Analysis of Driver Behavior in Various Events Using Electrodermal Activity Signal

Yedukondala Rao Veeranki[1], Toleti Sai Shanmukh Kailash[1], Luis Roberto Mercado Diaz[2], Hugo F. Posada-Quintero[2]

Abstract*—Inappropriate driver behavior is a leading cause of traffic accidents, contributing to 94% of crashes, according to the National Motor Vehicle Crash Causation Survey. Factors such as individual driving styles, risk-taking tendencies, and non-compliance with traffic regulations increase the likelihood of conflicts and accidents, emphasizing the need for a deeper understanding of driver behavior. Despite existing studies on driving behavior, there is a lack of precise and comprehensive classification methods that effectively correlate physiological signals with driving actions. Current models do not fully capture the cognitive aspects of driving, limiting their applicability in enhancing traffic safety and accident prevention. The primary goal of this study is to identify the most relevant features that contribute to analyzing driving behavior and utilize them to classify driver actions accurately. We conducted feature extraction using 52 distinct features to analyze driving behavior. Following this, feature selection was performed using the Random Forest Recursive Feature Elimination method to identify the 10 most significant features. These important features were then used for driver behavior classification with machine learning models, ensuring improved accuracy and efficiency in identifying different driving patterns.

***Keywords*—Electrodermal Activity, Galvanic Skin Response, Decomposition, Random Forest Recursive Feature Elimination, Machine Learning**

I. INTRODUCTION

Traffic accidents remain a significant global concern, with most collisions attributed to human-related factors, including behavior, decision-making, and reaction time. Predominantly, crashes result from inappropriate driving behaviors such as non-compliance with traffic regulations, engagement in risky driving practices, and impaired judgment [1]. Variables such as age and cognitive abilities introduce significant risks to driving safety [2]. Given the inherent complexity and variability of human behavior, manual behavior analysis remains limited in scope and effectiveness. Machine learning presents a scalable solution by uncovering complex patterns within large datasets, thereby enhancing the accuracy of behavior prediction and facilitating real-time interventions. Integrating insights from psychology, physiology, and engineering can further advance the development of intelligent safety systems, enabling a more holistic approach to understanding and mitigating human-related risks [3].

Detecting driver stress is pivotal for enhancing road safety, prompting extensive research into various detection methodologies [4]. Traditional approaches have focused on physiological measurements, such as heart rate variability (HRV), electroencephalography (EEG), and electrodermal activity (EDA). For instance, a real-time cardiac measurement study demonstrated that heart rate changes effectively indicate driver stress and hazard anticipation [5]. However, HRV and EEG monitoring often require intrusive equipment, limiting their practicality in real-world driving scenarios.

In contrast, EDA, which measures skin conductance changes associated with sweat gland activity, offers a more feasible alternative [6]. This response indicates physiological arousal, providing insights into emotional and cognitive states. The EDA analysis not only provides a less intrusive means of monitoring but also delivers high accuracy in detecting driver stress, making it a superior choice for practical applications [6]. EDA signals consist of two components: the tonic and the phasic components. The tonic component represents the baseline skin conductance during a resting state, reflecting general arousal and alertness over time. The phasic component represents rapid fluctuations in conductance in response to stimuli or internal changes, known as skin conductance responses (SCRs) [7]. Phasic responses are more commonly emphasized due to their sensitivity to specific stimuli and ability to capture immediate reactions, making them helpful in studying emotional or cognitive responses [7]. Our research focused on the phasic component of EDA signals to classify driver states.

Recent studies have utilized EDA signals for driver stress detection using various feature extraction and classification techniques. Memar and Mokaribolhassan employed statistical features like mean and variance to classify stress levels during driving [8]. Similarly, Zontone et al [9], developed a low-complexity classification algorithm that incorporated adaptive filtering to mitigate motion artifacts in EDA signals, followed by statistical feature extraction and SVM classification, resulting in an accuracy of 87.40% in stress detection. Khai Ooi et al. developed an EDA emotion recognition framework that can accurately recognize stress and anger from neutral emotion, with Significant differences in classification accuracy of 85% and insignificant differences in classification accuracy of 70% [10]. Jiao et al. combined EDA and HRV signals to evaluate driver fatigue using wearable devices [11].

Building on these studies, our work presents a lightweight and interpretable framework for driver behavior classification using only basic time-domain features from EDA signals. Unlike prior approaches relying on complex preprocessing or deep models, we apply Random Forest Recursive Feature Elimination (RFRFE) to select the top 10 features from an initial set of 52. These selected features are then used to train machine learning classifiers for accurate and real-time stress detection.

The contributions of this study are twofold:

(i) Demonstrating that basic time-domain features can effectively differentiate driver emotional states with high accuracy when carefully selected.

(ii) Providing a lightweight and computationally efficient framework that facilitates practical real-time driver monitoring applications.

[1]Department of Electronics & Communication Engineering, National Institute of Technology Puducherry, Karaikal, PY 609 609 India
[2]Department of Biomedical Engineering, University of Connecticut, Storrs, CT 06269 USA
Corresponding Author: hugo.posada-quintero@uconn.edu

Fig.1 Block diagram of methodology used in the study

II. METHODOLOGY

EDA signals are collected from the Multimodal Physiological Dataset for Driving Behavior (MPDB) [12]. The EDA signals are decomposed into phasic and tonic components to isolate the rapid, event-related responses better. From the phasic component, a comprehensive set of time-domain features is extracted to capture the dynamic characteristics of the EDA signals. Subsequently, a Recursive Feature Removal based on Random Forest (RFRFE) method is employed for feature selection, followed by classification using machine learning algorithms. Fig.1 shows the block diagram of the methodology used in the study.

1. Collection of EDA Signals:

In this study, EDA signals are collected from the publicly available dataset [12]. The EDA signals were collected from 35 voluntary participants (26 males, 9 females, aged 20–60 years, average age 25.06 ± 7.90 years) from Tsinghua University participated in the experiment, which involved 150-minute event-related driving tasks. Participants were required to have a valid driving license (minimum grade C) and at least one year of driving experience. To minimize external factors that could affect physiological signals, participants were instructed to ensure adequate rest prior to the experiment and to refrain from consuming stimulants or drugs. During the experiment, data were collected while participants engaged in five different driving behaviors: smooth driving, acceleration, deceleration, lane-change, and turning. These physiological signals were recorded synchronously, and the data were later used for further analysis in this study [12]. Table 1 summarizes the demographic and EDA data of the participants

TABLE I. EDA AND PARTICIPANT DEMOGRAPHIC DATA

Category	Description
Number of Participants	35 (26 males, 9 females)
Age Range	20–60 years (Average age = 25.06 years, SD = 7.90)
Driving Experience	1–20 years (Average driving experience = 3.03 years, SD = 3.68)
Driving License	Grade C or above (People's Republic of China)
Experimental Duration	150 minutes of event-related driving tasks
Physiological Data Collected	EDA, EEG, ECG, Eye Tracking, EMG, GSR
Data Collection System	NeuSen W GSR Series Wireless GSR Acquisition System (Sampling rate: 1000 Hz)
Event Types	Smooth driving (control), acceleration, deceleration, lane-change, turning

2. Decomposition:

EDA signals are typically decomposed into tonic and phasic components to better understand autonomic nervous system responses. Greco et al. [13] developed cvxEDA, a quadratic programming-based decomposition technique designed to decompose tonic and phasic components of EDA. The method represents the EDA as the sum of a fast phasic component (r), a slow tonic component (t), and a Gaussian noise term (ϵ) according to the equation $y_{cvx} = r + t + \epsilon$. The phasic component is obtained through a convolution between a sparse, nonnegative SNA driver and an impulse response function (IRF), which is modeled using a Bateman function. The tonic component is estimated using cubic spline interpolation to determine the physiological characteristics such as temporal scale and smoothness. The method utilizes quadratic programming convex optimization to identify the optimal tonic and phasic components. Pre-processing involves z-score normalization of the EDA to offset individual baseline variations. The method has been evaluated for its ability to detect sympathetic nervous system (SNS) activity, robustness against noise, and effectiveness in separating stimulus inputs across multiple experimental sessions. The MATLAB implementation of the cvxEDA algorithm is available online, with parameters set to α = 0.008, $\tau1$ = 0.7s, $\tau2$ = 2s, and γ = 0.01 in this study [13].

3. Feature extraction:

In this study, a comprehensive set of 52 time-domain features [14] was extracted from the phasic component of EDA signals to capture the underlying emotional states of drivers. These features were selected to cover diverse signal characteristics, broadly grouped into categories such as statistical features (e.g., mean, standard deviation, skewness, kurtosis), signal power descriptors (e.g., energy, RMS, power), morphological descriptors (e.g., waveform length, average amplitude change, arc length), derivative-based measures (e.g., mean and standard deviation of first- and second-order differences), complexity measures (e.g., mobility, complexity, chaos, hazard), and entropy measures (e.g., Shannon entropy). This wide feature space ensures that subtle temporal and structural variations in the EDA signal relevant to emotional states are captured. The list of features are shown in Table II.

TABLE II. CATEGORIZATION OF TIME-DOMAIN FEATURES

Category	Selected Features
Statistical Features	Mean, Standard Deviation, Variance, Median, Skewness, Kurtosis, Coefficient of Variation (CV)
Power/Amplitude Measures	Energy, Power, RMS, Absolute Mean (MAD), Simple Square Integral (SSI), Log Detector (LOG), Mean Absolute Value (MAV)
Morphological Features	Average Amplitude Change (AAC), Difference Absolute Standard Deviation Value (DASDV), Arc Length (ARC), Perimeter-Area Ratio (APR), Energy-Perimeter Ratio (EPR), Dynamic Range (DRSC)
Derivative-based Features	Mean of First Derivative, Mean of Negative First Derivative, Std. of First Derivative (FDSC), Mean and Std. of Second Derivatives (SMSC, SDSC)
Peak/Extrema Features	Minimum (MIN_SC), Maximum (MAX_SC), Crest Factor, Margin Factor, Impulse Factor
Complexity Measures	Activity, Mobility, Complexity, Chaos, Hazard
Entropy Measure	Shannon Entropy (SE)

4. Feature selection:

Feature selection is vital in improving model performance by identifying the most relevant features while minimizing redundancy. In this study, the top 10 features are selected from the 52 extracted features using the technique Random Forest Recursive Feature Elimination (RFRFE) [15]. RFRFE leverages the feature importance scores from a Random Forest model to iteratively remove the least significant features. Applying RFRFE ensures that only the most informative features are retained, reducing dimensionality and computational complexity while enhancing classification accuracy. This refined feature set allows the machine learning model to focus on the most critical aspects of the EDA signal, improving both interpretability and reliability in driver behavior classification [15].

5. Classification:

The final step in our approach involves classifying driver behavior using machine learning models, including Support Vector Machine (SVM), Random Forest (RF), Linear Discriminant Analysis (LDA), and Multi-Layer Perceptron (MLP). These models are trained on the top 10 selected features from the EDA signal to categorize driving behaviors into five classes: Smooth Driving, Acceleration, Deceleration, Turning, and Lane Changing. SVM effectively handles high-dimensional data and finds optimal decision boundaries, while RF enhances robustness by combining multiple decision trees. LDA improves classification by maximizing class separability MLP, a neural network-based model, captures complex patterns within the data. We employed a One-vs-One (OvO) approach to enhance classification performance, which involves building separate models for each pair of behavior classes. This method breaks down the multi-class problem into simpler binary classification tasks, enabling the models to better learn the differences between specific behavior types. As a result, the OvO strategy improves precision and reduces confusion between similar classes, making it a valuable choice for multi-class driver behavior analysis.

III. RESULTS AND DISCUSSION

The EDA signals observed during various driving behaviors shows distinct patterns as illustrated in Fig.2. During smooth driving, a linear decrease in EDA is observed, likely reflecting a steady state of relaxation or lower stress, as the driver maintains a constant and controlled driving pace. In deceleration, a linear increase in EDA is seen initially, followed by a slower rise, which can be attributed to the driver's heightened physiological response as they anticipate a reduction in speed, followed by a period of gradual relaxation as the deceleration completes. Acceleration presents a rise in EDA up to a certain point, followed by a reduction, which may indicate an initial increase in stress or arousal as the driver accelerates, followed by a return to baseline once the task stabilizes. During turning, the EDA shows an exponential increase followed by a constant value, suggesting a rapid physiological response to the more demanding task of turning, followed by stabilization once the turn is completed. Finally, during lane changing, the EDA exhibits a bell-shaped curve, reflecting an initial increase in stress or alertness during the maneuver, followed by a decrease as the driver completes the lane change and regains a steady state. These EDA signal patterns are consistent with the expected physiological responses to these driving tasks, supporting the idea that changes in EDA are linked to the driver's cognitive and emotional state during different phases of driving.

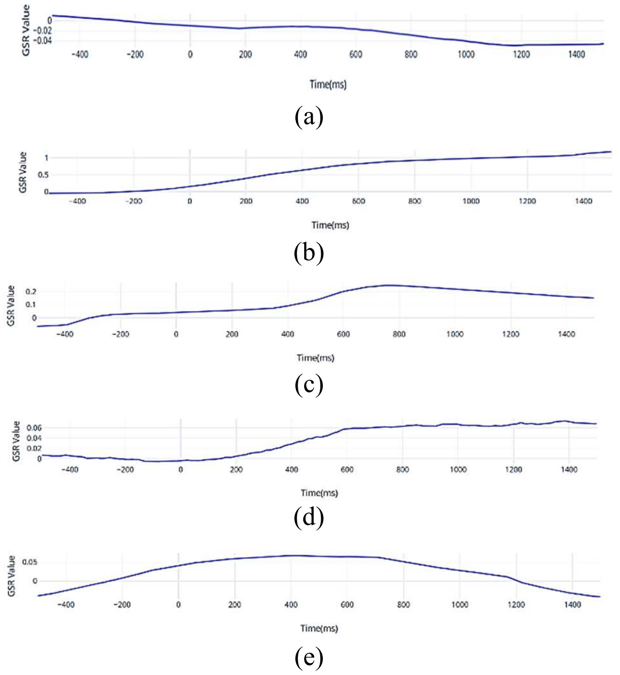

Fig. 2 Representative EDA signals during (a) Smooth driving, (b) Deceleration, (c)Acceleration, (d)Turning, and (e) Lane change

TABLE III. TOP 10 FEATURES FROM RFRFE

RFE Rank	Feature	RF Importance	Feature Name
1	F6	0.03	Kurtosis
2	F40	0.03	SMSC (Spectral Moments of Second Order)
3	F7	0.03	Skewness
4	F12	0.03	CrestFactor
5	F47	0.02	Chaos
6	F46	0.02	Complexity
7	F9	0.02	CoeffVar (Coefficient of Variation)
8	F48	0.02	Hazard
9	F39	0.02	FDSC (Frequency Domain Signal Complexity)
10	F30	0.01	MAVSLP2 (Mean Absolute Value Slope 2)

The top 10 features selected through RFE with RF importance are listed in Table III. It consist of a mix of statistical and signal complexity measures that are crucial for analyzing the underlying dynamics of the system. These features include Kurtosis, SMSC, Skewness, and CrestFactor, all with RF importance values of 0.03, which capture key statistical properties of the signal such as its distribution shape and peak characteristics. Other important features include Chaos, Complexity, Coefficient of Variation, Hazard, FDSC (Frequency Domain Signal Complexity), and MAVSLP2 (Mean Absolute Value Slope 2), with importance values ranging from 0.02 to 0.01. These features provide critical insights into signal variability, frequency domain characteristics, and non-linear dynamics, highlighting their relevance for improving model performance and understanding complex patterns in the data.

Table IV. Classification metrics of SVM

Class Pair	Accuracy	Precision	Recall	F1 Score	AUC
1-2	0.594758	0.593496	0.996587	0.743949	0.46983
1-3	0.617476	0.621094	0.990654	0.763505	0.576934
1-4	**0.677205**	**0.677205**	**1**	**0.80754**	0.505661
1-5	0.532688	0.545455	0.485714	0.513854	0.540183
2-3	0.557025	0.567867	0.646688	0.60472	**0.583366**
2-4	0.554269	0.554585	0.994778	0.71215	0.461683
2-5	0.596421	0.5	0.108374	0.178138	0.540369
3-4	0.55493	0.555887	0.953964	0.702448	0.568312
3-5	0.578544	0.423077	0.050926	0.090909	0.545631
4-5	0.547739	0.524096	0.679688	0.591837	0.569977

The Table IV presents the classification metrics of the SVM classifier for different class pairs, showing its performance across various evaluation metrics including accuracy, precision, recall, F1 score, and AUC. The highest F1 score (0.80754), accuracy (0.677205), precision (0.677205), and recall (1) are observed between class 1 and class 4, indicating a strong classification performance for this pair. In contrast, the recall for class 1 and class 5 is notably lower (0.485714), with the F1 score also being reduced (0.513854), reflecting a weaker performance. The area under the curve (AUC) is highest between class 2 and class 3 (0.576934), suggesting better overall model discrimination between these classes. Overall, the SVM classifier outperforms the other machine learning models in terms of F1 score, accuracy, precision, and recall, with class 1 and class 4 demonstrating the most reliable classification results.

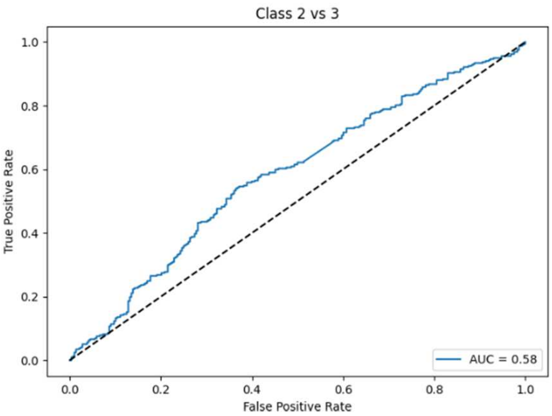

Fig.3 ROC Curve

The area under the curve (AUC) for Class 2 vs. Class 3 is around 0.58, representing the highest separability observed among all class combinations is depicted in Fig.3.

IV. Conclusion

This study proposes a robust and efficient framework for classifying driver behavior using Electrodermal Activity (EDA) signals. By decomposing the signals and focusing on the phasic component, 52 statistical, temporal, and complexity-based features were extracted. Random Forest Recursive Feature Elimination (RFRFE) was applied to select the top 10 most significant features, which were then used to train machine learning models including SVM, RF, LDA, and MLP. Among these, SVM outperformed others in accurately classifying five driving behaviors: Smooth Driving, Acceleration, Deceleration, Turning, and Lane Changing. One-vs-One analysis further highlighted the strong classification performance between Smooth Driving and Turning. The results demonstrate that the proposed approach is both accurate and computationally efficient, making it well-suited for real-time driver behavior and stress monitoring applications. This study is limited by a small, single-institution sample and pairwise metrics; future work will expand participants and include full multi-class evaluations.

References

[1] S. Singh, "Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey," *Traffic Saf. Facts - Crash Stats*, Art. no. DOT HS 812 506, Mar. 2018, Accessed: May 26, 2025. [Online]. Available: https://trid.trb.org/View/1507603

[2] K. J. Anstey, J. Wood, S. Lord, and J. G. Walker, "Cognitive, sensory and physical factors enabling driving safety in older adults," *Clin. Psychol. Rev.*, vol. 25, no. 1, pp. 45–65, Jan. 2005, doi: 10.1016/j.cpr.2004.07.008.

[3] F. A. Alijoyo *et al.*, "Enhancing AI interpretation and decision-making: Integrating cognitive computational models with deep learning for advanced uncertain reasoning systems," *Alex. Eng. J.*, vol. 99, pp. 17–30, Jul. 2024, doi: 10.1016/j.aej.2024.04.073.

[4] M. N. RASTGOO, B. Nakisa, A. Rakotonirainy, V. Chandran, and D. Tjondronegoro, "A Critical Review of Proactive Detection of Driver Stress Levels Based on Multimodal Measurements," *ACM Comput Surv*, vol. 51, no. 5, p. 88:1-88:35, Sep. 2018, doi: 10.1145/3186585.

[5] W.-Y. Chung, T.-W. Chong, and B.-G. Lee, "Methods to Detect and Reduce Driver Stress: A Review," *Int. J. Automot. Technol.*, vol. 20, no. 5, pp. 1051–1063, Oct. 2019, doi: 10.1007/s12239-019-0099-3.

[6] A. Affanni, R. Bernardini, A. Piras, R. Rinaldo, and P. Zontone, "Driver's stress detection using Skin Potential Response signals," *Measurement*, vol. 122, pp. 264–274, Jul. 2018, doi: 10.1016/j.measurement.2018.03.040.

[7] Y. R. Veeranki, L. R. M. Diaz, R. Swaminathan, and H. F. Posada-Quintero, "Non-Linear Signal Processing Methods for Automatic Emotion Recognition using Electrodermal Activity," *IEEE Sens. J.*, pp. 1–1, 2024, doi: 10.1109/JSEN.2024.3354553.

[8] M. Memar and A. Mokaribolhassan, "Stress level classification using statistical analysis of skin conductance signal while driving," *SN Appl. Sci.*, vol. 3, no. 1, p. 64, Jan. 2021, doi: 10.1007/s42452-020-04134-7.

[9] P. Zontone, A. Affanni, R. Bernardini, A. Piras, and R. Rinaldo, "Low-Complexity Classification Algorithm to Identify Drivers' Stress Using Electrodermal Activity (EDA) Measurements," in *Biomedical Engineering and Computational Intelligence*, J. M. R. S. Tavares, N. Dey, and A. Joshi, Eds., Cham: Springer International Publishing, 2020, pp. 25–33. doi: 10.1007/978-3-030-21726-6_3.

[10] J. S. Khai Ooi, S. A. Ahmad, Y. Z. Chong, S. H. Md Ali, G. Ai, and H. Wagatsuma, "Driver emotion recognition framework based on electrodermal activity measurements during simulated driving conditions," in *2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, Dec. 2016, pp. 365–369. doi: 10.1109/IECBES.2016.7843475.

[11] Y. Jiao, C. Zhang, X. Chen, L. Fu, C. Jiang, and C. Wen, "Driver Fatigue Detection Using Measures of Heart Rate Variability and Electrodermal Activity," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 6, pp. 5510–5524, Jun. 2024, doi: 10.1109/TITS.2023.3333252.

[12] X. Tao *et al.*, "A multimodal physiological dataset for driving behaviour analysis," *Sci. Data*, vol. 11, no. 1, p. 378, Apr. 2024, doi: 10.1038/s41597-024-03222-2.

[13] A. Greco, G. Valenza, A. Lanata, E. P. Scilingo, and L. Citi, "cvxEDA: A Convex Optimization Approach to Electrodermal Activity Processing," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 4, pp. 797–804, Apr. 2016, doi: 10.1109/TBME.2015.2474131.

[14] Y. R. Veeranki, N. Ganapathy, and R. Swaminathan, "Analysis of Fluctuation Patterns in Emotional States Using Electrodermal Activity Signals and Improved Symbolic Aggregate Approximation," *Fluct. Noise Lett.*, vol. 21, no. 02, p. 2250013, Apr. 2022, doi: 10.1142/S0219477522500134.

[15] H. Sanz, C. Valim, E. Vegas, J. M. Oller, and F. Reverter, "SVM-RFE: selection and visualization of the most relevant features through non-linear kernels," *BMC Bioinformatics*, vol. 19, no. 1, p. 432, Nov. 2018, doi: 10.1186/s12859-018-2451-4.