The International Journal of Interdisciplinary Organizational Studies
ISSN: 2324-7649 (Print), ISSN: 2324-7657 (Online)
Volume 19, Issue 1, January-June, 2024

PREDICTION OF ROAD ACCIDENTS USING MACHINE AND DEEP LEARNING TECHNIQUES

Dr. K. Saraswathi
Assistant Professor(Senior Grade), Department of Computer Technology – UG, Kongu Engineering College, Perundurai, India, saraswathiteaching@gmail.com

Dr. N. T. Renukadevi
Assistant Professor(Senior Grade), Department of Computer Technology – UG, Kongu Engineering College, Perundurai, India, renuka.kec@gmail.com

K.G.Akshaya
Department of Computer Technology- UG, Kongu Engineering College, Perundurai, India
akshayaani22@gmail.com

S.Kanishka
Department of Computer Technology – UG, Kongu Engineering College, Perundurai, India
kanishkasuresh04@gmail.com

Abstract— Globally, the occurrence of traffic collisions has shown a noticeable increment within recent years. It has become a major issue due to the vast amount of auto accidents that happen annually. It is horrifying and completely unbearable for its citizens to lose their lives in auto accidents. This paper aims to examine traffic collisions that occur at the federal, state, and local levels in India. By employing machine learning and deep learning methodologies, this research conducted a comprehensive investigation into traffic accidents and determined the precise accident sites. Combining deep learning and machine learning technology for intelligent accident detection is advised as part of an all-encompassing approach to enhancing traffic safety. In order to create a worldwide safety performance function (SPF) that can project collision rates for diverse roads in different places, the research investigates the idea of applying a machine learning technique. The study investigates the synergy of ML and DL in predicting and analyzing accident locations using a broad dataset supplied from Kaggle, which includes variables ranging from geospatial coordinates to weather conditions. This study focuses on forecasting automobile accidents based on road type, including asphalt, gravel, and earth roads. The study includes machine learning approaches such as data preprocessing and model training, as well as historical accident data, road infrastructure characteristics, and environmental variables. Given the increasing number of vehicular collisions and casualties in India, the issue of road accidents is a matter of utmost importance. India faces a number of challenges in ensuring road safety. Rapid urbanization, increased vehicle density, diverse road users, and insufficient infrastructure all contribute to a high accident rate. Understanding these limitations is crucial for developing effective treatments. The age group between 30 and 59 years old holds the highest susceptibility, wherein males encounter a larger percentage of both fatalities and injuries. Traffic accidents are more prevalent in bad weather and during business hours. The research technique entails gathering and combining several statistics, such as accident severity, number of victims, latitude, longitude, vehicle count, and vehicle characteristics. In this case, it will be good to investigate the frequency of accidents so that we may utilize this information to build ways to reduce them. RESNET, Random Forest, SVM, and LSTM are therefore used to increase the precision and effectiveness of event detection.

Keywords—machine learning SVM, Random Forest, LSTM, RESNET

I. INTRODUCTION

The field of artificial intelligence, known as machine learning, aims to create models and algorithms for computers to analyze and predict data without specific programming. Machine learning uses supervised learning to find correlations between inputs and outputs, and unsupervised learning to uncover patterns in unlabelled data. Feature selection and extraction are important for identifying relevant characteristics and converting raw data. During training, supervised learning models adjust parameters based on input data and labels, while validation or test sets evaluate model performance on unseen data.[21] Overfitting occurs when a model performs well on training data but poorly on new data, and underfitting occurs when a model lacks complexity to reveal patterns in training data. Striking a balance between overfitting and underfitting is crucial for effective models. Hyperparameters, established before training, play a significant role in improving model performance. Artificial intelligence encompasses computer science, psychology, and other disciplines. AI involves creating algorithms and statistical models. Machine learning is a key part of AI and aims to replicate human brain networks. It uses mathematical models and algorithms to process and improve performance over time. This shift has revolutionized problem-solving and decision-making [23]. Machine learning drives research into deep learning and creates applications in various industries. Deep learning is based on neural networks similar to the human brain. It transforms information through nonlinear processing units. Artificial neural networks are trained using concepts from the human brain. Deep learning is characterized by its layered architecture[1].

Deep learning models can be used for unsupervised tasks without labelled data. Autoencoders, a type of unsupervised model, are used for feature learning and dimensionality reduction. Deep learning models can identify precise features without much guidance, making them effective for solving dimensionality issues. They are useful for dealing with a large number of inputs and outputs. Deep learning enables end-to-end learning, mapping input to output without human feature engineering. This is beneficial for tasks like speech recognition and natural language processing. Training deep learning models, especially large ones, can be computationally demanding. GPUs and TPUs have been used to speed up the training process. Hyperparameters greatly impact model effectiveness, so tuning is important. Deep learning automatically captures hierarchical structures from unprocessed input. The hidden layers of a deep neural network learn features at different levels, helping the model identify complex patterns and correlations[2].

Traffic-related deaths and injuries in India are a significant public health issue that is expanding. India has a diverse traffic system and many accidents occur due to various factors, including noncompliance with traffic laws. The goal of this project is to identify the factors that contribute to traffic accidents and find a practical solution to enhance road safety. India has the world's second largest road network, leading to increasing traffic congestion and a high number of fatalities and injuries from automobile crashes. Without effective measures, road traffic deaths in India could reach 250,000 by 2025. Predicting the causes of accidents and analyzing accident frequency can help develop methods to reduce them. Implementing regulations and safety measures, along with improving road and vehicle safety, can significantly decrease the frequency and severity of accidents. The primary objective of traffic accident analysis is to proactively address risks, reduce accidents, and improve road safety for public welfare and infrastructure enhancement[3].

II. LITERATURE REVIEW

The survey was made on the different existing works. Some of them are listed below.

Barbosa et al. (2020) proposed "Machine learning applied to road safety modeling: A systematic literature review". The road safety model is crucial for improving safe mobility through collision prediction. However, it has limitations and assumptions. This investigation explores alternative strategies like machine learning to build collision prediction models (CPMs) categorized into crash frequency, collision frequency and severity, and accident severity classification. Various methods are used, including nearest neighbor classification, artificial neural networks, decision trees, evolutionary algorithms, and support vector machines. The study also examines factors affecting accident likelihood, particularly road-environmental factors[1].

Chen et al. (2019) conducted a study entitled "Comparing Machine Learning and Deep Learning Methods for Real-Time Crash Prediction". The road safety model plays a substantial role in enhancing safe mobility by facilitating the creation of a collision prediction model and examining factors that affect the frequency of collisions. Nevertheless, it is crucial to comprehend the constraints associated with this approach, including specific assumptions and the need for preliminary link function design. Nonetheless, the utilization of statistical techniques in this model offers possibilities for exploring alternative strategies, such as the implementation of machine learning techniques. Even though it was more straightforward than other models, the Naive Bayes model did a great job. The study's results are especially noteworthy as they offer preliminary understanding of how well ML and DL models function[2].

Salahuddin et.al., (2023) published the article “DCOP: Deep Learning for Road Safety System”. The road safety model is crucial for predicting collisions and studying mobility factors, but has limitations and link function design. Naive Bayes offers initial insight into machine learning and deep learning models. The ultimate objective is to precisely predict the steering angle for an autonomous vehicle, thus enabling it to navigate safely and accurately. The DCOP classification approach utilized in this paper is a useful tool for investigating, discussing, testing, and assessing item detection and prediction algorithms. This categorization method can greatly increase the safety and dependability of autonomous cars in the transportation system of the future[3].

Peppes et al. (2022) published a study titled "Driving Behaviour Analysis Using Machine and Deep Learning Methods for Continuous Streams of Vehicular Data". This research paper proposes a framework for analyzing auto data using machine learning and available resources. It uses clustering to evaluate driver behavior, and deep learning to compare supervised machine learning techniques. The aim is to reduce emissions and environmental impact, while improving fuel consumption and emissions. Data from various autos is processed and assessed using clustering methodologies to identify whether or not the driver's behavior is environmentally friendly, then a comparison, using the given labeled dataset, between deep learning and supervised machine learning techniques[4].

Sattar et al. (2023) introduced a study entitled "A Comparative Study of Optimized Deep Learning-Based Road Accident Severity Prediction Models". Traffic accidents remain a significant cause of fatalities and injuries worldwide. It is crucial to accurately classify accident types and severities to expedite post-accident procedures and establish comprehensive road safety regulations. The objective of this study is to explore the use of deep learning methods in predicting the severity of injuries sustained in collisions within the Eastern Province of Saudi Arabia. The research involved the training and evaluation of four models: a multilayer perceptron, an artificial neural network with radial basis function, and a tabular data learning network. To tackle class imbalance, the training dataset was oversampled using the synthetic minority oversampling technique, and Bayesian optimization was employed to optimize the models. The utilization of the SMOTE technique led to improved accuracy, recall, and precision in the ANN utilizing RBF and TabNet models. Nonetheless, the effect of oversampling strategies on model performance was not consistently positive. By employing LASSO regularization and permutation importance analysis, significant features were identified. Once again, it should be noted that the use of oversampling strategies did not always lead to enhanced model performance based on the data. In summary, the feature "Number of Injuries Major" consistently demonstrates a critical role as a predictor in deep learning models, independent of the selection procedures applied. These findings demonstrate the significant relationship between the population with severe injuries and the degree of injuries received in incidents, as well as the discovery's possible importance in informing traffic safety regulations[5].

Hossain et al., (2021) proposed "A Comparative Study of Machine Learning Algorithms to Predict Road Accident Severity". Worldwide, road accidents are a major problem that cause injuries and fatalities in addition to a variety of direct and indirect expenses. Governments and international entities have established accident prevention mechanisms, technologies, and regulations. The utilization of extensive data and artificial intelligence holds considerable potential in the advancement of an auspicious approach aimed at forecasting or diminishing the likelihood of automobile accidents. The majority of current research looks at how road layout, environment, and weather affect traffic accidents. Human elements like as drinking, drugs, age, and gender are frequently overlooked when assessing the severity of accidents. A study was done to look at what influences and how serious an accident will be. The study examined a number of machine learning model metrics. There were two categories for severity: fatal/serious/minor/non-injury and grievous/non-grievous. In both tests, Random Forest fared better than other algorithms. Individual algorithms were less accurate than ensemble machine learning approaches. These results shed light on the causes and consequences of accidents[6].

Thakali et al. (2017) released the article "Development of a Global Road Safety Performance Function Using Deep Neural Networks". This study examines the use of a machine learning approach with the main objective of developing a comprehensive road safety performance function (SFP) that can estimate collision rates on a range of routes and locations. In pursuit of this goal, a practical deep learning framework is introduced, proposing the deep belief network (DBN) as a potential replacement for traditional regression models in crash simulation. Three real-world accident datasets covering six different highway categories—distance, number of lanes, access control, and location (rural vs. urban)—are used in this study's empirical research. The study conducts many tests that focus on network topology, training methods, data amount, and generalization capability in order to evaluate the DBN's performance in comparison to standard regression models. The obtained experimental results demonstrate that, when trained on different crash datasets, the DBN model may attain prediction performance comparable to the locally calibrated negative binomial (NB) model[7].

Muresan et al. (2018) proposed "An Improved Deep Belief Network Model for Road Safety Analyses". Anticipating crashes is a crucial part of evaluations related to traffic safety. Approaches based on regression are commonly employed to predict crashes. The calibration process underlying these approaches is typically time-consuming and requires extensive knowledge and aptitude in the subject matter; it is not easily automated. This investigation presents a fresh alternative to standard methodologies in the form of a machine learning (ML)-based strategy. Two training steps are completed by a deep neural network called a regularized subconscious belief network. The project begins with unsupervised learning and proceeds to initialize a Bayes neural network, which is then fine-tuned, using the weights from the first step. It is anticipated that this approach will require less extensive human involvement and yield higher accuracy. This study's primary objective is to show how well our new model predicts crashes. We accomplish this goal by conducting a thorough examination of two specific case studies that incorporate comprehensive collision data from an extensive 800 kilometer stretch of Highway 401, as well as other major highways located throughout the vast province of Ontario, Canada. Our study's main goal is to thoroughly examine and contrast the effectiveness and overall performance of our state-of-the-art machine learning method with those of other well-known models, such as the Bayesian neural networks (Bayesian NN), kernel regression (KR), and negative binomial (NB) model. Furthermore, we have decided to go deeper into this study project by thoroughly investigating and assessing additional critical attributes and characteristics, such as the potential impact and significance of different training data volumes and configurations[8].

Hossain et al. (2023) proposed "A study on road accident prediction and contributing factors using explainable machine learning models: analysis and performance." Around the world, car crashes claim millions of lives every year, leaving society with heavy financial and economic costs. Previous research has primarily concentrated on classifying road accidents as a predictive issue, striving to anticipate whether a traffic collision will occur in the future. However, the aforementioned investigations, thus far, have not undertaken a comprehensive analysis that delves deeply into the intricate and linked links among all of the variables that influence the occurrence of traffic accidents. While a multitude of studies have indeed evaluated and appraised the importance and relevance of various factors pertaining to road accidents and the extent of their impact on the severity of such incidents, it is regrettable to note that only a limited number of these aforementioned studies have undertaken a thorough examination of a specific subgroup of ensemble machine learning models while also focusing on a dataset that is exclusive to road accidents transpiring within the territorial confines of New Zealand (NZ). Hence, our research utilized the latest NZ road accident dataset to assess multiple machine learning models for predicting accident severity. Additionally, we used an interpretable machine learning technique to assess the importance of several parameters related to traffic incidents. A variety of machine learning models, including Extreme Gradient Boosting, Light Gradient Boosting Machine, Random Forest, Adaptive Boosting, and Categorical Boosting, were examined. The Ministry of Transport in New Zealand supplied the data analysis, which included traffic events from 2016 to 2020. A comparative analysis found that Random Forest was the most accurate classifier, with an F1-Score of 81.04%, recall of 81.42%, accuracy of 81.45%, and precision of 81.68%. Subsequently, we conducted Shapley value analysis to explain the performance of the Random Forest model on both a global and local scale. Globally, this method ranked the contributions of different features to severity classification, while locally it examined feature usage within the model. Furthermore, we examined the relationship between the target variable's attributes and prediction using the Shapley Additive ExPlanation dependency plot. Our research demonstrates how the kind of road and the number of vehicles involved in a collision influence the degree of injury. The SHAP analysis's top-ranked features are utilized to retrain and evaluate the ML models. According to the results, the DJ, AdaBoost, and Cat Boost models outperform by 6%, 5%, and 8%, respectively[9].

Elwahsh et al. (2023) suggested a work entitled A technique rooted in Deep Learning is proposed in order to enhance Road Maintenance Systems through the lens of Climate Change. Road Maintenance Systems (RMS) hold utmost importance in ensuring the safety and functionality of roads. Weather-related incidents and the consequent damage pose a significant threat to these RMS, thereby raising concerns about the influence of climate change on them. In order to confront this challenge, a proposed strategy known as RMSDC is introduced with the objective of enhancing road maintenance systems through the integration of deep learning and climate adaptation. The dataset is partitioned into training and test sets in order to facilitate the application of the multivariate classification method. The RMSDC strategy combines sensor data, weather data, and Convolutional Long Short-Term Memory (ConvLSTM) techniques. But the effects of climate change are already being seen in underdeveloped nations, where they are making road networks more susceptible to landslides, floods, and harsh weather. As a result, road networks must adjust to changing circumstances, especially in developing countries with few financial resources. We offer a novel and useful RMSDC method that leverages deep learning algorithms based on climate change estimates to address this problem. The temporal correlations between the input features are efficiently captured by the ConvLSTM block while calculating the root-mean-square deviation (RMSD). We utilize two measures, mean absolute difference and root-mean-square error, to evaluate the effectiveness of RMSDC in downscaling climatic variables in comparison to other frameworks. In empirical evaluations, RMSDC typically performs better than competing approaches, producing a root mean squared error. These numerical discoveries demonstrate the efficacy of RMSDC in tackling the challenges of road network maintenance systems, leading to proactive solutions for road upkeep that enhance environmental sustainability, generate cost savings, and enhance traffic safety[10].

Vaiyapuri et al. suggested a paper employing deep learning to forecast the severity of traffic accidents and do cognitive analyses. Road accidents occur often all around the world. India is among the nations with the highest number of accidents. Predicting collision injury severity is a fascinating topic of traffic safety research. Conventional statistical models involve established correlations and underlying assumptions, which, if broken, might result in erroneous conclusions. Many accidents occur in India as a result of numerous situations and infractions of traffic laws. The purpose of this suggested research is to discover the contributing variables or characteristics in traffic accidents. Based on the severity of the injuries, this work proposes a potential solution to the problem the proposed model aims to optimize traffic accidents, resulting in the overall improvement of the system. Furthermore, for cognitive analysis, a feature selection technique is utilized to identify the significant attributes of traffic accidents. The incorporation of feature selection enhances the accuracy percentage of algorithms[12].

Komol et al. (2021) submitted a study is being conducted which utilizes machine learning techniques to evaluate the extent of accidents that involve vulnerable road users. Fatalities in traffic accidents are a prevalent issue in the transport sector. Every year, a large number of people are killed in traffic accidents, and vulnerable road users (VRU) are particularly vulnerable. This investigation centres its attention on the utilization of categorization techniques based on machine learning to replicate the extent of harm incurred by road users who are at a higher risk, including pedestrians, cyclists, and motorcyclists. This study will specifically examine the essential variables linked with various VRU groups, including walkers, cyclists, motorcyclists, and all VRU groups. The essential determinant of accident severity between these VRU groups was analysed by detecting similarities and differences in key characteristics connected with each VRU group. Crash data from Queensland, an Australian state, was gathered from the years 2013 to 2019. The study examined accidents using supervised machine learning techniques including K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and random forest models (RF). Numerous factors were included by the models, including traffic flow, speed, road features and locations, weather, conditions for drivers and vehicles, and characteristics of other road users. Each model was trained separately and evaluated for a variety of vulnerable road user (VRU) categories. The Random Forest model showed high accuracy in testing for Motorcyclist, Cyclist, Pedestrian, and Unified VRU. This model was then used to classify and compare road crash characteristics. Moreover, the model illustrated the partial dependence of each individual and combined VRU on the severity levels of crash characteristics, highlighting how traffic accident parameters vary with the severity of VRU collision. The comparison study findings suggest that motorcyclists are more prone to involvement in serious crashes, followed by pedestrians and cyclists[11].

Wang et al. (2017) released the article "Deep Learning for Real-Time Crash Prediction on Urban Expressways". Real-time prediction of crash risk is expected to play a crucial role in modern traffic management systems, with the goal of assessing the likelihood of an accident occurring in real-time. However, due to the challenging traffic conditions in urban areas, previous research has predominantly focused on freeways rather than urban arterials. In this dissertation, a novel approach is presented for predicting real-time crash risk in vascular systems using a long short-term memory convolutional neural network (LSTM-CNN), which combines the strengths of LSTM and CNN. Specifically, CNN captures time-invariant characteristics, while LSTM handles persistent dependencies in the data. The designated region for the case study is comprised of four metropolitan arterials in Orlando, Florida, and multiple data sources, such as weather, traffic patterns, and signal timings, are included for the analysis of collision risk. Various methods of data preparation are employed, with the synthetic minority oversampling technique (SMOTE) used to address data imbalance by oversampling crash events. The LSTM-CNN model is trained and tested on separate datasets, utilizing a range of performance indicators. Additionally, five benchmark models—including Bayesian Logistics Regression, XG Boost, LSTM, CNN, and Sequential LSTM-CNN—are developed for comparative purposes. The experimental results demonstrate that the suggested LSTM-CNN model outperforms state-of-the-art techniques in terms of AUC, sensitivity, and false alarm rate. Consequently, this study highlights the promising potential of LSTM-CNN for real-time prediction of arterial crash risk[13].

Pezoa et al. (2021) proposed a study titled "A deep learning approach for real-time crash prediction using vehicle-by-vehicle data." With the goal of estimating the probability of accidents happening in real time, real-time crash risk prediction is anticipated to be a major component of sophisticated traffic management systems. However, because of the difficult traffic circumstances in these places, prior study has mostly concentrated on highways rather than urban arterials. The long short-term memory convolutional neural network (LSTM-CNN) we construct in this dissertation may be used to detect vascular accident threats in real time. Combining the best features of both methods, this model uses CNN to extract time-invariant qualities and LSTM to resolve long-term relationships in the data. We studied four urban arterials in Orlando, Florida, as a case study to evaluate the risk of accidents. This has been accomplished by utilizing a variety of data sets, including those on traffic, weather, and signal timing. In order to address the problem of data imbalance, we have also used a variety of data preparation strategies including the synthetic minority over-sampling approach (SMOTE) to oversample the crash events. Using a variety of measures, the LSTM-CNN model has been improved on training data and verified on test data. In addition, five benchmark models have been developed for comparison: Sequential LSTM-CNN, XG Boost, CNN, LSTM, and Bayesian Logistics Regression. These findings imply that LSTM-CNN may be able to forecast accident risk in real time on arterials with promising outcomes[14].

Liu et.al., (2019) recommended “spatiotemporal deep learning approach for citywide short-term crash risk prediction with multi-source data”. Through the use of several data sources, this study seeks to evaluate the effect of deep learning on the prediction of short-term accident risk in an urban setting. The procedure is shown using data from Manhattan, New York City. Data on temperature, land use, aspects of the road network, GPS data from taxis, and collision data are all included. STCL-Net, a citywide short-term crash risk assessment network, is created. We execute and compare nine prediction tasks with different grid sizes and resolutions. The findings indicate that prediction ability is improved by higher spatiotemporal resolution. STCL-Net outperforms benchmark techniques in terms of false alarm rate and prediction accuracy. These findings show that the improved spatiotemporal properties are effectively captured by the suggested spatiotemporal deep learning technique. Furthermore, the comparison tests show that machine-learning models outperform econometric models in predicting daily crash risk, while econometric methods outperform them in predicting weekly crash risk. The outcomes of this research can be of assistance to transportation safety experts when choosing suitable methodologies for various accident risk prediction tasks[15].

Lei et.al., (2017) published the article "A deep learning approach to forecast the immediate risk of traffic accidents" is the title of this research. Due to the enormous traffic accidents brought on by the rapid rise of urbanization, there have been fatalities as well as huge financial losses. The capacity to anticipate the possibility of a traffic collision is essential for proactive damage minimization and accident avoidance. However, due to human behavior, the complex traffic environment, and the unavailability of real-time traffic data, it is difficult to anticipate traffic accidents with high spatiotemporal precision. We collected a variety of traffic-related data from the same city for this study, such as air pollution, traffic flow, weather, and accidents. With this data, we created a deep learning model to predict the likelihood of traffic accidents using a recurrent neural network. By implementing a traffic accident warning system, the predicted accident risk can be effectively utilized. Utilizing a Granger causality analysis, we were able to ascertain the relative significance of each component taken into account in our model and define the predictive power hierarchy: traffic flow, traffic accidents, geographic position, weather, air quality, holidays, and time period. According to this list, the most important element influencing the likelihood of traffic accidents is traffic flow. Traffic forecasting and control organization can be improved by incorporating the suggested strategy into an intelligent traffic management system[16].

Wang et.al., (2015) proposed a paper “Large-scale transportation network congestion evolution prediction using deep learning theory”. For transportation academics and professionals who want to locate bottlenecks and reduce congestion, understanding how a particular location's traffic congestion affects a broader transportation network is essential. Conventional research uses simulation tools or mathematical equations to model the dynamics of traffic congestion. However, because of erroneous assumptions and the laborious process of parameter calibration, the majority of these approaches have their limits. Transportation data is becoming more common as Intelligent Transportation Systems (ITS) and the Internet of Things (IoT) grow, leading to a number of data-driven studies on transportation phenomena. It is often known that one of the most promising methods for managing enormous volumes of high-dimensional data is deep learning theory. The aim of this endeavor is to investigate huge transportation networks using the idea of deep learning. By using taxi GPS data, a deep Recurrent Neural Network architecture and Restricted Boltzmann

Machine are used to forecast and clarify the development of traffic congestion. A numerical analysis is currently being carried out in Ningbo, China to validate the suggested approach's effectiveness and efficiency. When using Graphic Processing Units (GPUs) in a parallel computing environment to implement the concept, the prediction accuracy can reach up to 88% in less than 6 minutes. A map-based platform could visually display anticipated patterns of congestion evolution both in terms of time and geography, enabling proactive measures to identify vulnerable linkages and reduce congestion[13].

Chang et.al., (2023) presented a paper “Enhancing travel time prediction with deep learning on chronological and retrospective time order information of big traffic data”. Smart mobility is redefining the logistics industry by applying sustainable solutions in response to increased environmental concerns and extensive usage of big data. This work introduces the bi-directional isometric-gated recurrent unit (BDIGRU), a novel deep learning approach for intelligent transportation planning. Finding viable data, choosing appropriate techniques for intelligently predicting such data, and investigating accessible procedures for prediction are some of the questions these searches aim to answer. For route planning and travel time prediction, our approach is included into a deep learning neural network architecture. Through a recursive attention mechanism based on temporal ordering, the suggested method directly learns high-level characteristics from massive amounts of traffic data. We utilize our technique to forecast stochastic journey time under different traffic situations, including congestions, by using a computer algorithm using stochastic gradient descent. Next, we calculate the best vehicle path to reduce trip time while accounting for future unpredictability. We empirically demonstrate that the proposed BDIGRU method enhances predictive accuracy and determines optimal vehicle routes compared to conventional methods[18].

Kumar et.al., (2017) published the article "Road traffic accidents in India: issues and challenges”. The study aims to analyze traffic incidents in India at different levels. Previous studies have demonstrated that the distribution of road fatalities and injuries in India is influenced by age, gender, month, and time. It has been observed that men are more prone to accidents between the ages of 30 and 59 compared to women, and this age group is the most susceptible among all age categories. Furthermore, there is a higher occurrence of road accidents during unfavourable weather conditions and regular working hours. By analyzing various road accident scenarios at the state and city levels, significant disparities in the probability of fatalities are identified between states and municipalities. To be specific, 16 out of the 35 states and union territories in India exhibit higher mortality rates compared to the rest of the country. In comparison to their rural equivalents, urban regions see significantly fewer traffic accidents, yet over half of them have higher death rates. Road safety conditions in India continue to deteriorate, despite advances seen in many industrialized and emerging nations, like China. It is projected that by 2025, there would be 250,000 road traffic fatalities if more efforts and new initiatives are not implemented. Consequently, the alarming increase in road deaths and injuries in India must be recognized and addressed[19].

Singh et.al., (2012) propose a paper on “The Neglected Epidemic: Road Traffic Crashes in India”. In India, the issue of road traffic fatalities and injuries is a significant and escalating matter of public health. On a weekly basis, approximately 2,500 individuals lose their lives, while an additional 9,000 sustain injuries as a result of traffic accidents. With about 127,000 deaths each year, India now has the unenviable distinction of having the largest number of road fatalities in the world, surpassing China. While many wealthy and growing countries, including China, are generally improving, India's situation is deteriorating. Traffic accidents now claim the lives of five times more individuals than they did thirty years ago. In the absence of escalated efforts and novel initiatives, it is projected that India's overall number of road traffic fatalities will reach approximately 250,000 by 2025. The increasing number of traffic fatalities and injuries thus makes it vital that we recognize and address this issue. There are various measures to lessen traffic accidents, like implementing speed and alcohol limits, enforcing seat belt and helmet usage, and improving road safety.. This is an urgent call to action for the federal, state, and local governments. You can save hundreds of lives by acting right now.[20]

III. METHODOLOGY

A. Support Vector Machine (SVM)

Regression and classification issues are handled using the supervised machine learning method known as Support Vector Machine (SVM). In order to carry out a classification test, SVM employs the creation of a hyperplane that effectively separates all data points belonging to one category from those belonging to the other category. It is important to note that in certain scenarios, there may exist multiple hyperplanes capable of achieving this segregation.

1) Pros of Support Vector Machine:

- When there is a substantial margin of class dissociation, support vector machines perform similarly.
- Performs well in multidimensional contexts.
- The system works best with multiple dimensions and instances.

2) Cons of Support Vector Machine:

- Large data sets render the support vector machine approach ineffective.
- Performs poorly when the data set contains a lot of noise and the target categories overlap.
- When there are more attributes per data point than training data specimens, the support vector machine may not perform properly.

3) Applications of Support Vector Machine: Some of the applications of SVM are,

a) Face observation: It is used to recognize faces based on the model and classifier.

b) Text and hypertext organisation: In this case, the classification strategy is utilized to locate key information, or more precisely, information required for text organization.

c) Grouping of portrayals: It is also used in gathering portrayals to compare the content and then take necessary action.

d) Protein fold and remote homology spotting: It is used to classify amino acid sequences into functional and structural groupings. It is a bioinformatics issue.

Choosing the right classifier is critical, for instance,

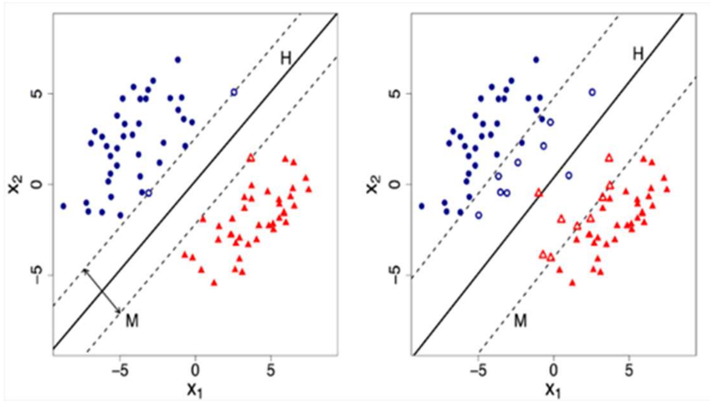

Fig. 1. Overview of Support Vector Machine

The example above contains two hyperplanes: HP1 and HP2. The majority of you would argue that HP2's margin makes it the best, but this is incorrect because each data point should be on the correct side of the margin. HP1 is the best classifier in this situation.

The supporting vectors are the points that are precisely on the margins, as determined by the SVM, which aims to discover the one that best separates the two categories by maximising the distance to points in either category[4].

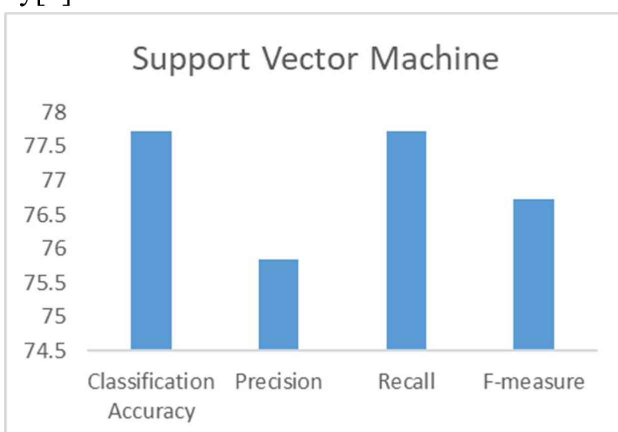

Fig. 2. The Support Vector Machine's visual depiction includes recall, precision, accuracy and F-measure

B. Random Forest

A variety of machine learning tasks, including regression and classification, may be carried out by the Random Forest method, which is widely utilized in the field. The basis of this methodology lies in the concept of ensemble learning, which entails the amalgamation of multiple classifiers for the purpose of addressing intricate problems and augmenting the model's effectiveness. The classifier based on Random Forests uses a variety of decision trees that are employed on diverse subsets of a provided dataset, thereby deriving the average results and ultimately enhancing the anticipated accuracy of the dataset [22]. Because trees in random forests grow concurrently with no interaction, the supervised learning algorithm and bagging method known as "random forest" are employed in machine learning to achieve regression. Random Forest models offer valuable insights into the significance of features, thereby assisting in the comprehension of the degree to which each feature contributes significantly to the predictions. The interpretability of individual decision trees within the Random Forest ensemble is not as straightforward as that of simpler models, however, the metrics that measure the aggregate feature importance act as a vital tool in acquiring a more profound comprehension of the underlying characteristics of the dataset.

1) Reason for Using the Random Forest

- Training is faster compared to other methods.
- Generates accurate forecasts with massive data sets.
- Maintains accuracy, even when the majority of the data is missing.

2) Features of a Random Forest

- It operates efficiently on massive data sets.
- Efficiently addresses a missing data.
- An accurate prediction can be made without altering the hyperparameter.

3) Pros of Random Forest

- Random forest is a highly dependable learning system. It creates a highly accurate classifier for a wide range of data sets.
- The generalization error is measured objectively as the forest grows.
- Effective estimate technique maintains accuracy even with missing data.

4) Cons of Random Forest

- Random forests have been shown to overfit in noisy classification or regression tasks on certain datasets.
- Random forests prioritize higher-level features in data with categorical variables of varied levels. As a result, the random forest's variable significance ratings are untrustworthy for this type of information.
- Unbalanced datasets, where one class has a substantial quantitative advantage over the others, might provide problems for Random Forests[5].

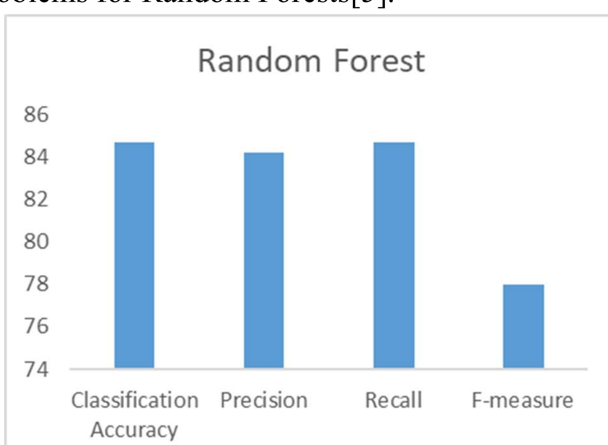

Fig. 3. The Random Forest visual depiction includes recall, precision, accuracy and F-measure

C. Long Short Term Memory(LSTM)

LSTMs, a kind of recurrent neural network, may be used to learn about Long Short-Term Memory dependencies and a wider context. Its strong insensitivity to gap length distinguishes it from other RNNs, embedded Markov models, and sequential learning techniques. It may be used for data classification, processing, and forecasting based on time series in a variety of applications, such as robotic control, handwriting identification, voice recognition, machine translation, video games, and healthcare. Four parts make up a typical long short-term memory unit: an input gate, an output gate, a forget gate, and a cell.

1) Types of LSTM:

a) Univariate LSTM Models: LSTMs can be used to simulate the problems of forecasting univariate time series. These are problems with a single series of data, hence a model is required to learn from prior observations and forecast the next value in the sequence.

b) Multivariate LSTM Models: A multivariate LSTM is a high-performance deep learning model that can handle several input features and capture their dependencies. It performs better than standard LSTMs in terms of flexibility, durability, and accuracy.

c) Multi-Step LSTM Models: Long short-term memory recurrent neural networks possess the ability to manage numerous variables, automatically acquire features from sequential data, and generate sequences of varying lengths suitable for multi-step forecasting, as opposed to other machine learning techniques.

d) Multivariate Multi-Step LSTM Models: The multivariate LSTM model uses multiple input features and predicts a large number of future steps. This paper presents a multivariate LSTM model with numerous phases for forecasting flood runoff in the coming week.

2) Applications

- Robotics management
- Time series forecasting
- Speech identification
- Learning rhythms
- Hydrological rainfall-runoff modelling
- Composing music
- Learning grammar
- Recognizing handwriting
- Recognizing human actions
- Translation from sign language to many languages
- Prediction of protein subcellular localization
- Time series anomaly detection
- Drug design
- Airport passenger management
- Short-term traffic forecast

3) Structure of LSTM

LSTM consists of four neural networks and multiple memory cells organized in a sequential configuration. An LSTM unit is comprised of a cell, an input gate, an output gate, and a forget gate. Three gates control the information flow into and out of the cell, and it stores values for arbitrarily long periods of time. The LSTM approach performs well for the classification, analysis, and forecasting of time series with arbitrary periods.

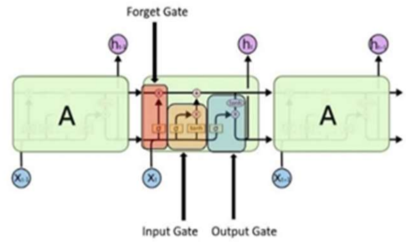

Fig. 4. Structure of LSTM

4) Pros of LSTM

Handling Long Sequences: LSTMs exhibit efficacy in handling data sequences with protracted interdependencies. They possess the ability to retrieve information from preceding time steps over an extended duration, hence making them useful in fields like time series analysis and natural language processing (NLP).

a) Preventing the Vanishing Gradient Issue: LSTMs offer a solution to the issue of the vanishing gradient, a common occurrence when deep networks, specifically RNNs, are being trained. The gating functionalities of LSTMs, including the forget gate, enable them to control the movement of input and gradients throughout the network, thereby preventing gradients from becoming excessively small during the training process.

b) Managing Sequences Variable-Length: By constantly changing its internal state, LSTMs are able to accommodate input sequences of varying length. In several real-world situations when the duration of the receiving data fluctuates, this is helpful.

c) The Memory Cell: A memory cell of an LSTM may store and retrieve information for extended periods of time. LSTMs are ideal for activities requiring prior context memory because of this memory cell's ability to retain important information while removing unnecessary information.

d) Controlled Gradient Flow: LSTMs offer ways to modulate gradient flow during backpropagation. For example, if gradients must be propagated back in time, the forget gate can keep them from disappearing. As a result, LSTMs can successfully collect information from previous time steps.

5) Cons of LSTM

a) Computational Complexity: LSTMs have a higher computational cost than other neural network topologies such as feedforward networks and basic RNNs. Training LSTMs can be time- and resource-intensive.

b) Overfitting: When there is insufficient training data, LSTMs and other deep learning models are prone to overfitting. Regularization techniques such as dropout can help to reduce this issue.

c) Hyperparameter Tuning: The number, learning rate, and length of the sequence are examples of LSTM hyperparameters. It might be challenging and lengthy to identify the best set of hyperparameters for a specific situation.

d) Limited Interpretability: LSTMs are known as "black-box" models since it is difficult to specify how they make decisions. This can be problematic in instances when interpretability is important[6].

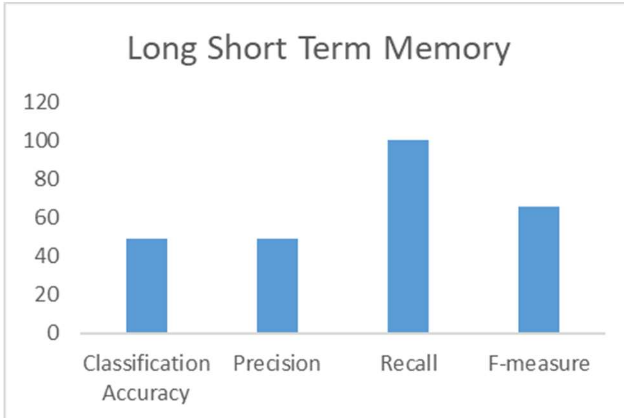

Fig. 5. The graphic representation of Long Short Term Memory with accuracy, precision, recall and F-measure

D. Residual Network(RESNET)

The term ResNet refers to a residual network. ResNet is an artificial neural network that included an identity shortcut connection, allowing the model to bypass one or more layers. ResNet allows you to build extremely deep neural networks that can increase image identification performance. This method allows you to train the network on hundreds of levels without compromising performance. To overcome the vanishing/exploding gradient issue, this architecture introduces Residual Blocks. In this network, we employ a technique called skip connections. This strategy enables you to train the network on hundreds of levels while maintaining performance. This architectural design employs Residual Blocks to tackle the problem of vanishing/exploding gradient. Within this network, we implement a technique referred to as skip connections.

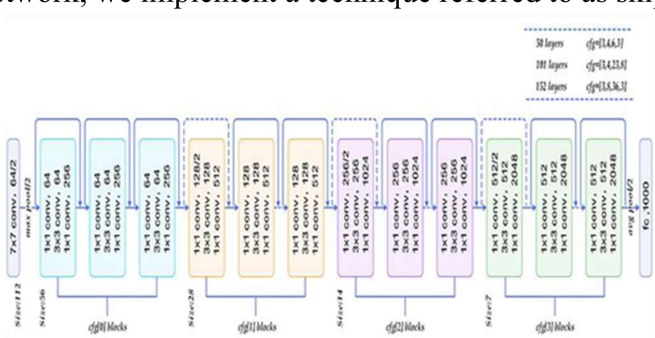

Fig. 6. Architecture of ResNet

1) Pros of RESNET

• ResNet enhances deep neural network performance by increasing neuronal layers and minimizing error rates.

2) Cons of RESNET

• Architecture has become more sophisticated.

• Implement Batch normalization layers as ResNet relies heavily on them.

Numerous computer vision applications, such as the categorization of images, the identification of objects, the delineation of semantic boundaries, and the generation of visual content, have extensively employed ResNet and its various iterations. The adaptability and efficacy displayed by ResNet have elevated it to the status of a cornerstone within the realm of deep learning for computer vision, with pre-trained ResNet models frequently utilized either as feature extractors or fine-tuned for specific tasks.

The overarching impact of ResNet, stemming from its groundbreaking advancements in residual learning and the integration of skip connections, cannot be overstated. Technological advancements have dramatically boosted the capabilities of deep neural networks in computer vision, reaching previously unattainable heights. This has firmly established its role as an essential framework in both current deep learning research and its real-world implementations[7].

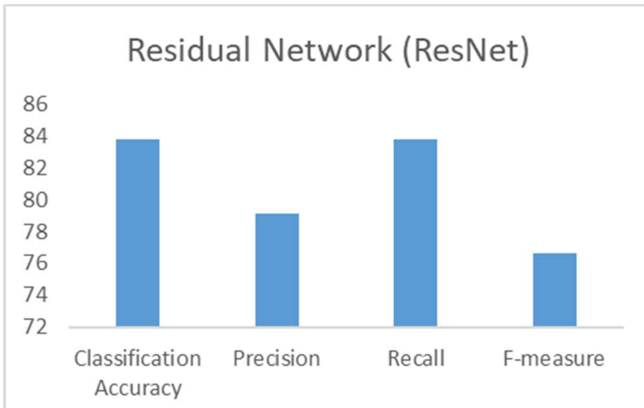

Fig. 7. The graphical representation of Residual Network results

E. Hybridization of Random Forest and LSTM

The merging of Random Forest (RF) and Long Short-Term Memory (LSTM) networks necessitates a meticulous approach in order to fully exploit their respective advantages and attain road safety by means of hybridization. The process commences with a comprehensive preprocessing of the data, wherein historical accident data, weather conditions, and traffic patterns are formatted in a manner that is suitable for the algorithms employed. Subsequently, feature engineering is executed, whereby significant features are extracted and engineered to encompass crucial predictors of road safety, including both temporal and categorical variables. Following this, model training ensues, involving the training of a Random Forest model to capture the intricacies associated with static features, while simultaneously training an LSTM model to effectively comprehend the temporal dependencies present in sequential data, such as time-series accident patterns. The process of hybridization then proceeds to integrate the predictions derived from both models, thereby leveraging the predictive capabilities of RF for static features, whilst also harnessing the capacity of LSTM to accurately capture temporal dynamics. This integration facilitates a more comprehensive comprehension of road safety conditions.

Upon completion of the training of the hybrid model, the evaluation and validation stages assume paramount importance, wherein performance metrics, including accuracy and precision, are employed to assess its predictive capabilities. The robustness and generalizability of the hybrid model across a diverse range of road safety scenarios are ensured through the implementation of rigorous testing, which incorporates cross-validation techniques. Continuous monitoring and the utilization of iterative improvement mechanisms, such as hyperparameter tuning and feature selection, serve to further refine the accuracy and reliability of the hybrid model. Through this iterative process, the hybridization of RF and LSTM for road safety analysis not only optimizes predictive accuracy, but also augments the model's adaptability to the real-world challenges encountered in the realm of road safety, ultimately leading to the creation of safer road environments[8].

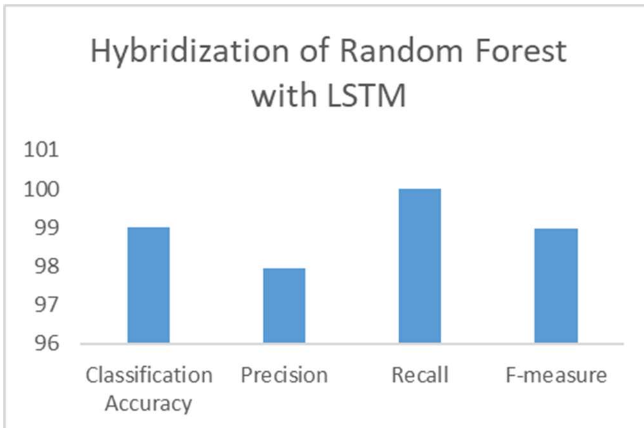

Fig. 8. The graphic representation of Random Forest and LSTM with accuracy, precision, recall and F-measure

TABLE I. RESULTS OF ALL CLASSIFICATION ACCURACY

Technique	Classification Accuracy
SVM	0.7772
Random Forest	0.8466
LSTM	0.4880
RESNET	0.8381
Hybridization of Random Forest and LSTM	0.99

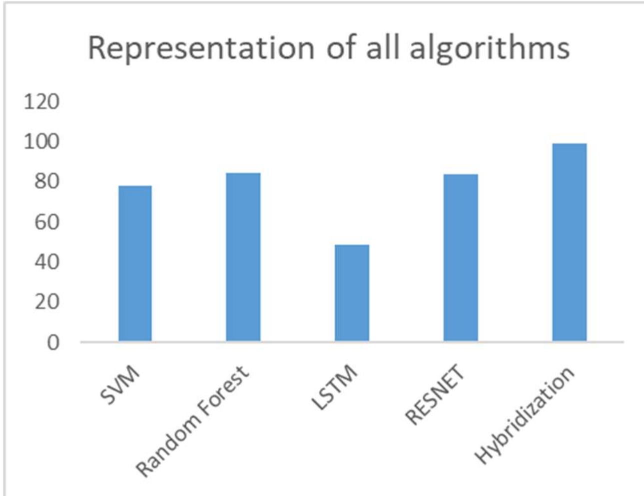

Fig. 9. Representation of all Classification Accuracy

IV. RESULT

In this section, we compute the predictions of all models on the training dataset. The data is categorized using a machine learning method. It divided the original dataset into multiple training and testing subsets. The objective of this research is to estimate how catastrophic an accident can be. This is very crucial so that emergency services and traffic authorities can swiftly deliver the proper support.

The dataset used for this study provides information on 12316 people who were victims of an accident. Random Forest, in contrast to other algorithms, has shown favorable outcomes. With an accuracy of 99%, the hybridization of Random Forest and LSTM delivered the best results[9].

V. CONCLUSION

As machine learning and deep learning are used to traffic accident prediction, traffic management optimization, and proactive preventive measures, there is potential for improving emergency response systems. By identifying high-risk regions and times, authorities can more effectively allocate resources, execute targeted safety measures, and improve overall road safety. By combining the RESNET and LSTM algorithms, we reached the highest accuracy of 0.99. To address specific road safety concerns, solutions are being implemented at the national, state, and metropolitan city levels. Increased political will and commitment on the federal, state, and local levels to improve and secure road travel for Indian motorists. Additional research and analysis will be conducted to identify significant road safety issues and develop effective life-saving remedies. More effort and innovative ideas are required to improve road safety in India. The overall number of road traffic deaths in India is expected to exceed 250,000 by 2025. There are proposed strategies for infrastructure management, standardization, policymaking, and law and regulation enforcement. In India, it is recommended to implement road safety management. Personnel and posting levels have been improved. Hospital resources are being negotiated based on surveillance data. The study will look at socioeconomic factors to determine how TI-related mortality affects people. Future research might focus on the impact of road communication on public health concerns[10].

REFERENCES

[1] Silva, Philippe Barbosa, Michelle Andrade, and Sara Ferreira. "Machine learning applied to road safety modeling: A systematic literature review." Journal of traffic and transportation engineering (English edition) 7, no. 6 (2020): 775-790.

[2] Theofilatos, Athanasios, Cong Chen, and Constantinos Antoniou. "Comparing machine learning and deep learning methods for real-time crash prediction." Transportation research record 2673, no. 8 (2019): 169-178.

[3] Hyoju, Binod, AB Emran Salahuddin, Haneen Heyasat, Omar Hisham Rasheed Al-Sadoon, and Ahmed Dawoud. "DCOP: Deep Learning for Road Safety System." In International Conference on Interactive Collaborative Robotics, pp. 93-104. Cham: Springer Nature Switzerland, 2023.

[4] Alexakis, Theodoros, Nikolaos Peppes, Konstantinos Demestichas, and Evgenia Adamopoulou. "A distributed big data analytics architecture for vehicle sensor data." Sensors 23, no. 1 (2022): 357.

[5] Hijazi, Hussam, Karim Sattar, Hassan M. Al-Ahmadi, and Sami El-Ferik. "Comparative Study for Optimized Deep Learning-Based Road Accidents Severity Prediction Models." Arabian Journal for Science and Engineering (2023): 1-21.

[6] Ahmed, Shakil, Md Akbar Hossain, Md Mafijul Islam Bhuiyan, and Sayan Kumar Ray. "A comparative study of machine learning algorithms to predict road accident severity." In 2021 20th International Conference on Ubiquitous Computing and Communications (IUCC/CIT/DSCI/SmartCNS), pp. 390-397. IEEE, 2021.

[7] Pan, Guangyuan, Liping Fu, and Lalita Thakali. "Development of a global road safety performance function using deep neural networks." International journal of transportation science and technology 6, no. 3 (2017): 159-173.

[8] Pan, Guangyuan, Liping Fu, Lalita Thakali, Matthew Muresan, and Ming Yu. "An improved deep belief network model for road safety analyses." arXiv preprint arXiv:1812.07410 (2018).

[9] Ahmed, Shakil, Md Akbar Hossain, Sayan Kumar Ray, Md Mafijul Islam Bhuiyan, and Saifur Rahman Sabuj. "A study on road accident prediction and contributing factors using explainable machine learning models: analysis and performance." Transportation research interdisciplinary perspectives 19 (2023): 100814

[10] Elwahsh, Haitham, Alaa Allakany, Maazen Alsabaan, Mohamed I. Ibrahem, and Engy El-Shafeiy. "A Deep Learning Technique to Improve Road Maintenance Systems Based on Climate Change." Applied Sciences 13, no. 15 (2023): 8899.

[11] Komol, Md Mostafizur Rahman, Md Mahmudul Hasan, Mohammed Elhenawy, Shamsunnahar Yasmin, Mahmoud Masoud, and Andry Rakotonirainy. "Crash severity analysis of vulnerable road users using machine learning." PLoS one 16, no. 8 (2021): e0255828.

[12] Vaiyapuri, Thavavel, and Meenu Gupta. "Traffic accident severity prediction and cognitive analysis using deep learning." Soft Computing (2021): 1-13.

[13] Yang, K., X. Wang, M. A. Quddus, and R. Yu. "Deep learning for real-time crash prediction on urban expressways using highly imbalanced big data." J. Transp. Res. Board (2017).

[14] Basso, Franco, Raúl Pezoa, Mauricio Varas, and Matías Villalobos. "A deep learning approach for real-time crash prediction using vehicle-by-vehicle data." Accident Analysis & Prevention 162 (2021): 106409.

[15] Bao, Jie, Pan Liu, and Satish V. Ukkusuri. "A spatiotemporal deep learning approach for citywide short-term crash risk prediction with multi-source data." Accident Analysis & Prevention 122 (2019): 239-254.

[16] Ren, Honglei, You Song, JingXin Liu, Yucheng Hu, and Jinzhi Lei. "A deep learning approach to the prediction of short-term traffic accident risk." arXiv preprint arXiv:1710.09543 (2017).

[17] Ma, Xiaolei, Haiyang Yu, Yunpeng Wang, and Yinhai Wang. "Large-scale transportation network congestion evolution prediction using deep learning theory." PloS one 10, no. 3 (2015): e0119044.

[18] Chen, Claire YT, Edward W. Sun, Ming-Feng Chang, and Yi-Bing Lin. "Enhancing travel time prediction with deep learning on chronological and retrospective time order information of big traffic data." Annals of Operations Research (2023): 1-34.

[19] Singh, Sanjay Kumar. "Road traffic accidents in India: issues and challenges." Transportation research procedia 25 (2017): 4708-4719.

[20] Singh, Sanjay K. "The neglected epidemic: road traffic crashes in India." Metamorphosis 11, no. 2 (2012): 27-49.

[21] Saraswathi Kuppusamy, Renukadevi Thangavel , "Deep Non-linear and Unbiased Deep Decisive Pooling Learning–Based Opinion Mining of Customer Review" , January 2023, Cognitive Computation 15(2):1-13.

[22] M N Kavitha , S S Saranya , E Dhinesh , L Sabarish , A Gokulkrishnan, "Hybrid ML Classifier for Loan Prediction System", Proceedings International Conference on Sustainable Computing and Data Communication Systems (ICSCDS - 2023), 25 April 2023

[23] Hemalatha, D., Poorani, S.,"Machine learning techniques for heart disease prediction", Journal of Cardiovascular Disease Research, Vol.12, Issue 1, 2021.