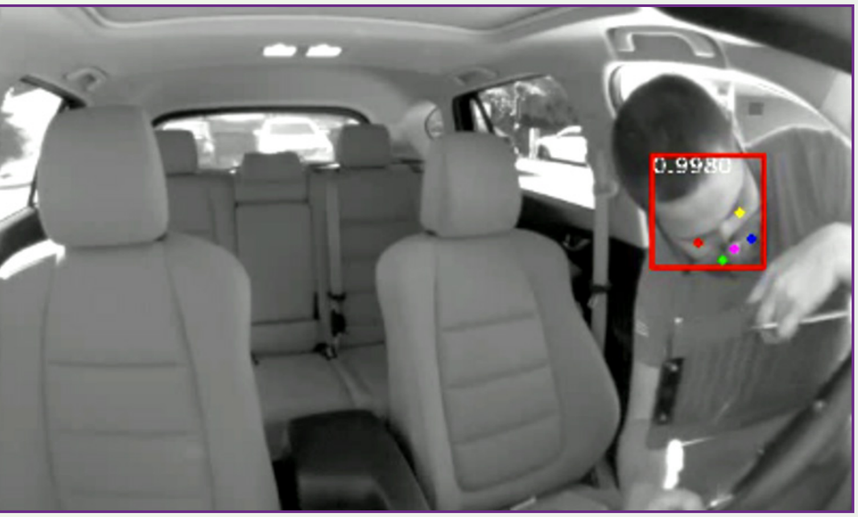
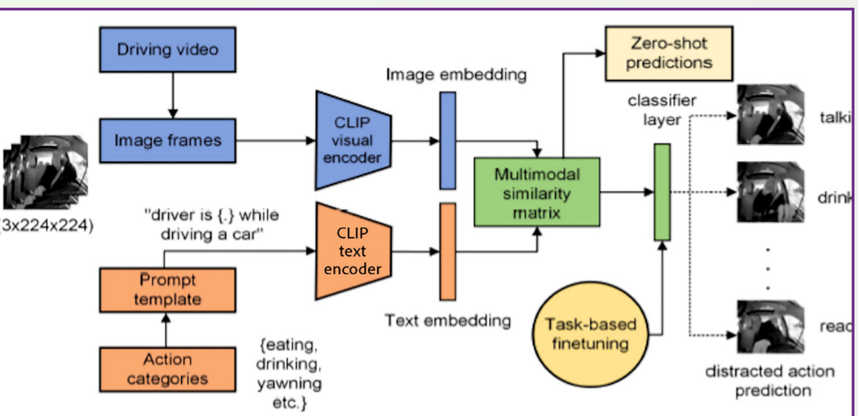
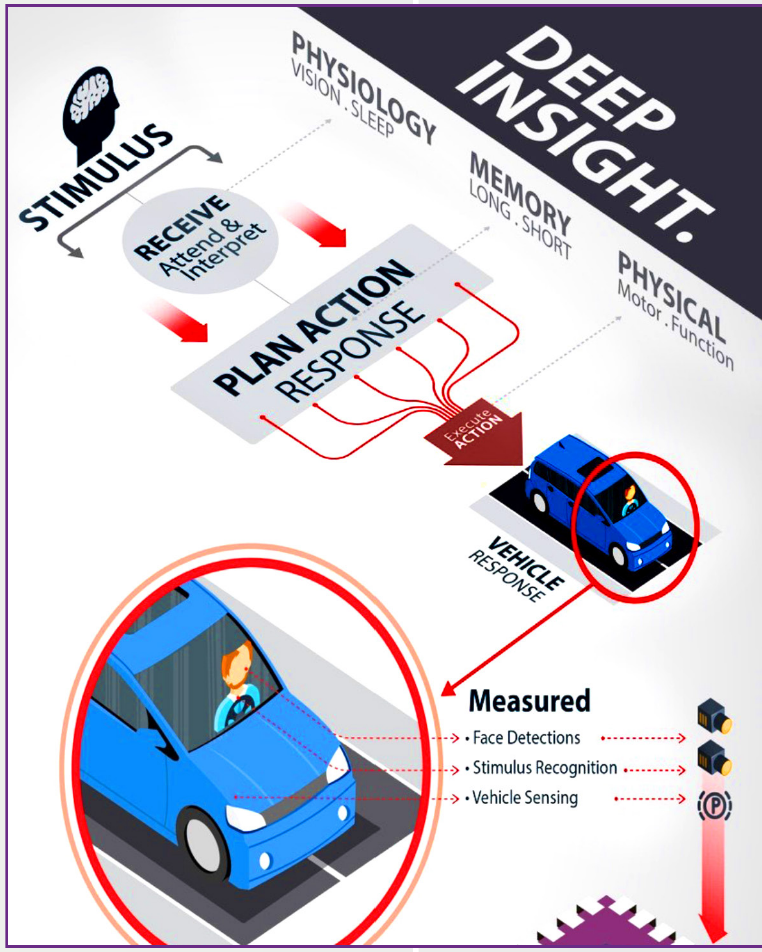
Exploratory Advanced Research Program

Deep InSight: A Driver-State Estimation Platform for Processing Naturalistic Driving Data

Research Summary Report







U.S. Department of Transportation
Federal Highway Administration

Notice

This document is disseminated under the sponsorship of the U.S. Department of Transportation (USDOT) in the interest of information exchange. The U.S. Government assumes no liability for the use of the information contained in this document.

Non-Binding Contents

Except for the statutes and regulations cited, the contents of this document do not have the force and effect of law and are not meant to bind the States or the public in any way. This document is intended only to provide information regarding existing requirements under the law or agency policies.

Quality Assurance Statement

The Federal Highway Administration (FHWA) provides high-quality information to serve Government, industry, and the public in a manner that promotes public understanding. Standards and policies are used to ensure and maximize the quality, objectivity, utility, and integrity of its information. FHWA periodically reviews quality issues and adjusts its programs and processes to ensure continuous quality improvement.

Disclaimer for Product Names and Manufacturers

The U.S. Government does not endorse products or manufacturers. Trademarks or manufacturers' names appear in this document only because they are considered essential to the objective of the document. They are included for informational purposes only and are not intended to reflect a preference, approval, or endorsement of any one product or entity.

Cover photo credits clockwise from top left:

© REACTOR Lab.
© REACTOR Lab.
Source: FHWA.

Technical Report Documentation Page

1. Report No. FHWA-HRT-24-137	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Deep InSight: A Driver-State Estimation Platform for Processing Naturalistic Driving Data		5. Report Date December 2024
		6. Performing Organization Code
7. Author(s) Anuj Sharma (ORCID: 0000-0001-5929-5120), Soumik Sarkar, Chinmay Hegde (ORCID: 0000-0003-4574-8066), Koray Ozcan (ORCID:0000-0003-3758-9616), Senem Velipasalar (ORCID:0000-0002-1430-1555), Matthew Rizzo (ORCID:0000-0003-3015-0184), Jennifer Merickel (ORCID:000-0002-0394-9110), and Yaw Adu-Gyamfi (ORCID: 0000-0002-1924-9792)		8. Performing Organization Report No.
9. Performing Organization Name and Address Iowa State University 813 Bissell Rd. Ames, IA 50011		10. Work Unit No.
		11. Contract or Grant No.
12. Sponsoring Agency Name and Address Office of Research, Development, and Technology Federal Highway Administration 6300 Georgetown Pike McLean, VA 22101-2296		13. Type of Report and Period Covered Research Summary Report; August 2019–June 2024
		14. Sponsoring Agency Code HRT-10
15. Supplementary Notes The contracting officer's representative was Ana Maria Eigen (HRSO-20; ORCID: 0000-0003-4056-361X). Editorial review, layout, and design were performed by Schatz Publishing Group, 11950 W. Highland Ave., Blackwell, OK 74631.		
16. Abstract Extracting information from naturalistic driving datasets is crucial to understanding driver behavior and distractions. To overcome some of the challenges involved with annotating the data, detecting objects and behaviors, and storing large datasets, the researchers set out to develop a robust platform that can automatically estimate driver state, address detection challenges in naturalistic driving study (NDS) videos, serve as a repository for models, and enhance current and future NDS data. Deep InSight is a comprehensive artificial intelligence platform for data management, modeling, and enhanced annotations. The platform incorporates recurrent neural network models trained to automatically detect and estimate driving behaviors and deal with challenges, such as when a driver looks to the side or down. The researchers also examined video quality enhancement using deep learning models. The cloud-based platform leverages high-performance computing to support end-to-end model development and deployment, which could make it easier for researchers to work together. The Deep InSight platform serves as a repository that research teams can use to test and compare their models with archived models from other research teams working with the same datasets.		
17. Key Words Naturalistic driving, annotations, driver behavior, distraction, head position, maneuvers, models, Deep InSight		18. Distribution Statement No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161. https://www.ntis.gov
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 20 / 22. Price N/A

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized.

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.755	cubic meters	m^3
		NOTE: volumes greater than 1,000 L shall be shown in m^3		
		MASS		
ounces		28.35	grams	g
pounds		0.454	kilograms	kg
short tons (2,000 lb)		0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
		ILLUMINATION		
foot-candles		10.76	lux	lx
foot-Lamberts		3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
poundforce		4.45	newtons	N
poundforce per square inch		6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2,000 lbs)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	2.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbfin^2

*SI is the symbol for International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.
(revised March 2003)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

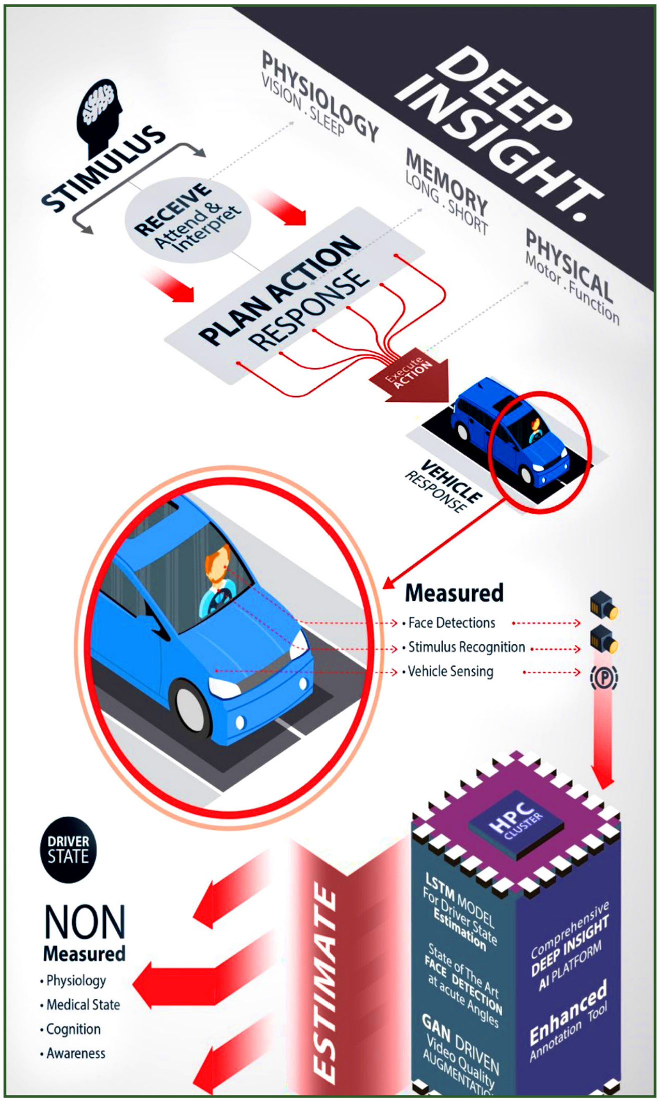
AI	artificial intelligence
BiGRU	bidirectional gated recurrent unit
CLIP	Contrastive Language-Image Pretraining
CNN	convolutional neural network
CV	computer vision
CVAT	Computer Vision Annotation Tool
EAR	Exploratory Advanced Research
EMA	energy-maximization algorithm
FHWA	Federal Highway Administration
GAN	generative adversarial network
GRU	gated recurrent unit
HPC	high-performance computing
LSTM	long short-term memory
MBHL	Mind and Brain Health Labs
ML	machine learning
NDS	naturalistic driving study
RNN	recurrent neural network
SHRP2	second Strategic Highway Research Program
SynDD	synthetic distracted driving
TRL	Technology Readiness Level

Naturalistic driving studies (NDS) are rich data sources for studying driver behavior and transportation safety. Recent NDS data suggest that driver state (e.g., distracted, attentive, sleepy, or angry) contributes to crash risk. A study by Dingus et al. found that driver-related factors exist in almost 90 percent of crashes.(1) Accessing NDS data to evaluate ways to identify driver states and mitigate their dangers is invaluable to researchers. However, NDS datasets are often large and require manual data reduction techniques to identify driver behaviors. Manually annotating the data—where a human reviewer looks for observable indicators of distracted behavior—is time-consuming and expensive and, therefore, limits the study sample size and the ability to understand the role of driver-state variables on crash risk. Automating the annotation process would help researchers fully utilize the available data. A system for storing, mining, visualizing, and analyzing large naturalistic datasets would further overcome the existing challenges and enhance research in this area.

The Federal Highway Administration's (FHWA) Exploratory Advanced Research (EAR) Program supported the project Deep InSight: Deep Extraction of Driver State from Naturalistic Driving Dataset to accomplish the following:(2)

- Develop a robust platform that can automatically detect and estimate driving behaviors.
- Address detection challenges in NDS videos.
- Serve as a repository for models.
- Enhance current and future NDS data.

A research team from Iowa State University, Syracuse University, the University of Missouri, and the University of Nebraska Medical Center designed Deep InSight, a cloud-based, artificial intelligence (AI), driver-state estimation platform to demonstrate an effective approach to analyze large datasets related to driver behavior (figure 1). The platform provides enhanced frame-by-frame video annotations and models for driver-state estimation and behavior.



HPC = high-performance computing; LSTM = long short-term memory; GAN = generative adversarial network.

Figure 1. Illustration. Stages and tools in the Deep InSight platform design.(3)

The platform's end-to-end framework provides the following:

- Data storage and a repository of publicly available and privately collected NDS datasets.
- Integrated tools for custom data annotation and machine-learning (ML) modeling that permit data analysis and inference.

- Recurrent neural network (RNN) models trained to automatically detect and estimate driver behaviors and address detection challenges, such as when a driver looks to the side or down.
- A comprehensive platform that is cost efficient, accessible, scalable, and secure.

The research team also experimented with enhancing the quality of naturalistic driving videos. These videos are typically low resolution and noisy, limiting the accuracy of models trained using clear images. Through their attempts, the researchers learned valuable lessons that could influence future work in this area.

Near the end of the project, Deep InSight was assessed at a Technology Readiness Level (TRL) 4, meaning the components were validated in a laboratory environment.[(4)] The TRL assessment also provided recommendations for next steps, including additional operational and user requirements.

Naturalistic driving studies (**NDS**) are rich data sources for studying driver behavior and transportation safety. Recent **NDS** data suggest that driver state (e.g., distracted, attentive, sleepy, or angry) contributes to crash risk.

The research team wanted the Deep InSight platform to help transportation researchers study driver behavior and analyze NDS data quickly and efficiently without needing to set up or manage the system architecture. The researchers proposed the following features for the platform:

- Large-scale NDS data storage securely shared across multiple universities. Approved users can access and view the data but cannot download the data to limit security concerns.
- An automated annotation pipeline consisting of enhanced ML-augmented data annotation features for expediting model development. This pipeline allows human annotators from across the world to easily access and verify datasets and labels.
- Models that can be easily stored, transferred, and executed without depending on the computing environment.
- Programming language choices for model development without the responsibility of setting up virtual machines or handling data storage and network management.
- NDS data samples for standard training, validation, and test sets across which in-house models and third-party models can be evaluated for benchmarking their performance.

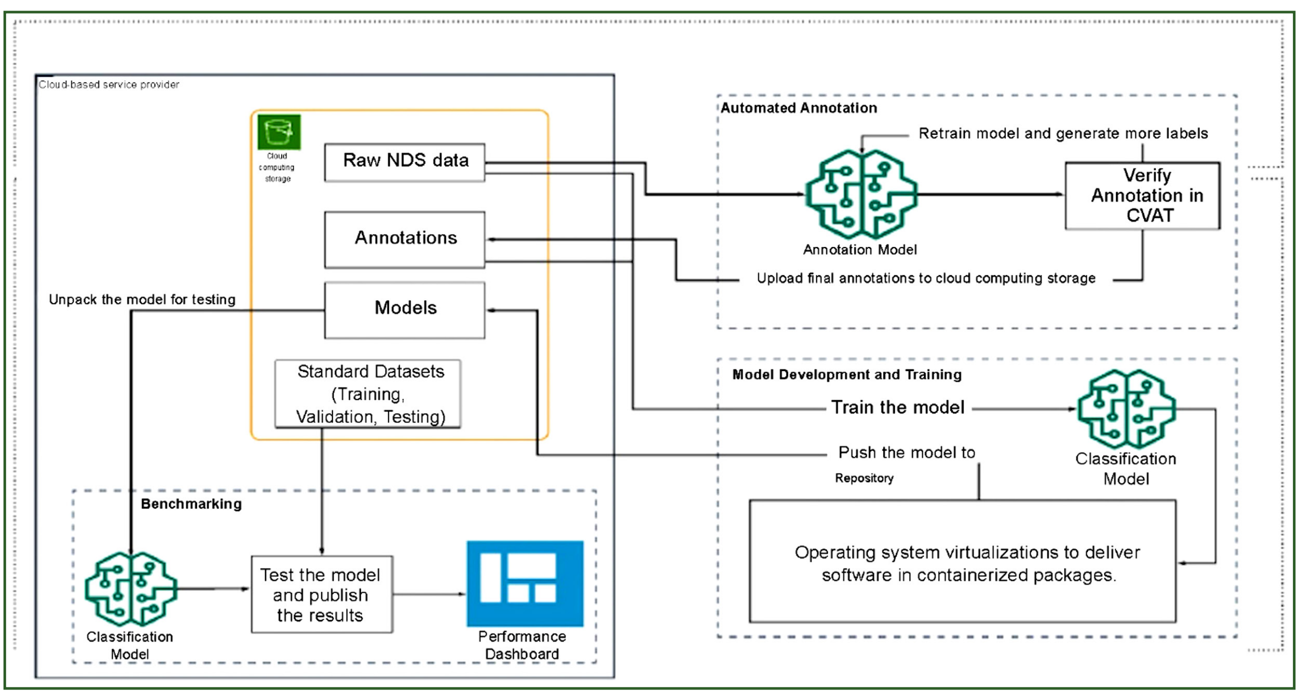
The Deep InSight project consisted of several stages to develop and test the platform and its integrated analysis tools.

PLATFORM DEVELOPMENT

Storage

The first step was to find a system capable of storing large volumes of data. The system also had to be cost efficient and provide security and privacy. The research team chose a cloud-based end-to-end pipeline that would allow researchers to store, annotate, process, visualize, analyze, and benchmark their results via the Internet.

The system stores data as objects in resources called buckets, and a single object can store up to 5 terabytes of data. The Deep InSight framework includes individual buckets for different datasets (raw, processed, and personal health information), annotations, and ML models with separate access controls to protect and secure the data. The ML component of Deep InSight enables data annotation and model training, testing, and deployment. The ML component uses an application that enables researchers to build and train models without having to learn additional cloud computing skills. Figure 2 shows the overall workflow of the Deep InSight framework.



Source: FHWA.

CVAT = Computer Vision Annotation Tool.

Figure 2. Diagram. Deep InSight framework workflow.[(5)]

Datasets

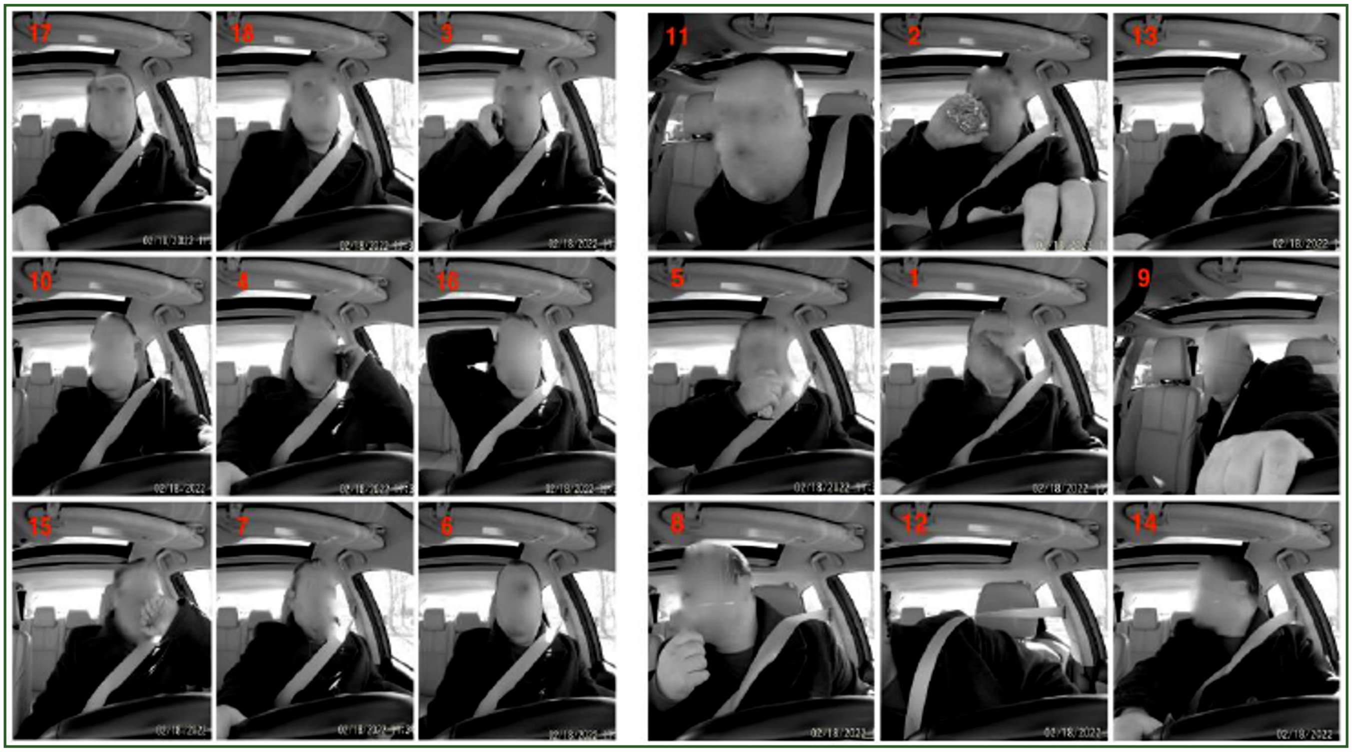
The research team used a naturalistic driving dataset collected by Mind and Brain Health Labs (MBHL) at the University of Nebraska Medical Center to develop, test, and train models for Deep InSight.[6] MBHL collected data from 143 participants who had driven a total of more than 500,000 mi. Similar to the second Strategic Highway Research Program (SHRP2) datasets, the MBHL dataset includes multiple cameras located inside and outside the vehicle and data collected from other vehicle sensors, such as the Global Positioning System, accelerometer, and gyroscope.[7]

In addition, the researchers generated a synthetic distracted driving (SynDD) dataset. The SynDD1 dataset was designed for ML models to identify and analyze distracted behaviors and gaze zones exhibited by drivers.[8] The research team collected high-resolution video data from 100 diverse participants in a stationary vehicle using 3 in-vehicle cameras placed on the dashboard, near the rearview mirror, and on the top right-side window corner.

For the distracted activity, each participant continuously performed 18 distracted driver behaviors (table 1 and figure 3) for a short time interval. For the gaze activity, each participant was instructed to gaze at 1 of 11 different zones in the car, such as the speedometer and control panel. The sequence and duration of each activity were random for each participant. After each participant completed one set of activities, the researchers had them repeat the set while wearing a hat or sunglasses. The researchers manually annotated the dataset for each activity, specifying the activity's start and end times. The SynDD1 dataset was used to evaluate the performance of ML algorithms to classify distracting activities and driver gaze zones and is available for other researchers to use.[8]

Table 1. Distracted driving activities.

Number	Distracted Driver Behavior
1	Normal forward driving
2	Drinking
3	Phone call (right hand)
4	Phone call (left hand)
5	Eating
6	Texting (right hand)
7	Texting (left hand)
8	Hair/makeup
9	Reaching behind
10	Adjusting control panel
11	Picking up object from floor (driver)
12	Picking up object from floor (passenger)
13	Talking to passenger at the right
14	Talking to passenger in back seat
15	Yawning
16	Hand on head
17	Singing with music
18	Shaking or dancing with music



Source: FHWA.

Figure 3. Photos. Study participant performing 18 distracted driving behaviors.[9]

ANNOTATIONS AND LABELING

NDS data provide researchers with valuable insights into real-world driving behaviors. However, these datasets often contain very large amounts of data. For example, the SHRP2 NDS collected 2 petabytes of continuous naturalistic driving data over 3 yr from more than 3,400 vehicles and 3,500 drivers.[7] The extensive data poses challenges for researchers trying to extract and validate variables for analysis. In this phase of the Deep InSight project, the research team demonstrated a process that automatically provides custom annotations to maximize the use of NDS data.

An integral part of computer vision (CV) models, annotations are required to train ML models for feature learning and accurate prediction. The research team set up Deep InSight with an automated annotation and validation pipeline integrated with open-source tools to reduce the manual labor and time required to label the datasets. The researchers used the Computer Vision Annotation Tool (CVAT) to allow frame-by-frame video annotation and labeling.[10] Human annotators verified the model annotations and made any necessary corrections. The modified annotations were used as the training dataset to improve the model's performance. When the model achieved satisfactory accuracy, it was converted to images on a cloud-based platform that allows developers to build, run, and share containerized applications. The images were saved in the Deep InSight repository. The researchers used this automated annotation process to model driver head position and advanced behavioral analysis inside the vehicle in the next phase.

MODELS

The researchers developed several driver-state estimation models to detect head position, driver distraction, and the environment surrounding the driver, such as lane changes, traffic, and traffic control devices. Specifically, the researchers used CV for face detection and recognition at acute angles, such as when the driver is looking down or to the side, and for advanced behavior analysis inside the vehicle.

RNN models (e.g., a neural network that uses sequential or time-series data) trained on the SynDD1 dataset were used to automatically detect and estimate driver behaviors and address detection challenges, such as when a driver looks to the side or down.[8] RNNs are ideal for applications that involve complex interactions and input from multiple sensors, such as those involved in driver-state evaluations. The models the researchers developed required tracking combinations of cues over many video frames from multiple camera views and merging them with vehicle sensor data over time.

Deep InSight made it easier for the researchers to manually check the automated annotations and verify the model's performance. Figure 4 shows the model deployment pipeline. The models the researchers developed were adapted to work within a containerized software package and made available through the Deep InSight platform.

Driver Head Pose Models

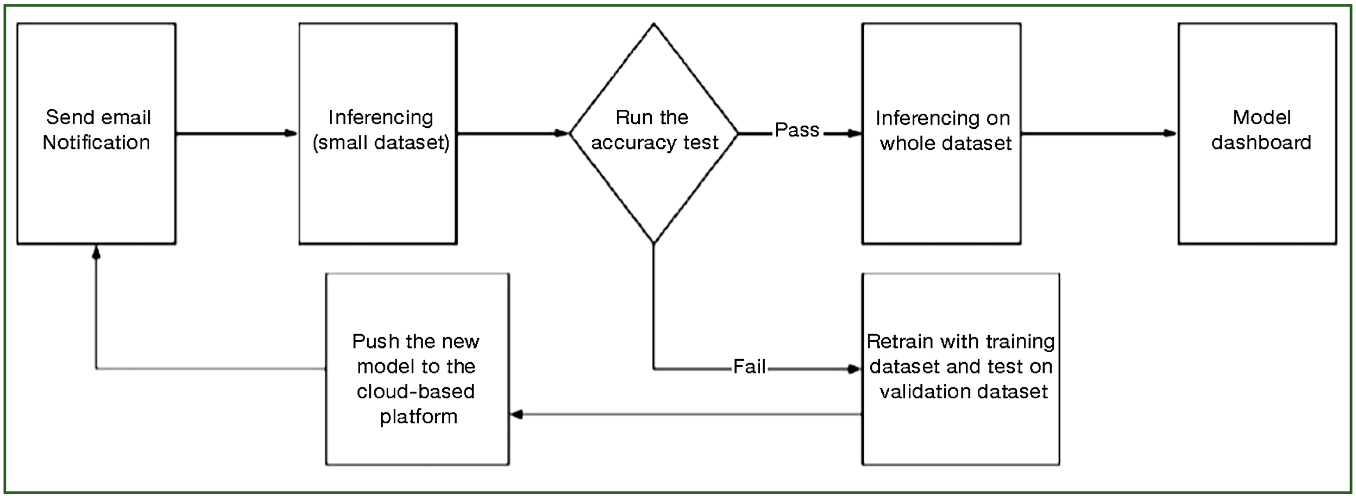
Head pose estimation is a key step in detecting distracted and drowsy driving behaviors because the orientation of a driver's face and head helps researchers understand whether a driver's attention is focused on the road. NDS data often consist of low-resolution videos captured from more challenging camera angles than when a camera is directly facing the driver. In this stage of the project, the researchers' goal was to develop deep learning-driven models for head pose estimation using lower resolution NDS videos from a camera capturing a side view of the driver. The researchers also wanted their model to address the challenges of driver head poses at extreme angles, such as when a driver looks to the side or down.

Methodology

First, the researchers conducted an extensive literature review to establish the current state-of-the-art benchmark performance. Next, they selected a single-stage face detector that could process images in realtime and identify key points on the driver's face.[11] Five facial key points were obtained for each image (figure 5). Then the researchers compared three different approaches to classify a driver's head pose into one of the following categories:

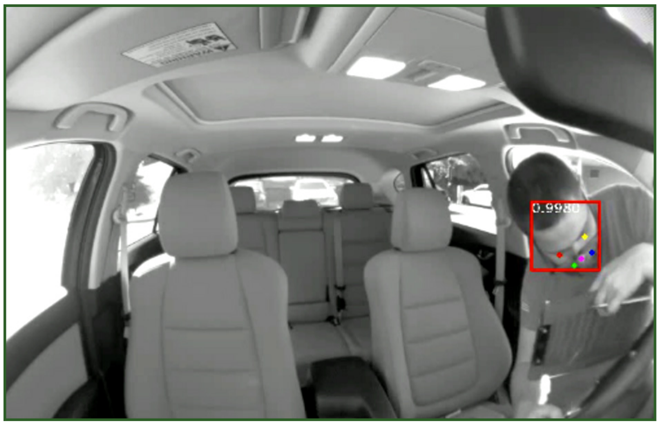
- Looking forward.
- Looking left.
- Looking right.

The test data were collected during day and nighttime settings, and each of the three classes was represented by the same number of video frames.



Source: FHWA.

Figure 4. Diagram. Model deployment pipeline.[5]



© REACTOR Lab.

A. Driver looking down.

© REACTOR Lab.

B. Driver looking to the side and down.

Figure 5. Photos. Facial key points detected under challenging circumstances, such as acute angles and blocked features.[12]

The sequence of facial key points was then sent to a point-processing module and used to compare the performance of three proposed models to each other as well as to the performance of three state-of-the-art baseline models. The three proposed models were as follows:

- PointNet-based approach: A point cloud is a set of three-dimensional (3D) points. This approach was originally introduced to classify 3D point clouds obtained from sensors, such as light detection and ranging and stereo cameras.[13] The researchers adapted PointNet to combine a sequence of facial key points extracted from the images with point cloud processing.
- DeepSet approach: DeepSets were designed for instances where input, and possibly output, data were in sets.[13] This approach has proven efficient for point cloud classification. With DeepSets, changing the order of the input set will not affect the output.
- Bidirectional gated recurrent unit (BiGRU)-based approach: A gated recurrent unit (GRU) is an RNN that uses an update gate and a reset gate to determine what information should be passed to the output.[14] (The update decides which information from past processing steps to pass on, and the reset decides what past information to forget.) A BiGRU consists of two GRU units with two directions (forward and backward) for input, allowing it to capture information from the past and the future of the sequence.[13] The BiGRU was used to encode the sequential information.

Results

To compare the approaches and the baselines, the researchers used leave-one-driver-out cross validation on a dataset using nine different drivers. This method used one driver as a validation set and the rest in the training set. The baseline methods included two video classification methods and one image classification method using convolutional neural network (CNN)-based approaches because most existing face-detection methods use CNN-based techniques. All three of the approaches the researchers tested outperformed the baselines. Of the three approaches, the BiGRU-based approach performed the best in terms of overall accuracy, outperforming the best performing baseline approach by 11 percent.

Using facial key points helped overcome the challenge posed by using a camera capturing a side view of the driver rather than a front view. The researchers were able to develop and test a model with high accuracy for acute-angle face tracking.

Driver Behavior Analysis Models

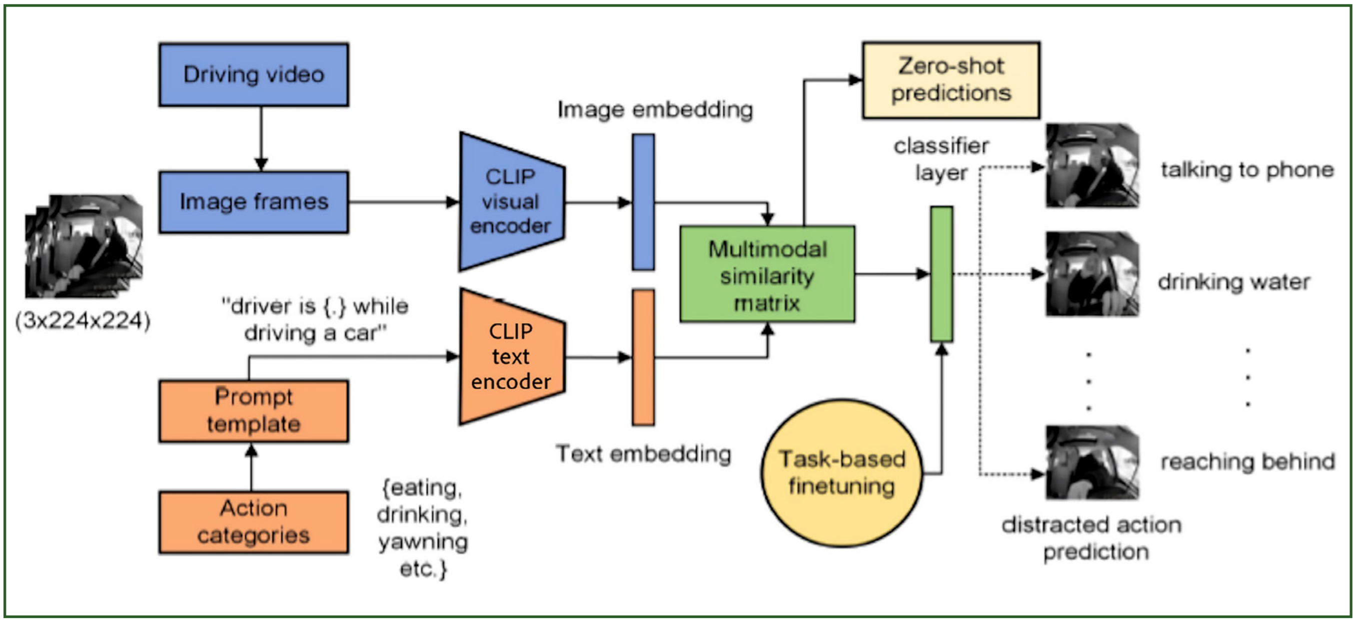
Transportation researchers use CV techniques to automatically analyze NDS video data and categorize various aspects of driver behavior.

For this stage of the project, the research team conducted a literature review to gain insights into CV techniques used for NDS video data and the performance of the CV techniques in classifying driver behavior.

The research team found that CNN and deep neural networks were the techniques most frequently used to predict driver gaze and distracted behavior. These techniques achieved high accuracy and were not affected by the sensor positioning or whether single or multiple cameras were used. For detecting drowsiness, deep-learning techniques performed better than traditional methods. Long short-term memory (a type of RNN capable of learning long-term dependencies in sequential data) and CNN were most often used for detecting lane changes.

Conventional CV methods used to detect distracted driving behaviors often demand extensive supervision through abundant annotated training data. Recently, vision-language models have introduced large-scale pretraining on visual-textual data, enabling adaptation to unsupervised task-specific learning, such as recognizing distracted activities. Vision-language and image-text pretraining models like Contrastive Language-Image Pretraining (CLIP) have demonstrated significant potential for learning visual concepts from natural language supervision.(15) The pretraining in these models can be adapted for specific tasks, such as recognizing distracted driving activities.

The Deep InSight researchers developed a driver activity recognition framework based on CLIP to detect distracted driving, which they called DriveCLIP. Using the pretrained network allowed for zero-shot transfer (e.g., using knowledge from the pretraining to recognize and classify new concepts without labeled examples) for tasks involving driving datasets. The researchers fine tuned a linear classifier and used it to address tasks, such as object detection or action recognition, in the distracted activity task. Figure 6 shows the complete DriveCLIP process.



Source: FHWA.

Figure 6. Flowchart. DriveCLIP architecture for distraction detection.(16)

The researchers tested the DriveCLIP approach on three distinct datasets and conducted the same tests using traditional deep-learning models for comparison. DriveCLIP performed better than the traditional deep models on all three datasets.

Further investigations into the CLIP approach used single- and multiframe CLIP images and integrated a model known as VideoCLIP.[17] These models input images and predict distraction based on actions performed by drivers. The researchers trained and tested the models on two datasets. The single-frame CLIP outperformed traditional models, while the multiframe-based model achieved higher accuracy on one dataset and VideoCLIP on the other.

The researchers explored the impact of different camera angles and integrated sensor information (e.g., gyroscope) on detecting and classifying distracted driving. After exploring various camera angles, the researchers found that the driver-facing view yielded the best performance. Behaviors like eating, yawning, and singing were challenging for frame-based models to distinguish clearly but were handled better by VideoCLIP. The results showed that this framework offered state-of-the-art performance on zero-shot transfer and video-based CLIP for predicting the driver's state on two public datasets.

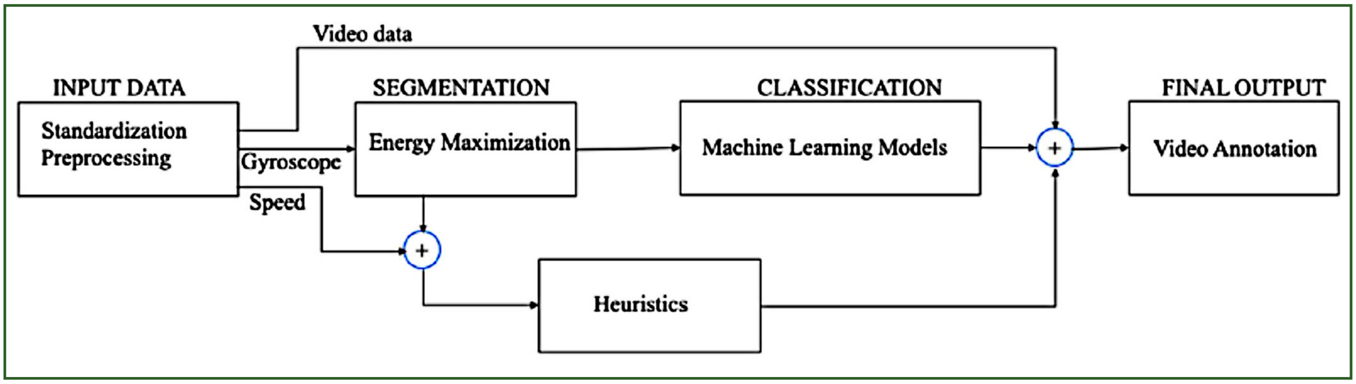
Driver Maneuver Models

Another feature the research team set out to add to the Deep InSight platform was an end-to-end pipeline for automatically annotating NDS videos frame by frame into various driving events, including lane changes, left–right turns, horizontal curve maneuvers, and stop and lane-keeping events. Previous approaches treated vehicle maneuvering as a classification issue, but time series segmentation is also important to these tasks. To address this need, the researchers developed an energy-maximization algorithm (EMA) capable of extracting driving events of varying durations and frequencies from continuous data. Heuristic algorithms (arrived at through trial and error) were used to classify events with highly variable patterns like stops and lane keeping. Four ML models were implemented to classify segmented driving events, and their accuracy and transferability were assessed over multiple data sources.[18]

Methodology

The model's development involved the following five-stage methodology (figure 7):

1. Data preprocessing: The input data was normalized through standard preprocessing.
2. Event segmentation: EMA was used to segment events and nonevents.
3. ML classification: Four ML algorithms were used to classify the events.
4. Heuristics classification: A heuristic algorithm classified nonevents as lane-keeping and stop events.
5. Model output: The final output resulted in frame-by-frame video annotation.



Source: FHWA.

Figure 7. Flowchart. Driver maneuver model development.[18]

Results

The tests indicated that the gyroscope reading is a good parameter to use in extracting driving events. It showed consistent accuracy across all four models. All four models had comparable accuracies to studies that used similar models. The highest accuracy model achieved 98.99 percent, followed by the other models at 97.75 percent, 97.71 percent, and 97.65 percent.

The researchers concluded that implementing a segmentation-classification pipeline significantly improved both the accuracy of driver maneuver detection and the transferability of shallow and deep ML models across diverse datasets. Future work should consider using video data for analyzing distracted driver behavior by using predictive models—such as eye detection and object detection models—to better understand driver behavior.

VIDEO QUALITY ENHANCEMENT

One of the main drawbacks of NDS datasets is low-resolution video captured from affordable sensors and cameras. Nonprofessional-grade cameras result in challenging image conditions, such as low light and “noise” (i.e., random variations in brightness or color that can obscure image details). Models trained using clear images are less accurate when applied to poor-quality videos. CV algorithms that can enhance the brightness of these video files without losing critical details are desirable.

Quality may be further degraded when images are compressed for storage. Some image compression formats discard chunks of an image to reduce its size. The greater the compression, the more obvious the visual losses become, exhibiting chunks across the image known as compression artifacts. Deep-learning models perceive these chunks as features, which affect the model’s ability to classify or predict data. When low light and compression exist in the same image, working with the images becomes even more challenging.

The researchers explored the use of deep-learning generative adversarial network (GAN) models to enhance the video quality.[19] Using two neural networks—the generator and the discriminator—GANs create new data instances that are virtually indistinguishable from the real data in the training set. Given some input and noise, the generator creates images similar to the original image. The discriminator receives both the original and newly generated data and evaluates how realistic the input from the generator seems. As the training continues, the generator learns to improve the images it creates until it can fool the discriminator.

The research team enhanced the quality of low-light videos in the datasets and conducted experiments on vehicle detection and artifact generation. The team used CycleGAN, a GAN variant that consists of two sets of generator-discriminator pairs, to translate daytime images to nighttime images and vice versa to test for detection in low-light scenarios.[20]

The results of the vehicle detection experiments showed that applying denoising and enhancement methods to the NDS data was not straightforward. The team had some success but found that enhancing the low-light images increased the overall brightness and the visibility of noise in the images. The artifact generation experiments resulted in randomly generated light source artifacts in images despite the model accurately translating most of the image features. The researchers determined that well-known low-light enhancement models or denoising models cannot be directly applied to improve detection performance. The lessons learned from these experiments could influence future work on processing low-light images with more accuracy using GAN-based models to generate more accurate low-light images from well-annotated daytime images.

CONCLUSIONS AND NEXT STEPS

The research team built Deep InSight, a comprehensive, cloud-based platform that encourages collaboration and multidisciplinary research on NDS video data in a centralized and standardized destination. The platform is cost efficient, accessible, scalable, and secure. Deep InSight's repository of publicly available and privately collected NDS datasets significantly enhances the scope and impact of NDS by providing integrated tools for data annotation and ML modeling to analyze data and investigate driver behavior.

The advanced models for face detection and recognition, driver behaviors, and driver vehicle maneuvers developed for this project are available through the platform. By serving as a repository for models, research teams can use the platform to test their models and compare them with archived models from other research teams working with the same datasets.

The Deep InSight platform is currently assessed at TRL 4, meaning it has been validated in a laboratory setting. Plans to enhance the platform's functionality and usability include several critical expansions:

- Develop additional operational requirements, such as training and labeling time and computational resources, to optimize the platform's performance.
- Test the platform with NDS datasets from several universities to provide diverse data inputs and validation scenarios.
- Enrich performance metrics to provide comprehensive insights into the operational environment, helping to fine-tune the platform's functionality.
- Measure the framework's impact on research agility, evaluating how much more efficient research processes can become by implementing Deep InSight.
- Address data collection limitations in traditional NDS, including delayed data processing, limited scalability, narrow focus, and insufficient data-mining analysis.
- Enable a new Naturalistic Health and Mobility Data (NMHD) platform. NMHD will provide a cutting-edge, privacy-aware, and economical approach for observing subject mobility and health data.

The research team built **Deep InSight**, a comprehensive, cloud-based platform that encourages collaboration and multidisciplinary research on **NDS** video data in a centralized and standardized destination. The platform is cost efficient, accessible, scalable, and secure.

Aboah, A., Y. Adu-Gyamfi, S. V. Gursoy, J. Merickel, M. Rizzo, and A. Sharma. 2023. “Driver Maneuver Detection and Analysis Using Time Series Segmentation and Classification.” *Journal of Transportation Engineering, Part A: Systems* 149, no. 3. https://doi.org/10.1061/JTEPBS.TEENG-7312, last accessed October 4, 2023.

Chai, W., J. Chen, J. Wang, S. Velipasalar, A. Venkatachalapathy, Y. Adu-Gyamfi, J. Merickel, and A. Sharma. 2023. “Driver Head Pose Detection from Naturalistic Driving Data.” *IEEE Transactions on Intelligent Transportation Systems* 24, no. 9: 9368–9377. https://doi.org/10.1109/TITS.2023.3275070, last accessed October 4, 2023.

Chai, W., J. Wang, J. Chen, S. Velipasalar, and A. Sharma. 2024. “Rethinking the Evaluation of Driver Behavior Analysis Approaches.” *IEEE Transactions on Intelligent Transportation Systems* 25, no. 8: 9958–9966. https://ieeexplore.ieee.org/document/10417763, last accessed September 19, 2024.

Hasan, M. Z., A. Joshi, M. Rahman, A. Venkatachalapathy, A. Sharma, C. Hegde, and S. Srakar. 2022. “DriveCLIP: Zero-Shot Transfer for Distracted Driving Activity Understanding Using CLIP.” Presented at *Machine Learning for Autonomous Driving Workshop at the 36th Conference on Neural Information Processing Systems* (NeurIPS 2022). New Orleans, LA: Neural Information Processing Systems. https://dr.lib.iastate.edu/server/api/core/bitstreams/6caedeff-3a7d-46bc-aa80-64fbcc5435cc/content, last accessed October 4, 2023.

Hasan, M. Z., J. Chen, J. Wang, A. Joshi, S. Velipasalar, C. Hegde, A. Sharma, and S. Srakar. 2024. “Vision-Language Models Can Identify Distracted Driver Behavior From Naturalistic Videos.” *IEEE Transactions on Intelligent Transportation Systems* 25, no. 9: 11602–11616. https://ieeexplore.ieee.org/document/10492662, last accessed September 19, 2024.

Naphade, M., S. Wang, D. C. Anastasiu, Z. Tang, M-C Chang, Y. Yao, L. Zheng, et al. 2023. “The 7th AI City Challenge.” Presented at the 2023 *IEEE/CVF Computer Vision and Pattern Recognition Conference*. Vancouver, BC: IEEE. https://openaccess.thecvf.com/content/CVPR2023W/AICity/papers/Naphade_The_7th_AI_City_Challenge_CVPRW_2023_paper.pdf, last accessed October 4, 2023.

Naphade, M., S. Wang, D. C. Anastasiu, Z. Tang, M-C Chang, Y. Yao, L. Zheng, et al. 2022. “The 6th AI City Challenge.” Presented at the 2022 IEEE/CVF Computer Vision and Pattern Recognition Conference. New Orleans, LA: IEEE. https://openaccess.thecvf.com/content/CVPR2022W/AICity/papers/Naphade_The_6th_AI_City_Challenge_CVPRW_2022_paper.pdf, last accessed October 4, 2023.

Rahman, M. S., A. Venkatachalapathy, A. Sharma, J. Wang, S. V. Gursoy, D. Anastasiu, and S. Wang. 2023. “Synthetic Distracted Driving (SynDD1) Dataset for Analyzing Distracted Behaviors and Various Gaze Zones of a Driver.” *Data in Brief* 46: 108793. https://www.sciencedirect.com/science/article/pii/S2352340922009969, last accessed October 4, 2023.

Rahman, M. S., A. Venkatachalapathy, and A. Sharma. 2022. *Synthetic Distracted Driving (SynDD1) Dataset*. Version 2. Mendeley Data. https://data.mendeley.com/datasets/ptcp7rp3wb/4, last accessed October 4, 2023.

Venkatachalapathy, A., M. S. Rahman, A. Raj, J. Merickel, A. Sharma, J. Wang, and S. Velipasalar. 2023. “Deep InSight: A Cloud Based Big Data Analytics Platform for Naturalistic Driving Studies.” *International Journal of Automotive Engineering* 14, no. 3: 66–76. https://www.jstage.jst.go.jp/article/jsaeijae/14/3/14_20234415/_article, last accessed July 2, 2024.

REFERENCES

1. Dingus, T. A., F. Guo, S. Lee, J. F. Antin, M. Perez, M. Buchanan-King, and J. Hankey. 2016. "Driver Crash Risk Factors and Prevalence Evaluation Using Naturalistic Driving Data." *Proceedings of the National Academy of Sciences* 113, no. 10: 2636–2641. https://www.pnas.org/doi/full/10.1073/pnas.1513271113, last accessed June 6, 2024.

2. FHWA. n.d. "Deep InSight: Deep Extraction of Driver State from Naturalistic Driving Dataset" (web page). https://highways.dot.gov/research/projects/deep-insight-deep-extraction-driver-state-naturalistic-driving-dataset, last accessed June 6, 2024.

3. FHWA. 2022. *The Role of Artificial Intelligence and Machine Learning in Federally Supported Surface Transportation: 2022 Updates*. Report No. FHWA-HRT-22-026. Washington, DC: Federal Highway Administration. https://www.fhwa.dot.gov/publications/research/ear/22026/22026.pdf, last accessed July 2, 2024.

4. Deshmukh Towery, N., E. Machek, and A. Thomas. 2017. *Technology Readiness Level Guidebook*. Report No. FHWA-HRT-17-047. Washington, DC: Federal Highway Administration. https://www.fhwa.dot.gov/publications/research/ear/17047/17047.pdf, last accessed June 18, 2024.

5. Venkatachalapathy, A., M. S. Rahman, A. Raj, J. Merickel, A. Sharma, J. Wang, and S. Velipasalar. 2023. "Deep InSight: A Cloud Based Big Data Analytics Platform for Naturalistic Driving Studies." *International Journal of Automotive Engineering* 14, no. 3: 66–76. https://www.jstage.jst.go.jp/article/jsaeijae/14/3/14_20234415/_article, last accessed July 2, 2024.

6. University of Nebraska Mind and Brain Health Labs. n.d. *Naturalistic Driving Dataset* (dataset).

7. Virginia Tech Transportation Institute. 2020. "InSight Data Access Website: SHRP2 Naturalistic Driving Study" (website). https://insight.shrp2nds.us/, last accessed April 29, 2024.

8. Rahman, M. S., A. Venkatachalapahy, and A. Sharma. 2022. *Synthetic Distracted Driving (SynDD1) Dataset* (dataset). Version 4. https://data.mendeley.com/datasets/ptcp7rp3wb/4, last accessed June 24, 2024.

9. Rahman, M. S., A. Venkatachalapathy, A. Sharma, J. Wang, S. V. Gursoy, D. Anastasiu, and S. Wang. 2023. "Synthetic Distracted Driving (SynDD1) Dataset for Analyzing Distracted Behaviors and Various Gaze Zones of a Driver." *Data in Brief* 46: 108793. https://www.sciencedirect.com/science/article/pii/S2352340922009969, last accessed July 9, 2024.

10. CVAT.ai Corporation. 2024. "Open Data Annotation Platform" (website). https://www.cvat.ai/, last accessed June 24, 2024.

11. Deng, J., J. Guo, Y. Zhou, J. Yu, I. Kotsia, and S. Zafeiriou. 2019. *Retinaface: Single-Stage Dense Face Localisation in the Wild*. https://doi.org/10.48550/arXiv.1905.00641, last accessed June 28, 2024.

12. FHWA. 2022. *The Role of Artificial Intelligence and Machine Learning in Federally Supported Surface Transportation: 2022 Updates*. Publication No. FHWA-HRT-22-026. Washington, DC: Federal Highway Administration.

13. Chai, W., J. Chen, J. Wang, S. Velipasalar, A. Venkatachalapathy, Y. Adu-Gyamfi, J. Merickel, and A. Sharma. 2023. "Driver Head Pose Detection from Naturalistic Driving Data." *IEEE Transactions on Intelligent Transportation Systems* 24, no. 9: 9368–9377. https://ieeexplore.ieee.org/document/10128799, last accessed June 28, 2024.

14. Kostadinov, S. 2017. "Understanding GRU Networks." *Towards Data Science*. December 16, 2017. https://towardsdatascience.com/understanding-gru-networks-2ef37df6c9be, last accessed June 28, 2024.

15. Radford, A., J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, et al. 2021. “Learning Transferable Visual Models from Natural Language Supervision.” In *Proceedings of the 38th International Conference on Machine Learning*. Netherlands: ML Research Press. https://arxiv.org/pdf/2103.00020, last accessed July 10, 2024.

16. Hasan, M. Z., A. Joshi, M. Rahman, A. Venkatachalapathy, A. Sharma, C. Hegde, and S. Sarkar. 2022. “DriveCLIP: Zero-Shot Transfer for Distracted Driving Activity Understanding Using CLIP.” Presented at the *36th Conference on Neural Information Processing Systems*. New Orleans, LA: NeurIPS Foundation.

17. H. Xu, G. Ghosh, P-Y. Huang, D. Okhonko, A. Aghajanyan, F. Metze, L. Zettlemoyer, et al. 2021. “VideoCLIP: Contrastive Pre-Training for Zero-Shot Video-Text Understanding.” In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 6787–6800. Kerrville, TX: Association for Computational Linguistics.

18. A. Aboah, Y. Adu-Gyamfi, S. V. Gursoy, J. Merickel, M. Rizzo, and A. Sharma. 2022. “Driver Maneuver Detection and Analysis Using Time Series Segmentation and Classification.” *Journal of Transportation Engineering, Part A: Systems* 149, no. 3. https://ascelibrary.org/doi/full/10.1061/JTEPBS.TEENG-7312, last accessed July 11, 2024.

19. Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. 2014. “Generative Adversarial Nets.” *Advances in Neural Information Processing Systems* 3, no. 11. https://proceedings.neurips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf, last accessed June 21, 2024.

20. Zhu, J.-Y., T. Park, P. Isola, and A. A. Efros. 2017. “Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks.” In *Proceedings of the IEEE International Conference on Computer Vision*. Piscataway, NJ: IEEE, 2223–2232.

Getting Involved With the EAR Program

To take advantage of a broad variety of scientific and engineering discoveries, the EAR Program involves both traditional stakeholders (State department of transportation researchers, University Transportation Center researchers, and Transportation Research Board committee and panel members) and nontraditional stakeholders (investigators from private industry, related disciplines in academia, and research programs in other countries) throughout the research process.

Learn More

For more information, see the EAR Program website at https://highways.dot.gov/research/research-programs/exploratory-advanced-research/exploratory-advanced-research-overview. The site features information on research solicitations, updates on ongoing research, links to published materials, summaries of past EAR Program events, and details on upcoming events.

EAR Program Results

As a proponent of applying ideas across traditional research fields to stimulate new problem-solving approaches, the EAR Program strives to develop partnerships with the public and private sector. The program bridges basic research (e.g., academic work funded by National Science Foundation grants) and applied research (e.g., studies funded by State DOTs). In addition to sponsoring projects that advance the development of highway infrastructure and operations, the EAR Program is committed to promoting cross-fertilization with other technical fields, furthering promising lines of research and deepening vital research capacity.

EXPLORATORY ADVANCED **RESEARCH**

U.S. Department of Transportation
Federal Highway Administration

Recommended citation: Federal Highway Administration, *Deep InSight: A Driver-State Estimation Platform for Processing Naturalistic Driving Data* (Washington, DC: 2024)
https://doi.org/10.21949/1521583

FHWA-HRT-24-137
HRT-10/12-24(WEB)E