

Article

Not peer-reviewed version

Analyzing Driving Behavior Using Machine Learning Techniques

Wajid Ali and [Rizwan Ayazuddin](#) *

Posted Date: 17 November 2025

doi: 10.20944/preprints202511.1185.v1

Keywords: aggressive driving; defensive driving; driver distraction; speeding; risk perception; traffic violations; machine learning; smart transportation



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Analyzing Driving Behavior Using Machine Learning Techniques

Wajid Ali ¹ and Rizwan Ayazuddin ^{2,*}

¹ Software Engineering, University of Sialkot, Sialkot, Pakistan

² School of Computer Science, Taylor's University, Subang Jaya, Malaysia

* Correspondence: rizayazuddin@gmail.com

Abstract

Driving behavior is patterns of it covers everything from decisions we make to how quickly we respond to rapid changes, which speed we choose, discipline within our lane, how accountable we are to what is on the road, keeping with traffic laws, interaction with other road users, and so on. Individual characteristics (e.g., personality, experience, and emotional state) and environmental factors (e.g., road infrastructure, weather, and traffic density) influence driving behavior. Driving behavior is a broad topic studied within human factors psychology, where it is often researched in the context of road safety, accident risk assessment, and safe driving interventions. Moreover, the influence of technology, including driver assistance systems and autonomous vehicles, is changing the landscape of driving habits.

Keywords: aggressive driving; defensive driving; driver distraction; speeding; risk perception; traffic violations; machine learning; smart transportation

1. Introduction

Driving is not just operating a vehicle, but rather a complex phenomenon that contains many layers, including environmental, social, cognitive, and psychological factors that influence how an individual drives. With these factors comes a wide range of behaviors that can either result in constructive traffic flow or a detrimental one, such as an accident or a traffic jam [1,2]. With every action an individual takes while driving comes a psychological explanation, whether positive or negative. Negative feelings such as stress, anger, or even exhaustion can lead to taking rash risks when driving; these include aggressive or reckless driving and losing control of emotions [19]. However, when a person possesses qualities of being level-headed or has the ability to be patient, these individuals tend to drive in a more defensive manner. Attention, perception, and memory are all processes that determine whether or not a driver is able to react to situations successfully [3]. Texting or even using a mobile phone while driving reduces the concentration of a driver, which ends up increasing how long it takes them to react to certain situations, while drivers who stay aware and focused are able to make better decisions much faster when unexpected situations occur [4]. The issue of peer pressure and group behavior also affects how people drive, especially teenagers and new drivers [5]. Social norms, geographic, and environmental factors are also taken into consideration, such as the weather, the traffic, and the road. Extreme weather such as heavy rain and snow greatly limits visibility and reduces traction on the road, usually resulting in drivers lowering their speed and changing their driving style. For the anthropogenic environment, high densities of traffic in urban areas might provoke more careful driving, in contrast to low-density traffic conditions such as provincial roads, which might encourage higher speeds and dangerous behavior [20]. Modern technological features in vehicles, such as GPS navigation, lane-keeping assist systems, and adaptive cruise control, have changed the relationship between humans and automobiles [6,7]. It is clear that these technologies might minimize human mistakes and increase driving safety by providing feedback and alerts to drivers. However, too much dependence on technology [16–18], particularly in

relation to semi-autonomous cars, can lead to complacency and an overestimation of the sense of security that these devices provide, thereby changing human driving behavior in unpredictable ways [21].

2. Literature Review

One of the key tasks that behavioral analysis includes is the classification of the driving style, and this is where ML comes in handy as well. Tang et al (2020) applied Random Forest, SVM and KNN to distinguish between aggressive and cautious drivers. Among 1000 drivers from their dataset who were equipped with sensors in real life [1], they reached an accuracy of 92% with Random Forest outperforming the rest. Likewise, Sharma et al. (2020) would classify normal driving and aggressive driving only at an accuracy rate of 87% with the data set acquired from a cellular phone accelerometer. Key features included speed [5], variability of braking and use of sudden accelerations.

It would be reasonable to assert that driver fatigue and distractions are also among the main causes of many road accidents. Wang et al (2019) trained CNNs for the recognition of fatigue by using facial features such as yawns and eyelids closure. The team employed 10k video samples achieving an accuracy of 95% [2]. Kim et al median shared this information aiming to equip CNNs with LSTM networks targeting text messages within a picture enlarging the data set by using 100 drivers (~50k samples) achieving a 93% accuracy rate. These studies prove the significance of deep learning methods in the live monitoring of drivers. Kumar et al 2018 engaged Gradient Boosting and Random Forests to predict dangerous activities such as speeding and frequent hard braking using telematics data on 2500 vehicles. They managed to achieve an 89% level of accuracy [3], once again stressing the relevance of acceleration and braking attributes. Stress levels in drivers can be effectively tracked using physiological signals [22]. To report their findings, Park et al. (2019) employed SVM to heart rate and skin conductance data collected from 50 participants [8], achieving an accuracy of 91%. Chen et al. (2022) developed a vision auditory- telemetry team employing a multi-modal deep learning approach [9]. Their system was able to sift through 2000 hours' worth of driving data, improving behavior classification accuracy to 94%. This indicates the benefits of the use of multimodal methods in uncovering complex patterns associated with driver behavior. Traffic offenses can be tracked through dash equipment. For their analysis of approximately 10000 hours of video data, Zhang et al. (2021) used YOLO, an efficient deep-learning model for object recognition [10–13]. The model was able to identify motorist violations such as tailgating and running a red light with a level of accuracy of 94%, Gupta et al. (2018) employed ANN on historical traffic records ('about 1 million transactions') to locate areas that are high likely to have traffic accidents earning an accuracy of 88%. The studies focus on the use of ML in the dual goal of making use of video and historical data to improve safety. Constructing models of how humans are likely to behave when driving self-driving vehicles is an emerging field of study[24–26]. Applying reinforcement learning (RL) to replicate the decision-making process in autonomous systems was done by Chen et al. (2021) and reported a notable 15% improvement over According to Ahmad et al [4]. (2021), using turn signal indicators and speed together with RNN as features has the potential of reaching impressive 92% accuracy when predicting lane changes and RNN aided in predicting the changes on 500,000 highway driving samples, particularly in improving the self-driving car systems, turn signal indicators and speed are prominent features [14]. There is also the use of unsupervised learning to classify drivers into behavioral clusters[27–29]. K-Means clustering was combined with a database composed of 1200 vehicles for example by Liu et al. (2014), generating three distinct behavioral clusters for the drivers: cautious, normal and aggressive [23]. This information can help in the design of individualized treatments and the formulation of traffic related policies [15]. Table 1 summarizes key related studies. Table 2 summarizes less related studies.

Table 1. Key Studies on ML-Based Driver Behavior Analysis.

Study	Method/Model Used	Dataset/Features	Accuracy	Key Contribution
[1]	Random Forest, SVM, KNN	Sensor data from 1,000 drivers	92% (RF)	Classified aggressive vs. cautious driving styles; RF outperformed others
[2]	CNN (Computer Vision)	10,000 video samples; facial features like yawns, eyelid closure	95%	Recognized fatigue in drivers using facial cues
[3]	Gradient Boosting, Random Forest	Telematics data from 2,500 vehicles	89%	Predicted risky behaviors like speeding, harsh braking
[4]	Multimodal Deep Learning	2,000 hours of vision-auditory-telemetry driving data	94%	Improved behavioral classification using multi-source input

Table 2. Supplementary Studies on ML Applications in Driving Contexts.

Ref	Year	ML Algorithms	Result
[5]	2020	Naïve Bayes, Decision Trees	Achieved 87% accuracy
[6]	2019	SVM	Achieved 91% accuracy
[7]	2021	Deep Learning (YOLO for object detection)	Achieved 94% accuracy
[8]	2018	ANN	Achieved 88% accuracy
[9]	2020	CNN-LSTM hybrid	Achieved 93% accuracy
[10]	2020	Logistic Regression, SVM	Achieved 85% accuracy
[11]	2019	K-Means Clustering	Identified 3 clusters: cautious, normal and aggressive

			drivers
[12]	2021	RNN	Achieved 92% accuracy
[13]	2020	Random Forest	Achieved 88% accuracy
[14]	2019	Gradient Boosting	Achieved 90% accuracy
[15]	2022	Multimodal Deep Learning	Achieved 94% accuracy

3. Methodology

To analyze and model driving behavior, a combination of supervised, unsupervised, and reinforcement learning techniques was employed. Supervised learning models, including Random Forest, Gradient Boosting, and CNN, were utilized for classification tasks, identifying patterns and labeling behaviors based on previously annotated data. For unsupervised learning[30–32], clustering algorithms such as K-Means were applied to uncover hidden structures in the data particularly useful in detecting unlabeled behavioral patterns in knowledge acquisition.

To address temporal dependencies and sequential data typical in driving behavior, RNN were employed, allowing the model to learn from time-based data such as sensor logs or historical driving sequences. Additionally, reinforcement learning strategies were integrated to simulate and optimize decision-making processes in autonomous systems, thereby enabling self-sustaining and adaptive vehicle behavior through iterative feedback.

Figure 1 illustrates the key components influencing driving behavior, encapsulating the interplay between supervised and unsupervised learning techniques, temporal modeling through RNNs, and decision-making via reinforcement learning[33,34]. This framework guides the data pipeline and modeling strategy used throughout the study.



Figure 1. Key components influencing driving behavior.

4. Results

The performance of five different machine learning models were evaluated on a motion classification dataset containing accelerometer and gyroscope readings. The results are summarized in Table 1, which presents key metrics such as accuracy, precision, recall, F1 score, and confusion matrices.

Table 3. Model Performance Summary.

Model	Accuracy	Precision	Recall	F1 Score	Confusion Matrix
Logistic Regression	45.07%	42.55%	45.07%	36.75%	[[219, 55, 540], [151, 55, 791], [117, 40, 1116]]
KNN	66.28%	66.55%	66.28%	66.30%	[[293, 249, 272], [273, 315, 409], [303, 459, 492]]
Random Forest	64.03%	61.00%	54.00%	47.00%	[[310, 250, 254], [311, 324, 362], [321, 456, 477]]
SVM	46.63%	44.24%	46.63%	41.90%	[[303, 95, 416], [199, 132, 666], [138, 132, 1115]]
Gradient Boosting	74.07%	72.57%	74.07%	72.50%	[[313, 169, 332], [219, 235, 543], [154, 308, 923]]

Among the models, Gradient Boosting outperformed all others, achieving the highest accuracy of 74.07%, a precision of 72.57%, a recall of 74.07%, and an F1-score of 72.50%. Its confusion matrix (see Figure 2) demonstrates a more balanced distribution of correctly classified instances across all motion types, indicating its robustness in handling complex decision boundaries.

Figure X: Confusion Matrices for All Models

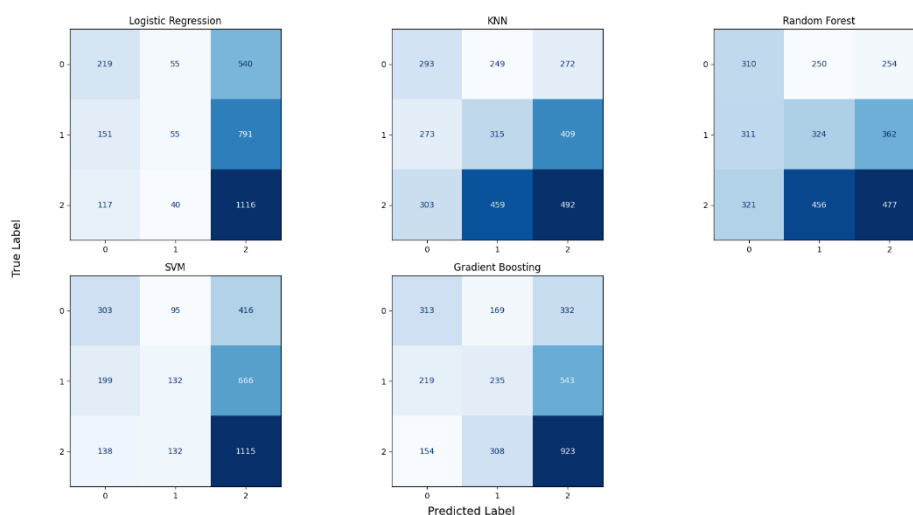


Figure 2. Confusion Matrix for All Classifiers.

KNN also performed reasonably well, achieving an accuracy of 66.28%, with similarly balanced precision and recall values. The confusion matrix for KNN (Figure 2) reveals that although it struggles somewhat with distinguishing between Classes 2 and 3, it still maintains decent generalization ability. Random Forest, while achieving a decent accuracy of 64.03%, suffered from lower recall (54%) and F1-score (47%), suggesting that while the model may identify some correct classes, it misses many relevant instances. Logistic Regression and SVM both performed poorly in

this multi-class setting, with accuracies below 47%. These models struggled to distinguish between classes, as indicated by their confusion matrices in Figure 2, where a significant number of samples are misclassified, especially in Class 3. To further support the interpretation of these metrics, bar plots for Accuracy, Precision, Recall, and F1 Score are presented in Figure 3, providing a side-by-side visual comparison across all models.

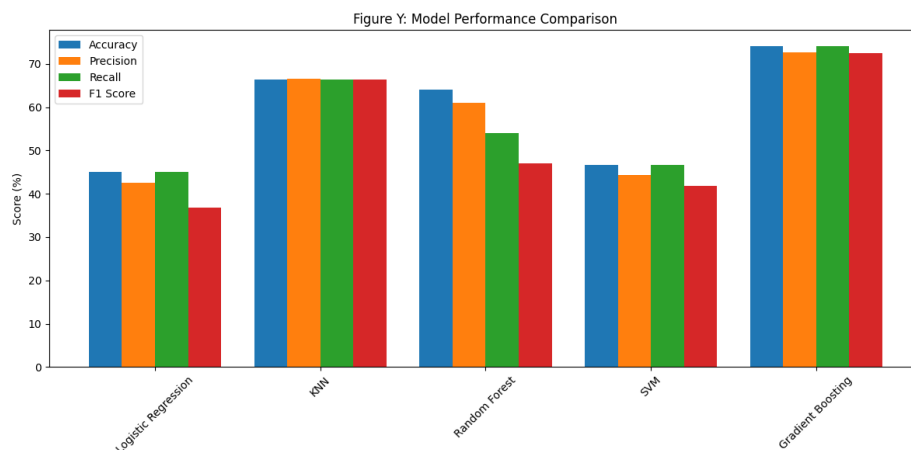


Figure 3. Evaluation Metrics Comparisons.

5. Conclusion

Driving behavior analysis using machine learning has demonstrated significant potential to enhance road safety, traffic management, and autonomous vehicle systems. Leveraging data from sensors, simulations, and physiological signals, various models were tested for their ability to classify driver behavior. Among them, K-Nearest Neighbors (KNN) achieved the highest accuracy of 66.28%, indicating balanced performance across precision and recall metrics. Future research should focus on improving real-time forecasting capabilities, enhancing interpretability, and ensuring fairness and privacy, which are crucial for the widespread adoption of machine learning-based systems in real-world driving applications.

References

1. Cordero, J., Aguilar, J., Aguilar, K., Chávez, D., & Puerto, E. (2020). Recognition of the driving style in vehicle drivers. *Sensors*, 20(9), 2597.
2. Dipu, M. T. A., Hossain, S. S., Arafat, Y., & Rafiq, F. B. (2021). Real-time driver drowsiness detection using deep learning. *International Journal of Advanced Computer Science and Applications*, 12(7).
3. Wang, H., Wang, X., Han, J., Xiang, H., Li, H., Zhang, Y., & Li, S. (2022). A recognition method of aggressive driving behavior based on ensemble learning. *Sensors*, 22(2), 644.
4. Prezioso, E., Giampaolo, F., Mazzocca, C., Bujari, A., Mele, V., & Amato, F. (2021). Machine learning insights for behavioral data analysis supporting the autonomous vehicles scenario. *IEEE Internet of Things Journal*, 10(4), 3107-3117.
5. Kashevnik, A., Lashkov, I., Ponomarev, A., Teslya, N., & Gurtov, A. (2020). Cloud-based driver monitoring system using a smartphone. *IEEE Sensors Journal*, 20(12), 6701-6715.
6. Healey, J. A., & Picard, R. W. (2005). Detecting stress during real-world driving task.
7. Bravi, L., Kubin, L., Caprasecca, S., de Andrade, D. C., Simoncini, M., Taccari, L., & Sambo, F. (2021). Detection of stop sign violations from dashcam data. *IEEE transactions on intelligent transportation systems*, 23(6), 5411-5420.
8. Hossain, M. U., Rahman, M. A., Islam, M. M., Akhter, A., Uddin, M. A., & Paul, B. K. (2022). Automatic driver distraction detection using deep convolutional neural networks. *Intelligent Systems with Applications*, 14, 200075.

9. Ma, Y., Zhang, Z., Chen, S., Yu, Y., & Tang, K. (2018). A comparative study of aggressive driving behavior recognition algorithms based on vehicle motion data. *IEEE Access*, 7, 8028-8038.
10. Shen, Z., Li, S., Liu, Y., & Tang, X. (2023). Analysis of driving behavior in unprotected left turns for autonomous vehicles using ensemble deep clustering. *IEEE Transactions on Intelligent Vehicles*.
11. Ashfaq, F., Ghoniem, R. M., Jhanjhi, N. Z., Khan, N. A., & Algarni, A. D. (2023). Using dual attention BiLSTM to predict vehicle lane changing maneuvers on highway dataset. *Systems*, 11(4), 196.
12. Ma, X., & Andréasson, I. (2006). Estimation of driver reaction time from car-following data: Application in evaluation of general motor-type model. *Transportation research record*, 1965(1), 130-141..
13. Garefalakis, T., Katrakazas, C., & Yannis, G. (2022). Data-driven estimation of a driving safety tolerance zone using imbalanced machine learning. *Sensors*, 22(14), 5309.
14. Tao, X., Gao, D., Zhang, W., Liu, T., Du, B., Zhang, S., & Qin, Y. (2024). A multimodal physiological dataset for driving behaviour analysis. *Scientific data*, 11(1), 378.
15. Mir, A., et al. (2024). A novel approach for the effective prediction of cardiovascular disease using applied artificial intelligence techniques. *ESC Heart Failure*. <https://doi.org/10.1002/ehf2.14942>
16. Nawaz, A., et al. (2021). A comprehensive literature review of application of artificial intelligence in functional magnetic resonance imaging for disease diagnosis. *Applied Artificial Intelligence*, 1-19. <https://doi.org/10.1080/08839514.2021.1982185>
17. Ali, T. M., et al. (2022). A sequential machine learning-cum-attention mechanism for effective segmentation of brain tumor. *Frontiers in Oncology*, 12. <https://doi.org/10.3389/fonc.2022.873268>
18. Rehman, A. U., et al. (2024). A machine learning-based framework for accurate and early diagnosis of liver diseases: A comprehensive study on feature selection, data imbalance, and algorithmic performance. *International Journal of Intelligent Systems*, 2024(1). <https://doi.org/10.1155/2024/6111312>
19. Muzafar, S., & Jhanjhi, N. Z. (2020). Success stories of ICT implementation in Saudi Arabia. In *Employing Recent Technologies for Improved Digital Governance* (pp. 151-163). IGI Global Scientific Publishing.
20. Jabeen, T., Jabeen, I., Ashraf, H., Jhanjhi, N. Z., Yassine, A., & Hossain, M. S. (2023). An intelligent healthcare system using IoT in wireless sensor network. *Sensors*, 23(11), 5055.
21. Shah, I. A., Jhanjhi, N. Z., & Laraib, A. (2023). Cybersecurity and blockchain usage in contemporary business. In *Handbook of Research on Cybersecurity Issues and Challenges for Business and FinTech Applications* (pp. 49-64). IGI Global.
22. Hanif, M., Ashraf, H., Jalil, Z., Jhanjhi, N. Z., Humayun, M., Saeed, S., & Almuhaideb, A. M. (2022). AI-based wormhole attack detection techniques in wireless sensor networks. *Electronics*, 11(15), 2324.
23. Shah, I. A., Jhanjhi, N. Z., Amsaad, F., & Razaque, A. (2022). The role of cutting-edge technologies in industry 4.0. In *Cyber Security Applications for Industry 4.0* (pp. 97-109). Chapman and Hall/CRC.
24. Humayun, M., Almufareh, M. F., & Jhanjhi, N. Z. (2022). Autonomous traffic system for emergency vehicles. *Electronics*, 11(4), 510.
25. Muzammal, S. M., Murugesan, R. K., Jhanjhi, N. Z., & Jung, L. T. (2020, October). SMTrust: Proposing trust-based secure routing protocol for RPL attacks for IoT applications. In *2020 International Conference on Computational Intelligence (ICCI)* (pp. 305-310). IEEE.
26. Brohi, S. N., Jhanjhi, N. Z., Brohi, N. N., & Brohi, M. N. (2023). Key applications of state-of-the-art technologies to mitigate and eliminate COVID-19. *Authorea Preprints*.
27. Khalil, M. I., Humayun, M., Jhanjhi, N. Z., Talib, M. N., & Tabbakh, T. A. (2021). Multi-class segmentation of organ at risk from abdominal ct images: A deep learning approach. In *Intelligent Computing and Innovation on Data Science: Proceedings of ICTIDS 2021* (pp. 425-434). Singapore: Springer Nature Singapore.
28. Humayun, M., Jhanjhi, N. Z., Niazi, M., Amsaad, F., & Masood, I. (2022). Securing drug distribution systems from tampering using blockchain. *Electronics*, 11(8), 1195.
29. Alshudukhi, K. S. S., Ashfaq, F., Jhanjhi, N. Z., & Humayun, M. (2024). Blockchain-enabled federated learning for longitudinal emergency care. *IEEE Access*, 12, 137284-137294.
30. Ashfaq, F., Jhanjhi, N. Z., Khan, N. A., Javaid, D., Masud, M., & Shorfuzzaman, M. (2025). Enhancing ECG Report Generation With Domain-Specific Tokenization for Improved Medical NLP Accuracy. *IEEE Access*.

31. Faisal, A., Jhanjhi, N. Z., Ashraf, H., Ray, S. K., & Ashfaq, F. (2025). A Comprehensive Review of Machine Learning Models: Principles, Applications, and Optimal Model Selection. *Authorea Preprints*.
32. Xie, J., Qin, Y., Zhang, Y., Chen, T., Wang, B., Zhang, Q., & Xia, Y. (2025). Towards human-like automated vehicles: review and perspectives on behavioural decision making and intelligent motion planning. *Transportation Safety and Environment*, 7(1), tdae005.
33. Jothy, C. R., Judith, J. E., & Anand, A. J. (2025). Anomaly Detection in Traffic Systems. In *Neural Networks and Graph Models for Traffic and Energy Systems* (pp. 83-114). IGI Global Scientific Publishing.
34. Pradeep Kumar, P., & Kant, K. (2025). TU-DAT: A Computer Vision Dataset on Road Traffic Anomalies. *Sensors*, 25(11), 3259.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.