

Abnormal Driving Detection using GPS Data

Charles Boateng Florida Atlantic University Boca Raton, USA Email: cboateng2023@fau.edu	Kwangsoo Yang Florida Atlantic University Boca Raton, USA Email: yangk@fau.edu	Seyedeh Gol Ara Ghoreishi Florida Atlantic University Boca Raton, USA Email: sghoreishi2021@fau.edu	Jinwoo Jang Florida Atlantic University Boca Raton, USA Email: jangj@fau.edu
Muhammad Tanveer Jan Florida Atlantic University Boca Raton, USA Email: mjan2021@fau.edu	Joshua Conniff Florida Atlantic University Boca Raton, USA Email: jconniff@health.fau.edu	Borko Furht Florida Atlantic University Boca Raton, USA Email: bfurht@fau.edu	Sonia Moshfeghi Florida Atlantic University Boca Raton, USA Email: smoshfeghi@fau.edu
David Newman Florida Atlantic University Boca Raton, USA Email: dnewman@fau.edu	Ruth Tappen Florida Atlantic University Boca Raton, USA Email: rtappen@fau.edu	Jinnan Zhai Florida Atlantic University Boca Raton, USA Email: jzhai@fau.edu	Monica Rosselli Florida Atlantic University Boca Raton, USA Email: mrossell@fau.edu

Abstract—Given a GPS dataset comprising driving records captured at one-second intervals, this research addresses the challenge of Abnormal Driving Detection (ADD). The study introduces an integrated approach that leverages data preprocessing, dimensionality reduction, and clustering techniques. Speed Over Ground (SOG), Course Over Ground (COG), longitude (lon), and latitude (lat) data are aggregated into minute-level segments. We use Singular Value Decomposition (SVD) to reduce dimensionality, enabling K-means clustering to identify distinctive driving patterns. Results showcase the methodology’s effectiveness in distinguishing normal from abnormal driving behaviors, offering promising insights for driver safety, insurance risk assessment, and personalized interventions.

Keywords—GPS, Singular Value Decomposition

I. INTRODUCTION

Given a GPS dataset comprising driving records captured at one-second intervals, this research addresses the challenge of Abnormal Driving Detection (ADD). The study introduces an integrated approach that leverages data preprocessing, dimensionality reduction, and clustering techniques to classify drivers as normal or abnormal based on their GPS driving records. Applying this approach clearly distinguishes between normal and abnormal driving behaviors, which is significant for domains such as driver safety, insurance risk assessment, and personalized interventions. Figures 1 and 2 shows how we start with a matrix representing driving records with three distinct driving features: Speed Over Ground (SOG), Course Over Ground (COG), and Latitude (Lat). Using Singular Value Decomposition (SVD), we reduce the dimensionality of this matrix to two latent features x_1 and x_2 . The first column of the SVD decomposition reveals that rows 1, 2, and 6 share similarity, while rows 3, 4, and 5 exhibit a distinct pattern. We subsequently apply the K-means clustering algorithm to classify drivers into two distinct homogeneous groups. K-means groups rows 1, 2, and 6 into one cluster and rows 3, 4, and 5 into another. This toy example illustrates the

$$\begin{bmatrix} \text{SOG} & \text{COG} & \text{Lat} \\ 9 & 9 & 9 \\ 9 & 9 & 9 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 9 & 9 & 9 \end{bmatrix} \xrightarrow{\text{SVD}} \begin{bmatrix} -0.57 & -0.57 \\ -0.82 & 0.40 \\ -0.06 & 0.03 \\ -0.06 & 0.03 \\ -0.06 & 0.03 \\ -0.57 & 0.40 \end{bmatrix} \cdot \begin{bmatrix} -0.57 & -0.57 \\ -0.82 & 0.40 \\ -0.06 & 0.03 \end{bmatrix}$$

Fig. 1: Dimensionality Reduction using SVD



Fig. 2: Outcome of K-means clustering (best in colors)

capability of our methodology to distinguish normal and abnormal driving patterns through dimensionality reduction and clustering.

A. Application Domain

The ADD problem holds significance across various domains, spanning transportation safety, data analytics, and societal benefits. This section discusses the critical applications and societal implications of this problem.

One of the significant applications of abnormal driving detection is the enhancement of transportation safety. In the modern world, ensuring the safety of drivers and passengers is of utmost importance. ADD systems serve as vigilant watchdogs, capable of identifying potentially hazardous situations

such as sudden braking, aggressive maneuvers, or erratic speed changes. Recognizing these anomalies can reduce the risk of accidents and save lives on the road. For example, ADD systems can be used to develop driver assistance systems that warn drivers of potential hazards and provide corrective interventions. These systems can also be used to identify high-risk drivers and provide them with targeted training or interventions to improve their driving safety. [1].

It also provides insights into driver's conduct. This enables us to identify patterns such as distracted driving, drowsy driving, or even signs of impaired driving due to substances. Fleet managers can utilize this technology to optimize routes, reduce fuel consumption, and promote safe driving practices among their drivers [2].

An often-overlooked application is in the realm of elderly and dementia care. For older individuals, especially those with dementia, maintaining independence while ensuring their safety is a delicate balance. ADD systems can be integrated into vehicles to monitor driving behavior. In cases where deviations from standard patterns are detected, caregivers or family members can be alerted, ensuring prompt intervention when necessary. [3].

Another vital application is that insurance companies can use ADD to assess insurance risk. Insurance companies typically use a variety of factors to determine a driver's risk level, including driving history, vehicle type, and location. ADD data can provide additional insights into a driver's behavior, such as their hard brakes and acceleration, lane change frequency, and oversteering patterns. This information can help insurance companies to assess risk and set premiums accordingly more accurately [4], [5].

In crisis scenarios such as natural disasters or medical emergencies, efficient and rapid response is critical. ADD can play a pivotal role in optimizing emergency response. By monitoring the driving behavior of emergency service vehicles, traffic conditions, and road accessibility, authorities can make informed decisions about resource allocation and routing. This ensures that emergency responders reach their destinations swiftly, saving time and potentially lives. Overall, ADD has the potential to revolutionize the way we conduct emergency response by reducing response time, saving lives, and protecting property in the event of a crisis.

In the context of urban populations and smart city initiatives, ADD contributes to effective traffic management. By identifying traffic bottlenecks, accidents, or congested areas in real-time, city planners can make data-driven decisions to alleviate traffic congestion and enhance overall mobility. This reduces commuting times and minimizes fuel consumption and environmental impact.

The ADD problem carries significant weight across various domains. Its societal and practical applications are profound, from saving lives on the road and fostering safer driving practices to facilitating independent living for older individuals and enabling more efficient emergency responses. As we delve deeper into this research, we aim to harness the power of data analytics and machine learning to address these critical

challenges effectively.

B. Problem Definition

In our formulation of the ADD problem, we start with a dataset comprising driving segments, each containing timestamped spatial points and GPS-derived features. This dataset is treated as a matrix, where each row represents different driving features corresponding to a timestamped point record. To distill meaningful insights from the multitude of features, we apply Singular Value Decomposition (SVD) to the matrix, thereby reducing its dimensionality to two essential features, denoted as x_1 and x_2 . Subsequently, we employ the K-Means clustering algorithm to partition the dataset into distinct clusters based on these derived features. The ADD problem can be defined as follows:

Input: A dataset of driving segments, each comprising timestamped spatial points and GPS-derived features.

- D represents the set of driving features for each segment. $D = d_1, d_2, d_3, \dots, d_N$, where N = number of instances
- Feature X is the set of derived driving features including the normalized change for the speed over ground (S), course over ground (C), longitude (lon), and latitude (lat).
- $M = X(t) - X(t - 1)$, which represents the difference between two successive time points of a driving feature.
- Normalization change $M_n = (M - \mu)/\sigma$
- $x(d)$ is the feature vector for the driving instance $d \in D$ where n_i is the number of time points of a driving instance d_i , and M_j^i is a normalized change in a driving feature for instance i at the j^{th} point of the trajectory

$$X = \begin{bmatrix} S_1^i & C_1^i & \dots & lon_1^i \\ S_2^i & C_2^i & \dots & lon_2^i \\ \vdots & \vdots & \ddots & \vdots \\ S_{n_i}^i & C_{n_i}^i & \dots & lon_{n_i}^i \end{bmatrix}$$

Output: A set of cluster labels that assigns each driving instance to a specific cluster based on its driving behavior pattern. Cluster Assignments = $\{c(d_1), c(d_2), c(d_3), \dots, c(d_4)\}$, where $c(d_i) \in \{1, 2\}$ denotes the cluster assignment for instance d_i .

Objective: Minimize the dissimilarity between driving records by partitioning them into two distinct homogeneous groups, enabling the accurate classification of normal and abnormal driving behaviors.

Our contributions: Our research addresses the unique challenges posed by the Abnormal Driving Detection (ADD) problem using real-world GPS data extracted from actual drivers. This problem has been largely unexplored in the context of ADD, where existing algorithms are primarily designed for other domains. Unlike synthetic datasets typically used in research, our work leverages authentic GPS data, offering a more accurate representation of real driving behaviors and patterns. Building upon the foundation of authentic data, we introduce novel solutions to effectively detect abnormal driving behaviors from real-world GPS data. Our approach involves a two-step process:

(a) Dimensionality Reduction with SVD: We employ Singular Value Decomposition (SVD) to reduce the dimensionality of the [6] GPS data matrix, transforming it into a compact representation with two principal features, denoted as x_1 and x_2 .

(b) Clustering with K-Means: We apply the K-Means clustering algorithm to identify [7] distinct clusters within the dataset. This step aims to group similar driving behaviors and pinpoint potential anomalies.

In summary, our contributions encompass the introduction of real-world GPS data, the development of novel algorithmic solutions, the introduction of efficiency-enhancing techniques, and a thorough evaluation of our approaches. This collective effort advances the field of abnormal driving detection using GPS data and paves the way for more accurate, scalable, and practical solutions.

C. Related Work

In the domain of abnormal driving detection, prior research efforts have predominantly revolved around the use of generative data models, such as Hidden Markov Models (HMMs) and Gaussian Mixture Models (GMMs) [8], [9]. These models aim to simulate driving behaviors by relying on assumed patterns and statistical distributions. While these generative data-based approaches have contributed valuable insights, they introduce synthetic elements into the data, potentially failing to capture the complexities of real-world driving behaviors.

Another notable approach to abnormal driving detection is the use of support vector machines (SVMs) [10]. SVMs are a class of machine learning algorithms that can be used for classification and regression tasks. To detect abnormal driving behaviors, SVMs can be trained on a dataset of normal and abnormal driving data. Once trained, the SVM can be used to predict whether a new driving behavior is normal or abnormal. However, SVMs can be sensitive to the dimensionality of the feature space. If the feature space is too high-dimensional, it can be difficult to train and deploy the SVM classifier.

In contrast to generative data-based approaches and SVMs, our study utilizes authentic GPS data acquired from sensors embedded in actual vehicles driven by human operators. This approach affords a considerable advantage in terms of data authenticity and reliability. By utilizing genuine GPS data collected from real driving experiences, we ensure that our analysis aligns with the intricate and nuanced nature of authentic driving behaviors. This authenticity enhances the accuracy and robustness of our abnormal driving detection system.

Our research introduces a novel and integrated methodology for detecting abnormal driving behaviors. By harnessing data preprocessing techniques, dimensionality reduction through Singular Value Decomposition (SVD), and K-means clustering, we achieve significant advancements in accurately discerning normal and abnormal driving behaviors. The adoption of minute-level segments in the feature aggregation

process allows for a fine-grained analysis of driving patterns, improving our capacity to identify subtle deviations from typical behavior.

D. Scope and Outline

Scope: Our research focuses on the development of an abnormal driving detection system utilizing GPS data. We analyze GPS-derived features and timestamped spatial points to identify and classify abnormal driving behavior. The project entails data preprocessing, feature extraction, clustering, and anomaly detection techniques.

Outline: The rest of the paper is organized as follows: Section II outlines the methodologies and algorithms employed for ADD using GPS data. It covers data preprocessing, feature extraction, clustering, and anomaly detection techniques. Section III describes the results of experiments and evaluations conducted to assess the performance of our abnormal driving detection system using real-world GPS datasets. Section V discusses potential directions for future research and enhancements in the field of abnormal driving detection.

II. METHODOLOGY

In this chapter, we present the methodology used for abnormal driving detection using GPS data. We begin by introducing the key features that form the basis of our approach. These features are essential for characterizing various driving behaviors. We then describe the algorithmic approach used for abnormal driving detection.

A. Speed Change (*Hard-braking and Hard Acceleration*)

Speed Change measures the rate at which the vehicle's Speed Over Ground (SOG) changes between successive time points. It is calculated as the difference in SOG values between two consecutive time points divided by the time interval (1 second). This metric helps identify abrupt changes in speed, such as hard braking or hard acceleration. Such sudden changes can indicate potentially risky driving behavior.

B. COG Change (*Oversteering and Lane-Changing*)

COG Change quantifies the change in Course Over Ground (COG) between two successive timepoints. Like Speed Change, it detects rapid changes in the vehicle's direction of travel. An abrupt shift in COG suggests oversteering or lane-changing maneuvers, which can be indicative of aggressive driving behavior.

C. Vehicle Direction

Vehicle Direction captures the direction in which the vehicle moves based on changes in longitude and latitude coordinates. It calculates the arctangent of the difference in latitude and longitude values between two successive time points. Changes in vehicle direction can indicate maneuvers such as weaving or swerving.

D. Turn Sharpness

Turn Sharpness assesses the sharpness of turns made by the vehicle. It is calculated based on changes in longitude and latitude between consecutive time points. It measures the Euclidean Distance traveled in the 2D plane during a time interval of 1 second. Sharp turns can imply erratic driving, making this metric valuable for detecting abnormal driving patterns.

Algorithm 1 Abnormal Driving Detection Algorithm

Input:

- Dataset D with driving records
- Desired segment duration δ

Output:

- Labeling result L
- 1) **Preprocess the data:**
 - a) Segment D based on the desired segment duration, δ .
 - b) Calculate matrix M.
 - 2) **Perform SVD (Singular Value Decomposition):**
 - a) Apply SVD to matrix M_n .
 - b) Select the top two components from SVD (i.e., x_1 and x_2).
 - 3) **Perform K-Means Clustering on the components**
 - 4) **Output:**
 - Return L
-

Algorithm 1 is designed to identify and label abnormal driving behavior from a dataset of driving records. It takes two main inputs: the dataset, which contains timestamped spatial points and GPS-derived features, and a desired segment duration to define the time window for analysis. The algorithm proceeds through several key steps to achieve its objective. First, it preprocesses the data by segmenting the dataset based on the specified segment duration and calculates a matrix from the segmented data. Next, it employs SVD to reduce the dimensionality of the data and transforms the dataset into a more compact form, extracting two significant components. To categorize driving behavior, the algorithm employs K-Means clustering on these two components. The final output of Algorithm 1 is a labeling result (L), which indicates whether a particular driving segment is classified as normal or abnormal based on its characteristics in the feature space.

III. EXPERIMENTAL EVALUATION

A. Experimental Layout

In this section, we present the experimental layout, data analysis procedures, and key findings from our investigation into driver behavior profiling and classification. Our dataset consists of older drivers, typically between the ages of 65 and 85 years, with a focus on individuals who may have conditions like dementia. These drivers participated in a research project funded by the National Institutes of Health (NIH), which spanned over three years. The project involved the collection

of driving behavior data using in-vehicle telematic sensors. These sensors recorded a variety of information, including precise GPS data, which serves as the core of our research into driver behavior profiling and classification. This rich dataset provides valuable insights into the driving patterns and behaviors of elderly drivers.

B. Experimental Results and Analysis

1) *Experiment Results:* In our experiment, we aimed to identify unique driving behavior clusters within individual driving segments. We focused on the precise identification of unique driving behavior clusters within individual driving segments, with a clear demarcation between training and testing phases. The testing phase involved the selection of two drivers for whom we had prior knowledge of the results, ensuring rigorous validation of our model's ability to detect abnormal driving behavior. In the first stage of the experiment, we divided each driving instance into segments. This process laid the groundwork for clustering these segments. Figures 3 and 4 shows the outcomes of the clustering process. Figure 3 provides insight into the segments of a driving instance, showing the occurrences of labels 0 and 1 that categorize distinct driving behaviors within the segments. Each label corresponds to a specific cluster, facilitating a comprehensive understanding of driving patterns within these segments. Figure 4 showcases segments from a different driver in our test dataset, revealing the occurrences of labels 0 and 1. It effectively distinguishes driving behavior patterns within these segments.

2) *Day-of-Driver Clusters:* Our experiment advanced by grouping driving instances by day, using a predefined threshold. Figure 5 is an illustrative representation of this phase. Figure 5 demonstrates the clustering results by day, with clustering decisions driven by the number of abnormal instances within each driver's driving instance. The threshold set allows for the categorization of each day's behavior pattern, providing insights into the consistency or variability of driver behavior daily.

3) *Driver Classification:* In the final phase of our experiment, we implemented a precise method for classifying drivers as either "normal" or "abnormal". This critical classification process was rooted in the careful evaluation of a driver's behavior over time, specifically focused on the accumulation of abnormal instances across multiple days.

Our approach was straightforward yet effective: if a driver consistently exceeded a predefined threshold for the number of abnormal driving instances within a specified timeframe, they were categorized as "abnormal". In contrast, drivers who predominantly exhibited normal driving instances, maintaining stability over time, were rightfully identified as "normal."

This classification strategy leveraged the accumulation of abnormal instances over days, which offered a panoramic view of a driver's behavior. By considering both the segment-level analysis and the broader temporal dimension, our approach ensured the comprehensive identification and classification of abnormal driving patterns.

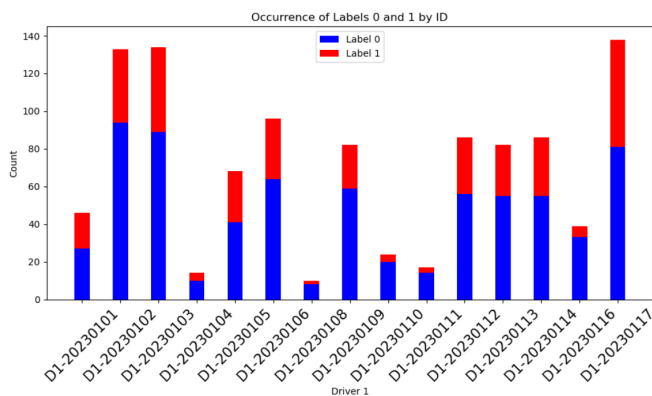


Fig. 3: Driver 1 cluster segments

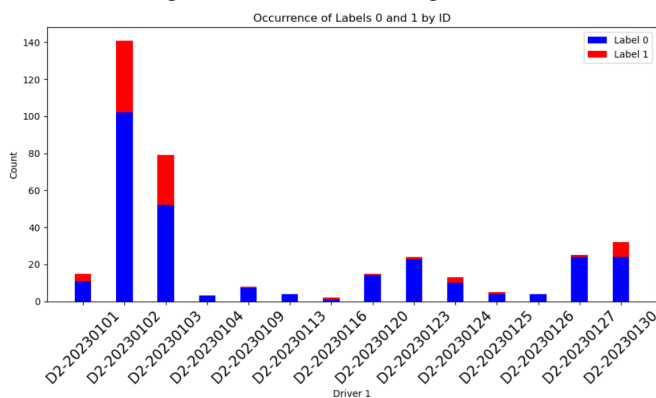


Fig. 4: Driver 2 cluster segments

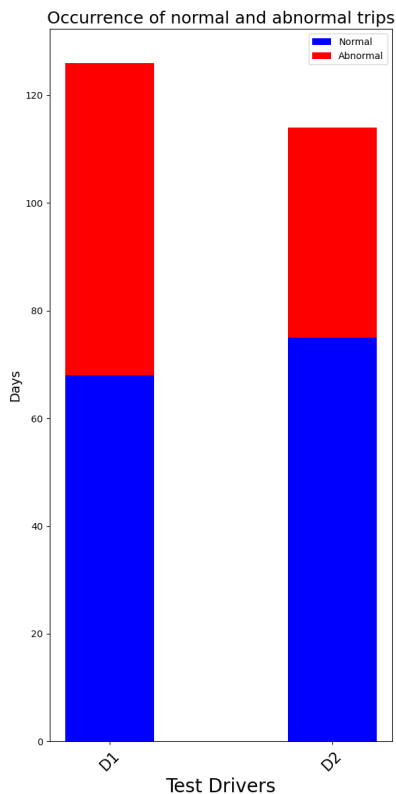


Fig. 5: Clustering drivers on each day

Our holistic approach to driver classification considered segment-level clustering, time-based analysis, and the accumulation of abnormal instances over multiple days. This strategy provided us with a thorough and accurate means of distinguishing between normal and abnormal driver behavior, a fundamental step in our research into driver behavior profiling and classification.

4) *Aligning with the Data:* One notable observation during our analysis was the strong alignment between our test data and the actual driving behaviors of the individuals under study. During the validation phase of our experiment, our clustering and classification techniques achieved an astonishing level of precision. The driving behaviors of two test drivers, whose data was deliberately excluded from our training dataset, served as a crucial point of validation. One of these drivers had consistently exhibited normal driving behaviors, while the other displayed a distinctive pattern of abnormal driving behaviors. Despite the intentional exclusion of their data from the training phase, our approach accurately categorized these test drivers. The driver with a history of normal behaviors was precisely classified as “normal”, while the driver characterized by abnormal behaviors was accurately marked as “abnormal”.

This extraordinary alignment between our test results and the known driver behaviors underscores the robustness and reliability of our approach. It not only validates the effectiveness of our model but also demonstrates its capacity to provide a precise representation of real-world driving behaviors

IV. DISCUSSION

Our research demonstrates that GPS data can be used to enhance road safety and gain valuable insights into driver behavior, particularly among elderly drivers. By developing an effective approach for identifying and classifying abnormal driving patterns, we have paved the way for the development of cutting-edge solutions to reinforce road safety for older drivers. Our findings have significant implications for both road safety and our understanding of driver behavior. By proactively identifying drivers at risk and implementing targeted interventions, we can prevent accidents and improve the safety of our roadways. Additionally, our research can help us to understand better the driving patterns of older adults, which can be used to develop personalized interventions to improve their driving safety and promote independent living. Overall, our research constitutes a seminal contribution to the field of abnormal driving detection using GPS data. We have established a robust foundation for future explorations within this domain. Our novel approach possesses the transformative potential to reshape the way we safeguard the lives of all drivers, with a particular focus on elderly drivers.

V. CONCLUSION AND FUTURE WORK

In this section, we outline potential directions for future research and enhancements to our driver behavior profiling and

classification system. While our experiment primarily focused on GPS-derived features, there is ample room for improvement by incorporating data from other vehicle sensors. Integration of sensor data from accelerometers, gyroscopes, and even cameras could provide a more comprehensive view of driver behavior. Future work could explore the fusion of multi-modal sensor data for enhanced accuracy in profiling. One aspect to consider in future research is the sensitivity of K-means clustering to initial cluster centroids. Different initializations can yield distinct results, and understanding this variability is vital. Robustness in clustering algorithms needs to be explored, and the development of methods to mitigate this sensitivity should be a priority.

Additionally, our research did not explicitly address the influence of external factors such as weather conditions, road quality, or traffic density on driving behavior. Incorporating data from other vehicle sensors, such as accelerometers, gyroscopes, and cameras, could provide a more comprehensive view of a driver's behavior in varying external conditions.

Future work should aim to move beyond classification and towards a driver-centric risk assessment system. Such a system would not only classify behavior but also evaluate the risk associated with each driver's actions. This would be especially valuable for insurance companies seeking to refine risk assessment models and ultimately improve road safety.

To validate the scalability and robustness of our approach, we recommend large-scale deployments across diverse driving environments and populations. This would allow for a more comprehensive evaluation of the impact of external factors and further refine our methods.

In this way, our research not only serves as a promising proof of concept but also highlights critical areas for future exploration and development in the realm of driver behavior profiling and classification. By delving into these directions, we can enhance our understanding of driver behavior and contribute to safer roadways for all.

VI. ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation CAREER under Grant No. 1844565, and the National Institutes of Health under Grant No. 1R01AG068472.

REFERENCES

- [1] Fonseca, J., Aparicio, J., Pereira, F. L., Santos, B. S. (2018). A review on the applications of driving monitoring systems in vehicular networks. *IEEE Communications Surveys Tutorials*, 20(2), 1082-1111.
- [2] Nawaz, T., Abbas, H., Hussain, M. (2020). Advanced driver assistance systems: Challenges, solutions, and future prospects. *IEEE Transactions on Intelligent Transportation Systems*, 21(7), 2733-2744.
- [3] Gorr, W., Olligschlaeger, A., Thompson, Y., Harney, D. (2016). Spatial and temporal analysis of crash risk at non-recurrent congestion locations. *Accident Analysis Prevention*, 87, 113-123.
- [4] Sarkar, S., Chien, S., Ban, X. (2016). Data mining techniques for anomaly detection in traffic data. *IEEE Transactions on Intelligent Transportation Systems*, 17(2), 471-480.
- [5] D. Zhou, J. Wang, J. Lu, and S. Gao, "Abnormal driving detection based on a deep learning approach," *Sensors*, vol. 19, no. 19, p. 4305, 2019.
- [6] Anandarajan, Murugan, et al. "Semantic Space Representation and Latent Semantic Analysis." *Advances in Analytics and Data Science*, 2019.
- [7] "Immunophenotype Discovery, Hierarchical Organization, and Template-Based Classification of Flow Cytometry Samples." 2016.
- [8] J. Wang, S. He, and D. Zhang, "Abnormal driving detection using hidden Markov models," in *Proceedings of the 2011 IEEE International Conference on Intelligent Transportation Systems*, 2011, pp. 113-118.
- [9] Y. Zhang, W. Wang, and M. Li, "Abnormal driving detection using Gaussian mixture models," in *Proceedings of the 2012 IEEE Intelligent Vehicles Symposium*, 2012, pp. 1119-1124.
- [10] S. Wang, P. Wang, J. Zhao, and L. Zhang, "Abnormal driving detection based on feature selection and support vector machines," in *Proceedings of the 2013 IEEE Intelligent Transportation Systems Conference*, 2013, pp. 2581-2586.