# Multi-Head Attention Machine Learning for Fault Classification in Mixed Autonomous and Human-Driven Vehicle Platoons

Theodore Wu[1], Satvick Acharya[1], Abdelrahman Khalil[2], Ahmad F. Aljanaideh[3], Mohammad Al Janaideh[2], Deepa Kundur[1]
[1]Department of Electrical and Computer Engineering, University of Toronto; [2]Department of Mechanical Engineering, Memorial University; [3]Bentley University

## Motivation

- Connected Autonomous Vehicle (CAV) platoons enable high awareness in autonomous vehicles through wireless communication networks.
- Vulnerable to physical faults and cyber attacks, and must also coexist with unpredictable human drivers.
- Need for a fault classification mechanism to bridge existing fault detection and fault mitigation methods.

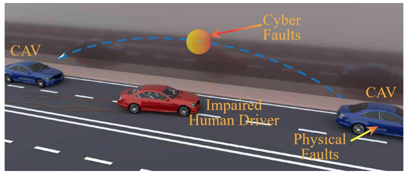



**Figure 1:** An illustration of a faulty CAV platoon operating alongside an impaired human driver.

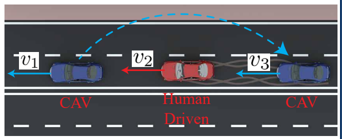
## Healthy Platoon Model

- Require a simulation model to generate training data.
- Simulate CAV with a cruise control law.
- Simulate human driver with the Intelligent Driver Model.

**CAV Model:**

$$\frac{V_i(s)}{V_i^*(s)} = \frac{\delta_i s + \epsilon_i}{s^3 + \alpha_i s^2 + \beta_i s + \gamma_i}$$

$$i \in \{1, 3\}$$




**Human Driver Model:**

$$a_2(t, v_2) = a_{\max}\left[1 - \left(\frac{v_2(t)}{v_2^*(t)}\right)^{\lambda} - \left(\frac{s_2^*(t, v_2)}{s_2(t)}\right)^2\right]$$

$$s_2^*(t, v_2) = s_0 + \max\left(0, v_2(t)T + \frac{v_2(t)\Delta v_2(t)}{2\sqrt{b^* a_{\max}}}\right)$$

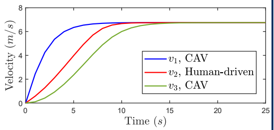



**Figure 2:** The mixed CAV and human-driven vehicle platoon model (Top) and an example of the simulated velocity response (Bottom).

## Faulty Platoon Simulation

- Simulate three classes of CAV faults and two types of human driver abnormalities by altering the healthy platoon model.
- Full coverage of faults and abnormalities for physical, cyber, and human-driver layers of the mixed vehicle platoon.

**CAV Faults:**

- Actuator Fault
- False Data Injection (FDI)
- Denial-of-Service (DoS)

**Human Driver Abnormalities:**

- Distracted Driver
- Drunk Driver

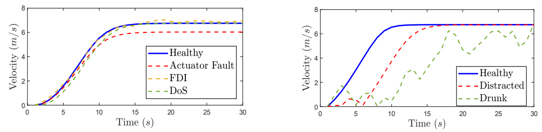



**Figure 3:** Faulty vehicle responses under each fault class. (Left) CAV faults in third vehicle. (Right) Human driver abnormalities in human vehicle.

## Multi-Head Attention Machine Learning

- Leverage large amounts of simulation data and train a machine learning model to identify the fault class.
- Propose a Multi-Head Attention Machine Learning (MHA-ML) model that uses scaled-dot-product attention to identify the most informative timesteps in the input velocity sequence.

$$\mathcal{A}(Q, K, V) = \text{softmax}\left(\frac{1}{\sqrt{d_k}} Q K^T\right) V$$

- Only using platoon velocities allows the MHA-ML approach to be platoon model dynamics independent.
- MHA-ML is trained on 5000 samples of 500s runs across each fault class and achieves 92.4% validation accuracy.

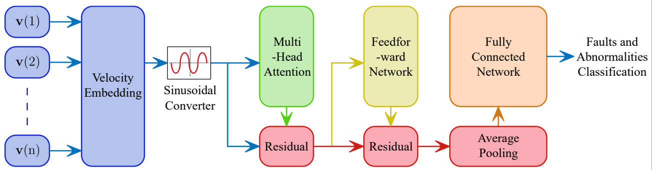



**Figure 4:** Proposed multi-head attention network architecture for mixed CAV platoon fault and abnormality classification.

## Experimental Results

- Lab platoon of Quanser Qbots used as a real-world test.
- Internal PI controller for CAV vehicles, human-driven vehicle controlled via external keyboard.
- Compare MHA-ML with an LSTM baseline network.

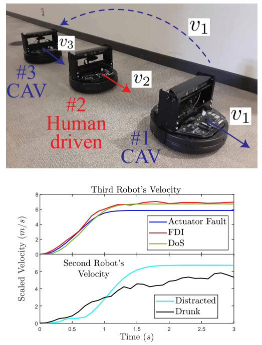



**Figure 5:** Experimental platoon setup (Top) and an example of the emulated faults (Bottom).

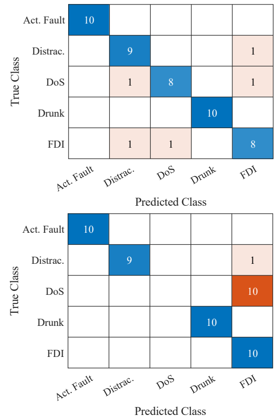



**Figure 6:** Confusion matrix predictions on experimental data for MHA-ML (Top) and LSTM baseline (Bottom).

## Discussion

- MHA-ML scores 90% test accuracy and outperforms the LSTM baseline network that scores 78% test accuracy.
- Most significant source of misclassification between FDI and DoS fault classes due to similarity in responses.

## Conclusion

- We introduce a new MHA-ML deep learning architecture for CAV fault classification.
- Our algorithm can extend CAV fault classification to a mixed vehicle environment.
- Validation on real-world experimental platoon with unknown model dynamics shows generalizability of MHA-ML.

The Edward S. Rogers Sr. Department
of Electrical & Computer Engineering
UNIVERSITY OF TORONTO