Ausaf Ali Khan

Using AI to study impact of driving patterns

Master’s thesis of mathematical information technology

November 17, 2023

University of Jyväskylä

Department of Mathematical Information Technology

Author: Ausaf Ali Khan

Contact information: engr.ausafalikhan@gmail.com

Supervisors: Prof. Vagan Terziyan, Atri Vainikainen

Title: Using AI to study impact of driving patterns

Työn nimi: Tekoälyn hyödyntäminen liikennekäyttäytymisen tutkimisessa

Project: Master’s thesis

Study line: Information Technology

Page count: 7+69

Abstract:

This thesis examines the ways in which artificial intelligence (AI) can be used to study the impact of driving patterns, aiming to find a correlation between the variables influencing the driver's decision-making process by using data that can be gathered in various driving environments and terrains. This analysis will be helpful in developing a system that helps drivers modify their driving habits for increased vehicle efficiency and reduced damage to the environment. A T10G device from Aplicom Oy, containing important interfaces to the vehicle sensors via a CAN bus interface and on-device sensors that measure vehicle speed, latitude longitude and timestamp data, is used to analyze driving behavior. Data has been collected on multiple journeys both in-city and on highway in Finland and mapped onto the Finnish Transport Infrastructure Authority’s publicly available API which contains detailed mapping of Finnish road and street networks as well as winter and summer speed limits with Geo coordinates, thus providing an accurate picture of driving behavior along the aforementioned path. This study expresses how incorporating machine learning is a fundamental shift in driving that will make it safer, more efficient, and environmentally friendly. The driving experience of the future will see more involvement from human-machine interaction based on sustainability, safety, and accountability.

Keywords: Artificial intelligence, machine learning, driving behavior, driving efficiency, telematics.

Suomenkielinen tiivistelmä:

Tässä opinnäytteessä tutkitaan tapoja hyödyntää tekoälyä (AI) kuljettajan ajokäyttäytymisen analysointiin. Tavoitteena on löytää korrelaatio niiden eri tekijöiden väliltä jotka vaikuttavat kuljettajan tekemiin päätöksiin hyödyntäen kuljettajan ajoympäristöstä ja tiestöstä saatavilla olevaa ja kerättyä tietoa. Tämän pohjalta voidaan jatkossa kehittää järjestelmä joka auttaa kuljettajia kiinnittämään huomiota ajotapaan ja parantamaan ajamisen tehokkuutta ja vähentämään ajamisesta aiheutuu ympäristön kuormitusta.

Aplicom Oy on toimittanut työtä varten T10G-telematiikkalaitteen, josta löytyy mittalaitteet ajoneuvon nopeuden, sijainnin ja ajan seuraamiseksi. Lisäksi laitteesta löytyy CAN-liitäntä, jonka avulla päästään lukemaan ajoneuvon omia antureita. Näiden tietojan avulla kuljettajan ajotapaa voidaan analysoida. Opinnäytteen lähtötiedoiksi ajotietoa on kerätty useilla ajokerroilla Suomessa kaupunkiliikenteestä ja moottoritieajosta. Analyysissa kerätyt tiedot on yhdistetty Trafin julkisen Digiroad-aineiston kanssa josta on saatu tiestön yksityiskohtaista nopeusrajoitus- ja geometriatietoa. Muodostetusta tietomallista voidaan muodostaa tarkka kuva kuljettajan ajokäytöksestä kyseisellä tieosuudella.

Opinnäyte havainnollistaa osuvalla tavalla kuinka koneoppimista voidaan hyödyntää liikenteessä kuljettajan avustamiseksi tehden ajamisesta turvallisempaa, tehokkaampaa ja ympäristöystävällisempää. Tulevaisuuden liikenteessä tullaan hyödyntämään entistä enemmän tietokoneen ja kuljettajan yhteistyötä jotta vastuullisuuden, turvallisuuden ja kestävän kehityksen tavoitteet voidaan saavuttaa.

Avainsanat: Tekoäly, koneoppiminen, kuljettajan ajokäyttäytyminen, taloudellinen ajaminen, telematiikka

Preface

This master's thesis presents my research on the topic “Using AI to study impact of driving patterns” conducted under the umbrella of the Department of Mathematical Information Technology at University of Jyväskylä and Aplicom Oy.

Throughout this journey I have been able to gain great learning experiences and I would like to thank my supervisors Professor Vagan Terziyan and Atri Vainikainen. Their guidance and support have made it possible for me to accomplish this valuable research and create an impact in my field.

I would also like to take this chance and thank my colleagues for the wonderful discussions and sharing their knowledge and wisdom with me.

Jyväskylä, November 17, 2023

Ausaf Ali Khan

Glossary

ADAS	Advanced Driver Assistance Systems
CAN	Controller Area Network
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
MQTT	Message Queuing Telemetry Transport
OTAP	Over-The-Air Programming
RTC	Real-Time Clock

List of Figures

List of Tables

Contents

1 Introduction

In this thesis, the emphasis is on analyzing driving behaviors in different scenarios and identifying patterns in driving style of a driver for improved vehicle efficiency. It is commonplace nowadays for automotive companies to utilize technology to introduce digital interfaces for reading engine data as well as other electronic components to evaluate vehicle performance and diagnosis.

The approach used to perform such data analytics starts by first clarifying the goal that has to be achieved. In this thesis, the goal is to utilize the data collected with Aplicom T10G device in order to identify patterns or events that can be useful in analyzing driving behavior. Another important goal is to provide a thorough overview of different machine learning models with suitable driving parameters as input to process and how such models can help in identifying and correcting the driving behavior keeping in mind different terrains, external environment factors, driver's state, and vehicle data. Such statistical and logical analysis will be beneficial in creating an assisting system for a driver to correct their driving behavior for improved vehicle efficiency and reduced damage to the environment.

The collected data will be correlated, and an overview of appropriate machine learning algorithms will be provided which can be valuable in analyzing different scenarios where driving data is available. This would not only analyze the current state of the vehicle and driver but also predict an appropriate driving style in the given driving conditions such that the driver will be able to adapt to better driving practices and improve their driving score. It would also reduce the chances of wear and tear on the vehicle and improve fuel efficiency.

This research will prove to yield numerous benefits including increased work efficiency, reduced carbon emissions, improved vehicle diagnostics and eco-driving. Furthermore, a reward system can be introduced which can be based on a driving score that will increase or decrease based on driving efficiency.

This research aims to provide a framework of all the relating factors that could affect drivers' performance as well as a vehicle's efficiency and provide correlations among varied factors which could help in better-informed decision making. With the help of this system, visual feedback can be provided to the driver which would not only detail vehicle diagnostics data but also provide an insight on the driver's performance throughout the ride, potentially displaying statistics at the end of the trip which would differentiate between current performance and an ideal one for better efficiency.

2 KEY CONCEPTS

This chapter focuses on explaining some of the key terms and concepts which are important in order to thoroughly understand the research.

2.1 ADAS Systems

According to the National Highway Traffic Safety Administration (NHTSA), Advanced Driver Assistance Systems (ADAS) are technological advancements that introduce automation or improvements to vehicle systems, with the primary goal of enhancing safety and driving experience. These innovations assist drivers in avoiding accidents, maintaining control of their vehicles, and overall, enhancing road safety.

The European Automobile Manufacturers' Association (ACEA) describes that ADAS encompass a variety of safety technologies designed to aid vehicle drivers. These systems work towards making driving safer by enhancing road safety and reducing accident severity. They can assist drivers in accident avoidance or, when necessary, minimize the impact of accidents.

The Society of Automotive Engineers (SAE) defines ADAS systems in agreement with the ACEA in that ADAS encompass a wide array of electronic systems engineered to support drivers during their journeys. These systems can automate tasks, issue warnings or alerts, and enhance vehicle performance in various ways to improve safety and convenience.

Automotive OEM (Original Equipment Manufacturer) represent ADAS as integrated technologies within vehicles that leverage sensors, cameras, and radar to provide driver support. These technologies may include features like adaptive cruise control, lane-keeping assistance, and traffic sign recognition, all working together to enhance both driver safety and convenience.

2.2 Controller Area Network

Due to its durable and robust features, CAN (Controller Area Network) can be defined as a communication protocol and bus system that was originally created in the 1980s, primarily for the automotive industry. Microcontrollers and other devices can communicate in real time thanks to this two-wire, differential serial communication technology. With CAN High (CANH) and CAN Low (CANL) cables, differential signaling is used to lessen electromagnetic interference and improve noise resistance. It functions as a multi-master system, providing flexibility and distributed networking by enabling multiple devices on the bus to initiate communication. Due to its priority-based message arbitration mechanism, which gives higher-priority messages preference over lower-priority ones, CAN provides deterministic communication, ensuring predictable transmission and reception times that are essential for real-time applications.

Retransmission of messages is made possible by robust error detection and handling methods, which identify and report mistakes. The scalability of CAN enables both regular and extended message frames with 11-bit and 29-bit IDs, and can handle a range of communication rates, from kilobits to megabits per second. It is effective for little data packets thanks to its reduced transmission overhead. CAN has widespread acceptance in industrial automation, aircraft, medical devices, and other industries that require real-time, dependable communication despite its automobile industry roots.

3 METHODOLOGY

3.1 Research Methods and Data Collection

Telematics device play a pivotal role in vehicle performance monitoring and data collection. The driving behavior analysis is performed using Aplicom's T10G device which features important interfaces to vehicle and sensors that can be utilized for collecting data from in-vehicle sensors via a CAN bus interface. When assessing driving behavior, the CAN interface is very crucial. It offers immediate access to a multitude of data from the internal network of the car. Engine metrics, including RPM, fuel consumption, and engine load, can be monitored via the CAN interface. These parameters are essential for assessing driving effectiveness and vehicle health.

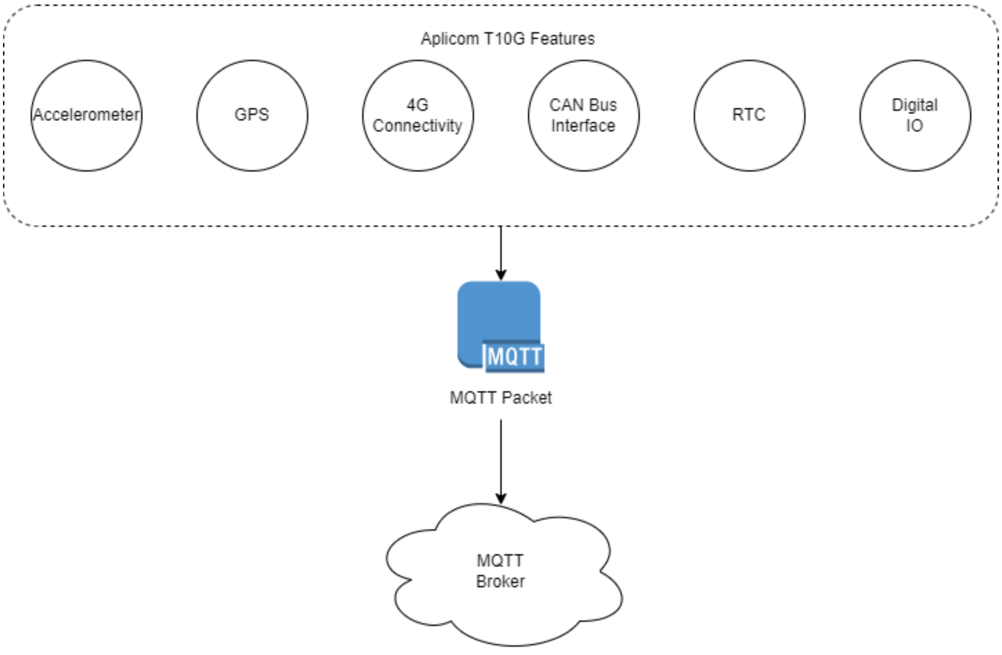

Figure 1: Data Collection

These devices have on board GNSS and GPS capability that provides accurate location of the vehicle and makes it possible to retrieve data from vehicle sensors at the current location with timestamps. With the help of 4G communication the data is sent from vehicle to the cloud service and helps in performing data analysis remotely.

When the vehicle is operating, the device gathers data and uses the MQTT (Message Queuing Telemetry Transport) protocol to communicate it to a cloud server. After collection, the data is processed and examined to address research concerns about driving habits and vehicle efficiency.

Chapter 4 presents a comprehensive study that sheds light on important aspects of driving behavior, vehicle performance, and road safety. The research questions posed are based on the vast variety of data that can be retrieved using the telematics device .

3.2 Data Processing

These devices being highly configurable via cloud service makes it easy to control the device remotely and program the device using the OTAP service. The data sent over the cloud contains unique device identifiers and is in the form of snapshots that contains GPS positioning, speed, acceleration metrics and data collected from vehicle's CAN interface. The device is also capable of handling events based on the provided configuration and performing certain actions in the response.

Table 1 below shows a few of the many factors and their sources and impact on driver's decision-making process and vehicle efficiency.

Sensors	**Collected Data**	**Corelation**
Accelerometer	Speed	Fuel Consumption, Speed limit
GPS	Position	Identifying terrain
Engine	RPM, Fuel Level	Speed, Diagnostics
Brake	ABS	Harsh braking

Table 1. Corelation of sensors data and vehicle efficiency

Figure 2 below highlights how the process of data collection goes for different drivers and its feeding to appropriate machine learning models which can help determine the patterns in their data and find out accurate correlations between distinct factors. This machine learning model will now be capable of taking input fresh driving data and comparing it with the patterns identified and helping in determining the ideal values of different driving parameters.

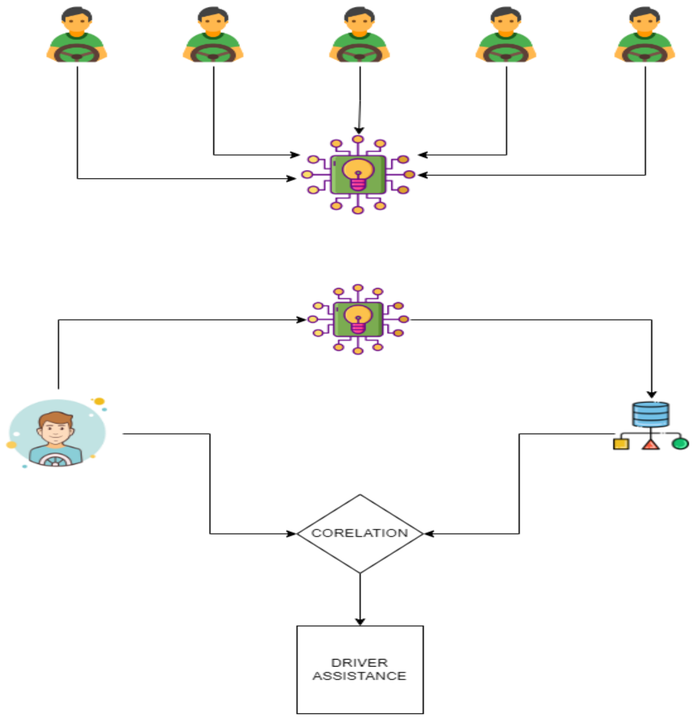

Figure 2. Data Processing

The T10G device contains a 3-axis accelerometer which helps in determining vehicle speed and whether the driver is abiding by the location's speed limit. In order to compare with the speed limitations imposed by the local Finnish authorities, the device is intended to gather real-time speed data along with GPS positions and timestamps. The Finnish Transport Infrastructure Authority offers a publicly accessible API that includes a thorough map of the country's street and road networks. Additionally, this database includes GPS-coordinated summer and winter speed limits.

Using the T10G telematics device, data has been gathered during multiple in-city and on-highway trips in order to present a thorough overview. There is travel data from highways and within cities in this collection.

A more thorough examination of the real-world implementation of this service, emphasizing crucial elements of safe driving practices, is provided in Chapter 5.

With the help of the accelerometer and braking data retrieved from the CAN interface, it becomes possible to detect harsh breaking events which not only tell about the psychological behavior of the driver but also helps in determining the driving score. With sudden changes in the accelerometer data, events such as harsh braking or veering off road can be identified and can help in alerting the driver in real-time using the performance evaluation matrix.

This analysis will also allow the determination of the change in speed in congested city areas, at intersections or at bus stops. Small changes of a few kilometers per hour during a short span of time can be negatively scored and indicate that the driving is not steady. Moreover, on highways there should be a maximum speed limit which should not be exceeded even momentarily and can be considered as highly dangerous and thus have a higher negative score. Maintaining a constant driving speed within the speed limit can be considered as efficient driving and point towards a positive score. However, such situations require careful consideration of the surrounding environment of the vehicle in order to avoid accidents and maintaining safe distance from surrounding vehicles.

These analyses are also helpful for companies that employ heavy vehicle drivers and can determine their vehicle performance based on driving style for timely predictive maintenance of the vehicles as well as optimizing the routes that the vehicle follows and result in lower fuel consumption and maintenance costs.

The study explores these situations' dynamics and how they relate to the parameters under investigation. Beyond that, an investigation into corner case situations and constraints is conducted to make sure the analysis stays thorough and considers the intricacies of the real world. Later chapters will go into further detail regarding the appropriateness of particular parameters in various scenarios. The goal is to present a comprehensive understanding of the driving behavior analysis process while taking into account its nuances and real-world differences. To identify the best method for evaluating particular parameters, the collected data and its suitability is explained using a variety of machine learning techniques.

4 RESEARCH QUESTIONS

Based on the extensive data gathered from various sources, including the on-device sensors, the internal sensors of the car via CAN bus interface, this study identifies three key research problems in the field of driving behavior analysis. The following research is based on these questions, which are derived from the rich tapestry of data. The research questions are addressed meticulously below, exploring the many facets of driving behavior, vehicle performance, and road safety. The study provides insightful information that has broad consequences for vehicle efficiency as well as safe driving behaviors.

4.1 How do current approaches, methods and algorithms automatically recognize, evaluate efficiency and forecast driver activity based on available online monitoring data?

The application of diverse machine learning models designed for specific circumstances is often necessary to comprehend driving behavior patterns. In this analysis, we investigate several machine learning models, each specially designed to clarify pertinent driving parameters. Decision trees are one of these models, and they work well for identifying distinct, rule-based patterns in driving behavior. When it comes to managing intricate relationships between parameters, random forests shine, offering a more nuanced perspective of driver behavior. Classifying driving patterns is a strong suit for Support Vector Machines (SVMs), but sequential dependencies in data, like those in time-series analysis, are best handled by recurrent neural networks (RNNs). Using this heterogeneous collection of machine learning models, we hope to fully understand driving behavior and reveal insights that may be hidden with a single method.

4.1.1 Decision Trees and Random Forests

Driving patterns can be categorized using decision trees and its ensemble equivalent, random forests. These models can identify particular driving behaviors, such as erratic braking, forceful acceleration, or prolonged idling, and then initiate feedback based on these categories.

- Decision Trees:

  i. Classification: Decision trees (Breiman et al, 1984; Quinlan 1986; 1993) are versatile tools for categorizing different driving patterns through iterative data splitting based on input features like speed, acceleration, and brake usage. This makes them well-suited for pinpointing specific behaviors in driving data.

  ii. Interpretability: Decision trees (DT) are inherently transparent, offering a structured hierarchical representation of decisions and conditions graphically (Nagalla et al., 2017). This interpretability aids in understanding the driving features contributing to a particular behavior, facilitating insights into, for instance, the characteristics of aggressive acceleration.

  iii. Feedback Generation: After training on historical driving data, decision trees can be employed in real-time to recognize a driver's actions. When a specific behavior is detected following the decision tree's logic, immediate feedback can be delivered, possibly in the form of visual or auditory alerts.

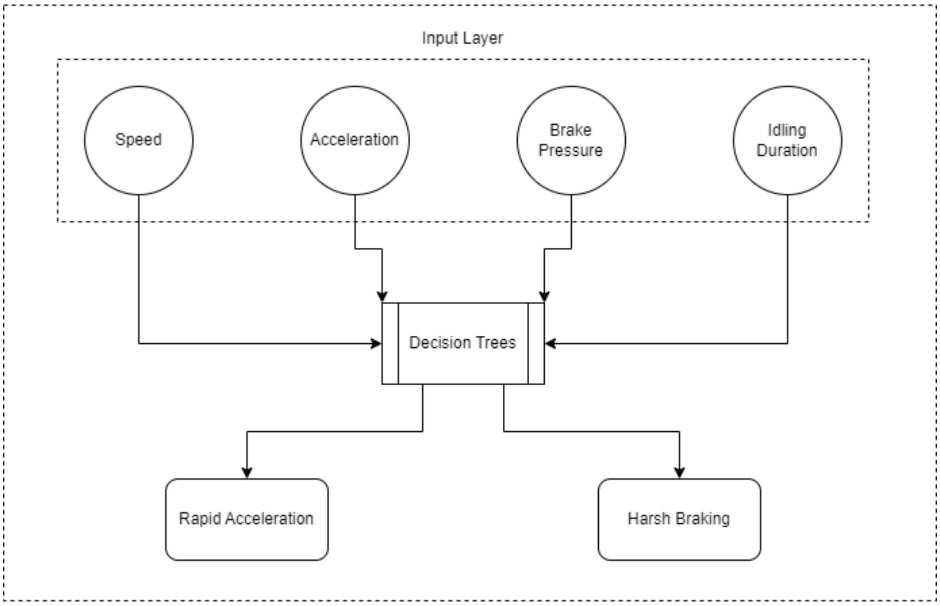

Figure 3: Decision Trees

Figure 2 elucidates the technical flow encompassing input features, the Decision Trees algorithm as the middleware processing unit, and the resultant output. The algorithm systematically processes a compiled dataset, employing training procedures to develop models. It conducts evaluations based on input parameters such as brake usage, idle duration, speed, and acceleration. The algorithm discriminates driving behavior by predicting outcomes, particularly emphasizing forceful acceleration and harsh braking. Notably, a high probability prediction of "Erratic Braking" implies a driver propensity for abrupt and frequent braking, while a forecast of "Forceful Acceleration" denotes rapid and forceful acceleration tendencies. The technical intricacies lie in the algorithm's ability to discern patterns within the input data, facilitating nuanced predictions of distinct driving behaviors.

- Random Forests (Ensemble of Decision Trees):

    i. Ensemble Approach: Random forests (RF), as ensemble learners, combine the predictions of multiple decision trees. These trees are constructed using distinct subsets of training data and features. This ensemble approach enhances the robustness and accuracy of pattern classification (Breiman, 2001).

ii. Handling Variability: Variability in driving patterns can pose challenges for a single decision tree model. Random forests excel in handling this variability as they capture different facets of the data through multiple trees. This adaptability enables them to account for various driver behaviors and environmental conditions.

iii. Feedback Customization: Random forests facilitate more nuanced and personalized feedback. Rather than providing generic responses for broad categories like "aggressive driving," the ensemble can differentiate between degrees of aggressiveness, offering tailored feedback based on the detected severity of the behavior.

iv. Reducing Overfitting: Random forests are less prone to overfitting compared to individual decision trees and often outperform them (Dietterich, 2000). Overfitting occurs when models learn noise in the data, leading to poor generalization. By combining the predictions of multiple trees, random forests mitigate overfitting concerns and offer more reliable results.

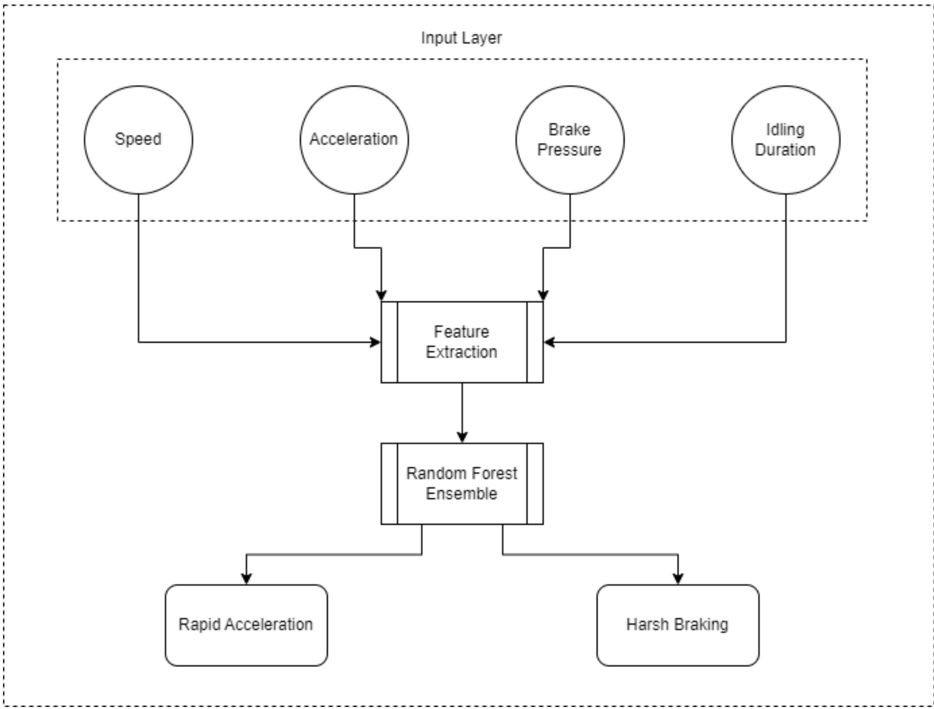

Figure 4: Random Forest

The diagram illustrates the integral components of the system: input factors, the middleware algorithm, and the targeted output. In the context of Random Forest Trees (RFT) algorithm, the input parameters, including brake usage, idle duration, speed, and acceleration, undergo a feature extraction process to create a comprehensive dataset. The RFT algorithm then leverages this dataset for model training, employing an ensemble of decision trees to evaluate and predict driving behavior.

The RFT algorithm's proficiency lies in its ability to discern complex patterns within the input data, employing a multitude of decision trees that collectively contribute to the final output. For instance, specific decision trees focus on rapid acceleration, while others scrutinize instances of harsh braking. Each decision tree encapsulates a set of rules tailored to its assigned aspect of driving behavior analysis.

The system's output, consequently, provides a detailed report elucidating specific driving behaviors, notably emphasizing forceful acceleration and harsh braking incidents. This granular feedback, rooted in the Random Forest algorithm's intricate analysis, serves as a technical tool for discerning and improving driving practices.

It not only identifies problematic driving patterns but also provides actionable insights for enhancing overall driving safety.

Nagalla et al. (2017) utilized and compared the results of data extracted using decision trees, random forests and support vector machines to predict drivers' gap acceptance behavior at intersections that lacked signals. The data included the crossing vehicle type, the conflicting vehicle, vehicle speed and size of the spatial gap. Results showed that SVMs are not sensitive to class imbalance, but the CART algorithm's generated decision tree yielded important information about the driver's decision-making process. Decision trees and random forests both established the relevance of several factors influencing the driver's choice. Moreover, validating the models using skill scores revealed that, while the random forest model beat the SVM and DT models, the SVM and DT models performed relatively similarly.

In summary, decision trees and random forests serve as valuable tools for the classification of driving behaviors and the generation of adaptive feedback. Decision trees offer transparency and interpretability, while random forests provide robustness and adaptability by aggregating insights from multiple trees. These models enhance driving efficiency by accurately identifying specific behaviors and delivering targeted feedback to encourage safer and more fuel-efficient driving practices.

4.1.2 Support Vector Machines (SVMs)

Support Vector Machines (SVMs) serve as effective tools for categorizing driving behaviors and recognizing deviations from ideal conduct, thus contributing to driver safety and offering relevant feedback. SVMs excel in binary classification tasks, distinguishing between optimal and non-optimal driving patterns based on learned decision boundaries in feature space.

- Categorizing Driving Patterns:

i. Ideal vs. Aberrant Behavior: SVMs are trained on labeled datasets encompassing examples of both exemplary and non-standard driving behaviors. These behaviors encompass adherence to safe speeds, smooth maneuvering, and compliance with traffic rules. The SVM's learning process enables it to differentiate between these categories, establishing a clear boundary between them.

ii. Speed Detection: A common application of SVMs in this context is the identification of speeding incidents. By analyzing data such as vehicle speed and road conditions, SVMs ascertain when a driver surpasses speed limits. This is instrumental in promoting road safety and adherence to traffic regulations.

- Generating Feedback and Alerts:

i. Real-time Surveillance: SVMs operate in real-time, continuously assessing incoming sensor data as the driver operates the vehicle. This enables swift detection of deviations from the expected behavior and research indicates they can be of great benefit in real-time crash risk evaluation (Yu & Abdel-Aty, 2013).

ii. Feedback Activation: Upon identifying deviations, SVMs can trigger feedback responses. Feedback may manifest through visual cues on the dashboard, auditory alerts, or tactile feedback via the steering wheel or pedals (Chen & Chen, 2017). For instance, if a driver consistently exceeds speed limits or executes sharp turns, the SVM may initiate visual warnings or issue audible alerts to foster safer driving habits.

iii. Tailored Feedback: SVMs can be fine-tuned to customize feedback based on the seriousness of the deviation. For instance, a minor speeding incident might lead to a subtle visual indicator, while a more significant violation could prompt a more immediate and pronounced warning.

- Elevating Driver Safety:

i. The primary objective of utilizing SVMs in this capacity is to heighten driver safety and encourage responsible driving conduct. By promptly identifying and addressing deviations from recommended behavior, SVM-based systems contribute to the reduction of accidents, the preservation of lives, and the prevention of traffic infractions (Wang et al., 2019).

ii. Furthermore, SVMs can be seamlessly integrated into comprehensive telematics and driver assistance systems, supplying valuable data for analysis and reporting on driving performance. This data can be harnessed for coaching and educational programs designed to enhance driver skills and behaviors.

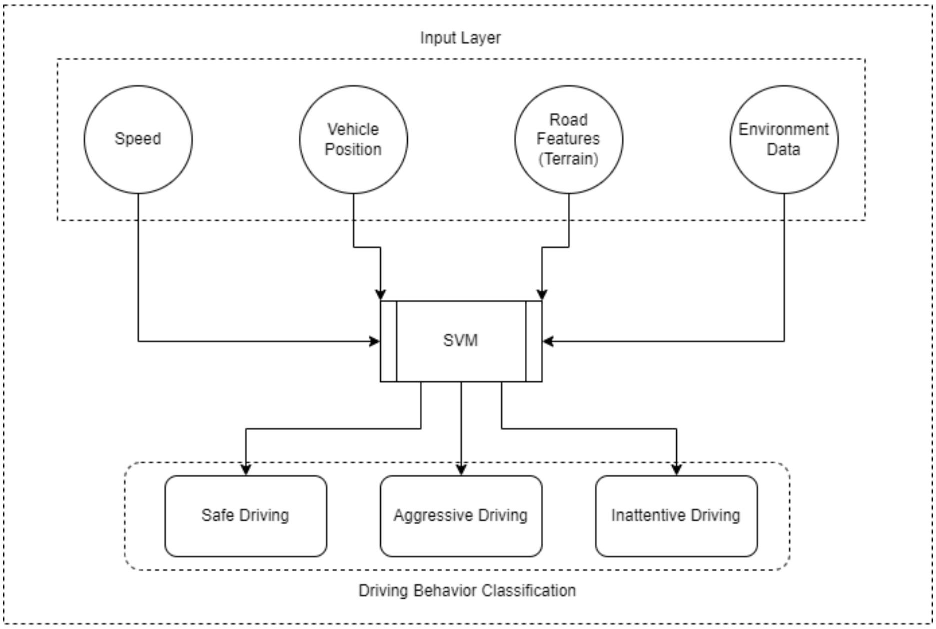

Figure 5: Support Vector Machines (SVM)

Figure 4 delineates the fundamental components of the driving behavior analysis system, incorporating input factors, the middleware algorithm, and the targeted output. In the realm of Support Vector Machines (SVMs) analysis, discerning the pertinent input factors is pivotal. These factors encapsulate road features, vehicle position, interactions with other vehicles, pedestrian activity, adherence to traffic signs and signals, speed dynamics, obstacle detection, turn signals, traffic flow, and environmental conditions. SVMs, renowned for their efficacy in classification tasks, assimilate these factors as input parameters. The algorithm's middleware functions involve intricate mathematical transformations and hyperplane delineations within the feature space, enabling the classification of driving behavior patterns. The anticipated outcomes, facilitated by SVM analysis, encompass categorizations such as safe driving, aggressive driving, and inattentive driving. SVMs excel in discerning complex relationships within high-dimensional datasets, rendering them adept at capturing nuanced patterns in driving behavior based on the amalgamated input features. This systematic analysis establishes a robust framework for real-time assessment and prediction of diverse driving behavior patterns.

Yu and Abdel-Aty (2013) investigated the SVM model's predictive capabilities for real-time crash risk evaluation. Results showed that the SVM model's classification accuracy can be enhanced by using a smaller sample size and that there is a need for a procedure for variable selection before the SVM model estimation.In summary, Support Vector Machines (SVMs) are valuable instruments for classifying driving patterns and detecting variations from ideal behavior. Their capacity to provide immediate feedback and activate alerts positions them as pivotal tools for enhancing driver safety, road traffic safety and endorsing prudent driving practices. Whether it involves identifying speeding incidents or pinpointing sharp turns, SVMs play an essential role in fostering safer road environments.

4.1.3 Deep Neural Networks

Deep neural networks, specifically CNNs and RNNs, offer robust capabilities for processing sequential driving data obtained from sensors and delivering adaptive feedback. These neural network architectures are well-suited to handle diverse data types, enhancing the comprehension of driving patterns and enabling context-aware responses.

- Convolutional Neural Networks (CNNs):

  i. Spatial Data Processing: CNNs excel in the processing of spatial data, making them invaluable for analyzing images or video streams from in-vehicle cameras or external sensors (Lecun et al., 1998). Through convolutional layers, CNNs automatically extract pertinent features from images, enabling them to identify patterns and objects.

  ii. Object Recognition: In the driving context, CNNs are adept at detecting objects like vehicles, pedestrians, traffic signs, and road markings. This capability is instrumental for assessing how the driver interacts with the surroundings and for identifying potential hazards or rule violations. Yan et al. (2016) utilized a CNN model to generate labels identifying driver's actions based on the movements of skin-like regions in driver images taken from the Southeast University Driving-posture Dataset with a 97.76% mean average precision, proving the effectiveness of the CNN model in recognizing driver actions.

iii. Visual Feedback: By analyzing real-time camera inputs, CNNs can offer visual feedback to drivers. For instance, they can detect lane deviations and issue alerts or provide guidance during parking maneuvers by recognizing obstacles and presenting visual cues on a display.

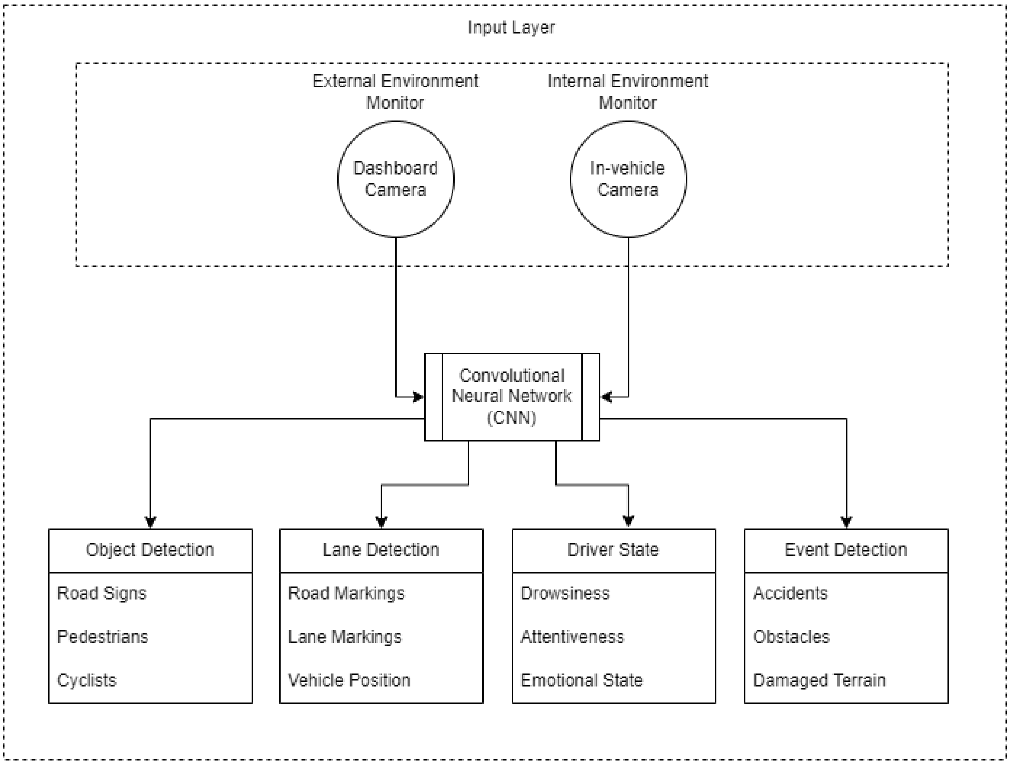

Figure 6: Concurrent Neural Network (CNN)

In the intricately designed system represented Figure 5, Convolutional Neural Networks (CNNs) serve as the central processing hub, extracting invaluable insights from both in-vehicle and external visual sensors. The in-vehicle camera data provides a comprehensive view of the driver's behavior, including emotional states and attentiveness, while external environmental imaging captures the surrounding conditions and potential obstacles. The CNN middleware, characterized by convolutional layers, engages in sophisticated feature extraction, unraveling hierarchical patterns within the input images. Through a meticulous training algorithm, the CNN learns intricate relationships from datasets, becoming adept at discerning object categories, identifying road markings for precise lane detection, and evaluating the driver's state, encompassing factors like drowsiness and emotional cues. Event detection capabilities further enable the CNN to pinpoint critical incidents such as accidents and damaged terrain. The resulting output encapsulates object recognition, lane detection, driver state assessment, and event detection, collectively contributing to a holistic understanding of the driving scenario. This intelligent middleware plays a pivotal role in enhancing safety through collision prediction, real-time assessments, and insightful recommendations, elevating the driving experience to new realms of efficiency and security.

- Recurrent Neural Networks (RNNs):

    i. Sequential Data Handling: RNNs are explicitly designed for working with sequential data, rendering them ideal for processing time-series sensor data commonly encountered in driving datasets. Such data may encompass readings from accelerometers, gyroscopes, GPS, and other sensors. Carvalho et al. (2017) determined that certain RNN configurations upon data collected using a smartphone accelerometer provide results with a high accuracy for classifying driving events.

    ii. Temporal Relationships: RNNs capture temporal relationships in data by maintaining hidden states that evolve as the data unfolds. This capability allows them to recognize patterns in driving behavior that develop sequentially, such as the rhythm of acceleration and braking.

- Integration and Contextual Awareness:

    i. Both CNNs and RNNs can be seamlessly integrated into a unified system to harness their complementary strengths. For instance, CNNs can process visual data from cameras to detect road conditions and objects, while RNNs can process time-series data to comprehend driver responses to visual cues. Virmani and Gite (2017) utilized both CNN and RNN models to analyze driver conduct along with a combined CNN with Long-Short Term Memory (LSTM) to give improved results in lesser response time.

    ii. Contextual awareness is a pivotal aspect of employing deep neural networks for driving feedback. The system can adapt its feedback based on the specific driving context, whether it involves highway driving, city traffic, or challenging weather conditions. This ensures that the feedback provided aligns with the driver's immediate circumstances and needs.

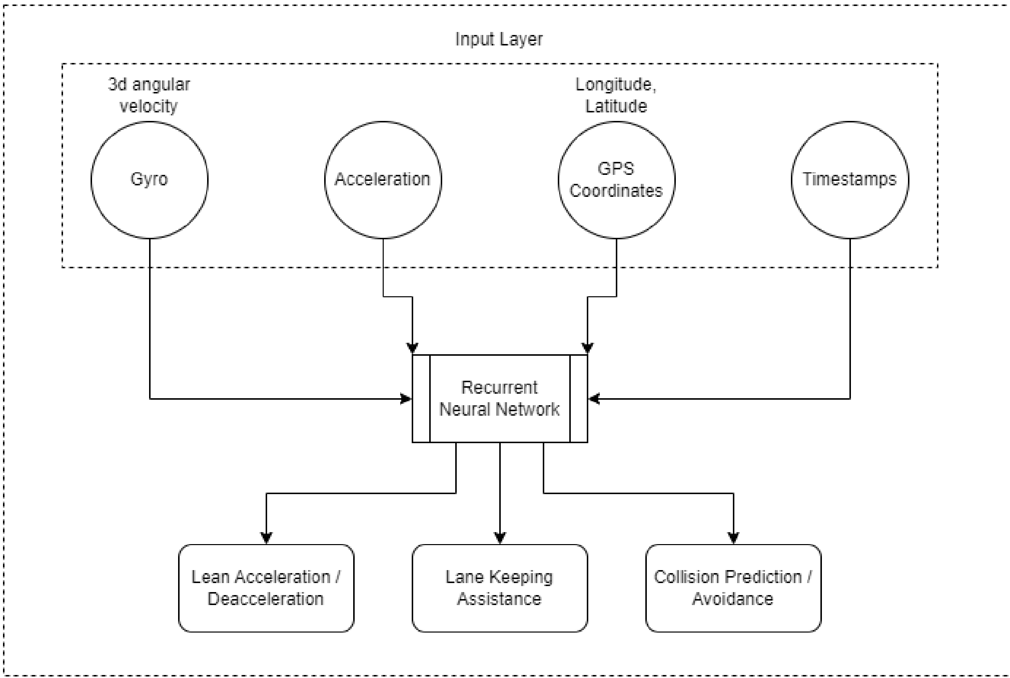

Figure 7: Recurrent Neural Network (RNN)

The schematic in Figure 6 delineates the integral components of the system, encompassing input factors, the intermediary algorithmic processes, and the targeted output. Within this framework, the Recurrent Neural Network (RNN) assumes a pivotal role in conducting a nuanced analysis of driving behavior patterns. The RNN algorithm meticulously processes the sequential dataset acquired from diverse sources, notably accelerometers, gyroscopes, and GPS. This sequential dataset is employed for model training, facilitating the algorithm to discern intricate patterns inherent in the driving environment.

The RNN's proficiency lies in its ability to recognize temporal dependencies and patterns within sequential data, making it particularly suited for analyzing driving behavior, which unfolds over time. The algorithm dynamically evaluates the predicted driving behavior based on the sequential input datasets, enabling a comprehensive understanding of the driver's interactions with the vehicle and the surrounding environment.

The input factors, derived from sensors such as accelerometers and gyroscopes, capture the nuanced dynamics of the vehicle's movement and orientation. Simultaneously, GPS data provides spatial context, enhancing the algorithm's capacity to comprehend the driver's position and navigation choices. The amalgamation of these factors forms a rich sequential dataset that serves as the foundation for the RNN's analytical processes.

In terms of output, the RNN algorithm yields predictive insights into driving behavior patterns. These outputs may encompass assessments of adherence to speed limits, smoothness of acceleration and deceleration, lane-keeping proficiency, and responses to dynamic elements like traffic and pedestrians. Moreover, the algorithm has the potential to suggest areas for improvement in driving behavior, thereby contributing to enhanced safety. By leveraging the capabilities of RNN-based analysis, this system empowers users to make informed decisions based on a sophisticated understanding of their driving habits and fosters a proactive approach to road safety.

In summary, deep neural networks, encompassing CNNs and RNNs, play a pivotal role in the analysis of driving data. While CNNs are adept at processing spatial data from various sensors, RNNs excel in handling sequential data from a myriad of sources. These networks empower the delivery of adaptive feedback by discerning objects, patterns, and temporal dependencies, ultimately contributing to safer and more efficient driving practices.

4.1.4 Clustering Algorithms

Clustering algorithms, such as k-means or hierarchical clustering, offer valuable capabilities in grouping similar driving patterns together. This clustering process helps identify shared behaviors among drivers and can facilitate the provision of feedback by drawing comparisons between an individual driver's behavior and those of others with similar driving profiles.

- Utilizing Clustering Algorithms for Driving Behavior:

i. Grouping Similar Patterns: Clustering algorithms work by partitioning a dataset into groups or clusters based on the similarity of data points. In the context of driving behavior, these algorithms have been utilized to examine various features and characteristics of how drivers operate their vehicles (Yang et al., 2022). Similar driving patterns, such as consistent speed, smooth acceleration, or adherence to traffic rules, tend to result in data points clustering together within the same group.

ii. Behavioral Insights: By categorizing drivers into clusters, these algorithms provide a means to gain insights into common behaviors. For instance, one cluster might represent cautious and law-abiding drivers, while another might include drivers who tend to speed or take risks. This clustering helps in understanding the prevailing driving tendencies within a dataset. Ping et al. (2019) utilized an unsupervised spectral clustering algorithm to determine the fuel efficiency of driving behavior based on naturalistic driving data. Results showed that this can effectively identify the connection between driving behavior and fuel consumption, allowing the latter's feature prediction which can be used practically in advanced driving assistance systems.

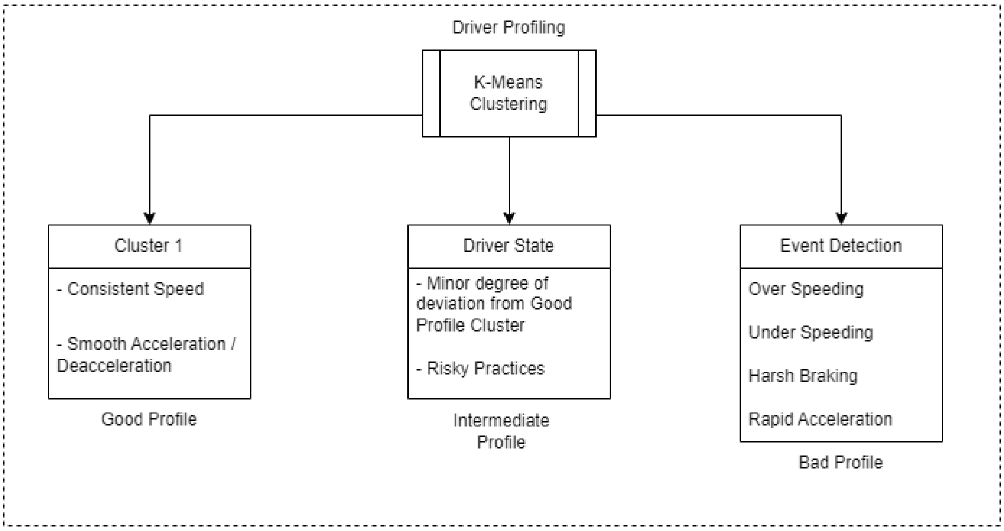

Figure 8: K-Means Clustering Algorithm

- Feedback and Comparison:

    i. Identifying Deviations: Once a driver is categorized into a specific cluster, their driving behavior can be compared to the norms and commonalities within that cluster. This comparison allows for the detection of deviations from typical behavior. For example, if a driver in a cluster known for cautious driving suddenly exhibits aggressive behavior, such as excessive speeding or abrupt lane changes, this can trigger feedback alerts.

    ii. Tailored Feedback: The feedback provided can be tailored based on the nature and degree of deviation. For instance, if a driver typically falls within a cluster of cautious drivers and temporarily deviates from this pattern, the feedback may consist of gentle reminders to return to safer driving practices. Conversely, if a driver consistently displays risky behavior compared to their cluster peers, the feedback may be more assertive and emphasize the need for improvement.

- Enhancing Driving Behavior and Safety:

i. The use of clustering algorithms in this manner contributes to enhancing driving behavior and overall safety. By placing drivers into clusters based on similarities in behavior, it becomes possible to not only identify deviations but also offer targeted feedback aimed at correcting or improving those behaviors (Ping et al., 2019).

ii. Over time, this feedback loop can lead to safer and more responsible driving practices. It can encourage drivers to align their behavior with the norms of their respective clusters, promoting adherence to speed limits, safe following distances, and courteous driving habits.

- Privacy Considerations:

  i. It's crucial to acknowledge the importance of privacy in this process. While clustering algorithms are effective for feedback and behavior improvement, it's imperative to handle data in compliance with privacy regulations, ensuring the anonymization of driver data and the protection of personal information throughout the analysis and feedback process (Nasr Azadani & Boukerche, 2022). Benyahya et al. (2022) analyzed the potential data privacy breaches and security considerations as well as their respective mitigation techniques and debated how best to limit them while maximizing the benefits of Automated City Shuttles (ACS) that would provide efficient and accessible transportation in smart cities.

In conclusion, clustering algorithms like k-means and hierarchical clustering play a crucial role in identifying common driving behaviors and enabling tailored feedback based on comparisons with similar driving profiles. This approach contributes to safer roads, encourages responsible driving practices, and can lead to improved overall driver behavior and safety.

4.1.5 Reinforcement Learning Algorithms for Adaptive Driving Feedback:

Reinforcement learning (RL) algorithms, including the advanced variant known as deep reinforcement learning (DRL), provide a real-time and adaptable method for delivering feedback based on a driver's actions. These algorithms have the capacity to learn and enhance driving behaviors through a process involving the reinforcement of certain actions while penalizing others.

- Distinctive Features of Reinforcement Learning:

    i. Trial-and-Error Learning: Reinforcement learning algorithms function on the principle of learning through experimentation. They acquire knowledge by actively engaging with the driving environment, taking actions, and receiving feedback in the form of rewards or penalties.

    ii. Agent and Environment: In the context of driving, the driver and the vehicle constitute the agent, while external factors such as road conditions, traffic, and environmental elements form the environment (Hu et al., 2019). The agent makes driving decisions within this environment, seeking to optimize cumulative rewards over time.

- Real-time Feedback and Adaptation:

    i. Immediate Feedback: Reinforcement learning excels in delivering real-time feedback. As the driver takes actions, the RL system continually assesses these actions and provides instantaneous feedback based on their consequences.

ii. Adaptive Behavior: RL algorithms excel at adapting driving behaviors by optimizing expected cumulative rewards. Actions that lead to positive outcomes and safer driving practices are encouraged with rewards, while actions that pose risks or result in suboptimal outcomes are discouraged with penalties. Shan et al. (2020) utilized this approach to adaptably balance between the path tracking accuracy and the passenger experience in fully autonomous vehicles.

- Reward and Penalty Mechanisms:

i. Reward Signals: In the driving context, reward signals can be designed to promote desired behaviors (Pandey et al., 2010). For instance, maintaining safe following distances, adhering to speed limits, and executing smooth lane changes can be positively reinforced. Rewards can also be utilized to encourage eco-friendly driving, such as minimizing fuel consumption.

ii. Penalties: Conversely, RL algorithms have the capability to impose penalties for unsafe or inefficient actions. Aggressive driving behaviors like sudden braking, rapid acceleration, or tailgating can incur negative rewards. These penalties serve as deterrents against risky actions and promote safer and more efficient driving practices.

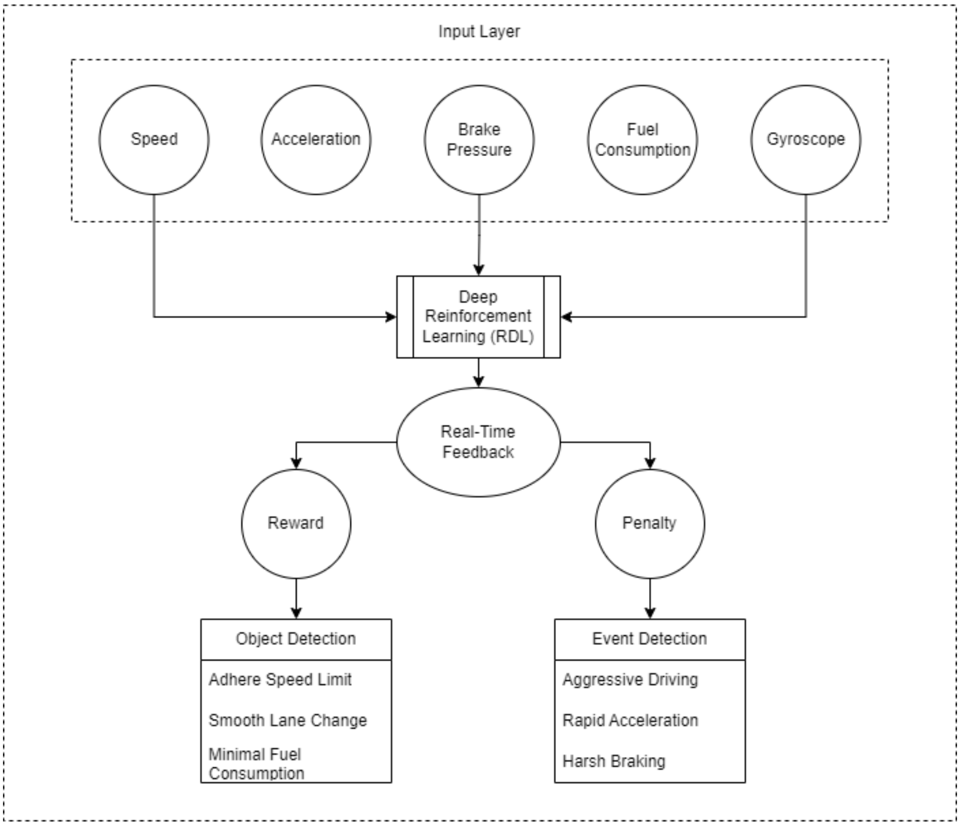

Figure 9: Re-inforcement Learning

The schematic illustration in Figure 8 delineates the intricate interplay between input factors, the middleware algorithm's operational facets, and the coveted output in the realm of driving behavior pattern analysis. In this technical framework, the reinforcement learning algorithm assumes a pivotal role. It meticulously processes the amassed dataset, embarks on the training of intricate models, and subsequently assesses the predicted driving behavior. The algorithm derives insights from a multifaceted input array, encompassing parameters such as speed, acceleration, fuel consumption, brake utilization, and gyroscopic datasets. The culminating output manifests predominantly as a comprehensive driver score, a nuanced amalgamation of rewards and penalties. This score serves as a comprehensive indicator, furnishing valuable feedback on driving proficiency. By delineating areas for potential enhancement, the algorithm contributes to the cultivation of safer and more adept driving practices.

- Learning and Optimization:

i. Deep Reinforcement Learning: Deep reinforcement learning (DRL), which integrates RL with deep neural networks, is especially effective for intricate tasks like driving. DRL models can acquire intricate driving strategies and behaviors by processing extensive sensor data, including information from cameras and other sources (Kiran et al., 2022).

ii. Continuous Enhancement: RL systems consistently refine driving behavior through an iterative learning process. As experience accumulates, the algorithm becomes progressively skilled at making decisions that prioritize safety, efficiency, and adherence to traffic rules.

- Safety and Adaptability:

i. RL algorithms prioritize safety by minimizing actions that result in adverse outcomes. Marchesini et al. (2022) adapted this principle to propose a safety-oriented search complementing deep RL algorithms such that the policy is biased toward safety in an evolutionary penalty optimization.

RL algorithms prioritize safety by minimizing actions that result in adverse outcomes (Marchesini et al., 2022). They can dynamically adjust driving behavior to respond to evolving road conditions, traffic scenarios, and unexpected incidents.

4.1.6 Hidden Markov Models (HMMs)

Hidden Markov Models (HMMs) are a suitable choice for modeling sequences of data, making them valuable for the analysis of driving behavior patterns over time. HMMs are especially adept at capturing the shifts or transitions between various driving states and can offer feedback based on these transitions.

- Modeling Sequential Data with HMMs:

  i. Sequential Nature: Driving behavior unfolds sequentially, involving a sequence of actions and events as time progresses. HMMs are specifically designed to model such sequential data, where each state within the model represents a particular aspect or condition of the driving behavior.

  ii. State Transitions: HMMs excel at capturing the transitions that occur between states. In the context of driving, these states can represent different driving behaviors or conditions, like safe driving, aggressive maneuvers, abrupt braking, or lane changes (Li et al., 2016). The transitions between these states signify shifts in driving behavior.

- Analyzing Driving Patterns:

  i. Recognizing Driving States: HMMs can be trained on historical driving data to identify and characterize distinct driving states. These states might encompass safe driving practices, risky behaviors, or specific driving situations such as highway driving or urban traffic.

  ii. Transition Analysis: Through the analysis of transitions between these states, HMMs reveal patterns in a driver's conduct. For example, they can detect when a driver transitions from a state of safe driving to one marked by risky behaviors, such as sudden acceleration (Li et al., 2016). This transition analysis provides insights into driving habits.

- Providing Feedback Based on Transitions:

  i. Feedback Mechanisms: HMMs can activate feedback mechanisms in response to identified transitions. For instance, if a driver frequently shifts from a safe driving state to a risky behavior state, the system can deliver real-time feedback. This feedback could take the form of visual or auditory alerts, encouraging the driver to adopt safer driving habits.

ii. Tailored Guidance: HMMs also enable tailored guidance. When transition analysis uncovers that a driver often switches to risky behavior during specific road conditions or traffic scenarios, the feedback can be personalized to address these particular situations, offering targeted recommendations on how to enhance driving performance.

- Advancing Driving Behavior:

i. Systems based on HMMs contribute to the improvement of driving behavior by raising awareness of transitions and their significance. By offering feedback regarding these transitions, drivers gain insights to make informed decisions that promote safer and more responsible driving practices.

ii. Furthermore, HMMs play a role in creating driver profiles and uncovering behavioral trends. This information proves valuable for long-term driver enhancement initiatives and for evaluating driving practices across various conditions, assisting individuals in becoming more conscientious and skilled drivers.

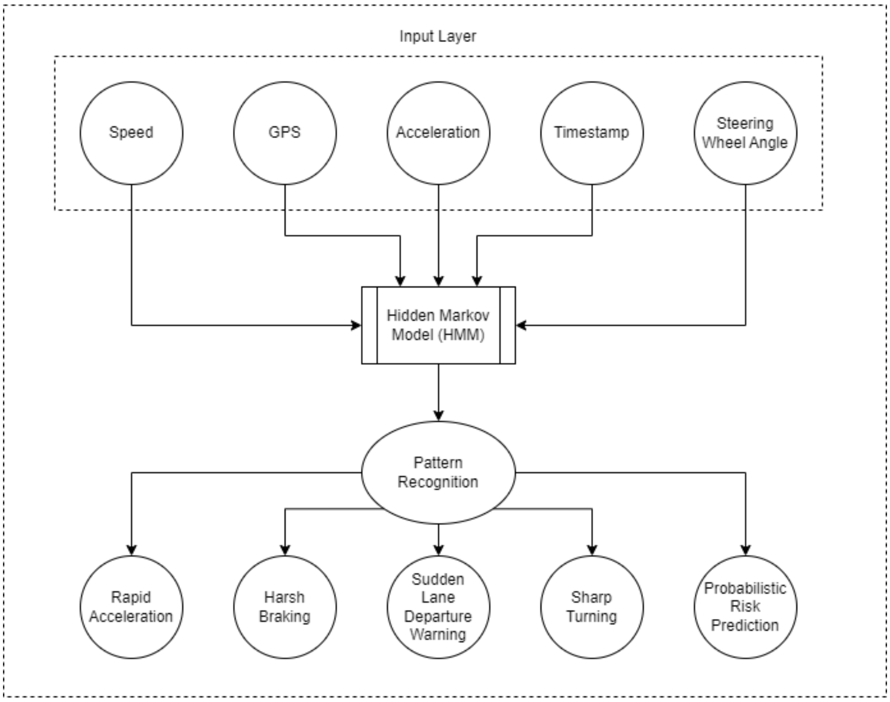

Figure 10: Hidden Markov Models (HMM)

In the realm of Hidden Markov Models (HMMs) for driving behavior analysis, the diagram in Figure 9 showcases a symbiotic relationship among input factors, the middleware algorithm, and outputs. The reinforcement learning paradigm within the algorithm processes data, training HMM models with features like speed, acceleration, GPS, timestamp, and steering angle.

As a probabilistic graphical model, the HMM adeptly captures latent states underlying driving behaviors. Transition probabilities represent temporal dependencies, while emission probabilities link states to observable features. Outputs include risk predictions and insights into specific driving habits, such as rapid accelerations, harsh braking, lane shifts, and sharp turns.

The HMM's ability to model temporal dependencies enables it to grasp sequential patterns, offering a holistic view of driving behavior. This technical middleware enhances safety insights, indicating areas for improvement. In summary, the HMM serves as a powerful tool for probabilistic modeling, decoding intricate patterns from diverse inputs and contributing to intelligent transportation systems and safety applications.

Li et al. (2016) utilized this model to measure the stability and the risk of driver behavior at intersections, discovering that in the dilemma zone driver behavior is of lower stability and higher risk as compared to that in other areas around intersections. In summary, Hidden Markov Models (HMMs) are a suitable choice for modeling and comprehending sequential driving data. Their strength lies in capturing transitions between different driving states and offering valuable insights that enhance our understanding of driving patterns (Zhang et al., 2014). By triggering feedback mechanisms tied to these transitions, HMMs contribute to the cultivation of safer and more responsible driving practices, ultimately fostering improvements in driver behavior over time.

4.1.7 Gradient Boosting Algorithm

Gradient boosting algorithms like XGBoost and LightGBM are versatile tools applicable to both classification and regression tasks related to the analysis of driving patterns. Renowned for their impressive predictive accuracy, these algorithms play a pivotal role in furnishing feedback based on anticipated results within the realm of driving behavior analysis.

- Employing Gradient Boosting for the Analysis of Driving Patterns:

    i. Dual Capabilities in Classification and Regression: Gradient boosting algorithms, notably XGBoost and LightGBM, demonstrate their versatility by accommodating a wide array of driving pattern analysis tasks. These encompass classification undertakings such as the identification of hazardous driving behaviors and regression assignments like the prediction of factors such as fuel efficiency or driver performance metrics (Mousa et al., 2019).

ii. Emphasis on Predictive Precision: An inherent strength of gradient boosting algorithms lies in their remarkable predictive accuracy (Mousa et al., 2019). They are adept at capturing intricate relationships within driving data, enabling highly accurate forecasts and classifications. This precision is especially valuable when it comes to scrutinizing and comprehending driving behavior.

- Feedback Informed by Predicted Outcomes:

i. Tailored Guidance: XGBoost and LightGBM are equipped to offer feedback to drivers based on their anticipated driving outcomes. For instance, if the algorithm anticipates an elevated probability of aggressive driving behavior, the system can issue real-time feedback or alerts to promote safer and more responsible driving practices.

ii. Personalized Recommendations: These algorithms can generate individualized recommendations aimed at enhancing specific facets of driving behavior. If, for instance, the prediction suggests excessive fuel consumption, the system can provide suggestions for adopting eco-friendly driving techniques to optimize fuel efficiency.

- Exploring Feature Importance:

i. Comprehending Driving Patterns: XGBoost and LightGBM extend feature importance analysis, facilitating a deeper comprehension of which factors or characteristics exert the most significant influence on driving patterns. This insight proves invaluable in identifying the primary determinants of particular behaviors and tailoring feedback accordingly.

ii. Detecting Risk Factors: In the context of driver safety, these algorithms have the capacity to spotlight critical risk factors contributing to unsafe driving practices (Mousa et al., 2019). By pinpointing these factors, the system can offer focused feedback and interventions to mitigate risks.

- Real-time and Continuous Feedback:

    i. Swift Response: The real-time capabilities inherent to gradient boosting algorithms empower immediate responses to driving behaviors. Should a driver exhibit behavior aligned with unsafe or inefficient driving patterns, the system can promptly dispense feedback or alerts, facilitating prompt corrective actions.

    ii. Persistent Enhancement: By delivering feedback rooted in predicted outcomes, these algorithms play a pivotal role in perpetually refining driving behavior. Over time, drivers can cultivate safer, more efficient habits as they receive guidance and insights shaped by the algorithm's prognostications.

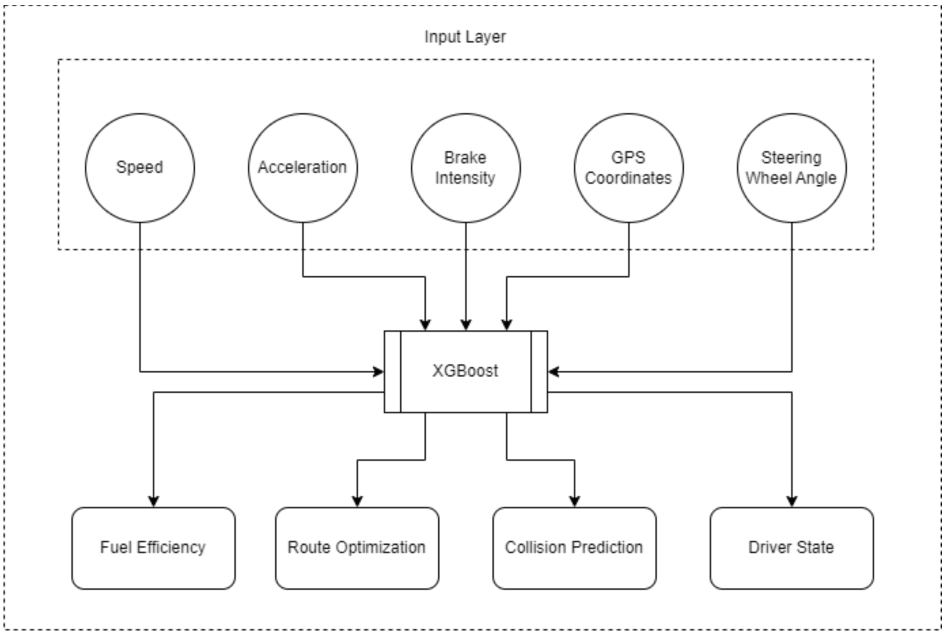

Figure 11: Gradient Boosting Algorithm

As shown in Figure 10, in applying Gradient Boosting algorithms to driving behavior analysis, key input factors include speed, acceleration, GPS coordinates, brake intensity, and steering angle. These factors are processed by the middleware algorithm, which orchestrates an ensemble of decision trees, iteratively refining the model to discern driving patterns. The output encompasses practical insights such as fuel efficiency estimates, route optimization suggestions, collision predictions, and driver statistics. By analyzing speed dynamics, acceleration patterns, and steering behavior, the algorithm transforms complex input data into actionable outcomes, promoting safer driving practices.

Mousa et al. (2019) developed an XGB classifier to identify factors contributing to crash and near-crash (CNC) events and evaluated its performance against three other machine learning algorithms. Their results showed that the XGB model outperformed the other three algorithm models, giving an 85% detection accuracy and identifying intersection influence and driver behavior as the biggest contributors to detecting CNCs.

In summary, gradient boosting algorithms like XGBoost and LightGBM are versatile assets for both classification and regression tasks pertaining to driving patterns. Their exceptional predictive accuracy enables highly precise feedback and recommendations grounded in projected results, fostering safer, more conscientious, and more efficient driving practices. Moreover, their feature importance analysis and real-time feedback capabilities render them indispensable components for the advancement of driving behavior and safety.

4.1.8 Auto Encoders

Autoencoders represent a type of neural network architecture employed for tasks such as reducing the dimensionality of data and learning essential features. When applied to the analysis of driving behavior, autoencoders become valuable assets. They possess the capability to extract pertinent features from intricate driving datasets and play a pivotal role in identifying distinctive patterns that influence the efficiency or inefficiency of driving (Xie et al., 2018).

- Autoencoders for Feature Extraction:

  i. Dimension Reduction: Autoencoders are structured to compress complex, high-dimensional driving data into a more simplified representation, often referred to as the latent space or encoding. This process simplifies the data while preserving its critical attributes. In the context of analyzing driving behavior, this dimensionality reduction simplifies the intricate data, making it easier to recognize significant patterns.

  ii. Feature Discovery: Autoencoders can learn and uncover relevant features from the input data. As they adapt and train on the driving dataset, they acquire the ability to emphasize distinct attributes and characteristics that exert a substantial influence on driving efficiency (Xie et al., 2018). These acquired features may encompass elements like speed variations, patterns of acceleration, lane-keeping behaviors, or responses to traffic signals.

  iii. Customization: Autoencoders have also been trained to ignore the impact of features considered irrelevant in various scenarios, such as trees (Wang et al., 2022). This indicates the wide range of applications of autoencoders.

- Discerning Driving Patterns:

  i. Efficiency vs. Inefficiency: Through the utilization of autoencoders, it becomes feasible to differentiate between driving patterns that contribute to efficiency and those that lead to inefficiency. For example, the neural network can recognize that maintaining a consistent speed within legal limits positively correlates with fuel efficiency, whereas frequent rapid acceleration and deceleration negatively impact it.

ii. Contextual Awareness: Autoencoders also consider the context in which driving patterns occur. They can discern, for instance, that aggressive acceleration might be appropriate on highways but not in congested urban traffic, underscoring the importance of context-aware driving behavior analysis.

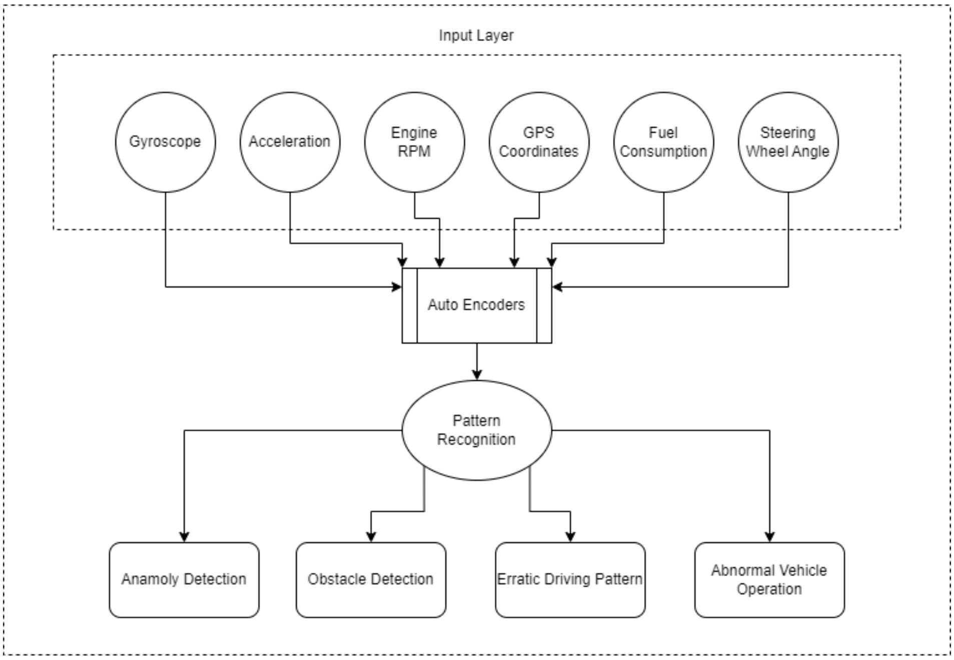

Figure 12: Auto Encoders

As shown in Figure 11, in the context of Auto Encoder algorithms for driving behavior analysis, pivotal input factors such as Engine speed, acceleration, GPS, Gyroscope, fuel consumption, and steering angle collectively contribute to the middleware's neural network. This Auto Encoder middleware excels at compressing and reconstructing these inputs, creating a latent space representation that captures underlying patterns. The reinforcement algorithm processes the dataset, facilitating model training and evaluation.

The Auto Encoder's proficiency lies in results like Anomaly detection, Obstacle detection, recognition of erratic driving patterns, and identification of abnormal vehicle operation. Anomalies are deviations from learned patterns, detected when the reconstructed output significantly differs from the actual input. This nuanced approach contributes to safer driving by assessing behavior and suggesting areas for improvement. In summary, the Auto Encoder's adeptness at learning intricate patterns positions it as a crucial middleware in driving behavior pattern analysis.

- Enhancing Driving Efficiency:

    i. Feedback and Optimization: Once pertinent features and patterns are identified, autoencoders can play a pivotal role in offering feedback and guidance to drivers. For instance, if the analysis reveals that smoother acceleration and braking patterns result in improved fuel efficiency, the system can provide real-time feedback to encourage such driving behaviors.

    ii. Personalization: Autoencoders permit the tailoring of feedback to align with individual driving habits and preferences. Drivers can receive personalized recommendations designed to suit their unique driving scenarios and requirements, further enhancing their driving efficiency.

In summary, autoencoders, as a neural network architecture, excel in dimensionality reduction and feature learning. Their aptitude for extracting essential features from driving data and revealing influential patterns has a significant impact on encouraging safer, more efficient, and context-aware driving behaviors. Autoencoders empower the creation of intelligent systems capable of delivering real-time feedback and recommendations, ultimately promoting improved driving practices and sustainability on the road.

4.2 Can real-time feedback and coaching improve driving efficiency and what is the most effective method for delivering feedback/assistance?

Onboard vehicle systems that use historical data can provide real-time coaching and feedback. There are several studies showing how real-time feedback and coaching can directly impact driving efficiency. Drivers can receive quick feedback on their driving style and fuel efficiency via sophisticated telematics systems. Additionally, drivers may adopt more fuel-efficient habits on the road as a result of this feedback. Wickramanayake and Bandara (2016) proposed a method of enhancing the fuel economy of fleet vehicles using a real-time driver behavior monitoring and feedback system with feedback being provided through mobile devices based on individual driver profiles and other real-time data including route information, traffic and weather data. They estimated that up to 20% further fuel savings can be achieved by drivers adopting more efficient driving behavior. Additionally, fleet managers can use previous data to build driver improvement and behavior-targeting training programs.

The effectiveness of the distribution mechanism chosen will depend on the real-time feedback and coaching. The feedback and coaching can directly contribute to the change in driving patterns that might affect vehicle efficiency and environmental factors, in order to maximize driving efficiency. Driving style changes including erratic acceleration, abrupt braking, excessive speeding, and extended idling can significantly reduce fuel efficiency and increase emissions, which have an impact on the environment. Real-time feedback systems are used to address these problems, giving drivers direction and data in real-time so they can change their driving practices in favor of more environmentally-friendly ones. Rolim et al. (2017) assessed how real-time feedback impacted the driving behavior of bus drivers and the variables influencing fuel consumption. Their research results provided valuable insights for bus companies to improve their operational strategies and update their training programs.

Several methodologies exist to deliver real-time feedback which include visual displays on the dashboard or heads-up display, audio alerts, haptic feedback through resistance or vibrations in the steering wheel, smartphone applications, and linked vehicle systems. To accommodate variations in driving patterns, machine learning algorithms can be employed to provide adaptive feedback. These algorithms can analyze the historical driving data and the real-time sensors inputs to tailor the feedback and coaching based on the individual driving behaviors and contextual factors (Rolim et al., 2017).

Real-time feedback aims to reduce not only fuel consumption but also greenhouse gas emissions and air pollution in order to protect the environment. It makes recommendations for actions that are customized to the particular circumstances and variances in driving styles, such as slower acceleration on uphill routes or prudent use of air conditioning to save energy.

Real-time feedback systems' efficacy also depends on user acceptance and involvement. Picco et al. (2023) studied the responses of 628 Dutch drivers on the potential use of a system monitoring and providing feedback on their driving. Results indicated that most drivers were neutral to positive about the prospect of receiving feedback after monitoring their driving and that the driver acceptability could be increased if personalization of the monitoring and feedback device was allowed. Increased driver receptivity to coaching is brought on by clear, unobtrusive feedback that produces measurable results, such as better fuel economy and less environmental impact as well as the option to personalize the system. Therefore, the most successful strategy incorporates a range of feedback modalities, adaptive algorithms, an emphasis on user interaction and system personalization, all geared at encouraging environmentally friendly driving practices that benefit both drivers and the environment.

4.3 How can past driving data be used to develop more accurate and effective models to reduce costs and improve fuel/driving (which one is better here?) efficiency?

Past driving data serves multiple purposes including improving driving efficiency, providing a proactive approach to maintenance prediction, enhancing fuel efficiency and thus reducing costs, and even optimizing traffic and routing. This can ultimately lead to the development of better and more accurate models to reduce costs and improve efficiency.

4.3.1 Proactive Maintenance Prediction:

Manufacturers of automobiles and fleet managers can benefit significantly from this wonderful opportunity to approach maintenance with a forward-thinking mindset by analyzing past driving data (Lee et al., 2020). This entails using predictive modeling and data analysis to predict component failures or maintenance requirements far more accurately than with typical preventative care techniques.

- Enhanced Predictive Accuracy:
  A vehicle's performance and behavior over time can be learnt from historical driving data. Numerous factors are included in this data, including as wear and tear patterns, engine performance measures, and a thorough maintenance history. Through the use of sophisticated analytics and machine learning algorithms to this plethora of data, organizations can acquire a profound comprehension of the degradation and failure of different components over time and can identify possible problems well in advance of serious failures (Chen et al., 2019).

- Optimized Maintenance Scheduling:
  Optimizing maintenance schedule is one of the main advantages of proactive maintenance prediction. Organizations may precisely schedule maintenance tasks by using past data-driven insights, as opposed to waiting for an unex-

pected component failure (Chen et al., 2019). In order to minimize needless early replacements and to prevent failures and disruptions, this entails determining the optimal time to maintain or replace a component. As a result, maintenance schedules can be adjusted to better suit the vehicle's real condition, increasing both their effectiveness and affordability. Wolf et al. (2023) provide a standardized description model to help efficiently introduce predictive maintenance models in the automotive industry.

- Mitigating Financial Burden:
  Forecasting maintenance needs in advance can significantly reduce the financial strain brought on by unexpected malfunctions and expensive repairs (Chen et al., 2021). Organizations that experience unexpected component failure not only have to pay for the cost of replacing or repairing the damaged parts, but they also have to pay for other expenses including vehicle downtime, lost productivity, and possible damages or injuries. Organizations can save these monetary losses by using previous data to anticipate and prevent failures. Since the organizations cover both the direct costs of repairs and the indirect costs of interruptions to operations, the cost reductions are significant.

- Increased Operational Efficiency:
  Operational efficiency declines when unanticipated breakdowns occur in vehicles. Reduced productivity results from idle cars and thrown off work schedules. Proactive maintenance prediction, on the other hand, reduces the likelihood of unplanned downtime by ensuring that vehicles stay in optimal functioning order (Mesgarpour et al., 2013). As a result, vehicles are able to continue operating and fulfilling their intended functions without interruption, which increases operational efficiency.

- Prolonged Vehicle Lifespan:
  Historical driving data helps extend the life of automobiles by proactively addressing maintenance needs. Components are properly maintained and replaced

when necessary, minimizing the wear and tear brought on by usage over time (Mesgarpour et al., 2013). As a result, vehicles can continue to perform effectively and efficiently for a longer period of time, giving businesses a great return on their investment and lowering the need for vehicle replacements.

In a nutshell, proactive maintenance prediction is a data-driven method that makes use of past driving information to foresee and avoid component failures or maintenance requirements. Through proactive measures, maintenance schedules are improved, financial burdens are reduced, operational efficiency is increased, and vehicle lifespans are extended. Organizations can save time and money while assuring the dependability and lifespan of their fleets of vehicles by switching from reactive to preventative care methods.

4.3.2 Fuel Efficiency Enhancement:

In the effort to maximize fuel efficiency in vehicles, historical driving data is crucial. This information is used as a starting point for developing sophisticated algorithms and models that aim to steadily lower fuel use. Organizations can find valuable insights that improve fuel efficiency by carefully examining past driving patterns, speed changes, acceleration and deceleration rates, and a variety of other pertinent characteristics. Here is a closer look at the advantages and techniques used:

- In-Depth Analysis of Driving Patterns:
  Past driving statistics reveals an extensive number of details about how automobiles are driven over time. This includes information on variables including the frequency of violent acceleration, quick deceleration, and idle times. Organizations can acquire a thorough understanding of driving habits and patterns that affect fuel efficiency by carefully studying this data.

- Identifying Fuel-Consuming Habits:

Analyzing historical data has several benefits, one of which is the capacity to pinpoint fuel-consumption patterns. For instance, data analysis may show that frequent braking and strong acceleration are closely related to higher fuel use. With this understanding, motorists can intentionally modify their behavior to develop more relaxed driving practices. This entails accelerating gradually, anticipating stops, and keeping constant speeds. These behavioral adjustments can result in immediate fuel savings. Recent research has utilized a framework for predicting energy consumption in vehicles using a machine learning-based framework on driving data compiled from fifty-five electric taxis in Beijing (Zhang et al., 2020).

- Route Optimization:
  The optimization of a route also heavily relies on historical driving data. For example, fleet operators can use historical data to pinpoint the routes that will use their cars' fuel most efficiently. This entails looking at previous travel routes, traffic patterns, and the effect of different route selections on fuel use. Organizations can cut fuel usage and total transportation expenses by selecting the best routes based on historical data (Bozorgi et al., 2017). Liu et al. (2023) constructed energy consumption estimation models for 33 vehicle types based on large quantities of historical driving data which were utilized in eco-routing optimization simulation experiments under various traffic conditions and several road networks. Results indicated up to 11.50% energy-saving improvement as compared to the conventional path planning approach.

- Cost Savings and Environmental Benefits:
  Historic driving statistics provides data-driven insights that can result in real cost savings. Reducing fuel expenses immediately affects an organization's bottom line. Additionally, environmental footprint can be reduced by optimizing fuel usage, which can create a win-win situation for both cost-conscious businesses and eco-conscious people by lower greenhouse gas emissions.

- Continuous Improvement:
  Organizations can continuously improve their strategies over time by conducting data analysis on a regular basis and tracking the results of fuel efficiency measures. The continuous fuel savings and further operational optimization are made possible by this iterative procedure (Linda & Manic, 2012).

In conclusion, improving fuel efficiency through analysis of past driving data entails a variety of steps, including detailed examination of driving patterns, identification of fuel-consuming habits, route optimization, real-time feedback, and coaching. Organizations can significantly cut costs, reduce their environmental effect, and constantly improve their fuel economy tactics by utilizing historical data. Linda and Manic (2012) utilized historical vehicle performance data combined with GPS information on fixed routes to model the most fuel-efficient driving behavior. By comparing current vehicle state with the optimum state from the model, an optimum control action is generated which provides the most fuel-efficient cruise control option. Data-driven strategies such as this help the bottom line while also helping to make transportation more environmentally friendly and sustainable.

4.3.3 Optimized Traffic and Routing:

In order to create extremely effective navigation systems, it is essential to make use of historical traffic and route data. Through the use of insights from vast historical traffic patterns and route choices, these systems are made to make travel more efficient (Bozorgi et al., 2017). For fleet operators as well as individual drivers, the integration of historical data into navigation and routing systems yields numerous significant benefits.

- Intelligent Traffic Management:
  An in-depth study of traffic patterns and congestion points that have evolved over time is possible with the use of historical traffic data. Through the analysis of this abundant historical data, navigation systems are able to make informed

choices about route planning and optimize travel routes (Bozorgi et al., 2017). This intelligence makes travel more predictable and efficient by reducing delays brought on by traffic congestion.

- Time Savings:

Optimized traffic and routing systems significantly save time for both individual drivers and fleet operators. By avoiding traffic congestion and selecting the most efficient routes, travel times are reduced (Bozorgi et al., 2017). This is particularly crucial for fleet operators, where time saved directly impacts the efficiency of deliveries, service appointments, and overall operational productivity.

- Fuel Conservation:

Efficient routing based on historical traffic data also has an impact on fuel conservation (Zhang et al., 2020). Avoiding stop-and-go traffic and choosing the shortest, most efficient routes directly translates into fuel savings. This not only reduces fuel expenses but also contributes to a reduction in overall fuel consumption and greenhouse gas emissions, making it an environmentally responsible practice.

- Cost Reductions:

Cost savings are a significant benefit of optimized traffic and routing systems. Reduced travel times and fuel consumption lead to lower operational costs for fleet operators. For individual drivers, less time spent on the road and lower fuel expenses result in personal cost savings. These cumulative cost reductions make optimized traffic and routing a valuable investment.

- Data-Driven Decision-Making:

The incorporation of historical traffic data into navigation systems embodies a data-driven approach to travel. This means that decisions regarding routes and travel times are based on concrete historical insights rather than guesswork. This data-driven decision-making not only leads to efficiency gains but also instills confidence in the chosen routes (Bozorgi et al., 2017).

In conclusion, optimized traffic and routing systems leverage historical traffic and route data to provide intelligent and efficient navigation. By avoiding traffic congestion, minimizing travel times, conserving fuel, reducing costs, and enhancing customer satisfaction, these systems offer numerous benefits for both individual drivers and fleet operators. The data-driven approach ensures that travel decisions are grounded in historical insights, leading to more predictable and efficient journeys on the road.

5 Practical Exploration: Evaluating Overspeeding, Underspeeding, Harsh Braking, and Rapid Acceleration

The Aplicom T10G device is capable of calculating the speed with the help of an accelerometer. Furthermore, with the help of GNSS, the GPS coordinates are also available with the device in order to have full data sets of speed at different locations with accurate timestamps throughout the journey. The goal is to use the device to collect real-time data on speed along with GPS coordinates and timestamps to compare with the speed limits set by the local Finnish authorities. In Finland, the Finnish Transport Infrastructure Authority provides a publicly available API containing detailed mapping of Finnish road and street networks. This database also contains speed limits for both summer and winter seasons with Geo coordinates.

Comprehensive travel data has been gathered over multiple journeys, including in-city and on-highway scenarios, using the T10G telematics device. This dataset offers a thorough summary of travel data collected from urban and highway settings. The device continuously collects vehicle speed data with corresponding GPS coordinates and a valid timestamp and sends the data via MQTT to the cloud. Each data entry contains latitude, longitude, speed and timestamp.

This study's methodology takes a multimodal approach to processing geographical data, starting with the careful mapping of data obtained from the device or through Digiroad API. The gathered information, which mainly takes the form of data snapshot along with gps coordinates, is next organized and formatted into a CSV file.

Figure 13 showcases an all-inclusive collection of information obtained by T10G device. At various places throughout the route, the data snapshot structure provides a full summary of the collected data and connects with particular GPS locations at specific timestamps.

Server time	2023-03-17 19:14:04
unitID	
fieldSelector	00a0df
eventID	112
eventInfo	00
dataValidity	f4
snapshotTime	17.03.2023 19:14:03
gpsTime	17.03.2023 19:14:03
gpsLatitude	61.731841
gpsLongitude	24.714098
numOfSatellites	12
speed	95
maxSpeed	105
heading	256
dInStatus	1
mainPwrVoltage	14355
intBattVoltage	0
trip1	24.655
dOutStatus	0
snapshotCount	287
runtime	3:4:47
acctime	1:25:27
din1Time	1:13:59
voltagetime	3:1:51

Figure 13: Data Snapshot Structure

One of the most important aspects of the data preparation technique is the adjustment of errors in arctan computations. In the setting of geodetic coordinates, where precision in longitudinal and latitudinal values is crucial, this adjustment becomes extremely relevant. A corrective procedure is used to address this, involving use of 1-degree geodesic coordinates. This tactical adjustment improves the accuracy of calculated coordinates, providing a more realistic depiction of geographic places. This is an essential step in order to guarantee consistency and correct irregular geodesic shifts.

The dataset goes through to a transformational algorithm after the correction stage. In the geometric space, this procedure is essential to mapping geodetic coordinates onto a Euclidean plane. The transformation offers a more logical framework for calculating distances and relationships between spatial points, acting as a basic element for further analytical procedures.

After being translated into a Euclidean space, the equated dataset is subjected to additional analytical inspection in order to derive meaningful insights. Events like excessive acceleration calculations and overspeeding detection are carried out with precision. The UTM

(Universal Transverse Mercator) zone for Central Finland is EPSG:32635 and WGS84 (World Geodetic System 1984) is EPSG:4326.

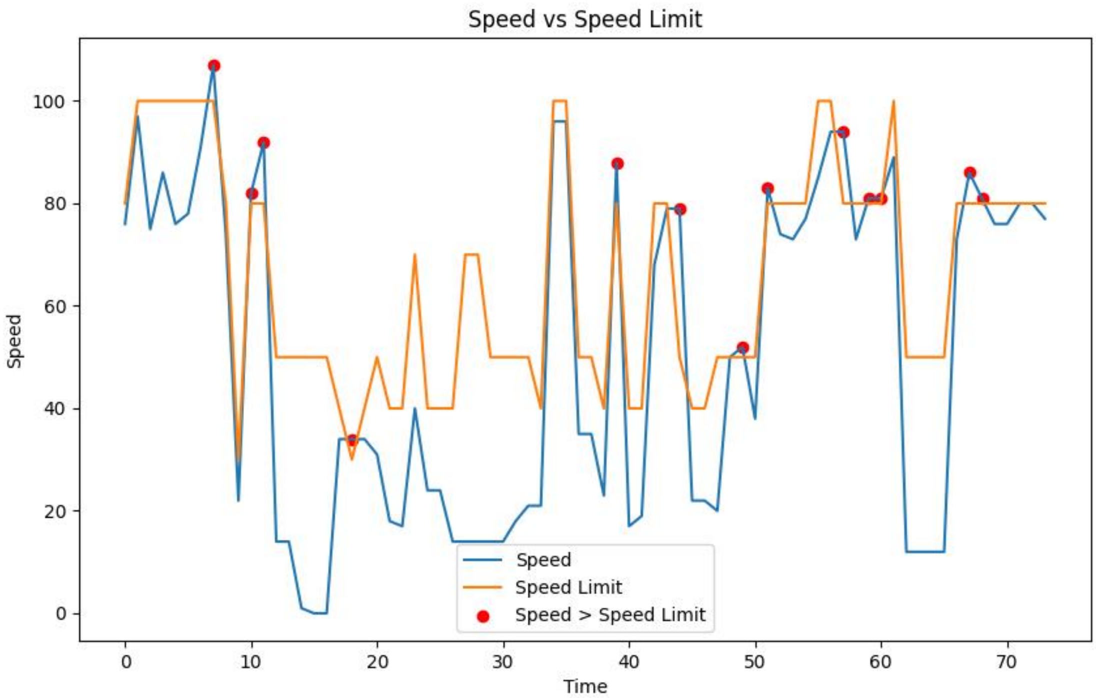

Figure 14: Graphical view of speed limit violations (Speed vs Speed Limit)

In figure 14, different red dots highlight instances of observed speed limit violations on the plotted speed versus speed limit analysis graph. This visual representation displays spots where legally required thresholds were crossed, thus capturing departures from the published speed restrictions. These points are important because they raise awareness of possible violations of traffic laws and raise questions about obeying speed limits. This analytical method offers a useful perspective on driver behavior, enabling a thorough comprehension of speed limit compliance and the identification of particular areas where enforcement actions or interventions may be necessary to encourage safer driving habits.

Figure 15: Speed Limit Violations – Over Speeding (Geo-coordinates)

(Map Data from OpenStreetMap under ODbL License, https://www.openstreetmap.org/copyright)

Figure 15 above shows a map displaying the location coordinates where the speed was above the permitted speed limit and therefore over-speeding is detected (indicated by red circles). The collected device data containing GPS coordinates are mapped with the DigiRoad API speed limits defined resulting in availability of both speed and speed limits data at specific geo coordinates. In this way, it became possible to identify when the driver exceeded the speed limit specified by authorities.

In this case, real-time data gathering, analysis, and API integration allows the monitoring and improving of driving style while supporting wider traffic control and road safety initiatives.

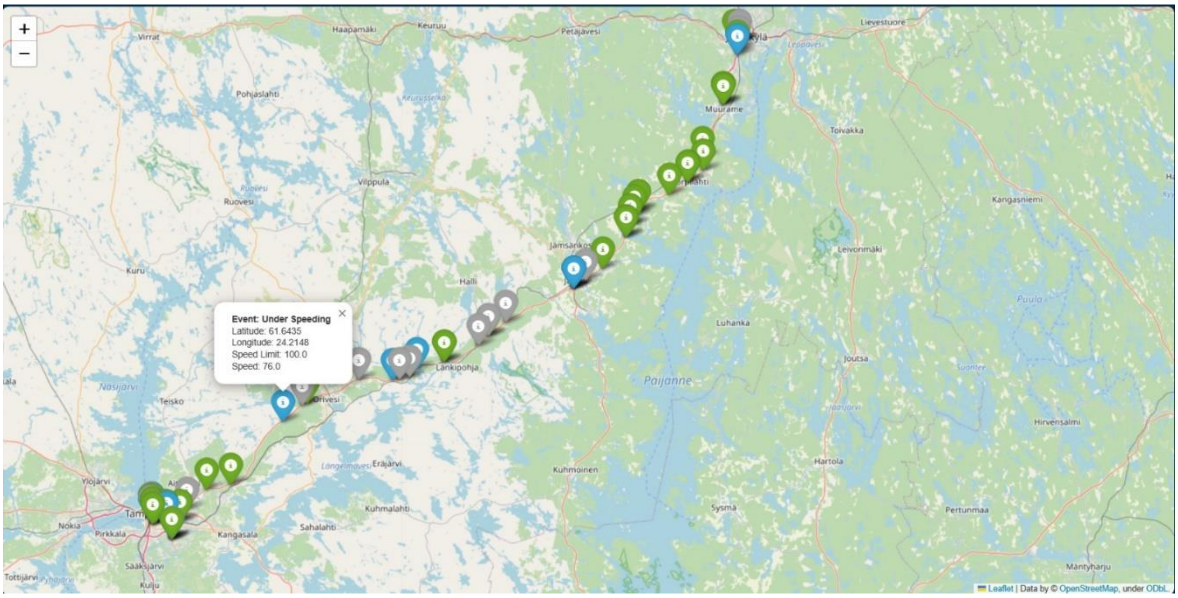

Figure 16: Underspeeding Detection (Geo-coordinates)

(Map Data from OpenStreetMap under ODbL License,
https://www.openstreetmap.org/copyright)

The figure 16 shows the locations on the GPS map where underspeeding incidents occurred throughout the trip. A crucial element uncovered by the data gathered with T10G device is underspeeding, which is defined as vehicle speeds that are below ideal or specified boundaries. With the availability of speed and speed limits data for gps coordinates, it become possible to effectively identify the underspeeding events. The blue markers on the map identify spots where vehicle went underspeeding while the gray markers represent the converging data that points towards sudden drop in speed.

Underspeeding is hazardous because it can make travel times longer, disrupt traffic flow generally, and even jeopardize safety. This highlights how crucial it is to comprehend underspeeding incidents and take appropriate action in order to maximize transportation efficiency and uphold traffic safety regulations.

Figure 17: Rapid Acceleration Events

(Map Data from OpenStreetMap under ODbL License,
https://www.openstreetmap.org/copyright)

The GPS map in the figure 17 illustrates the scattered occurrences of rapid acceleration events during the course of the trip. One noteworthy characteristic identified from the extensive dataset gathered by device is rapid acceleration, which is defined as sudden and significant increases in vehicle speed.

Rapid acceleration occurrences are recognized, which leads to a critical analysis of its possible disadvantages. Frequent or excessive rapid acceleration can result in less fuel efficiency, more wear and tear on car parts, and an increased risk of collisions. This is because abrupt increases in speed have the potential to surprise other drivers and jeopardize general road safety.

Figure 18: Rapid Acceleration Convergence

(Map Data from OpenStreetMap under ODbL License,
https://www.openstreetmap.org/copyright)

The figure 18 provides a closer view of the consecutive data points which are used to evaluate the rapid acceleration. This representation is the outcome of considering a series of data points rather than just one isolated snapshot. The red marker identifies the rapid acceleration event, whereas orange markers represent the convergence points where were consectively evaluated in order to recognize rapid acceleration event. The green markers represent the journey trajectory.

To compute rapid acceleration points, the latitude and longitude values within the data CSV file underwent a sorting process based on geodesic coordinates. Subsequently, these sorted coordinates were employed to derive acceleration and time values. The acceleration threshold values were then applied to identify points exceeding the specified limits. These identified points constitute rapid acceleration points, and the neighboring points within close proximity of the acceleration value of each rapid acceleration point are designated as moderation points.

Figure 19: Harsh Braking Events

(Map Data from OpenStreetMap under ODbL License,
https://www.openstreetmap.org/copyright)

The figure 19 represents harsh braking events during the entire journey. The GPS map provides a clear visual depiction of the locations of hard braking incidents that occur over the trip. These occurrences, as inferred from the data obtained by means of the deployment of device, corresponding to sudden and violent stops at particular points in the journey.

The data CSV file was subjected to a thorough analysis in order to identify hard braking points. This involved carefully sorting the latitude and longitude values according to their geodesic coordinates. These sorted coordinates then provided the basis for deriving the matching time and acceleration values. Finding data points that exceeded predetermined negative acceleration threshold values was the next stage. Extensive examination was conducted to confirm that these negative acceleration readings clearly indicated the vehicle's deceleration phase, supporting the brake value computation.

The identified points with severe braking behavior were identified, and another aspect of this study involved moderation point definitions. These moderation points captured data points that were near the hard braking points that were found, more precisely, those that were within a small range of the acceleration linked to the main harsh braking point. This sophisticated method offers a finer-grained comprehension of the dynamics of the vehicle during braking incidents, making an extensive evaluation of braking patterns possible.

Figure 20: Harsh Braking Convergence

(Map Data from OpenStreetMap under ODbL License, https://www.openstreetmap.org/copyright)

The figure 20 represents a closer view to harsh braking event at a certain gps location during the journey. The light-red marker shows harsh braking event whereas the light-blue markers show the approaching deacceleration that represents the moderation area of convergence in the case of harsh braking.

Important insights into the variability and central patterns of driving behavior measurements have been obtained from the practical analysis. Even though statistical analysis has been effectively utilized in this study to assess driving events, a potential direction for future research is to perform the same analysis with the help of suitable machine learning models described in Chapter 4, and compare the performance in different aspects which could produce high quality insights. This comparison analysis may take into account factors such as processing efficiency, sensitivity to changes in driving circumstances, and event detection accuracy. Furthermore, investigating ensemble approaches that integrate machine learning and statistical techniques may provide a synergistic solution by utilizing the advantages of each methodology.

6 Future Work

Promising future directions exist in driving behavior analysis research with the potential to lower environmental impact, increase road safety, and improve driving pleasure in general.

6.1 Comparative Analysis of Statistical and Machine Learning Approaches

Develop a more comprehensive understanding of the strengths and limits of both techniques by performing a thorough comparative analysis. A precise evaluation of the sensitivity and performance trade-offs can be achieved by analyzing the event detection accuracy for both statistical techniques and machine learning models.

Machine learning algorithms can demonstrate effectiveness in predicting and classifying driving events due to their capacity to identify subtle trends within large, intricate datasets. Examining the degree to which each approach adjusts to changes in driving circumstances, such as weather, traffic volume, and kind of road, may help create more resilient models that work in a variety of situations.

In particular, a direct comparison between the outcomes produced by statistical methods and machine learning models would be very informative to further deepen our understanding and evaluate the robustness of our findings. Furthermore, investigating ensemble techniques that combine the results of machine learning and statistical analysis may offer a hybrid approach. Combining the benefits of both approaches, ensemble techniques may increase the overall accuracy and reliability of driving event detection.

6.2 Driver Profiling - based on parameters / scoring system / reward system

Parameters: Establish a thorough set of factors for driver profiling such as braking patterns, speed, acceleration, and compliance with traffic laws.

Scoring System: Establish a system of scoring that measures driving behavior and allows drivers to track and enhance their profiles.

Reward System: Establish a system of rewards for safe and responsible driving, providing drivers who get high ratings with real benefits such as lower insurance premiums, fuel subsidies and road tax reductions.

6.3 Emotion Recognition

To determine a driver's emotional state, integrate emotion recognition technology into automobiles. This may be useful in determining stress levels and emotional reactions to road conditions.

Conduct behavior studies in conjunction with psychologists and behavior scientists to investigate the relationship between driving behavior and emotional states. It might be possible to spot trends that could result in more specialized interventions.

6.4 External Environment Monitoring

Increase the vehicles' capacity to observe their surroundings. This can involve monitoring air quality for pollution levels, climate, and traffic patterns.

Data Integration: Examine the combined information from multiple sources to determine how external factors affect driving safety and behavior. This may have an impact on safety features, driver alerts, and route planning.

6.5 GUI Vehicle Dashboard (Real-time analysis/feedback)

Real-Time Analysis: Create an intuitive dashboard in cars that offers real-time driving behavior analysis. Drivers are able to view feedback regarding their performance and driving patterns.

Behavior Insights: Provide drivers with safety advice, eco-driving guidelines, and customized recommendations together with visuals and insights to assist them understand their driving habits.

6.6 Centralized Driving Pattern Recognition for Organizations (Commerical Vehicles) - Route Optimization

Establish a single, centralized platform that will allow fleet managers of commercial vehicles to assess and identify driving trends throughout their whole fleet.

Route optimization: Make use of this information to plan commercial vehicles' routes in a way that maximizes efficiency, minimizes fuel consumption, and lessens environmental effect. Additionally, based on driving habits, this can entail predictive maintenance.

To accomplish these objectives, it is imperative to leverage cutting-edge technology, data analytics, and behavioral insights. Furthermore, ethical and data privacy concerns must be met in all facets of driving behavior analysis.

7 Conclusion

A new era in our comprehension of the dynamic interplay between humans and automobiles has emerged from the junction of driving behavior analysis and machine learning algorithms. This research has involved a thorough exploration of the complex network of data produced by modern cars, which are powered by cutting-edge sensors and telematics systems.

The complex nature of driving behavior is one important insight. Algorithms for machine learning have highlighted the significance of customized strategies for various situations. Nuanced insights regarding driver actions have been uncovered thanks to the use of decision trees, which can interpret discrete, rule-based patterns, and random forests, which are adept at managing complicated interactions between factors. Sequential dependencies in time-series data were captured by recurrent neural networks (RNNs), whereas support vector machines (SVMs) excelled in the classification of driving patterns.

Several current models, approaches and algorithms can be used to automatically recognize, evaluate efficiency and forecast driver activity based on available online monitoring data. Among these models are decision trees, which are effective in recognizing unique, rule-based patterns in driving behavior. Random forests excel at managing complex relationships between parameters and provide a more nuanced understanding of driver behavior. Research has shown that ensembles such as random forests often outperform individual decision trees partially due to reduced overfitting (Dietterich, 2000). Support Vector Machines (SVMs) excel at classifying driving patterns, but recurrent neural networks (RNNs) perform best when dealing with sequential dependencies in data, such as those in time-series analysis. Our goal is to gain a comprehensive understanding of driving behavior and uncover insights that might be obscured by a single approach by utilizing this diverse set of machine learning models.

As to how historical data can be utilized in coaching and real-time feedback systems, numerous studies demonstrate the immediate influence that coaching and real-time feedback may have on increasing productivity. Thanks to advanced telematics systems, drivers can quickly obtain feedback on their driving style and fuel efficiency. The most effective approach combines a variety of feedback modalities, adaptive algorithms, and a focus on user interaction and system customization, all of which are intended to promote eco-friendly driving behaviors that are advantageous to both drivers and the environment.

With regards to how historical data can be used to create more precise and efficient models, its various purposes, such as boosting fuel economy and hence cutting expenses, predicting maintenance needs proactively, optimizing traffic and routing, and improving driving efficiency, provide a basis for more precise and superior models that will save expenses and improve fuel and driving efficiency.

It became clear that certain cases required specialization. The machine learning technologies enabled people to drive responsibly, whether it was through driver profile and reward systems that offered individualized incentives or through eco-friendly driving systems that encourage sustainability and responsible conduct.

A common topic was real-time empowerment, which was made possible by the advancement of GUI vehicle dashboards. These displays have given drivers the autonomy to make wise decisions by giving them immediate feedback and insights into their driving habits. But enormous power also comes with great responsibility, and a strong ethical framework is necessary to guide the proper gathering and use of driving data.

Looking ahead, there is a lot of promise for AI-driven driving behavior analysis. The knowledge gathered from this study highlights the fact that machine learning is a fundamental shift in driving that will make it safer, more effective, and ecologically friendly.
The driving experience of the future is expected to go beyond simple transportation and instead involve human-machine interaction based on sustainability, safety, and accountability.

Bibliography

Benyahya, M., Collen, A., Kechagia, S., & Nijdam, N. A. (2022). Automated City Shuttles: Mapping the key challenges in cybersecurity, privacy and standards to future developments. *Computers & Security*, *122*, 102904. https://doi.org/10.1016/j.cose.2022.102904

Bozorgi, A. M., Farasat, M., & Mahmoud, A. (2017). A time and energy efficient routing algorithm for electric vehicles based on historical driving data. *IEEE Transactions on Intelligent Vehicles*, *2*(4), 308–320. https://doi.org/10.1109/tiv.2017.2771233

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984). *Classification and Regression Trees*. Chapman and Hall, New York, NY.

Breiman, L. (2001). Random forests. *Machine Learning* 45(1): 5–32.

Carvalho, E., Ferreira, B. V., Ferreira, J., de Souza, C., Carvalho, H. V., Suhara, Y., Pentland, A. S., & Pessin, G. (2017). Exploiting the use of recurrent neural networks for driver behavior profiling. *International Joint Conference on Neural Networks (IJCNN)*. https://doi.org/10.1109/ijcnn.2017.7966230

Chen, C., Liu, Y., Sun, X., Cairano-Gilfedder, C. D., & Titmus, S. (2021). An integrated deep learning-based approach for automobile maintenance prediction with GIS Data. *Reliability Engineering & System Safety*, *216*, 107919. https://doi.org/10.1016/j.ress.2021.107919

Chen, C., Liu, Y., Sun, X., Cairano-Gilfedder, C. D., & Titmus, S. (2019). Automobile maintenance prediction using Deep Learning with GIS data. *Procedia CIRP*, *81*, 447–452. https://doi.org/10.1016/j.procir.2019.03.077

Chen, H., & Chen, L. (2017). Support Vector Machine Classification of Drunk Driving Behaviour. *International Journal of Environmental Research and Public Health*, *14*(1), 108. https://doi.org/10.3390/ijerph14010108

Chen, S., Yao, H., Qiao, F., Ma, Y., Wu, Y., & Lu, J. (2023). Vehicles driving behavior recognition based on Transfer learning. *Expert Systems with Applications*, *213*, 119254. https://doi.org/10.1016/j.eswa.2022.119254

Dietterich, T.G. (2000). Ensemble methods in machine learning. *Proceedings of the 1st International Workshop on Multiple Classifier Systems, Cagliari, Italy*, pp. 1–15.

Digiroad Data. Finnish Transport Infrastructure Agency. Retrieved April 28, 2023, from https://vayla.fi/en/transport-network/data/digiroad/data

Elamrani Abou Elassad, Z., Mousannif, H., Al Moatassime, H., & Karkouch, A. (2020). The application of machine learning techniques for driving behavior analysis: A conceptual framework and a systematic literature review. *Engineering Applications of Artificial Intelligence*, *87*, 103312. https://doi.org/10.1016/j.engappai.2019.103312

Hu, B., Li, J., Yang, J., Bai, H., Li, S., Sun, Y., & Yang, X. (2019). Reinforcement learning approach to design practical adaptive control for a small-scale intelligent vehicle. *Symmetry*, *11*(9), 1139. https://doi.org/10.3390/sym11091139

Kim, B., Seo, J., Lim, J., & Baek, Y. (2018). Design and implementation of driving information collection system for Driver Behavior Analysis. *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*. https://doi.org/10.1145/3210240.3210818

Kiran, B. R., Sobh, I., Talpaert, V., Mannion, P., Sallab, A. A., Yogamani, S., & Perez, P. (2022). Deep Reinforcement Learning for Autonomous Driving: A Survey. *IEEE Transactions on Intelligent Transportation Systems*, *23*(6), 4909–4926. https://doi.org/10.1109/tits.2021.3054625

Lattanzi, E., & Freschi, V. (2021). Machine learning techniques to identify unsafe driving behavior by means of in-vehicle sensor data. *Expert Systems with Applications*, *176*, 114818. https://doi.org/10.1016/j.eswa.2021.114818

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324. https://doi.org/10.1109/5.726791

Lee, J., Ni, J., Singh, J., Jiang, B., Azamfar, M., & Feng, J. (2020). Intelligent Maintenance Systems and predictive manufacturing. *Journal of Manufacturing Science and Engineering*, *142*(11). https://doi.org/10.1115/1.4047856

Li, J., He, Q., Zhou, H., Guan, Y., & Dai, W. (2016). Modeling driver behavior near intersections in Hidden Markov model. *International Journal of Environmental Research and Public Health*, *13*(12), 1265. https://doi.org/10.3390/ijerph13121265

Linda, O., & Manic, M. (2012). Improving vehicle fleet fuel economy via learning fuel-efficient driving behaviors. *2012 5th International Conference on Human System Interactions*. https://doi.org/10.1109/hsi.2012.28

Liu, Y., Chen, Q., Li, J., Zhang, Y., Chen, Z., & Lei, Z. (2023). Collaborated eco-routing optimization for continuous traffic flow based on energy consumption difference of multiple vehicles. *Energy*, *274*, 127277. https://doi.org/10.1016/j.energy.2023.127277

Marchesini, E., Corsi, D., & Farinelli, A. (2022). Exploring safer behaviors for deep reinforcement learning. *Proceedings of the AAAI Conference on Artificial Intelligence, 36*(7), 7701–7709. https://doi.org/10.1609/aaai.v36i7.20737

Mesgarpour, M., Landa-Silva, D., & Dickinson, I. (2013). Overview of telematics-based prognostics and health management systems for commercial vehicles. *Communications in Computer and Information Science*, 123–130. https://doi.org/10.1007/978-3-642-41647-7_16

Mousa, S. R., Bakhit, P. R., & Ishak, S. (2019). An extreme gradient boosting method for identifying the factors contributing to crash/near-crash events: A naturalistic driving study. *Canadian Journal of Civil Engineering*, *46*(8), 712–721. https://doi.org/10.1139/cjce-2018-0117

Nagalla, R., Pothuganti, P., & Pawar, D. S. (2017). Analyzing gap acceptance behavior at unsignalized intersections using support vector machines, decision tree and random forests. *Procedia Computer Science, 109*, 474–481. https://doi.org/10.1016/j.procs.2017.05.312

Nasr Azadani, M., & Boukerche, A. (2022). Driving behavior analysis guidelines for intelligent transportation systems. *IEEE Transactions on Intelligent Transportation Systems, 23*(7), 6027–6045. https://doi.org/10.1109/tits.2021.3076140

NHTSA. Driver assistance technologies. Retrieved October 15, 2023, from https://www.nhtsa.gov/equipment/driver-assistance-technologies

OpenStreetMap (n.d.). [Tampere to Jyväskylä, Finland - OpenStreetMap under ODbL license]. Retrieved November 15, 2023, from https://www.openstreetmap.org/#map=6/62.259/25.752

Pandey, P., Pandey, D., Kumar, S. (2010). Reinforcement Learning by Comparing Immediate Reward. *IJCSIS International Journal of Computer Science and Information Security, 8*(5). https://doi.org/10.48550/arXiv.1009.2566

Picco, A., Stuiver, A., de Winter, J., & de Waard, D. (2023). The use of monitoring and feedback devices in driving: An assessment of acceptability and its key determinants. *Transportation Research Part F: Traffic Psychology and Behaviour, 92*, 1–14. https://doi.org/10.1016/j.trf.2022.10.021

Ping, P., Qin, W., Xu, Y., Miyajima, C., & Takeda, K. (2019). Impact of driver behavior on fuel consumption: Classification, Evaluation and prediction using Machine Learning. *IEEE Access*, *7*, 78515–78532. https://doi.org/10.1109/access.2019.2920489

Quinlan, J.R. (1993). C4.5: *Programs for Machine Learning*. Morgan Kaufmann, San Mateo, CA.

Quinlan, J.R. (1986). Induction of decision trees. *Machine Learning* 1(1): 81–106.

Rolim, C., Baptista, P., Duarte, G., Farias, T., & Pereira, J. (2017). Real-time feedback impacts on eco-driving behavior and influential variables in fuel consumption in a Lisbon urban bus operator. *IEEE Transactions on Intelligent Transportation Systems*, *18*(11), 3061–3071. https://doi.org/10.1109/tits.2017.2657333

Shan, Y., Zheng, B., Chen, L., Chen, L., & Chen, D. (2020). A reinforcement learning-based adaptive path tracking approach for autonomous driving. *IEEE Transactions on Vehicular Technology*, *69*(10), 10581–10595. https://doi.org/10.1109/tvt.2020.3014628

Virmani, S., & Gite, S. (2017). Performance of convolutional neural network and recurrent neural network for anticipation of driver's conduct. *8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. https://doi.org/10.1109/icccnt.2017.8204039

Wang, F., Zhang, J., & Li, S. (2019). Recognition of dangerous driving behaviors based on support vector machine regression. *2019 Chinese Control Conference (CCC)*. https://doi.org/10.23919/chicc.2019.8865491

Wang, T., Luo, Y., Liu, J., Chen, R., & Li, K. (2022). End-to-end self-driving approach independent of irrelevant roadside objects with auto-encoder. *IEEE Transactions on Intelligent Transportation Systems*, *23*(1), 641–650. https://doi.org/10.1109/tits.2020.3018473

Wickramanayake, S., & Bandara, H. M. N. D. (2016). Poster. *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services Companion*. https://doi.org/10.1145/2938559.2948863

Wolf, Y., Sielaff, L., & Lucke, D. (2023). A standardized description model for predictive maintenance use cases. *Procedia CIRP*, *118*, 122–127. https://doi.org/10.1016/j.procir.2023.06.022

Xie, J., Hilal, A. R., & Kulic, D. (2018). Driving maneuver classification: A comparison of feature extraction methods. *IEEE Sensors Journal*, *18*(12), 4777–4784. https://doi.org/10.1109/jsen.2017.2780089

Yang, K., Haddad, C. A., Yannis, G., & Antoniou, C. (2022). Classification and evaluation of Driving Behavior Safety Levels: A driving simulation study. *IEEE Open Journal of Intelligent Transportation Systems*, *3*, 111–125. https://doi.org/10.1109/ojits.2022.3149474

Yan, S., Teng, Y., Smith, J. S., & Zhang, B. (2016). Driver behavior recognition based on deep convolutional neural networks. *12th International Conference on Natural Com-*

putation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD). https://doi.org/10.1109/fskd.2016.7603248

Yu, R., & Abdel-Aty, M. (2013). Utilizing support vector machine in real-time crash risk evaluation. *Accident Analysis & Prevention*, *51*, 252–259. https://doi.org/10.1016/j.aap.2012.11.027

Zhang, J., Wang, Z., Liu, P., & Zhang, Z. (2020). Energy consumption analysis and prediction of electric vehicles based on real-world driving data. *Applied Energy*, *275*, 115408. https://doi.org/10.1016/j.apenergy.2020.115408

Zhang, X., Zhang, X., & Rong, J. (2014). A Study of Individual Characteristics of Driving Behavior Based on Hidden Markov Model. *Sensors and Transducers*, *167*(3), 194–202.