Enhanced Datasets and AI Models for Monitoring of Grade Crossings

Gasser Ali, Ph.D.
Principal Investigator
Civil Engineering Department
University of Texas Rio Grande Valley

Constantine Tarawneh, Ph.D.
Director, UTCRS, Co-Principal Investigator
Mechanical Engineering Department
University of Texas Rio Grande Valley

Silvia Flores-Osuna and Felix Chavez
Undergraduate Research Assistants
Computer Science Department
University of Texas Rio Grande Valley

Darren Espinoza
Undergraduate Research Assistant
Mechanical Engineering Department
University of Texas Rio Grande Valley

A Report on Research Sponsored by

University Transportation Center for Railway Safety (UTCRS)
University of Texas Rio Grande Valley (UTRGV)

September 2025

Technical Report Documentation Page

1. Report No. UTCRS-UTRGV-O1CY24	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Enhanced Datasets and AI Models for Monitoring of Grade Crossings		5. Report Date September 30, 2025 6. Performing Organization Code UTCRS-UTRGV
7. Author(s) Gasser Ali, Constantine Tarawneh, Silvia Flores-Osuna, Felix Chavez, and Darren Espinoza		8. Performing Organization Report No. UTCRS-UTRGV-O1CY24
9. Performing Organization Name and Address University Transportation Center for Railway Safety (UTCRS) University of Texas Rio Grande Valley (UTRGV) 1201 W. University Dr. Edinburg, TX 78539		10. Work Unit No. (TRAIS) 11. Contract or Grant No. 69A3552348340
12. Sponsoring Agency Name and Address U.S. Department of Transportation (USDOT) University Transportation Centers Program 1200 New Jersey Ave. SE Washington, DC, 20590		13. Type of Report and Period Covered Project Report June 1, 2024 – August 31, 2025 14. Sponsoring Agency Code USDOT UTC Program
15. Supplementary Notes		

16. Abstract

Safety at grade crossings is a major concern due to the number of accidents every year. Many innovative technologies have been proposed to automate the monitoring of crossings. The goal of this research is to investigate the use of Artificial Intelligence (AI) and Deep Learning (DL) to monitor grade crossings and detect various hazardous conditions such as vehicles, pedestrians, cyclists, animals, warning lights, and others. Limitations of previous work show the need to improve the size and balance of the data. The work in this proposal aims to address limitations in the current model and to make new advances by (1) increasing the number of photos in the dataset using real video streams; (2) using captures from a train simulator videogame environment; (3) addressing the issue of imbalanced dataset for training and validation; and (4) hyper-optimizing the model for accuracy and real-time performance. This research relates directly to the strategic research goal of UTCRS of reducing fatalities and injuries at highway-rail grade crossings (HRGCs); and relates to the railway operation systems research area of autonomous systems for grade crossing safety. The outcomes of this research should advance knowledge in automated monitoring of hazards at grade crossings, and result in a model that can be implemented in cameras for automated hazard monitoring at grade crossings.

17. Key Words Artificial Intelligence, Computer Vision, Railroad Grade Crossings, Safety, Hazards		18. Distribution Statement This report is available for download from https://www.utrgv.edu/railwaysafety/research/operations/index.htm	
19. Security Classification (of this report) None	20. Security Classification (of this page) None	21. No. of Pages 32	22. Price

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Artificial Intelligence (AI)

Computer Vision (CV)

Convolutional Neural Network (CNN)

Deep Learning (DL)

Federal Highway Administration (FHWA)

Disclaimer

Acknowledgements

The authors wish to acknowledge the University Transportation Center for Railway Safety (UTCRS) for funding this research under the USDOT UTC Program Grant No. 69A3552348340.

1. Introduction

Rail-Highway grade crossings pose a critical safety concern due to the risk of collisions between trains and vehicles or pedestrians. According to the US Department of Transportation (USDOT) and the Federal Railroad Administration (FRA), there were 2,246 incidents, 266 fatalities, and 744 injuries in 2024 (US Department of Transportation 2024). 96% of rail-related fatalities in the past 10 years are linked to highway-rail grade crossings and trespassing (US Department of Transportation (USDOT) Federal Railroad Administration (FRA) 2025). There is a strong need to improve safety at grade crossings to eliminate fatalities, injuries, damage to property, and disruptions to rail operations. The rail infrastructure in the US is large, with approximately 140,000 miles of track (The American Society of Civil Engineers (ASCE) 2025). Considering the vast amount of grade crossings and the wide variety of conditions and traffic at each one, ensuring safety at all crossings is a difficult endeavor requiring lots of resources. Thus, there is an ongoing need for effective solutions to monitor grade crossings and prevent accidents.

Many safety measures have already been implemented at crossings. This includes warning signs, crossbucks, gates, and active warning devices such as bells and flashing lights. Additionally, active warning systems, such as gates, bells, and flashing lights, may be automated according to incoming train traffic. While these systems are essential and may be installed with site-specific criteria depending on traffic (Zayandehroodi et al. 2025), they do not detect vehicles and pedestrians. To address this need, many solutions have been proposed over the past decades to improve railway safety using different methods of sensing and detection. These solutions range from systems monitoring traffic or pedestrians to detecting hazards at crossings, stations, or railways. Many of the proposed systems use methods such as Light Detection and Ranging (LIDAR) laser detectors (Amaral et al. 2016) and radar systems (Hari Narayanan et al. 2011). Other proposed methods involve the use of cameras for video surveillance (Salmane et al. 2013; Sheikh et al. 2004; Shin et al. 2021; Zhang et al. 2018). The cameras used may be regular cameras, stereo cameras (Hosotani et al. 2009; Yoda et al. 2006), or thermal cameras (Vivek et al. 2023).

In later research, deep learning was introduced to detect hazards in videos. Deep learning enabled significant scientific and technological breakthroughs in many fields, including railway safety in general, monitoring of grade crossings, and monitoring traffic conditions at crossings (Guo et al. 2022; LeCun et al. 2015; Oh et al. 2022a). Advancing AI, Internet-of-Things, big data, robotics, and other innovative technologies are essential to modernize the railway infrastructure as an intelligent system with active safety capabilities (Qin et al. 2023). Given that cameras are relatively inexpensive and robust, this research capitalizes on deep learning to detect hazards using video feeds from the cameras.

Overall, there is a need for a hazards monitoring systems that: (1) can be readily usable for any grade crossing, as opposed to being purpose built for a specific grade crossing or requiring setting customization to work properly; (2) can detect and classify various types of hazards and vehicles, which can be valuable for data collection towards supporting further operational analytics; (3) can function in any lighting (day or night), weather, and environment conditions; (5) is cost effective by using non-expensive equipment; and (6) is automated and requires no human intervention.

In previous research, the authors collected data for hazards at grade crossing and trained a neural network (Ali et al. 2024; Espinoza et al. 2024). However, the research was limited due to (1) the limited size of the dataset; (2) the scarcity of data for some labels; and (3) more experimentation was needed to try more advanced and established CNNs models. These limitations affected the accuracy of the model. Accordingly, this research continues the previous work performed to address limitations and explore methods to create better performing models for hazards detection such as rail crossings.

2. Goal and Objectives

The goal of this research is to investigate the application of AI to automatically monitor grade crossings and detect hazards that may result in accidents. Such hazards may include vehicles, trailers, pedestrians, bicyclists, and animals on the grade crossing. The requirements of the resulting model include robustness by being able to detect various hazards, adapt to different crossings, and being functional under different lighting and weather conditions.

The goal of the paper is achieved by following two objectives. The first objective is data collection. Developing AI models requires large datasets for training. In the case of monitoring crossings, there is a need for a large dataset of images with labels indicating different conditions and hazards at crossings. The authors create a database of multi-labelled images relating to grade crossings and associated hazards. In previous research, the authors collected images and built a dataset (Ali et al. 2024; Espinoza et al. 2024). However, it was limited and struggled with scarcity of some labels. This research adds model images to it using real data and synthetic data.

In the second objective, a deep artificial convolutional neural network is trained to detect hazards crossings. The authors experiment with an established computer vision classification model which is ResNet (He et al. 2020; Li et al. 2017; NVIDIA NGC 2023). Transfer learning is performed. Multiple model configurations are tested.

The model presented in this research is proposed as a robust and cost-effective solution that can use feeds from cameras installed at grade crossings and provide uninterrupted monitoring of safety hazards to raise warnings of hazards and record data logs for further analysis to improve safety at crossings.

3. Background

Rail grade crossings are inherently a safety concern due to the risk of collision between trains and vehicles or pedestrians. An overview of the incidents, fatalities, and injuries over the past 20 years based on data from the USDOT (US Department of Transportation 2024) is shown in Figure 1. Many incidents are reported every year and result in injuries and fatalities. There is a need to reduce the number of accidents by exploring processes and innovative technologies to improve safety at crossings.

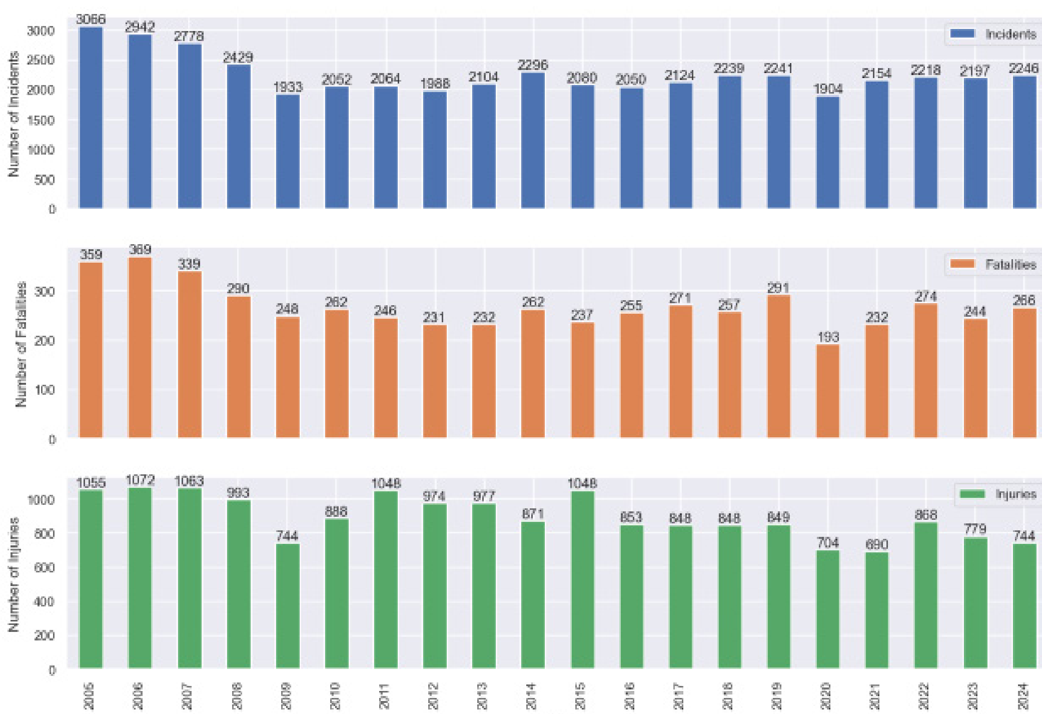

Figure 1. Number of incidents, fatalities, and injuries at highway-rail grade crossing over the past 20 years.

Many grade crossings have already implemented safety measures to help prevent accidents. These measures include passive measures such as warning signs, crossing gates, flashing lights, and, in some locations, gatekeepers. However, a major flaw in these measures is their inability to account for human error. For instance, many impatient drivers ignore crossing gates and, in some cases, even drive into them to avoid waiting, which contributes to the many fatalities and casualties at level crossings (Starčević et al. 2016). In response, more active measures are being introduced to automatically detect hazards, secure crossings, and increase awareness.

Given the number of unforeseen accidents with traditional methods, enhancing safety more actively at grade crossings is a major goal for rail operators and agencies. Many innovative methods and technologies have been proposed to improve safety at rail crossings. Previous research in this domain is summarized in Table 1. Proposed technologies involve the use of hardware such as depth or stereo cameras (Hosotani et al. 2009; Yoda et al. 2006), thermal cameras (Vivek et al. 2023), or in some instances, a combination of radars and cameras (Wang et al. 2024). Over the previous years with advances in computing, a combination of computer vision and deep learning has been used to improve the accuracy of detecting hazards and help with the current limitations of traditional methods. Computer vision, deep learning techniques, and a combination of both have gained traction in grade crossing safety research because of their increased accuracy and low cost, when compared to other technologies (Oh et al. 2022b). While traditional computer vision methods have been used to monitor movement through CCTV footage (Sheikh et al. 2004; Shin et al. 2021; Zhang et al. 2018), they often require manual fine-tuning and may struggle to generalize across different environments. In contrast, deep learning techniques, particularly convolutional neural networks, have recently gained traction because they can automatically learn complex patterns from large volumes of training data with minimal manual intervention (O'Mahony et al. 2020). Deep learning has led to many breakthroughs in speech recognition, natural language processing, and visual object recognition and detection (LeCun et al. 2015). Similarly, the railway industry has embraced deep learning. Because both traditional computer vision methods and deep learning have their strengths, recent research in railway safety has applied a combination of both in a plethora of scenarios. This includes monitoring of fall, slip, and trip incidents at stations, the automation of train stops operations, detection of incoming trains at level crossings, and the monitoring of traffic conditions near railway crossings (Alawad et al. 2020; Etxeberria-Garcia et al. 2020; Guo et al. 2022; Murshed et al. 2022). Convolutional Neural

Networks, a subset of deep learning, have also been used to detect defects in rail surfaces, monitor train vibrations, detect trespassing, and address several other railway safety challenges (Oh et al. 2022a). While previous research has demonstrated success in detecting several types of hazards at rail grade crossings, these methods often face limitations related to cost, scalability, and limited environmental adaptability. In addition, many methods are associated with high initial costs to be implemented. While sensors are individually inexpensive, the cost of deploying and maintaining them across an entire rail network can be substantial. Moreover, sensors often struggle to detect hazards at greater distances. To address these issues, some researchers have proposed the use of pilot vehicles placed in front of trains to identify hazards. However, to keep this cost-effective, this solution is only being implemented on trains traveling to dangerous areas, leaving other trains vulnerable (Wang et al. 2024a). Some methods are also designed to detect only one type of hazard which is either pedestrians, vehicles, or incoming trains. This lack of versatility can be problematic in real-world settings, where multiple types of hazards may occur. Furthermore, many of these systems do not offer reliable monitoring under varying environmental conditions, adding to their lack of generalizability in real-world applications. To summarize, although many methods and technologies have been proposed in literature, there are some limitations that hinder their robustness and versatility. Deep learning is a promising method to overcome previous limitations. Accordingly, this research investigates the use of deep learning and computer vision for enhancing safety at grade crossings.

Table 1. Overview of previous research.

Reference	**Summary**
(Sheikh et al. 2004)	Detection of moving objects in “danger zone” at grade crossings using video cameras and computer vision.
(Yoda et al. 2006)	Detecting pedestrians on grade crossings using a multi-point stereo camera system at grade crossing corners.
(Hosotani et al. 2009)	Detecting pedestrians using a system of two stereo cameras at corners of crossings.
(Hari Narayanan et al. 2011)	Detection of railway obstructions using MIMO radars to mimic phased array radars in a cost-effective manner.
(Salmane et al. 2013)	Detection and tracking of objects using a video-based system to calculate the level of risk in hazardous situations at level crossings.

(LeCun et al. 2015)	Discusses deep learning, how it works, how it has improved recognition systems, and the future of deep learning.
(Amaral et al. 2016)	Detection of obstacles at level crossings using 2D laser scanners to scan objects.
(Zhang et al. 2018)	Detection of near misses at railway-grade crossings using video surveillance and computer vision.
(Alawad et al. 2020)	Detection of fall, trip, and slip events at railway stations using a CNN-based computer vision framework for real-time risk management.
(Etxeberria-Garcia et al. 2020)	Explores the applicability of existing deep learning and visual odometry techniques in the railway domain. An autonomous train stop use case is suggested.
(Shin et al. 2021)	Detection of maintenance signs and maintenance workers on railways to enhance safety using existing tunnel-monitoring systems on trains and computer vision algorithms.
(Guo et al. 2022)	Assessment of traffic congestion conditions at railway grade crossings using computer-vision based object detection.
(Oh et al. 2022a)	Examines AI applications for railway safety, specifically deep learning approaches
(Murshed et al. 2022)	Automating level crossings by detecting incoming trains using computer vision and Raspberry Pi microcontrollers.
(Vivek et al. 2023)	Uses thermal imaging along with deep learning to improve detection of obstacles on railway tracks under various weather conditions
(Greitans 2023)	Presents a cost-effective obstacle-detection approach for level crossings using a combination of a radar and a camera that can perform reliably even in poor weather conditions.
(Wang et al. 2024)	Proposes using a small, self-driving vehicle equipped with a camera and sensors in front of trains traveling through dangerous areas to help detect anomalies and prevent accidents.

4. Methodology

This research capitalizes on deep learning methods to create a model that can detect hazards at crossings. This approach can be described in two major parts as shown in Figure 2: dataset creation: where data is collected, cleaned, and labelled; and model developed where several network designs are selected, trained, validated, and compared. The following subsection explains each part of the process in detail.

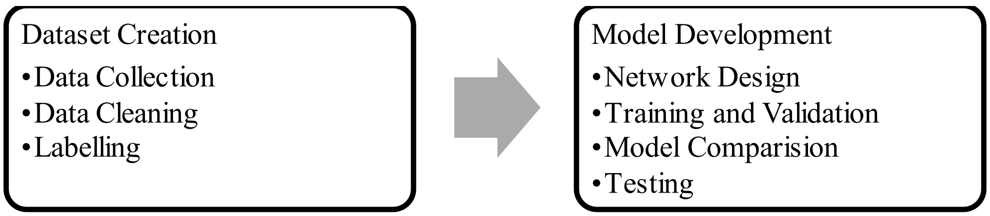

Figure 2. Overview of the methodology.

4.1. Data Collection

A large dataset of images was needed to train the model. In general, having more data leads to better model training and evaluation. The dataset is required to represent diverse situations that count happen at grade crossings. In addition, the images need to represent a variety of crossings with different scenarios, and lighting and weather conditions. The process of data collection started by downloading videos that are publicly available online such as a YouTube channel called "*Virtual RailFan, Inc.*" (Virtual Railfan, Inc. 2023). Live streams were downloaded into video files with a resolution of 640 × 360 pixels. Considering that the videos have an unneeded large number of frames per second, the number of images collected was reduced by selecting one frame out of five. A Python package "*ytb-dlp*" was used to enable video downloading and conversion. An example of two frames is shown in Figure 3.

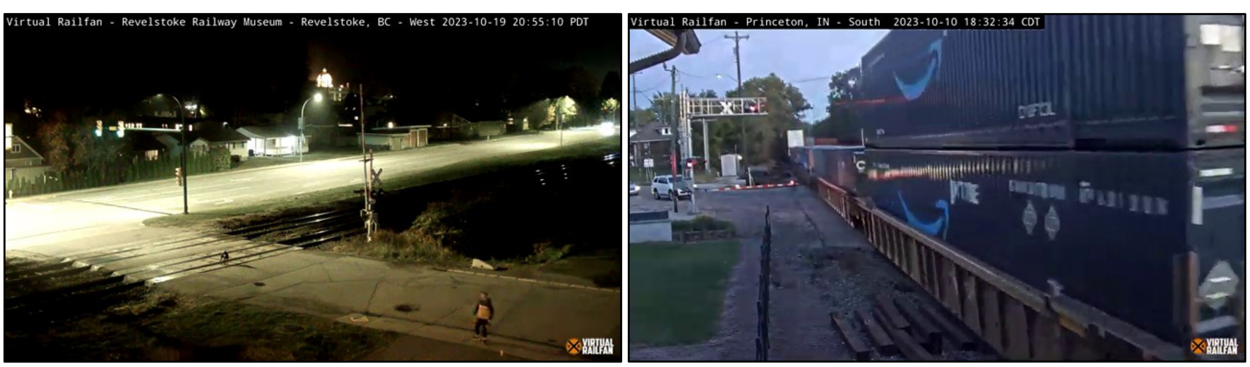

Figure 3. Examples of frames extracted.

4.2. Data Cleaning

The frames collected in the previous step included many images that are repetitive and would create additional benefits to keep. For instance, there were many images where the grade crossings are empty. There was a need to remove duplicate images if they share a high degree of centrality. This was achieved by using an existing package called "*Image Duplicator (Imagededup)*" (Jain et al. 2019). The method relies on encoding images using a CNNS, namely MobileNetV3 (Howard et al. 2019). The encoded images are compared to finding duplicates based on a cosine similarity threshold between the encodings. In this research, a similarity threshold of 95% was selected. This approach was tested and found to be effective in removing large numbers of duplicates.

4.3. Synthetic Data

The data collected from camera feeds have some limitations due to the lack of images from some angles. In addition, there are some simultaneous conditions that do not occur often in real images. In general, collecting additional images with various situations and environments improves the quality of the dataset and ultimately the performance of a trained model. Additional images were collected as screenshots from a video game called train simulator. Several scenes and vehicles were simulated. The collected screenshots were added to the dataset from real sources, which serves to increase the variety of scenes and situations in the dataset. The images are highly realistic, offer various angles, and are intended to improve the performance of the model. Some examples are shown in Figure 4.

Figure 4. Screenshots from train simulator video game.

4.4. Data Labelling

After collecting the dataset, the images were manually labelled according to the objects in each image. The authors selected 13 labels which are: Rail Track, Train, Grade Crossings, Grade Crossing Gate Down, Red Light on Grade Crossing, Train on Grade Crossing, Vehicle waiting on Grade Crossing, Trailer, Vehicle on Grade Crossing, People on Grade Crossing, Animals, Bicyclists, and Animal on Grade Crossing. The labels are nonexclusive, meaning that each image may have multiple labels, which creates a multi-label classification approach. The labels were manually assigned to each image. An open source software, Label Studio, was used to conveniently streamline the process of labeling the images (Label Studio 2023). The software is used to create and manage a database of labels to be used in the model development phase.

4.5. Model Selection: ResNet

The authors selected an established and widely used CNN design, ResNet, short for "Residual Networks", as the base of the model in this research. It was introduced in 2015 by He et al. (He et al. 2015, 2016). Traditional deep learning networks in early research suffered from increased training difficulties as the number of hidden layers is increase. In the ResNet design, an innovative approach was proposed by using repeating building blocks with residual shortcut connections, as shown in Figure 5 (Zhang et al. 2023). The addition of residual shortcuts reduces the training difficulty and results in better performance and generalization (He et al. 2020; Li et al. 2017).

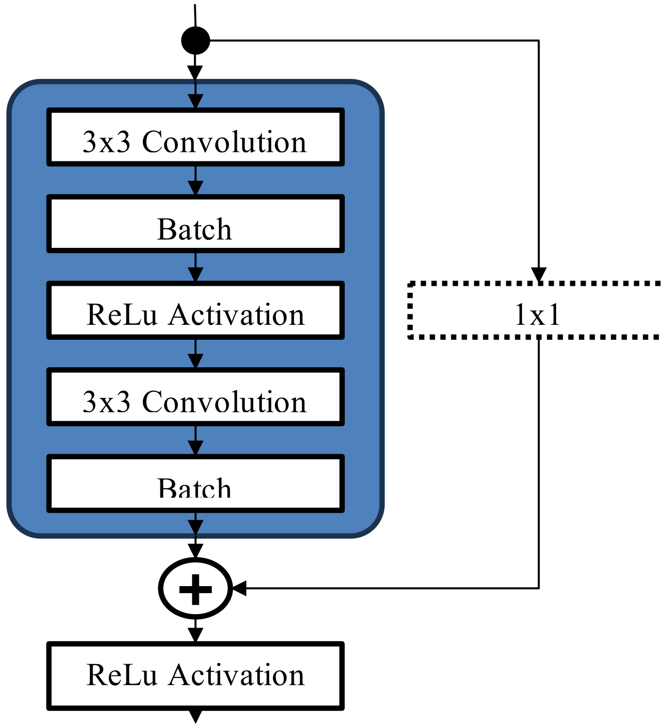

Figure 5. Building block in ResNet.

ResNet has been successfully used in many applications such as in medical image processing (Xu et al. 2023), facial expression recognition (Li and Lima 2021), pose estimating to reduce accidents involving field workers (Lee et al. 2023), improving construction site safety (Lee and Lee 2023), and many others. The model implemented in this research is based on ResNet V1.5 and implemented in PyTorch (Imambi et al. 2021; NVIDIA NGC 2023). There are several versions of ResNet based on the number of building blocks and accordingly the number of parameter layers. The versions are ResNet 18, 34, 50, 101, and 152, which are all tested in this research and compared.

4.6. Transfer Learning / Fine Tuning

Transfer learning describes that a model is trained for a task and then utilized for a different activity with some relevance to the former (Gupta et al. 2022). In other words, a model is trained to perform image classification as an example of a task, then the same model is later used to perform a different images classification task by modifying and retraining the model. Transfer learning has been successfully applied in many applications in the deep learning field, including for computer vision and others. For example, regarding ResNet, transfer learning has been successfully used with ResNet to achieve satisfactory model performance in classifying medical x-ray images (Showkat and Qureshi 2022). Transfer learning offers many benefits in reusing pre-trained models by requiring less training time and data while offering good performance and generalization.

Accordingly, pre-trained ResNet models are used in this research by modifying and training them for the purposes of this research. The weights for ResNet model used in this research were pre-trained using the ImageNet dataset (Deng et al. 2009). ImageNet is a large dataset with more than 14 million images. Previous researchers have trained the ResNet model with each set of layer configurations using the ImageNet dataset considering 1,000 outputs. The accuracies are shown in Table 2 (PyTorch 2025a) indicating high accuracy.

Since the models are trained with 1,000 classes, while the dataset in this research describes 13 labels, the pre-trained networks are modified to accommodate that output requirement. The base ResNet models are designed such that the last layer is a fully connected layer with 1,000 neurons. Accordingly, the last layer in the base models was modified to have 13 neurons which is the number of labels in this research, and the models were further trained to minimize the losses.

Table 2. Accuracies of pre-trained models. (PyTorch 2025a)

Model	**Dataset**	**Top 1 Accuracy**	**Top 5 Accuracy**	**Number of Parameters**
ResNet18	ImageNet 1k V1	69.76%	89.08%	11.7M
ResNet34	ImageNet 1k V1	73.31%	91.42%	21.8M
ResNet50	ImageNet 1k V1	76.13%	92.86%	25.6M
ResNet50	ImageNet 1k V2	80.86%	95.43%	25.6M
ResNet101	ImageNet 1k V1	77.37%	93.55%	44.5M
ResNet101	ImageNet 1k V2	81.89%	95.78%	44.5M
ResNet152	ImageNet 1k V1	78.31%	94.05%	60.2M
ResNet152	ImageNet 1k V2	82.28%	96.00%	60.2M

4.7. Data Preprocessing and Augmentation

Before data augmentation, all images were resized to have a height and width of 250 pixels and cropped to remove the borders into a height and width 200 pixels. This step reduced the size of the images to avoid the unneeded computational burden of processing high-definition images and is applied to the entire dataset once. During the training and testing, each batch processed by the neural network model undergoes further data preprocessing by normalizing the batch with a mean of 0.5 and standard deviation of 0.25.

One of the common problems in deep learning projects is the lack of data which may ultimately limit the performance of the trained models. Data augmentation is widely used to increase the amount of training data. Traditional data augmentation methods involve performing image transformation and color modifications. Transformations may include random translation, rotation, zoom, shear, etc. This approach is relatively easy to implement is proven to be successful in increasing training datasets (Chlap et al. 2021; Mikołajczyk and Grochowski 2018; Perez and Wang 2017).

In this research, traditional data augmentation methods are applied to the training dataset to artificially increase the number of training images and therefore improve the model. The data augmentation transformations are randomly applied to the images repeatedly for each new batch during the training process. The authors implement an established augmentation procedure, AutoAugment, which was proposed by Cubuk et al. (2019a; b). The method executes a policy of random image transforms that was optimized in previous research to maximize training performance.

4.8. Training Parameters

Different sizes of the ResNet were trained and compared, which include ResNet18, 34, 50, 101, and 152. The networks were trained with an early-stopping criteria such that the training is stopped when there is no improvement in the validation score for 10 epochs. Each model was trained separately until its stopping criteria was triggered. The batch size is set to include 16 images per batch. The networks were trained using the Adam optimizer, which a well-established stochastic optimization method for neural networks (Kingma and Ba 2017). The loss function used is binary cross entropy (BCE), as shown in Equations (1) and (2), which is applicable for multilabel classification problems. The loss uses logits from the neural network and internally applies sigmoid activation as shown in Equation (3) (PyTorch 2025b). A positive weight (p) is assigned as the ratio between negative and positive values for each label separately, to address the imbalance between positive and negative values.

$$loss = mean(l_n) \tag{1}$$

$$\mathrm{l_n} = -\big(p\, y_n \log \sigma(x) + (1 - y_n) \cdot \log\big(1 - \sigma(x_n)\big)\big) \tag{2}$$

$$\log \sigma(x) = \log\left(\frac{1}{1 + e^{-x}}\right) \tag{3}$$

4.9. Evaluation Metrics

Various metrics are used to evaluate the model for its classification performance. Due to the high imbalance between positives and negatives in most labels in the dataset, special emphasis is given to the balanced accuracy metric as it can tackle this imbalance. The following explains the calculations of the metrics. Actual Positive (AP) and Actual Negative (AN) refer to the values assigned in the labelling process, or the ground truth. Predicted Positive (PP) and Predicted Negatives (PN) are the predicted outputs from the neural networks. By comparing the predicted outcomes of the model with the ground truth, the outcomes of the models are assigned as True Positive (TP), True Negative (TN), False Positive (FP), or False Negative (FN). Several metrics can be calculated using TP, TN, FP, and FN. Accuracy is calculated as shown in Equation (4). The accuracy results indicate the ratio of TP and TN the model can achieve. However, accuracy may be misleading with unbalanced datasets such as in this dataset. True Positive Rate (TPR), True Negative Rate (TNR) which is also called the specificity, False Positive Rate (FPR), and False Negative (FNR) are calculated as shown in equations (5), (6), (7), and (8). It is important to ensure that the FNR metric of the model, the type II error, is low. FN implies that the model indicates that there are no hazards on the grade crossings while there are hazards on the grade crossings, which must be avoided. Balanced accuracy is calculated as shown in Equation (9). Balanced accuracy is an important metric to note to the unbalanced nature of the dataset. Finally, the Predicted Positive Rate (PPR) which is also called the precision, is calculated as shown in Equation (10) and the F_1 score is calculated as shown in Equation (11). The F_1 score is the harmonic mean of precision and recall. All the metrics are calculated for each model alternative and will be shown in the results section.

$$Accuracy = \frac{TP + TN}{AP + AN} = \frac{TP + TN}{Total} \tag{4}$$

$$TPR\ (Recall\ or\ Sensitivity) = \frac{TP}{AP} \tag{5}$$

$$TNR\ (Specificity) = \frac{TN}{AN} \tag{6}$$

$$FPR\ (Type\ I\ Error) = \frac{FP}{AN} = 1 - TNR \tag{7}$$

$$FNR\ (Type\ II\ Error) = \frac{FN}{AP} = 1 - TPR \tag{8}$$

$$Balanced\ Accuracy = \frac{TPR + TNR}{2} \tag{9}$$

$$PPV\ (Precision) = \frac{TP}{TP + FP} = \frac{TP}{PP} \tag{10}$$

$$F_1 Score = 2\ \frac{PPV \times TPR}{PPV + TPR} = 2\ \frac{Precision \times Recall}{Precision + Recall} \tag{11}$$

5. Results and Analysis

In this section, the results present the outcomes of data collection and processing, training and comparison of the models. The section is presented in the following sub-section (1) Dataset, (2) Training, (3) Evaluation; and (4) Model Comparison.

5.1. Dataset

Images were collected from video streams of grade crossings. Duplicate images were removed using deep learning as previously described in the methodology section. The images were labelled manually. Ultimately, the dataset contained 4,947 labelled images, which include 4,699 real images and an additional 248 synthetic images added to improve increase the number of images with infrequent labels. The dataset contains many images with a variety of objects and scenes to improve the success of the training process. The distribution of the labels is shown in Figure 6. Images can have more than one label, such that the model can perform multi-labelling. The highest number of labels is associated with “Rail Track” and “Train”. Many images with no rails or grade crossings were included in the data set to increase the performance of the model. In addition, it is noted that some labels are relatively infrequent compared to others, such as having animals or bicyclists on grade crossings, compared to having vehicles on grade crossings. Such events are infrequent, and it is tedious to find enough images of those events. Still, they are included in the dataset to investigate how the model will be able to handle those events. Overall, the most critical labels such as vehicles or trains on grade crossings, and related to the grade crossing gate and lights, have a suitable number of images with a relatively good balance between positives and negatives.

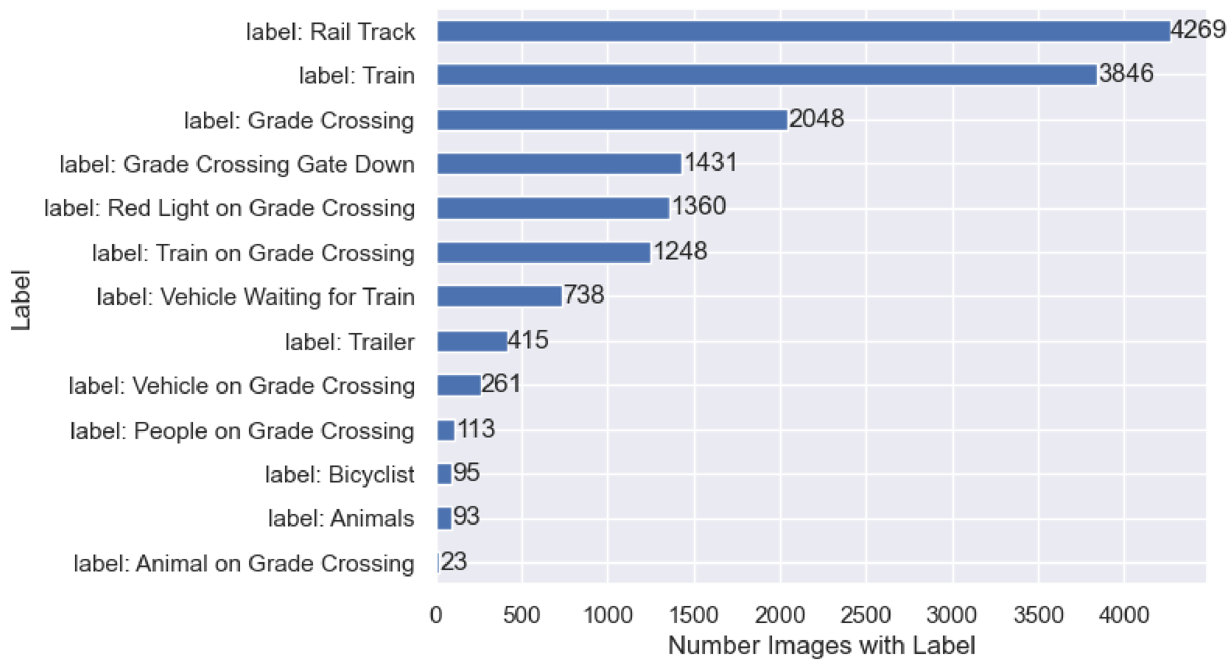

Figure 6. Count of labels.

An adjacent matrix of the labels is shown in Figure 7 and quantifies the number of images associated with the label combinations. The diagonal represents the number of images for each label, which matches the numbers in Figure 6. The non-diagonal represents the number of images associated with at least the two labels matching the x-axis and the y-axis. It is again noted that there are some label combinations that have a high number of associated images, which includes the combinations of images associated with the following labels: grade crossing, rail track, grade crossing gate down, train, red light on grade crossing, and vehicle on grade crossing. Other labels, which as animals, trailers, animals on grade crossing, people on grade crossing, and bicyclists, have relatively limited combinations. Some combinations, such as having animals and vehicles on grade crossings at the same time, are non-existent. Such situations are highly infrequent relative to others. This limitation may influence the performance of the model and will be evaluated later in the results.

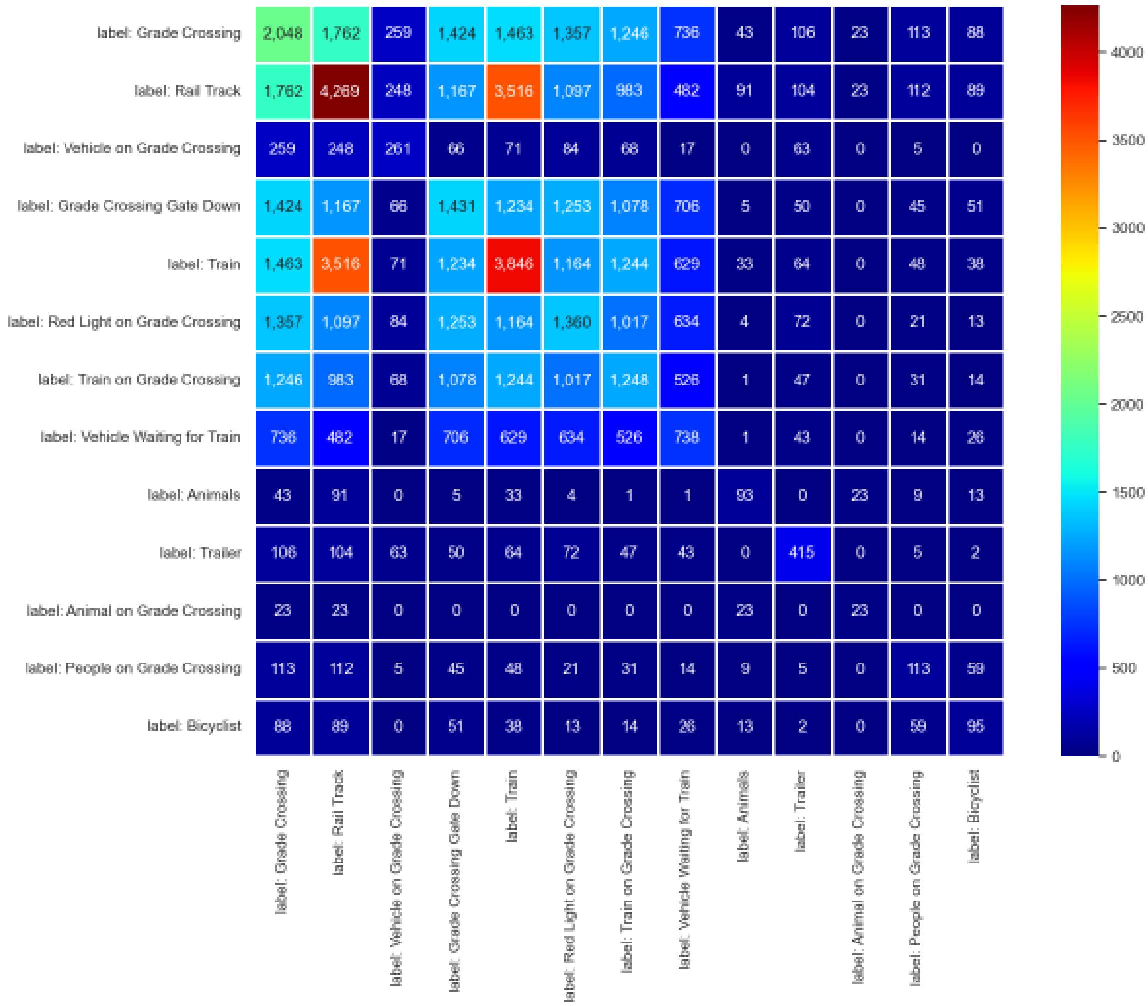

Figure 7. Adjacency matrix represents the label relationships.

5.2. Preprocessing

The labelled images were further preprocessed. As mentioned in the methodology section, all images were resized and cropped to have a height and width of 200 pixels. This step was applied to the entire dataset. During the training process, data augmentation steps were performed to improve the performance of the model. The augmentation was performed using AutoAugment (Cubuk et al. 2019a; b). Three examples of images are shown in in Figure 8. The image on the right demonstrates a clear example of random rotation, which is one of the many random images processing methods applied to the images. The images were then used to train and test the models.

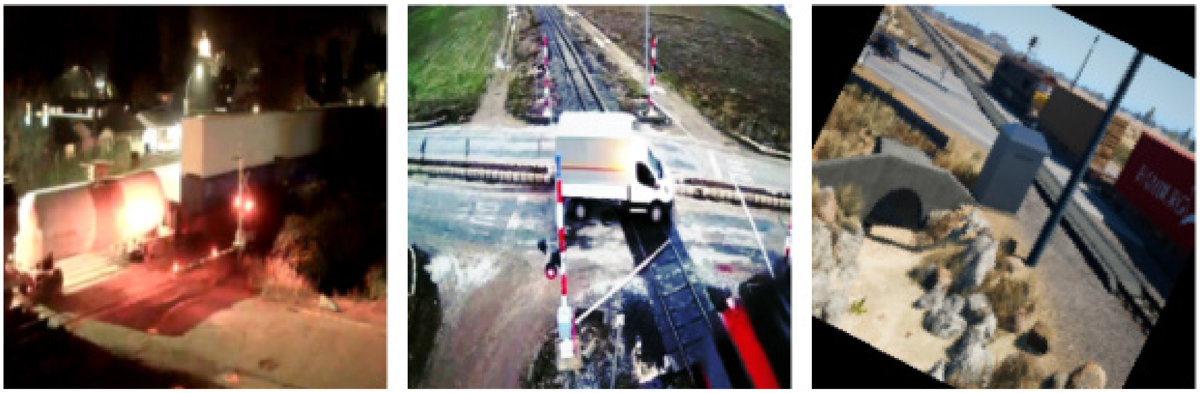

Figure 8. Example of Inputs into the Model.

5.3. Training/Validation/Testing

The data was split into training, validation, and testing with 70%, 15%, and 15% respectively. Ten model alternatives were evaluated, which include transfer learning versions of the ResNet 18, 34, 50, 101, and 152 by training all the layers and training the last layer only. All models were trained separately with early stopping criteria triggered after no improvement in testing score for ten epochs. The training, validation, and testing losses history is shown in Figure 9. The trends in the losses show that the models were trained successfully and reached their stopping criteria. Table 3 shows that epoch number and losses at the epoch with the lowers testing loss. The lowest training loss is associated with ResNet152. However, the same model is also associated with comparatively high validation loss in comparison with its train loss which may indicate overfitting. This potential issue may be associated with wrong predictions in labels with high weights, which is investigated further in the following subsection. The second-best model considering training loss is the ReNet50 model. It has a low training loss compared to other models except ResNet152, and a comparatively acceptable validation loss compared to other models. However, still, the training loss is very high compared to ResNet152. Analysis of the results based on the loss values alone is not conclusive, especially considering that the evaluation is based on a multi-labelling problem. Accordingly, more analysis is performed in the following subsection to explore the multi-labeled performance on the models to evaluate all the models and select the best performing model.

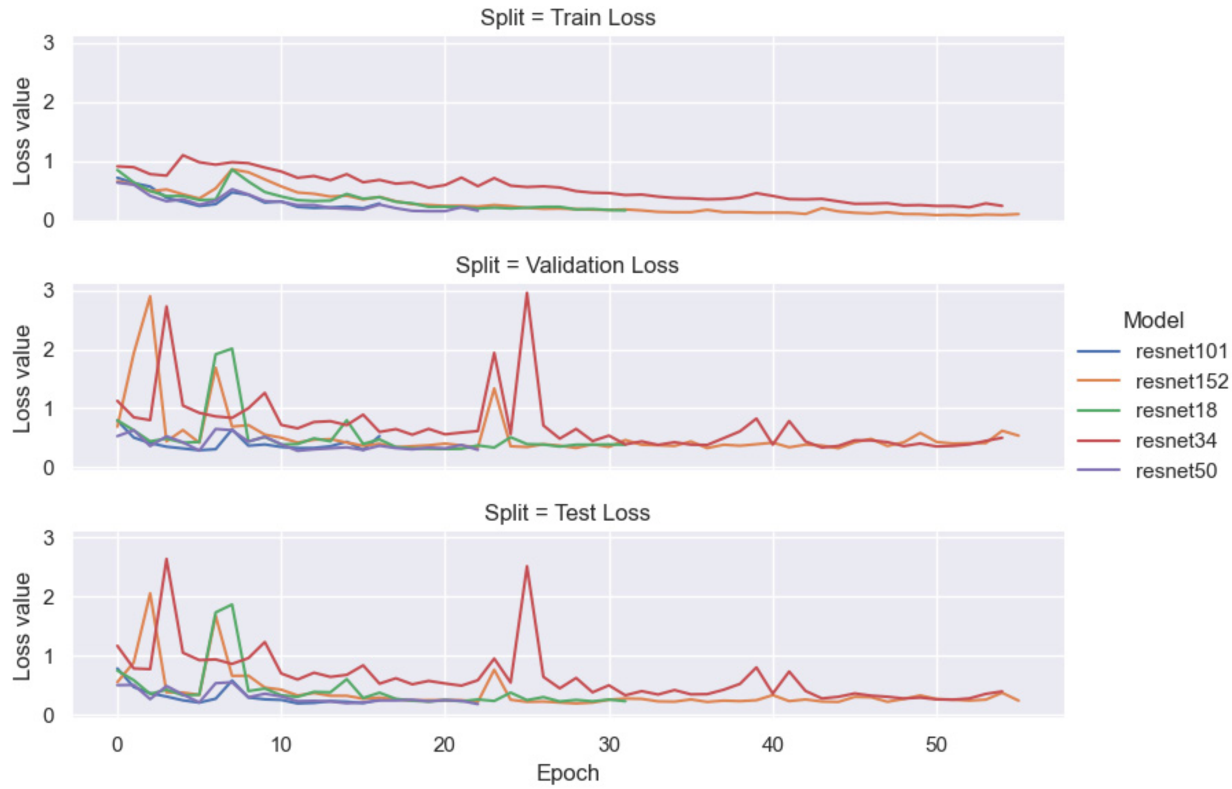

Figure 9. Training (top) and testing (bottom) loss for all models.

Table 3. Epoch number and losses of the models.

Model	Epoch	Train Loss	Validation Loss	Test Loss
resnet18	31	0.1613	0.3845	0.2341
resnet34	54	0.2385	0.5032	0.3995
resnet50	**22**	**0.1572**	**0.2998**	**0.1859**
resnet101	16	0.2789	0.5320	0.2589
resnet152	**55**	**0.0997**	**0.5386**	0.2440

5.4. Model Comparison

The models are compared according to their balanced accuracy in respect to each label separately. The results are shown in Table 4. The balanced accuracy metric has the advantage of handling the class imbalance in each label. It is important to evaluate the models in this research using balanced accuracy compared to regular accuracy because of the high imbalance between positive and negative outcomes. The models have an acceptable balanced accuracy in most labels exceed 85% and many exceeding 90%. Still, some labels, particularly "*animal on grade crossing*" and "*animal*", have a relatively very small number of positives which makes the results inconsistent in those labels. Overall, it is seen that the ResNet152 model has the highest validation accuracies in many labels compared to other models. Accordingly, based on the validation balanced accuracy in this subsection, and the losses investigated in the previous subsection, ResNet152 is selected as the best performing model.

Table 4. Validation balanced Accuracy for all models.

Label	Resnet 18	Resnet 34	Resnet 50	Resnet 101	Resnet 152
Rail Track	88.35%	87.65%	86.79%	88.41%	**91.72%**
Train	87.02%	82.70%	85.90%	84.00%	**88.93%**
Grade Crossing	92.81%	89.18%	91.38%	93.21%	**93.72%**
Grade Crossing Gate Down	92.83%	89.20%	92.81%	91.96%	**94.13%**
Red Light on Grade Crossing	92.32%	89.05%	92.69%	94.15%	**95.02%**
Train on Grade Crossing	87.88%	85.78%	88.67%	88.15%	**89.93%**
Vehicle Waiting for Train	91.85%	88.70%	90.28%	91.21%	**93.92%**
Trailer	95.28%	90.62%	**94.98%**	94.11%	94.04%
Vehicle on Grade Crossing	**93.40%**	91.57%	93.26%	93.90%	91.47%
Animals	94.54%	**94.79%**	95.28%	96.74%	92.61%
Bicyclist	**95.57%**	94.95%	92.79%	95.50%	92.58%
People on Grade Crossing	88.28%	90.93%	88.35%	**92.82%**	87.52%
Animal on Grade Crossing	99.32%	99.32%	99.32%	99.32%	99.32%
AVERAGE	92.27%	90.34%	91.73%	92.58%	**92.69%**

* Highest accuracy for label

5.5. Testing

The ResNet152 model is selected as the best candidate model based on this validation performance in the last subsections. However, since the validation performance was used for early training stop and for model selection. There is a need to test the model accuracy on an additional testing set that is not used in the previous tasks. A testing set of 15% of the data was withheld for this reason. The performance metrics from this testing set represent the performance of the model on completely unseen data. Several metrics are calculated as previously explained in the methodology section. The results are shown in Table 5 and continued in Table 6. The model has an acceptable balanced accuracy, exceeding 90% in most labels. Notably, the model achieved a balanced accuracy of 94% to detect vehicles at grade crossings. The F1 score is also shown in in Table 6. While the F1 score is acceptable for many labels, it is also low for many other labels, notably because they have a low number of actual positives. The F1 score is the harmonic mean of the PPV (precision) and TPV (recall). In this case, it is affected by the low PPV in some labels. The PPV is the ratio of TP to Predicted Positives (PP). The F1 score is therefore not considering the ability of the model to identify TNs in a highly imbalanced dataset, which is reflected in the balanced accuracy metric. Still, the low F1 score highlights the need for more positive images for the labels with low F1 to improve the performance of the model.

Table 5. Testing metrics.

Label	**AP**	**AN**	**PP**	**PN**	**TP**	**TN**	**FP**	**FN**	**Total**
label: Rail Track	645	98	611	132	605	92	6	40	743
label: Train	599	144	539	204	527	132	12	72	743
label: Grade Crossing	311	432	303	440	289	418	14	22	743
label: Grade Crossing Gate Down	229	514	227	516	212	499	15	17	743
label: Red Light on Grade Crossing	215	528	218	525	207	517	11	8	743
label: Train on Grade Crossing	206	537	217	526	183	503	34	23	743
label: Vehicle Waiting for Train	118	625	114	629	106	617	8	12	743
label: Trailer	54	689	62	681	48	675	14	6	743
label: Vehicle on Grade Crossing	28	715	54	689	26	687	28	2	743
label: People on Grade Crossing	21	722	42	701	20	700	22	1	743
label: Bicyclist	15	728	27	716	14	715	13	1	743
label: Animals	13	730	33	710	12	709	21	1	743
label: Animal on Grade Crossing	2	741	10	733	2	733	8	0	743

Table 6. Testing metrics (continued).

	Accuracy	**TPR**	**TNR**	**FPR**	**FNR**	**PPV**	**F1**	**Balanced Accuracy**
label: Rail Track	93.81%	93.80%	93.88%	6.12%	6.20%	99.02%	96.34%	93.84%
label: Train	88.69%	87.98%	91.67%	8.33%	12.02%	97.77%	92.62%	89.82%
label: Grade Crossing	95.15%	92.93%	96.76%	3.24%	7.07%	95.38%	94.14%	94.84%
label: Grade Crossing Gate Down	95.69%	92.58%	97.08%	2.92%	7.42%	93.39%	92.98%	94.83%
label: Red Light on Grade Crossing	97.44%	96.28%	97.92%	2.08%	3.72%	94.95%	95.61%	97.10%
label: Train on Grade Crossing	92.33%	88.83%	93.67%	6.33%	11.17%	84.33%	86.52%	91.25%
label: Vehicle Waiting for Train	97.31%	89.83%	98.72%	1.28%	10.17%	92.98%	91.38%	94.28%
label: Trailer	97.31%	88.89%	97.97%	2.03%	11.11%	77.42%	82.76%	93.43%
label: Vehicle on Grade Crossing	95.96%	92.86%	96.08%	3.92%	7.14%	48.15%	63.41%	94.47%
label: People on Grade Crossing	96.90%	95.24%	96.95%	3.05%	4.76%	47.62%	63.49%	96.10%
label: Bicyclist	98.12%	93.33%	98.21%	1.79%	6.67%	51.85%	66.67%	95.77%
label: Animals	97.04%	92.31%	97.12%	2.88%	7.69%	36.36%	52.17%	94.72%
label: Animal on Grade Crossing	98.92%	100.00%	98.92%	1.08%	0.00%	20.00%	33.33%	99.46%

Considering that the goal of the mode is to identify hazards and avoid potential accidents, it is important to avoid False Negatives (FN). A FN means that the model output indicates that there are no hazards (such as vehicle, train, or bicyclist) of the grade crossing, while there is a hazard on the grade crossing. This situation may result in an avoidable accident and must be minimized. According to Table 6 if can be seen as the False Negative Rate (FNR), which is the type II error, relatively high in some labels. For the label "Vehicle on Grade Crossing", the FNR is 7.14%. Lowering the FNR of the model can be achieved by calibrating the threshold to consider positives. However, this adjustment is a tradeoff that would increase FP. The need to evaluate the thresholds of the model is achieved by plotting the Receiver Operating Characteristic (ROC) curve as shown in Figure 10. The ROC curve shows the TPR vs. FPR. It indicates the performance of the classifiers using a variety using a range of threshold values. The ROC Area Under the Curve (AUC) measures the performance of the model and can range from zero to one where a random classifier would have a score of 0.5. The ResNet152 model can achieve high AUC scores, exceeding 0.96 in all labels, as shown in Figure 10, which indicates that the model has a good performance considering various thresholds.

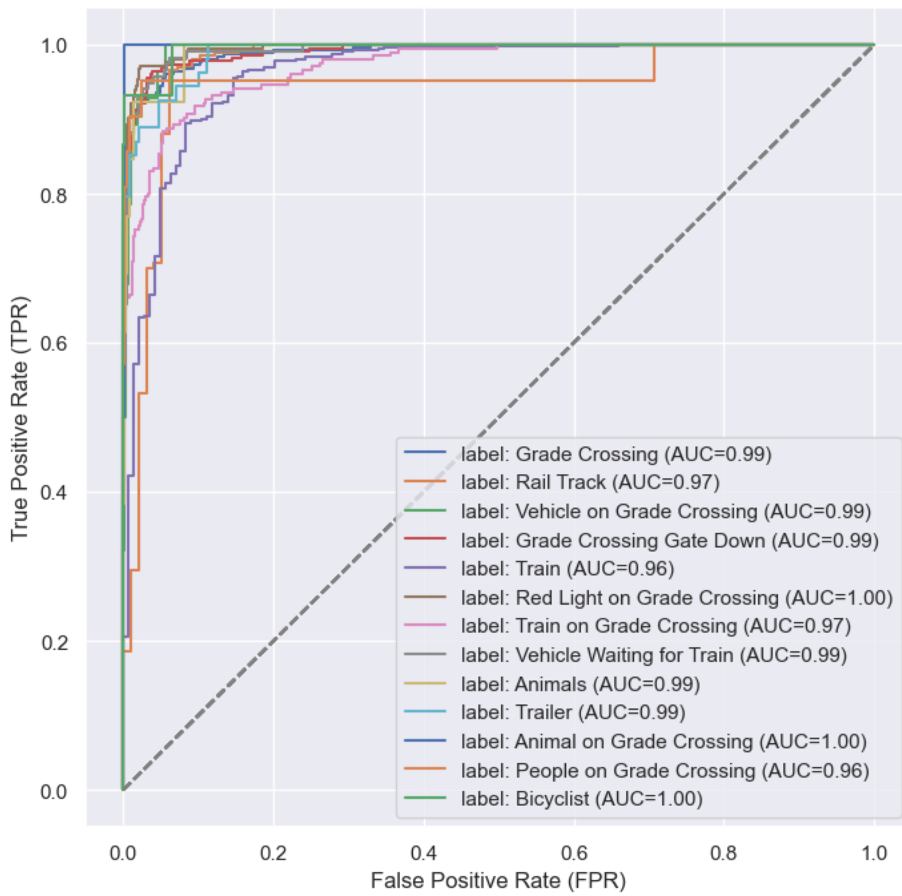

Figure 10. ROC Curves.

Finally, a selection of successful examples of model outcomes is shown in Figure 11. The examples depict that the model can correctly identify various situations in the images, such as vehicles, humans, and traffic lights, in different scenarios, environments, and lighting conditions, which achieves the goal of this research.

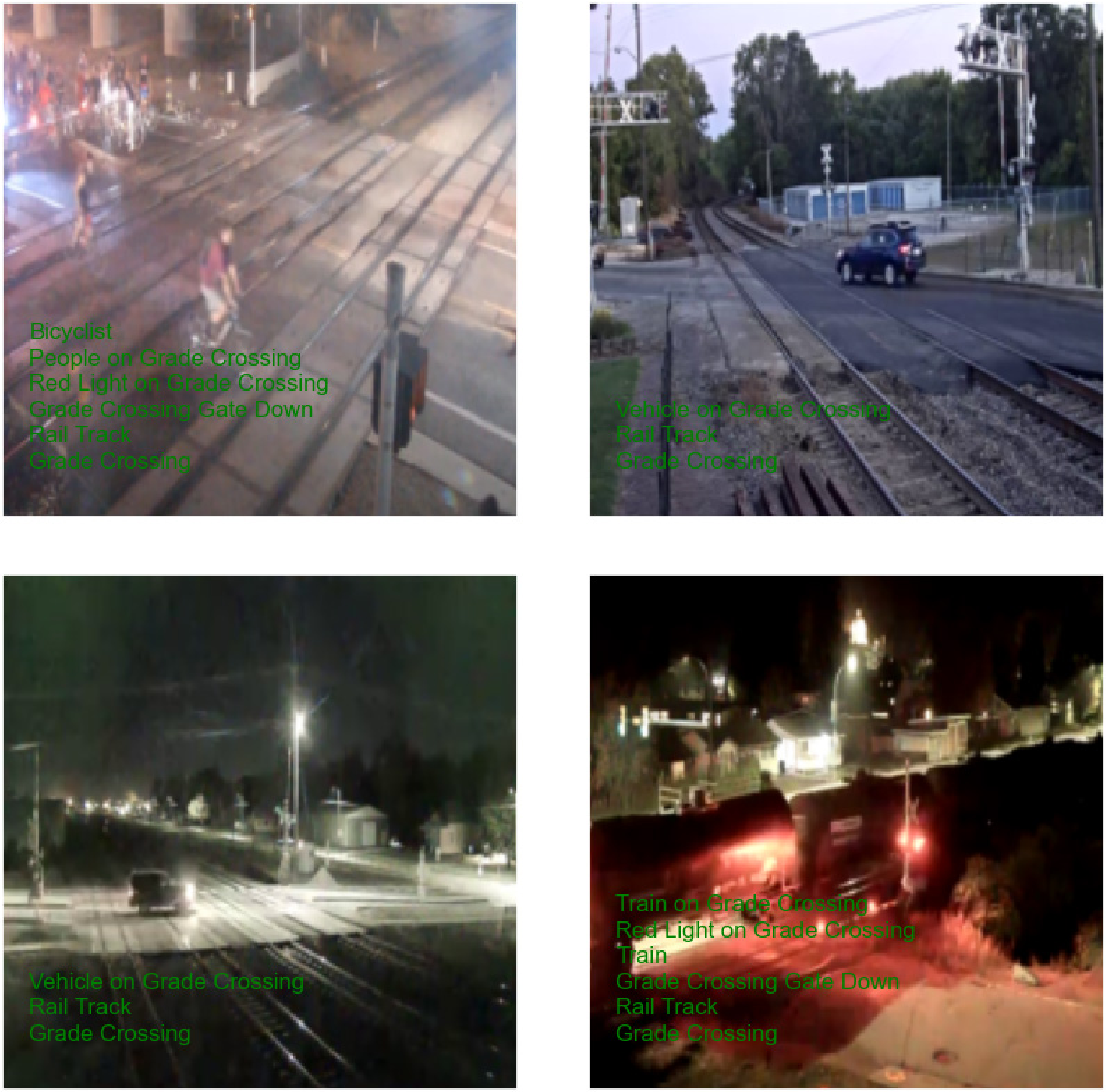

Figure 11. Examples of model outputs.

6. Conclusion

The goal of this research was to investigate the application of AI and computer vision to detect hazards at grade crossings. This goal was achieved by (1) collecting and labelling a large dataset of images and (2) developing a CNN model using the collected dataset. Several versions of the established ResNet model were adopted and finetuned. ResNet152 was selected as the best candidate. Testing results indicate that the model has a high balanced accuracy, exceeding 90% in most labels. Notably, the model achieved a balanced accuracy of 94% to detect vehicles at grade crossings. The model also has a high ROC AUC score exceeding 0.96 in all labels.

The contribution of this model is to improve safety at grade crossings by detecting hazards using camera feeds. The model can use feeds from regular cameras including existing CCTV cameras. Furthermore, the model is trained to handle any grade crossing with various scenarios, lighting, and weather conditions. The model does not require human intervention to customize it for new locations and can readily detect various hazards. Overall, the model is intended to be a robust, cost-effective, and automated tool to monitor crossings.

This model can be implemented in several ways. First, the model can be deployed at high-traffic grade crossings as an additional safety monitoring system. It may be connected to positive train control systems to raise alarms when a hazardous situation is detected while a train is expected. As such, the model may alert train crew and initiate stopping mechanisms. Second, the model may be implemented to assist field inspectors to monitor hazards at crossings and report summary statistics of hazardous events. The data may be further analyzed to identify high-risk crossings, evaluate the effectiveness of safety plans, and to support decision making.

However, the limitations of the model are noted in some labels that have a very high imbalance due to the low number of positives. These labels are associated with relatively infrequent cases such as having bicyclists or animals on grade crossings. Such limitations can be addressed in future work by collecting more images for these labels or synthesizing the images. Furthermore, the authors used ResNet as the based model. Other established or emerging models can be tested in future work to pursue better performance.

7. References

Alawad, H., S. Kaewunruen, and M. An. 2020. "A Deep Learning Approach Towards Railway Safety Risk Assessment." *IEEE Access*, 8: 102811–102832. https://doi.org/10.1109/ACCESS.2020.2997946.

Ali, G., C. Tarawneh, F. Chavez, and D. Espinoza. 2024. *Grade Crossing Monitoring Using Deep Learning*. University Transportation Center for Railway Safety (UTCRS) Tier-1 ….

Amaral, V., F. Marques, A. Lourenço, J. Barata, and P. Santana. 2016. "Laser-Based Obstacle Detection at Railway Level Crossings." *Journal of Sensors*, 2016 (1): 1719230. https://doi.org/10.1155/2016/1719230.

Chlap, P., H. Min, N. Vandenberg, J. Dowling, L. Holloway, and A. Haworth. 2021. "A review of medical image data augmentation techniques for deep learning applications." *J Med Imag Rad Onc*, 65 (5): 545–563. Wiley. https://doi.org/10.1111/1754-9485.13261.

Cubuk, E. D., B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le. 2019a. "Autoaugment: Learning augmentation strategies from data." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 113–123.

Cubuk, E. D., B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le. 2019b. "AutoAugment: Learning Augmentation Policies from Data." arXiv.

Deng, J., W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. 2009. "Imagenet: A large-scale hierarchical image database." *2009 IEEE conference on computer vision and pattern recognition*, 248–255. Ieee.

Espinoza, D., G. Ali, and C. Tarawneh. 2024. "AI-Based Hazard Detection for Railway Crossings." *ASME/IEEE Joint Rail Conference*, V001T05A004. American Society of Mechanical Engineers.

Etxeberria-Garcia, M., M. Labayen, M. Zamalloa, and N. Arana-Arexolaleiba. 2020. "Application of Computer Vision and Deep Learning in the railway domain for autonomous train stop operation." *2020 IEEE/SICE International Symposium on System Integration (SII)*, 943–948.

Greitans, K. M. 2023. "A study on Automated railway level crossing control system using FMCW radar for accident prevention." *2023 International Conference on Applied Electronics (AE)*, 1–6.

Guo, F., Y. Wang, and Y. Qian. 2022. "Computer vision-based approach for smart traffic condition assessment at the railroad grade crossing." *Advanced Engineering Informatics*, 51: 101456. https://doi.org/10.1016/j.aei.2021.101456.

Gupta, J., S. Pathak, and G. Kumar. 2022. "Deep Learning (CNN) and Transfer Learning: A Review." *J. Phys.: Conf. Ser.*, 2273 (1): 012029. IOP Publishing. https://doi.org/10.1088/1742-6596/2273/1/012029.

Hari Narayanan, A., P. Brennan, R. Benjamin, N. Mazzino, G. Bochetti, and A. Lancia. 2011. "Railway level crossing obstruction detection using MIMO radar." *2011 8th European Radar Conference*, 57–60.

He, F., T. Liu, and D. Tao. 2020. "Why resnet works? residuals generalize." *IEEE transactions on neural networks and learning systems*, 31 (12): 5349–5362. IEEE.

He, K., X. Zhang, S. Ren, and J. Sun. 2015. "Deep Residual Learning for Image Recognition." arXiv.

He, K., X. Zhang, S. Ren, and J. Sun. 2016. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Hosotani, D., I. Yoda, and K. Sakaue. 2009. "Development and Long-Term Verification of Stereo Vision Sensor System for Controlling Safety at Railroad Crossing." *Computer Vision Systems*, Lecture Notes in Computer Science, M. Fritz, B. Schiele, and J. H. Piater, eds., 154–163. Berlin, Heidelberg: Springer.

Howard, A., M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam. 2019. "Searching for MobileNetV3." arXiv.

Imambi, S., K. B. Prakash, and G. R. Kanagachidambaresan. 2021. "PyTorch." *Programming with TensorFlow: Solution for Edge Computing Applications*, EAI/Springer Innovations in Communication and Computing, K. B. Prakash and G. R. Kanagachidambaresan, eds., 87–104. Cham: Springer International Publishing.

Jain, T., C. Lennan, Z. John, and D. Tran. 2019. "Imagededup." Python. idealo.

Kingma, D. P., and J. Ba. 2017. "Adam: A Method for Stochastic Optimization." arXiv.

Label Studio. 2023. "Open Source Data Labeling." *Label Studio*. Accessed December 3, 2023. https://labelstud.io/.

LeCun, Y., Y. Bengio, and G. Hinton. 2015. "Deep learning." *Nature*, 521 (7553): 436–444. Nature Publishing Group. https://doi.org/10.1038/nature14539.
Lee, J., T. Kim, S. Beak, Y. Moon, and J. Jeong. 2023. "Real-time pose estimation based on ResNet-50 for rapid safety prevention and accident detection for field workers." *Electronics*, 12 (16): 3513. MDPI.
Lee, J., and S. Lee. 2023. "Construction Site Safety Management: A Computer Vision and Deep Learning Approach." *Sensors*, 23 (2): 944. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/s23020944.
Li, B., and D. Lima. 2021. "Facial expression recognition via ResNet-50." *International Journal of Cognitive Computing in Engineering*, 2: 57–64. https://doi.org/10.1016/j.ijcce.2021.02.002.
Li, S., J. Jiao, Y. Han, and T. Weissman. 2017. "Demystifying ResNet." arXiv.
Mikołajczyk, A., and M. Grochowski. 2018. "Data augmentation for improving deep learning in image classification problem." *2018 International Interdisciplinary PhD Workshop (IIPhDW)*, 117–122.
Murshed, R. U., S. K. Dhruba, Md. T. I. Bhuian, and Mst. R. Akter. 2022. "Automated Level Crossing System: A Computer Vision Based Approach with Raspberry Pi Microcontroller." *2022 12th International Conference on Electrical and Computer Engineering (ICECE)*, 180–183.
NVIDIA NGC. 2023. "ResNet v1.5 for PyTorch." *NVIDIA NGC Catalog*. Accessed May 11, 2025. https://catalog.ngc.nvidia.com/orgs/nvidia/resources/resnet_50_v1_5_for_pytorch.
Oh, K., M. Yoo, N. Jin, J. Ko, J. Seo, H. Joo, and M. Ko. 2022a. "A Review of Deep Learning Applications for Railway Safety." *Applied Sciences*, 12 (20): 10572. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/app122010572.
Oh, K., M. Yoo, N. Jin, J. Ko, J. Seo, H. Joo, and M. Ko. 2022b. "A Review of Deep Learning Applications for Railway Safety." *Applied Sciences*, 12 (20): 10572. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/app122010572.
O'Mahony, N., S. Campbell, A. Carvalho, S. Harapanahalli, G. V. Hernandez, L. Krpalkova, D. Riordan, and J. Walsh. 2020. "Deep Learning vs. Traditional Computer Vision." *Advances in Computer Vision*, Advances in Intelligent Systems and Computing, K. Arai and S. Kapoor, eds., 128–144. Cham: Springer International Publishing.
Perez, L., and J. Wang. 2017. "The Effectiveness of Data Augmentation in Image Classification using Deep Learning." arXiv.
PyTorch. 2025a. "Models and pre-trained weights — Torchvision main documentation." Accessed May 15, 2025. https://web.archive.org/web/20250515181131/https://docs.pytorch.org/vision/master/models.html#table-of-all-available-classification-weights.
PyTorch. 2025b. "BCEWithLogitsLoss." *PyTorch 2.7 documentation.* Accessed May 24, 2025. https://docs.pytorch.org/docs/stable/generated/torch.nn.BCEWithLogitsLoss.html.
Qin, Y., Z. Cao, Y. Sun, L. Kou, X. Zhao, Y. Wu, Q. Liu, M. Wang, and L. Jia. 2023. "Research on Active Safety Methodologies for Intelligent Railway Systems." *Engineering*, 27: 266–279. https://doi.org/10.1016/j.eng.2022.06.025.
Salmane, H., L. Khoudour, and Y. Ruichek. 2013. "Improving safety of level crossings by detecting hazard situations using video based processing." *2013 IEEE International Conference on Intelligent Rail Transportation Proceedings*, 179–184.
Sheikh, Y. A., Y. Zhai, K. Shafique, and M. A. Shah. 2004. "Visual monitoring of railroad grade crossing." E. M. Carapezza, ed., 654. Orlando, FL.

Shin, D., J. Jin, and J. Kim. 2021. "Enhancing Railway Maintenance Safety Using Open-Source Computer Vision." *Journal of Advanced Transportation*, 2021: e5575557. Hindawi. https://doi.org/10.1155/2021/5575557.

Showkat, S., and S. Qureshi. 2022. "Efficacy of Transfer Learning-based ResNet models in Chest X-ray image classification for detecting COVID-19 Pneumonia." *Chemometrics and Intelligent Laboratory Systems*, 224: 104534. https://doi.org/10.1016/j.chemolab.2022.104534.

Starčević, M., D. Barić, and H. Pilko. 2016. "Safety at level crossings: Comparative analysis." *Road and Rail Infrastructure IV*, 861–868. Zagreb: Gra\djevinski fakultet Sveučilišta u Zagrebu.

The American Society of Civil Engineers (ASCE). 2025. *Rail Infrastructure Report Card*. USA: The American Society of Civil Engineners (ASCE).

US Department of Transportation. 2024. "Highway-Rail Grade Crossing Incidents, Fatalities and Injuries (2.08)." *Transportation.gov*. Accessed March 3, 2025. https://data.transportation.gov/stories/s/Highway-Rail-Grade-Crossing-Incidents-Fatalities-a/bda5-32at/.

US Department of Transportation (USDOT) Federal Railroad Administration (FRA). 2025. "Highway-Rail Grade Crossing Safety and Trespass Prevention." Accessed May 6, 2025. https://web.archive.org/web/20250415074105/https://railroads.dot.gov/crossing-safety-trespass-prevention/crossing-safety-trespass-prevention.

Virtual Railfan, Inc. 2023. "Virtual Railfan, Inc. – We bring the trains to you!" Accessed December 3, 2023. https://virtualrailfan.com/.

Vivek, V., J. Hemalatha, T. P. Latchoumi, and S. Mohan. 2023. "Towards the development of obstacle detection in railway tracks using thermal imaging." *NNW*, 33 (5): 337–355. https://doi.org/10.14311/NNW.2023.33.019.

Wang, S., X. Li, Z. Chen, and Y. Liu. 2024. "A Railway Accident Prevention System Using an Intelligent Pilot Vehicle." *IEEE Transactions on Intelligent Transportation Systems*, 25 (6): 5170–5188. https://doi.org/10.1109/TITS.2023.3331901.

Xu, W., Y.-L. Fu, and D. Zhu. 2023. "ResNet and its application to medical image processing: Research progress and challenges." *Computer Methods and Programs in Biomedicine*, 240: 107660. https://doi.org/10.1016/j.cmpb.2023.107660.

Yoda, I., K. Sakaue, and D. Hosotani. 2006. "Multi-point Stereo Camera System for Controlling Safety at Railroad Crossings." *Fourth IEEE International Conference on Computer Vision Systems (ICVS'06)*, 51–51.

Zayandehroodi, M., Mojaradi ,Barat, and M. and Bagheri. 2025. "Evaluating the effectiveness of safety countermeasures at highway–railway grade crossing based on a machine learning framework." *Traffic Injury Prevention*, 26 (1): 120–127. Taylor & Francis. https://doi.org/10.1080/15389588.2024.2387713.

Zhang, A., Z. C. Lipton, M. Li, and A. J. Smola. 2023. *Dive into deep learning*. Cambridge University Press.

Zhang, Z., C. Trivedi, and X. Liu. 2018. "Automated detection of grade-crossing-trespassing near misses based on computer vision analysis of surveillance video data." *Safety Science*, Railway safety, 110: 276–285. https://doi.org/10.1016/j.ssci.2017.11.023.