Advanced Distracted Driving Detection: Deep Learning Methods for Cell Phone Usage Recognition

DDI 2024 Presentation
Oct 24, 2024

Zifei Wang (Presenter), Meitang Li, Huizhong Guo, Carol Flannagan, Jim Sayer, Shan Bao

Department of Industrial and Manufacturing Systems Engineering , University of Michigan-Dearborn,
University of Michigan Transportation Research Institute

Contents

Introduction & Background

- Driver distraction has been a serious public safety concern. In 2022, 3308 people were killed in distraction-involved crashes (NCSA, 2024).

- Cell phone use, as a major type of in-vehicle distraction, is associated with increased crash potential. Texting, for example, diverts drivers' eyes, hands, and attention from driving (Simons-Morton et al., 2014).

- Numerous distraction detection systems have been developed, including:
  - Image based classification (e.g., head, eye-gaze, posture, hand position, etc.)
  - Kinematic based identification (e.g., mean shifts of lateral speeds, steering wheel reversal, etc.)

Research Goal: Identify whether drivers are engaged in secondary tasks (e.g., cell phone use) through computer vision and machine learning methods.

The proposed methodology to identify driver distraction includes analysis of:

1. In-cabin video image, and
2. Vehicle kinematic features

Method – Data

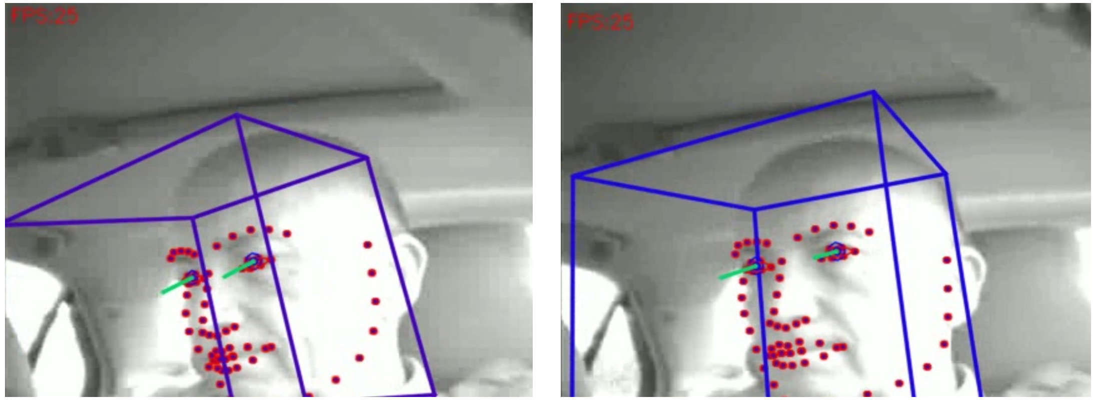
- This study used data from UMTRI's **Integrated In-Vehicle Based Safety System Study (IVBSS)** (Sayer et al., 2011).

- A set of **distracted** driving episodes and matched **undistracted** data from the same driver were extracted and used.
  - 932 events of manually labeled distracted driving events from 58 drivers are considered.

- Three types of data were extracted from IVBSS:
  - Face videos (head and eye position),
  - Cabin videos (hand movements), and
  - Vehicle kinematic features (longitudinal speed, longitudinal acceleration, steering angle, and lateral speed) at 10 Hz.

Method – Video Image-Based Detection

- ***OpenFace:***
  - A neural network-based face recognition tool in Python;
  - Distraction identification:
    1) Detect driver face position using the face video data and calculate **head pose**;
    2) Determine **eye-gaze location** relative to head pose via coordinate transformation (gaze behavior).
    3) If the calculated relative eye gaze location is facing **downward** or **sideward**, the algorithm outputs as distracted.



(a) *distracted (looking off the road)* ***(b)*** *non-distracted (looking on the road)*

Limitation: ignore distraction when driver's face looking straight.

- ***YOLOv4:***
  - Part of the algorithm family of one-stage object detectors (*You only look once*);
  - Cell phone detection:
    1) Detect **cell phone location** relative to the driver's body
    2) If the cell phone is detected **close to or in the driver's hand**, the algorithm outputs the driver is distracted

Method – Kinematic Feature-Based Detection

Three machine learning methods were applied on the **vehicle kinematic features**:

1) Hidden Markov Model (HMM);
2) Support Vector Classification (SVC);
3) Long Short-Term Memory (LSTM).

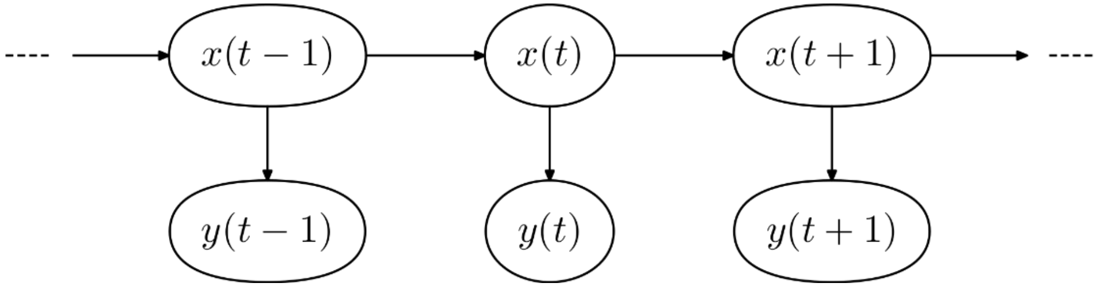
Kinematic space:

- longitudinal speed,
- longitudinal acceleration,
- steering angle, and
- lateral speed

Method – Kinematic Feature-Based Detection

Three machine learning methods were applied on the **vehicle kinematic features**:

1) **Hidden Markov Model (HMM):** a statistics-based model aiming at inferencing hidden states from influenced observations. The HMM:
   - Takes input of a sequence of time-series driving statistics: $\{\dots, y(t-1), y(t), y(t+1), \dots\}$;
   - Decides the corresponding states of distraction: $\{\dots, x(t-1), x(t), x(t+1), \dots\}$;



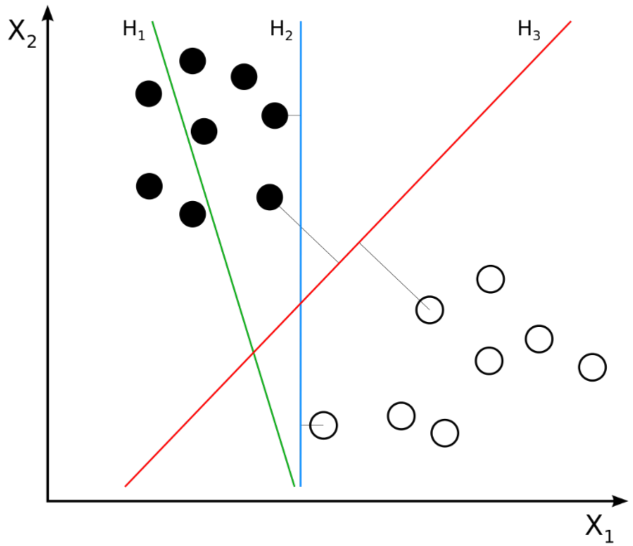
   - The hidden state is a discrete variable indicating the state of distraction:
     - {Distracted, Not Distracted}
     - {Pre-distracted, Distracted, Post-distracted}
   - Each driver is assigned an HMM specifically trained on their driving data.

Kinematic space:
- longitudinal speed,
- longitudinal acceleration,
- steering angle, and
- lateral speed

Method – Kinematic Feature-Based Detection

Three machine learning methods were applied on the **vehicle kinematic features**:

2) **Support Vector Classification (SVC):** a supervised machine learning algorithm that can separate data points based on their possible differences:
   - Each data sample is composed of either 5, 10 or 15 timestamps;
   - The associated distraction state is 5 timestamps after the last one from the input data sample, to be able to make early detections;
   - Each driver is assigned an SVC specifically trained on their driving data.



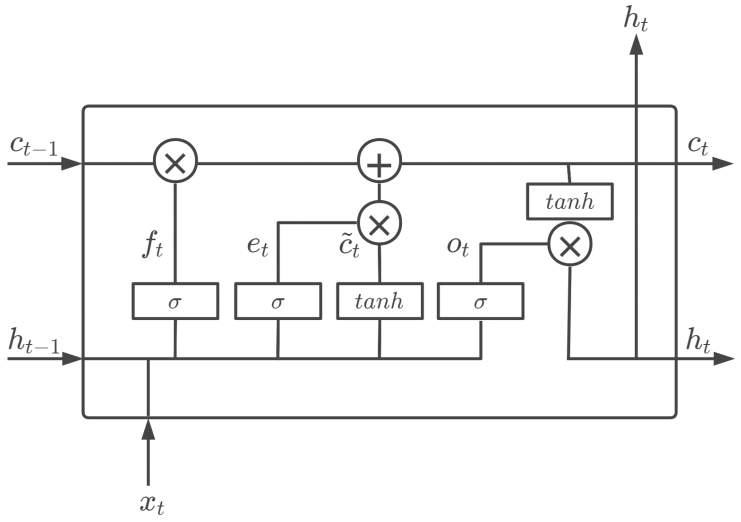
SVC Model Structure (Graham, S. S., & Hopkins)

Kinematic space:
- longitudinal speed,
- longitudinal acceleration,
- steering angle, and
- lateral speed

Method – Kinematic Feature-Based Detection

Three machine learning methods were applied on the **vehicle kinematic features**:

3) **Long Short-Term Memory (LSTM):** a recurrent neural network architecture to model sequences and long-term dependencies effectively. LSTM cells include three main gates:
   - Forget (f_t), Input (e_t), and Output (o_t) gates, which regulate the flow of information and help preserve the error that can be backpropagated through time and layers;
   - This architecture helps to mitigate the vanishing error problem, allowing the model to learn long-term dependencies in sequential data.



LSTM Model Structure (Fu et al., 2022)

Kinematic space:
- longitudinal speed,
- longitudinal acceleration,
- steering angle, and
- lateral speed

Results – Video Image-Based Detection

- This study included 300 IVBSS video clips, within which 296 clips are valid, including:
  - **932 pre-defined events** from 58 drivers
- Each event is detected through OpenFace or YOLOv4, see performance in Table 1.
- In the combined results, if either model was detected “distracted” for the data sample:
  - final output will be "distracted."

Table 1 *Video Processing Results*

	Accuracy	F1	Precision	Recall
OpenFace	0.621	0.666	0.711	0.725
YOLOv4	0.344	0.122	0.449	0.086
Combined	0.628	0.676	0.710	0.739

Results – Kinematic Feature Prediction

1. HMM Results

- The four vehicle kinematic features were analyzed through machine learning models, with parameters tuned via 3-fold cross-validation.
- Table 2 shows the best models of:
  - Two hidden states: {Distracted, Not Distracted} and
  - Three hidden states: {Pre-Distracted, Distracted, Post-Distracted}.

Table 2 *HMM Results*

	Accuracy	F1	Precision	Recall
2 states	0.431	0.370	0.580	0.348
3 states	0.417	0.340	0.572	0.311

Kinematic space:
- longitudinal speed,
- longitudinal acceleration,
- steering angle, and
- lateral speed

- Due to poor performance, the HMM was not considered further.

Results – Kinematic Feature Prediction

2. SVC Results

- Table 3 shows the performance of SVC models for three different setups:

***Table 3** SVC Results*

	Accuracy	F1	Precision	Recall
5 timestamps	0.751	0.788	0.720	0.931
10 timestamps	0.761	0.788	0.733	0.916
15 timestamps	0.784	0.806	0.745	0.935

Kinematic space:
- longitudinal speed,
- longitudinal acceleration,
- steering angle, and
- lateral speed

- SVC models have better performance than video processing and HMM results.
- The SVC model with **15 timestamps** achieved best performance.
  - However, biased results is noticed based on the 20% gap between precision and recall.

Results – Kinematic Feature Prediction

3. LSTM Results

- Table 4 shows the performance of LSTM models for three different setups:

***Table 4** LSTM Results*

	Accuracy	F1	Precision	Recall
5 timestamps	0.813	0.829	0.778	0.920
10 timestamps	0.826	0.841	0.790	0.932
15 timestamps	0.845	0.854	0.810	0.937

Kinematic space:
- longitudinal speed,
- longitudinal acceleration,
- steering angle, and
- lateral speed

- LSTM models have best performance so far, with average accuracy > 81%.
- The LSTM model with **15 timestamps** achieved best performance.
  - While not promoting recall, the model increased precision by 6%
    → better ability to deal with biased database.

Results – Model Improvement

1. Combination of Video Data and Kinematic Data

- To improve accuracy, the combined video prediction results was included as an input.
- For SVC models, adding the video data did not improve data significantly.
- For LSTM models, an average of 6% increment was found when adding the video data.
  - LSTM with 15 timestamps and video results achieved best performance.

***Table 5** Combination of Video Data and Driving Data Results*

	Accuracy	F1	Precision	Recall
SVC 5	0.751	0.788	0.720	0.931
SVC 5+*	0.754	0.791	0.723	0.933
SVC 10	0.761	0.788	0.733	0.916
SVC 10+	0.771	0.801	0.733	0.935
SVC 15	0.784	0.806	0.745	0.935
SVC 15+	0.788	0.810	0.746	0.937
LSTM 5	0.813	0.829	0.778	0.920
LSTM 5+	0.870	0.869	0.839	0.923
LSTM 10	0.826	0.841	0.790	0.932
LSTM 10+	0.887	0.883	0.857	0.931
LSTM 15	0.845	0,854	0.810	0.937
LSTM 15+	0.910	0.904	0.882	0.944

*: "+" indicates adding video prediction as a variable

Results – Model Improvement

2. Extra Kinematic Variables

- Extra variables were added to the feature space for better predictions:

Old Kinematic Space:
- longitudinal speed,
- longitudinal acceleration,
- steering angle,
- lateral speed

New Kinematic Space:
- longitudinal speed,
- longitudinal acceleration,
- steering angle,
- lateral speed
- lateral acceleration
- lane offset
- yaw rate
- the variance within 1 second for above
- traffic
- video prediction

Results – Model Improvement

2. Extra Kinematic Variables

- Extra variables were added to the feature space for better predictions:

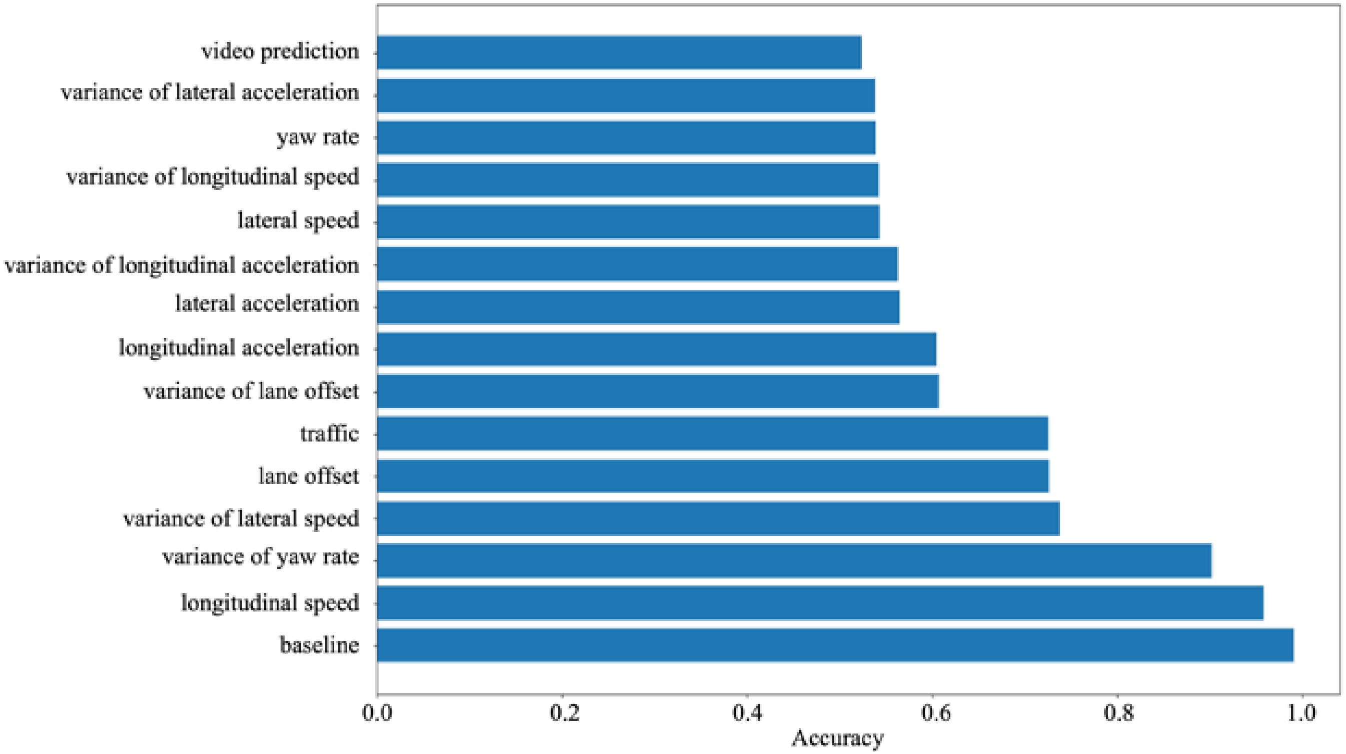
Table 6 *Additional Driving Data Results*

	Accuracy	F1	Precision	Recall
SVC (10,old)	0.771	0.801	0.733	0.935
SVC (10,new)	0.772	0.813	0.747	0.953
SVC (15,old)	0.788	0.810	0.746	0.937
SVC (15,new)	0.797	0.826	0.768	0.944
LSTM (5,old)	0.870	0.869	0.839	0.923
LSTM (5,new)	0.911	0.904	0.895	0.923
LSTM (10,old)	0.887	0.883	0.857	0.931
LSTM (10,new)	0.944	0.940	0.930	0.958
LSTM (15,old)	0.910	0.904	0.882	0.944
LSTM (15,new)	0.957	0.954	0.946	0.969

New Kinematic Space:
- longitudinal speed,
- longitudinal acceleration,
- steering angle,
- lateral speed
- lateral acceleration
- lane offset
- yaw rate
- the variance within 1 second for above
- traffic
- video prediction

Results – Feature Importance

- A permutation importance analysis was conducted using the best LSTM.
- “Baseline”: the model's accuracy without permutation (reference point).
- Video prediction was the most important feature,
  - followed by the variance of lateral acceleration and yaw rate.



Discussion & Conclusion

- **Video-Only Processing Limitations:**
  - IVBSS dataset was monochromatic and of low resolution; better results expected with colored and higher resolution videos.
- **Kinematic Data Prediction Performance:**
  - HMM: 43% accuracy, possibly due to inconsistency of distraction event length.
  - SVC (15 timestamps): 78.4% accuracy, 93.5% recall, but only 74.5% precision, indicating a bias in learning.
  - LSTM (15 timestamps): 83% accuracy, improving precision to 81.0% while maintaining 93.7% recall, reducing bias.
- **Video Prediction Result as Input:**
  - SVC performance did not significantly improve with the video prediction results.
  - LSTM gained 6% increase in accuracy with video input, achieving better precision and recall.

Discussion & Conclusion

- The top 3 important kinematic features are: variance of lateral speed, yaw rate, and variance of longitudinal speed, which indicates that distraction is more related to control instability, especially in lateral directions.
- The integration of the computer vision and machine learning methods significantly enhanced the robustness of cell phone-related distraction predictions.
- The findings of the research provide valuable insights for better in-vehicle distraction detection system design, promoting a safer road environment for all.

Acknowledgement

The authors would like to acknowledge Federal Highway Administration (FHWA) for sponsoring and supporting this project.

Any opinions, findings, and conclusions, or recommendations expressed in this paper/presentation are those of the authors and do not necessarily reflect the views of FHWA.

Reference

1. National Center for Statistics and Analysis. (2024, April). Distracted driving in 2022 (Research Note. Report No. DOT HS 813 559). National Highway Traffic Safety Administration.
2. Simons-Morton, B. G., Guo, F., Klauer, S. G., Ehsani, J. P., & Pradhan, A. K. (2014). Keep your eyes on the road: Young driver crash risk increases according to duration of distraction. Journal of Adolescent Health, 54(5), S61-S67.
3. Sayer, J., LeBlanc, D., Bogard, S., Funkhouser, D., Bao, S., Buonarosa, M. L., & Blankespoor, A. (2011). Integrated vehicle-based safety systems field operational test: Final program report (No. FHWA-JPO-11-150; UMTRI-2010-36). United States. Joint Program Office for Intelligent Transportation Systems.
4. Fu, X., Meng, H., Wang, X., Yang, H., & Wang, J. (2022). A hybrid neural network for driving behavior risk prediction based on distracted driving behavior data. PLoS one, 17(1), e0263030.
5. Graham, S. S., & Hopkins, H. R. (2022). AI For social justice: New methodological horizons in technical communication. Technical Communication Quarterly, 31(1), 89-102.

Thanks for Listening!

Contact:
Zifei Wang (Presenter), zifwang@umich.edu
Dr. Huizhong Guo, hzhguo@umich.edu
Dr. Shan Bao, shanbao@umich.edu