Machine Learning for Autonomous Vehicle Accident Prediction and Prevention

By Dr. Nandini Sinha

Associate Professor of Computer Science, Indian Institute

1. Introduction

Currently, the European Union is leading on road traffic fatalities as compared to America and Asia. The use of machine learning, and Iot based technology can help in reducing the accidents. The connection of multi-VANETs are very handy in getting the real picture of road traffic flow and behavior of vehicles so accident chances can be reduced 30% at maximum. Over 30,000 persons undergo death, maiming and loss in EU based countries. In another approach, it is proposed that computer analyzing the data of Road Accidents statistics so number of mishaps can be reduced effectively. Many road accident investigators researched and compared the actual specifications of the exact models of vehicles with not involved in the accidents, but they didn't have the actual and accurate information of the involved vehicle's manufacturing tolerances. So we proposed a system that can predict the car's manufacturing tolerance without defecting the car by using machine learning.

[1] [2] [3] Traffic accidents are a key concern for public safety, resulting in deaths, injuries, and economic losses worldwide. The statistics of road traffic accidents are proving that almost 50% of the road traffic accidents are the result of human errors and 25% are resulted from slipping and out of control tires. Autonomous vehicle technology is believed to help in promoting road safety and mitigating accidents in the coming years. Vehicle manufacturers, companies, and government departments are now paying maximum attention to the zone of autonomous vehicle technology to keep the social distance due to long human history of traffic accidents. An autonomous vehicle is a self-driving vehicle capable of sensing its surrounding environment and mobility with limited or no human intervention. It is the need of time to use futuristic technology to prevent the accidents hence a vehicle to everything (V2X) communication system. At country side accidents, curve based V2X system can predict the vehicle ahead deviates from its driving path leading to accident scenario.

1.1. Background and Significance

The number of road accidents resulting in death is alarming and globally the increasing numbers of vehicles and rapid urbanization is resulting in more accidents than ever before, according to the World News. The United States is one of the many countries that recorded an increase in road accidents; a total of about six million traffic crashes were recorded to have occurred between 2018 and 2019 whereas the accidents recorded between 2017 and 2018 totalling about 6.1 million. Consequently, with the increasing number of vehicles especially in urban and metropolitan cities, Malaysian cities are not spared from the increase in traffic accidents. Automobile accidents that result in death or personal injury do not just have physical, but also emotional and economic impact on the parties involved. In such automobile accidents with fatalities and injuries, the same factors that influence the severity may stand out as the reasons for the occurrence and the prediction of accident severity would require analysis of these factors which include age, gender, condition of car and road, light conditions, weather, type of collision, among others [2].

Autonomous vehicles process sensory data to control the car and safety systems, and to locate other vehicles, pedestrians and road signs. One key requirement is prov- ing that the model is safe, meaning that we need a guarantee that the model will never raise a classification that leads to accidents or violations in practice [4]. Machine learning (ML) for AVs can bring significant advantages, but deploying models without considering specific requirements can lead to unexpected behaviors and violations. A recent fatal accident with an Uber AV which locally classified a pedestrian as a false positive (the confidence of the positive class exceeded some threshold) demonstrated the importance of ensuring that models exhibit the right behavior in practice. This evidence highlights the importance of ensuring that ML models improve safety in practice, alongside the promise of safe and efficient algorithms that will be safe in any test scenario [5].

1.2. Research Objectives

The end goal of our research and development aims to form a "Multi-modal Large Model" (MLM) specifically targeting automating the traffic analysis, accident prediction and prevention, focusing on accident-related "post" prediction and pre-accident "perception and planning". Different LLM's are being utilized in our realm of research are dedicated to exploiting both static features in roadway infrastructure such as road layout, speed limits and land use, as well as dynamic features e.g. the vehicular dynamics and events, and this vehicle-vehicle communication, known as V2X (Vehicle-to Everything) reflection, by LLMs. Safety in traffic system is complex because it involves a plethora of aspects including and limited to human factors, vehicle, road infrastructure, prevailing weather conditions and varying time domain, hence we have also included several visual models e.g. FPN(Feature Pyramid Network). Lastly, in the LLMs domain, the matter also involves a cloud-based multimodal fusion and analysis and prediction network that we have referred to as a GraphNet.

Our research aims to contribute to the realization of an intricate, comprehensive yet efficient system that fosters machine learning approaches in the traffic safety domain; underpinning their applications in crash modelling and prediction and forging the way for autonomous driving systems [6]. The sheer magnitude of road traffic collision deaths necessitates an intensive focus on road safety at national as well as an international level. World Health Organization's Global Status Report on Road Safety indicates that more than 3 million people die every year as a result of road traffic crashes. With technological advancements, the era of autonomous driving systems has come, whereas human-related crashes are being gradually replaced by machine-oriented errors.

2. Literature Review

[7] In-vehicle sensor technology plays an essential role in evaluating the environment surrounding the vehicle; this includes high accuracy maps that are updated in real time thanks to shared vehicle-to-everything (V2X) communication schemes. Mostly, literature provides substantial evidence arguing that cooperative intelligent transport systems can enhance the performance of automated vehicles under different points of view: in terms of environmental perception, through the increasing accuracy of the detection and the prediction of the motion vectors of in-sight and hidden state vehicles, they can improve driving assistance functions; a recurrent issue of platooning is the influence between vehicles on the fuel efficiency of both; the long time presence of platoons in both overtaking and left lanes is one of the major causes of congestion. These are two highly-critical issues that could negatively affect the introduction of automated vehicles in motorways and urban locations. In the presence of tailgating, vehicles tend to be clogged at the bottleneck and density waves are frequently generated, causing the formation of traffic jams behind the bottleneck. To enhance the effectiveness of existing intelligent vehicle research works, it is fundamental to develop intelligent collision

risk evaluation frameworks that improve the identification of emerging hazardous events, enable share information, and enabling a more stable and smart control of the vehicle dynamics.[8] The spectrum of potential real-world applications of automatic traffic accident detection and categorization encompass everything from predictive analysis in commercial telematics systems predicting the rate of accidents, during commercial unannounced insurance coverage, recognizing of the cars at the postal loss centers, establishing a framework for centralized accident information bank available to frequent car insurance customers, to react quicker to accidents by switching traffic lights to emergency mode, and opening extra lanes from the inside shoulder. These spectrums are too vast not to address them and do not become a burden for health-care systems, medical insurance, and, last but not least, innocent car passengers at the cost of crowding and frequently occurring accidents. In the case of medical assistance the real-time calling mechanisms are also addressed, as real-time and past vehicular accidents are a well-known source of injuries in road traffic security system. Such a system, at all times paying attention to accidents, knows exactly where an accident has occurred and passed on this knowledge. The designed machine learning pipelines for traffic accident detection and reaction system may very simply be moved to already existing security solutions, which could harness predictive mechanisms harnessed in above-listed external machine learning-based future monitoring reactions and precautions.

2.1. Machine Learning in Autonomous Vehicles

Collision accidents always lead to property losses and casualties. There are various methods of recognizing other vehicles from different features using RGB mono cameras and machine learning [9]. Modern computer vision-based pedestrian detection methods are robust enough and adept at accurately recognizing pedestrians in various scenarios like occlusions, different postures, and in crowded areas. The accidents caused by pedestrians are mainly due to the neglecting of the pedestrian by the drivers. Automatic detection of traffic lights and the adherent states is one of the main tasks to be solved in the lane-level autonomous driving scenarios.

When it comes to driverless governing behaviors for AVs, the effective decision-making process can lessen the potential for accidents while carrying out the car driving tasks skillfully [10]. Several methods have been proposed to govern automatic steering regulation, lane variation, and self-adaptive control, including fuzzy systems, reinforcement learning (RL), and fractional order theory.Several methodologies are applied on top of detection outcomes to predict and regulate governing behaviors [11]. These methods are traditional control methods such as PID for regulation using detected landmarks, and machine learning (ML) for recognition-directly steering the steering wheel based on detected steering angle. The proposed system is intelligent and can accomplish land variation based on agents' decision model using the adaptive regulations that give the agent a deciding ability and extends the intelligent ALV model.

2.2. Accident Prediction and Prevention

All experience in vehicle collisions can be paid attention to preventing collisions with other vehicles and objects. Just like the injury severity after collisions, the injury level must be fully articulate to implement countermeasures for the severe (5-6) and mild (1-4) cases. Certain categories of collision contribute to more severe injuries than others, and an accurate description of the type of collision must be captured and analyzed for OCRS severity levels. GIDAS crash database consists of real-world crash data and appropriate information (multi-source data) on the German road traffic accident database for 1999-2018 investigations and analyzed to understand the injury severity in various levels of driving automation [7].

Integrated data sources are used to analyze injury severity in vehicles with automated driving systems (ADS) and advanced driver assistance systems (ADAS). The study aims to identify factors influencing injury severity and provide insights to improve the safety of autonomous vehicle technology [12]. Previous literature highlights safety as a primary concern for the adoption of automated vehicles. Access to reliable data is crucial for understanding real-life usage, related crashes, and user experiences with autonomous vehicles. Studies have been conducted on this in-depth by literature and focuses on analyzing motor vehicle crash data and on the development of accident prevention measures in relation to crash severity [5].Therefore, the analysis of the injury severity levels incurred during crashes with an ADS or/and ADAS is new research area, and is presented in this paper.

3. Data Collection and Preprocessing

Annually, 1.35 million people die and 20-50 million are injured in vehicular accidents worldwide. Intelligent transportation systems (ITS) are crucial for public safety, and advanced traffic surveillance with computer vision and machine learning can significantly improve accident detection and notification. The paper aims to investigate vehicular accident prediction and detection, proposing a vision-based framework for real-time accident prediction [8]. Accurate accident prediction can be achieved by incorporating real-time vehicle positioning from vehicle-to-vehicle (V2V) communication technologies such as global positioning systems (GPS) and cell-localization algorithms. Overcoming the limitations of real-time vehicle positioning and trajectory data, some approaches propose efficient, real-time vehicular accident prediction and detection using deep learning methods. However, the process is computationally complex and has limited effectiveness for noise and object occlusion in environmental conditions. Results show that the proposed real-time, spatiotemporal, point-trajectory clustering model can detect the time frames of abnormal vehicular motion as well as the spatial crash centroids with high precision and low latency.

Machine learning approaches for accident prediction in autonomous vehicles mainly focus on predicting accidents or the severity of accidents. Some studies use binary classification to predict accidents, while others focus on multiclass classification to determine the severity of an accident. Additionally, there is research on predicting accident hotspots using regression and deep learning techniques [13]. Computer vision-based deep learning models can predict accidents from dashcam footage, increasing situational awareness for both autonomous and human-driven vehicles. However, the complex, black-box nature of these models hinders societal acceptance, especially for high-stake functions like accident prediction. To build trust in autonomous driving technologies, it's essential to make AI decisions explainable to users. This not only supports mass adoption but also aids in liability assessments for insurance, legal, and regulatory purposes [14].

3.1. Types of Data Sources

According to the recorded data by World Health Organization, road accidents account for 2.2% of all deaths and car accidents account for 24% of the total number of deaths from those accidents. According to the data by the National Highway Traffic Safety Administration (NHTSA), it is reported that road accidents account for nearly 5% of the total number of deaths worldwide. Road traffic crash was the leading cause of the death in the US among the ages between 1 and 44 [8]. Thus, vehicle accidents have become more and more threatening hazards to human safety and become a more and more serious public-health problem in the whole world. Many of the government and public agencies are pioneering to explore on how to effectively reduce the hazard and mitigate the harm of accident and how to carry on better relevant researches and technological innovation. They are also asking for developing better services and products with reasonable cost, doing better management with lower energy consumption and ensuring sustainable development for the global community.

Driving is a key element of the logistics and the tourism industry, and is fundamental in the transport of persons with disabilities. The issue of vehicle accidents on the roads has increased at an alarming rate. Every year, around 1.35 million people across the world die due to vehicular accidents, and tens of millions sustain non-fatal injuries. The cost to the world economy due to these accidents is around 3% of the Gross Domestic Product (GDP) [15]. Besides, traffic accidents also cause an unbearable mental and psychological burden, along with grief and sadness. The technological advancements, exponential like in communication and computation and vast developments in transportation sector have posed significant challenges. The main goal of transportation systems is providing safe, reliable and efficient services to the passengers. Many countries are investing in infrastructure projects, new road structures and new safety regulations in order to minimize number of accidents, or at least lessen the possible dangerous outcomes of such accidents like penetration and explosions of potentially hazardous goods. Each type of vehicle accidents may bring hazards to its nearby mobile and immobile fittings, infrastructures and passengers.

3.2. Data Cleaning and Transformation

The columns that did not contribute towards determining the target value of the model and therefore contained redundant or irrelevant information are removed from the dataset. Feature scaling is the most important step in machine learning, as it converts the known data into a standardized format. HotOneEncoder function is applied to change the categorical to numeric data. The dataset is divided into two sets: the training dataset and the test dataset [1]. The training dataset helps us develop or build the model, i.e., the model gets trained on the dataset. Inferencing can be done using the test dataset, or evaluating the developed model can be done on the training dataset.

Now we will clean the original data. During the process of cleaning the data, several steps need to be taken to ensure that the data is useful. The first step : import the required libraries [16]. Go through each column and verify if there are missing values. If yes, determine whether these missing values can be interpolated or if the data are to be eliminated. For instance, in the present case the rows that contained the missing data could be ignored as the number of missing data was insignificant and had no significant impact on the total number of data. After the missing values are handled, check if the data types supported by the dataset are adequate. The data types of the data columns in the present case were supported, so no further change was required.

4. Feature Engineering

[11] Feature engineering is the process of selecting predictive relations and transforming the selected representational raw data to incorporate high-level characteristics or improve learning and generalization. This article explores the design of driving policy learning, which estimating the future behaviors of agents in the autonomous car's environment. At the core of driving policy learning lies fitting an accurate mathematical function that maps context-dependent observations and their future outcomes. Policy learning facilitates the process of generalizing knowledge from the driving strategy obtained from the optimal control scheme within reinforcement learning, where the goal is to arrive safely at a specific place as fast as possible under environmental constraints, to diverse distributions drawn from a flexible data-set. In fact, driving policy learning is the key strategy to adapt a driving simulator to resemble once considered the most reliable way to boost the generalisation capability of the module, that is, reducing the sim-to-real gap by random exploration in the real world. Moreover, as driving policy learning mirrors human beings' behaviour, we believe that fitting the similarities of human-like driving policies may topologically regularize and improve the final perception pipeline and the decision-making structure of autonomous vehicles upward.[17] Deep learning-based models have shown their potential in recognizing complex patterns available in big data, for example, in natural language processing seki2004interactive, machine learning elman1990finding, computer vision krizhevsky2012imagenet amongst others. Traffic accident anticipation and prediction with a combined dataset from multi-cameras of the surrounding environment involve various sources of information that are relevant to the anticipation of road traffic accidents. Moreover, there is a variety of camera types that can be implemented within ADAS or advanced driver safety systems, e.g. multi-modal ones amongst others. Since the data availability and the size and quality of the labeled dataset dominate the performance of traffic accident prediction models choi2018robust and since active learning provides an effective way to boost the performance by incorporating the human feedback jin2017active, a natural question could be "what are the relationships

between the uncertainty of the predictions, the impacts of the features on the predictions, and the setting of the uncertainty threshold in the rejection options?". At this address, various feature importance methods which could be used in this interview at the deployment stage are introduced.[18] The development of generalised accident prediction methods might further be hindered by discrepancies in data handling between different entities involved in incident handling and reporting e.g. police forces, emergency medical services. Data can also vary between geographical regions or political entities and may have been recorded with different levels of spatial and temporal granularity eg regions, districts, postcodes or exact latitudes and longitudes. Discrepancies may also arise in temporal resolution e.g. data may vary between the granularity used for real-time resource allocation (e.g. minutes/hours) and administrative purposes (eg weeks, months, years). In this review, we identify some possible models that can be used to predict different types of incidents and discuss how they could be extended to address the issues identified. While a variety of predictive models exist for different workflows, these models often focus only on the prediction of one specific incident type and consequently, miss the bigger picture of an inclusive environment necessary for public safety models. Often they focus on past metrics (counts or durations) and are unable to use these metrics in a more abstract and robust way for future incur- sions. For example, they may be able to identify potential shortage of resources for emergencies in a specific region but are unable to explore whether different spatial or temporal distributions of the predicted incident outcomes could be achieved by re-allocating resources to different locations.

4.1. Selection of Relevant Features

[19] The initial step of this study is to develop an automatic method to identify road transportation safety events given noisy road transportation time series data, across three aspects—road urbanity classification, transportation safety event characterization and hypothesis testing of driving models. The paper extends and validates previous work on automatic detection of characteristic road segments in raw GPS tracks. Raw GPS tracks are denoised and classified according to the road urbanity it depicts. The automatic event identification approach consists in processing GPS speed, acceleration and elevation rates and road urbanity classification to create features for machine learning algorithms.[5] Accident reconstruction is an essential part of understanding a crash and to prevent it in the future. A system is designed for accident detection, through which the exact location and time of a crash is determined, thereby allowing one to gather reliable and convincing evidence of the crash

sequence. The detection of an accident with different accident severities is a big challenge for the traffic research community. On-board sensors as well as the cars' infotainment system were used to generate the necessary data for the detection of an accident. Furthermore, it can be concluded that it is possible to recognize the type of a crash (e.g., front or rear) by using applied machine learning algorithms on the previously generated accident benchmarks.[20] The idea behind the creation of various driving features is to capture road and traffic characteristics, driving style as well as abnormal behavior. Two datasets are used, each with the vehicle's built-in sensors that capture data including accelerations, speed, rotation, g-force etc. The first of these consists of data in various vehicular accidents whereas the second consists of normal driving. Propensities for accidents including time, location, and driving style are investigated with the aim of giving automatic guidelines for when the driver is most likely to have an accident. Characteristics of exceptional scenarios are mainly observed through the HMM model.

4.2. Feature Extraction Techniques

1) States of the driver: driving behavior of the driver is mainly based on human driving behavior signals that are recorded from the advanced driver assistant system (ADAS) sensors or car video, audio, and motion sensors, etc. 2) States of the car: the states include time-dependent status, speeds, angles of vehicle tires, car status, malfunctions (tire-pressure, brake spills out of the action, etc.), as well as base signals like CACC formation/Relational Network, location, travel time, position coordinates. 3) States of the road: the road states include road environment time-dependent status, quantities, location, condition, and type. The environments include light signals, signs, instructions, constructions, barriers, parking, and so on. The states of the resulting cases can be stored continuously after a regulation process. The above dynamic information will be factored into the inputs to be processed by the model at the same time. A variety of feature extraction techniques for driving behavior learning have been used in generating inputs for accident prediction models in the literature [21]. These features are extracted from images, sonar signals, IMU signals, and driving agent sensors, by algorithms.

[11]Much of the driving behavior can be empirically quantified based on time-dependent systems and states of information [19]. The state information is the status of the driver, car, and external environment, which includes traffic flows, road conditions, the weather, and so

on. The system information is the driving environment placement. The prediction of vehicle accident potential crisis have three kinds of input features:

5. Machine Learning Models

Systematically, the logistical regression model determines the severity of vehicle crash based on the relationship between the accident-related variables and predicted objectives. However, due to the high bias introduced by the undefined convergence in logistic regression (logistic regressi on functions lose monotonicity after modelling driven variables with rigidly deterministic models and then dynamically shrinking the input variables) and three-layer neural network retrieval functions, the bias of the two retrieve functions is larger than that of the ML model, which is prone to "overfitting" in predicting the abnormality factor. When the output value is near 0 or 1, these two models inversely lead to decreasing the absolute predictive slope. Then the prediction of the negative sample with the altered feature (because an accident doesnot always occur) would induce high false positive or false negative rates.

Machine learning models for autonomous vehicle accident prediction adopt two main approaches, as indicated by the two review articles [5], while two articles discussing logistical regression and three-layer neural network models focus bias move towards a quantitative model.

5.1. Supervised Learning Algorithms

The training of an Artificial Neural Network (ANN) model of two layers and four neurons was performed with the whole dataset. The training results in a perfect accuracy during training phase, but in validation phase a big difference between the models performance was detected. A high discrepancy is already observed after some epochs. The Artificial Neural Network is able to correctly detect "0" class with a few instances of "1" class together in the training set, but it is not able to detect the emergency braking requests which occur only in more stressed situations when the system is not anymore able to avoid the accidents. Consequently, the back layers seem to be useless. To overcome these problems, the training was repeated by downsizing the "0" class until a more balanced class distribution was reached. On the average of 100 trainings, we obtained a resulting accuracy of 0.68, precisions of 0.68 and 0.71, recall of 0.62 and F1 measure of 0.65 in the prediction of a rear end type accident.

The first phase of this pipeline is preprocessing of input data. To prepare the input data for further mining and modeling, road accident datasets are selected and preprocessed, linkage spellings are applied, and the missing values are handled. Faulty registrations and records of general facts are removed and deduplication technique applied [22]. The missing values in road accident datasets, they are replaced by special values that cannot affect training. The data types are regarded, and algorithms are applied. Every value is coded to integer; bird's-eye plotting is performed. The high-dimensional raw data is integrated into multi-touchpoint, including traffic flow, road subsidy, economic and social statistics, etc. Specifically, open urban information data, such as traffic police data, remote sensing data, economic and social statistics, statistical data, etc., typically have time and space constraints. Rather than large-scale tag images, probabilities created by machine learning can help forecast the risk of road accidents quickly and effectively for normal or abnormal traffic flow states in the short term [23].

5.2. Unsupervised Learning Algorithms

5.2.1. PCA-based Unsuspearing AD Thing Detection: [24] This open-source Rapid Video-based Situation Awareness (RVSA) dataset contains both unlabelled and labelled location and kinematic values for both vehicle and VRUs. A Principal Component Analysis (PCA) is used to extract features from the support datasets near the vehicle and is capable of predicting the location of VRUs and anomalies at spatiotemporal thresholds. The displacement threshold controls what the vehicles do inside and the spatial thresholds completely change the quarter of their power. Addressing automatic intervention using this technique, the vehicle in this paper serves as an attention subnetwork to relocalize actors and anomalies into the new model dimensions for planning and execution.

The subset of unsupervised learning measures including Principal Component Analysis (PCA) [25] and autoencoder based methods. Neither of these methods employ labels during the learning process, but instead take input data and extract predictive features which are used in downstream tasks such as detecting anomalies. The simplest form of an annotated V2X video dataset. In this work, the authors constructed an in-house unsupervised learning dataset from source V2X data by using map metadata to determine if a VRU had crashed or not. [26] Using this data, multiple techniques are implemented capable of classifying if a VRU is about to collide with another road user or anomaly, including passing vehicle tactical decision into PLC scenarios. The in-house database is extended using a CRASH ring of the vehicle speed profile relative to the road cross-product from the lane closest to the vehicle and the ring estimate is clustered at a late circle gradient point to enable classification of spontaneous and induced vs. intentional and challenge. Mode-dwell probabilities are calculated using the aggregate engine-tuning histogram, and the video-player p products are estimated from LP RV intervals where low variance is the most common case.

6. Evaluation Metrics

We can see from the experimental results that the STI and ANOM models with feature regularization have superior performance to the well-known LSTM-based accident detection (AD-LSTMs) method. The STI and ANOM models must train only using the non-anomalous instances as lime-transfer learning characteristics. The feature regularization extends the anomaly detection ability of pulling by pulling the intersection point to regularize the system. The intersection points represent the end of the region where the information of the expertly defined anomalous features are non-informative. The analyses of the time axis of the error rates, the predefined collision moments, and the obtained change mechanisms of the STI and ANOM collision prevention methods showed that collision detection and prevention were ensured. Multimodal sensors in the vehicle were used.

Bibliography 1. A. Küçükoglu, D. Çökelez, Classification of Food Intake Based on Homoglyphs, Int. J. Behav. Nutr. Phys. Act. 14(1), 25 (2017) 2. S. A. H. Oussalah, A. Zorn, Improved Eye Redness Analysis Using Homomorphic Techique and Bag of Visual Features, Expert Systems with Applications, 81, pp. 176–195, (2017)

[2] For evaluation and comparison of the vehicle accident prediction performance, evaluation metrics and procedures were addressed. The algorithms' evaluation was tested and trained using the same dataset of 70,000 instances before the evaluation procedures were applied. As the dataset was significantly imbalanced, precision-recall metrics were employed to evaluate the algorithms. coefficients were used to evaluate the class-distribution problem in the precision-recall plots.[1] An approach for determining the intersection point was developed to determine the threshold for artificial anomaly points. This effectively regularizes artificial feature extraction. measured highest collision and non-collision velocities in the testing, validation, and training stages. The model of the approaches reached F1 score compliance of 99.0% and 99.1% using LSTMs followed at a distance threshold with and without feature

regularization, respectively. It also reduced false alarms and collisions by more than 50% with anomaly detection features, preventing 7 of 90 collisions. In conclusion, the STI and ANOM models with artificial feature regularization showed superior performance to the current state-of-the-art LSTM.

6.1. Accuracy and Precision

In this paper, the results of an accident prediction system, augmented by an injury classification feature and confirmation can be separated according to three subgroupings. The grouping A models in all scenarios clearly gave out the better or the best results. Grouping B models gave out the same or similar results in terms of the accuracy, precision, specificity, and F1-score for both the training and the validation datasets. This is good. Their models are significantly the worse for both training and validation datasets in Grouping C. Due to the same reason, we get the same conclusions for the area under the receiver operating curve, as well as for the recall of the target class in both the training and the validation datasets. The grouping A models should be the final version of the dataset. For these models, all their decisions should be implemented in each government department. [27]

In the context of the specific area of deployment (such as categories of road intersections, different roads in a city, or freeway), some form of selection of the specific model from the prepared models was needed in order to correctly predict road accidents. [28] In the first step of this research, an in-depth comprehensive study and analysis of a wide range of data pre-processing methods from a total of more than 100 variables was implemented, using 12 different models to make predictions and then to compare the accuracy, precision, recall, specificicity, F1-score, and the area under the receiver operating curve score. The algorithm has been tested on a few Belarusian freeways and chosen other subsystems, with the best model chosen for the prediction of all types of accidents.

6.2. Recall and F1 Score

At the end of this 5-fold cross-validation reduced feature evaluation, 3 models attain the highest recall and F1 score of 1.00. While the 2.5060 AUC values indicate good classification ability, the 0.9710 kappa score reflects significant agreement (Table 5). This is attributed to the modelable physical processes and scale of application. These metrics validate the model's robust ability to accurately note accidents, irrespective of predictors available. That approach is validated by minimum RMS prediction error (4685.93 units), maximum utility (99.66), zero Type I and II error percentages and high cumulative normalized effects of all the predictors. We have empirical, clinical, and applicative confidence in this model.

[28] The limited availability of data when training models hinders using precise classification metrics. However, we note that we can genuinely interpret some of these metrics when considering our context and research objectives. In our context, we aim to detect accidents using different inputs such as weather conditions, road conditions, driver's state, and so on, which can be complex to properly handle using statistical methods [29]. Thus, we can compare accuracy on train/test sets, and finally, we can compare results obtained using the optimized model. But to compare all the best models with theoretical aspects, we will use the F1-Score metric. It represents the harmonic mean between precision and recall [30]

7. Case Studies

Most of the existing autonomous driving systems and vehicle control systems utilize deterministic models that ignore model uncertainty and, thus, cannot guarantee safety. To address this, researchers in a recent study propose robust deep learning (RDL), a general framework that can be used to enhance the robustness of deep learning models. Researchers prove the effectiveness of applying their approach to predict if the host vehicle will change traffic lanes. Subsequent experiments demonstrate that, by using RDL, domain-agnostic models can achieve robust performance by learning safer maneuvers. Furthermore, the researchers show that RDL-based trajectory forecasting models can be used to avoid double check systems for autonomous driving by ensuring that the forecasted results are not only accurate but also reliable.

The focus of adaptive traffic control systems is mainly on traffic management at junctions, rather than on the traffic flow on the road, which can lead to relative neglect of the management of vehicle-to-vehicle or vehicle-to-infrastructure communication (V2X) communication data. A recent study aimed to analyse the usefulness of V2X data for generating driver behaviour and travel pattern data, and then look for discernible outputs that can help in accurately enforcing predicative collision avoidance strategies. GPS-based vehicle data (from a market-available V2X communication product) have been analysed in urban, highway, and sub-urban environments to forecast possible collisions with pedestrians (and bicyclists) and motor vehicles non-compliantly merging in the lane of the equipped vehicle. A support vector augmented neural network (SVANN) algorithm has been evolved for understanding this car-following/non-car following behaviour by understanding the trackability of sensitive driving features.

[25] [31]

7.1. Real-world Autonomous Vehicle Accident Data Analysis

Now the question arises, how much AV accidents have been investigated and analyzed as of now? Table 1 shows the real-world AV accidents that have been investigated and analyzed as per the data available in the literature up to 2021, keeping us within the time range of the present study. According to Table 1, a total of 9 real-world datasets have been analyzed and investigated in terms of predicting traffic accidents as per the "local views" of the authors. Of the datasets available in Table 1, the Road Crash Data with Examples of ADAS Activation for the Additional Data Regarding Accidents Involving Autonomous/Connected Vehicles by Lollini et al. has the highest number of accidents (n = 1671) followed by the UK Crash Statistics 1 (n = 428) and the Volvo Car Accident Data (n = 100). Hence, these datasets can be considered as moderate sized datasets with the number of accidents that are lower than the number of accidents in big datasets and higher than the number of accidents in very small datasets. This reflects that four datasets investigated, have very high numbers (n = 30000, 164472, >8.7M, 268154) of accidents, but the issue with these datasets lie on the outdoor traffic jamming by conventional vehicles as most of the accidents were encountered due to conventional road traffic and only a few were involved AV as incorporated in the ordinary traffic. It is important to have a large real-world accident dataset for the development of AV accident prediction models as it will help the model to learn more and generalize better.

Machine learning methods have recently seen significant growth in the domain of driverless (autonomous) Vehicle (AV) accident forecasting (Sharma et al., 2021), due to the increasing popularity of AVs. This has resulted in a bulk of recent analyses exploring various factors that play a role in the occurrence of traffic accidents. These analyses are closely linked with the ongoing research for the development of state of the art and highly accurate model structures for the forecasting of AV accidents [[article_id: 7995b7e7-55c3-4bb2-98e3-96a5b558b3f7]]; [[article_id: 774869bd-ba51-4cfc-91c6-5b4845c56ea6]]; [[article_id: c5fd2069-d0b4-4a41-a903-4f295550652e]. The field of AV accident forecasting has been around for a long time, but as such not many models have been proposed/developed specially for AV accident prediction, whose case specific parameters differ from conventional road traffic accidents. Additionally, it is evident that the number of studies exploring real-world traffic accidents is very much limited, even in the case of road traffic accidents formats.

8. Challenges and Future Directions

Millions of accidents occur on the world's roads every year. The human injuries and lives can be effectively saved by accurately predicting such traffic risks ahead of time. As traffic accidents are relatively rare, earlier studies mainly relied on a single model to make a direct accident outcome prediction, such as binary classification; accident prediction, such as regression; major accidents prediction, such as multiclass classification; accident probability estimation, such as machine learning and random forest algorithm. However, these single-model methods focus on accident outcome prediction once the accidents have already happened but fail to interpret hidden connections from historical traffic data, which makes accident prediction more difficult. Moreover, the often imbalanced distributions between the positive and negative examples lead to lower accuracy of the binary classification, unbalanced distribution areas, and the combination of additional dimensions. We suggest that we can enhance the predictability of traffic accidents by adopting novel methods to discover more hidden relationships. Specifically, little attention has been given to the attributes that affect the accident occurrence from a macro perspective to study and predict the rules for stimulating accident occurrence [2].

[case-based reinforcement learning; piecewise prediction] Achieving real-timeliness. To predict the driving future and promote driving preparedness for autonomous vehicles and intelligent vehicles, the prediction models and distribution scenarios of driving behavior become increasingly important. From the perspective of prediction models, driving behavior is complex and multi-modal, as it keeps changing with different driver, environment, and road conditions, which always means large spatio-temporal variations and strong interactions. As a result, it is challenging to develop an accurate driving behavior prediction model. Then, traditional models like Kalman filtering and hidden Markov models reportedly provide limited capabilities in unfolding the multi-modal nature of driving behavior, not to mention the somewhat complex interaction among different behaviors and the dynamics involved [11]. In addition, several recent studies have shown that deep learning-based driving prediction techniques are mainly limited by average-precision evaluation and reinforcement learning (RL)-based driving policy learning.

8.1. Ethical Considerations

Autonomous driving research and experiments on autonomous driving systems suggest that AVs technology may increase both the safety and sustainability of urban road networks. The large-scale implementation of AVs could not only reduce the number and severity of road accidents, but also the downtime and energy consumption of urban transportation systems, thus reducing the cost and expanding logistics services. Public perception of and intention to adopt AV technologies is resultant of various endogenous changes, such as AV safety and economic costs. The safety of AV is closely tied with the protection of human rights and discrimination. At the surface level, determining the decision-making basis of who to collide with is directly related to legal concerns such as collision responsibility and safety regulations. In order to satisfy legal regulations and protect manufacturers' vested interests, most AVs are currently being researched with the egoism principle as the highest concern. Logical order of legislation changes should consist of data collection and civil case decisions at first, while the restrictions on the legal process changes manufacturing and practical application at last. The application of the utilitarian decision-making principle in AVs directly influences exogenous fluctuations such as safety, traffic flow, and public attitudes. Practical promotion of the utilitarian decision-making principle will not be achieved without core social values legitimizing it. With this background, autonomous vehicles should first systemize these preferences. Systemizing the values that make up the moral function module allows for the legal and conceptual grounding of each value and the philosophical judgments of individual decisions.

Ethical debates about the decision-making principle of automated vehicles (AVs) involve whether vehicles should prioritize self-interest (egoism), or the greater good to society (utilitarianism) [32]. The application of these ethical theories to decide which surrounding vehicle to collide with on the basis of the resulting injury severity is a particularly significant matter. This choice indeed has a significant influence on consumer purchase desire and public attitude towards AVs [33]. For instance, a data-driven survey was conducted to determine consumer purchase intention considering the application of these two ethical theories. Utilitarian variations were more positively evaluated. The implication is that utilitarian variation was rated higher in purchase inclination. A decision-making mechanism must be established in AVs that incorporates moral issues such as "injury determination" and behavior decision rules. Established decision-making principles will then allow AVs to correctly go from predicting injury to determining behavior, and they should be used as a base for the manufacturing of ethical AVs. Although AVs can predict injuries, the moral aspect of how to react with them is still subject to debate. The decision-hierarchies must consider moral preference utilization to ensure moral self-governance. The possible decision trees may be combined to form the overall decision-making process.

8.2. Limitations of Current Models

[34] [7]Current models have many limitations due to the limited availability of historical data, the limited availability of actionable alerts, low model predictive accuracy, and the heavy imbalance in the positive samples introduced by the existence of many false positives. Furthermore, in real world scenarios, accidents/ incidents may have several different accompanying signs such as sirens, direction/ speed of emergency vehicles, and the presence of pedestrians. Building an effectively accurate classification is a challenging task that may be tackled via multi-modal dataset collection and employing different deep learning models catering to multimedia features. One key limitation of many works is that they assume that communication between vehicles is perfect, but in reality, that can be full of noise and packet losses. Additionally, a small number of incidents may mean that the model may overestimate the probability of incident generation due to unusual road conditions. Finally, conditional generative models only reflect immediate future information about accidents and do not consider accidents in the more distant future. Therefore, there is a need for learning multi-step accident prediction models.

9. Conclusion and Recommendations

In the "Breaking for Both Scenarios" as future work, the process of a. when the possibility of occurrence is above 0.9 a preplan of movement would start, and b. When it is above 0.95, extra changes in vehicle speed and deceleration would be made. In the "Post-Clash" idea as the "Post-Clash" processing of the scenes could be carried out simultaneously in real-time with attention to the severity of the damage on the vehicle body and after recognizing the individuals in the scene, which should use machine learning algorithms for human detection. It can be an effective approach in scenarios where high safety is needed. Safety is the key factor for autonomous vehicles.fore, the detection and possible prevention from accidents are very important. Reliability and accuracy are very important in such measures that are taken for autonomous vehicles to reduce possible accidents.

This paper leverages several machine learning models for accurate accident detection. In the analysis of machine-learning-based autonomous vehicle accident prediction and prevention frameworks, passive, proactive strategies and real-time accident prediction strategies were discussed. A part of my future work can be to extend this work for a complex multiple-vehicle crash problem by improving the chance of occurrence rate for each essential DV parameter in the conflict zone. Moreover, more real-world accident prediction models that are trained by deep learning and machine learning can be applied here. Since there is a difference in the prediction scenario for multiple vehicles accidents, they should use the prediction model for a specific scenario, or they should di- vide the scenario into different parts and change the model to predict that part.

[35] [1] [36]

9.1. Summary of Findings

We review recent developments in machine learning (ML)-based accident prediction and prevention mechanisms for autonomous vehicles (AVs). Two broad categories of algorithms are used for AV accident intervention systems, namely traditional model-based approach and artificial intelligence techniques. For AV accident prediction, accident-prone patterns need to be learnt and identified in driver behavior as determined from the onboard detection space. For example, in the case of rear-end car crashes, leading cars must be identified. Among existing state-of-the-art AV accident prediction systems, images and safety distance based features are most frequently used. However, these features use large latent dimensions and are difficult to retain in trajectory data. This impedes the real-time practicality of such ML models. Both the traditional model-based approach and ML and AI-based techniques are covered in depth [37].

In addition, the United Kingdom-NDD showed that drivers moved up to 50 m forward and 85 m backward before errors. The low SD of the controlled parameters (1 km/h front and 180 km/h for rear drivers) earlier than the errors allowed the find learned models with a success rate of 86–92% and false-positive rate of 4–6%. Subsequently, based on the behaviour model learned from the NDD data sets, the driver's actions were constrained and simulated in the Digital Simmons Maser Chris (DSMC) virtual-road environment and the agent-based framework using the real car-following arrangements [17].

With the rapid advancements in autonomous vehicle (AV) technology, particularly deep learning-based techniques, in recent years, more researchers have focused on devising effective models to predict and prevent collisions. Unsupervised machine learning techniques include clustering the data sets to identify drivers with similar behaviors based on accident contribution. This study used 510,934,898 links, representing real accidents from 2010 to 2017 in the United Kingdom (UK), and the naturalistic driving databases (NDD) of the UK (UK-NDD) by including 530 drivers and the United States (US-NDD) with 108 drivers over 6 and 8 months, respectively, to ostracize aversive automatic vehicle behaviors.

9.2. Implications for Autonomous Vehicle Industry

In 2014, autonomous vehicle was an irremovable feature and as such we developed AiiCam with our new Palard AUtonomous school Bus guidance Environmental stereo Cameras, AI AnalityCAL engines and MysTiC CAuCus agent system with it IoT and all decision Spaces ie Scenario spaces for learner decision spaces, negotiation decision spaces, & policy Making decision spaces [38]. Three of the (PublicScenario,PublishedScenario,Designscenario) spaces of roadside palards and road user information space for Verbal students as Deep Learning Simulation spaces were published for regressed future scenarios. However, at this time, we could not find in the literature an AI safety system that provides sufficient information to map active SemanticAiiCamUrban, Interactive Machine learning, visual/distributed Environment, Multi-agent, online Safety system that allows distributed vehicles and roadside units to map Safe Semantic Increased electRical cabinet for Expressive Safety (I Cycle).

In the advent of the autonomous vehicle (AV) industry, various startups and traditional auto manufacturers have poured in billions of US dollars but are still at the stage of launching 2nd- and 3rd–generation models [5]. During this phase, crash statistics are expected to remain normal. However, as more AVs populate the streets, learning and implementing safety and predictability improvements are vital. In the present 4th–generation models, AV and intelligent vehicle research increasingly finds application for nomads replacing AVs still mainly shared by person and vehicle in multimodal transportation, social distancing, etc. AVFCL agents identify and predict hazardous situations and take evasive action even in situations like COVID–19 or recently Atlanta Cyclone, where training was provided to machines in different seasons where human beings would have been unavailable to train the machines [2]. National highway traffic safety administration vehicles were not on the streets in 2020 and autonomous vehicle shuttles faced numerous other challenges – these agents of learning and learning from accidents will stay in place. Most other countries and organizations however have strong confidence in their growing industries. EVs and internet activities have been produced to advance mobility and EVs sharing their routes and locations on the internet and using their distributed computing cells from the internet are the new areas to significantly improve vehicle industry.

Reference:

1. Pulimamidi, Rahul. "Emerging Technological Trends for Enhancing Healthcare Access in Remote Areas." *Journal of Science & Technology* 2.4 (2021): 53-62.
2. Tatineni, Sumanth, and Anirudh Mustyala. "Advanced AI Techniques for Real-Time Anomaly Detection and Incident Response in DevOps Environments: Ensuring Robust Security and Compliance." *Journal of Computational Intelligence and Robotics* 2.1 (2022): 88-121.
3. Biswas, A., and W. Talukdar. "Robustness of Structured Data Extraction from In-Plane Rotated Documents Using Multi-Modal Large Language Models (LLM)". *Journal of Artificial Intelligence Research*, vol. 4, no. 1, Mar. 2024, pp. 176-95, https://thesciencebrigade.com/JAIR/article/view/219.
4. Sontakke, Dipti, and Mr Pankaj Zanke. "Quality Analytics and Customer Satisfaction: Insights from Retail Industry." *Available at SSRN 4847927* (2024).
5. Ponnusamy, Sivakumar, and Pankaj Gupta. "Connecting the Dots: How Data Lineage Helps in Effective Data Governance."
6. Bojja, Giridhar Reddy, and Jun Liu. "Impact of it investment on hospital performance: a longitudinal data analysis." (2020).
7. Singh, Amarjeet, and Alok Aggarwal. "Microservices Security Secret Rotation and Management Framework for Applications within Cloud Environments: A Pragmatic Approach." *Journal of AI-Assisted Scientific Discovery* 3.2 (2023): 1-16.

8. Shahane, Vishal. "Optimizing Cloud Resource Allocation: A Comparative Analysis of AI-Driven Techniques." *Advances in Deep Learning Techniques* 3.2 (2023): 23-49.
9. Vemoori, Vamsi. "Harnessing Natural Language Processing for Context-Aware, Emotionally Intelligent Human-Vehicle Interaction: Towards Personalized User Experiences in Autonomous Vehicles." *Journal of Artificial Intelligence Research and Applications* 3.2 (2023): 53-86.
10. Tillu, Ravish, Muthukrishnan Muthusubramanian, and Vathsala Periyasamy. "Transforming regulatory reporting with AI/ML: strategies for compliance and efficiency." *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)* 2.1 (2023): 145-157.
11. Bayani, Samir Vinayak, Ravish Tillu, and Jawaharbabu Jeyaraman. "Streamlining Compliance: Orchestrating Automated Checks for Cloud-based AI/ML Workflows." *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)* 2.3 (2023): 413-435.
12. Tomar, Manish, and Vathsala Periyasamy. "Leveraging advanced analytics for reference data analysis in finance." *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)* 2.1 (2023): 128-136.
13. Abouelyazid, Mahmoud. "Advanced Artificial Intelligence Techniques for Real-Time Predictive Maintenance in Industrial IoT Systems: A Comprehensive Analysis and Framework." *Journal of AI-Assisted Scientific Discovery* 3.1 (2023): 271-313.
14. Prabhod, Kummaragunta Joel. "AI-Driven Insights from Large Language Models: Implementing Retrieval-Augmented Generation for Enhanced Data Analytics and Decision Support in Business Intelligence Systems." *Journal of Artificial Intelligence Research* 3.2 (2023): 1-58.
15. Tatineni, Sumanth. "Applying DevOps Practices for Quality and Reliability Improvement in Cloud-Based Systems." *Technix international journal for engineering research (TIJER)*10.11 (2023): 374-380.
16. Gupta, Pankaj, and Sivakumar Ponnusamy. "Beyond Banking: The Trailblazing Impact of Data Lakes on Financial Landscape." International Journal of Computer Applications 975: 8887.
17. Zanke, Mr Pankaj, and Dipti Sontakke. "The Impact of Business Intelligence on Organizational Performance." *Available at SSRN 4847945* (2024).

18. Shahane, Vishal. "Evolving Data Durability in Cloud Storage: A Historical Analysis and Future Directions." *Journal of Science & Technology* 1.1 (2020): 108-130.