Insights from the U.S. DOT Intersection Safety Challenge Stage 1B

System Assessment and Virtual Testing Primary Track Competition

www.its.dot.gov/isc

Final Report – June 2025
FHWA-JPO-25-157

Source: U.S. DOT

U.S. Department of Transportation

Produced by Noblis, Inc.
U.S. Department of Transportation
Office of the Assistant Secretary for Research and Technology
Intelligent Transportation Systems Joint Program Office

Notice

This document is disseminated under the sponsorship of the Department of Transportation in the interest of information exchange. The United States Government assumes no liability for its contents or use thereof.

The U.S. Government is not endorsing any manufacturers, products, or services cited herein and any trade name that may appear in the work has been included only because it is essential to the contents of the work.

Technical Report Documentation Page

1. Report No. **FHWA-JPO-25-157**	**2. Government Accession No.**	**3. Recipient's Catalog No.**

4. Title and Subtitle

Insights from the U.S. DOT Intersection Safety Challenge Stage 1B:

System Assessment and Virtual Testing Primary Track Competition

5. Report Date

June 2025 (FINAL)

6. Performing Organization Code

7. Author(s)

Stephen Scarano, Claire Silverstein, Anand Seshadri, Haley Townsend,
Peiwei Wang, Karl Wunderlich, Blake Thompson

8. Performing Organization Report No.

9. Performing Organization Name and Address

Noblis, Inc.
500 L'Enfant Plaza, S.W., Suite 900
Washington, D.C. 20024

10. Work Unit No. (TRAIS)

11. Contract or Grant No.

693JJ321D000021

12. Sponsoring Agency Name and Address

Intelligent Transportation Systems Joint Program Office (ITS JPO)
1200 New Jersey Avenue, S.E.
Washington, D.C. 20590

13. Type of Report and Period Covered

Final Report

14. Sponsoring Agency Code

HOIT-1

15. Supplementary Notes

Work Performed for: Norah Ocel (ITS JPO; Task Order Contracting Officer's Representative [TOCOR])

16. Abstract

In response to growing concerns regarding the safety of vulnerable road users at intersections, the U.S. Department of Transportation launched the Intersection Safety Challenge in Spring 2023 with the vision to transform intersection safety through the development of innovative intersection safety systems that can identify, predict, and mitigate unsafe conditions involving vehicles and vulnerable road users in real time. The Challenge began with Stage 1, which had two substages—Stage 1A: Concept Assessment and Stage 1B: System Assessment and Virtual Testing. For Stage 1A, participants submitted design concepts for their proposed systems. The U.S. Department of Transportation received 120 innovative concept papers and selected 15 for prize awards. The 15 winners of Stage 1A were invited to participate in the Stage 1B Primary Track data science competition using multi-sensor data collected at the Turner-Fairbank Highway Research Center Smart Intersection facility.

This report summarizes technical results and key takeaways from the U.S. DOT Intersection Safety Challenge Stage 1B: Primary Track competition. Algorithms submitted by participating teams were assessed on their ability to quickly and accurately perform three key technical elements of intersection safety system operations: (1) the detection, localization, and classification of vulnerable road users and vehicles; (2) path prediction; and (3) conflict prediction among identified vulnerable road users and vehicles. Overall, the Stage 1B results are promising for future prototyping.

17. Keywords

intersection safety, artificial intelligence, sensor fusion, conflict prediction, pedestrian

18. Distribution Statement

19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 54	**22. Price**

Acknowledgements

The authors would like to thank the teams that participated in the U.S. Department of Transportation (U.S. DOT) Intersection Safety Challenge Stage 1B: System Assessment and Virtual Testing prize competition, and congratulations to the ten winners.

Thanks also to the Leidos Saxton Transportation Operations Laboratory (STOL) team for instrumenting the intersection and collecting the data used for Stage 1B.

Table of Contents

List of Tables

List of Figures

Executive Summary

The U.S. Department of Transportation (U.S. DOT) launched the Intersection Safety Challenge ("the Challenge") in Spring 2023. Its goal was to transform intersection safety through the innovative application of emerging technologies and identify and mitigate unsafe conditions involving vehicles and vulnerable road users (VRUs) at intersections. The Challenge kicked off with Stage 1A: Concept Assessment for which U.S. DOT received 120 innovative concept papers on proposed intersection safety systems (ISS) and selected 15 for prize awards. These winners were announced at the 2024 Transportation Research Board (TRB) Annual Meeting and were invited to participate in Stage 1B: System Assessment and Virtual Testing Primary Track Competition (U.S. DOT, 2024a).

In Stage 1B Primary Track, the participating teams tackled a series of technical challenges utilizing U.S. DOT-provided real world sensor data collected on a controlled test intersection at the Federal Highway Administration (FHWA) Turner-Fairbank Highway Research Center (TFHRC). The data was collected from multiple roadway sensors of different types (e.g., cameras, Light Detection and Ranging [LiDAR]) and covered multiple scenarios, including non-conflicts, potential conflicts, and actual collisions between vehicles and VRUs. Please note that surrogate VRUs were used when necessary, and no one was harmed in the data collection process. The Stage 1B training data is now available for download from the U.S. DOT Intelligent Transportation Systems (ITS) DataHub (U.S. DOT, 2024b).[1] This report summarizes technical results and key takeaways from the Stage 1B.

Overall Insights from Stage 1B Primary Track

Stage 1B of the Challenge was a data science competition focused on better understanding the accuracy and reliability of three key technical elements of ISS operations: (1) the detection, localization, and classification of VRUs and vehicles; (2) path prediction of identified VRUs and vehicles; and (3) conflict prediction among identified VRUs and vehicles. **Key insights from the results are summarized below:**

- **Challenge results are encouraging for ISS prototyping.** Many teams were capable of suitably accurate detection, classification, and localization. Path and conflict prediction results were also promising, but the Stage 1B data science competition was limited in its ability to assess these ISS capabilities. Additional testing is needed to understand if these mechanisms can operate efficiently and reliably as part of an ISS in the real world.

[1] To view a sample of the Intersection Safety Challenge Stage 1B data and for instructions to access the full dataset, please visit: https://data.transportation.gov/Roadways-and-Bridges/Intersection-Safety-Challenge-Stage-1B-Sample/vq7s-mv3v/about_data.

- **Teams showed minor improvements with additional time between First and Final Data Submissions.** Teams were required to submit a First Data Submission within 72 hours of the test data release, as well as a Final Data Submission within 6 weeks of the test data release. Teams generally showed minimal to minor improvements to system performance between their First and Final Data Submissions across all metrics. This result suggests that most teams were able to rely on relatively rapid algorithmic analysis of the data rather than extensive manual visual inspection to obtain their results.
- **Performance on road user subclasses differed by technical element, but all teams struggled most in detecting and predicting movement for the surrogate child.** Lower performance on all metrics with respect to the Child road user subclass points to a high-risk need for further research, development, and testing.
- **Road user speed appeared to impact technical performance differently, depending on the metric.** Teams typically performed a little better on path prediction for fast road users compared to slower road users, perhaps since slow movement may indicate or precede a change in movement direction.
- **Nighttime and vehicle left turns point to areas for ISS improvement.** Teams performed better on day and right-turn runs compared to night and left-turn runs respectively across all metrics.
- **Sensor fusion outperformed individual sensor approaches.** Teams using two or more sensor types in their systems clearly performed better than teams relying on a single sensor type.
- **Teams recognized and adapted to data quality challenges.** Although teams identified inherent flaws and unexpected inconsistencies in the data (reflective of potential real-world situations), they generally adapted and used the imperfect data to build a robust ISS.
- **Teams acknowledged the value of the data challenge.** Teams viewed the challenge's premise as a crucial call-to-action and many outlined ambitions to extend their ISS work beyond this initiative.

Intersection Safety Challenge Next Steps

Results from the U.S. DOT Intersection Safety Challenge Stage 1B Primary Track point to potential future activities to address research gaps. These future activities include diving deeper to understand the potential of a near-term, deployed ISS:

- **Conflict Prediction Reliability and Speed.** Stage 1B focused on assessing technical accuracy and rewarded quick responses but was limited in its ability to guarantee that conflict prediction reliability, timing, and speed were sufficient to take mitigating actions. These are important elements to assess in any ISS prototype given this will directly inform conflict mitigation strategies.
- **Conflict Mitigation Strategies and Effectiveness.** Prototyping must assess the practicality of the final key element of an ISS: conflict mitigation, as this was not assessed in the Stage 1B data science challenge. Given ISS response speed, the ISS prototyping stage could explore possible conflict mitigation strategies and their effectiveness at preventing VRU-vehicle conflicts.
- **Cost Effectiveness.** Certain ISS components may be driving overall ISS costs; understanding what they are is crucial to identifying strategies to reduce their costs. Further research is needed to understand if an ISS can be deployed cost-effectively today, and if so, under what conditions and for which intersection geometries.
- **Potential for Broader Deployment.** ISSs offer potential for broader deployment but further prototyping and field testing are needed to assess how they perform in real-world settings with their

specific sensor suites installed at the intersection(s). These ISS prototypes can also shed light on whether narrow or general-purpose ISSs are more practical for near-term deployment.

1. Introduction

Roadway intersection safety is a growing issue, especially for vulnerable road users (VRUs). Each year, roughly one-quarter of traffic fatalities and about one-half of all traffic injuries in the United States are attributed to crashes at intersections (FHWA, 2024). Additionally, according to data from the National Highway Traffic Safety Administration (National Center for Statistics and Analysis, 2024a-c), the number of pedestrians killed in traffic crashes in 2022 (7,522) was the highest since 1981. While pedestrian fatalities saw a slight decrease (3%) in the first half of 2024 compared to the same period in 2023, they remain significantly higher than a decade ago (e.g., 4,779 pedestrian fatalities in 2013).

Improving the safety of pedestrians, bicyclists, and other VRUs is of critical importance to achieving the United States Department of Transportation's (U.S. DOT) objectives of the addressing fatalities and serious injuries across the transportation system. To improve safety at a national scale, cost-effective safety solutions are required that can set the stage for broader, nationwide deployment. U.S. DOT recognizes that technology development and integration is one of many potentially cost-effective approaches for improving safety at intersections, including alternative intersection geometric design and changes to local traffic safety policies.

1.1. Purpose and Organization of the Report

This report summarizes technical results and key takeaways from the U.S. DOT Intersection Safety Challenge Stage 1B: System Assessment and Virtual Testing Primary Track competition.

This report is organized into 4 main sections:

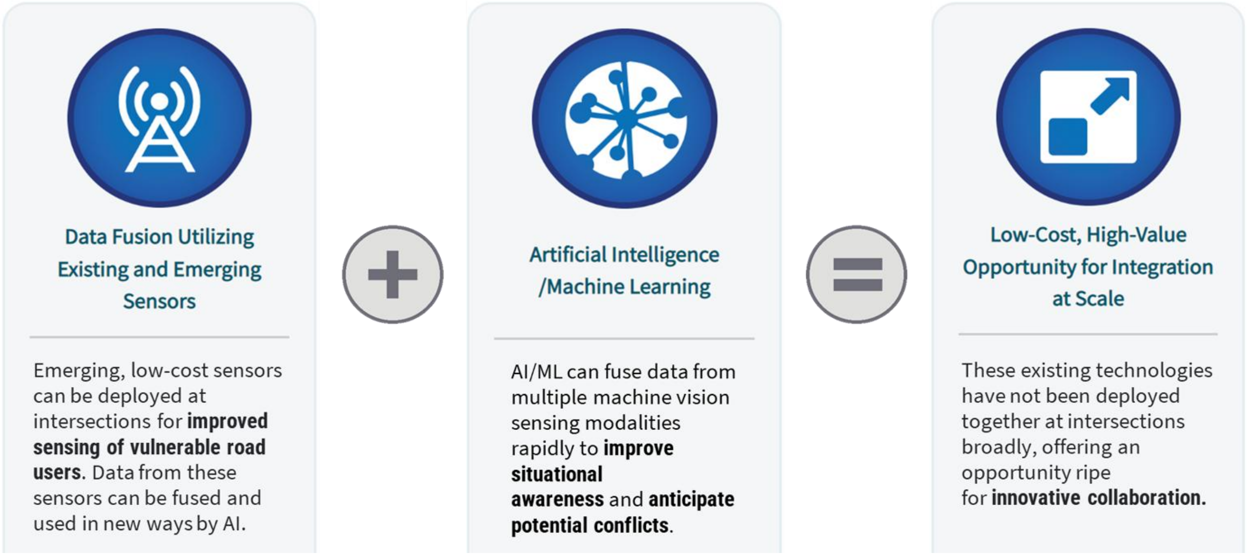
- **Section 1: Introduction.** The introduction includes background information on the U.S. DOT Intersection Safety Challenge, including its vision, structure, and the focus of Stage 1B, as well as an overview of the Stage 1B data that the participating teams used.
- **Section 2: Stage 1B Technical Evaluation Process.** This section discusses the U.S. DOT technical evaluation process for Stage 1B, including the training, testing, and validation data used; submission requirements; and evaluation metrics for the three technical elements assessed in Stage 1B: (1) detection, classification, and localization; (2) path prediction; and (3) conflict prediction between VRUs and vehicles.
- **Section 3: Stage 1B Technical Evaluation Results.** This section summarizes the aggregated technical evaluation results from Stage 1B and is organized by the three technical elements assessed in Stage 1B: (1) detection, classification, and localization; (2) path prediction; and (3) conflict prediction between VRUs and vehicles.
- **Section 4: Key Takeaways.** This section highlights key takeaways for each of the technical elements assessed in Stage 1B — (1) detection, classification, and localization; (2) path prediction; and (3) conflict prediction between VRUs and vehicles — as well as overall key takeaways from Stage 1B that point to gaps to address in potential follow-on prototyping activities.

1.2. Background on the U.S. DOT Intersection Safety Challenge

To better understand the technologies that could enhance intersection safety, the U.S. DOT published the *Enhancing the Safety of Vulnerable Road Users at Intersections; Request for Information (RFI)* in the Federal Register on September 16, 2022 (U.S. DOT, 2022). The RFI sought information on a conceptual VRU and vehicle warning system building on existing and emerging vehicle automation technologies—including machine vision, perception, sensor fusion, real-time decision-making, artificial intelligence (AI), and vehicle-to-everything (V2X) communications. A summary of responses and insights from the RFI can be found in the *Summary Report for RFI: Enhancing the Safety of Vulnerable Road Users at Intersections* (Townsend et al., 2023). Insights from the RFI helped to inform U.S. DOT on the status of technologies that can be used to improve safety at roadway intersections and laid the foundation for the Intersection Safety Challenge (hereafter "the Challenge").

1.2.1. Challenge Vision

The U.S. DOT Intersection Safety Challenge, which publicly launched in April 2023, aims to *transform intersection safety through the innovative application of emerging technologies including machine vision, sensor fusion, and real-time decision-making to identify and mitigate unsafe conditions involving vehicles and vulnerable road users*. **Figure 1** illustrates at a high level the proposed solution as the Challenge envisions it of leveraging emerging technologies to improve intersection safety at scale in a new way. Specifically, data fusion mixed with AI and machine learning (ML) points to a potentially low-cost, high-value opportunity for integration at scale.



Data Fusion Utilizing Existing and Emerging Sensors

Emerging, low-cost sensors can be deployed at intersections for **improved sensing of vulnerable road users.** Data from these sensors can be fused and used in new ways by AI.

Artificial Intelligence /Machine Learning

AI/ML can fuse data from multiple machine vision sensing modalities rapidly to **improve situational awareness** and **anticipate potential conflicts**.

Low-Cost, High-Value Opportunity for Integration at Scale

These existing technologies have not been deployed together at intersections broadly, offering an opportunity ripe for **innovative collaboration.**

Figure 1. Vision of the Challenge to Leverage Emerging Technologies to Improve Intersection Safety. Image Source: U.S. DOT

The Challenge seeks broad innovation in real-time roadway intersection safety systems (ISS) featuring anticipatory warning systems and other safety-countermeasures for both drivers and VRUs, including those without connectivity. Specifically, an ISS is anticipated to:

- Deploy existing and emerging, low-cost sensors (e.g., cameras, radar, Light Detection and Ranging [LiDAR], infrared) at intersections to improve sensing,
- Use multi-sensor data fusion/analytics to improve situational awareness and anticipate safety threats, and
- Issue warnings and/or modify control settings to improve safety.

Such ISS must be cost-effective to expedite their deployment at scale at the highest risk intersections throughout the nation. **Figure 2** shows a concept illustration of an ISS with the red dotted lines illustrating identified hazards and the yellow dotted lines illustrating warnings and communications to road users.

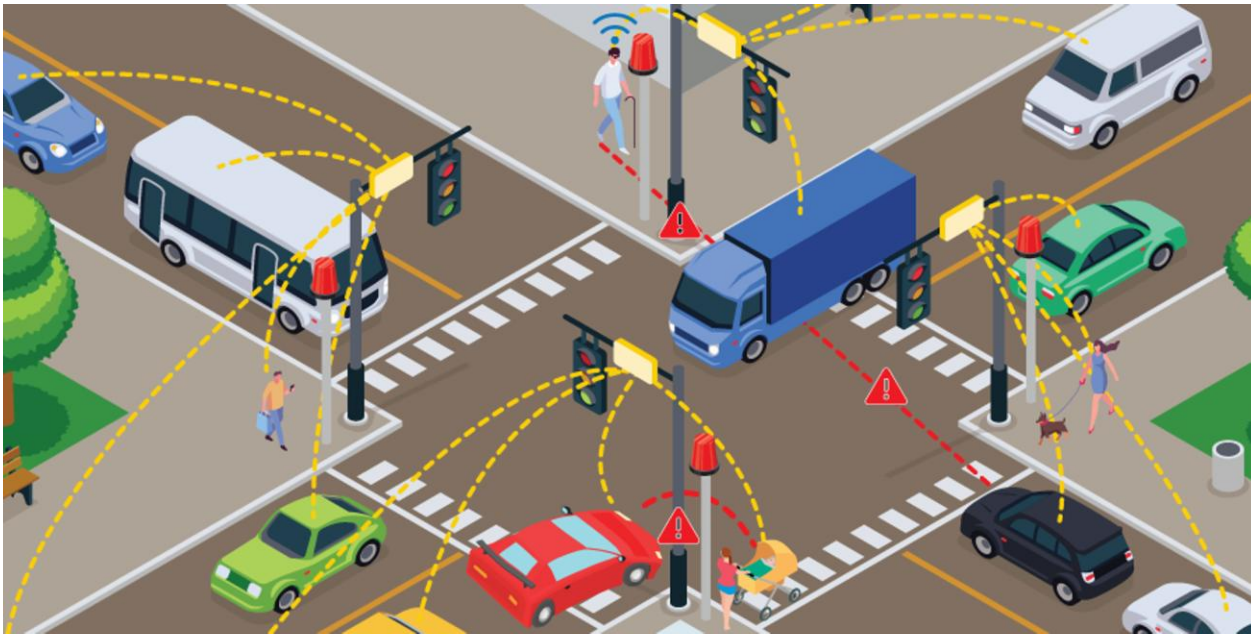


Figure 2. Concept Illustration of an Intersection Safety System (ISS). Image Source: U.S. DOT

Teams participating in the Challenge were asked to consider a range of relevant technologies, both existing and emerging, as part of their ISS concept aligned with the vision of the Challenge.

1.2.2. Challenge Structure

Overall Challenge Structure. The Challenge was organized into two sequential stages comprising the overall Prize Competition. Winners of Stage 1A: Concept Assessment were invited to participate in Stage 1B: System Assessment and Virtual Testing. This Challenge structure is illustrated in **Figure 3**. The overall Intersection Safety Challenge could set the foundation for future potential ISS prototyping.

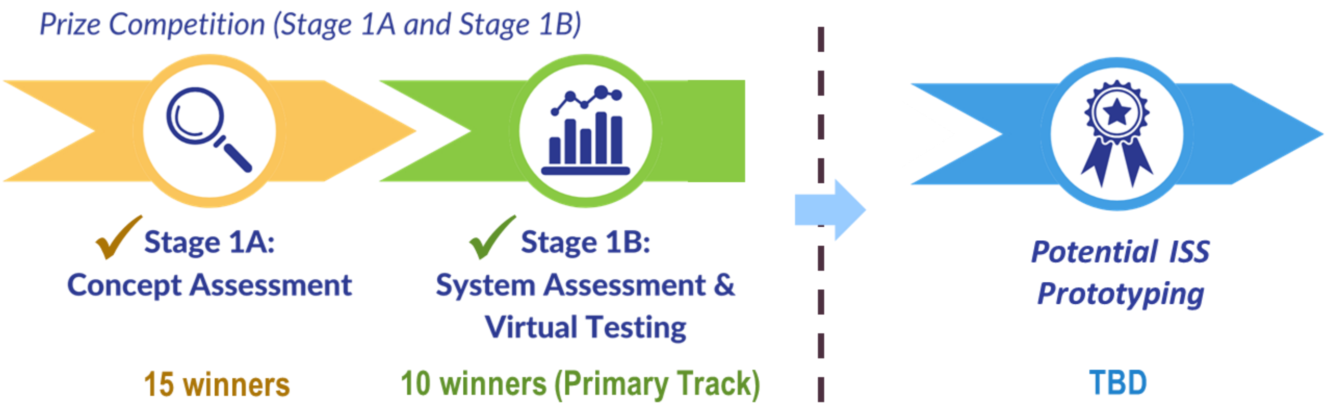

Prize Competition (Stage 1A and Stage 1B)
✓ Stage 1A:
Concept Assessment
15 winners
✓ Stage 1B:
System Assessment & Virtual Testing
10 winners (Primary Track)
Potential ISS Prototyping
TBD

Figure 3. Overview of Challenge Structure. Image Source: U.S. DOT

Stage 1A. In the first part of the Prize Competition (Stage 1A), each participating team submitted a Concept Paper of no more than 15 pages describing their ISS concept and the potential of this concept to address the vision and objectives of the Challenge. U.S. DOT announced the winners of Stage 1A during the Transportation Research Board (TRB) Annual Meeting on January 8, 2024 (U.S. DOT, 2024a). U.S. DOT received 120 innovative concept papers and selected 15 for prize awards. Each of the 15 winning teams in Stage 1A received a prize of $100,000 and an invitation to participate in the Primary Track of Stage 1B: System Assessment and Virtual Testing.

Stage 1B. During the second part of the Prize Competition (Stage 1B), participating teams tackled a series of technical challenges utilizing U.S. DOT-provided real world sensor data collected on a closed course at the Federal Highway Administration (FHWA) Turner-Fairbank Highway Research Center (TFHRC). Stage 1B is intended to provide further insight regarding elements of the ISS vision and to inform potential follow-on research activities related to the ISS concept. Stage 1B may include multiple Prize Competition “tracks” in which different entities or individuals may be eligible to participate. Stage 1B began with the Primary Track Prize Competition, in which only teams led by lead entities identified as Stage 1A winners were eligible to participate. U.S. DOT awarded 10 teams prize amounts ranging from $166,666 to $750,000 for a total of $4,000,000 in prize awards for the Stage 1B Primary Track (U.S. DOT, 2025).

Potential Future ISS Prototyping. Informed by the results of Stage 1B, U.S. DOT may choose to conduct future activities focused on ISS prototyping and physical testing. **Figure 4** depicts an ISS concept diagram and which components each stage of the Challenge focuses on.

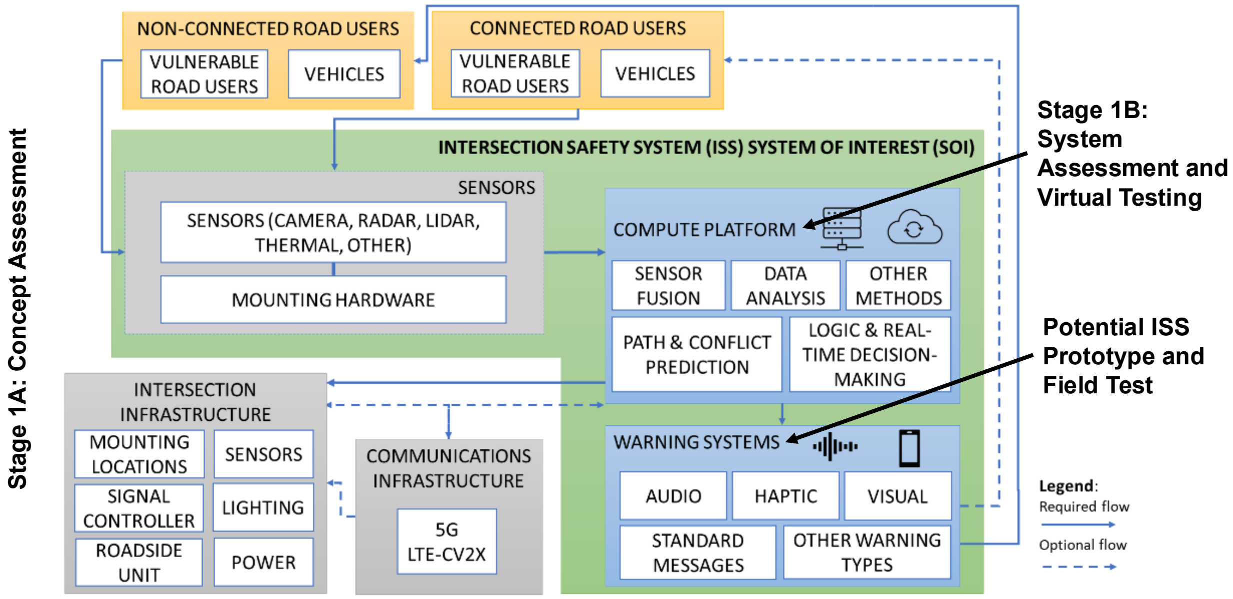


Figure 4. Focus of Each Stage of the Challenge on the ISS Concept Diagram. Image Source: U.S. DOT. *Acronyms: Long Term Evolution Cellular Vehicle-to-Everything (LTE-CV2X).*

Stage 1A: Concept Assessment of the Challenge focused on understanding teams' entire ISS concepts, including sensors, the compute platform (including sensor fusion and AI models), warning systems, and any additional infrastructure. Stage 1B: System Assessment and Virtual Testing focused on assessing the ISS compute and decision-making platforms against U.S. DOT-supplied sensor data. The potential follow-on prototyping stage is expected to focus on the entire end-to-end ISS prototype and its ability to mitigate potential conflicts with warnings.

1.2.3. Focus of Stage 1B

The Stage 1B Primary Track focused on three key technical elements of ISS operations (**Figure 5**): (1) the detection, localization, and classification of VRUs and vehicles; (2) path prediction; and (3) conflict prediction among identified VRUs and vehicles. Conflict mitigation, another critical element of the overall ISS concept, was not addressed in Stage 1B but may be a focus of potential future activities on ISS prototyping.

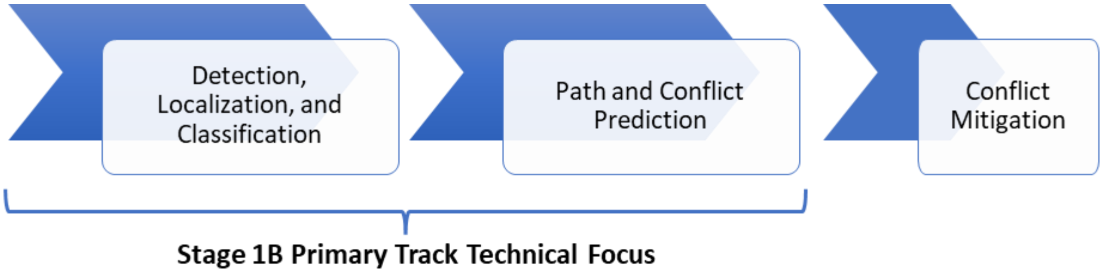

Detection, Localization, and Classification
Path and Conflict Prediction
Conflict Mitigation
Stage 1B Primary Track Technical Focus

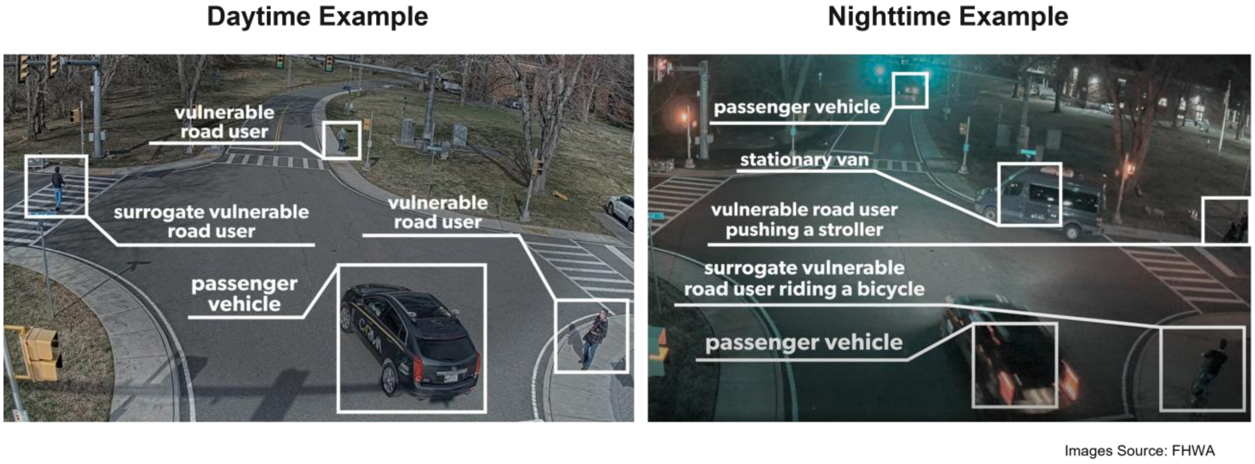
Figure 5. Stage 1B Primary Track Technical Focus of ISS Operations. Image Source: U.S. DOT.

The Stage 1B teams developed, trained, and improved their algorithms for the detection, localization, and classification of VRUs and vehicles using U.S. DOT-supplied sensor data collected at a controlled test roadway intersection. Further, teams used these data and algorithms to predict future intersection conditions, including the prediction of potential conflicts between VRUs and vehicles in and around the intersection.

1.3. Overview of Stage 1B Data

Data for Stage 1B was collected at the FHWA TFHRC Smart Intersection facility. The data includes non-conflicts, potential conflicts, and actual collisions between vehicles and VRUs. Potential conflict refers to scenarios that intentionally created a risk of or actual occurrence of two road users (i.e., VRUs and vehicles) occupying the same location on the intersection facility at the same time (e.g., a collision). Surrogate VRU systems—programmable motorized mannequin systems that mimic real human actors—were used to ensure safety. Please note that no one was harmed during the data collection process.

Various scenarios were orchestrated in a controlled test bed. Different props—such as a wheelchair, stroller, walker, bike, and electric scooter—were used in data collection to ensure a variety of VRUs were represented in the data. Additional variability in the data collection scenarios was introduced based on road user speeds, time of day and lighting conditions, and other factors. **Figure 6** shows two example data runs with different road user types, props, test scenarios, and times of day from the perspective of a single closed-circuit television (CCTV) camera.



Images Source: FHWA

Figure 6. Day and Night Examples of Collected Data from the Perspective of a Single CCTV Camera. Image Source: FHWA.

1.3.1. Sensor Types

The Challenge Dataset includes data from 20 roadside sensors and traffic control devices (see **Table 1**). Data from the roadside sensors is intended to serve as input to AI/ML algorithms for the detection, classification, localization, and trajectory/conflict prediction of the moving road users. Roadside sensors at the TFHRC West Intersection included eight CCTV visual cameras, five thermal cameras, two LiDAR sensors, and four radar sensors. Intrinsic calibration was performed for all visual and thermal cameras.

Extrinsic calibration was performed for specific pairs of roadside sensors. Additionally, the traffic signal phase and timing data and vehicle and/or pedestrian calls to the traffic signal controller (if any) were also provided. Data collection from all sensors was coordinated using Robot Operating System (ROS) to start and stop recording from multiple sensors simultaneously, while applying timestamps from each device's local clock, which were synced to a local Network Time Protocol (NTP) server.

Teams that participated in Stage 1B: System Assessment and Virtual Testing of the Intersection Safety Challenge were invited to use any combination of sensors that they believed would achieve the best results.

Table 1. General Information on Sensors, Controllers, and Associated Data.

Sensor Type	Number of Sensors	Type of Data Collected
Traffic Signal Controller	1	SAE J2735 signal phase and timing (SPaT) data (in .pcap format), and vehicle and pedestrian call data (in .csv format) from traffic signal controller.
Visual Camera	8	Video data (in .mp4 format) covering the entire intersection area.
Thermal Camera	5	Thermal video data (in .mp4 format) covering the intersection area, with a focus on crosswalks.
LiDAR	2	Raw LiDAR data (in .pcap format) covering the main intersection and parts of the approaches.
Radar	4	Object data (in .JSON format) recorded from Message Queue Telemetry Transport (MQTT) stream.

The **Traffic Signal Controller (TSC)** recorded traffic signal phasing and timing information, as well as vehicle and pedestrian calls. The TSC was located at the West Intersection of the TFHRC. This controller adheres to Advanced Transportation Controller (ATC) 5.2b and National Transportation Communications for Intelligent Transportation Systems Protocol (NTCIP) standards.

1.3.2. Road User Types and Subclasses

This dataset included a wide range of road user types to capture those that are likely to be encountered at intersections.

Surrogate VRUs are mannequins with varying extremity movements and props sitting on top of a small platform/sled connected to a belt, which was pulled back and forth by a motor unit. The surrogate VRUs fall into three distinct categories: adult pedestrian, child pedestrian, and adult bicyclist (see **Figure 7**).

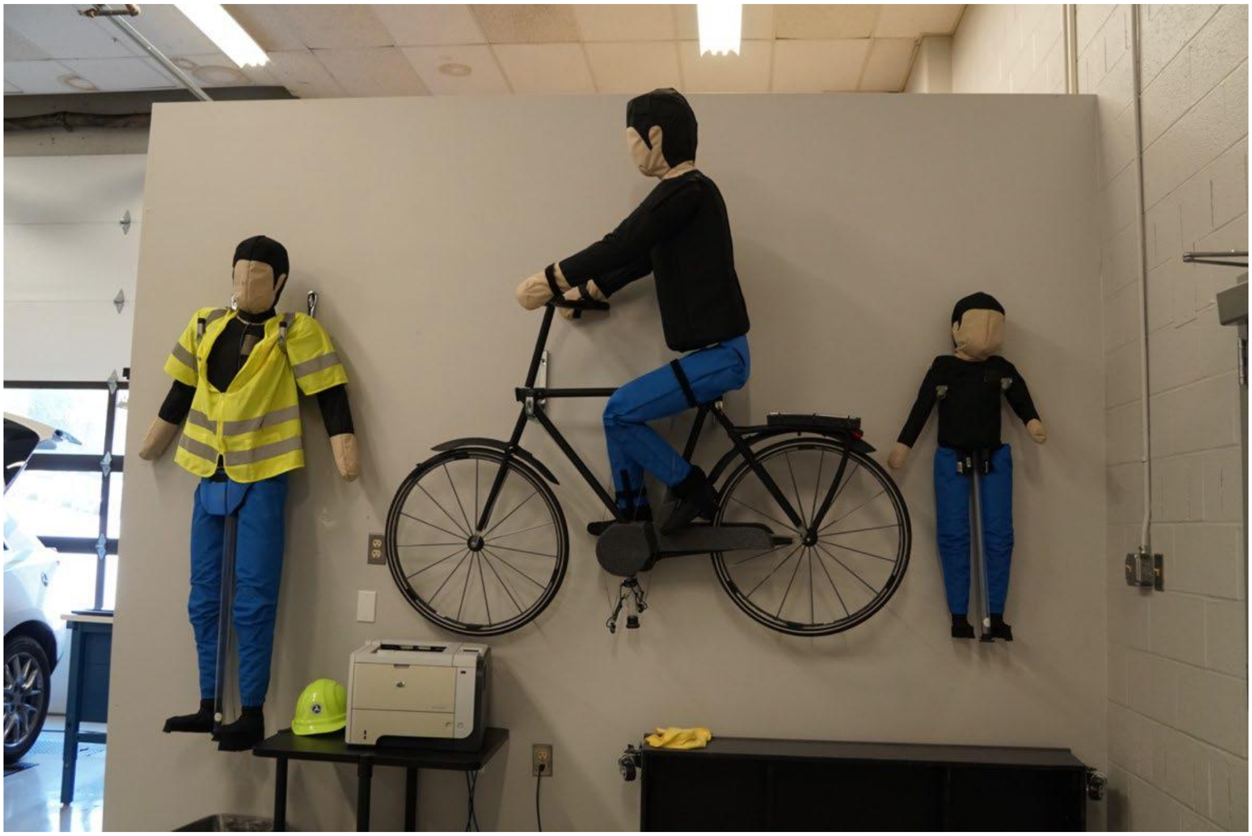


Figure 7. Surrogate Adult (left), Child (right), and Adult Bicyclist (middle). Image Source: FHWA.

Non-Surrogate VRUs are comprised of real human adult actors who utilized a variety of mobility-enhancing and mobility-impeding props. Note that these non-surrogate VRUs were never put at risk of being involved in a potential conflict or actual collision. To encompass as many types of VRUs potentially encountered on intersections as possible, more than 10 different props were used during data collection. These props were selected to allow actors to travel at varying speeds across the intersection and to ensure that a wide variety of VRUs were represented in the data, including VRUs with varying mobility enhancements and limitations.

Props could be reused non-consecutively and could be swapped between actors. Props included items that pedestrians could be holding or using such as a cane, stroller, walker, crutches, or a large box. A manual and an electric scooter were used to represent the growing use of e-scooters and other micromobility devices. A manual pedal bike and an electric bike were used as well. Two wheelchairs, one electric and one manual, were also used. These example props are shown in **Table 2**, which illustrates the broad range of road user subclasses included within this dataset:

Table 2. Road User Subclasses.

Road User Subclass Name	Description	Road User Subclass Abbreviated Name Used Throughout This Report
Passenger Vehicle	Includes the grey and black Sport Utility Vehicles (SUVs) described in the metadata document (U.S. DOT, 2024b) [2]	Vehicle
Vehicle Other	Includes any other moving vehicle not described in the Passenger Vehicle subclass	N/A
VRU Child	Child-sized surrogate[3] VRU	Child
VRU Adult	Includes the surrogate VRU adult and real actor VRU adult without mobility enabling or impeding device/apparatus/prop	Adult
VRU Adult Using Wheelchair	Includes a VRU adult in a manual or motorized wheelchair	Adult Wheelchair User
VRU Adult Using Bicycle	Includes the surrogate VRU adult riding a bicycle or a real actor VRU adult riding a motorized/electric bicycle or manual bicycle	Adult Bicyclist
VRU Adult Using Non-Motorized Device/Prop Other	Includes a VRU adult walking with a cane, pushing a walker, walking with crutches, pushing a wheelchair, pushing a stroller, carrying a large cardboard box, or carrying an umbrella	Adult Non-Motorized Device User
VRU Adult Using Scooter or Skateboard	Includes a VRU adult riding an electric scooter, manual scooter, or a skateboard	Adult Scooter/Skateboard User
VRU Other	Any other VRU detected that does not fall within the above subclasses	N/A

[2] The metadata document can be accessed at: https://datahub.transportation.gov/Roadways-and-Bridges/Intersection-Safety-Challenge-Stage-1B-Sample/vq7s-mv3v/about_data

[3] The Surrogate Child VRU had an approximate height of 1.15 m (3 ft 9 in). For comparison, the Surrogate Adult VRU had an approximate height of 1.8 m (5 ft 10 in).

1.3.3. Data Collection Experimental Scenario Features

The data collection efforts were conducted for various potential conflict-based experimental scenarios and non-conflict-based experimental scenarios. These scenarios were determined and refined by a group of subject matter experts (SMEs) from U.S. DOT and their contract support teams.

Each data collection experimental scenario included a range of experimental conditions, where certain factors varied within a set of desired levels/values, such as vehicle speeds, VRU types and speeds, VRU props, and time of day light conditions. Moreover, additional variability was introduced in the data collection execution for both potential conflict- and non-conflict-based scenarios to improve the alignment with real world-equivalent intersection scenarios as much as possible. These variabilities included, but were not limited to, crosswalks used by the surrogate VRU, intended conflict directness, driving style of vehicle operators, and starting location of VRUs. Additionally, in both potential conflict- and non-conflict-based scenarios, data collection team members (who acted as real VRUs) varied across the data collection efforts. Variations in real actor VRU physical traits were considered in choosing team members to act as real VRUs to help increase real-world applicability of the collected data.

High-level descriptions of the experimental scenarios executed at the TFHRC West Intersection include the following example vehicular, surrogate VRU, and VRU movements/behaviors (note that these are not exhaustive):

- Various types of vehicle left turns, originating from multiple approaches of the intersection
- Various types of vehicle right turns, originating from multiple approaches of the intersection
- Vehicles traveling straight through the intersection, originating from multiple approaches of the intersection
- Various intersection crossing movements of the surrogate VRUs, originating from multiple approaches of the intersection
- Various intersection crossing movements of the real actor VRUs, originating from multiple approaches of the intersection

Additionally, stationary objects were placed at various locations on the intersection to obstruct the view of some of the sensors during some data collection experimental scenarios so various VRUs and/or vehicles would be blocked from these sensors' fields of view, creating occlusion.

As mentioned above, the details of the conditions executed for each scenario, as well as additional variabilities introduced, were not provided to the participants as they were considered part of the Stage 1B challenge. Sensor data was organized by each "run," which was defined as the roughly one to two minutes during which various scripted and non-scripted vehicular and VRU actions/movements/behaviors occurred on/adjacent to the described intersection.

2. Stage 1B Technical Evaluation Process

Two judging criteria were used by U.S. DOT in the Stage 1B Primary Track: (1) Technical Merit and (2) Deployment Suitability and Alignment with Challenge Vision. Technical Merit was more important than Deployment Suitability and Alignment with Challenge Vision in the overall evaluation process. This report summarizes the aggregated results from the Technical Merit evaluation conducted by U.S. DOT. Note that all information below was provided to participating teams via a detailed evaluation plan. Technical Merit was based on scoring of team-generated structured data files returned in the First Data Submission and the Final Data Submission.

- *First Data Submission:* The First Data Submission was due on September 20, 2024, three days after the release of the test data (released on September 17, 2024). This submission was intended to test both the accuracy and response speed of the team's algorithmic approach without an extensive, potentially manual examination of the test dataset.
- *Final Data Submission:* The Final Data Submission was due on October 29, 2024, six weeks after the release of the test data. This submission was intended solely to test the accuracy of the team's approach when informed by more extensive consideration of the test dataset and potential enhancements and tailoring. Teams could choose to submit the same results from their First Data Submission for their Final Data Submission.

2.1. Data Splitting and Cutting for Evaluation

For purposes of the Intersection Safety Challenge Stage 1B Primary Track, the data was split into three groups: training, validation, and test data. Please note that these dataset group names were used for purposes of the Stage 1B Primary Track and may or may not reflect how these terms are used in other circumstances.

- **Training data** refers to data that was not labeled (i.e., unlabeled, without ground truth). The training data included only the sensor output data and no additional information about the ground truth. This data was provided to participants for architecting and training their systems. Training data was released to teams incrementally over a roughly two-month period in June and July of 2024.
- **Validation data** refers to data that was accompanied with corresponding labels (i.e., labeled, with ground truth). The validation data runs had similar types of sensor information as the training data runs but also with accompanying ground truth labels in the form of road users' 3D bounding boxes, classes, and subclasses annotated at relevant timestamps for each run. Furthermore, binary labels regarding conflict versus no conflict for each of the validation data runs were provided. This data was provided to participants for architecting and training their systems as well as for ensuring their systems produced the proper ground truth formatting expected for the test data. Validation data was released to teams incrementally over a roughly two-month period in June and July of 2024.
- **Test data** refers to data that was labeled, but these labels were not provided to the participating teams so that U.S. DOT could score the teams' ability to determine the labels accurately. Since the test data was used by U.S. DOT to evaluate the participants' algorithm(s), it was held out from the other data, meaning the participants were not able to see these test data runs in the training and validation data.

The test data included additional runs with similar information as the training and validation data. However, unlike the training and validation data, the test data was cut off at certain timestamps to exclude the potential conflict period. Therefore, the test data released to the teams reflected potential pre-conflict conditions at the intersection only. Teams were asked to predict vehicle and VRU paths and potential conflicts after the cutoff point within the provided test data. The test data was released to the participating teams on September 17, 2024.

The train_test_split model from the scikit-learn Python package was leveraged to split the 1,318 total data runs into the training dataset, the validation dataset, and the test dataset. Approximately 3 percent of the runs made up the validation set, 15 percent of the runs made up the test set, and the remaining 82 percent of the runs made up the training set. The training data and the test data were intended to look similar, but there was no guarantee that the distribution of potential VRUs, situations, conditions, etc. was perfectly balanced across the two sets.

2.2. Detection, Classification, and Localization Process

The first technical element that the participants' systems performed was detection, classification, and localization of moving road users of interest. Detection refers to the system understanding and indicating that a road user (vehicle or VRU) is within the evaluation area. Classification refers to the system being able to identify what class and subclass the road user belongs to. Finally, localization refers to the system being able to situate where in 3D space the road user is located, with respect to the ground truth frame of reference. Programmatically, these properties of the road users are represented by annotated 3D bounding boxes with fields: (subclass, x_center, x_length, y_center, y_length, z_center, z_length, yaw). **Figure 8** shows a visualization of an example ground truth bounding box on the frame of reference. Yaw, with a range of [0, 360) degrees, is measured counter-clockwise such that a detected object (road user) has a yaw of 0 degrees when it faces the positive Y direction and a yaw of 90 degrees when it faces the negative X direction.

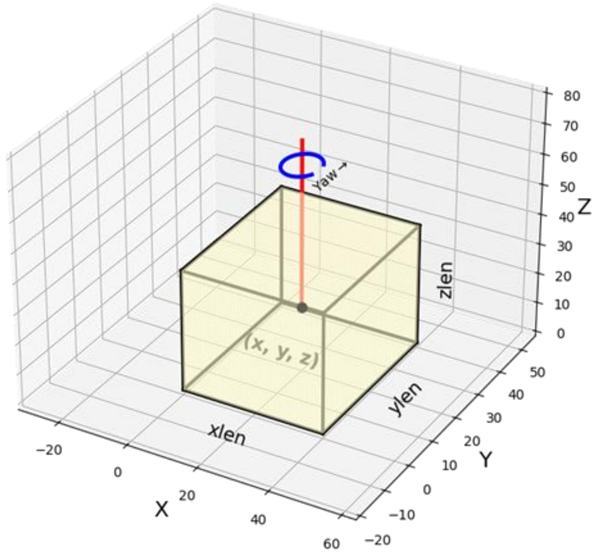


Figure 8. Ground Truth Bounding Box Example (Axes in meters). Image Source: U.S. DOT.

2.2.1. Submission Format for Detection, Classification, and Localization

Participants were expected to submit a .csv file for each run within the test data that included the following indices and columns:

- Indices: Timestamps
- Columns: [subclass, x_center, x_length, y_center, y_length, z_center, z_length, z_rotation]

The columns labeled "center" and "length" refer to the specific geometric center coordinate of the submission bounding box and length of the submission bounding box, respectively for each road user detected. Z_rotation (or yaw) refers to the rotation of the box around the z axis. Whenever a road user was detected within the evaluation region, it should have had an associated row in the submission data at the timestamp. If multiple road users were present in a given timestamp, multiple rows for the given timestamp should have been provided in the submission file.

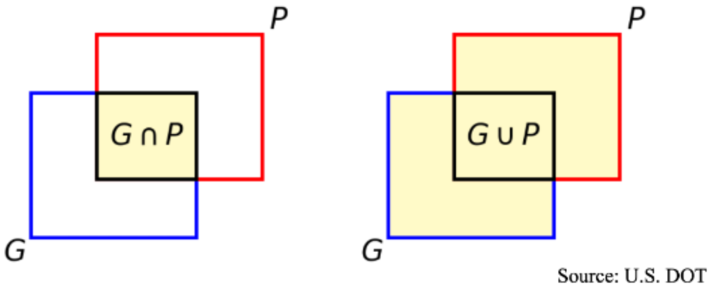
Additionally, participants were requested to use the provided timestamps for each test data run as well as the subclass names from **Table 2** in their submissions, or their final submissions would not be evaluated for possible award.

2.2.2. Evaluation Metrics for Detection, Classification, and Localization

Detection, classification, and localization were collectively evaluated using the mean average precision (mAP) at different intersection over union (IoU) thresholds. The IoU of a ground truth bounding box ***G*** and prediction bounding box ***P*** is calculated by:

$$IoU(G, P) = \frac{G \cap P}{G \cup P}$$

Example intersections and unions are shown in **Figure 9**.



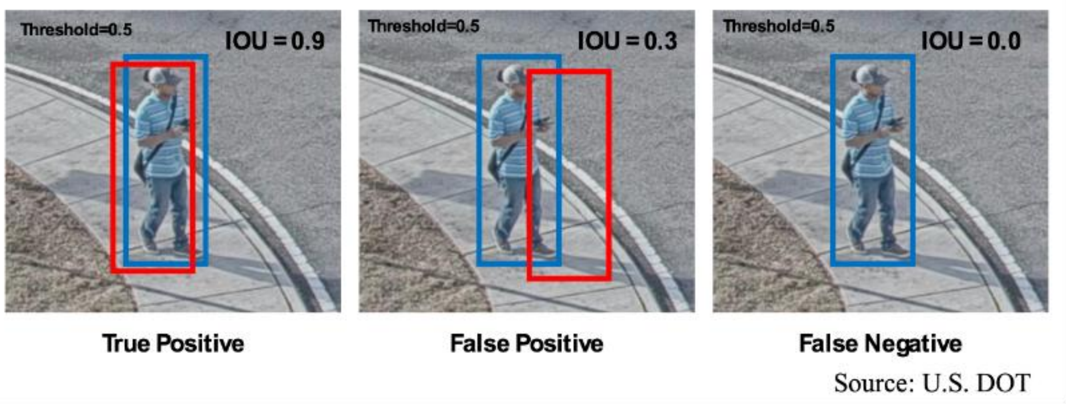
Source: U.S. DOT

Figure 9. 2D Intersection (left) and Union (right). Image Source: U.S. DOT.

Calculation of 3D IoU conformed to the following pseudocode algorithm:

1. Project both bounding boxes (ground truth and participant algorithm output predictions) into the *xy* plane.
2. Calculate the intersection points along the edges of the two projected boxes.
3. Use the intersection points to calculate the area of intersection in the *xy* plane.
4. Multiply the area by the intersection of the heights along the *z*-axis to compute the intersection volume.
5. If the intersection volume is greater than 0, then calculate the volume of each of the two bounding boxes; the union volume is the sum of the two bounding boxes' volumes minus the intersection volume.
6. Calculate IoU by taking the quotient of the intersection volume divided by the union volume.

Figure 10 shows three different example scenarios. The blue bounding box represents the ground truth while the red bounding box represents an example participant submission. For the sake of visual clarity, bounding boxes are shown as 2D in this example; however, as stated before, the actual bounding boxes were 3D bounding boxes.



Source: U.S. DOT

Figure 10. Example (2D) Predictions and Label Outcomes. Image Source: U.S. DOT.

The prediction calculation was converted to a confusion matrix prediction outcome by calculating IoU with respect to the ground truth and comparing it to a threshold. The relevant metrics calculated depended on three components of the confusion matrix: true positive (TP), false positive (FP), and false negative (FN). They are defined as:

- TP: IoU > threshold for the class
- FP: IoU < threshold for the class > 0
- FN: IoU = 0 when the object exists

Average Precision (AP), interpreted as the estimated area under the precision-recall curve (i.e., AUC-PR) was calculated for each class at a certain threshold α:

1. Report every detection of a VRU across all timestamps in each run along with associated confidence scores.
2. Sort the table of VRU detections in order of decreasing confidence.
3. Tabulate cumulative TP and FP going down the table.
4. Calculate row-wise precision and recall. The precision (P) at threshold *n* is computed as $P_n = \frac{TP_n}{TP_n+FP_n}$. The recall (R) at threshold *n* is computed as $R_n = \frac{TP_n}{TP_n+FN_n}$.
5. Plot precision-recall graph.
6. Calculate AP using 40-point interpolation method used in the Karlsruhe Institute of Technology and Toyota Technological Institute at Chicago Vision Benchmark (KITTI) (Zhang et al., 2022).

The range of thresholds α used by the COCO (Common Objects in Context), which are 10 thresholds ranging from 0.50 to 0.95 in steps of 0.05, were used here as well.

$$\alpha = \{0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95\}$$

To calculate the mAP, the evaluators averaged across the ten threshold values and across each class. C refers to the number of classes and c refers to a specific class:

$$mAP = \frac{1}{C}\frac{1}{10}\Sigma_{\mathrm{c}}\Sigma_{\mathrm{a}\in\alpha}\, AP_{c,a}$$

The ground truth labels contained both classes and subclasses (class was derived via subclass). Therefore, the evaluation took into consideration correct identification of *both class and subclass*, weighting both equally. If a subclass was correctly classified and localized, the class was also, awarding full points. If the submission correctly identified the class via the subclass (e.g., VRU) but incorrectly identified the subclass (e.g., Adult Using Bicycle instead of Adult Using Wheelchair), then half of the score was given. Formally,

$$Score_{overall} = \frac{1}{2}mAP_{class} + \frac{1}{2}mAP_{subclass}$$

2.3. Path Prediction Process

The second technical element of the evaluation was path prediction. Participants were given the test data (note that not all sensors' data was provided and/or usable for each test run) starting at the beginning (t_0) of a run until the cutoff timestamp (t_{cut}) before the potential conflict so they could understand the scene's context (e.g., road users present, their locations, paths, and behaviors). After the cutoff timestamp (t_{cut}), the sensor data was no longer available to participants so as not to give away the true paths and true conflict conditions. Participants were expected to provide path predictions for each road user up to time *T*, which was approximately 5.0 seconds after t_{cut}. Participants were able to provide *up to but no more than* three predicted paths per road user with corresponding confidence weights w_i for each. The confidence weights had to be positive $w_i > 0, \forall i$ and for each road user had to sum to one (i.e., $\sum_i w_i = 1$).

Figure 11 visualizes the setup of a path prediction scenario from the participants' point of view. This visualization represents one road user, although there were often multiple road users within a run. Although this example visualization is not in 3D, the actual paths themselves were in 3D with participants expected to provide the predicted future locations of the geometric centers (x, y, z) of the bounding boxes representing road users. In the example, the participant provided three predicted paths for a single road user as well as corresponding confidence weights of 0.1, 0.6, and 0.3.

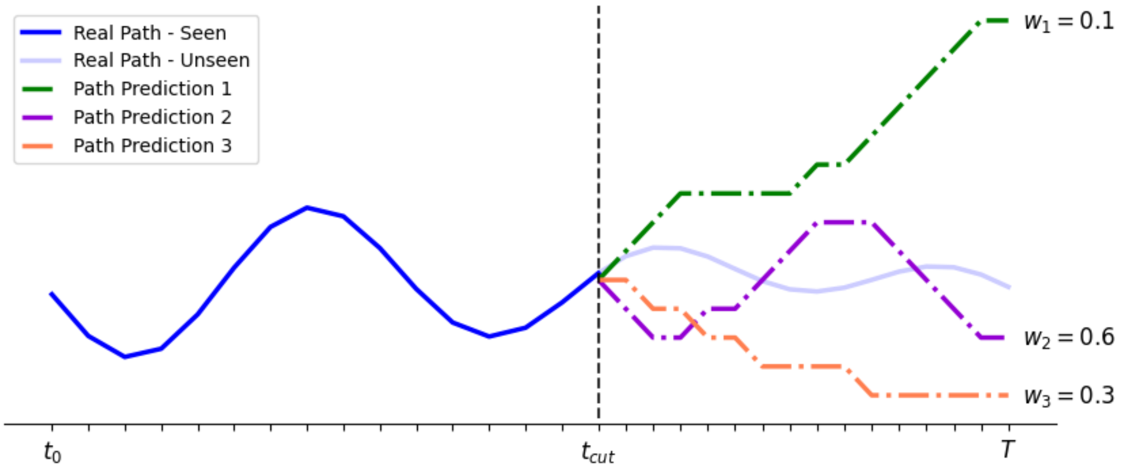


Figure 11. Example Path Prediction Task for One Road User. Image Source: U.S. DOT.

2.3.1. Submission Format for Path Prediction

Participants were expected to provide up to three predicted paths for road users after the cutoff time (t_{cut}). The column path_ID was used to group the locations of individual road users together to form their paths. Each path_ID must have had an associated confidence_score, a floating point value between 0.0 and 1.0, based on the likelihood of that specific road user continuing on the predicted path. Note that the confidence scores for the up to three predicted paths for each road user had to sum to 1.0 within machine precision.

Additionally, participants were requested to use the provided timestamps for each test data run as well as the subclass names from **Table 2** in their submissions, or their final submissions would not be evaluated for possible award.

Participants were expected to submit a .csv file for each run within the test data that included the following indices and columns:

- Indices: Timestamps
- Columns: [path_ID, subclass, x_center, y_center, z_center, x_length, y_length, z_length, confidence_score]

2.3.2. Evaluation Metrics for Path Prediction

The metric used to evaluate submissions for path prediction was a modified version of *Average Displacement Error (ADE)*. ADE can be thought of as the Euclidean or L2 distance between the prediction and the ground truth, averaged over all timestamps. For an entire run, the path prediction score was calculated as the weighted ADE score (calculated for class and for subclass).

Weighted ADE Score =

$$\Sigma_c I_c(c)^* \frac{1}{\text{T-t}_{\text{cut}}} \Sigma_{t=t_{cut}+1}^{T} \sqrt{(x_{c,t}\text{-}\Sigma_i(w_{c,i}{}^* \hat{x}_{c,i,t})^2 + (y_{c,t}\text{-}\Sigma_i(w_{c,i}{}^* \hat{y}_{c,i,t}))^2 + (z_{c,t}\text{-}\Sigma_i(w_{c,i}{}^* \hat{z}_{c,i,t}))^2}$$

- $x_{c,t}$, $y_{c,t}$, $z_{c,t}$ are the true coordinates of the center of the bounding box for road user class (or subclass) *c* at timestamp *t*
- $\hat{x}_{c,i,t}$, $\hat{y}_{c,i,t}$, and $\hat{z}_{c,i,t}$ are the predicted coordinates of the center of the bounding box for road user *c* given by predicted path *i* at timestamp *t*
- $I_c(c)$ is an indicator function that equals one if the correct class of road user is identified and 0 if not.

As with detection, classification, and localization, each of the path prediction submissions was weighted half based on class and half based on subclass.

$$Score_{overall} = \frac{1}{2} ADE_{class} + \frac{1}{2} ADE_{subclass}$$

2.4. Conflict Prediction Process

The third and final technical element of the evaluation was conflict prediction. Participants were expected to algorithmically predict whether a conflict would occur between two road users of interest in the specified test data runs after the cutoff timestamp (t_{cut}) as a binary variable. If yes, they were expected to include the timestamp at which they thought the conflict would occur and which road user subclasses would be involved in the conflict.

2.4.1. Submission Format for Conflict Prediction

Participants were expected to provide a single .csv file with one row for each test data run. For each data run, participants had to provide the binary classification of conflict or no conflict between two road users (based on the time-to-collision [TTC] for when two road users were identified as having a path intersection point). In addition to denoting this binary classification when submitting their results for the specified test data, the following attributes were expected to be included in the above-mentioned .csv file for runs in the test data for which a conflict was identified:

- The timestamp of the conflict
- Subclass of road user 1 involved in the conflict
- Subclass of road user 2 involved in the conflict

The CSV submitted was expected to include the following indices and columns:

- Indices: Run ID
- Columns: [Conflict_No_Conflict_Label, timestamp_conflict, road_user1_subclass, road_user2_subclass]

2.4.2. Evaluation Metrics for Conflict Prediction

Scoring for conflict prediction was based primarily on F2 score, with a small percent (10 percent) predicated on the incorporation of the timestamp and road user subclasses involved in a predicted conflict. The F2 portion of a participant team's score for this evaluation component was based on the following:

- TP: correctly labeled a conflict run
- True Negative (TN): correctly labeled a non-conflict run
- FP: incorrectly labeled a non-conflict run as a conflict run
- FN: incorrectly labeled a conflict run as a non-conflict run

The F2 score for all runs within the test dataset was computed as:

$$\mathrm{F}_2 = \frac{5\mathrm{TP}}{5\mathrm{TP} + 4\mathrm{FN} + \mathrm{FP}}$$

The ground truth binary conflict versus no conflict label was determined analytically using the surrogate safety measure of TTC. For two road users whose travel paths intersect, if their individual time-to-intersect ranges at time *t* had overlap, then TTC(*t*) was calculated for the pair of road users. If TTC was calculated for a pair of road users based on the condition stated above and less than or equal to a threshold of 1.5 seconds, then the label of “conflict” was assigned to that data run. If TTC was not calculated due to the condition above not being met or the calculated TTC was greater than the threshold, then the label of “no conflict” was assigned to that data run.

3. Stage 1B Technical Evaluation Results

The following section covers the final metric outcomes for all participating teams that submitted their results and met final submission standards: that is, teams which abided by the required road user subclass labeling and timestamp standards. Of the 13 participating teams, 11 submitted valid results while 2 did not, and therefore, received lower scores due to improper road user subclass labeling and/or timestamp usage for Sections 3.1 and 3.2. Please note that all results shared in this report have been anonymized and/or aggregated to protect individual team privacy.

For the 11 valid submissions, we compare object detection/classification/localization, path prediction, and conflict prediction submissions measured by mAP, ADE, and conflict F2, respectively. Additionally, the results from the 2 teams disqualified from Sections 3.1 and 3.2 are included in Section 3.3. Additional subsections describe final results by subclass and run scenario to illustrate for which parameters participant ISSs could benefit from future work.

3.1. Detection, Classification, and Localization Results

Detection, classification, and localization proficiency is captured by the mAP metric (see Section 2). The following results catalogue mAP scores across all runs and subclasses (3.1.1), by road user subclass (3.1.2), and by scenario (3.1.3, 3.1.4). See Section 2.2 for details of the mAP score evaluation.

3.1.1. Detection, Classification, and Localization Results Overall

Figure 12 displays a histogram of mAP score by valid submission. Of the 11 valid submissions' scores, the majority (8/11) exceed 0.5 across all subclasses and runs, defining a 0.65 spread (0.14 minimum to 0.79 maximum). The mean score and variance are 0.55 and 0.04, respectively.

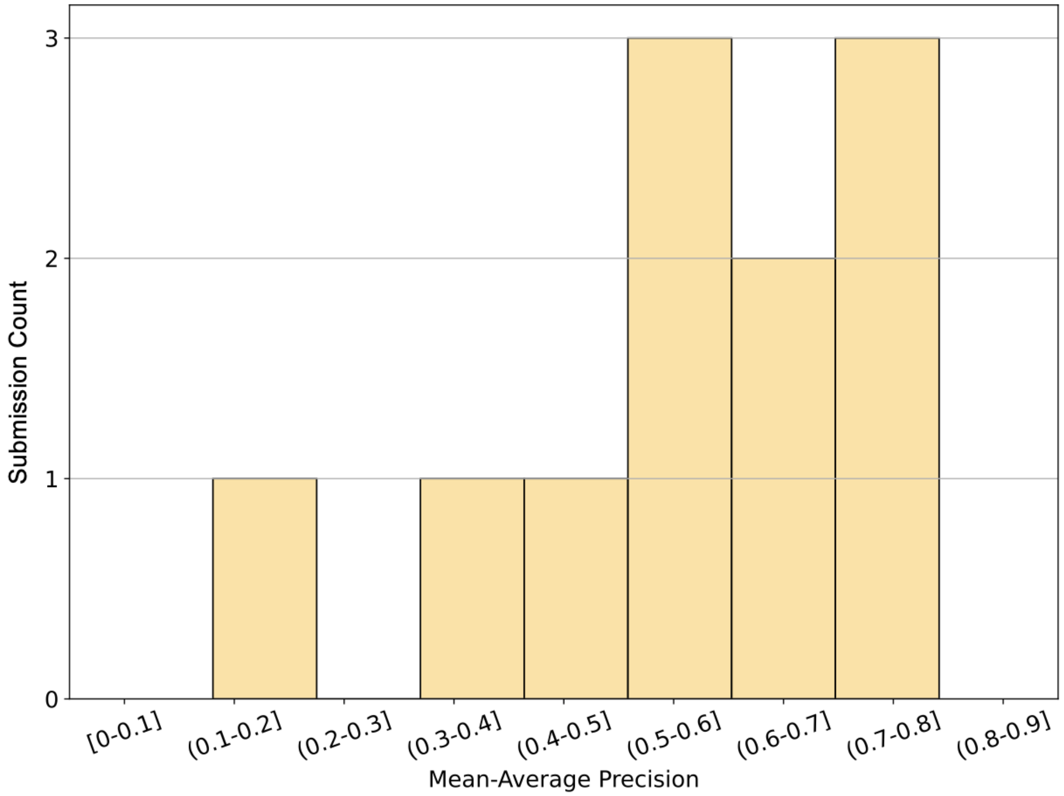


Figure 12. mAP Score Histogram for Valid Submissions (Final Results).

3.1.2. Detection, Classification, and Localization Results by Road User Subclass

Figure 13 displays a box plot of mAP scores by road user subclass type, where the *two-level* label describes the overall mAP. Ordered by mean score, the road user subclasses considered are Vehicle (0.71), Adult (0.67), Adult Wheelchair User (0.52), Adult Bicyclist (0.45), Adult Non-Motorized Device User (0.48), Adult Scooter/Skateboard User (0.45), and Child (0.29). **Table 3** shows the number of times the road user subclasses appeared in the dataset, for reference.

Table 3. Subclass Counts in Dataset.

Subclass	Count
Vehicle	149
Adult	98
Child	48
Adult Bicyclist	64

Subclass	Count
Adult Wheelchair User	30
Adult Non-Motorized Device User	50
Adult Scooter/Skateboard User	30

Participants’ performance for detecting, classifying, and localizing the Vehicle and Adult subclasses were better than their performance for the two-level overall submissions (0.71 and 0.67 vs. 0.56, respectively). Additionally, for these two road user subclasses, participants’ mAP scores had notably tighter variances (0.13 and 0.04) than for the other road user subclasses. Participants’ performance for the subclasses Adult Wheelchair User, Adult Bicyclist, Adult Non-Motorized Device User, and Adult Scooter/Skateboard User can be largely grouped together with respect to mean mAP score (0.45–0.52 range) and variance (0.09-0.12 range). Performance for detecting, classifying, and localizing the Child subclass is more than 0.16 below that of the next lowest subclass, performing notably below average and with slightly tighter variance (0.08).

Participants’ mAP score performance by road user subclass conforms largely to dataset representation (see **Table 3**) with some exceptions. Localization performance for Child lags, in part for this reason, but still falls short compared to other similarly sampled subclasses (Adult Wheelchair User, Adult Scooter/Skateboard User). Detection, classification, and localization results suggest that targeted increases in representation are a promising method to improve overall performance.

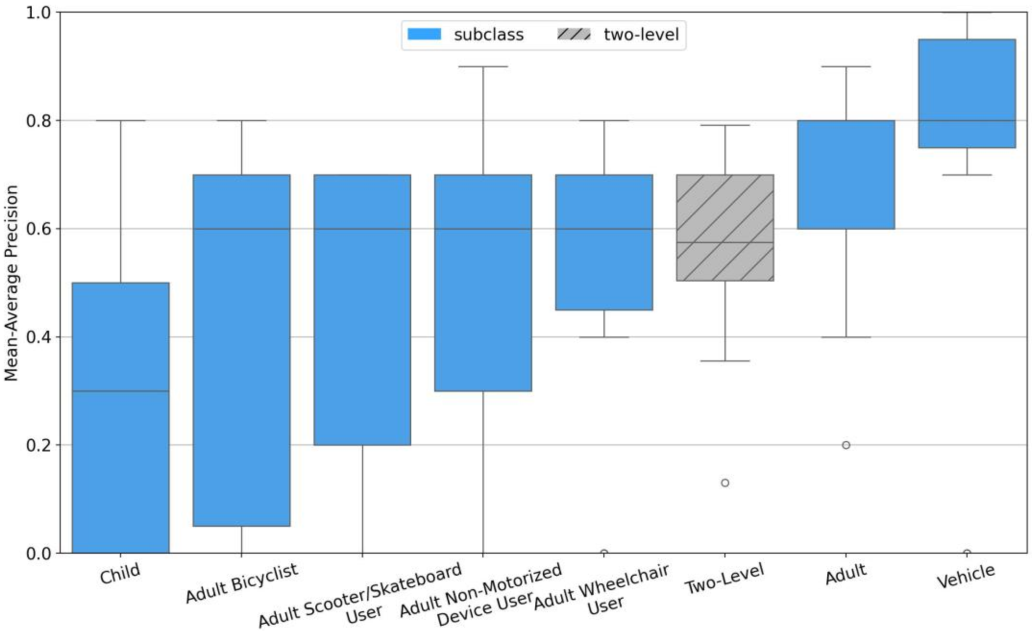


Figure 13. mAP Score by Road User Subclass (Final Results). Circles denote outliers.

Detection, classification, and localization performance by subclass is further partitioned by intended vehicle and VRU speed. **Table 4** displays overall mAP (across all subclasses) by runs containing a specified road user subclass intended to be moving at a particular speed, ordered by descending mAP with “Fast” intended speeds shown as orange rows and other speeds shown as blue in the table. The spread of average mAP is modest (0.40 minimum vs. 0.51 maximum), but scores slightly favored slow (0.45–0.51) as opposed to fast intended road user speeds (0.40–0.50). For runs containing a Child or Adult Bicyclist instance, participants typically performed worse, while for those containing a Vehicle, participants performed better.

This observation is potentially an outcome of the number of runs with each road user subclass and intended speed combination. mAP score of subclass-speed run group is positively correlated with run count (by independent T-test).

Table 4. mAP by Road User Subclass and Intended Speeds (Final Results).

(Intended Speed) & Road User Subclass	Count of Intended Speed by Road User Subclass Instances	mAP Score (Avg.)
(Average, *15 mph*) Vehicle	87	0.51
(Fast, *25 mph*) Vehicle	84	0.50
(Slow, *2 mph*) Adult	33	0.49
(Slow, *6 mph*) Adult Bicyclist	30	0.48
(Fast, *5 mph*) Adult	28	0.48
(Slow, *2 mph*) Child	31	0.45
(Fast, *15 mph*) Adult Bicyclist	25	0.43
(Fast, *3 mph*) Child	20	0.40

3.1.3. Detection, Classification, and Localization Results by Scenario

Like Section 3.1.2, mAP performance is scrutinized by groups of runs with particular conditions (a scenario) to determine and understand strengths and weaknesses of the evaluated dataset. **Table 5** orders run scenarios by descending mAP, but like Section 3.1.2 highlights that the mAP spread is relatively narrow (0.55 max vs. 0.47 min). On separate ends of the range are composite run groups: Non-Conflict (0.55) vs. Conflict runs (0.47) and Day (0.54) vs. Night (0.50) runs.

Table 5. mAP by Run Scenario (Final Results).

Scenario Groups	Count of Runs by Scenario Group	mAP Score (Avg.)
Non-Conflict	144	0.55
Day	100	0.54
Right-Turn	81	0.53

Scenario Groups	Count of Runs by Scenario Group	mAP Score (Avg.)
Non-Occlusion	85	0.52
Left-Turn	86	0.51
Night	85	0.50
Occlusion	60	0.48
Conflict	41	0.47

3.1.4. Detection, Classification, and Localization Improvements

All teams submitted First and Final Data Submissions of their detection, classification, and localization estimates. The First Data Submission captured the performance of the proposed ISS system within 72 hours of the test data release without the benefit of extensive appraisal of the data, which may unrealistically influence measures dependent on released data (mAP scores, for instance, are improvable through manual labeling). Comparing improvement in metrics between submissions may shed light on which VRU types, roadway scenarios, and overall predictive tasks have most improvement potential.

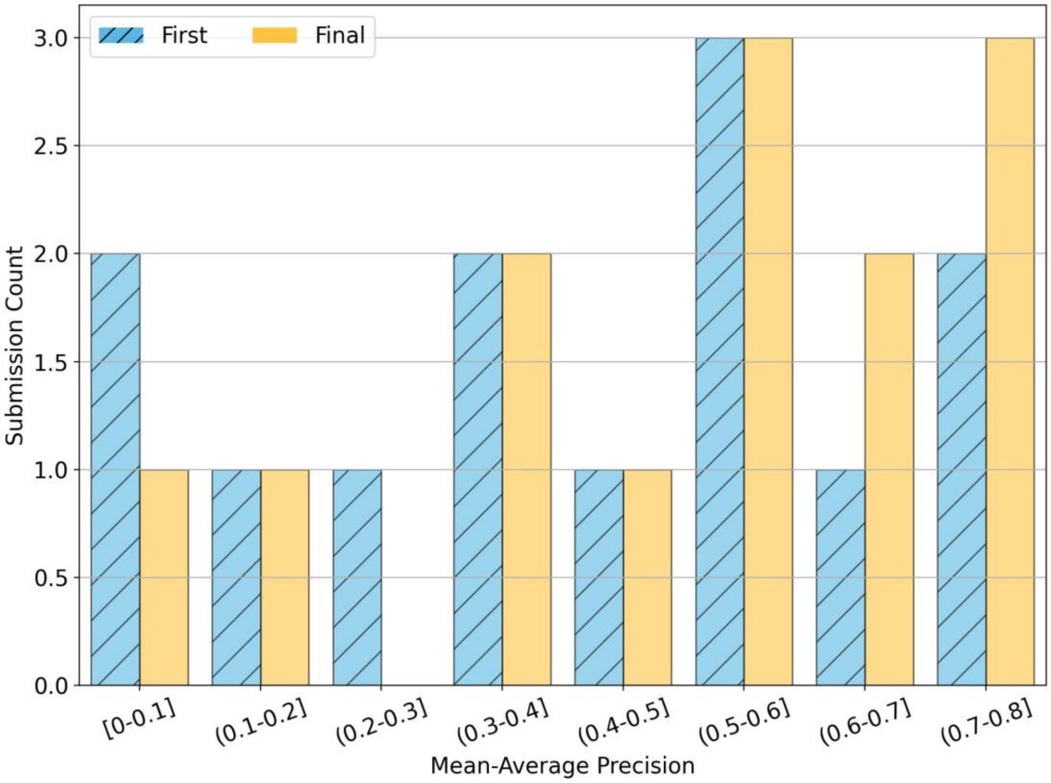


Figure 14. Distribution in mAP Score Improvements.

Figure 14 displays the distribution of overall First Data Submission alongside the Final Data Submission mAP scores. All changes between First and Final Data submissions are positive, such that the former 0.0–0.75 mAP range is replaced with a rightward-shifted 0.13–0.79 mAP range. Submissions in the bottom half of the distribution sport higher gains than those at the top: the bottom half of First Data Submission teams increased mAP by an average of 0.15 while the top half did so by 0.04. **Table 6** orders subclasses by descending mAP improvement.

Table 6. Subclass Ordered by mAP Improvement between First Data Submission and Final Data Submission.

Subclass	mAP Improvements
Adult	+ 0.18
Adult Wheelchair User	+0.13
Adult Scooter/Skateboard User	+0.13
Adult Non-Motorized Device User	+0.12
Vehicle	+0.05
Child	+0.04
Adult Bicyclist	+0.02

Figure 15 characterizes mAP improvements by road user subclass. All subclasses demonstrate positive deltas between First and Final Data Submissions, gaining 0.09 (~10% of total mAP) on average. Like the pattern observed in Section 3.1.2, several of the VRU subclasses cluster together apart from Adult Bicyclist, which improves the least. Besides a substantial average mAP gain (+0.05 more than closest subclass), Adult results demonstrate tighter variance.

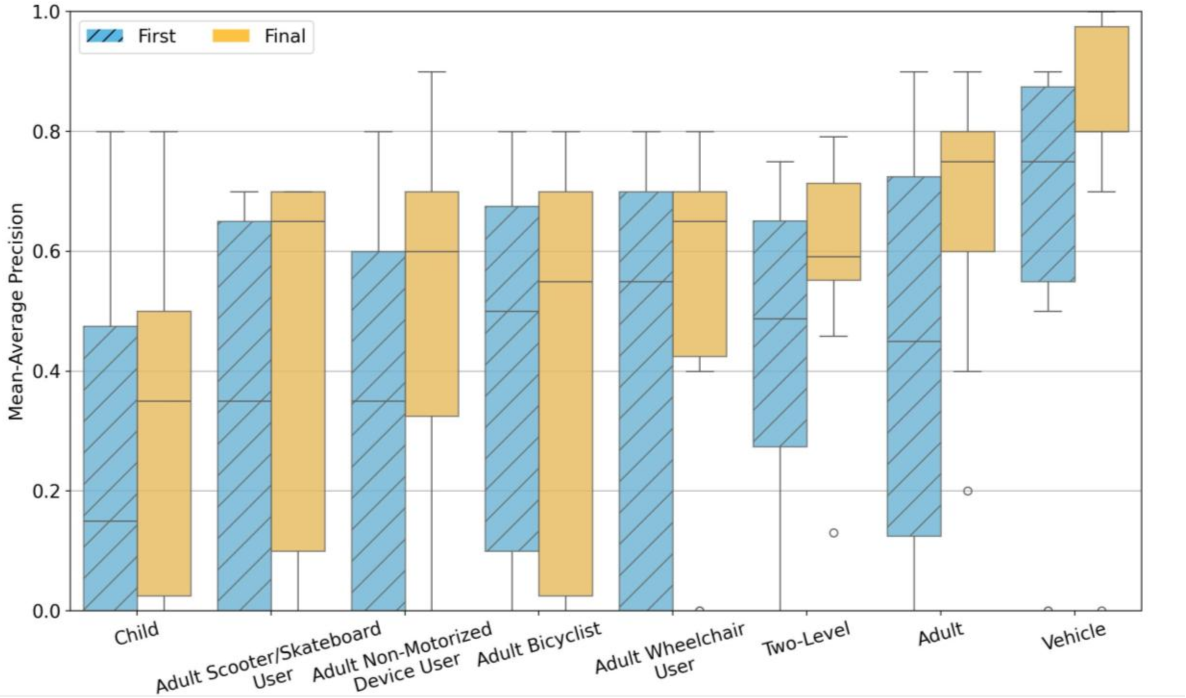


Figure 15. mAP Score by Road User Subclass (Improvements between First Data Submission and Final Data Submission).

3.2. Path Prediction Results

Path Prediction proficiency is captured by the ADE metric (see Section 2), the average distance in meters of the true paths from predicted paths. The following results catalogue ADE scores across all runs and road user subclasses (3.1.1), by subclass (3.1.2), and by run (3.1.3, 3.1.4). See Section 2.2 for details of ADE score evaluation.

3.2.1. Path Prediction Results Overall

Figure 16 displays a histogram of ADE score by valid submission. Of the 11 valid submissions, ADE scores fell primarily in the 2.21 m–9.01 m range (6.8 m spread), except for one 15.7 m ADE outlier (13.49 m spread). The top 4 submissions' ADEs are bounded by 4 m with a range of 2.21 m–3.81 m. The mean score and variance are 6.56 m and 15.9 m^2 respectively.

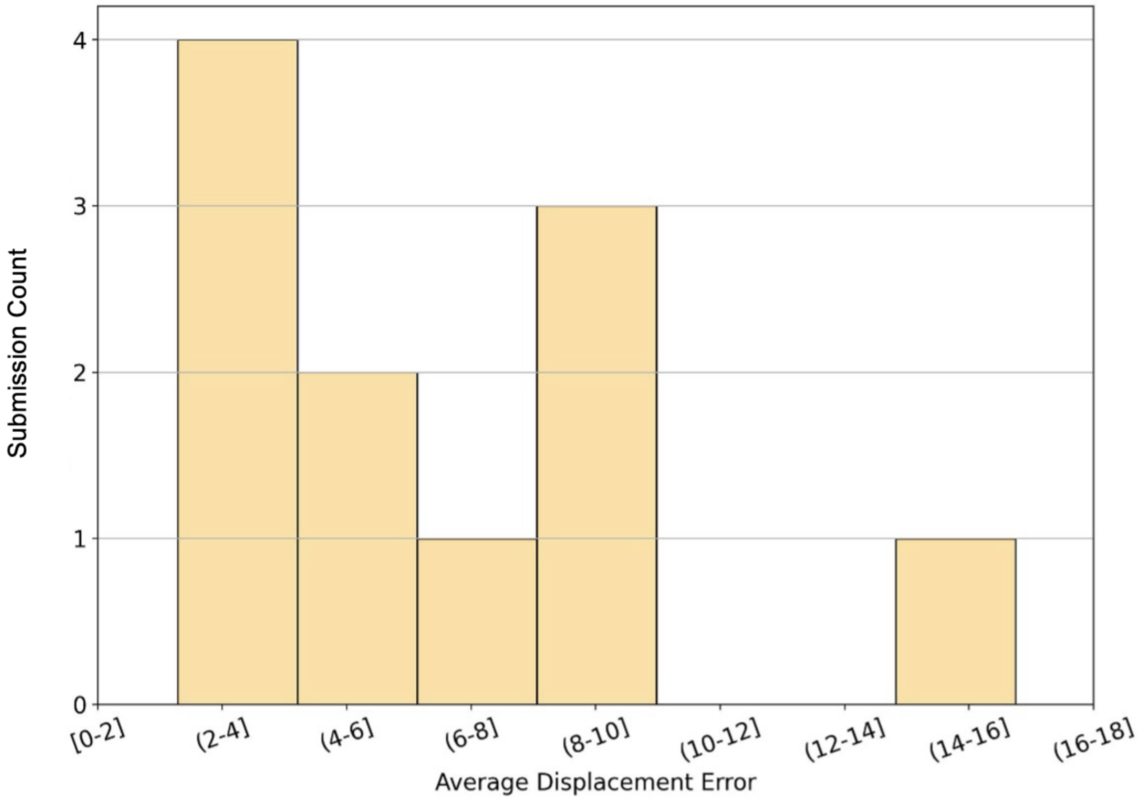


Figure 16. ADE Over All Runs and Road User Subclasses (Final Results).

3.2.2. Path Prediction Results by Road User Subclass

Figure 17 displays a boxplot of ADE scores by road user subclass type, where the *two-level* label describes the overall ADE across all subclasses. Ordered by mean score, the subclasses perform in nearly inverse order to that of their mAP performance (see Section 3.1.2): Adult Wheelchair User (1.4 m), Child (2.5 m), Adult Non-Motorized Device User (2.4 m), Adult Scooter/Skateboard User (3.1 m), Adult (6.2 m), Adult Bicyclist (6.3 m), and Vehicle (13.3 m).

All road user subclass scores (apart from Vehicles') outperform two-level, likely reflecting the comparative speed and unpredictability of motorized vehicles over VRUs; Adult Bicyclist is the second-lowest, likely for similar reasons. The score gap between vehicles and VRUs may be misleading, since the former subclass covers more ground in a shorter amount of time, expanding the ceiling of possible error. In contrast, VRUs move much slower, constricting the space of possible future paths. Path prediction scores are not normalized by speed.

The VRU subclasses Adult Wheelchair User, Adult Non-Motorized Device User, Adult Scooter/Skateboard User group together with respect to mean score (1.4–3.1 ADE range) and their collective outperformance of Adult may suggest an inverse relationship between path prediction score and sample size (opposite of the localization/sampling described in Section 3.1.2). Subclass performance conforms largely to dataset representation in **Table 3**. Classification results suggest that targeted increases in representation may challenge path prediction stability and reliability.

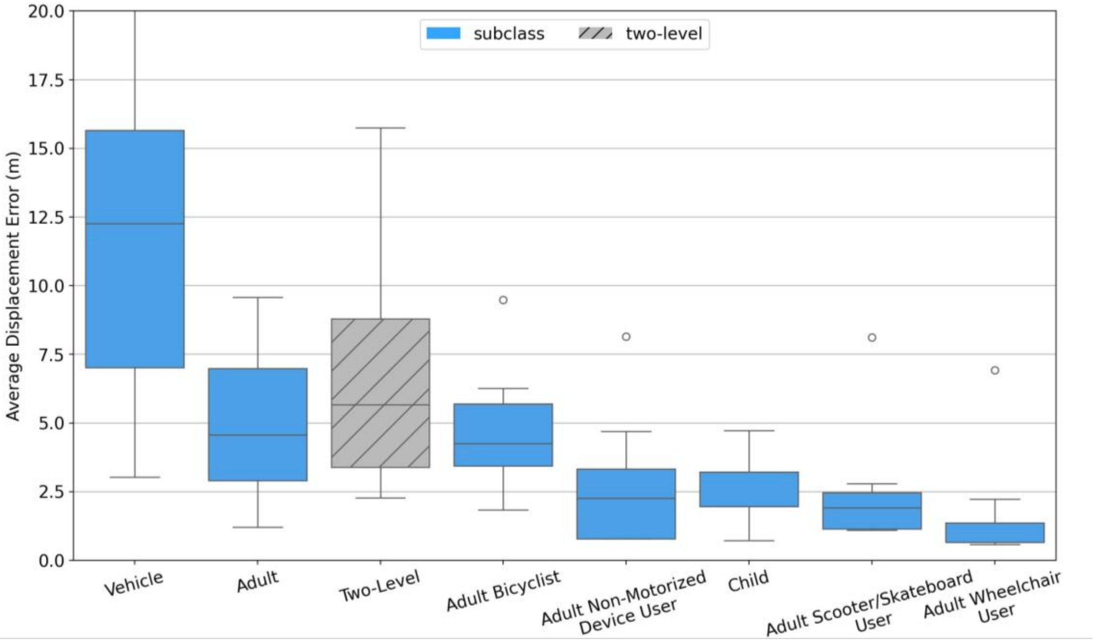


Figure 17. ADE by Road User Subclass (Final Results). Circles denote outliers.

Path prediction performance by road user subclass is further partitioned by speed. **Table 7** displays overall ADE (across all subclasses) by runs containing a specified subclass moving at a particular speed. The spread of means is modest (7.23 m max vs. 6.20 m min) where slow-speed runs are typically higher-error than fast-speed runs; excluding Adult Bicyclist, fast runs are contained to the bottom (6.20 m–6.49 m) of the error distribution. Fast Adult runs perform 0.86 m better than its respective slow runs, the highest fast/slow subclass gap. Overall, path prediction appears less sensitive to the number of runs in comparison to detection, classification, and localization (Section 3.1.2).

Table 7. ADE by Road User Subclass and Intended Speeds (Final Results).

(Intended Speed) & Road User Subclass	Number of Intended Speed by Road User Subclass Instances	ADE Score (Avg.) (m)
(Fast, 25 mph) Vehicle	84	6.20
(Fast, 3 mph) Child	20	6.39
(Fast, 5 mph) Adult	28	6.49
(Average, 15 mph) Vehicle	87	6.70
(Slow, 6 mph) Adult Bicyclist	30	7.13
(Slow, 2 mph) Child	31	7.17

(Intended Speed) & Road User Subclass	Number of Intended Speed by Road User Subclass Instances	ADE Score (Avg.) (m)
(Fast, 15 mph) Adult Bicyclist	25	7.23
(Slow, 2 mph) Adult	33	7.35

3.2.3. Path Prediction Results by Scenario

Like Section 3.1.3, path prediction performance is scrutinized by groups of runs with particular conditions (a scenario) to understand strengths and weaknesses of the evaluated algorithms. **Table 8** orders run scenarios by ascending ADE, but like Section 3.1.3 highlights that the ADE spread is narrow: the Conflict and Right-Turn run groups define a 1.39 m range, and since Right-Turn runs are an outlier group, the distribution largely rests in the 0.37 m range defined by Day and Conflict run groups. Echoing their detection, classification, and localization performance (Section 3.1.3), Right-Turn, Day, and Non-Conflict runs modestly outperform Left-Turn, Night, and Conflict runs respectively.

In contrast to detection, classification, and localization, run count appears less associated with path prediction performance.

Table 8. ADE by Run Scenario (Final Results).

Scenario Groups	Number of Runs by Scenario Group	ADE Score (Avg.) (m)
Right-Turn	81	5.42
Day	100	6.44
Non-Conflict	144	6.48
Occlusion	60	6.51
Non-Occlusion	85	6.54
Left-Turn	86	6.65
Night	85	6.67
Conflict	41	6.81

3.2.4. Path Prediction Improvements

Figure 18 displays the distribution of overall First Data Submission alongside that of Final Data Submission ADE scores. The majority of Final Data Submissions for path prediction improved performance, but 2 out of 11 teams performed slightly worse (+0.04 m and +0.17 m) in the Final Data Submission as compared to the First Data Submission. One team improved by ~2,000,000 m, but after excluding this outlier the mean ADE delta across teams is -1.90 m. Submissions in the bottom half of the distribution gained more ground than those at the top: -3.55 m vs. -0.81 m (again excluding the outlier value).

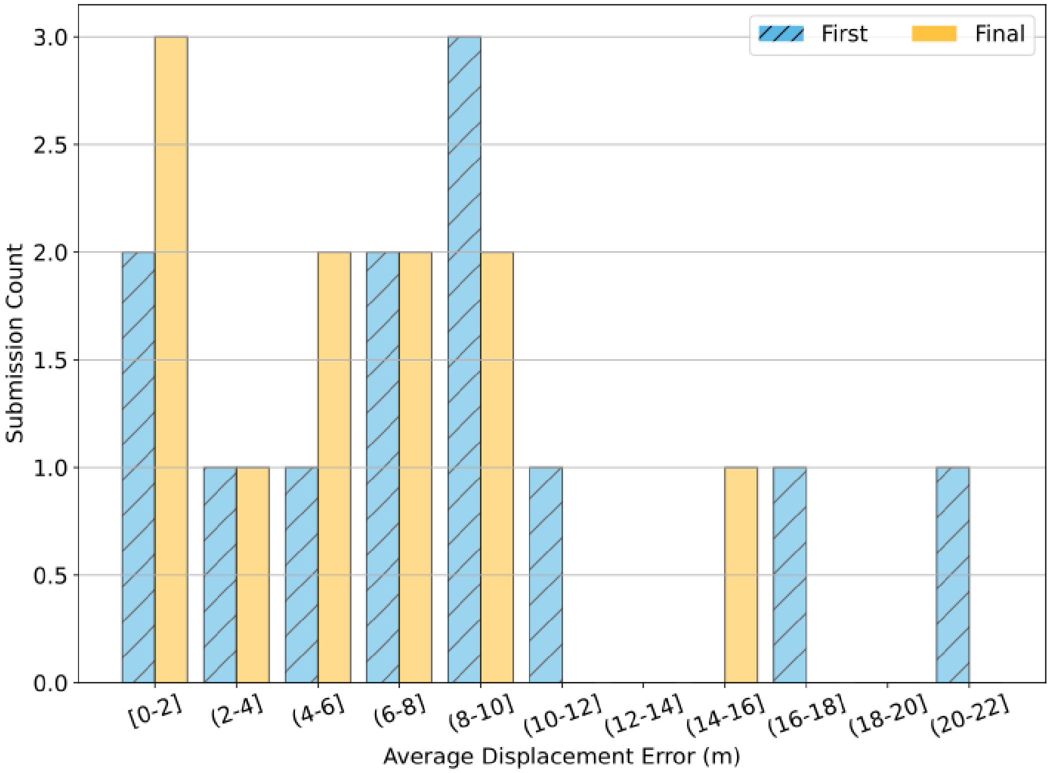


Figure 18. Distribution of ADE Improvements.

Table 9 orders subclasses by descending mean ADE improvement: Adult, the most improved class, also substantially reduced its error variance from 0.12 m^2 to 0.05 m^2 and reduced outliers. It is of note that outlier predictions may be most susceptible to manual review rather than technical improvements to the system. Additionally, subclasses with the least dataset occurrences improved the least between the First and Final Data Submissions.

Table 9. Subclass ADE Improvements.

Subclass	ADE Improvements
Adult	-3.11 m
Adult Bicyclist	-2.26 m
Vehicle	-1.83 m
Adult Non-Motorized Device User	-1.52 m
Child	-0.85 m

Subclass	ADE Improvements
Adult Wheelchair User	-0.53 m
Adult Scooter/Skateboard User	-0.14 m

Figure 19 characterizes ADE improvements by road user subclass. All subclasses demonstrate prediction error reduction between the first and final submissions.

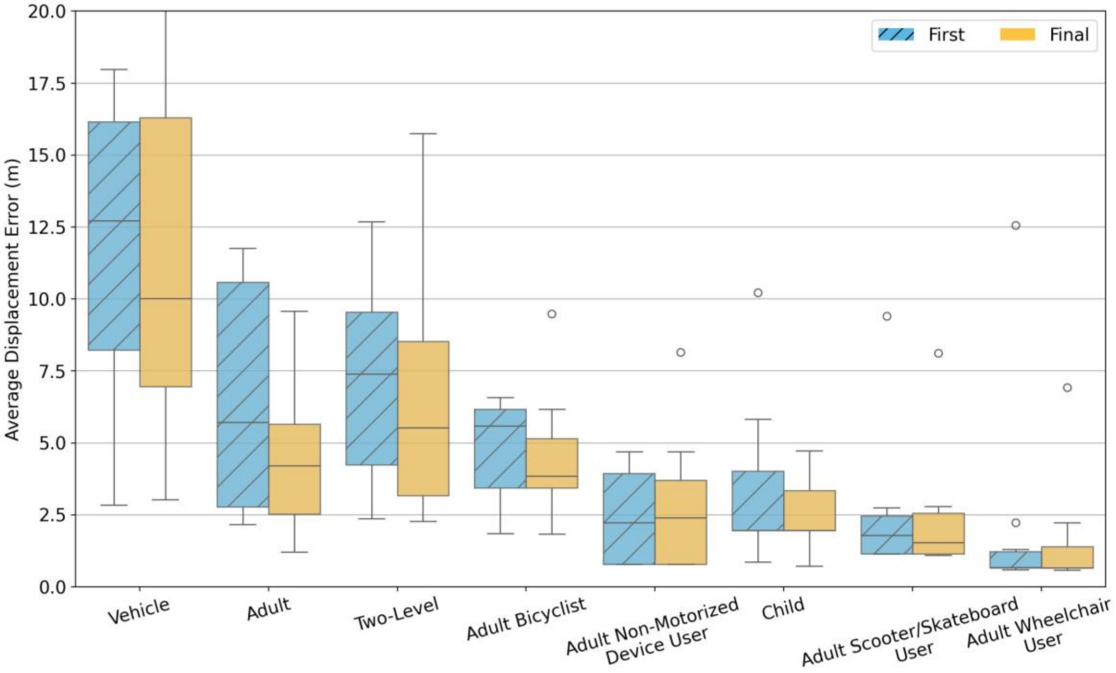


Figure 19. ADE Improvements by Road User Subclass.

3.3. Conflict Prediction Results

Conflict prediction performance is captured by Conflict Prediction F2 (Conflict F2) score, weighted in favor of true positive detections (see Section 2). The following results catalogue F2 across all runs and road user subclasses (3.1.1), by subclass (3.3.2), and by run (3.3.3). See Section 2.2 for details on F2 score evaluation.

3.3.1. Conflict Prediction – Overall Results

Figure 20 displays a histogram of Conflict F2 score by valid submission. As opposed to Sections 3.1.1 and 3.2.1, all 13 submissions are considered valid for this category, and their results are included for analysis. Submitted results define a 0.25–0.67 range (0.42 spread). The mean score and variance are 0.47 and 0.03, respectively.

Recall that each run is assigned a “Conflict” or “Non-Conflict” label dependent on whether the TTC for any VRU drops below 1.5 seconds (see Section 0). TTC explains 37 percent of variance in the number of successful predictions (ordinary least squares (OLS) regression).

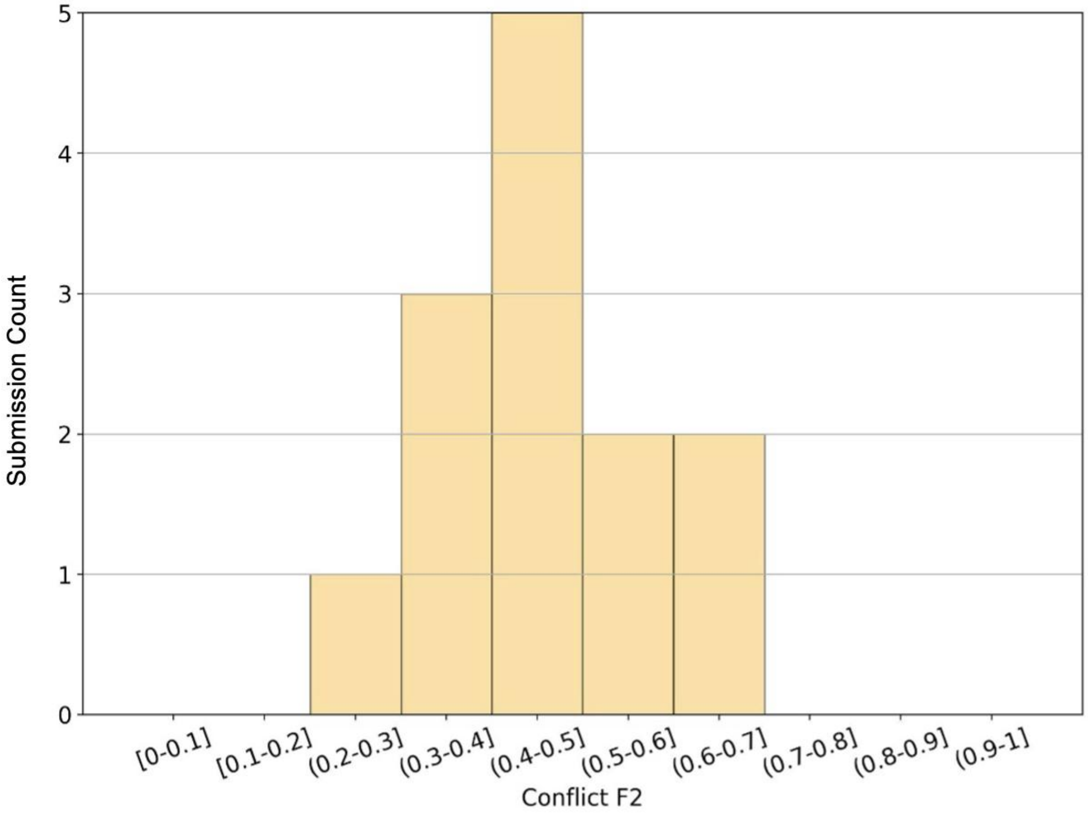


Figure 20. Conflict Prediction F2 Scores Overall (Final Results).

3.3.2. Conflict Prediction Results by Road User Subclass

F2 score does not subdivide by class since each run is labeled as “Conflict” or “Non-Conflict.” Therefore, road user subclass performance is evaluated solely by run group.

Table 10 displays Conflict F2 (across all road user subclasses) by runs containing a specified subclass moving at a particular speed (see Section 3.1.3). Run groups containing the same subclass perform comparably: they largely clump together in comparative ranking. Additionally, fast run groups modestly outperform their slow/average counterparts (+0.05 on average).

While Fast Child runs perform comparably (if poorly) with other road user subclass / speed groups, Slow Child runs perform significantly worse: 0.16 worse than the closest run group.

Table 10. Conflict F2 by Road User Subclass and Intended Speed (Final Results).

(Intended Speed) & Road User Subclass	Number of Intended Speed by Road User Subclass Instances	Conflict F2 (Avg.)
(Fast, 15 mph) Adult Bicyclist	25	0.63
(Slow, 6 mph) Adult Bicyclist	30	0.58
(Slow, 2 mph) Adult	33	0.55
(Fast, 25 mph) Vehicle	84	0.53
(Fast, 5 mph) Adult	28	0.47
(Average, 15 mph) Vehicle	87	0.47
(Fast, 3 mph) Child	20	0.42
(Slow, 2 mph) Child	31	0.26

3.3.3. Conflict Prediction Results by Scenario

Like Sections 3.1.3 and 3.2.3, conflict prediction performance is scrutinized by groups of runs with particular conditions (a scenario) to understand contextual strengths and weaknesses of the evaluated algorithms. The Conflict and Non-Conflict run groups are not shown, since the F2 score requires some number of true positive and negative runs for coherence.

Table 11 orders run scenarios by descending Conflict F2. Localization and path prediction trends generally hold steady in conflict prediction. Right-Turn, Occlusion, and Day run groups modestly outperform their complement groups (Left-Turn, Non-Occlusion, Night). Most run group scores fall within a relatively tight range (0.56–0.45); however, unlike Sections 3.1.3 and 3.2.3, one group performs as a low outlier (Non-Occlusion).

In contrast to localization, run count is not significantly correlated with conflict prediction performance.

Table 11. Conflict F2 by Run Scenario (Final Results).

Scenario Groups	Number of Runs by Scenario Group	Conflict F2 (Avg.)
Right-Turn	81	0.56
Occlusion	60	0.56
Day	100	0.50
Night	85	0.49
Left-Turn	86	0.45
Non-Occlusion	85	0.37

3.3.4. Conflict Prediction Improvements

Figure 21 displays the distribution of overall first-submission Conflict F2 alongside that of the final-submission. No teams performed worse between submissions, but 5 of 11 teams made no improvement. The mean gain is 0.09, where submissions in the distribution bottom gained more ground than those at the top: +0.17 vs. +0.01 for the top 6 teams and bottom 7 teams respectively, out of 13 teams total.

Overall, the range of Conflict F2 scores shift from 0.10–0.67 to 0.25–0.67 between the two submissions. It is worth noting that conflict prediction, like localization, may benefit from manual review.

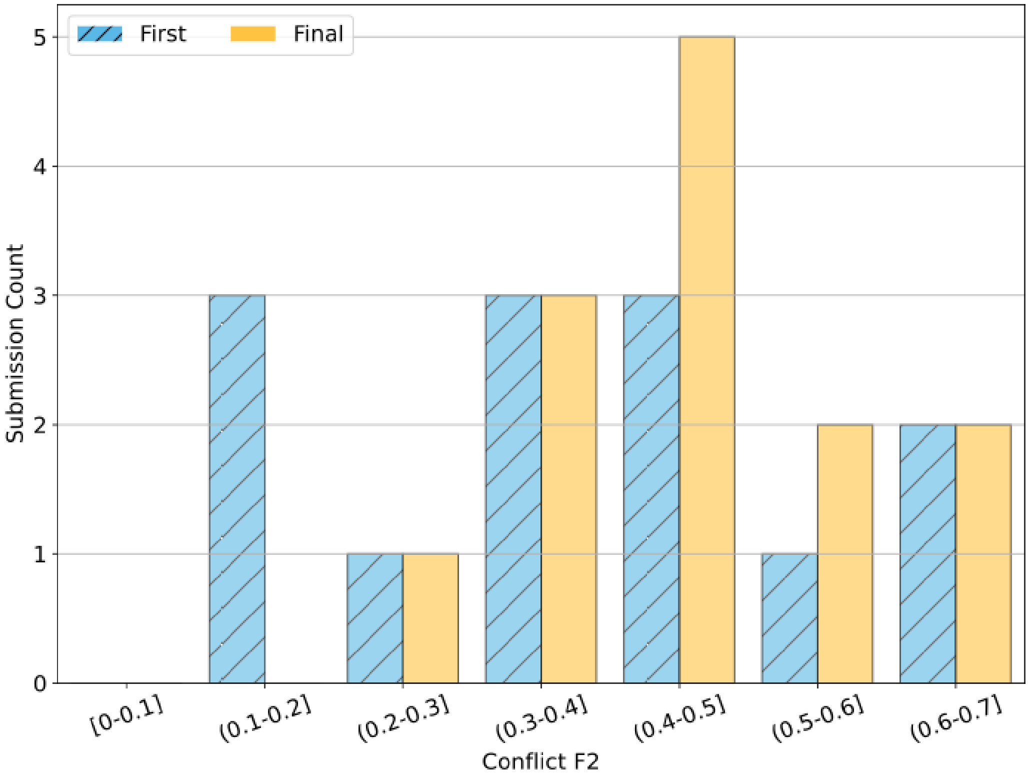


Figure 21. Overall Conflict F2 improvements

4. Key Takeaways

The U.S. DOT Intersection Safety Challenge Stage 1B (System Assessment and Virtual Testing Primary Track) sheds light on key technical elements expected to be part of a comprehensive ISS: (1) detection, classification, and localization of road users; (2) path prediction; and (3) conflict prediction between VRUs and vehicles. These data processing and decision-making steps are key inputs expected to inform downstream ISS interventions. Stage 1B demonstrated the performance of these technical capabilities against a limited but curated set of test data at a single four-way controlled test intersection at the FHWA TFHRC. The results from Stage 1B point to the maturity of these key technical elements of an ISS as well as potential research gaps to address in potential future activities focused on end-to-end ISS prototyping with conflict mitigation.

4.1. Key Takeaways for Detection, Classification, and Localization

Below are key takeaways from the detection, classification, and localization results:

- The average detection, classification, and localization mAP score was 0.67, with the best team's score being ~0.80. The mAP score was positively correlated to the number of subclass instances that participants were exposed to in the training and validation data, indicating the possibility for improvement with additional samples.
- The participants' mAP scores for the Vehicle subclass were higher than those for all VRU subclasses. Of the VRU road user subclasses, performance for Adult was comparable to that for the Vehicle subclass, and performance for the Child subclass ranked last as a low performance outlier (based on performance by subclass tracked with exposure/representation in the dataset).
- Detection, classification, and localization mAP performance disparity by scenario was narrow (0.08 range), though performances for Right-Turn, Occlusion, and Day scenario run groups were higher than those for the Left-Turn, Non-Occlusion, and Night scenario run groups—a pattern that held across all three technical evaluation metrics (see Sections 4.2 and 4.3).
- The performance for the weakest initial submissions improved the most, while initial top-performing teams gained modestly between the First and Final Data Submissions. Participants' performance for the Adult subclass was the most improved (nearly +0.20 mAP—or a 20 percent increase). Since Adult was the most common VRU subclass, the result indicated that robust sampling and additional exposure may yield further system improvements. Performance for the Adult Bicyclist and Child subclasses improved the least from the first to the final submissions.

4.2. Key Takeaways for Path Prediction

Below are key takeaways from the path prediction results:

- The average path prediction error (ADE) was ~6 m, with the best teams performing within ~2 m. For context, the approximate average vehicle length of a passenger car is ~5.79 m (Potts et al., 2023).
- Teams scored better on VRU subclass runs compared to Vehicle runs on path prediction (which is the opposite of the localization performance where teams scored better on vehicles compared to VRUs). Adult Bicyclist was the poorest scored VRU subclass; however, path prediction performance varied minimally between subclasses. Instead, ADE varied most by road user speed.
- By scenario, teams scored higher on the Right-Turn, Occlusion, and Day run groups compared to the Left-Turn, Non-Occlusion, and Night groups. The spreads of ADE scores by subclass and speed and scenario run groups were narrow (1.39 and 1.15 m respectively), but performance by subclass (Section 3.2.2) had a wider range (1.14–13.3 m range).
- Similarly to results seen for localization (Section 4.1), teams improved the most on predicting the paths of Adult runs across all subclasses (-3.11 m) while teams improved the least on Child runs (-0.85 m) from their First to Final Data Submissions. On average, submissions improved by roughly 1 m across all runs and subclasses.
- Notably, many outlier predictions in the first submissions were absent in the final submissions. This suggests that the additional time may have given teams a chance to manually correct their erroneous outlier path predictions from the first submission before submitting their final submission. More work is needed to determine if these improvements can be attributed to teams manually correcting predicted paths, which are comparably difficult to manually label, since manual labeling/correction is not a realistic expectation for real-world, real-time deployment of an ISS.

4.3. Key Takeaways for Conflict Prediction

Below are key takeaways from the conflict prediction results:

- Since conflict prediction cannot be evaluated by road user subclass—only by groups of runs containing a subclass—subclass-specific performance analysis is less straightforward. Results by road user subclass and speed (Section 3.3.2) suggest that road user subclass was more salient than road user speed. Teams performed better on runs containing Adult and Adult Bicyclist subclasses compared to those containing Vehicle and Child subclasses for conflict prediction.
- The conflict prediction performance range by scenario (~0.20 F2 spread) was wider than the performance range of localization and path prediction. Teams scored modestly higher on the Right-Turn, Occlusion, and Day run groups compared to the Left-Turn, Non-Occlusion, and Night groups (see Sections 4.1 and 4.2).
- The top-performing teams showed little to no improvement on conflict prediction between their First and Final Data Submissions. More work is needed to determine if improvements between the First and Final Data Submissions can be attributed to teams inspecting the test data and manually updating their conflict predictions, which would be unrealistic for a deployment setting.

4.4. Overall Key Takeaways for Stage 1B

The U.S. DOT Intersection Safety Challenge Stage 1B: System Assessment and Virtual Testing participating teams tested their sensor fusion and AI approaches utilizing the U.S. DOT-provided data

(U.S. DOT, 2024b).[4] Their technical performance was assessed against the test dataset based on scoring of their First and Final Data Submissions, with deadlines of 72 hours and 6 weeks after the test data was released, respectively. Key results and takeaways from the data challenge are summarized below:

- **Challenge results are encouraging for ISS prototyping.** Many teams were capable of suitably accurate detection, classification, and localization. Path and conflict prediction proved more difficult (as expected), at least in part because: (1) the teams had to use the U.S. DOT-supplied sensor data as opposed to data from their own installed sensors and devices, and (2) the teams could not update their predictions as time advanced since all sensor data was cutoff (t_{cut}) before the potential conflict. While these conditions allowed for a fair comparison across teams in the data science competition, they do not reflect how teams' ISSs would be expected to operate in real-time.
- **Teams showed minor improvements with additional time between First and Final Data Submissions.** Participants generally showed minor improvements to system performance between the First and Final Data Submissions across all metrics with some exceptions concentrated in path prediction evaluation. Teams were not given incremental scores or feedback following their First Data Submissions. Based on the results submitted, it is not clear what drove the minor improvements. In some cases, teams may have relied on an inspection of the test data to remove erroneous outliers manually. Detection, classification, and localization metrics were more susceptible to manual labeling than path prediction or conflict prediction.
- **Performance on road user subclasses differed by technical element, but all teams struggled most in detecting and predicting movement for the surrogate child.** Detection, classification, and localization scores were better for Vehicles, but path prediction scores were better for VRUs. The differing size and speed of the road user subclasses makes for difficult comparison, though scores were substantially worse for detection, classification, and localization of the Child subclass compared to all other subclasses, with minimal improvement from the First to Final Data Submission. Path prediction scores for the Child subclass fared better—though demonstrated minimal improvement—and runs containing the Child subclass were on average the lowest scoring for conflict prediction.
- **Road user speed appeared to impact technical performance differently, depending on the metric.** The impact of road user speed was dependent on metric. Teams typically performed better on detection, classification, and localization for slow road users compared to faster ones. For path prediction, on the other hand, teams typically performed better for fast road users compared to slower ones, perhaps since slow movement may indicate or precede a change in movement direction.
- **Nighttime and vehicle left turns point to areas for ISS improvement.** Teams performed better on day and right-turn runs compared to night and left-turn runs respectively across all metrics, though modestly. Pedestrian-vehicle conflicts have historically been more likely and more severe at night. Roughly 76 percent of pedestrian fatalities occur at night (FHWA, 2025).
- **Sensor fusion outperformed individual sensor approaches.** Teams using two or more sensor types in their systems clearly performed better than teams relying on a single sensor type. Visual

[4] To view a sample of the Intersection Safety Challenge Stage 1B data and for instructions to access the full dataset, please visit: https://data.transportation.gov/Roadways-and-Bridges/Intersection-Safety-Challenge-Stage-1B-Sample/vq7s-mv3v/about_data.

camera and LiDAR were the most commonly used sensor types. Occlusion of the scene did not noticeably hamper performance—likely the benefit of having access to multiple perspectives from different sensor types and placements.

- **Teams recognized and adapted to data quality challenges.** Although teams identified inherent flaws and unexpected inconsistencies in the data (reflective of potential real-world situations), they generally adapted and used the imperfect data to build a robust ISS.
- **Teams acknowledged the value of the data challenge.** Teams viewed the challenge's premise as a crucial call-to-action and many outlined ambitions to extend their ISS work beyond this initiative. Several teams noted that U.S. DOT Stage 1B data enhanced the development of their ISS.

4.5. Research Gaps to Address in Potential Prototyping Stage

Results from the U.S. DOT Intersection Safety Challenge Stage 1B: System Assessment and Virtual Testing are promising and point to research gaps to address in potential follow-on activities focused on ISS prototyping. These include diving deeper to understand ISS:

- **Conflict Prediction Reliability and Speed.** The Stage 1B data science competition was limited in its ability to test real-time ISS responses. Stage 1B tried to eliminate the possibility of teams manually labeling the test data by rewarding quick responses, constraining the submission windows, and cutting the sensor data so teams did not know whether a conflict occurred or not in the ground truth. However, these strategies did not guarantee that an ISS could reliably predict conflicts early enough and fast enough for the system to take mitigating actions. Therefore, conflict prediction reliability, timing, and speed are important elements to assess in any ISS prototype given this will directly inform conflict mitigation strategies.
- **Conflict Mitigation Strategies and Effectiveness.** As was shown in **Figure 5**, Stage 1B was designed as a data science competition focused on better understanding three key technical elements of ISS operations: (1) the detection, localization, and classification of VRUs and vehicles; (2) path prediction; and (3) conflict prediction among identified VRUs and vehicles. Prototyping must assess the practicality of the final key element of an ISS, conflict mitigation, as this could not be assessed in the Stage 1B data science challenge. Given ISS response speed, the ISS prototyping stage could explore possible conflict mitigation strategies and their effectiveness at preventing VRU-vehicle conflicts.
- **Cost Effectiveness.** While ISS cost effectiveness was included as a factor for consideration in the Deployment Suitability criterion for Stage 1A: Concept Paper and Stage 1B: System Assessment and Virtual Testing, it requires additional exploration when considering an entire end-to-end ISS prototype. Certain ISS components may be driving overall ISS costs, so understanding what they are is crucial to identifying strategies to reduce their costs. Further research is needed to understand if an ISS can be deployed cost effectively today, and if so, under what conditions and for which intersection geometries.
- **Potential for Broader Deployment.** Stage 1B assessed teams' systems under relatively narrow conditions (e.g., low speeds with only a few VRUs and vehicles) at a single test roadway intersection. Speeds and complexity are expected to be much higher in a real-world setting. Prototyping can assess how ISSs perform in real-world settings with teams using their own sensors installed at the intersection(s). These ISS prototypes can also shed light on whether narrow or general-purpose ISSs are more practical for near-term deployment. Please note that U.S. DOT anticipates potential future prototyping activities would be open to all interested entities without regard to participation in the Intersection Safety Challenge Stage 1A or Stage 1B.

References

1. FHWA. (last updated: 2024, July 26). “About Intersection Safety.” FHWA Highway Safety Programs. https://highways.dot.gov/safety/intersection-safety/about

2. FHWA. (last updated: 2025, February 26). “Nighttime Visibility for Safety.” FHWA Center for Accelerating Innovation. https://www.fhwa.dot.gov/innovation/everydaycounts/edc_7/nighttime_visibility.cfm

3. National Center for Statistics and Analysis. (2024a, June, Revised). *Overview of Motor Vehicle Traffic Crashes in 2022* (Traffic Safety Facts Research Note. Report No. DOT HS 813 560). National Highway Traffic Safety Administration. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813560

4. National Center for Statistics and Analysis. (2024b, July). *Pedestrians: 2022 Data* (Traffic Safety Facts. Report No. DOT HS 813 590). National Highway Traffic Safety Administration. https://crashstats.nhtsa.dot.gov/Api/Public/Publication/813590

5. National Center for Statistics and Analysis. (2024c, November). *Early Estimates of Motor Vehicle Traffic Fatalities and Fatality Rate by Sub-Categories through June 2024* (Crash Stats Brief Statistical Summary. Report No. DOT HS 813 661). National Highway Traffic Safety Administration. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813661

6. Potts, I. B., Harwood, D. W., Cook, D. J., Moran, R. A., & Busenbark, L. (2023). “Chapter 5: Updated Dimensions and Recommended Additions to the Green Book Design Vehicles.” In *Highway and Street Design Vehicles: An Update* (Figure 15. Page 37). National Academy of Sciences. https://nap.nationalacademies.org/read/27236/chapter/7

7. Townsend, H., Gatiba, A., Thompson, K., Wang, P., Wunderlich, K. (2023). *Summary Report on Request for Information (RFI): Enhancing the Safety of Vulnerable Road Users at Intersections* (No. FHWA-JPO-23-986). U.S. DOT ITS JPO. https://rosap.ntl.bts.gov/view/dot/66622

8. U.S. DOT. (2022, September 16). *Enhancing the Safety of Vulnerable Road Users at Intersections; Request for Information*. Federal Register. https://www.federalregister.gov/documents/2022/09/16/2022-20188/enhancing-the-safety-of-vulnerable-road-users-at-intersections-request-for-information

9. U.S. DOT. (2024a, January 8). “U.S. DOT Announces Winners of the Intersection Safety Challenge.” https://www.transportation.gov/briefing-room/us-dot-announces-winners-intersection-safety-challenge

10. U.S. DOT. (2024b, December 9). “Intersection Safety Challenge Stage 1B Sample” [Open-Access Data Portal]. ITS DataHub. https://its.dot.gov/data/

11. U.S. DOT. (2025, January 7). “U.S. DOT Announces Winners of the Intersection Safety Challenge Stage 1B: System Assessment and Virtual Testing.” https://www.transportation.gov/briefing-room/us-dot-announces-winners-intersection-safety-challenge-stage-1b-system-assessment-and

12. Zhang, H., Rogozan, A., & Bensrhair, A. (2022). An enhanced N-point interpolation method to eliminate average precision distortion. *Pattern Recognition Letters*, *158*, 111-116.

U.S. Department of Transportation
ITS Joint Program Office – HOIT
1200 New Jersey Avenue, SE
Washington, DC 20590

Toll-Free “Help Line” 866-367-7487

www.its.dot.gov

[FHWA-JPO-25-157]

U.S. Department of Transportation