Detecting Hazards at Rail Grade Crossings using Computer Vision and AI

Silvia Flores-Osuna[1]; Gasser G. Ali, Ph.D.[2]; Darren Espinoza[3];
Felix Chavez [4]; and Constantine Tarawneh[5]

[1]Undergraduate Student, Department of Computer Science, The University of Texas Rio Grande Valley, Edinburg, TX, USA 78539, Email: silvia.floresosuna01@utrgv.edu
[2]Assistant Professor, Department of Civil Engineering, University of Texas Rio Grande Valley, Edinburg, TX, USA 78539. Email: gasser.ali@utrgv.edu
[3]Graduate Student, Department of Mechanical Engineering, The University of Texas Rio Grande Valley, Edinburg, TX, USA 78539, Email: darren.espinoza01@utrgv.edu
[4]Undergraduate Student, Department of Computer Science, The University of Texas Rio Grande Valley, Edinburg, TX, USA 78539, Email: felix.chavez01@utrgv.edu
[5]Louis A. Beecherl, Jr. Endowed Professor, Sr. Associate Dean for Research and Graduate Programs, CECS, Director, UTCRS, Director, NSF CREST MECIS, Distinguished Teaching Professor, Mechanical Engineering Department, University of Texas Rio Grande Valley, Edinburg, TX, USA 78539, Email: constantine.tarawneh@utrgv.edu

ABSTRACT

The Federal Railroad Administration reported a total of 2,250 highway-rail incidents during 2024. Continuous efforts by transportation agencies aim to lower these numbers by spreading awareness, improving infrastructure, and installing warning devices. Efforts have been made to enhance safety at these crossings using a variety of innovative technologies; however, there remains a need for a robust and cost-effective system capable of detecting hazards at various crossings under diverse climate and lighting conditions. This paper explores the application of artificial intelligence and deep learning to develop a Convolutional Neural Network (CNN) for image-level multi-class classification of hazards at grade crossings, including stalled vehicles, pedestrians, and animals. An extensive dataset of grade crossing images was collected from publicly available sources and manually labeled. A MobileNetV3-based CNN model was trained using transfer learning across 13 hazard classes. Model performance was evaluated on unseen data by assessing balanced accuracy, F1 score, and Area Under Curve (AUC). Results indicate that the proposed model achieves balanced accuracy exceeding 90% for most hazard classes, with F1 scores greater than 0.8 and AUC values above 0.95, demonstrating reliable hazard classification performance. Overall, the findings support the potential of AI-driven computer vision systems to enhance safety at grade crossings.

INTRODUCTION

Grade crossings are accident-prone despite various existing safety measures such as warning signs, crossbucks, gates, and flashing lights. In 2024 alone, 2,260 incidents resulted in 261 fatalities and 762 injuries according to the US Department of Transportation (USDOT) and the Federal Railroad Administration (FRA) (2024). Furthermore, the FRA states that 96% of rail-related fatalities over the past 10 years have been due to highway-rail grade crossing and trespassing incidents (US Department of Transportation (USDOT) Federal Railroad Administration (FRA) 2025). With approximately 140,000 miles of rail track across the US, there

is a critical need for improved safety measures that can detect pedestrians, vehicles, and other items on the track.

Safety measures such as warning signs, barriers, and crossbucks have been set in place to address this need. However, despite their ability to warn passersby of incoming trains, they cannot detect hazards on the crossing. Various studies have been conducted aiming to solve this problem. These studies focus on the detection of pedestrians, vehicles, and objects on the track using various equipment. Some opt for stereo cameras (Hosotani et al. 2009), while others use video cameras (Sheikh et al. 2004) or a combination of CCTV cameras with radar (Greitans 2023). In other cases, technology such as MIMO radars (Hari Narayanan et al. 2011) or LIDAR laser detectors (Amaral et al. 2016) are utilized. However, many of these methods face several limitations. They often focus on one type of hazard, only function in "hot zones", or don't work very well in diverse lighting. Therefore, such measures are difficult to deploy at scale and do not support continuous, automated collection of hazard-related data, which is critical for safety analysis and informing safety decisions. Studies that meet these criteria disregard the potential for automated collection of data. Furthermore, there is a lack of public datasets containing labeled images of hazards at grade crossings, limiting reproducibility and hindering future research in the field.

More recently, deep learning has revolutionized the field due to its ability to achieve strong performance in object detection and image classification tasks. It has already been applied to railway safety, including grade-crossing and traffic condition monitoring (LeCun et al. 2015; Oh et al. 2022). Despite these advances, there remains a need for a robust, cost-effective model that can classify multiple hazard types under diverse environmental conditions while simultaneously enabling scalable data collection. To address this gap, this work builds upon prior research by the authors (Espinoza et al. 2024) by expanding the dataset and adopting a MobileNetV3 architecture, improving both generalization and suitability for deployment on low-cost cameras.

GOAL

The goal of this paper is to investigate the applicability of deep learning and image-level multi-class hazard classification in the prevention of accidents at grade crossings involving pedestrians, vehicles, and animals. Although safety measures such as warning signals and barriers already exist, they are not always sufficient, especially at low-income crossings where they are often missing. Deep learning-based computer vision methods offer an additional layer of safety by enabling automated classification of hazards at crossings and supporting timely safety interventions. This paper proposes a robust, cost-effective, MobileNetV3-based CNN that can operate effectively on video feeds collected at grade crossings in diverse environmental conditions while simultaneously supporting scalable data collection for safety analysis and future research. For model training and evaluation, the authors collected and manually labeled a large dataset of grade-crossing images.

METHODOLOGY

This model was developed in six steps: data collection, data preparation, data labeling, data augmentation, model selection, and model training. Each section is described in more detail below. Figure 1 demonstrates the order the methodology was performed.

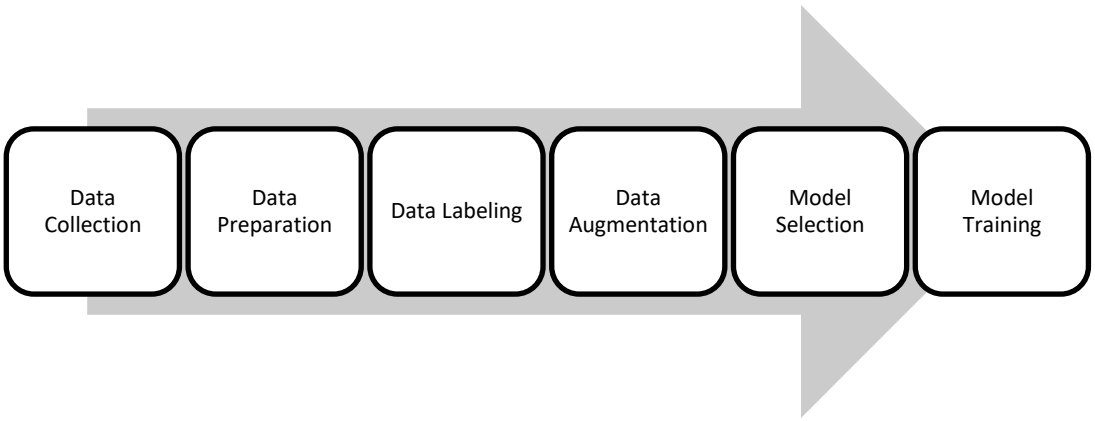

Figure 1. An overview of the methodology.

Data Collection. Machine learning models require vast datasets to learn from. However, because no suitable datasets depicting hazards at grade crossings were readily available, the authors manually collected 4,699 images. To do so, multiple public videos and livestreams depicting grade crossings were downloaded from various YouTube channels, such as "Virtual RailFan, Inc" by using 'ytb-dlp', a Python package. Figure 2 shows an image collected from the YouTube channel. These videos were downloaded in 640 x 360 pixels.

Figure 2. An image collected online from YouTube.

Data Preparation. Because the collected videos had an unnecessary number of frames, one out of every five frames were selected from each video. Despite this, a high number of similar images remained. For example, many of the frames contained empty grade crossings; this redundant data would not provide any additional benefits. To address this, the Python package "Image Duplicator" ("Imagededup", 2019) was utilized. This package makes use of user-selected algorithms to encode images; in this case, MobileNetV3 (Howard et al. 2019) was used. These encodings are then compared using a cosine similarity threshold set by the user, with duplicates determined when the similarity exceeds the set threshold. For this dataset, a threshold of 95% was used. Although this step reduced redundancy, the dataset still suffered from a lack of diversity. A diverse dataset is essential for reducing overfitting and improving the generalization and robustness of a model. The dataset collected, however, lacked variation in camera angles and different scenarios, which don't occur often in the real world. This limitation is discussed further in the conclusion.

Data Labeling. After cleaning the data, the dataset was manually labelled according to non-exclusive 13 labels: 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. Since several images contained multiple relevant labels, a multi-label classification approach was used. The labeling of the dataset was done manually by using the program Label Studio (Label Studio 2023). Examples of the model labels and output of the model are later shown in Figure 9.

Data Augmentation. To increase the model's generalizability and the data imbalance, data augmentation was performed. Data augmentation is a method used in machine learning to expand the dataset; traditional methods involve cropping, resizing, rotating, and slight alterations (Mikołajczyk and Grochowski 2018). It is a method that has proven successful in increasing training datasets, especially in the medical field (Chlap et al. 2021; Mikołajczyk and Grochowski 2018; Perez and Wang 2017). In this study, AutoAugment, a method introduced in 2019, was used (Cubuk et al. 2019). The chosen data augmentation policy for this paper was the ImageNet augmentation policy, which is implemented in the torchvision framework. This policy was chosen due to its demonstrated effectiveness across diverse image classification tasks and its ability to improve robustness, which is beneficial when working with class-imbalanced datasets. Some augmentations performed by this policy are rotations, which can prove useful if the camera is accidentally tilted. Figure 3 shows an example of an augmented image. The augmentation policy also includes transformations such as color changes, translations, and distortions.

Figure 3. An image from the dataset that has been rotated for the data augmentation process.

Model Selection. Due to its lightweight nature and efficiency on embedded systems, MobileNetV3 was selected as the primary model for this multi-label image classification task. MobileNetV3 is the third version of MobileNet, which improves upon its predecessors, MobileNet and MobileNetV2. MobileNet was initially introduced in 2017 as a lightweight architecture using depth-wise separable convolutions, which greatly reduced computation compared to standard convolutions at the cost of a reasonable amount of accuracy (Howard et al. 2017). MobileNetV3 would go on to improve accuracy and latency improvement methods (Howard et al. 2019). Beyond its performance, MobileNet has already been used in various applications, including but not limited to medical imaging (Goel and Singh 2024), object detection (Chiu et al. 2020), and facial recognition (Kashyap and Kumar 2023). Both its design choices and practical applications made MobileNetV3 the optimal choice.

Model Training. Because the model of choice was pretrained on ImageNet (Deng et al. 2009), a dataset with more than 14 million images, transfer learning was performed. Transfer learning is the process of using previous knowledge and applying it to a related task (Torrey and Shavlik 2010), which effectively cuts down on training time and the need for data. However, the dataset contained only 13 labels, while the model generated 1,000 outputs. To address this, the last layer of the model was modified to instead have 13 outputs, each one corresponding to the labels mentioned above. Furthermore, the data was split into 80% for training and 20% for validation to ensure proper model accuracy and prevent overfitting. To determine how well the model performed, the evaluation metrics used were the Area Under Curve (AUC), both accuracy and balanced accuracy, and the F1 score for each label.

RESULTS AND ANALYSIS

The authors' results are presented in two parts: an analysis of the data collected and an evaluation of the model by observing evaluation metrics.

Analysis of Collected Data. In total, there were 4,699 labeled images. Figure 4 demonstrates the amount of data for each label. As shown in Figure 4, there was a lack of images featuring animals or bicyclists on the railroad, with the majority containing trains and rail tracks. Despite this, images containing animals and cyclists were included to test how the model would perform in such cases.

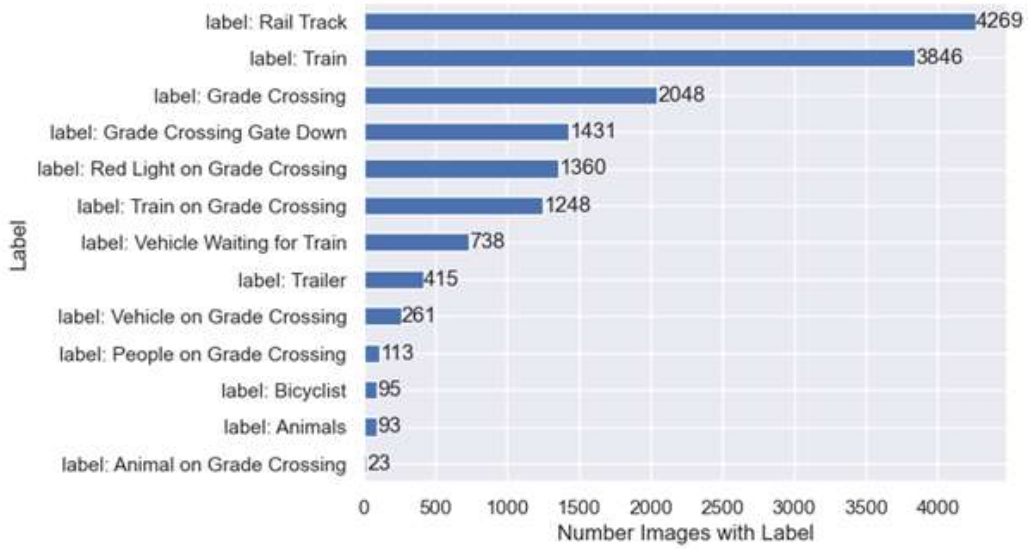

Figure 4. A bar graph that displays the frequency of each label.

Model Accuracy. As described in the methodology, the dataset was separated into two sections during model training: 80% for training, and 20% for validation. The training, validation, and testing losses are shown in Figure 5. As seen in the figure, all three decreased for every epoch and stopped below 0.1, indicating the model did a satisfactory job at learning and performed well at all levels. Furthermore, the AUC values for training and validation are shown in Figure 6. These values also improved at every epoch, with all labels achieving a validation AUC value close to 1. This demonstrates that the model was able to effectively classify hazards. Furthermore, Figure 7 illustrates the Receiver Operating Characteristic (ROC) curves for every label and corresponding AUC. The ROC AUC is a metric that ranges from 0 to 1, with 0.5 meaning the model is guessing

at random and 1 indicating perfect model prediction. Every label exceeds an AUC of 0.95, which signifies the model performs consistently, considering various thresholds. These values, coupled with the balanced accuracies and F1 scores (which reflect both precision and recall) displayed in Table 1 show that the model had good performance all around with most labels, with the categories "Animal on Grade Crossing", "Animal", and "Pedestrian on Grade Crossing" having the lowest balanced accuracy and F1 score, likely due to class imbalance and the relatively small number of labeled images for these events. To further analyze classification behavior, confusion matrices were created for each label, as shown in Figure 8, illustrating the distribution of true positives, true negatives, false positives, and false negatives across all classes. The matrices indicated high true positives for most labels, while minority classes indicated higher false negative rates, further highlighting the impact of class imbalance. Examples of the model outputs are shown in Figure 9.

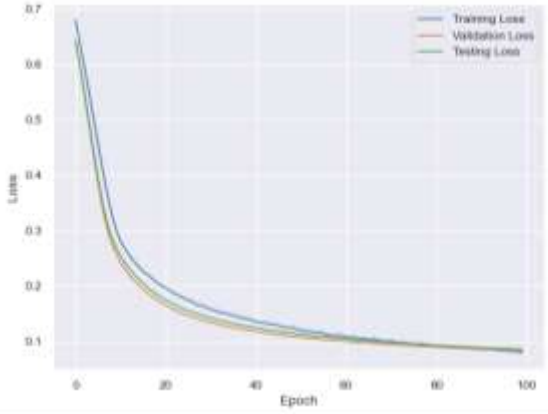

Figure 5. The training, validation, and testing loss of the model.

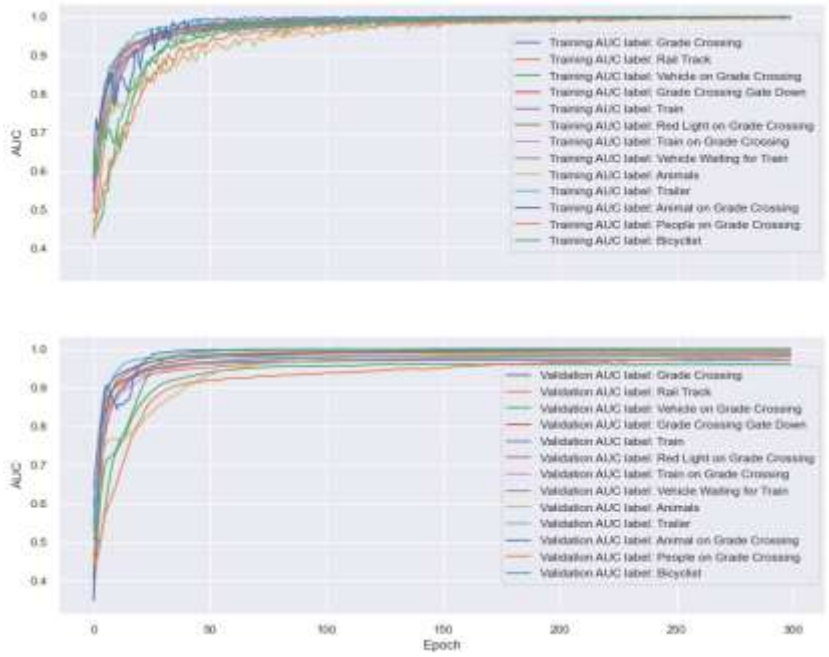

Figure 6. Both the AUC for training (top) and validation (bottom) for each label.

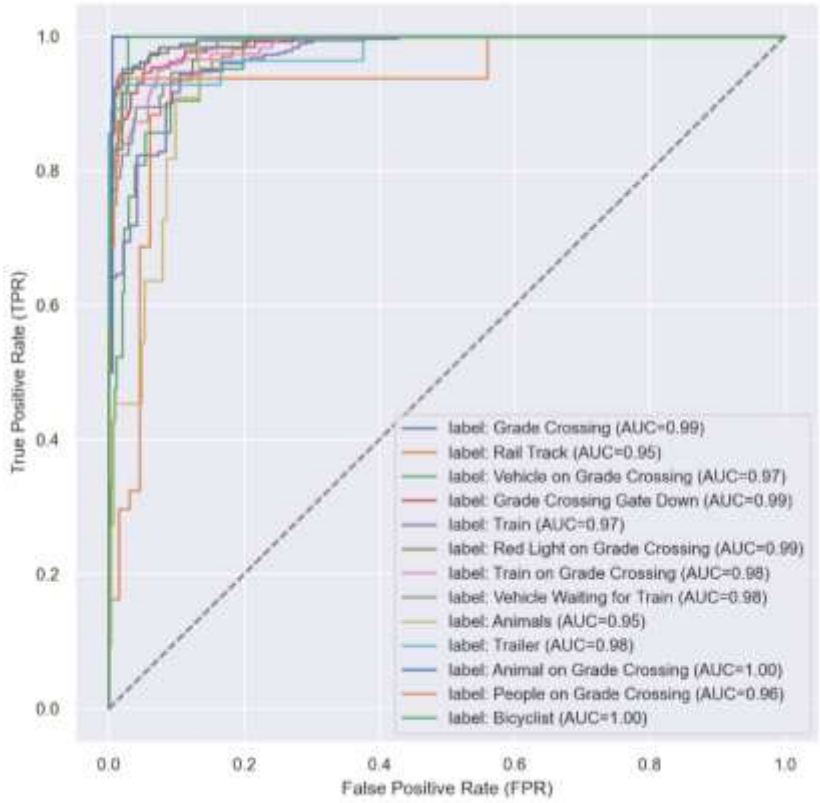

Figure 7. ROC curves for each class

Table 1. Evaluation metrics for each label.

Label	Accuracy	Balanced Accuracy	F1	AUC
Animal on Grade Crossing	99.57%	50.00%	0	1.00
Bicyclist	98.51%	70.83%	0.59	1.00
Grade Crossing	95.96%	95.42%	0.95	0.99
Red Light on Grade Crossing	97.23%	95.82%	0.95	0.99
Trailer	98.94%	92.74%	0.91	0.98
Grade Crossing Gate Down	95.32%	94.21%	0.92	0.99
Train on Grade Crossing	94.26%	90.88%	0.88	0.98
Rail Track	97.02%	90.03%	0.98	0.95
Vehicle Waiting for Train	95.96%	88.63%	0.83	0.98
Train	92.98%	87.85%	0.96	0.97
Vehicle on Grade Crossing	96.81%	75.63%	0.59	0.97
People on Grade Crossing	97.02%	56.25%	0.22	0.96
Animals	97.66%	63.31%	0.35	0.95

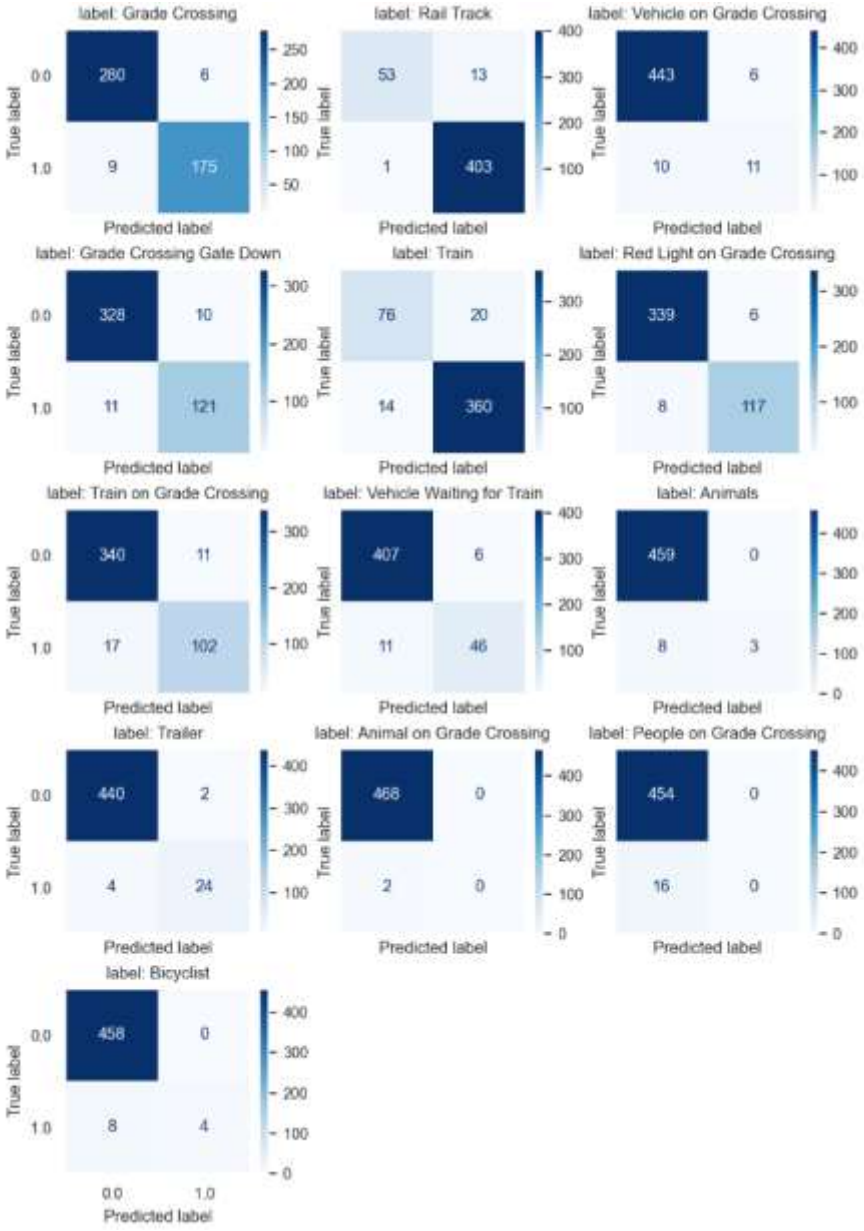

Figure 8. Confusion matrices for each label.

CONCLUSION

The goal of this paper was to explore the applicability of deep learning-based, image-level multi-class hazard classification at grade crossings, a field in critical need of safety improvements. This was accomplished by collecting and labeling a dataset of hazards on grade crossings, which was then used to develop a MobileNetV3 based CNN model designed to process video feed from inexpensive cameras installed at grade crossings. The model demonstrated strong performance with an AUC above 0.97 for all labels and a balanced accuracy above 90% for most labels. The resulting model can be used in two ways after implementation at grade crossings: to detect and classify hazards at grade crossings, and to collect data for field experts. This collected data can help identify high-risk crossings and evaluate and support decision-making for the implementation

of safety measures. However, the study is limited by the highly unbalanced data set. Footage of animals and bicyclists on the crossing was difficult to collect, as these are not events that happen often. This limitation can be addressed in future work by collecting more footage of these events.

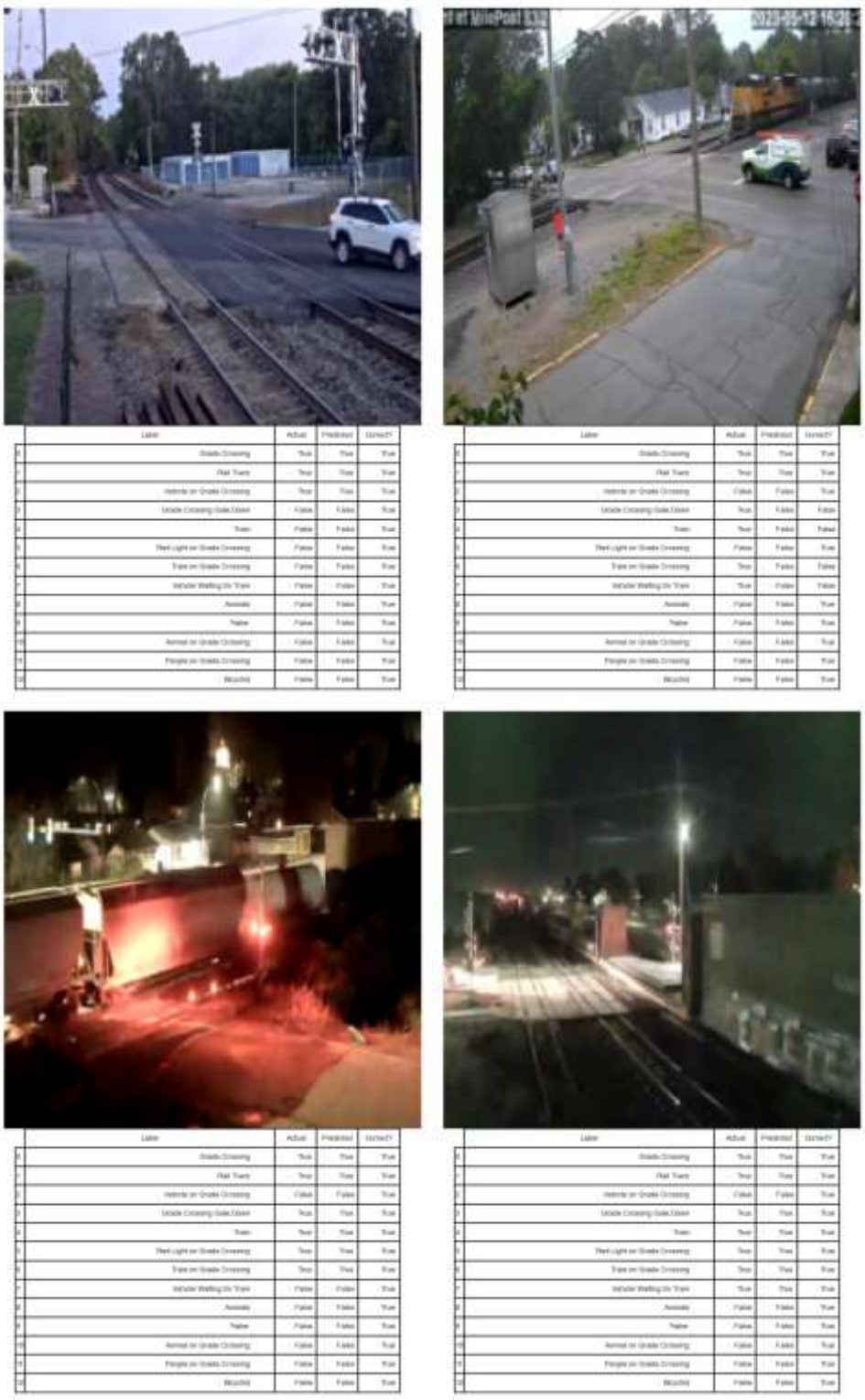

Figure 9. Examples of outputs from the model.

ACKNOWLEDGMENTS

The authors want to acknowledge the University Transportation Center for Railway Safety (UTCRS) at UTRGV for the financial support provided to perform this study through the USDOT UTC Program under Grant No. 69A3552348340.

REFERENCES

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." *Journal of Medical Imaging and Radiation Oncology*, 65 (5): 545–563. https://doi.org/10.1111/1754-9485.13261.

Cubuk, E. D., B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le. 2019. "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." In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255.

Espinoza, D., G. Ali, and C. Tarawneh. 2024. "AI-Based Hazard Detection for Railway Crossings." American Society of Mechanical Engineers Digital Collection.

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." 1314–1324.

LeCun, Y., Y. Bengio, and G. Hinton. 2015. "Deep learning." *Nature*, 521 (7553): 436–444. Nature Publishing Group. https://doi.org/10.1038/nature14539.

Mikołajczyk, A., and M. Grochowski. 2018. "Data augmentation for improving deep learning in image classification problem." In: *2018 International Interdisciplinary PhD Workshop (IIPhDW)*, 117–122.

Oh, K., M. Yoo, N. Jin, J. Ko, J. Seo, H. Joo, and M. Ko. 2022. "A Review of Deep Learning Applications for Railway Safety." *Applied Sciences*, 12 (20): 10572. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/app122010572.

Perez, L., and J. Wang. 2017. "The Effectiveness of Data Augmentation in Image Classification using Deep Learning." arXiv.

Torrey, L., and J. Shavlik. 2010. "Transfer Learning." *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*, 242–264. IGI Global Scientific Publishing.

US Department of Transportation (USDOT). 2024. "Highway-Rail Grade Crossing Incidents, Fatalities and Injuries (2.08)." *USDOT Federal Railroad Administration*. Accessed September 8, 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 | FRA." Accessed September 12, 2025. https://web.archive.org/web/20250415074105/https://railroads.dot.gov/crossing-safety-trespass-prevention/crossing-safety-trespass-prevention.