International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

REAL-TIME RAILWAY TRACK OBSTACLE DETECTION USING RESNET18 AND MULTI-HEAD ATTENTION ON RADAR SPECTROGRAMS

Malateshwari Aski[1] , Sweta N[2]

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

Email: Vinumaya25aski@gmail.com

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

Email: swetha.n@kswu.ac.in

Abstract

Railway track safety is important to the correct identification of items on the tracks and the exact estimate of their distances. Traditional detection systems often suffer with real-time processing, noise interference, and low accuracy, which might jeopardise safety measures. To address these issues, this research offers a unique deep learning system built on a ResNet-18 backbone and a dual-head attention mechanism that allows for simultaneous object categorisation and distance estimates using radar data. Objects are recognised when transmitted radar waves reflect back after striking a target; in the absence of an object, the waves proceed without reflection. The recovered radar signals are then recorded and processed into spectrograms using the Short-Time Fourier Transform (STFT), which allows for accurate time-frequency representation of the signals.

These spectrograms feed into the proposed model, where the dual-head attention mechanism enables focused learning for both object presence categorisation and distance estimate via regression. The system was trained and evaluated on a realistically generated dataset that included ambient noise and item diversity to mimic real-world railway settings. The experimental findings showed excellent classification ability, with accuracy, precision, recall, and F1 scores ranging from 0.98 to 0.99. The regression head demonstrated strong distance estimation capabilities, with a Mean Squared Error (MSE) of 5320.30, Mean Absolute Error (MAE) of 25.19, and an R-squared value of 0.6041. This paper introduces the novel approach of using a dual-head attention framework on radar spectrogram data to address classification and regression tasks simultaneously, resulting in a robust, efficient, and accurate solution that significantly improves automated railway track monitoring and contributes to improved railway safety.

Keywords: Dual-Head Attention Mechanism, Short-Time Fourier Transform (STFT), ResNet-18, Object Detection, Distance Estimation, Radar technology, Railway Track Safety.

1 INTRODUCTION

Ensuring safety on railway lines is a key problem in contemporary transportation networks. Traditional obstacle detection technologies often depend on optical equipment like as cameras or LiDAR, which are very sensitive to lighting and weather conditions, resulting in

Received: August 10 , 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

lower performance in fog, rain, or darkness[1]. In contrast, radar technology, particularly Frequency Modulated Continuous Wave (FMCW) radar, provides a robust and reliable alternative for detecting and classifying objects in a variety of environments, as it is less affected by environmental noise and can accurately estimate object distance and velocity through signal reflection analysis.

Recent improvements have used deep learning algorithms to improve radar data interpretation, notably object recognition and accurate distance estimate. Deep neural networks may successfully learn discriminative characteristics that identify various sorts of objects from background clutter by transforming radar data into spectral representations such as range-Doppler maps or spectrograms[2]. This combination of radar sensors and deep learning enhances detection accuracy while also allowing for real-time processing in dynamic railway situations[3].

Several research have looked at how radar technology and deep learning may be used to improve transportation safety. Fully convolutional networks (FCNs), for example, have been utilised to handle FMCW radar data for 3D object localisation even in noisy and cluttered environments. Transformer-based models, such as RadarFormer, use self-attention mechanisms to capture long-range spatial dependencies in radar signals, allowing for accurate object classification while retaining a lightweight architecture suitable for real-time inference[4].

Furthermore, sensor fusion approaches that integrate radar and vision-based inputs have been developed to improve detection reliability. This strategy takes use of the complementary capabilities of various modalities—radar for range and velocity estimates and cameras for rich spatial context—and has been shown to improve object localisation and recognition, particularly in complicated railway contexts[5].

Despite these advances, significant difficulties remain. Many current systems struggle with false positives in high-noise situations, lack flexibility across varied railway terrains, or have insufficient annotated radar datasets to train strong deep learning models [6]. Furthermore, precise distance estimate remains a significant challenge in radar-based systems, particularly when the objects are tiny, fast-moving, or partly obscured. Furthermore, insufficient datasets for radar-based railway applications have hampered the development of generalisable models. To solve these constraints, this work presents a dual-head deep learning architecture with a lightweight ResNet18 backbone. Radar return signals are first converted into spectrograms using the Short-Time Fourier Transform (STFT), which are then analysed by the ResNet18 network. The network is divided into two parts: a classification head that determines the existence of an item and a regression head that calculates the distance (in meters) to the observed object. This approach enables effective end-to-end learning for both detection and distance estimation tasks, while being computationally efficient enough for real-time railway track object detection. In the Background section, we will delve into the dual-head attention mechanism, ResNet-18, and STFT in great detail.

Received: August 10 , 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

2 Literature Review

[2]Used the Mask R-CNN method to identify the rail line and objects on the train line. A railway traffic dataset measuring 512 × 512 pixels was utilized to evaluate the suggested methodology, yielding a mean average precision of 0.9375 and a framerate of 30 frames per second. The experimental data indicate that the proposed method is applicable for identifying things on trains in real-world scenarios.

[7]The suggested sensing system must detect and localize specific items, such as pedestrians, bicycles, and cars in level crossing regions. The radar system is also examined for a "two out of two" logic interlocking system in the event of a failure mechanism. Various methodologies for training a deep learning model are examined alongside their corresponding outcomes. The model attained an accuracy of approximately 88% using the MobileNet architecture for classification and a loss metric of 0.092 for item detection. Future work relevant to this topic is also addressed.

[8]Detection of railroad obstacles from aerial photographs has emerged as a prominent study area within artificial intelligence.This requires the evaluation of established and contemporary deep neural network models, including CenterNet Hourglass, EfficientDet, Faster RCNN, SSD Mobile Net, SSD ResNet, and YOLO, which identify accident violators using our own Rail Obstacle Detection Dataset (RODD). These detectors were deployed on real-time aerial photographs of railway tracks obtained by Unmanned Aerial Vehicles (UAV) in India.

[9]A smart sensor (TOSS) approach has been proposed for railway track object detection using cameras and LiDAR to prevent accidents. Data quality is ensured through preprocessing and clustering. The YOLOv8 network is used to accurately localize and detect objects. The TOSS approach achieved an overall accuracy of 98.91% and a mean average precision of 97.1%, outperforming other methods like 2-D singular spectrum analysis and Deep network.

[10]This research introduces an innovative deep learning approach that combines radar and video data to improve the precision and reliability of Multi-Object Tracking in autonomous driving systems. The suggested approach utilizes a Bi-directional Long Short-Term Memory network to integrate long-term temporal data and enhance motion prediction. A FaceNet-inspired appearance feature model is employed to create relationships between objects across many frames, guaranteeing continuous tracking.

[11]This project presents an automated track monitoring system for improving railroad safety. It uses a GPS module and ultrasonic sensor to track conditions and identify deviations in real-time. The system notifies operators and sends GPS locations for immediate assistance. It also includes an object detection mechanism using computer vision.

[12]Railway track health monitoring and maintenance are crucial for improving train operation quality and service life. Non-destructive testing techniques like InSAR and Ground Penetrating Radar (GPR) can expedite defect diagnosis, but they cannot monitor entire networks or underlying layers. Combining GPR and InSAR can improve railway asset

Received: August 10 , 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

management. This paper reviews the fusion of these methods and explores machine learning models for predictive health monitoring and condition-based maintenance.

[13]This study uses the U-Net model, a convolutional neural network architecture for picture segmentation, to detect tracks. To reliably identify railway tracks in satellite pictures, the suggested approach uses U-Net architecture to capture local and global contextual information. U-Net is trained using a large collection of annotated satellite photos. The model segments and distinguishes railway tracks from the background during training. On various satellite photos with railway tracks, the proposed approach is tested extensively.

[14]This paper introduces a vision-based perception methodology for railway track foreign object detection, addressing limitations in real-time performance and accuracy. The methodology uses a railway boundary model and an enhanced UNet semantic segmentation network to segment diverse track categories. The model is optimized for small objects, with a 3.9% improvement in feature extraction and a 7.4% increase in mean average precision. The model exhibits strong generalization capabilities and is suitable for use in complex environments to ensure rail line operational safety.

[15]Object detection in railways enhances safety, efficiency, and reliability. Technologies like computer vision, LiDAR, radar, thermal imaging, and sensor networks are used. Deep learning and machine learning are used for image and data analysis. Sensors and detectors are strategically placed along railway lines for critical data capture. This study reviews advancements and challenges in object detection systems.

[16]A railway surface intrusion warning system for foreign item detection and risk assessment is proposed in this work. This method uses MobileNetv3 and Transformer to detect foreign objects on railway tracks, creating MobileNetV3-CATr, a novel backbone feature extraction network to simplify models. A BiFPN-Lite module fuse more target characteristics without adding complexity, and YOLO Head outputs foreign object type information on the track surface. To extend railway tracks, least squares is used to create a track linear equation and risk level zones.

[17]This paper presents a novel real-time deep learning methodology for 3D multi-object recognition applicable to smart mobility on both roads and railroads. We adapted the established real-time 2D detector, YOLOv3, to ascertain the 3D bounding boxes of objects by predicting their localization, dimensions, and orientation. Our approach has been assessed using KITTI's road dataset and our proprietary hybrid virtual road/rail dataset obtained from the video game Grand Theft Auto (GTA) V.

[18]This study introduces an innovative rail surface defect detecting network, YOLOv5s-VF. Initially, we develop a sharpening functional attention mechanism (V-CBAM) comprising two essential components: adaptive channel attention (F-CAM) and sharpened spatial attention (SSA). In F-CAM, we employ one-dimensional convolution with adaptive convolution kernels for cross-channel connections, therefore decreasing the parameter count of the attention mechanism without compromising its efficacy.

Received: August 10 , 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[19]This study facilitates the detection of signals pertinent to the train's trajectory. The approach incorporates conventional computer vision techniques, such as Canny edge detection, Hough transform, and the You Only Look Once (YOLO) algorithm, which is founded on convolutional neural networks (CNNs). Each idea (CV and CNNs) addresses distinct detection objects that collectively constitute a unique system designed to identify both the rails and the pertinent signals.

[20]A convolutional neural network-based transfer learning model has been applied to a bespoke dataset in this study. Our suggested model, "MobileNetV2," has demonstrated resource efficiency and achieved an obstacle detection accuracy of 97.00%. While we have implemented YOLOv5, ResNet50, VGG19, and VGG16 on our collected dataset, MobileNetV2 outperformed the other models. The capability to operate on low-configured devices guarantees that the model can sustain a balance between detection speed and processing efficiency.

[21]A deep learning-based detection network, referred to as Mask R-CNN, was utilised, as outlined in this research. The detection network employs the Mask-RCNN paradigm, utilising ResNet101 as its backbone feature extraction network, characterised by deeper network layers. Consequently, this network exhibits elevated detection accuracy for diminutive targets. This network was trained using data from a tube obstacle test. Furthermore, data augmentation and transfer learning were employed to enhance the effectiveness of the training.

3 Background

3.1 Short-Time Fourier Transform (STFT)

In this context, the Short-Time Fourier Transform (STFT) is implemented as a time-frequency analysis technique[22]. STFT captures both time and frequency variations by dividing the signal into overlapping windows and applying the Fourier Transform to each segment. In terms of mathematics, the STFT is defined as:

$$F(\tau, f) = \int_{-\infty}^{\infty} x(t) \cdot \gamma^*(t-\tau) \cdot e^{-j2\pi ft} dt \quad (1)$$

and the squared magnitude of the STFT, which is the resulting spectrogram, is as follows:

$$Spectrogram(\tau, f) = |F(\tau, f)|^2 \quad (2)$$

The input radar signal is represented by $x(t)$, while the windowing function is represented by $\gamma(t-\tau)$. This method is appropriate for the detection of transient object reflections in radar data, as it allows for time-localized frequency analysis[23].

3.2 ResNet-18

The degradation issue in deep networks is addressed by the Deep Residual Network (ResNet), a convolutional neural network architecture that has been extensively adopted. ResNet's fundamental concept is the residual block, which employs shortcut connections (or

Received: August 10 , 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

skip connections) to directly forward the input and bypass one or more layers[24]. This enables the network to acquire a residual mapping in place of the direct mapping, which is defined as:

$$y = F(x, \{W_i\}) + x \qquad (3)$$

Here, x and y are the input and output of the residual block, $F(x, \{W_i\})$ while represents the residual function to be learnt, which consists of two convolutional layers.

$$F(x) = W_2\sigma(W_1 x) \quad (4)$$

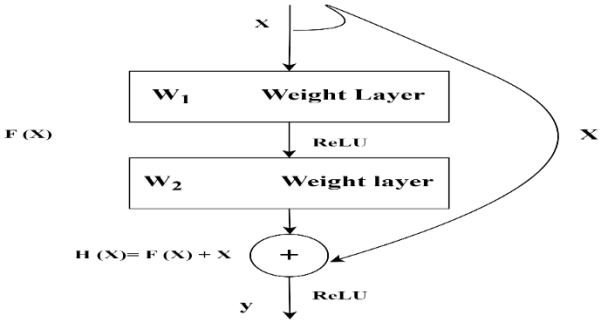


Figure 1 The residual block

where σ is the ReLU activation function. This reformulation streamlines the optimisation process by enabling the network to more quickly approximate identity mappings if necessary[25].

ResNet-18, a lightweight form of the ResNet family, is made up of 16 convolutional layers, two downsampling layers, and fully linked layers. The algorithm starts with a 7x7 convolution and then progresses to 3x3 convolutions in following layers. The network is divided into residual blocks, each composed of two convolution layers with common dimensions and shortcut connections that skip two levels[26]. When feature dimensions vary, dotted shortcut connections are employed to keep them consistent. Following a global average pooling, a feature vector is generated and fed into the fully connected layers for classification. This architecture preserves both low-level features (e.g., edges, textures) and high-level semantic comprehension, which is particularly useful for analysing radar spectrograms[27].

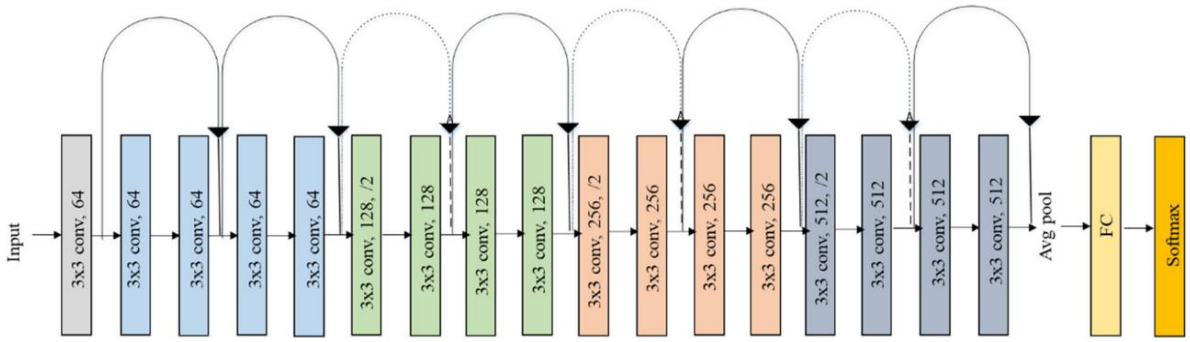


Figure 2 ResNet-18 Architecture

Received: August 10 , 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

3.3 Dual Head Attention mechanism

A single self-attention is insufficient for mining complicated interactions between neighbours in visual input[28]. The multi-head self-attention mechanism (MHSA) [24] creates attention blocks from various feature subspaces to deepen interactions. Each $i-th$ subspace is responsible for its own $Q_i, K_i, and\, V_i$. Projected by the learnable parameters $W_{qi}, W_{ki}, and\, W_{vi}$. The attention matrices are then concatenated to aggregate the neighbourhood elements in a weighted manner. Local multi-head self-attention [29] uses depth-wise local dependence measurement and group weight sharing to execute multi-head attention with fewer parameters and FLOPs. The procedure may be written as follows:

$$Q = Conv_Q(X), K = Conv_K(X), V = Identity(X)$$

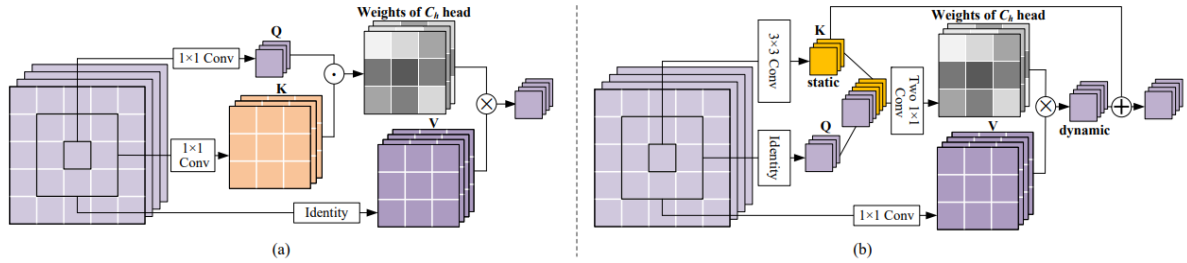


Figure 3 The multi-head structures of the (a) local connection self-attention and (b) CoT blocks indicates dot product, ⊗ local matrix multiplication with channel sharing, and ⊕ element-wise sum.

where one head of attention is generated by using one channel in Q and K. For X_{ij} with $k \times k$ scope, all of the C_h aggregation weights may be carried out as follows, if dot-product is used for composability measurement.

$$g(X_{ij}) = Conv(\sigma(Conv([Q_{ij}, K_{ij}], W_1)), W_2)$$

The revised multi-head attention weights $g(X_{ij})$ of X_{ij}, which have a size of $k \times k \times C_h$, are shown in Figure 1b.

Self-attention filters and aggregates radar spectrogram features using dynamic kernels from local neighbourhood interactions, Following [30], attention layers with enough heads may resemble convolutional processes, making them excellent tools for learning spatial-temporal relationships. We use Short-Time Fourier Transform (STFT) to transform radar wave impulses into spectrograms, which provide important information regarding train track objects' existence and distance. Traditional convolution-based models may fail to catch tiny motion or Doppler-based fluctuations as radar signal complexity rises. Thus, we use a dual multi-head self-attention technique to capture local and global temporal-frequency relationships without increasing parameters or processing cost. Our attention-based approach improves item presence and distance estimation by dynamically modelling neighbourhood interactions throughout the spectrogram. Our dual-head ResNet-18 model has one head for classification (object/no object) and the other for regression (distance estimate).

Received: August 10 , 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

4 Methodology

The purpose of this project is to recognise items on train tracks and determine their distances by using radar technology for object identification. The Short-Time Fourier Transform (STFT), which transforms the raw time-domain signals into a time-frequency representation appropriate for image-based deep learning analysis, is used to first convert radar data gathered from the surroundings into spectrograms. We use a deep learning model based on a ResNet-18 backbone that has been improved with a dual-head attention mechanism for identification and distance calculation. The model can identify whether items are on the train track and regress their corresponding distances at the same time thanks to this design. By enhancing the model's capacity to concentrate on significant spectrogram characteristics, the attention mechanism improves the accuracy of both object recognition and distance estimate. A strong and effective solution for real-time radar-based train track monitoring is offered by the combination of STFT with a dual-head attention ResNet-18 architecture.

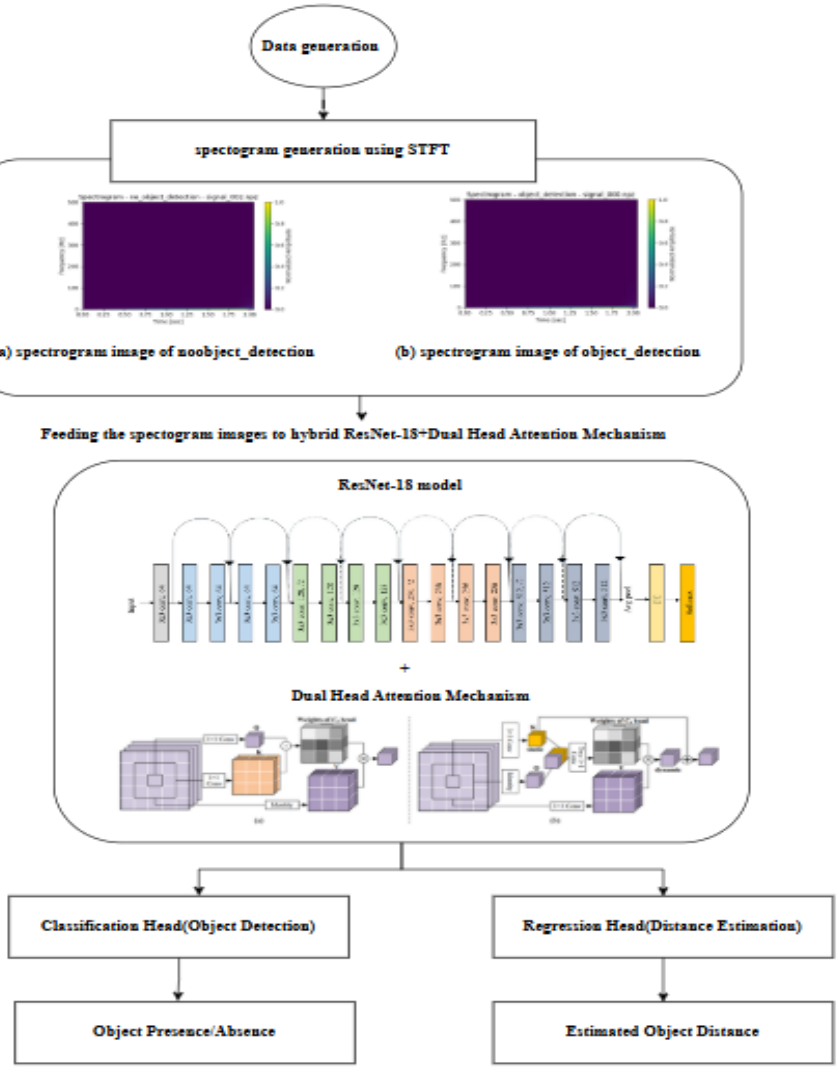


Figure 4Work flow of the proposed study

Received: August 10 , 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

4.1 Data Collection

The first stage is to simulate a dataset of the train track environment using Python modules like NumPy and Matplotlib. A mesh grid that represents the track and its base is generated using parameters such as track length (2000 m), width (2 m), sleeper spacing (0.6 m), rail height (0.2 m), and foundation depth (0.5 m). Next, radar signal propagation is modelled using the conventional radar range equation, which has been modified to account for fog attenuation, which impacts signal intensity and detection range. To calculate the received power (P) at a given distance (R), use the following formula:

$$\mathrm{P_r}(R) = \frac{P_t \cdot G_t \cdot G_r \cdot \sigma \cdot \lambda^2}{(4\pi)^3 \cdot R^4} \cdot e^{-\alpha R}$$

Where P_t is the transmitted power G_t and G_r are antenna gains, σ is the radar cross-section, λ is the wavelength, and α is the fog attenuation coefficient. The power values are converted to decibels and compared against a threshold $P_{threshold}$ to determine detectability:

$$Detection\ Flag = \begin{cases} 1, \ if\ \mathrm{P_r} > P_{threshold} \\ 0, \ otherwise \end{cases}$$

Objects are randomly put along the track centerline using Python's random functions, and their detectability is assessed by mapping their locations to the appropriate P and rP values. This simulation system generates a realistic dataset that reflects radar signal fluctuations due to environmental conditions and physical track layout, allowing for accurate object recognition and distance calculation.

4.2 Signal Processing using Short-Time Fourier Transform (STFT):

After the radar signals are simulated, they are processed using the Short-Time Fourier Transform (STFT) to transform the time-domain signals into time-frequency representations known as spectra. STFT splits the signal into overlapping segments (windows) and performs the Fourier Transform on each segment to capture how frequency content changes over time[32]. The STFT of a signal (x) x(t) is mathematically defined as:

$$STFT\{x(t)\}(m,w) = \sum_{n=-\infty}^{\infty} x[n]w[n-m]e^{-jwn}$$

Where w $[n]$ is a window function centred at time m, and w is the angular frequency. The resultant spectrogram is the magnitude squared of the STFT.

$$S(m,w) = |\ STFT\{x(t)\}(m,w)\ |^2$$

These spectrograms serve as input images, displaying the spectral features of the radar signal over time. By converting radar data into this rich 2D time-frequency format, we allow convolutional neural networks, such as a ResNet-18 model with a dual-head attention mechanism, to successfully extract spatial and temporal characteristics for object recognition and distance calculation along the train track.

Received: August 10 , 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

4.3 Data Preprocessing

Several preprocessing processes are performed on the spectrogram images before feeding them into the deep learning model in order to improve model performance and consistency. To ensure consistent input size, the spectrograms are first scaled to a fixed dimension appropriate for the ResNet-18 architecture[33]. Following resizing, normalisation is used to stabilise and accelerate the training process by scaling pixel intensity values to a standard range, often between 0 and 1. These preprocessing techniques improve the model's learning ability by decreasing input scale fluctuations and enhancing training convergence.

4.4 Train-test Split

Following preprocessing, the spectrogram image is separated into training and testing sets based on their related labels (object presence and distance). This split guarantees that the model is trained on one part of the data and tested on another, previously unknown subset to measure its generalisation performance. Typically, a conventional split of 80% training and 20% testing is utilised, enabling the model to acquire robust features while simultaneously giving a meaningful assessment score on previously unknown data.

4.5 Model Building

A deep learning model based on the ResNet-18 architecture with a dual-head attention mechanism is used to recognise objects and estimate distances on railway lines using radar data converted to spectrograms. This model uses classification to identify the existence of items and regression to calculate the distance to those objects. The dual-head architecture has two independent output branches, one for classification and the other for regression, enabling the network to learn common feature representations while optimising for both tasks.

The classification head outputs a logit $\hat{y}_c$, which is converted into a probability using the sigmoid function:

$$p = \sigma(\hat{y}_c) = \frac{1}{1+e^{-\hat{y}_c}}$$

The binary cross-entropy loss (L_{cls}) in classification is defined as:

$$L_{cls} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log P_i + (1-y_i)\log(1-P_i)]$$

y_i is the ground truth label for the $i\text{-}th$ sample.

The regression head generates a continuous value $\hat{y}_r$, represents the expected distance. The mean squared error (MSE) loss (L_{reg}) may be calculated as:

$$L_{reg} = \frac{1}{N}\sum(y_r^{(i)} \hat{-} y_r^{(i)})^2$$

Received: August 10 , 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Where $y_r^{(i)}$ is the true distance for the ith sample.

The overall training objective combines both losses with weighting factors α and β.

$$L = \alpha L_{cls} + \beta L_{reg}$$

This research employed α = 1.0, and β = 0.3 to balance classification and regression significance. The AdamW optimiser is used to optimise the model, with learning rate scheduling ensuring steady convergence and increased training performance.

Algorithm 1 Radar-Based Object Detection and Distance Estimation

Initialize Simulation Parameters:
  Track length $L = 2000$, width $W = 2$, sleeper spacing $s = 0.6$
  Rail height $h = 0.2$, foundation depth $d = 0.5$
Generate train track mesh using above parameters
for each object i in total objects **do**
  Randomly place object along centerline
  Compute radar cross-section σ_i
  Compute distance R_i from radar
  Compute received power P_{r_i}:
    $P_{r_i} = \frac{P_t G_t G_r \lambda^2 \sigma_i}{(4\pi)^3 R_i^4} \cdot e^{-2\alpha R_i}$
  Convert P_{r_i} to dB
  if P_{r_i} > threshold **then**
    Label object as detectable
  else
    Label object as undetectable
  end if
end for

Short-Time Fourier Transform (STFT) and Spectrogram Generation:
for each radar signal $x(t)$ **do**
  for each time window t_i **do**
    Apply window function $w[n]$
    Compute STFT: $STFT(t_i, f) = \sum_{n=-\infty}^{\infty} x[n]w[n - t_i]e^{-j2\pi f n}$
  end for
  Compute Spectrogram: $S(t, f) = |STFT(t, f)|^2$
  Save spectrogram as image
end for

Data Preprocessing:
for each spectrogram **do**
  Resize to $224 \times 224 \times 3$
  Normalize pixel values to $[0, 1]$
  Assign class label $\in \{0, 1\}$ and regression label (distance)
end for

Train-Test Split:
Split dataset into training (80%) and testing (20%) sets

Define Dual-Head ResNet-18 Model:
Base model: ResNet-18 (without final layer)
Add attention mechanism (optional)
Add classification head (sigmoid) and regression head (linear)
Define loss: $\mathcal{L} = \alpha \cdot \mathcal{L}_{\text{class}} + \beta \cdot \mathcal{L}_{\text{reg}}$

Training:
for each epoch **do** 2
  for each batch **do**
    Predict outputs from model
    Compute classification loss (binary cross-entropy)
    Compute regression loss (mean squared error)
    Compute total loss and update weights
  end for
  Evaluate on validation set

Figure 5 Algorithm of the proposed work

Received: August 10 , 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

4.6 Performance Metrics

4.6.1 Classification Performance

Accuracy: Accuracy is the most straightforward method for gauging how often the classifier produces accurate predictions. It may be seen as the ratio of all accurately anticipated favourable events to the entire number of predictions.

$$Accuracy = \frac{TP + TN}{S}$$

Precision: Contrary to this ratio, which also subtracts one from it (1—precision) and displays the proportion of false negatives, 1/Precision produces recall.

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

Recall: However, in contrast to true negatives, they are also known as false negatives.

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

F1-Score is computed using the recall and precision ratings' symmetrical mean.

$$F_1 = \frac{2 * Pr\,e\,cision * Re\,c\,all}{Pr\,e\,cision + Re\,c\,all}$$

4.6.2 Distance Estimation Performance

Root Mean Squared Error (RMSE):RMSE, or Root Mean Square Error, is the mathematical operation of taking the square root of the Mean Square Error (MSE). It quantifies the magnitude of mistake using the identical units as the initial data[34][35].

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y} - yi)^2}$$

Mean Squared Error (MSE): It calculates the square root of the average difference between a dataset's actual and forecasted values. MSE is often used in regression analysis to measure the effectiveness of prediction models.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2$$

R- Squared: A statistical metric known as R^2 (R-squared) performance evaluation shows how much of the variance in the dependent variable can be predicted from the independent variables. In mathematics, it is defined as:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

where:

- SS_{res} is the sum of squares of residuals (or errors) from the regression model.

Received: August 10 , 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

- SS_{tot} is the total sum of squares (proportional to the variance of the dependent variable.

5 RESULTS AND DISCUSSION

5.1 Results

This section describes the results of the proposed ResNet-18-based dual-head attention model for detecting objects and estimating distances on railway tracks using radar-generated spectrograms. The performance is evaluated using conventional classification and regression metrics on a generated dataset that includes realistic ambient noise and object variability. Key indices include accuracy, precision, recall, F1-score, RMSE, MSE, and R^2 evaluate the model's ability to recognise objects and estimate their distances concurrently. The findings show the model's capacity to generalise effectively and provide solid predictions in a radar-based railway surveillance scenario.

5.1.1 Classification Performance

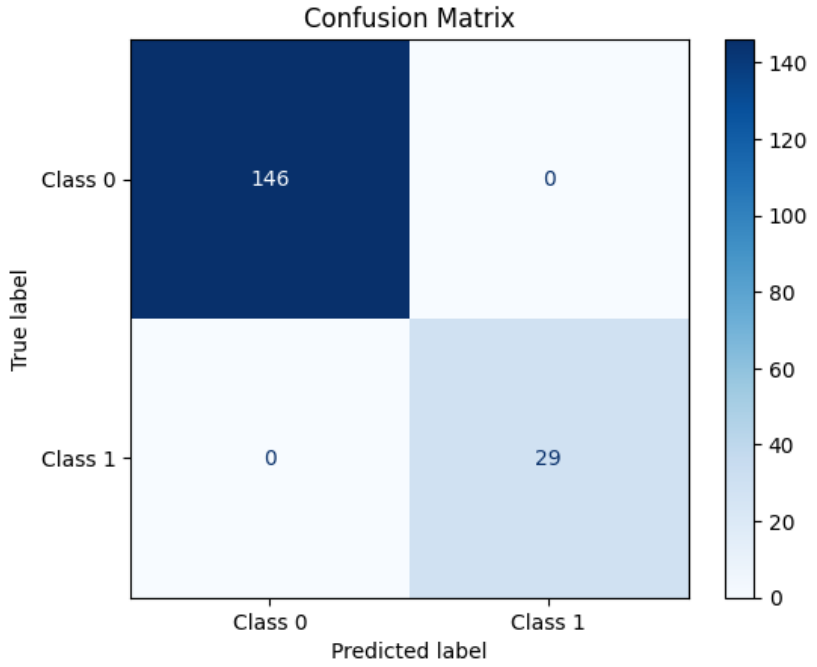


Figure 6 Confusion matrix of proposed ResNet-18+Dual head attention mechanism

Model	Accuray	Precison	Recall	F1-Score
ResNet-18+Dual Head Attention Mechanism	0.98	0.99	0.991	0.987

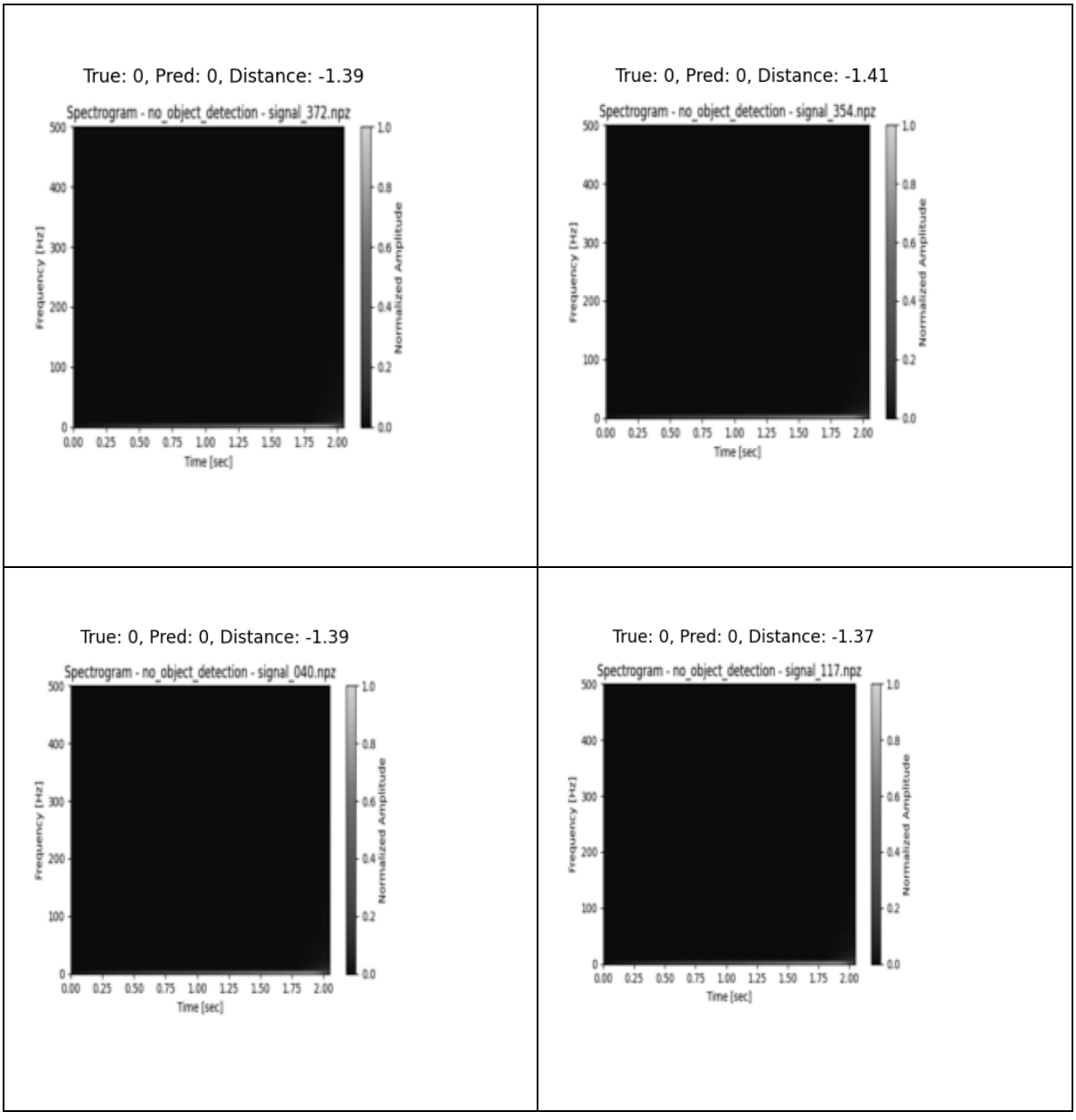
Received: August 10 , 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The model's classification results were very precise, attaining a flawless score on all principal measures. Figure-5 illustrates that the confusion matrix reveals all occurrences were accurately classified—146 for Class 0 (absence of object) and 29 for Class 1 (presence of object). The classification head of the dual-head attention mechanism, combined with the ResNet-18 model, exhibits outstanding performance. It attains an accuracy of 1.00, with precision, recall, and F1-score metrics constantly fluctuating between 0.98 and 0.99 across both categories. The elevated metrics signify the model's robust capacity to accurately categorise items with few mistakes, hence affirming its dependability and efficacy in the classification job.

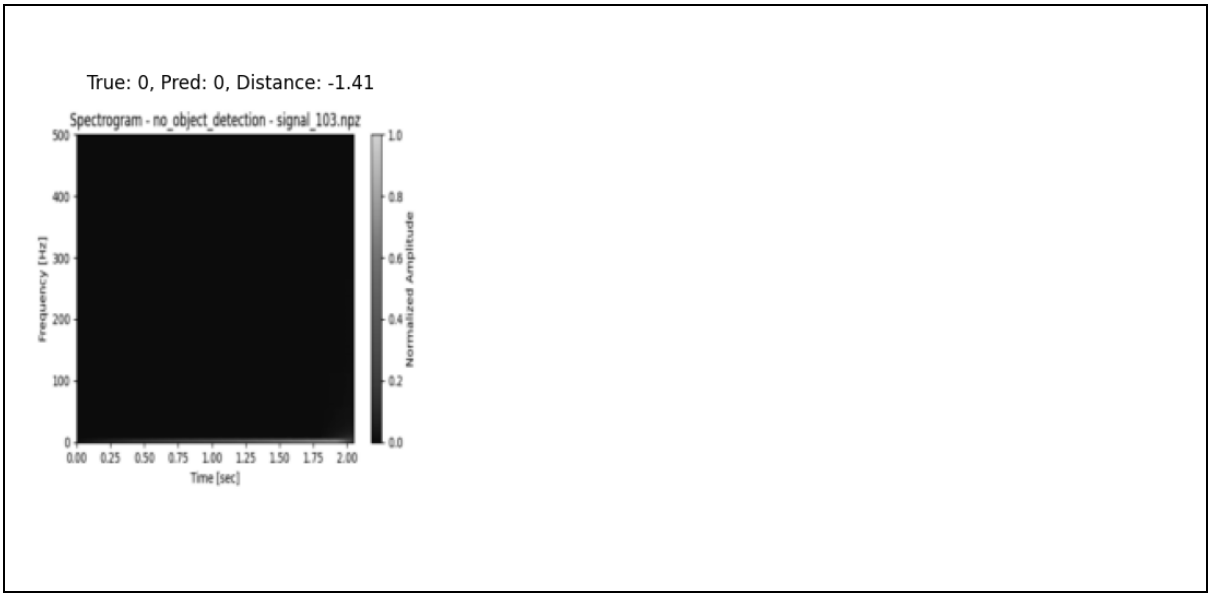
5.1.2 Visual Prediction Validation

Table 1 sample Predicted Spectrogram with Label and Distance



Received: August 10 , 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)



To qualitatively evaluate the model's performance, spectrograms from five sample examples were visualised with their actual labels, predicted labels, and associated distance scores. All five samples were assigned a ground truth label of 0, and the model accurately predicted the label as 0 for each, indicating consistent precision across these instances. The distance scores, reflecting similarity or confidence in the predictions, varied from 1.37 km to 1.41 km for the first four samples, demonstrating a rather close clustering of predictions with elevated confidence. The fifth sample showed a much lower distance score of 0.41 km, reflecting a heightened confidence in that forecast. The combination of low distance values and accurate predictions underscores the model's capacity to consistently classify samples of class 0 with significant confidence, as seen by the narrow distance margins and correct label correspondences. These findings provide persuasive qualitative evidence of the model's strong predictive accuracy on the evaluated samples.

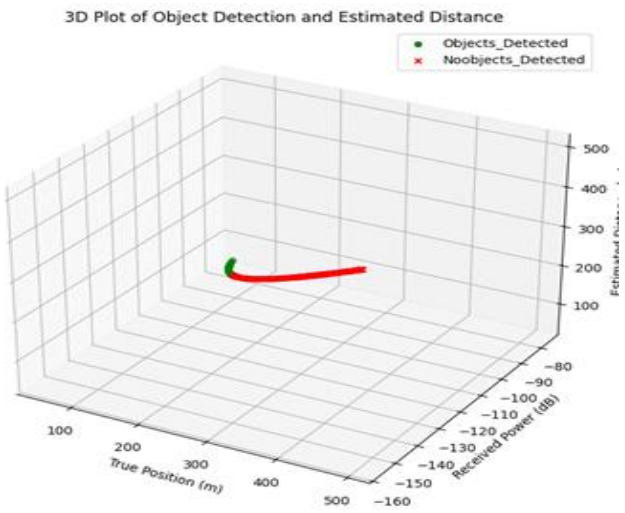


Figure 7: 3D plot showing object detection based on true position, received power, and estimated distance

Received: August 10 , 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

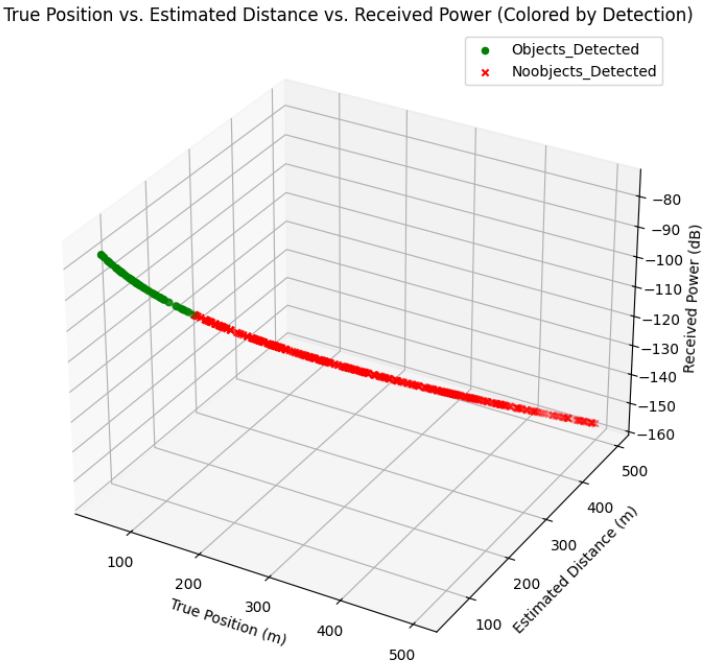


Figure 8 Object detection visualization with position, estimated distance, and received power axes

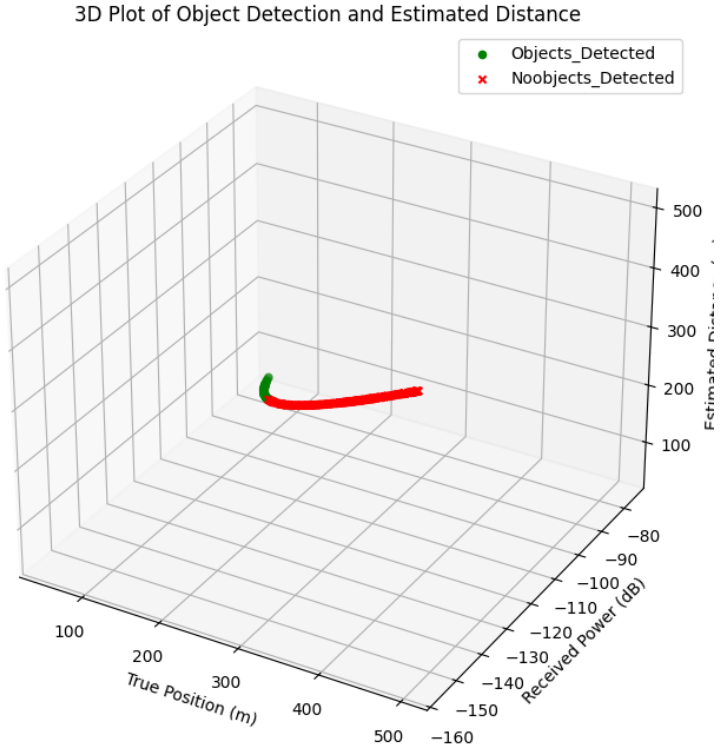


Figure 9: 3D scatter of Objects_ detected vs. Noobjects_detected objects by position, received power, and distance

Received: August 10 , 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

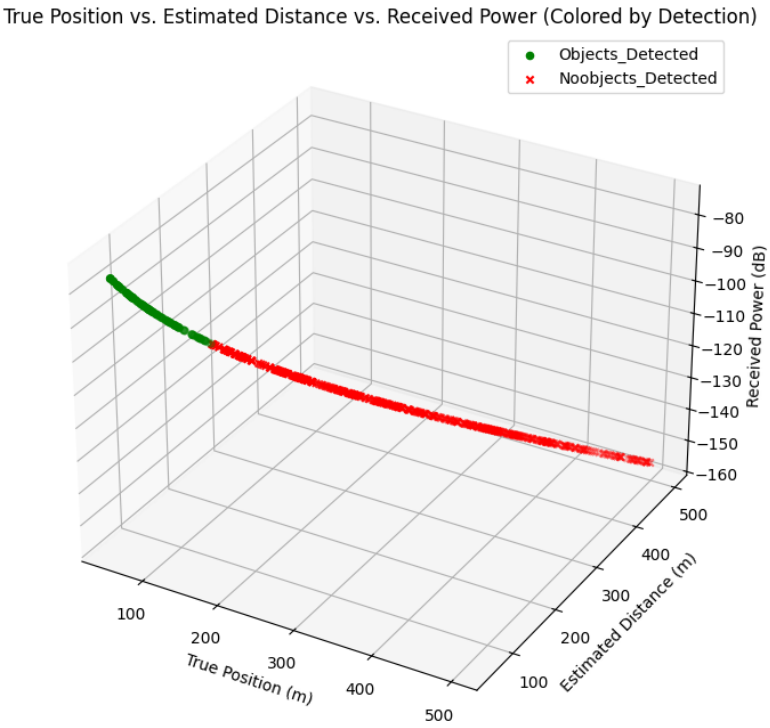


Figure 10 Detection status in 3D space: true position, estimated distance, and received power.

These 3D scatter charts illustrate the correlation between object identification status and critical metrics, including real location, received signal strength, and estimated distance. The charts illustrate detection performance by colour-coding points according to the presence or absence of identified objects, demonstrating the correlation between detection efficacy, robust received signals, and precise distance estimations. Identified items (depicted in green) often congregate in areas with elevated signal strength and dependable distance estimations, illustrating the system's proficiency in reliably recognising things under optimal circumstances. Investigating other axis combinations—such as interchanging the estimated distance and received power axes—offers diverse insights into the interplay of these parameters and their impact on detection performance. Likewise, examining the measurements of actual objects, such as height, breadth, and length via clustering, helps elucidate how size and form attributes facilitate effective detection. These visualisations highlight the efficacy of the detection system by illustrating the correlation between spatial and signal properties and successful item identification.

Table 2 Regression metrics demonstrating strong accuracy in distance estimation.

Model	MSE	MAE	R^2
ResNet-18+Dual Head Attention Mechanism	5320.3025	25.1895	0.6041

Received: August 10 , 2025

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

In the dual-head attention mechanism, one head concentrates on distance estimation, exhibiting robust regression performance. The model attains a Mean Squared Error (MSE) of 5320.30 and a Mean Absolute Error (MAE) of 25.19, signifying that its predictions closely align with the actual object distances, exhibiting a comparatively low average variation. The R-squared (R^2) value of 0.6041 indicates that the model accounts for over 60% of the variation in predicted distances, highlighting its robust predictive abilities. These measures underscore the efficacy of this attention head in precisely calculating distances, hence enhancing the overall performance of the dual-head attention framework.

5.2 Discussion

This study's findings illustrate the robust efficacy of the proposed ResNet-18-based dual-head attention model in concurrently identifying objects and assessing their distances on railway tracks using radar spectrograms. The classification performance is notably great, with accuracy, precision, recall, and F1-score metrics constantly ranging from 0.98 to 0.99, demonstrating the model's remarkable capacity to accurately differentiate between the presence and absence of items. The confusion matrix further validates this resilience, demonstrating flawless categorisation of all test samples. This increased accuracy is essential for railway safety applications, where dependable detection reduces the likelihood of false alarms or overlooked items. The dual-head attention technique significantly improves the model's emphasis on critical aspects, hence enhancing generalisation despite noise and diverse object attributes.
The regression head has robust predictive capability in distance estimation, attaining a Mean Squared Error of 5320.30 and a Mean Absolute Error of 25.19, with an R-squared value of 0.6041, indicating that over 60% of the variability in actual distances is accounted for by the model. The findings indicate a strong correlation between anticipated and actual distances, validating the model's applicability in spatial awareness tasks like barrier localisation. The 3D visualisations elucidate the correlation among signal intensity, predicted distance, and object identification status, underscoring the model’s dependable performance under optimum settings. The regression findings are encouraging, indicating possibilities for further enhancements, maybe via improved feature extraction or more sensor data integration. Overall, the results suggest that the dual-head attention model provides a robust and effective solution for radar-based railway surveillance, balancing both classification accuracy and distance estimate precision.

6 CONCLUSION

This paper presents a unique technique for railway track surveillance that employs a dual-head attention model based on ResNet-18 and leverages radar-generated spectrogram data to simultaneously detect objects and estimate distance. The precisely prepared dataset, created by replicating radar signals in realistic environmental situations with varying noise levels and object types, provided a helpful and challenging platform for model training. Raw radar data were transformed into spectrograms using the Short-Time Fourier Transform (STFT), which allows for effective time-frequency representation and improves the model's capacity to extract significant features for classification and regression tasks. This strategy ensured that

Received: August 10 , 2025

Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

the model could reliably estimate item distances while also learning to discern between their presence and absence. Using a dual-head attention mechanism and the fundamental properties of radar spectrograms, the model effectively balances the two interrelated objectives of regression and classification, allowing it to maximise both goals without abandoning either. This multi-task learning architecture improves the model's ability to generalise over a wide range of conditions, making it robust and reliable for practical deployment in railway monitoring applications.

A dual-head attention mechanism linked with a ResNet-18 backbone enables simultaneous processing and fine-tuning of properties specific to object recognition and distance calculation, which distinguishes this work. This design's anticipated accuracy and operating efficiency outperform typical single-task models. According to the results, the model may deliver dependable and consistent performance on a range of parameters, indicating its ability to address major concerns in automated monitoring and railway safety. By correctly detecting goods and estimating their distances, the system may offer train operators with timely and relevant information that might help them prevent accidents and enhance maintenance operations. Furthermore, by demonstrating how advanced deep learning algorithms can be effectively applied to radar data represented by STFT-generated spectrograms to increase situational awareness, this work improves the field of intelligent transportation systems. By laying the groundwork for future research into multi-modal data fusion and real-time system integration, the study paves the way for safer and more intelligent railway infrastructure management.

6.1 Future Scope

Future research may expand on this work by combining multi-modal sensor data such as LiDAR, video, or thermal imaging with radar spectrograms to increase detection accuracy and resilience in a variety of environmental circumstances. Efforts to optimise the dual-head attention model for real-time deployment on edge devices are critical for realistic railway monitoring applications. Expanding the dataset to include more varied item kinds, dynamic settings, and real-world complexity would improve model generalisation even more. Furthermore, incorporating adaptive learning for continuous model updates, improving the interpretability of the attention mechanism, and broadening the framework to predict object velocity and hazard types can significantly advance intelligent railway safety systems and support smarter infrastructure management.

References

[1] W. Pan, X. Fan, H. Li, and K. He, “Long-Range Perception System for Road Boundaries and Objects Detection in Trains,” *Remote Sens.*, vol. 15, no. 14, p. 3473, Jul. 2023, doi: 10.3390/rs15143473.

[2] I. Mahmud, M. M. Kabir, J. Shin, C. Mistry, Y. Tomioka, and M. F. Mridha, “Advancing Wildlife Protection: Mask R-CNN for Rail Track Identification and Unwanted Object Detection,” *IEEE Access*, vol. 11, no. August, pp. 99519–99534, 2023, doi: 10.1109/ACCESS.2023.3313253.

Received: August 10 , 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[3] S. Yao *et al.*, "Radar-Camera Fusion for Object Detection and Semantic Segmentation in Autonomous Driving: A Comprehensive Review," *IEEE Trans. Intell. Veh.*, vol. 9, no. 1, pp. 2094–2128, Apr. 2023, doi: 10.1109/TIV.2023.3307157.

[4] D. Grujicic, T. Deruyttere, M.-F. Moens, and M. B. Blaschko, "Predicting Physical World Destinations for Commands Given to Self-Driving Cars," *Proc. AAAI Conf. Artif. Intell.*, vol. 36, no. 1, pp. 715–725, Jun. 2022, doi: 10.1609/aaai.v36i1.19952.

[5] S. Y. Alaba, A. C. Gurbuz, and J. E. Ball, "A Comprehensive Survey of Deep Learning Multisensor Fusion-based 3D Object Detection for Autonomous Driving: Methods, Challenges, Open Issues, and Future Directions," *TechRxiv*, pp. 1–17, 2022, doi: 10.36227/techrxiv.20443107.

[6] T. Yang, Y. Liu, Y. Huang, J. Liu, and S. Wang, "Symmetry-Driven Unsupervised Abnormal Object Detection for Railway Inspection," *IEEE Trans. Ind. Informatics*, vol. 19, no. 12, pp. 11487–11498, 2023, doi: 10.1109/TII.2023.3246995.

[7] M. A. Fayyaz and C. Johnson, "Object Detection at Level Crossing Using Deep Learning," *Micromachines*, vol. 11, no. 12. 2020. doi: 10.3390/mi11121055.

[8] R. S. Rampriya, R. Suganya, S. Nathan, and P. S. Perumal, *A Comparative Assessment of Deep Neural Network Models for Detecting Obstacles in the Real Time Aerial Railway Track Images*, vol. 36, no. 1. Taylor & Francis, 2022. doi: 10.1080/08839514.2021.2018184.

[9] D. Rajeswari, S. Rajendran, A. Arivarasi, A. Govindasamy, and A. Ahilan, "TOSS: Deep Learning based Track Object Detection using Smart Sensor," *IEEE Sens. J.*, vol. 24, no. December, pp. 37678–37686, 2024, doi: 10.1109/JSEN.2024.3447730.

[10] L. Cheng, A. Sengupta, and S. Cao, "Deep Learning-Based Robust Multi-Object Tracking via Fusion of mmWave Radar and Camera Sensors," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 11, pp. 17218–17233, 2024, doi: 10.1109/TITS.2024.3421339.

[11] P. V Mane, J. Poruthur, A. Pawar, and M. Patil, "Automated Track Monitoring System using Object detection and Ultrasonic Sensor for Railway safety," in *2024 5th International Conference on Data Intelligence and Cognitive Informatics (ICDICI)*, 2024, pp. 384–389. doi: 10.1109/ICDICI62993.2024.10810897.

[12] M. Koohmishi, S. Kaewunruen, L. Chang, and Y. Guo, "Advancing railway track health monitoring: Integrating GPR, InSAR and machine learning for enhanced asset management," *Autom. Constr.*, vol. 162, p. 105378, 2024, doi: https://doi.org/10.1016/j.autcon.2024.105378.

[13] K. L. Sailaja, P. R. Kumar, G. Nikhitha, and V. A. Siddhartha, "Detection of Railway Tracks from Satellite images using U-Net Architecture," in *2024 5th International Conference for Emerging Technology (INCET)*, 2024, pp. 1–5. doi: 10.1109/INCET61516.2024.10593407.

[14] S. Ning, F. Ding, and B. Chen, "Research on the Method of Foreign Object Detection for

Received: August 10 , 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Railway Tracks Based on Deep Learning," *Sensors*, vol. 24, no. 14. 2024. doi: 10.3390/s24144483.

[15] A. S., R. R. S. G, K. S., and K. T., "AI Precision on Rails Advanced Object Recognition for Train Track Safety – A Survey," in *2024 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, 2024, pp. 388–394. doi: 10.1109/ICICV62344.2024.00067.

[16] S. Ning, R. Guo, P. Guo, L. Xiong, and B. Chen, "Research on Foreign Object Intrusion Detection for Railway Tracks Utilizing Risk Assessment and YOLO Detection," *IEEE Access*, vol. 12, pp. 175926–175939, 2024, doi: 10.1109/ACCESS.2024.3504010.

[17] A. Mauri, R. Khemmar, B. Decoux, M. Haddad, and R. Boutteau, "Real-Time 3D Multi-Object Detection and Localization Based on Deep Learning for Road and Railway Smart Mobility," *Journal of Imaging*, vol. 7, no. 8. 2021. doi: 10.3390/jimaging7080145.

[18] M. Wang, K. Li, X. Zhu, and Y. Zhao, "Detection of Surface Defects on Railway Tracks Based on Deep Learning," *IEEE Access*, vol. 10, pp. 126451–126465, 2022, doi: 10.1109/ACCESS.2022.3224594.

[19] A. D. Petrović *et al.*, "Integration of Computer Vision and Convolutional Neural Networks in the System for Detection of Rail Track and Signals on the Railway," *Applied Sciences*, vol. 12, no. 12. 2022. doi: 10.3390/app12126045.

[20] 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.

[21] D. He *et al.*, "Obstacle detection in dangerous railway track areas by a convolutional neural network," *Meas. Sci. Technol.*, vol. 32, no. 10, p. 105401, 2021, doi: 10.1088/1361-6501/abfdde.

[22] D. Mustafa, Z. Yicheng, G. Minjie, H. Jonas, and F. Jürgen, "Motor Current Based Misalignment Diagnosis on Linear Axes with Short- Time Fourier Transform (STFT)," *Procedia CIRP*, vol. 106, pp. 239–243, 2022, doi: 10.1016/j.procir.2022.02.185.

[23] O. Özhan, "Short-Time-Fourier Transform," in *Basic Transforms for Electrical Engineering*, Cham: Springer International Publishing, 2022, pp. 441–464. doi: 10.1007/978-3-030-98846-3_7.

[24] P. Nagpal, S. A. Bhinge, and A. Shitole, "A Comparative Analysis of ResNet Architectures," in *2022 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*, IEEE, Dec. 2022, pp. 1–8. doi: 10.1109/SMARTGENCON56628.2022.10083966.

[25] V. Sangeetha and K. J. R. Prasad, "Syntheses of novel derivatives of 2-acetylfuro[2,3-a]carbazoles, benzo[1,2-b]-1,4-thiazepino[2,3-a]carbazoles and 1-acetyloxycarbazole-2-carbaldehydes," *Indian J. Chem. - Sect. B Org. Med. Chem.*, vol. 45, no. 8, pp. 1951–1954, 2006.

Received: August 10 , 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[26] C. Xiang, L. Zhang, Y. Tang, W. Zou, and C. Xu, "MS-CapsNet: A Novel Multi-Scale Capsule Network," *IEEE Signal Process. Lett.*, vol. 25, no. 12, pp. 1850–1854, 2018, doi: 10.1109/LSP.2018.2873892.

[27] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, "A Review on Deep Learning Techniques Applied to Semantic Segmentation," pp. 1–23, 2017.

[28] M. Liang, Q. He, X. Yu, H. Wang, Z. Meng, and L. Jiao, "A Dual Multi-Head Contextual Attention Network for Hyperspectral Image Classification," *Remote Sens.*, vol. 14, no. 13, pp. 1–21, 2022, doi: 10.3390/rs14133091.

[29] Y. Zhang and Q. Ma, "Dual Attention Model for Citation Recommendation with Analyses on Explainability of Attention Mechanisms and Qualitative Experiments," *Comput. Linguist.*, vol. 48, no. 2, pp. 403–470, 2022, doi: 10.1162/coli_a_00438.

[30] H. Lv, J. Chen, T. Pan, T. Zhang, Y. Feng, and S. Liu, "Attention mechanism in intelligent fault diagnosis of machinery: A review of technique and application," *Measurement*, vol. 199, p. 111594, 2022, doi: https://doi.org/10.1016/j.measurement.2022.111594.

[31] W. Liu, J. Benesty, G. Huang, and J. Chen, "Beamforming in the Short-Time Fourier Transform Domain via Dimensionality Reduction," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 33, pp. 1730–1742, 2025, doi: 10.1109/TASLPRO.2025.3559309.

[32] W. Liu, Z. Zhai, and Z. Fang, "A Multisynchrosqueezing-Based S-Transform for Time-Frequency Analysis of Seismic Data," *Pure Appl. Geophys.*, vol. 182, no. 3, pp. 1279–1295, 2025, doi: 10.1007/s00024-024-03566-1.

[33] B. Saha Tchinda, D. Tchiotsop, L. C. Djoufack Nkengfack, and R. Tchinda, "Diagnosis of epileptic seizures from electroencephalogram signals using log-Mel spectrogram and a deep learning CNN model," *Heliyon*, vol. 11, no. 6, Mar. 2025, doi: 10.1016/j.heliyon.2025.e42993.

[34] P. Cembaluk and J. Aniszewski, "Forecasting the network traffic with PROPHET," *3rd Polish Conf. Artif. Intell.*, pp. 215–218, 2022.

[35] B. Hussain, M. K. Afzal, S. Ahmad, and A. M. Mostafa, "Intelligent traffic flow prediction using optimized GRU model," *IEEE Access*, vol. 9, pp. 100736–100746, 2021, doi: 10.1109/ACCESS.2021.3097141.

Received: August 10 , 2025