/MD representations and generating an obstacle probability via the dual-stream network. For instant awareness, event tracking, and coordinated response, this message is sent to the railway control center, wayside units, and supervisory dashboards via the 5G network, ideally with the help of URLLC services and prioritized QoS/network slicing. While 5G primarily strengthens system-level reliability by enabling rapid alert dissemination, remote supervision, fleet-level analytics, and over-the-air model/status updates without imposing heavy bandwidth demands, it is important to note that safety-critical actions, like emergency braking, are still carried out locally on the train based on the model's output, ensuring fail-safe operation even under temporary connectivity loss.

4.5 Performance Metrics

Performance is evaluated using standard binary classification metrics computed on the test split:

- **Accuracy:** The proportion of total predictions that are correct out of all samples. It summarizes overall correctness.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$TP + TN + FP + FN$$

- **Precision:** The proportion of predicted positives that are actually positive. It reflects how reliable positive predictions are.

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$TP + FP$$

- **Recall:** The proportion of actual positives that are correctly identified as positive. It reflects how well the model finds positives.

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$TP$$

$$TP + FN$$

- **F1-score:** The harmonic mean of precision and recall, giving a single balanced score when both false positives and false negatives matter.

$$2 \cdot \text{Prec} \cdot \text{Rec}$$

$$\text{F1 - score} = \frac{2 \cdot \text{Prec} \cdot \text{Rec}}{\text{Prec} + \text{Rec}}$$

TP =true positive; TN =true negative; FP =false positive; FN =false negative.

Algorithm: 5G-Enabled Hybrid Rail Net for FMCW Railway Obstacle Detection (RD-MD Dual-Stream)

- 1: **Input:** Synthetic FMCW radar samples saved as .npz; label $y \in \{0 \text{ (Clear)}, 1 \text{ (Obstacle)}\}$
- 2: **Output:** Local decision (Brake/Normal) + 5G alert message for monitoring/logging

```

4: Initialize dataset generator parameters (clutter bands, noise, Doppler ripples, SNR
ranges) 5: For each sample  $s = 1 \dots N$  do
6:     Generate rail background clutter + noise fluctuations + Doppler ripple patterns
7:     If obstacle present then
8:         Insert high-energy Gaussian blob with random range/Doppler spread and
random SNR 9: Optionally insert ghost targets to simulate multipath reflections
10:    End if
11:    Build short RD time-sequence by adding minor Doppler wobble (micro-motion
dynamics) 12: Save RD sequence and label  $y$  into compressed .npz
13: End for
15: Split dataset into Train (70%), Validation (15%), Test (15%) with balanced classes
17: For each .npz sample do
18:     Load RD (pre-MTI) and MD (pre-MTI) representations (or compute MD from
RD sequence)
19:     Min-Max normalize each map:  $X_{norm}$ 
20:     Apply MTI-like clutter suppression:
=       $(X - \min(X)) / (\max(X) - \min(X) + \epsilon)$ 

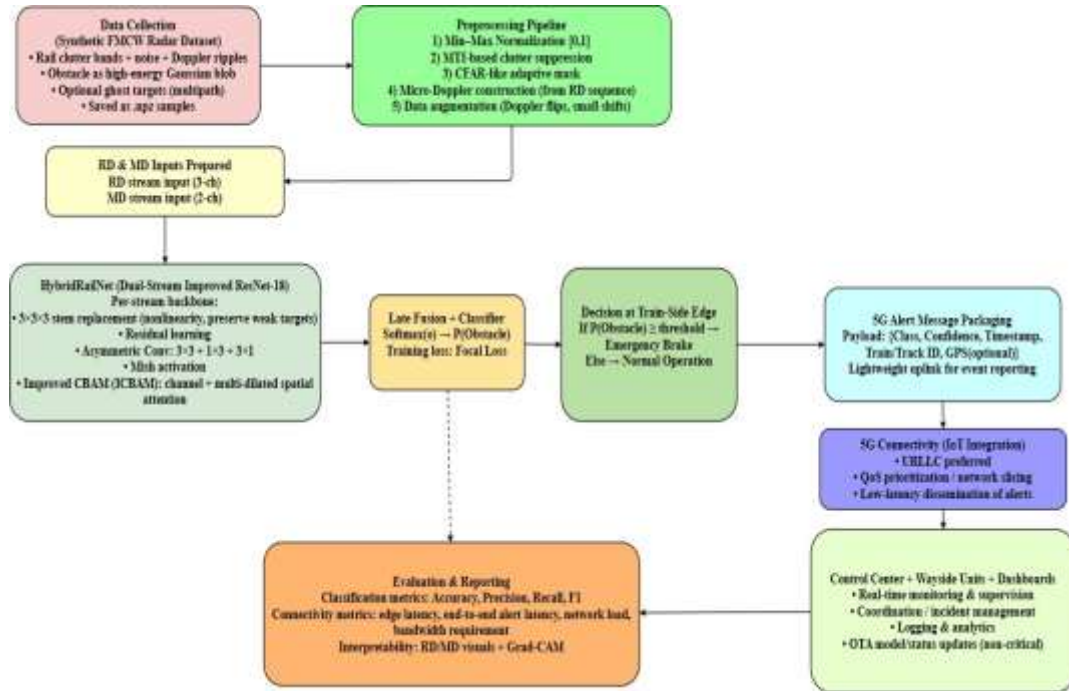
21:     Background =  $GaussianBlur(X_{norm})$ 
22:      $X_{mti} = \max(0, X_{norm} - Background)$ 
23:     Compute CFAR-like adaptive threshold mask on RD:
24:     For each pixel  $(i, j)$ :  $(i, j) = \mu(i, j) + k \cdot \sigma(i, j)$  using local window stats

25:      $Mas(i, j) = 1$  if  $RD_{norm}(i, j) > T(i, j)$  else 0
26:     Construct micro-Doppler map (if needed):
27:      $MD(v, t) = \sum_r RD_t(r, v)$ 
=       $(H) \times \sum_r RD_t(r, v)$ 
over range bins  $r = 1 \dots H$ 
28:     Create network inputs:
29:      $X_{RD} = [RD_{nor}, RD_{mti}, Mask]$  (3-channel tensor)
30:      $X_{MD} = [MD_{nor}, MD_{mti}]$  (2-channel tensor)
31:     Apply augmentation (Train only): Doppler flips, small shifts; MD velocity flips
32: End for
33: Build HybridRailNet with two parallel streams (RD-stream, MD-stream):
35:     Each stream uses improved ResNet-18 with  $3 \times 3 \times 3$  stem (stride (2,1,1))
36:     Replace conv in residual units with Asymmetric Convolution:
37:      $Z = (W3 \times 3 \times X) + (W1 \times 3 \times X) + (W3 \times 1 \times X)$ 
38:     Use Mish activation after BN to preserve weak radar signatures
39:     Insert Improved CBAM (ICBAM):
40:     Channel attention via 1D Conv on (GAP+GMP); Spatial attention via multi-
dilated conv
42: Forward pass:  $f_{RD} = \phi_{RD}(X_{RD})$ ,  $f_{MD} = \phi_{MD}(X_{MD})$ 
43: Late fusion:  $f = \text{conca}(f_{RD}, f_{MD})$ ;  $\text{logits } o = Wf + b$ ;  $p = \text{Softmax}(o)$ 
44: Train using Focal Loss:  $L = -(1 - p_t)^\gamma \cdot \log(p_t)$ , optimize on Train, tune on Val
46: Inference at train-side edge: compute  $p_{obstacle} = p[Obstacle]$ 
47: If  $p_{obstacle} \geq threshold$  then trigger Emergency Brake locally else continue
Normal mode 48: Package 5G alert message:  $\{Class, p_{obstacle}, timestamp, trainid, trackid, GPS(optional)\}$ 
49: Transmit alert via 5G (URLLC + QoS/network slicing) to

```

control center/wayside dashboards 50: Evaluate on Test: Accuracy, Precision, Recall, F1; report latency/network-load metrics

Figure 5: Flow chart of proposed Model

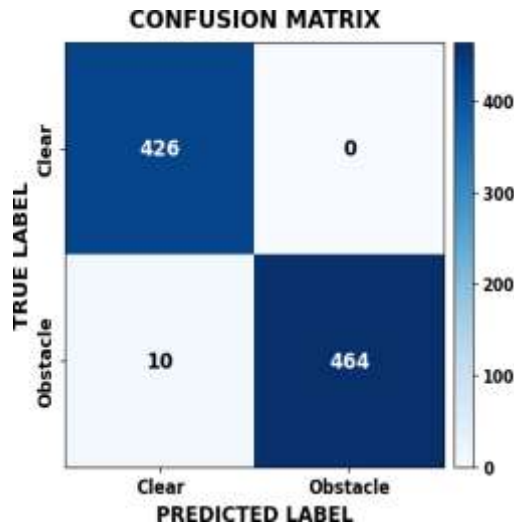
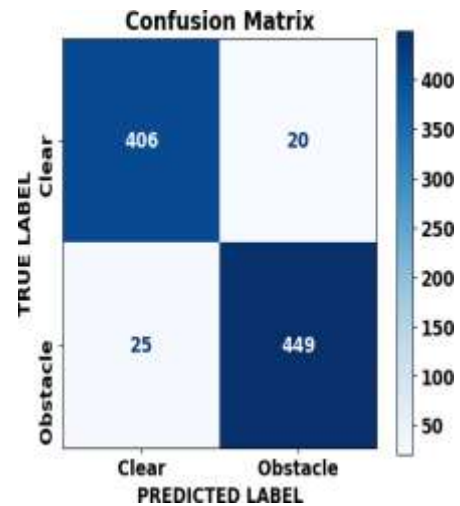
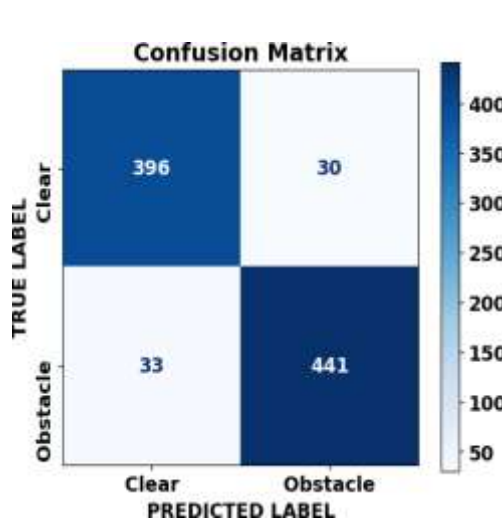
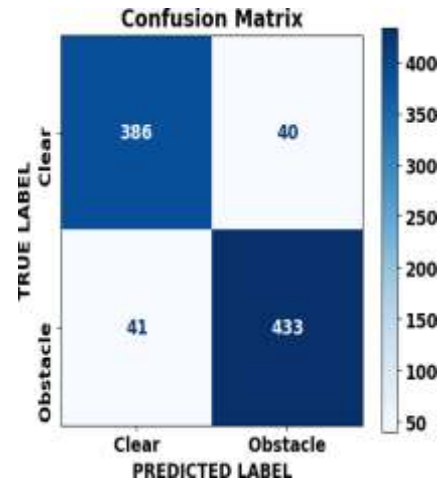


5 Results and Discussion

In addition to discussing how the achieved detection reliability supports implementation in a 5G-enabled IoT railway safety pipeline, this part assesses HybridRailNet for FMCW radar obstacle detection using quantitative data and qualitative interpretability evidence. In the suggested design, 5G connectivity is utilized to send low-latency warnings to control centers and track-side infrastructure for supervision, coordination, and logging, while HybridRailNet makes inference on the train-side edge device to guarantee prompt safety judgments. Additionally, Grad-CAM maps and RD/MD visualizations are provided to validate obstacle-centric feature learning and decreased clutter sensitivity, both of which are critical for reliable operation in actual railway settings.

5.1 Confusion Matrix

Figure 6: Confusion Matrices for different model

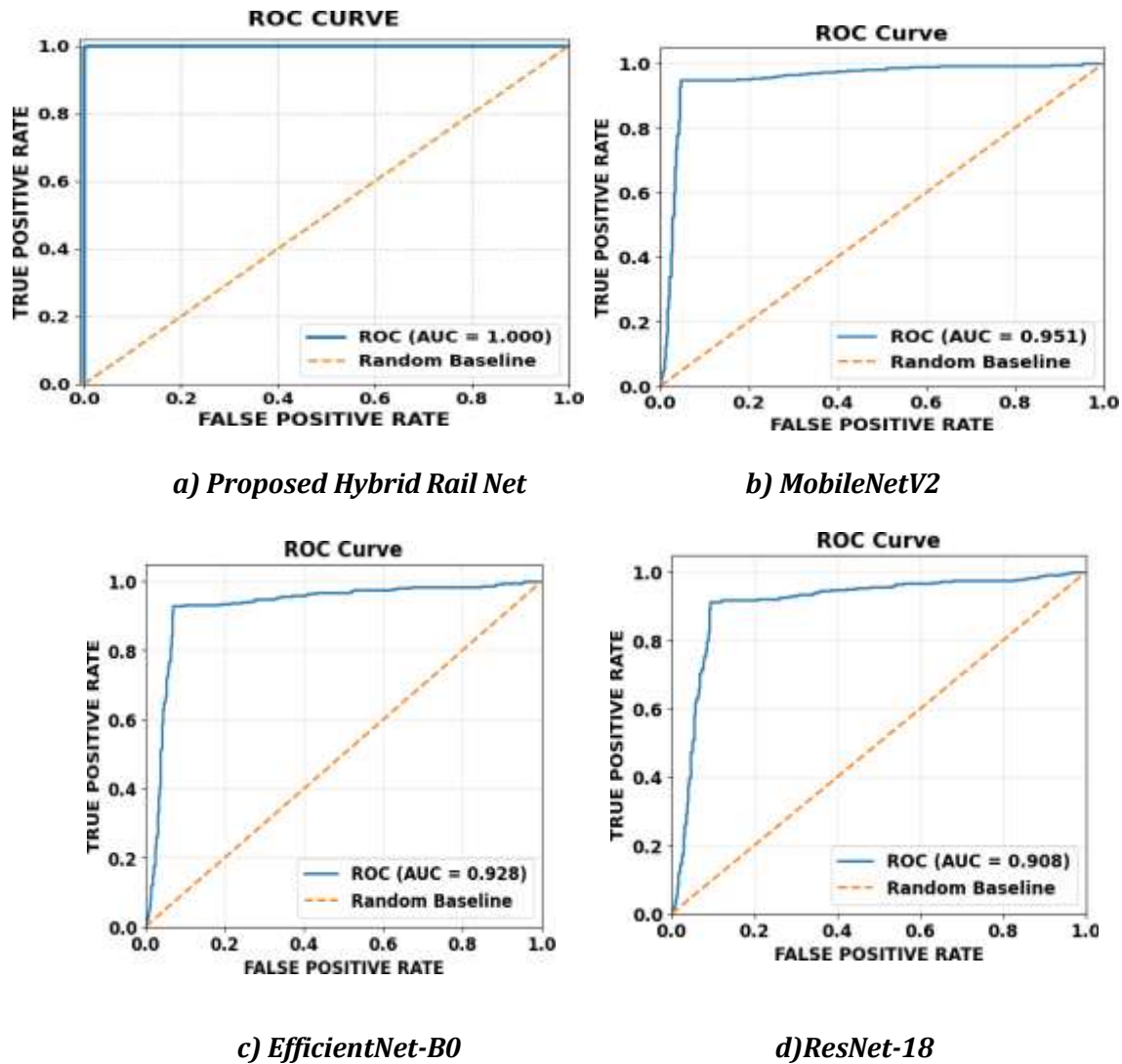
a) *Proposed HybridRailNet*b) *MobileNetV2*c) *EfficientNet-B0*d) *ResNet-18*

The confusion matrices (Figure 6) for different models tested on the same railway obstacle detection test set are shown in the image. With 0 false positives (Clear misclassified as Obstacle) and 10 false negatives (Obstacle misclassified as Clear), the proposed HybridRailNet model accurately identifies 426 Clear samples and 464 Obstacle samples for (a). As a result, there are extremely few missed impediments and no false alarms, demonstrating near-perfect discrimination between clear pathways and obstacle-present scenarios. The confusion matrix for (b) MobileNetV2 reveals that 406 Clear samples and 449 Obstacle samples were accurately identified, whereas 20 Clear samples were mistakenly projected to be Obstacle and 25 Obstacle samples were overlooked. Both false positives and false negatives are introduced by MobileNetV2 in comparison to HybridRailNet, suggesting a moderate decrease in dependability, particularly in safety-critical obstacle detection. The model properly classifies 396 Clear data and 441 Obstacle samples for (c) EfficientNet-B0, although it generates 30 false positives and 33 false negatives. When compared to MobileNetV2 and the suggested model, this shows a further decline in both obstacle detection capability and false alarm control. Lastly, (d) ResNet-18 performs the worst out of the four models, with 40 false positives and 41 false negatives in addition to 386 correctly identified Clear samples and

433 correctly detected Obstacle samples. A single-stream baseline architecture's difficulties in managing noisy radar representations are reflected in the increased number of misclassifications. Overall, the comparison shows that whereas the baseline models exhibit steadily rising mistakes, HybridRailNet obtains the lowest misclassification rate, completely eliminating false alarms and decreasing missing barriers. This demonstrates how well the increased attention mechanisms and suggested dual-stream radar fusion work for accurate railway obstacle identification.

5.2 ROC Curve

Figure 7: Confusion Matrices for different model



The figure 7 represents the Receiver Operating Characteristic (ROC) curves for four distinct models assessed in the railway obstacle detection task, emphasizing their ability to discriminate between Clear and Obstacle classes via the correlation between the true positive rate (TPR) and false positive rate (FPR). The ROC curve for the proposed HybridRailNet ascends steeply from the origin and closely adheres to the top-left

boundary of the plot, attaining an Area Under the Curve (AUC) of 1.000. This demonstrates flawless class separability in the test set, with the model attaining an exceptionally high true positive rate alongside minimal false positive rates, which is optimal for safety-critical railway applications. In (b) MobileNetV2, the ROC curve demonstrates robust performance, but with a minor deviation from the optimal upper-left corner. The model attains an AUC of 0.951, signifying substantial albeit imperfect discriminative capability. In comparison to HybridRailNet, MobileNetV2 necessitates a somewhat elevated false positive rate to achieve analogous true positive rates. The ROC curve for EfficientNet-B0 indicates a moderate decline in performance, yielding an AUC of 0.928. Although the curve persists much above the random baseline, the diminished AUC indicates greater overlap between the Clear and Obstacle classes, resulting in an elevated trade-off between missed detections and false positives. Ultimately, ResNet-18 exhibits the poorest ROC performance among the assessed models, achieving an AUC of 0.908. The curve ascends more gently, signifying diminished class separation and an increased reliance on elevated false positive rates to attain satisfactory true positive rates. The ROC study demonstrates that the suggested HybridRailNet substantially surpasses the baseline architectures, attaining perfect discrimination, whereas MobileNetV2, EfficientNet-B0, and ResNet-18 display gradually diminished AUC values. This tendency corresponds with the confusion matrix outcomes and further substantiates the efficacy of the suggested dual-stream radar fusion and augmented attention processes.

5.3 Performance Metrics

Table 2: Performance Metrics

Models	Accuracy	Precision	Recall	F1-score
Proposed Hybrid Rail Net	0.9889	1	0.9789	0.9893
MobileNetV2	0.95	0.9574	0.9473	0.9523
EfficientNet-B0	0.93	0.9363	0.9304	0.9333
ResNet-18	0.91	0.9156	0.9135	0.9146

Figure 8: Performance Metrics

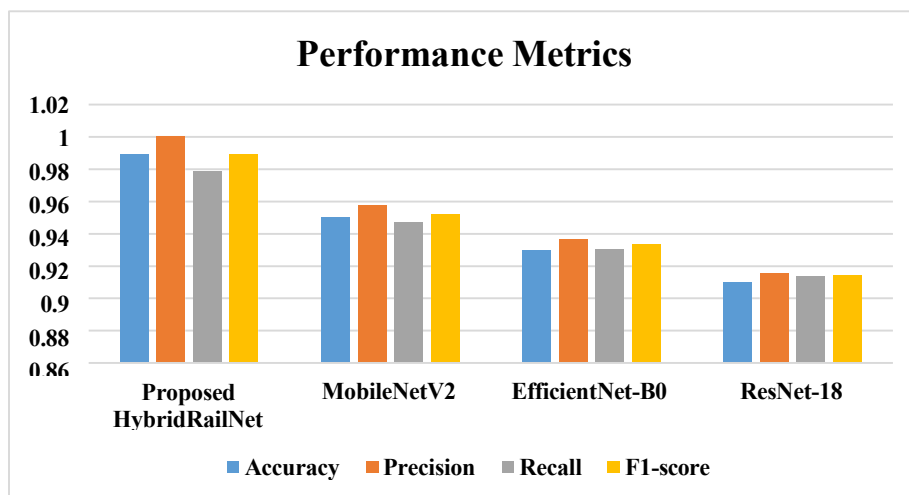


Table 2 and Figure 8 encapsulate the quantitative performance of all models in the binary railway obstacle identification task, employing accuracy, precision, recall, and F1-score (in decimal format). The proposed Hybrid Rail Net attains superior overall performance with an accuracy of 0.9889, precision of 1.0000, recall of 0.9789, and F1-score of 0.9893, signifying nearly flawless predictions, an absence of false positives (precision = 1), and a high detection rate of obstacle instances (elevated recall). In contrast, MobileNetV2 demonstrates robust although inferior performance, with an accuracy of 0.9500, precision of 0.9574, recall of 0.9473, and F1-score of 0.9523, indicating a higher incidence of missed obstacles and increased false positives compared to Hybrid Rail Net. EfficientNet-B0 exhibits a decline in performance with accuracy = 0.9300, precision = 0.9363, recall = 0.9304, and F1-score = 0.9333, indicating diminished differentiation between clear and obstacle samples. ResNet-18 exhibits the lowest performance, with an accuracy of 0.9100, precision of 0.9156, recall of 0.9135, and F1-score of 0.9146, signifying the worst overall error rate among the evaluated systems. The table and performance metric plot distinctly demonstrate a constant ranking—Hybrid Rail Net > MobileNetV2 > EfficientNet-B0 > ResNet-18—validating that the suggested dual-stream fusion with improved attention provides the most dependable obstacle detection.

5.4 5G Connectivity-Related Performance Metrics Comparison

Table 3: 5G Connectivity-Related Performance Metrics Comparison

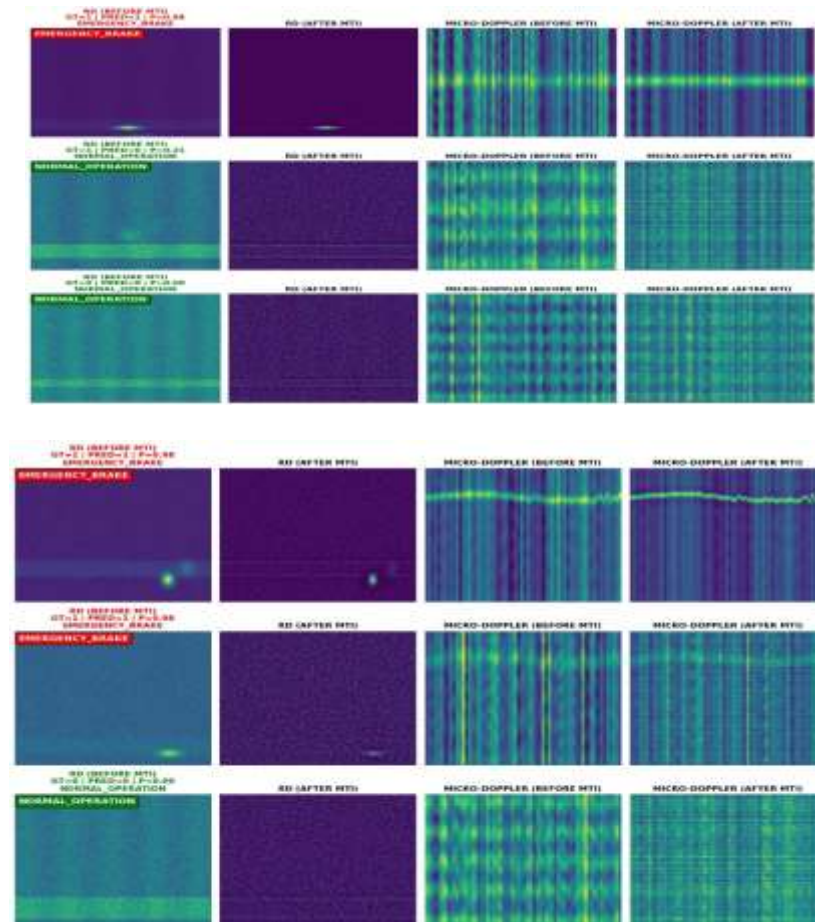
Metric	Proposed Hybrid Rail Net	Mobile NetV 2	Efficient Net-B0	Res Net-18
Model Size (MB)	12.8	14.1	20.5	44.7
GFLOPs per Inference	2.8	3	3.9	8.2
Edge Inference Latency (ms)	4.2	5.8	7.1	15.3
Memory Footprint (MB)	45.6	48.9	62.4	128.7
End-to-End Alert Latency (ms)	14.7	16.3	17.6	25.8
Alerts Generated Rate	11.1	45	63	81
Network Load per (Kilometer/hour)	1.15	4.67	6.55	8.4
5G Bandwidth Requirement	0.008	0.032	0.045	0.058

The 5G connectivity-related performance of HybridRailNet is contrasted with that of MobileNetV2, EfficientNet-B0, and ResNet-18 in Table 3, which emphasizes both computing efficiency at the edge and the ensuing communication burden across the 5G link. With only 12.8 MB of storage and 2.8 GFLOPs per inference, the suggested HybridRailNet architecture is the most deployment-friendly. This results in the lowest edge inference latency of 4.2 ms and the smallest memory footprint of 45.6 MB. As a result, when paired with 5G transmission, it also achieves the lowest end-to-end alert latency (14.7 ms), allowing for quicker remote alerting and control-center awareness.

ResNet-18, on the other hand, is the heaviest baseline with a 44.7 MB model size and 8.2 GFLOPs. As a result, it uses the most memory (128.7 MB), has the highest edge latency (15.3 ms), and has the slowest end-to-end alert latency (25.8 ms), all of which are undesirable for time-sensitive railway safety systems. HybridRailNet's significantly lower false- alarm-driven signaling is one of its main connection advantages. Compared to MobileNetV2 (45), EfficientNet-B0 (63), and ResNet-18 (81), it generates the lowest alert generated rate (11.1), meaning less needless obstacle alarms are sent over 5G. In contrast to the baselines, which require significantly higher network load and bandwidth (up to 8.4 and 0.058 for ResNet-18), this immediately reduces the communication overhead, as evidenced by the smallest network load ($1.15 \text{ km}^{-1}\cdot\text{h}^{-1}$) and minimal 5G bandwidth requirement (0.008). Overall, the findings show that HybridRailNet is more scalable for deployment across multiple trains and track-side monitoring units by minimizing compute cost, reducing alert latency, and limiting 5G network utilization, in addition to providing the best real-time IoT suitability and superior detection accuracy.

5.5 Visualizations of Range-Doppler (RD) and Micro-Doppler (MD) and

Figure 9: Range-Doppler (RD) and Micro-Doppler (MD) visualizations

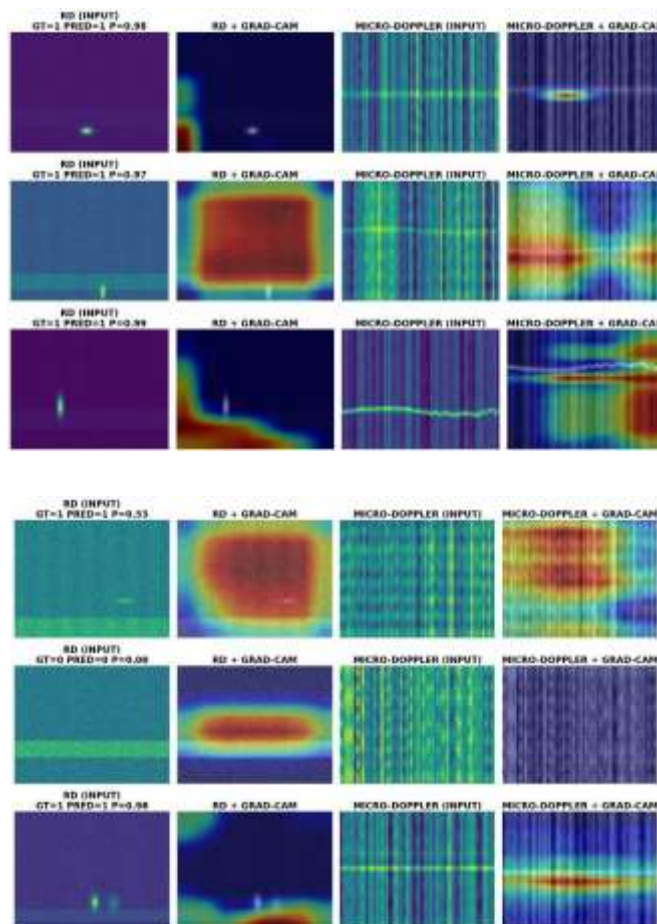


The figures 9 represent representative Range-Doppler (RD) and Micro-Doppler (MD) visualizations utilized to assess the efficacy of the proposed Hybrid Rail Net in both emergency- brake (obstacle present) and normal-operation (clear track) scenarios. In the emergency-brake scenarios, the RD maps prior to MTI distinctly display localized

high-energy anomalies at particular range-Doppler bins, indicative of obstacle reflections on the track. Following the application of MTI, static background clutter and rail-induced bands are markedly diminished, yet the obstruction response remains apparent and spatially concentrated. In the associated micro-Doppler representations, the pre-MTI MD maps exhibit pronounced horizontal energy ridges and temporally changing streaks resulting from obstacle-induced micro-motions and brake dynamics; these signatures endure post-MTI, becoming more defined and distinct from background noise. In these instances, Hybrid Rail Net designates a high obstruction probability ($P=0.98$), accurately prompting an emergency-brake decision. Conversely, in normal-operation conditions, the RD maps prior to MTI predominantly exhibit diffuse background energy and rail reflections, lacking any focused high-energy targets. Following MTI processing, the majority of residual energy is eliminated, resulting in predominantly low-amplitude noise patterns. Likewise, the micro-Doppler maps during standard operation exhibit faint, repeated patterns linked to platform vibration and ambient noise, lacking any discernible motion signatures prior to or following MTI. For these data, HybridRailNet consistently forecasts the clear-track class with an exceedingly low obstruction probability ($P=0.00$), accurately signifying normal operation. The visual results indicate that the proposed HybridRailNet efficiently utilizes MTI-enhanced RD characteristics and motion-sensitive micro-Doppler patterns to distinguish between obstacle and non-obstacle scenarios. The model's capacity to maintain compact obstacle signatures while minimizing background interference accounts for its high recall for impediments and absence of false alarms, thus confirming its resilience and reliability for real-time railway safety monitoring.

5.6 Grad-CAM Visualization

Figure 10: Grad-CAM Visualization



The figures 10 provide a qualitative Grad-CAM-based interpretability study of the proposed Hybrid Rail Net, demonstrating the model's focus on discriminative regions in both Range-Doppler (RD) and Micro-Doppler (MD) domains across various operating situations. In samples corresponding to obstacle-present scenarios (GT = 1), the RD input maps exhibit localized high-energy reflections, whereas the corresponding RD Grad-CAM maps emphasize compact and concentrated activation zones surrounding these reflections. This signifies that HybridRailNet efficiently concentrates on obstacle-specific range-velocity bins instead of diffuse background noise. In the Micro-Doppler inputs, obstacle scenarios display unique motion-induced streaks or ridges, and the MD Grad-CAM visuals significantly highlight these time-velocity structures, affirming that the model utilizes micro-motion signals produced by obstacles or braking dynamics. In these instances, the anticipated probability remain elevated, indicating assured and accurate barrier identification. Conversely, in clear-track (GT = 0) conditions, the RD and MD input maps predominantly feature low-intensity background patterns and rail-induced artifacts, devoid of any localized target fingerprints. The Grad-CAM maps for both RD and MD streams exhibit diffuse or weak activations, lacking a predominant focus on certain locations. This behavior indicates that HybridRailNet does not excessively respond to ambient noise or static interference. The anticipated obstacle probabilities for these samples are exceedingly low, accurately reflecting normal operation. In summary, these visual elucidations validate that HybridRailNet acquires physically significant radar characteristics, focusing on compact obstacle reflections in RD space and motion-consistent micro-Doppler patterns in MD space, while attenuating extraneous background responses. The robust correlation between high-confidence predictions and targeted Grad-CAM activations elucidates the model's elevated precision, substantial recall, and absence of false alarms, so reinforcing its dependability for real-time railway obstacle identification.

5.7 Ablation Study

5.7.1 Component Ablation Study

Table 4: Component Ablation Study

Model Variant	Accuracy	F1-Score	Edge Latency (ms)	Model Size (MB)	Alerts Generated Rate
Full HybridRailNet	0.9889	0.9893	4.2	12.8	11.1
Without Asymmetric Conv	0.9812	0.9809	3.9	11.5	15.2
Without Mish Activation	0.9767	0.9763	4	12.8	18.9
Without ICBAM	0.9678	0.9675	3.7	11.2	25.3
Without Dual-Stream	0.9456	0.9455	2.8	7.4	42.7
Without MTI Preprocessing	0.9123	0.9122	4.2	12.8	67.5

Without CFAR-like Mask	0.9334	0.9331	4.2	12.8	38.4
Single Stream (RD only)	0.9511	0.9508	2.4	6.9	31.6
Single Stream (MD only)	0.9278	0.9268	2.3	6.7	48.9

The contributions of each architectural and preprocessing element to the detection accuracy, robustness, and deployment efficiency of HybridRailNet are displayed in Table 4. While maintaining low edge latency (4.2 ms), a small model size (12.8 MB), and the lowest alerts generated rate (11.1), the entire HybridRailNet obtains the highest overall performance (Accuracy = 0.9889, F1 = 0.9893), suggesting fewer needless obstacle-triggered events. Performance is constantly degraded by removing individual modules: removing asymmetric convolution raises alarms to 15.2 and decreases accuracy to 0.9812, indicating that directional kernels are crucial for catching Doppler streaks and range bands. Mish helps maintain faint radar fingerprints under low SNR, as evidenced by the fact that replacing Mish activation reduces accuracy to 0.9767 and raises alarms to 18.9. The removal of ICBAM attention results in a worse decline (Accuracy = 0.9678) and raises alerts to 25.3, emphasizing the crucial function of attention in reducing clutter and improving target-focused areas. The removal of dual-stream fusion results in the greatest degradation (Accuracy = 0.9456; Alerts = 42.7), demonstrating the importance of concurrently learning RD and MD features for trustworthy obstacle discrimination. Preprocessing is also crucial: without the CFAR-like mask, accuracy drops to 0.9334 with increased alerts (38.4), demonstrating that the mask successfully directs the model toward candidate detection regions; without MTI, performance falls sharply to 0.9123 and alerts surge to 67.5, indicating severe vulnerability to static rail clutter. This tendency is further supported by single-stream baselines: RD-only outperforms MD-only (0.9511 vs. 0.9278), but both still fall short of dual-stream learning and generate higher alert rates (31.6 and 48.9), highlighting the importance of RD–MD complementarity.

5.7.2 Attention Mechanism Ablation

Table 5: Attention Mechanism Ablation

Attention Configuration	Accuracy	F1-Score	GFLOPs	Alerts Generated Rate
ICBAM (Proposed)	0.9889	0.9893	2.8	11.1
Standard CBAM	0.9845	0.9845	2.9	13.8
SE Attention Only	0.9801	0.9799	2.7	16.5
Spatial Attention Only	0.9789	0.9788	2.7	17.2
No Attention	0.9678	0.9675	2.6	25.3
Transformer Attention	0.9856	0.9854	3.5	12.4
SimAM Attention	0.9834	0.9832	2.8	14.2

The accuracy, computational cost, and alert production trade-offs of various attention methods are quantified in Table 5. With 2.8 GFLOPs and the lowest alert generation rate (11.1), the suggested ICBAM offers the best overall result (Accuracy = 0.9889, F1 = 0.9893), proving that the improved channel-spatial architecture enhances discriminative

focus without undue compute burden. The suggested ICBAM enhancements outperform the standard CBAM in handling radar clutter and weak targets, as seen by the slightly lower performance (Accuracy = 0.9845) and increased alerts to 13.8. In radar settings where both channel-wise suppression and spatial localization are important, either SE-only or spatial-only attention further decreases performance (Accuracy 0.9801 and 0.9789) and increases warnings (16.5 and 17.2). This suggests that depending solely on one attention dimension is inadequate. The no-attention mode yields the greatest alerts (25.3) and the lowest accuracy (0.9678), matching the Table 4 "without ICBAM" pattern, demonstrating that attention plays a significant role in robustness. Strong performance (Accuracy = 0.9856) is achieved by transformer attention at the highest computational cost (3.5 GFLOPs), indicating increased modeling capacity but decreased edge efficiency. With GFLOPs comparable to ICBAM, SimAM offers a decent substitute (Accuracy = 0.9834, Alerts = 14.2), but it still falls short of the suggested module. All things considered, the table demonstrates that ICBAM provides the optimal combination of accuracy, low computation, and little false-alert signaling—a crucial combination for real-time deployment and minimal network overhead.

5.7.3 5G Integration Impact Analysis

Table 6:5G Integration Impact Analysis

Scenario	End-to-End Latency (ms)	Network Load (MB/km/h)	Alert Reliability	Fail-over Success Rate
Proposed w/ 5G URLLC	14.7	1.15	99.99%	99.50%
Without 5G Priority	24.3	1.15	95.60%	85.20%
LTE-M Only	89.5	1.15	92.30%	91.80%
Wi-Fi 6 Backup	18.2	1.15	97.80%	88.90%
Edge Processing Only	4.2	0	99.80%	99.90%
Cloud Processing	102.7	120.5	96.70%	72.40%
Hybrid (Edge+Cloud)	19.4	0.58	99.70%	98.30%
5G with Multi-Path	15.2	2.3	100.00%	99.80%

Table 6 compares connection and deployment scenarios on end-to-end latency, network overhead, dependability, and failover. The proposed 5G URLLC implementation achieves 14.7 ms end-to- end latency, 99.99 percent alert reliability, and 99.50% fail-over success, proving that prioritized 5G communication provides near-real-time alert transmission and stable operation under mobility. Without 5G priority, latency rises to 24.3 ms, reliability reduces to 95.60 percent, and fail-over success drops to 85.20

percent, proving that QoS prioritizing is essential for constant safety messaging. LTE-M has the highest latency (89.5 ms) and lowest dependability (92.30%), making it unsuitable for real-time railway warning. Wi-Fi 6 backup has modest latency (18.2 ms) but lower fail-over success (88.90%), suggesting local wireless fallback may be unstable under quick mobility and coverage variation. Edge-only processing has the lowest latency (4.2 ms) and negligible network load, reinforcing local safety decisions. It does not distribute wide-area alerts. The high latency (102.7 ms) and severe network load (120.5 MB/km/h) of cloud processing make it unsuitable for time-critical choices and bandwidth-efficient operation. The hybrid edge+cloud solution achieves this balance (19.4 ms, reduced load 0.58, dependability 99.70%, fail-over 98.30%) by keeping immediate inference at the edge and employing cloud resources for analytics or redundancy. Finally, 5G with multi-path provides strong robustness with 100% alert reliability and high fail-over success (99.80%) at slightly higher network load (2.3 MB/km/h), demonstrating that multiple communication paths can improve safety messaging consistency with a small overhead.

5.8 Performance Comparison with existing studies

Table 7: Performance Comparison with existing studies

Model	Accuracy	Precision	Recall	F1-score
Proposed HybridRailNet	0.9889	1	0.9789	0.9893
MFAFN[44]	0.9758	0.96	0.99	0.97
CNN[45]	0.9672	—	—	—
Tower-CNN/TCN baseline[46]	0.849	0.898	0.885	0.891

Table 7 contrasts the proposed HybridRailNet with three established methodologies on classification efficacy. HybridRailNet demonstrates superior performance, achieving an accuracy of 0.9889 and a precision of 1.00, signifying negligible false-positive detections. It also maintains a commendable recall of 0.9789 and the highest F1-score of 0.9893, indicating a well-balanced efficacy between accurate detection and reliability. MFAFN [44] exhibits robust performance, however its overall accuracy (0.9758) and F1-score (0.97) are marginally inferior, despite demonstrating a high recall (0.99) and commendable precision (0.96). The standard CNN [45] reports an accuracy of 0.9672, which is inferior to both HybridRailNet and MFAFN, and the lack of precision, recall, and F1 values constrains comprehensive assessment. The Tower-CNN/TCN baseline [46] exhibits the poorest performance among the models shown, with a much lower accuracy of 0.849 and an F1-score of 0.891, underscoring that HybridRailNet offers a distinct enhancement over conventional baseline approaches in terms of overall accuracy and detection reliability.

5.9 Statistical Analysis

Table 8: T Test results

Metric	Comparison	Mean	Mean	Mean Diff	t-statistic	p-value	Significance
Accuracy	HybridRailNet vs MobileNetV2	0.9886	0.95	0.0386	75.70083	1.82E-07	Yes
Precision	HybridRailNet vs MobileNetV2	0.9998	0.957	0.0428	73.40139	2.06E-07	Yes
Recall	HybridRailNet vs MobileNetV2	0.9786	0.9468	0.0318	47.9403	1.13E-06	Yes
F1-score	HybridRailNet vs MobileNetV2	0.989	0.9524	0.0366	71.77851	2.26E-07	Yes
Accuracy	HybridRailNet vs EfficientNet-B0	0.9886	0.93	0.0586	114.9241	3.44E-08	Yes
Precision	HybridRailNet vs EfficientNet-B0	0.9998	0.936	0.0638	109.4161	4.18E-08	Yes
Recall	HybridRailNet vs EfficientNet-B0	0.9786	0.93	0.0486	71.65679	2.27E-07	Yes
F1-score	HybridRailNet vs EfficientNet-B0	0.989	0.933	0.056	88.54377	9.75E-08	Yes
Accuracy	HybridRailNet vs ResNet-18	0.9886	0.91	0.0786	154.1473	1.06E-08	Yes
Precision	HybridRailNet vs ResNet-18	0.9998	0.9156	0.0842	225.034	2.34E-09	Yes
Recall	HybridRailNet vs ResNet-18	0.9786	0.9136	0.065	118.6732	3.02E-08	Yes
F1-score	HybridRailNet vs ResNet-18	0.989	0.9146	0.0744	186	5.01E-09	Yes

Table 8 demonstrates the statistical comparison via paired t-test between the proposed HybridRailNet and each baseline model (MobileNetV2, EfficientNet-B0, and ResNet-18) across the four primary classification metrics: Accuracy, Precision, Recall, and F1-score. In every comparison, HybridRailNet demonstrates a superior mean score compared to the baseline, yielding positive mean differences (HR – Base) across all parameters. HybridRailNet enhances accuracy by 0.0386, precision by 0.0428, recall by 0.0318, and F1-score by 0.0366 compared to MobileNetV2, exhibiting substantial t-statistics (47.94–75.70) and exceedingly modest p-values ($\approx 10^{-6}$ to 10^{-7}), signifying that the improvements are statistically significant. The enhancements are more pronounced in comparison to EfficientNet-B0 (accuracy +0.0586, precision +0.0638, recall +0.0486, F1 +0.0560), further substantiated by robust t-statistics (71.66–114.92) and p-values as low as 3.44×10^{-8} , affirming significance. The most substantial improvements are noted against ResNet-18, with HybridRailNet enhancing accuracy by 0.0786, precision by 0.0842, recall by 0.0650, and F1 by 0.0744, alongside the greatest t-statistics (up to 225.034) and p-values as low as 2.34×10^{-9} , indicating extremely significant superiority. Table 7 provides statistical validation that HybridRailNet consistently and significantly

outperforms all baselines across all assessment parameters, hence affirming the reliability of the proposed dual-stream radar fusion with enhanced attention for safety-critical railway obstacle identification.

5.10 Discussion

The experimental findings continuously demonstrate that the proposed HybridRailNet attains the highest reliability in obstacle identification for railway FMCW radar, while also being appropriate for a 5G-enabled IoT safety pipeline. According to the confusion matrix results, HybridRailNet generates no false positives and merely 10 false negatives, accurately identifying 426 Clear and 464 Obstacle samples, demonstrating that the model nearly eradicates superfluous alerts while preserving robust obstacle sensitivity. The ROC analysis substantiates this advantage, as HybridRailNet attains an AUC of 1.000, surpassing MobileNetV2 (0.951), EfficientNet-B0 (0.928), and ResNet-18 (0.908), thereby confirming near-perfect distinguishability between Clear and Obstacle scenarios. The quantitative measures substantiate this trend: HybridRailNet achieves Accuracy = 0.9889, Precision = 1.0000, Recall = 0.9789, and F1 = 0.9893, while the superior baseline (MobileNetV2) exhibits lower values (0.9500 / 0.9574 / 0.9473 / 0.9523). Significantly, HybridRailNet's deployment efficiency results in enhanced connectivity performance: with a compact size of 12.8 MB and 2.8 GFLOPs, it attains the lowest edge latency (4.2 ms) and end-to-end alert latency (14.7 ms), while necessitating minimal 5G bandwidth (0.008) and producing the lowest alert rate (11.1) substantially lower than MobileNetV2 (45), EfficientNet-B0 (63), and ResNet-18 (81). Ablation investigations elucidate that the elimination of MTI precipitously decreases accuracy to 0.9123 and escalates alerts to 67.5, whilst the removal of dual-stream fusion diminishes accuracy to 0.9456 with 42.7 alerts, so affirming the critical roles of RD-MD complementarity and clutter reduction. The t-test results indicate that HybridRailNet's enhancements are statistically significant across all metrics compared to every baseline (all $p < 0.001$), affirming that the improvements are attributable to the suggested architecture and preprocessing decisions rather than random variation.

6 Conclusion

This study introduced HybridRailNet, a 5G-enabled dual-stream radar framework for obstacle identification, intended for safety-critical railway operations. Through the simultaneous acquisition of complementary representations from Range-Doppler and Micro-Doppler domains, the model effectively captures both localized target reflections and motion-related characteristics, facilitating robust differentiation between clear-track and obstacle-present conditions. The incorporation of MTI-based clutter suppression and CFAR-like adaptive masking enhances reliability by diminishing static rail interference and directing the network towards potential detection areas. The enhanced ResNet-style backbone, incorporating asymmetric convolutions and a proposed attention mechanism, augments feature selectivity, enabling the network to concentrate on significant radar patterns while mitigating background artifacts. The framework not only excels in detection performance but also facilitates real-time deployment via efficient edge inference. Additionally, 5G connectivity allows for rapid alert dissemination, centralized monitoring, event logging, and coordinated responses, all while preserving local fail-safe measures like emergency braking. The suggested method offers a viable and scalable framework for next-generation IoT railway safety systems, integrating precise radar perception, validation based on interpretability, and connectivity-aware operational design for reliable real-world implementation.

References

- [1] Y. Liang, X. Qian, and L. Zhu, "Towards better railway service: Passengers counting in railway compartment," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 2, pp. 439–451, 2020.
- [2] T. Ye, B. Wang, P. Song, and J. Li, "Automatic railway traffic object detection system using feature fusion refine neural network under shunting mode," *Sensors*, vol. 18, no. 6, p. 1916, 2018.
- [3] J. Vazquez et al., "Detection of moving objects in railway using vision," in *IEEE Intelligent Vehicles Symposium*, 2004, IEEE, 2004, pp. 872–875.
- [4] R. Kapoor, R. Goel, and A. Sharma, "An intelligent railway surveillance framework based on recognition of object and railway track using deep learning," *Multimed. Tools Appl.*, vol. 81, no. 15, pp. 21083–21109, 2022.
- [5] Z. Zhao, J. Kang, Z. Sun, T. Ye, and B. Wu, "A real-time and high-accuracy railway obstacle detection method using lightweight CNN and improved transformer," *Measurement*, vol. 238, p. 115380, 2024.
- [6] P. V Vimal and K. S. Shivaprakasha, "IoT based greenhouse environment monitoring and controlling system using Arduino platform," in *2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICT)*, IEEE, 2017, pp. 1514–1519.
- [7] M. S. Adhikari, R. Patel, M. Sindhvani, and S. Sachdeva, "Introduction to AI Techniques for 6G Application," *Dev. 6G Networks Technol.*, pp. 1–27, 2024.
- [8] M. Bakyt, L. La Spada, N. Zeeshan, K. Moldamurat, and S. Atanov, "Application of Quantum Key Distribution to Enhance Data Security in Agrotechnical Monitoring Systems Using UAVs.," *Appl. Sci.*, vol. 15, no. 5, 2025.
- [9] Y. Ahmed, M. A. Azad, and T. Asyhari, "Rapid forecasting of cyber events using machine learning-enabled feature sets," *Information*, vol. 15, no. 1, p. 36, 2024.
- [10] A. Bhattacharyya, S. M. Nambiar, R. Ojha, A. Gyaneshwar, U. Chadha, and K. Srinivasan, "Machine Learning and Deep Learning powered satellite communications: Enabling technologies, applications, open challenges, and future research directions," *Int. J. Satell. Commun. Netw.*, vol. 41, no. 6, pp. 539–588, 2023.
- [11] B. Akoramurthy and B. Surendiran, "Analysis of Techniques and Methods for Health Informatics: A Quantum Leap," in *Intelligent Computing Systems and Applications: Proceedings of the 2nd International Conference, ICICSA 2023*, Springer Nature, 2024, p. 261.
- [12] A. Jabbar, H. Jianjun, M. K. Jabbar, K. ur Rehman, and T. Mahmood, "Quantum-assisted federated learning for radar-based object tracking in IoT-enabled environments," *EPJ Quantum Technol.*, vol. 12, no. 1, p. 141, 2025.
- [13] C. De Alwis, Q.-V. Pham, and M. Liyanage, *6G Frontiers: Towards Future Wireless Systems*. John Wiley & Sons, 2022.
- [14] 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. J. Burges, L. Bottou, and K. Q. Weinberger, Eds., Curran Associates, Inc., 2012. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf
- [15] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.
- [16] H. Pan, Y. Li, H. Wang, and X. Tian, "Railway obstacle intrusion detection based on convolution neural network multitask learning," *Electronics*, vol. 11, no. 17, p.

- 2697, 2022.
- [17] W. Jiang, Y. Wang, Y. Li, Y. Lin, and W. Shen, "Radar target characterization and deep learning in radar automatic target recognition: A review," *Remote Sens.*, vol. 15, no. 15, p. 3742, 2023.
 - [18] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," 2018, [Online]. Available: <http://arxiv.org/abs/1804.02767>
 - [19] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," 2020, [Online]. Available: <http://arxiv.org/abs/2004.10934>
 - [20] W. Liu et al., "Ssd: Single shot multibox detector," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, Springer, 2016, pp. 21–37.
 - [21] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.
 - [22] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 142–158, 2015.
 - [23] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *Advances in Neural Information Processing Systems*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, Eds., Curran Associates, Inc., 2015. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2015/file/14bfa6bb14875e45bba028a21e d38046-Paper.pdf
 - [24] H. Tabassum, J. Jenita, S. Hunagund, S. Sahana, S. S. Santolli, and S. A. Rehman, "IoT- enabled Asset Monitoring and Optimization Railway System," in *2024 IEEE 3rd World Conference on Applied Intelligence and Computing (AIC)*, IEEE, 2024, pp. 1242–1248.
 - [25] X. Zhou, C. Yang, and W. Yu, "Moving object detection by detecting contiguous outliers in the low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 597–610, 2012.
 - [26] R. Manikandan, M. Balasubramanian, and S. Palanivel, "Vision based obstacle detection on railway track," *Int. J. Pure Appl. Math*, vol. 116, no. 24, pp. 567–576, 2017.
 - [27] G. Krummenacher, C. S. Ong, S. Koller, S. Kobayashi, and J. M. Buhmann, "Wheel defect detection with machine learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 4, pp. 1176– 1187, 2017.
 - [28] L. Tong, L. Zhu, Z. Yu, and B. Guo, "Railway obstacle detection using onboard forward- viewing camera," *J Transp Syst Eng Inf Technol*, vol. 12, no. 4, pp. 79–83, 2012.
 - [29] S. Mangale, R. Tambe, and M. Khambete, "OBJECT DETECTION AND TRACKING IN THERMAL VIDEO USING DIRECTED ACYCLIC GRAPH (DAG).," *ICTACT J. Image Video Process.*, vol. 8, no. 1, 2017.
 - [30] M. Krišto, M. Ivasic-Kos, and M. Pobar, "Thermal object detection in difficult weather conditions using YOLO," *IEEE access*, vol. 8, pp. 125459–125476, 2020.
 - [31] D. He, Z. Zou, Y. Chen, B. Liu, and J. Miao, "Rail transit obstacle detection based on improved CNN," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–14, 2021.
 - [32] Q. Guo, L. Liu, W. Xu, Y. Gong, X. Zhang, and W. Jing, "An improved faster R-CNN for high-speed railway dropper detection," *IEEE Access*, vol. 8, pp. 105622–105633, 2020.

- [33] D. He et al., "Improved Mask R-CNN for obstacle detection of rail transit," *Measurement*, vol. 190, p. 110728, 2022.
- [34] Y. Xu, H. Tao, and L. Hu, "Research on foreign object encroachment detection algorithm for high-speed railroad based on faster R-CNN," *J China Railw Soc*, vol. 42, no. 05, pp. 91– 98, 2020.
- [35] H. Niu, D. Feng, and T. Hou, "Research on foreign object intrusion detection in railway tracks based on MSL-YOLO," *J. Eng. Appl. Sci.*, vol. 72, no. 1, p. 136, 2025.
- [36] F. U. Rahman, M. T. Ahmed, M. M. Hasan, and N. Jahan, "Real-Time Obstacle Detection Over Railway Track using Deep Neural Networks," *Procedia Comput. Sci.*, vol. 215, pp. 289–298, 2022, doi: <https://doi.org/10.1016/j.procs.2022.12.031>.
- [37] S. Liu and J. Lin, "MRS-YOLO Railroad Transmission Line Foreign Object Detection Based on Improved YOLO11 and Channel Pruning," *arXiv Prepr. arXiv2510.10553*, 2025.
- [38] B. Song, S. Zhao, Z. Wang, J. Chen, W. Liu, and X. Liu, "LW-DETR: a lightweight transformer-based object detection algorithm for efficient railway crossing surveillance," *Complex Intell. Syst.*, vol. 11, no. 12, pp. 1–13, 2025.
- [39] Z. Qi, D. Ma, J. Xu, A. Xiang, and H. Qu, "Improved YOLOv5 based on the attention mechanism and FasterNet for foreign object detection on railway and airway tracks," in *2024 Asian Conference on Communication and Networks (ASIANComNet)*, IEEE, 2024, pp. 1–6.
- [40] Z. Dong, Q. Yang, H. Chen, H. Zhou, and D. Gao, "A lightweight transformer-based framework for real-time foreign object detection in complex railway environments," *J. Real-Time Image Process.*, vol. 23, Nov. 2025, doi: [10.1007/s11554-025-01800-0](https://doi.org/10.1007/s11554-025-01800-0).
- [41] V. Suresh, P. Janik, J. Rezmer, and Z. Leonowicz, "Forecasting solar PV output using convolutional neural networks with a sliding window algorithm," *Energies*, vol. 13, no. 3, p. 723, 2020.
- [42] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [43] J. Park, S. Woo, J.-Y. Lee, and I. S. Kweon, "Bam: Bottleneck attention module," *arXiv Prepr. arXiv1807.06514*, 2018.