Comparison between Video Image and Smart Wrist Band based Driving Stress Detection Systems

Project Members: Bishop Harsch, Austin Sternberg, Wrenn Zahniser **Project Advisor:** Dr. JungYoon Kim **Affiliation:** Kent State University SCI Lab

Choose Ohio First

1. The Problem

- Driver stress is a major contributing factor to road accidents worldwide
- Traffic congestion, proximity of large vehicles, complex urban intersections, and unpredictable road conditions all contribute to reducing reaction time and situational awareness.
- Everyday driving would require non invasive stress detection procedures to allow for practical use.
- This research investigates whether driver stress can be predicted using two non-invasive sensing modalities. One of which being dashcam video combined with OBD vehicle data. The other, Galvanic Skin Response (GSR) data collected from a wrist-worn sensor.
- The target prediction is binary classification: Normal Driving (0) versus Traffic Driving (1)

2. Introduction

- This study evaluates and compares two non-invasive sensing modalities for real-time driving stress detection, with the goal of determining which approach is more effective in identifying normal and traffic driving conditions.
- Video-Based Pipeline: A dashcam records the forward road scene. YOLO v11, fine-tuned on the BDD100K dataset (100K diverse driving images), performs object detection on each frame. Extracted spatial and traffic features are combined with On-Board Diagnostics (OBD) vehicle telemetry data for classification.
- GSR-Based Pipeline: A wrist-worn GSR sensor captures the driver's skin conductance at 4 Hz. The raw GSR signal is decomposed into tonic and phasic components, from which 60+ statistical features are extracted.
- Both pipelines are synchronized at 4 Hz. Six machine learning classifiers are evaluated on each modality: XGBoost, Random Forest, LightGBM, K-Nearest Neighbors, Multilayer Perceptron, and Deep Neural Network.

Dataset Overview

- Data collected during real on-road driving sessions with a wrist GSR sensor and a dashcam recording the forward road scene.
- 21 Normal driving scenarios and 12 Traffic driving scenarios were recorded.
- Labels: Normal Driving = 0 | Traffic Driving = 1 | Synchronized at a 4 Hz sampling rate

3. Existing Research

Author & Year	Objective	Approach	Key Finding	Dataset
Healey & Picard (2005) MIT Media Lab	Detect real-world driver stress from physiological signals	ECG, EMG, GSR, respiration; 3-level stress classification with 5-min windows	>97% accuracy; skin conductance most correlated with stress	Self-collected; Boston real-road drives (24 drives)
Siam et al. (2023)	AI-based mental stress detection from non-invasive biosignals	KNN, SVM, DT, LR, RF, MLP pipeline on physiological features	Random Forest: 98.2% accuracy on stressed vs. relaxed classification	DriveDB (PhysioNet)
Roha et al. (2023)	Multimodal biosignal fusion for driver stress classification	ECG, EDA, PPG, respiration; deep kernel learning fusion; CARLA driving simulator	ECG+PPG + RF: F1=0.97; PPG-only reaches F1=0.90	Self-collected (N=10 drivers; 25-min protocol)
Zhao et al. (2024)	Self-supervised learning for cognitive load detection with scarce labels	Self-supervised multimodal model; cascaded attention; uncertainty-aware module	+9.2% F1 over baselines in cross-subject settings via linear probing	DriveDB & CogPilot datasets
Ziaratnia et al. (2024)	Remote, non-contact stress estimation from facial video, no wearables	CCT-LSTM: facial expression + rPPG features; 7-fold cross-validation	Task classification: 83.2% Acc; stress-level classification: 80.5% Acc	UBFC-Phys (facial video + physiological signals)
Bustos et al. (2025) MIT Media Lab / UOC	Estimate driver reported stress from road-scene video only, no physiological data	Single-frame baselines (RF, SVM, CNN), Temporal Segment Networks (TSN, TSN-w, TSN-LSTM)	TSN-w achieves best avg accuracy of 0.77; TSN-w, TSN-LSTM, and Transformers statistically equivalent, all outperforming single-frame baselines	AffectiveROAD - 9 drivers, 676 min real-world video, stress annotated at 4 Hz (low/medium/high)

4. Workflows

Fig 4.1: Video/OBD pipeline

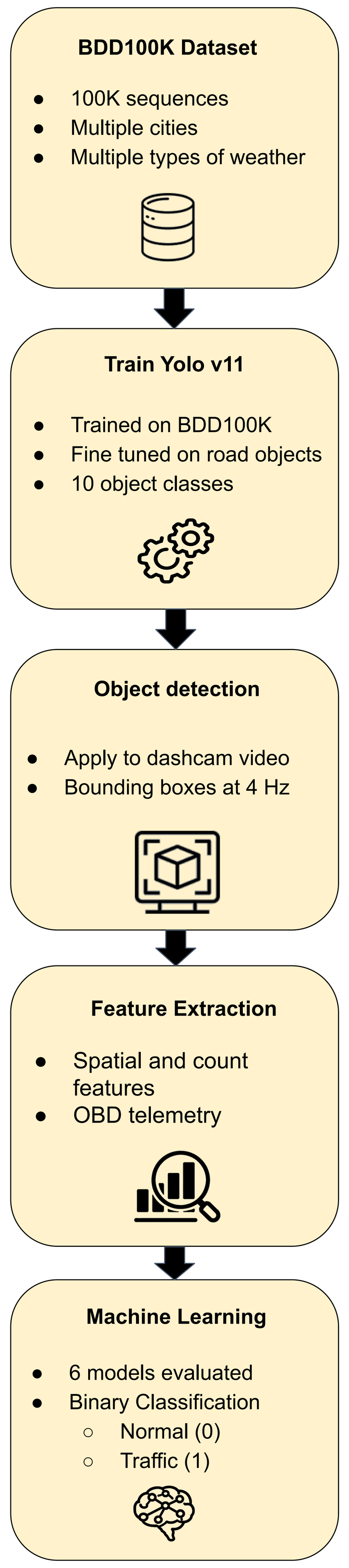
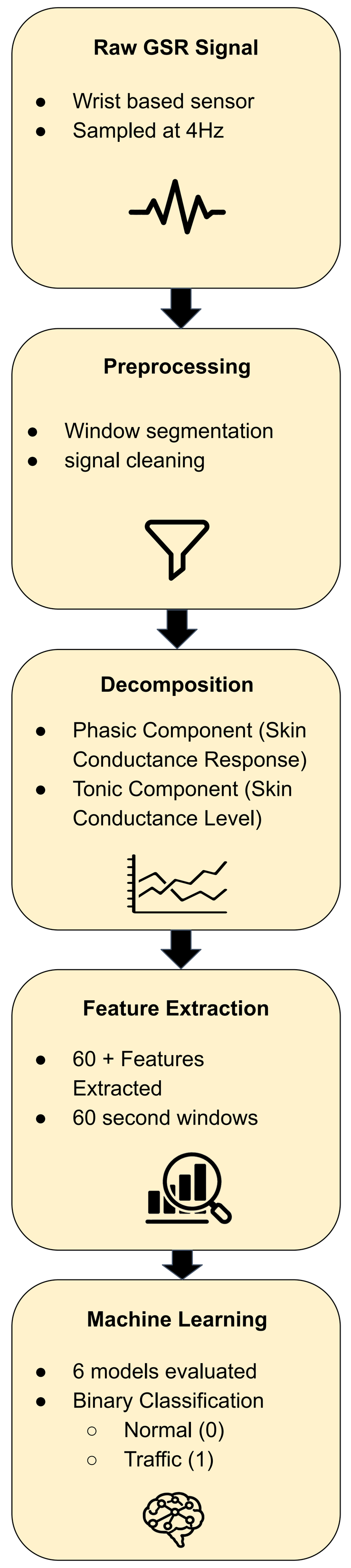


Fig 4.2: GSR pipeline



5. Feature Extraction

Video / OBD Key Features

- RPM
- Vehicle Speed
- Number of Objects Detected
- Objects in Center Zone - close leading vehicles
- Central View Occlusion Ratio - forward view blocked by vehicles

GSR / EDA Key Features

- Raw Signal Mean - overall conductance level higher in traffic
- Raw Signal Min - baseline conductance
- Tonic Starting Value
- Phasic Mean
- Peak Count - frequency of GSR elevations
- Inter-peak Interval - reduced recovery time between elevations

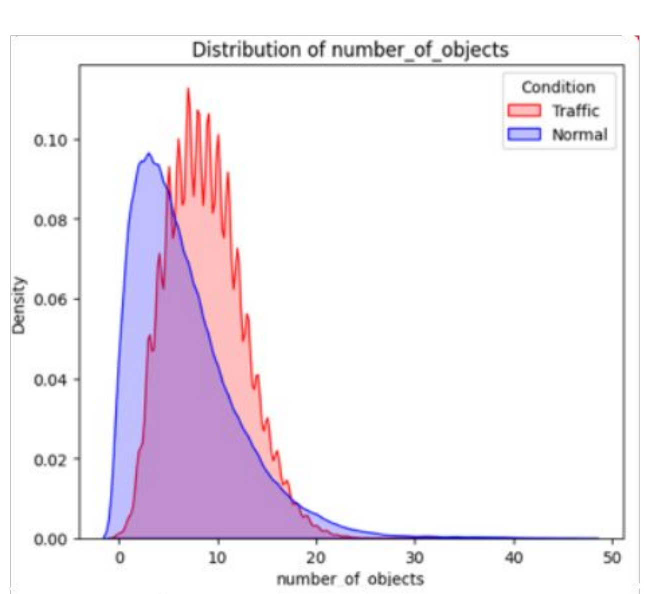


Fig. 5.1: Example of Feature histograms for Video/OBD features showing distribution differences between Normal (purple) and Traffic (pink) conditions

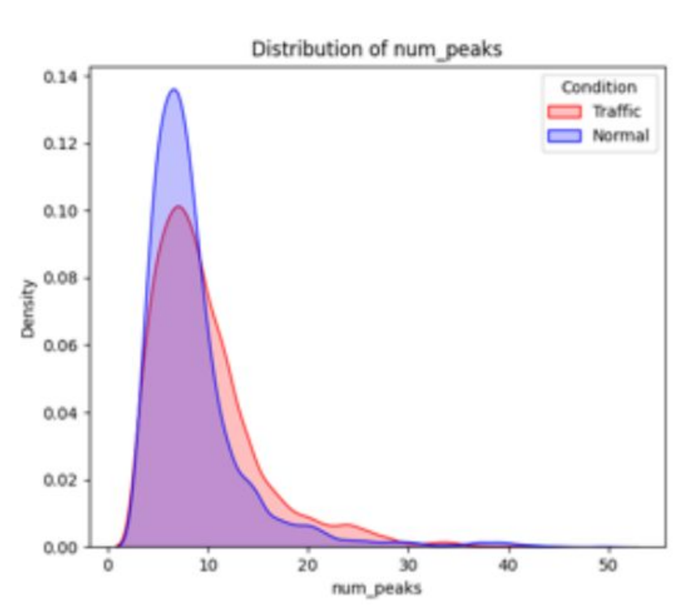


Fig. 5.2: Example of GSR feature histograms showing distribution differences between Normal (purple) and Traffic (pink) conditions

6. Results

Six classifiers were evaluated per modality. LightGBM achieved the highest AUC in both experiments. Results below show the best performing model for each modality.

Video/OBD Results - LightGBM (Best AUC)

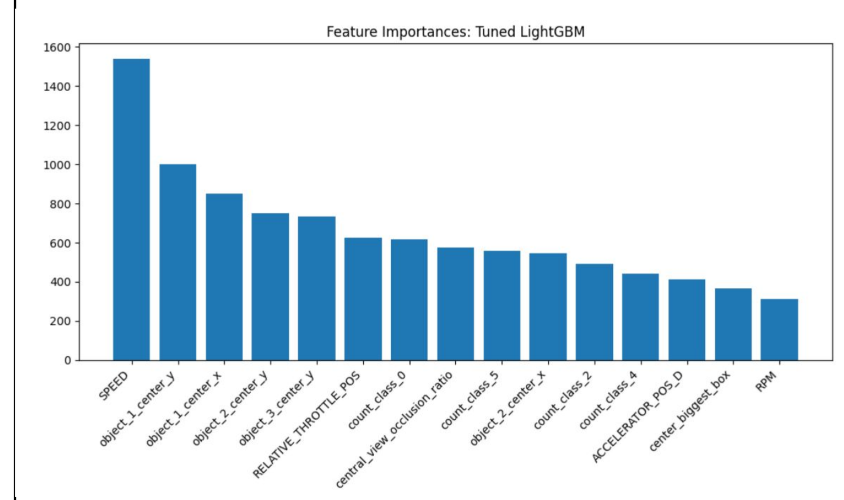


Fig. 6.1: Video Features broken down by importance

GSR Results - LightGBM (Best AUC)

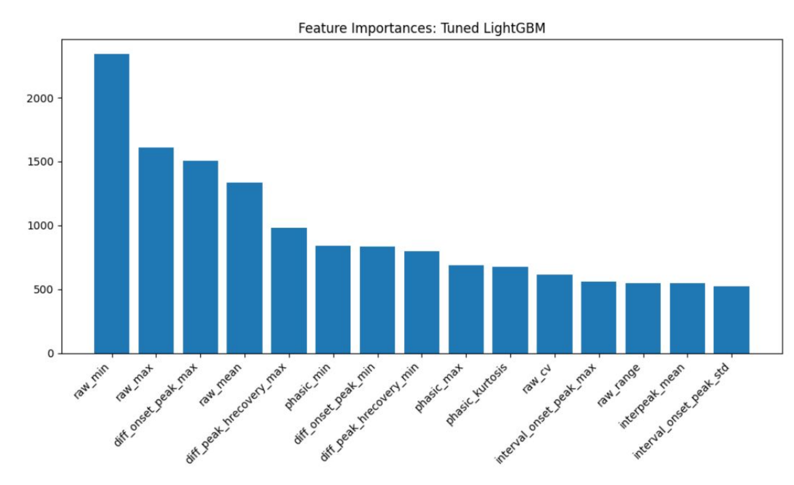
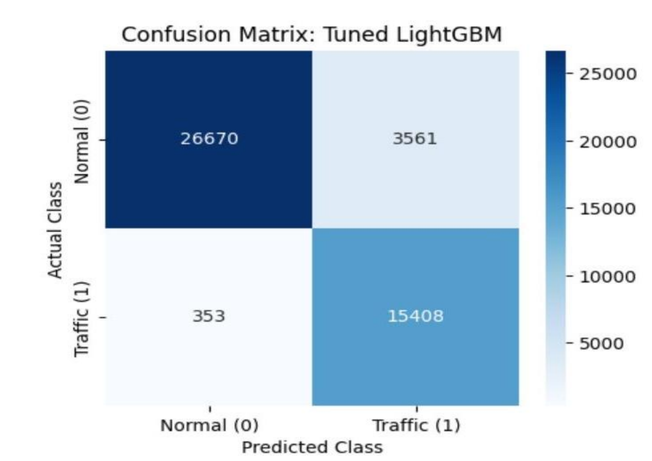
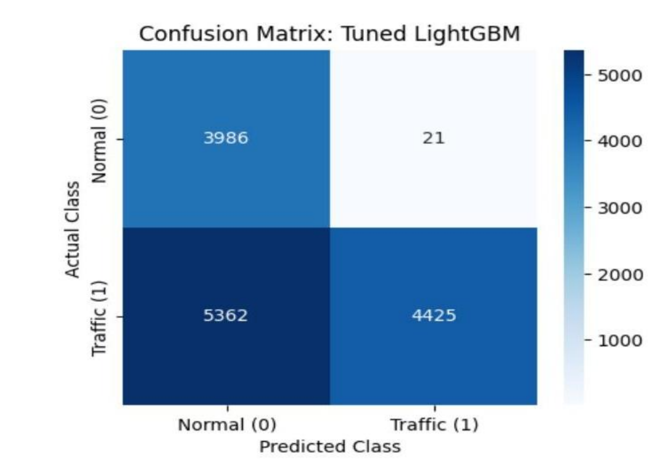


Fig. 6.2: GSR Features broken down by importance



LightGBM					
Accuracy	Precision	Recall	Specificity	F1 Score	AUC
0.9149	0.8123	0.9776	0.8822	0.8873	**0.9823**

Fig. 6.3: Confusion Matrix for LightGBM (Video)



LightGBM					
Accuracy	Precision	Recall	Specificity	F1 Score	AUC
0.6098	0.9953	0.4521	0.9948	0.6218	**0.9328**

Fig. 6.4: Confusion Matrix for LightGBM (GSR)

ROC Curve Comparison — Video Classifiers

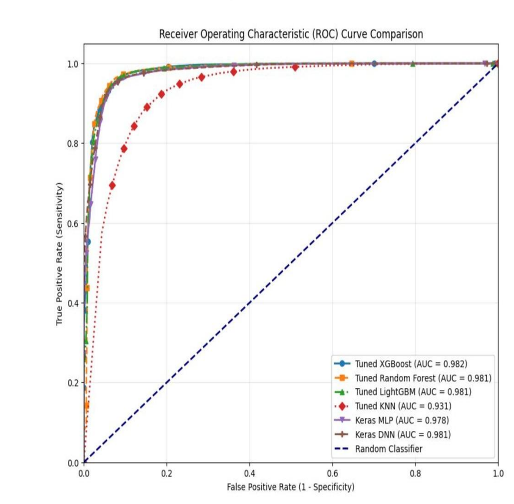


Fig. 6.5: ROC Curves for all Video classifiers.

ROC Curve Comparison — GSR Classifiers

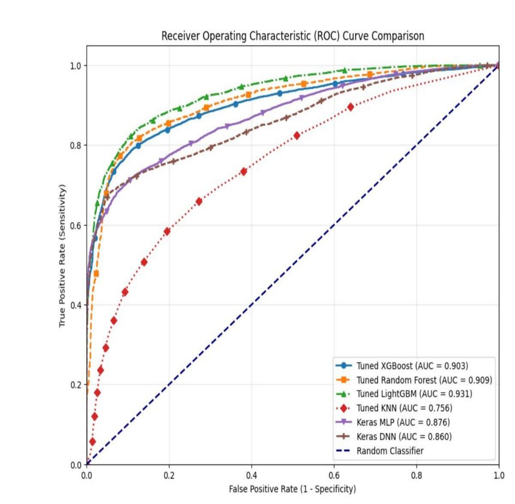


Fig. 6.6: ROC Curves for all GSR classifiers.

7. Future Work

- Combine both approaches: combine road-scene video features with physiological GSR signals in a single multimodal model to leverage the strengths of each.
- Expand the dataset: collect additional driving sessions to better balance Normal and Traffic scenarios, and introduce more gradual stress labels to indicate specific levels of stress.
- Additional physiological features: supplement GSR with additional biosignals such as skin temperature and wrist accelerometer data to improve signal robustness
- Real-time deployment: implement the best-performing model to be used for live driver stress monitoring

8. Conclusion

- Video-based detection proved highly effective at identifying when a driver is in traffic, correctly classifying over 91% of driving windows using features extracted from dashcam footage and vehicle data alone with no wearable sensors required.
- GSR-based detection showed strong potential as a stress indicator, but was held back more by the class imbalance
- Both modalities capturing meaningful stress-related information points toward a combined approach as the most promising direction. Pairing the reliability of video with the physiological data of GSR could produce a more accurate driver system that is still practical.