Utilizing Behind-the-Wheel Behavior for Driver Authentication

Jonathan Voris
jvoris@nyit.edu
Computer Science Department

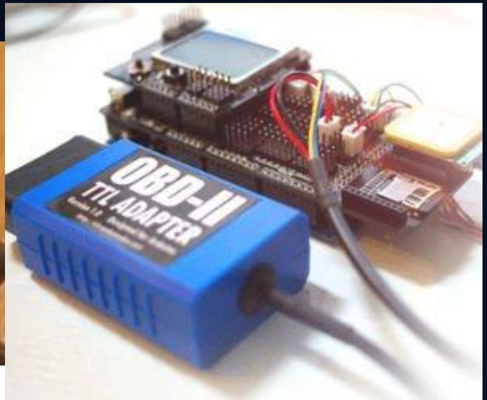
N. Sertac Artan
nartan@nyit.edu
Electrical & Computer Engineering Department

Wenjia Li
wli20@nyit.edu
Computer Science Department

New York Institute of Technology

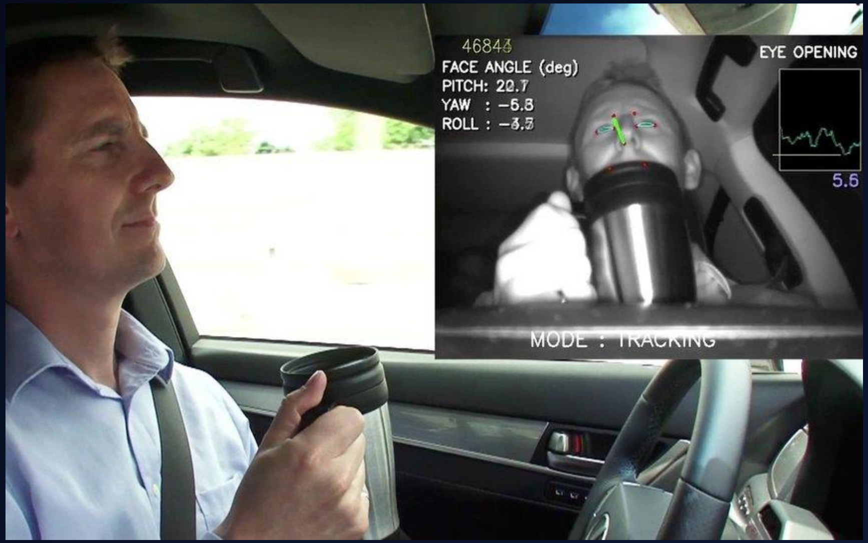
Driver Data Collection

- Amount of driver data being recorded is increasing
- Many new devices and applications

Sensing Application: Driver Authentication

- Vehicles can verify driver identity by measuring distinctive characteristics
- Potential applications to transportation security and safety

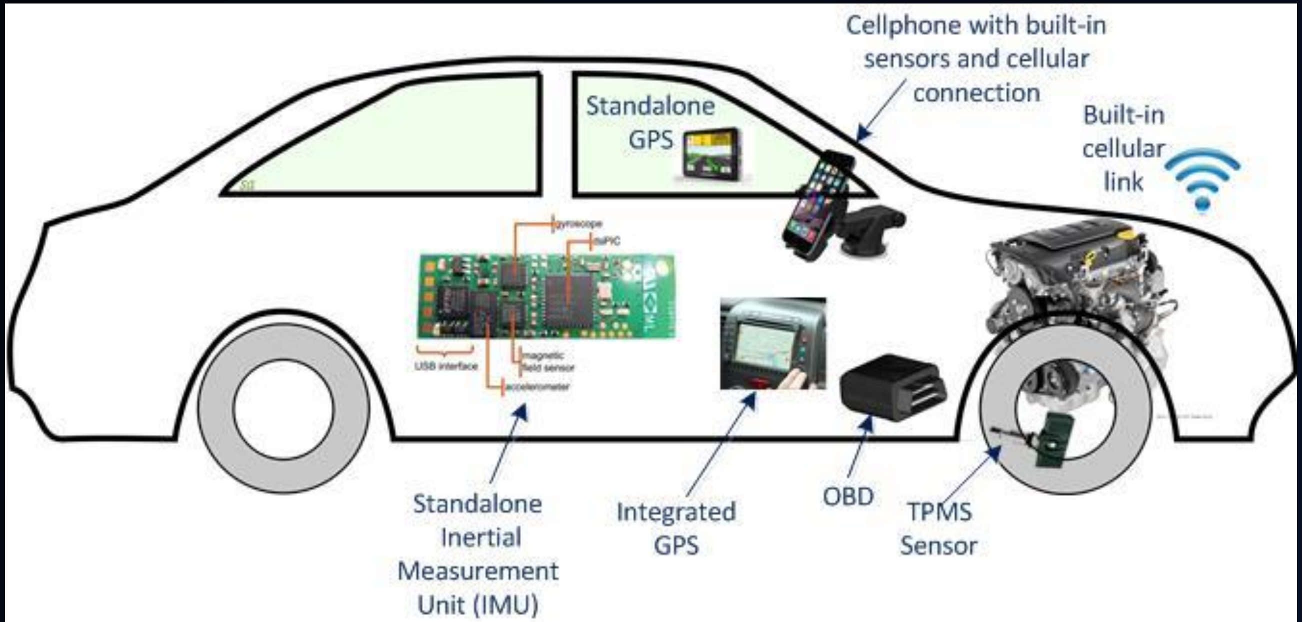


Potential Privacy Issues

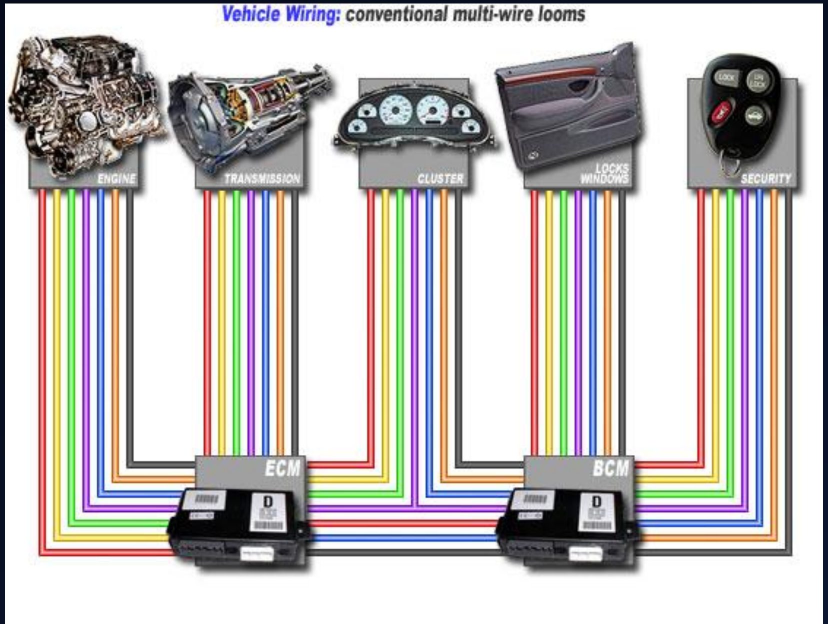
- Devices may record a variety of sensitive information including:
  - Geolocation
  - Audio
  - Images
  - Instantaneous engine readings

Potential Security Issues

- Modern cars controlled by Electronic Control Units (ECUs) connected by a Controller Area Network (CAN bus)



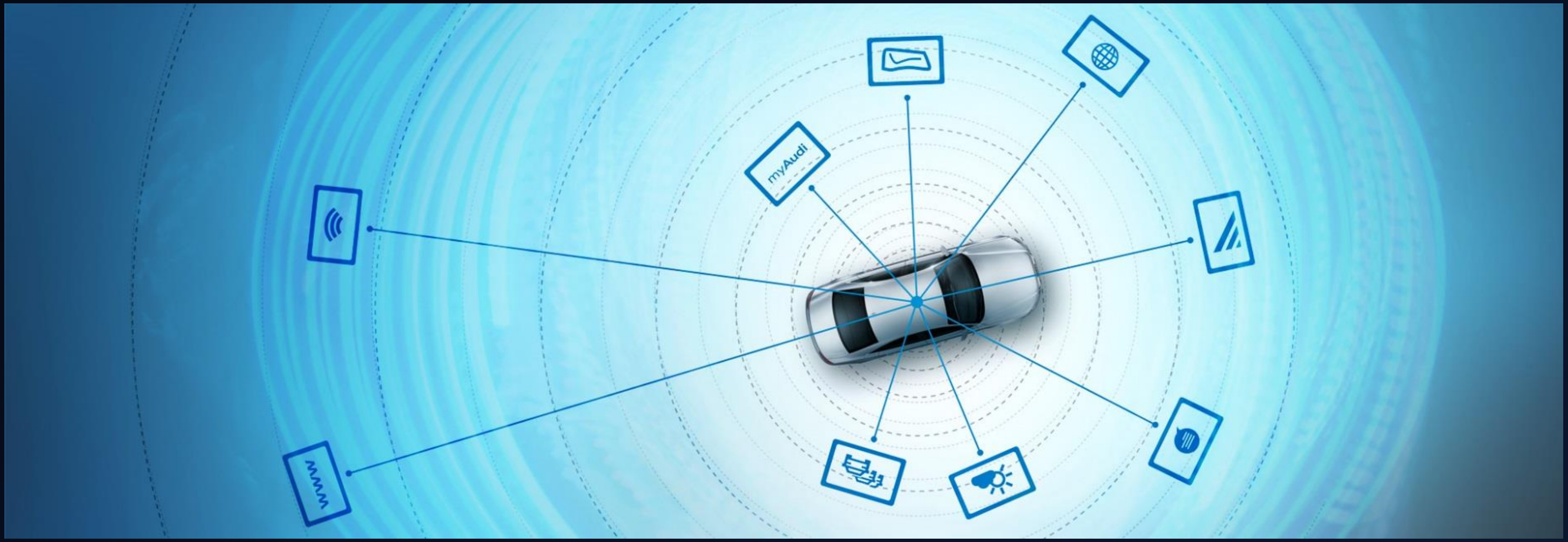
Sensors and communication interfaces available on a modern vehicle.



Examples of CAN Connections [1]

Potential Security Issues

- Devices connect to a vehicle’s CAN bus via an on-board diagnostics (OBD)-II port
- Increases attack surface of critical components
- Many devices also feature a wireless uplink

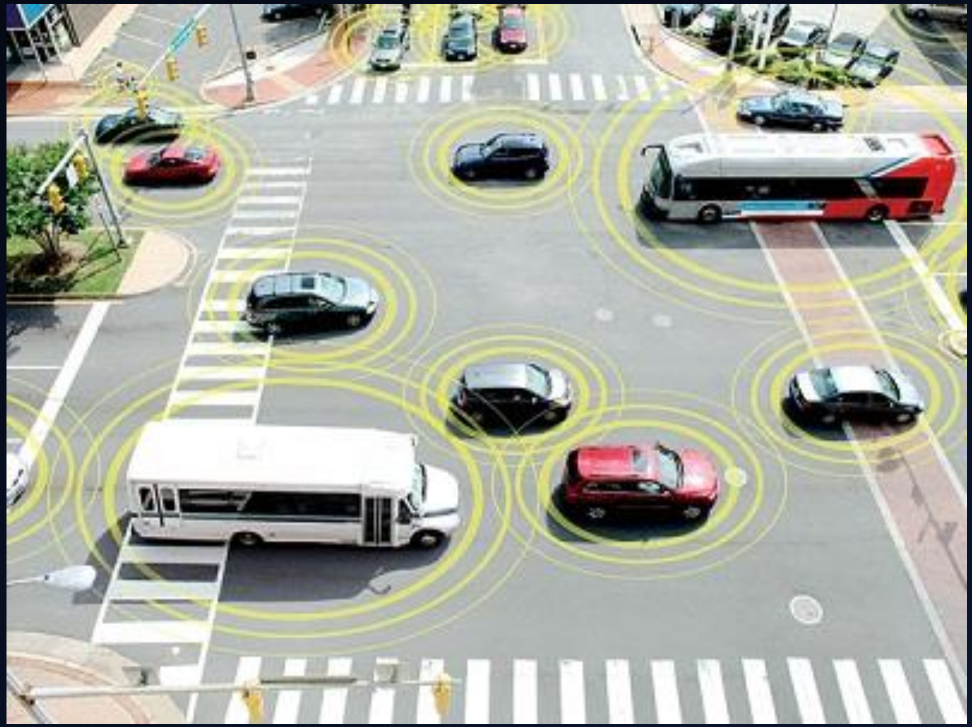


Threat Model

- Situations where token based authentication could be bypassed:
  - A single-owner vehicle is stolen
  - A vehicle is driven by an uninsured driver
  - An unlicensed driver operates a taxi or limo
  - A car sharing service is used by someone who isn't a member
- Adversary with no special knowledge of individual's driving behavior
- Possibility of mid-session attacks
  - Carjacking

Driving Data Dilemma

- Research challenge: how to enable emerging driving applications such as driver identification while ensuring
  - Driver privacy
  - Vehicular security



Solution Idea: Behind-the-Wheel Behavior Modeling

- Decouple sensing from critical vehicle systems
- Measure involuntary driving habits to discern driver identity
- Potential modalities:
  - Steering behavior
  - Speed control characteristics
  - Indicator usage
  - Contextual road features

Related Work

- Authentication via behavioral biometrics in other domains
- Desktops and laptops
  - OS interactions [Payne ‘13]
  - File system usage [Ben Salem ‘14][Voris ‘15]
  - Stylometry [Stolerman ‘14]
- Mobile devices
  - Touchscreen dynamics [Xu ‘14][Scindia ‘16]
  - Application usage [Voris ‘16]
  - Device movement [Sitova ‘15]

Related Work

- Use of driving characteristics to categorize drivers by:
  - Level of drowsiness [Hartley '00]
  - Degree of aggressiveness [Jensen '11]
- Issues with prior driver identification work:
  - Require intrusive sensors such as EEG [Nakanishi '11]

    or dashboard cameras [Ji '04]
  - Privacy issues with some sensors such as geolocation [Tang '08]
  - May require access via a OBD-II board, exposing vehicle control network to attack [Salemi '15]

Advantages of Behind-the-Wheel Behavior Modeling

- Driver identity verification would eliminate fraud
- Deviations from past driving patterns can detect safety issues
- Would not require direct access to a vehicle's CAN bus

Preliminary Evaluation

- Developed a simulated driving task on a desktop computer using the OpenDS driving simulator and a Logitech G27 Steering Wheel

Preliminary Study Design

- Recruited 10 test subjects from university students and staff
- Completed 4 laps each with a 5 minute duration
- Collected raw data at 40 ms interval
  - Coordinates within simulation
  - Steering wheel position
  - Pedal positions

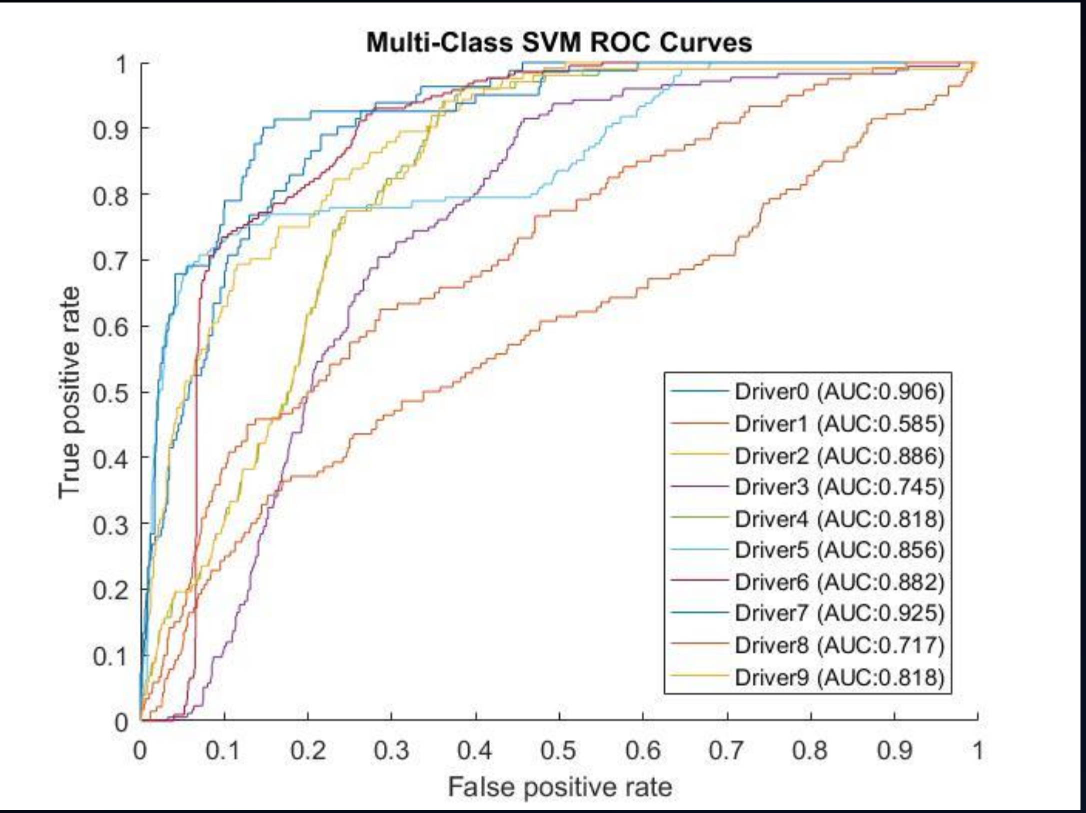
Feature Extraction

- Grouped raw data into 10 second samples to extract features:
  - Euclidean distance traveled
  - Average vehicle speed
  - Standard deviation of steering position
  - Average change of brake pedal position
  - Average change of gas pedal position

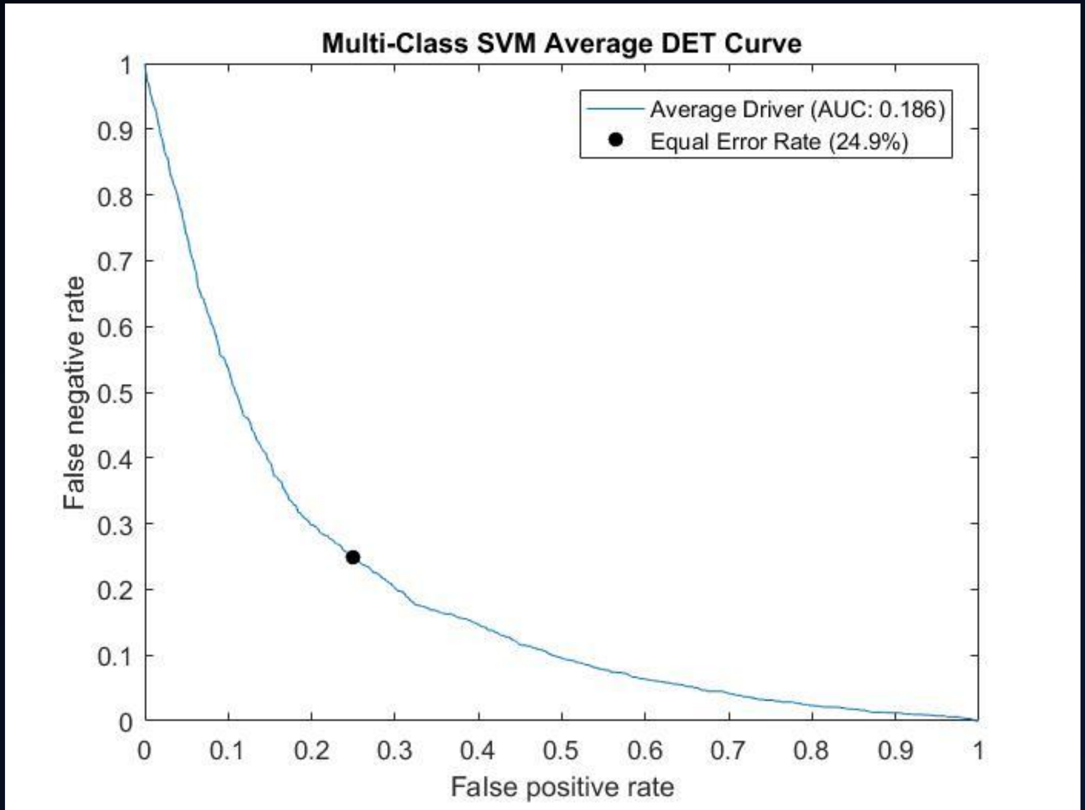
Multiclass Modeling

- Applied several machine learning techniques to driving features
  - Decision Tree
    - With Boosting: Random Forest
  - Support Vector Machine
  - k-Nearest Neighbor
    - With Boosting: Random Subspace
- Data labeled by driver for training and model verification
- Plotted the true positive classification rate against the false positive classification rate to obtain a Receiver Operator Characteristic (ROC) Curve
  - Measuring the area covered by an ROC curve provides the Area Under the Curve (AUC)
- Plotted the false negative classification rate against the false positive classification rate to obtain a Detective Error Tradeoff (DET) Curve

Multiclass Modeling Results

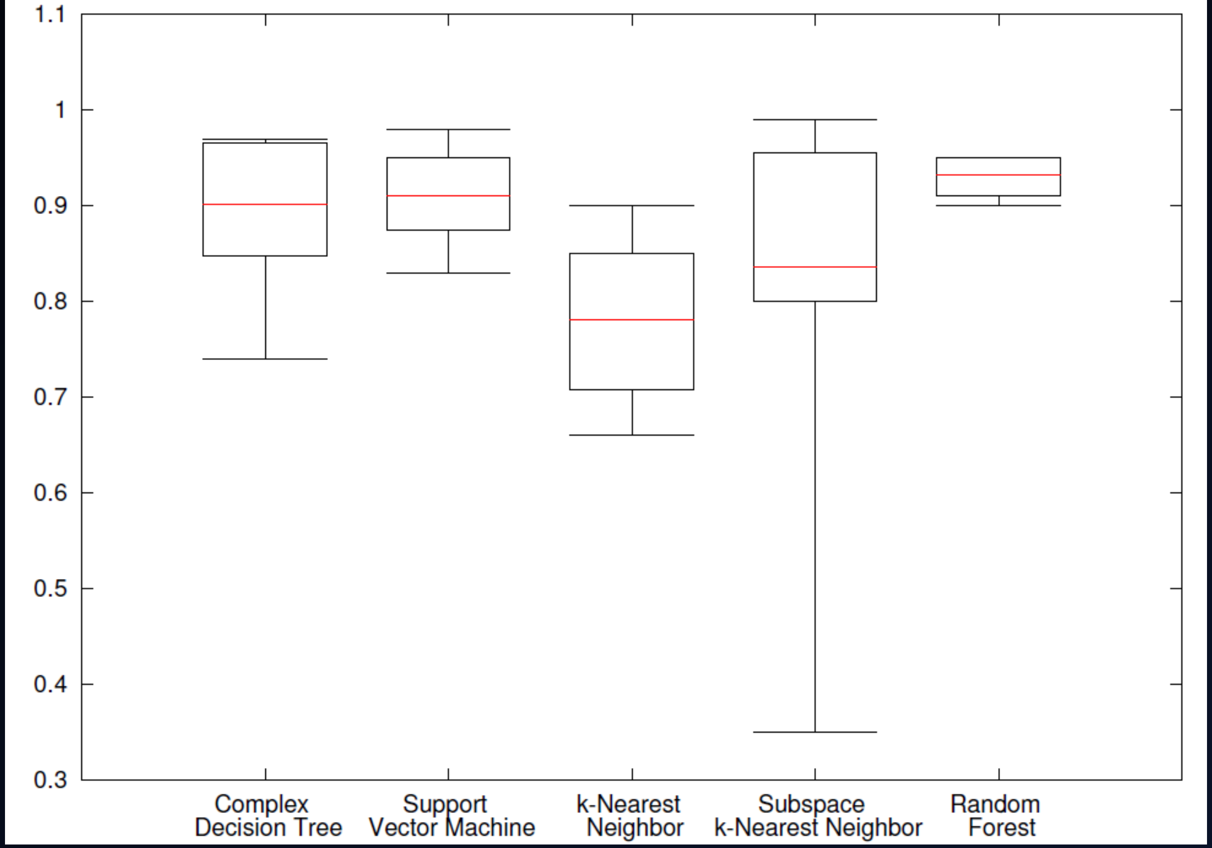


ROC Curves for Multi-Class SVM Classification of All Study Participants



Average DET Curve for Multi-Class SVM Classification.

Multiclass Modeling Comparison



Comparison of AUC Values for Multiclass Modeling Techniques

Feature Analysis

- Good behavioral modeling features should be:
  - Highly consistent for any given driver
  - Highly distinct between any given drivers
- Can be measured using Fisher's separation function:

$$s = \frac{\sum_{i=1}^{c} n_i (u_i - u)^2}{\sum_{i=1}^{c} n_i \sigma_i}$$

Feature Analysis Results

- Compared extracted and raw features

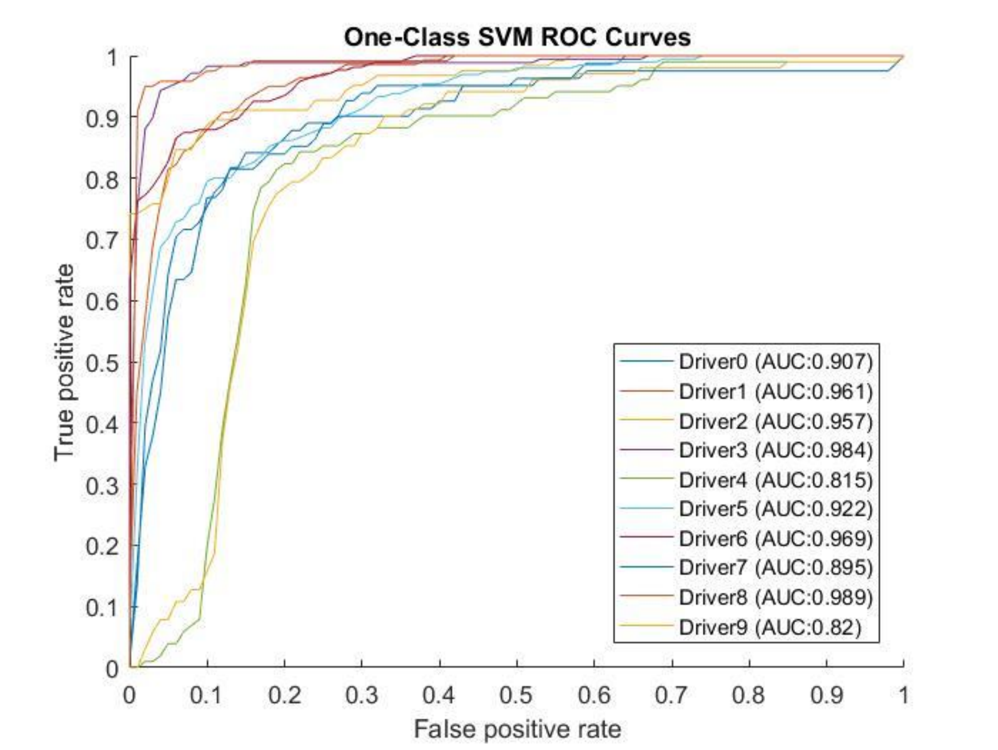
Feature	Fisher Score	Classification Contribution
Average Change in Accelerator Pressure	0.122	3.84%
Distance Traveled	0.101	0.23%
Average Speed	0.082	0.26%
Average Change in Brake Pressure	0.052	1.76%
Standard Deviation of Steering Position	0.039	0.60%
Average X Axis Position	0.037	0.46%
Average Z Axis Position	0.022	1.32%
Average Y Axis Position	0.020	0.00%
Average Z Axis Rotation	0.019	0.00%
Average Y Axis Rotation	0.018	-0.46%
Average X Axis Rotation	0.017	-0.03%
Average W Axis Rotation	0.014	0.07%

Fisher Scores for Driving Features

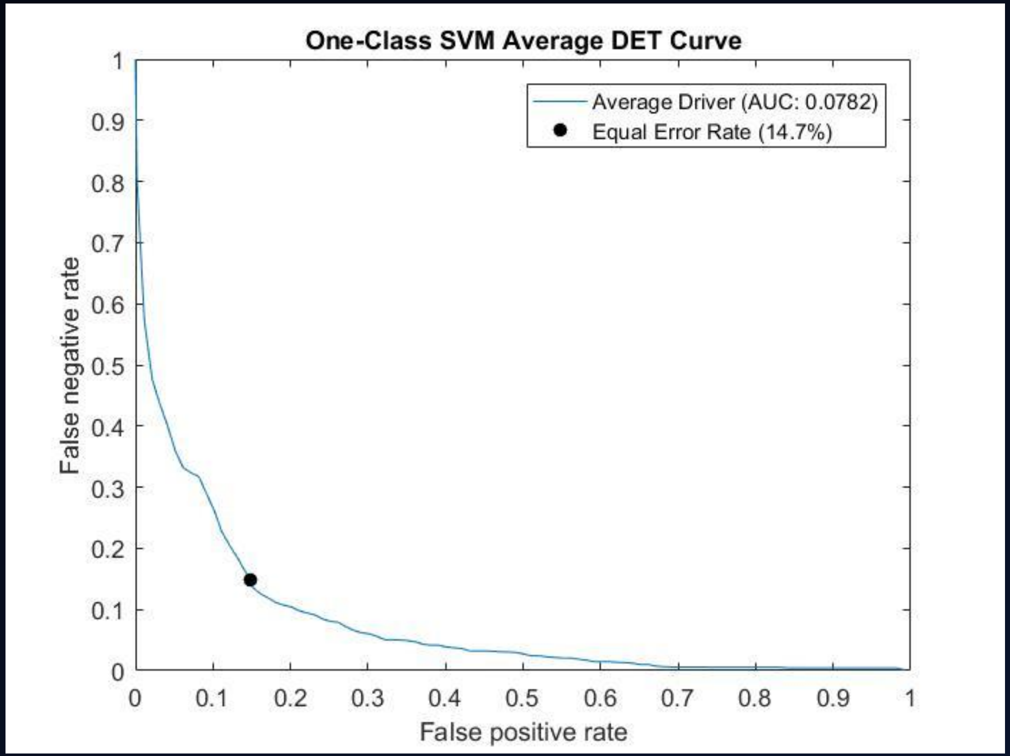
One-Class Modeling

- Multiclass modeling performed for algorithm comparison
  - Requires all user’s data for training
- One-Class training more appropriate to driver modeling
  - More scalable to busy driving environments
  - Other driver’s data might not be available

One-Class Modeling Results



ROC Curves for One-Class SVM Classification of All Study Participants



Average DET Curve for One-Class SVM Classification

Time To Detection

- How long to detect an unauthorized driver?
- Modeling sampling rate of 10 seconds
- Set acceptable false positive rate to one per 46-minute driving day
  - Requires a maximum per-sample FP rate of 0.362%
  - At this FP, TP rate is 19.5%, or 80.5% chance to evade detection per sample
- Samples required for 95% detection confidence: 14

Time To Detection

- Samples required for 95% detection confidence: 14

$$
\begin{aligned}
0.805^x &< 0.05 \\
x &< \log(0.05)/\log(0.805) \\
x &< 13.81
\end{aligned}
$$

- Average time to detection: 2 minutes and 20 seconds

Conclusion

- Novel applications such as driver authentication offer benefits to transportation systems
- Authenticating drivers by modeling their behind-the-wheel behavior seems like a promising approach
  - Prevents token theft and relay attacks
  - Can be performed throughout a session
  - Care must be taken to do so in an unobtrusive and privacy-conscious fashion
- Future work:
  - More comprehensive study with broader population currently underway
  - Analysis additional modeling features and algorithms
  - Susceptibility of behavioral driver authentication to attack

Thank you!