Evaluation of Intelligent Transportation System (ITS) for Reducing Traffic Accidents

Alok Rarotiya

Assistant Professor, Department of Civil Engineering, Medi-Caps University, Indore, Madhya Pradesh, India, alok.rarotiya@medicaps.ac.in

Traffic congestion and accidents remain critical challenges in urban transportation systems worldwide. This study proposes an intelligent framework for traffic flow prediction and accident reduction using a combination of Internet of Things (IoT) technology, machine learning algorithms, and feature selection techniques within Intelligent Transportation Systems (ITS). Real-time sensor data from IoT devices, combined with historical traffic statistics, forms the basis for predictive modeling. Data preprocessing and Particle Swarm Optimization (PSO) are employed to enhance the quality and relevance of input data, selecting key features for improved model accuracy. The classification models, built using Linear Regression, Multilayer Perceptron (MLP), and Extreme Gradient Boosting (XGBoost), achieve high precision, with XGBoost demonstrating superior performance with an accuracy of 98%, an MSE of 0.0021, and an RMSE of 0.007. These results validate the effectiveness of the proposed approach in forecasting traffic patterns, enabling informed decision-making for traffic management, infrastructure planning, and accident prevention. This research highlights the potential of integrating machine learning with IoT for developing sustainable and efficient ITS solutions.
Keywords: Traffic, Transportation, Machine learning, Error.

1. Introduction

Traffic congestion is a state in which vehicular movement is marked by diminished speeds, heightened vehicle queues, and extended journey durations. Traffic congestion has substantially increased since the 1950s. The main source of traffic congestion is the overwhelming quantity of private vehicles, exacerbated by insufficient public transit options. The lack of public transport in the modern period has resulted in a notable rise in traffic congestion. The inadequacy of public transport is a primary and direct cause of traffic congestion. Insufficient infrastructure is a substantial obstacle to properly mitigating traffic congestion. The restricted availability of public transportation, including buses and railroads, raises issues over travel quality, especially in developing countries. Severe traffic congestion may sometimes lead to incidents with modest repercussions, such as vehicle damage, injuries to drivers or passengers, or even deaths [1]. The first stage in traffic prediction is the collecting of data on current traffic conditions. The acquired traffic data is used to provide combinations of demographic, occupational, transit rate, and trip cost statistics. The objective of transportation forecasting is to predict future utilisation patterns of a certain mode of transportation by people or vehicles. Advanced approaches permit the creation of innovative ways to improve the accuracy and dependability of existing estimates, while also enabling the collection of dynamic data, extensive datasets, and diverse data kinds. Traffic predictions are extensively used in several domains, including public transit, engineering, and planning, especially for assessing infrastructure capacity. To develop a traffic model for the existing dynamic circumstances, traffic data is amalgamated with other established datasets, such as demographic information, meteorological data, economic indicators, travel statistics, and other relevant factors, according to particular analytical needs [2]. Historically, traffic volume prediction mostly depended on data collected from sensors strategically positioned along highways. A selection of machine learning methods may be used to predict future traffic patterns on roadways. This is accomplished by using current traffic data acquired from photoelectric and piezoelectric sensors, in conjunction with historical traffic data from alternate routes. Future traffic forecasts are essential for the development, execution, and improvement of transportation systems. It offers benefits to the management of operations and controls. Accurate economic analysis is crucial; nevertheless, traffic forecasting serves several other purposes, including system design, planning, air quality evaluation, and pollution reduction. The elevated expenses associated with overdesign and under design may be ascribed to several factors, including the intrinsic flaws in traffic flow forecasting and safety assessments. Traffic forecasting is an essential and foundational component of Intelligent Transportation Systems (ITS). It is essential to provide transportation-related information services, including traffic management and guidance. A mining methodology grounded on experimental analysis was developed to improve the accuracy of traffic forecasts in real-time datasets. The road transport network comprises junctions. Relocating it is a considerable challenge. The interdependence of one-way and two-way traffic flows, together with the timing of traffic signals, is apparent [3]. ITS data comprise a plethora of information about traffic movement, including various sorts of traffic. The road environment data encompasses several components, including the road topology network, road surface data capabilities, crucial speed limit signs, and driver information. We must develop a strategy to artificially include data related to critical situations, climatic conditions, public transport, and certain road environments that are now lacking in the system. Traffic predictions are generated using mathematical data samples, primary data from the transportation system, and their chronological order [4]. Daily congestion squanders individuals' time and incites their dissatisfaction. The government concurrently faces significant challenges from many accidents owing to its obligation to maintain clean roads until destinations are reached, mitigate pollution throughout various locations, and ensure pedestrian safety. Congestion, a global concern, affects several aspects of life. All drivers who have experienced a traffic collision on a roadway know that the primary factor contributing to these incidents is universally accepted. Unexpected interruptions may occur due to traffic complications, including collisions and road layouts. Insufficient signal time allocation intensifies the limitations of small inner roadways, resulting in reduced traffic speeds. The insufficient number of road lanes is the principal factor contributing to the global increase in traffic accidents, regardless of specific conditions. Although total eradication of traffic may be unattainable, forecasting methods exist that may provide time saves, cost reductions, and enhanced road safety. A goal of a sustainable and intelligent transportation system is to minimise the overall incidence of traffic accidents. One method to do this is by using high-quality traffic data [5]. The traffic data used for real-time and historical flow prediction is being enhanced by an expanding variety of sensor sources, including as inductive loops, radar systems, wireless global positioning systems, and social media. Data assumes an increasingly vital role in modern transportation management and control. This requires the investigation of methods for using deep learning systems and large datasets to predict traffic patterns. Non-parametric approaches include K-nearest neighbours, Support Vector Regression, and Artificial Neural Networks. Hybrid methodologies use a combination of parametric and non-parametric methods. It is impractical to advocate one strategy over another because of their suitability for various environments and datasets. Examples of the use of the Internet of Everything include intelligent automotive driving systems, smart cities, intelligent activity control, and climate prediction applications [6]. Inventions related to the Internet of Everything, including computerised sensor devices, intelligent networked accessories, industrial monitoring equipment, and various motorised systems, may be classified into several categories. A wireless sensor network (WSN) is a self-sufficient distributed system composed of autonomous devices. These gadgets use sensors to assess environmental or physical parameters. These autonomous nodes, when integrated with switches and a gateway, form a wireless sensor network (WSN). Sensor systems facilitate the social flow of information necessary for intelligent environments. This applies universally across many environments, including but not limited to structures, utilities, enterprises, residences, vessels, and the automation of transportation networks. To implement late countermeasures against psychological oppressors and conduct guerilla warfare, it is essential to possess suitable sensor systems that can be expanded and adjusted to coordinate capabilities. In circumstances such as these, the installation of wire or cables is often unfeasible. We want a sensor network that is uncomplicated in terms of installation and maintenance. The passage aims to facilitate communication between Wireless Sensor Networks (WSNs) and open communication systems throughout the whole internet. The portal for this technology comprises a centralised control unit, a database (DB), a WSN module, a WLAN access point, and a GSM module. Establishing a connection Segments of the Portal Suitable estimate hubs transmit data via wireless connections to a central portal, which subsequently offers internet access for data gathering, processing, analysis, and display. Enhance the network's isolation and reliability by using switches to provide an extra communication link between the end hubs and the entrance point. This may be achieved by establishing the connection between the two places. This text references a WSN module, a WLAN access point, and a GSM module as examples of ECMs. The differences between a VANET and a MANET are examined comprehensively. Vehicle-to-vehicle (V2V) and vehicle-to-road (V2R) communications provide the basis of the flag passing mechanism in vehicular ad hoc networks (VANET) [7]. Nevertheless, the communication speeds of the autos must be increased. The Internet of Vehicles (IoV) is differentiated from the Internet of

Things (IoT) to address this challenge and enhance the most beneficial attributes of Vehicular Ad Hoc Networks (VANETs). Astute researchers have identified that the IoV utilises solutions founded on the IoT. These systems are used throughout several administrations, including in vehicles. Others have suggested using a Markov Decision Process (MDP)-based algorithm inside the Internet of Vehicles (IoV) architecture to rectify these limitations and facilitate enhancements. A physiologically based intelligence system was developed to address the challenges of phantom scarcity and high flexibility. Numerous organisations specialising in cloud computing provide various implementation services for the IoV architecture. Experts have acknowledged the gaming methodology for elucidating price competition inside the IoV framework. This is because the ultimate objective is to determine suitable cloud management. In discussing the Internet of Everything (IoE) [9], two primary areas must be considered: The Source and the Host. The source is tasked with gathering data from the periphery and storing it locally, whilst the host is responsible for uploading the data to the cloud. The elements of the Internet of Everything are shown in Fig. 1.

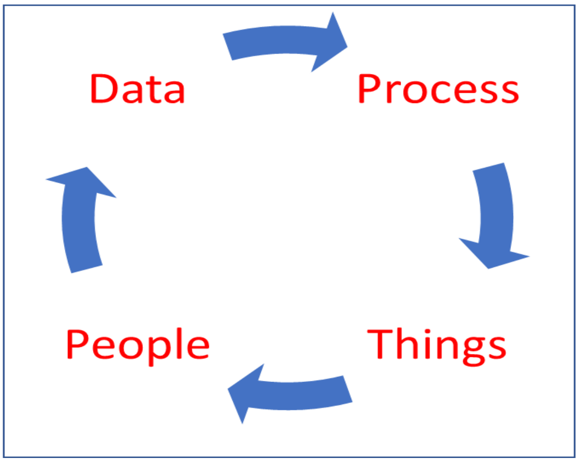

Fig. 1 Components of analysis

- People in system

In the context of the "internet of things," the term "community" denotes the collective of persons engaged in the acquisition and dissemination of information from one person to another. In recent years, there has been significant interest in information exchange and connection building via online platforms. Individuals interact and share information using a diverse array of technology in the contemporary world. The vast majority of individuals were engaged participants on at least one social networking platform. The rapid advancement of internet technology across several domains has enabled a growing number of individuals to interact with each other in significant ways.

- Things

This role is significant within the realm of the Internet of Things, since it has the capacity to collect vital data from a diverse array of internet-connected gadgets and appliances. The data collected by these devices is often used to tackle urgent issues in the customer's vicinity. In anticipation of a worst-case scenario, such as a significant deterioration in an individual's health, pertinent information is sent to specialists or physicians. The data collected by the sensor is then sent, stored, and evaluated for potential use during an emergency.

- Data available in system

The acquisition of pertinent data from individuals and other organisations often use devices connected to the Internet. Typically, internet-connected devices collect data and then render that data available to the public. All data will be securely stored on the sensor server while processing by the analysts. To make optimal decisions, it is essential to examine the data collected from various physical equipment that observe the natural environment. Assessing a patient's average glucose level, including both elevated and diminished readings, enables the social insurance office to ascertain the patient's comprehensive glucose status.

- Processes of system

Verification of data, connections, and membership necessitates the implementation of procedures, which are critically significant. Furthermore, there are methodologies used to extract pertinent information from the interconnected presence of data, people, and objects inside a network. Disseminating pertinent data and enhancing the efficacy of the IoE system both rely on precise methodologies and connections, respectively. The processes use advanced enhancements in remote system technology to provide essential information to the intended recipient. In contemporary Internet of Everything frameworks, 5G mobile technologies are widely used to transmit data activated by web-connected physical objects. An effectively organised network of supplementary elements, data, and people enables firms to get significant insights from advancements in the Internet of Everything (IoE). This is facilitated by the Internet of Everything. Recently, there has been an increase in the use of online communities and wearable technology to promote human service promises to prospective customers. This article delineates a strategy for estimating traffic flows that is based on the Internet of Things, machine learning, and feature selection. Real-time sensor data is acquired by IoT devices deployed on roadways or inside automobiles. The dataset comprises real-time data gathered via the Internet of Things and historical traffic statistics. The input dataset is kept in a centralised cloud. The data undergoes preprocessing to eliminate noise and identify outliers. Feature selection is essential for enhancing accuracy and reducing the Root Mean Square Error. The experimental study used the UCI traffic dataset to apply linear regression, MLP, and XGBoost algorithms. The accuracy of traffic flow prediction using XGBoost is 99%, with a Mean Squared Error (MSE) of 0.0021 and a Root Mean Squared Error (RMSE) of 0.007. Yachao Jia et al. established a mining network using big data approaches to assess urban Intelligent Transportation Systems (ITS) frameworks. The notion of flexible traffic data, driven by the use of big data, significantly influences the trajectory of transportation development. Big data mining ideas include several technologies, such as improved efficiency, distribution predictability, and real-time functionalities. Real-time model monitoring and the development of predictive systems may concurrently gather traffic data to guarantee safe driving and improve the processing capacity of the traffic system [8]. The emergence of extensive data has catalysed a substantial change in technology. Saad and his associates performed a comparative examination of several data mining approaches used for traffic forecasting. Smart cities are significantly reliant on Wireless Sensor Network (WSN) principles. Data mining technologies are used to aid in the collection of data from sensors. This study aims to improve decision-making via the observation of traffic characteristics. The paper offers a detailed elucidation of the use of data mining methods in sustaining the operational efficacy of traffic lights. Clustering methods are used to enhance the utilisation of sensors in the signage [9]. At now, deep neural networks (DNNs) are used for predicting traffic patterns. Yachao et al. provide an innovative technique that amalgamates a CNN-LSTM framework. The LSTM technique is applied to portray spatial relationships, whereas the CNN approach is employed to characterise temporal relationships. Additional features of the data include the weather conditions during the prior time. The system proves its effectiveness by using traffic data associated with certain geographical regions and time periods. The data indicate that using the RF approach for feature selection results in a 90% enhancement in accuracy [10]. Traffic congestion is a common problem seen during summer vacation. XianglongLuo et al. provide an innovative predictive model that integrates Support Vector Regression (SVR) with Density Functional Theory (DFT). In this respect, we use historical data to provide projections by assessing traffic statistics acquired by the DFT methodology and a specified threshold value. A Support Vector Regression (SVR) method is proposed for the purpose of forecasting residual sequences. The methodology was assessed using data gathered from toll gates in Jiangsu Province, China. In contrast to the conventional prediction technique, our approach produced more accurate results, establishing its use for forecasting holiday traffic [11]. Traffic forecasts are essential for Intelligent Transportation Systems (ITS). The expected result is crucial for both drivers and traffic control systems. The forecast is crucial because of the non-linear characteristics of the traffic data. Hashemi et al. used several machine learning techniques, including RF, NBC, CN2, and CT, to predict traffic patterns by the application of numerous category IDs in their research. The system's mentioned route properties may be efficiently governed by an if-then rule. This system employs a classification algorithm to forecast traffic. The results of this system demonstrate that RF and CT have enhanced predictive efficacy. The NBN illustrates the correlations among different IDs used for prediction. The CN2 approach is appropriate and beneficial for understanding traffic patterns [12]. Individuals often use social networking services to facilitate interpersonal interactions. A significant segment of the population extensively utilises Facebook, Instagram, Twitter, WhatsApp, and several other social media services. The use of social media platforms allows users to enhance their everyday lives. Instances of augmented information include meteorological data, traffic conditions in various locations, and real-time occurrences in the affected areas. Urban congestion is a substantial problem. The traffic congestion has caused a considerable delay for automobiles. Individuals are remaining informed about their needs via social media throughout this time of interruption. This work by Kovuru Sridevi et al. use real-time prediction algorithms to evaluate data obtained from social media posts and forecast traffic values. A specialised website is created with the objective of global data measurement. We collect data obtained from Twitter media and classify it according to its geographical location. The author use sentiment analysis derived from Twitter data to classify messages as good, neutral, or negative. The appropriateness of the RF approach for traffic assessment depends on the level of traffic congestion. Random Forest methodologies are proficient for both classification and prediction tasks. This method attains an accuracy rate of 88%. A key feature of this system is its capacity to suggest an alternate route with less traffic congestion, hence improving passenger convenience. Ultimately, it was concluded that the total number of tweets from each area is closely associated with the overall number of individuals interested in that specific place. Automobile accidents are the leading cause of worldwide death. Data mining methods are often used in many applications because of the intrinsic characteristics of traffic data. The principles of prediction and decision-making are widely used across several fields. Poonam Rani and colleagues performed an examination of data mining approaches using the WEKA tool. The authors use the J48, NBC, DT, and RF algorithms to assess the dataset and provide predictions. The RF approach outperforms other algorithms and functions swiftly, even with large datasets [13]. Heuristic and statistical methods are efficient for traffic forecasting. Both linear and non-linear models may be identified in the patterns of traffic data. Chunjiao Dong et al. outline the advantages of traffic flow prediction models and evaluate their performance outcomes in their research. The Support Vector Regression (SVR) approach efficiently combines heuristics with statistical models, using the significant strengths of both linear and non-linear models. The input values are obtained from the observed spatial-temporal connections. This model precisely predicts future traffic patterns for both elevated and diminished volumes. The amalgamation of the two models might improve the accuracy rate by an average of 9.04%, hence augmenting their applicability to various scenarios. The precision of forecasts improves when both geographical and temporal data are used, especially in situations marked by significant traffic congestion [14]. A vital component of Intelligent Transportation Systems (ITS) is the accurate prediction of traffic data. A crucial approach for evaluating the system's efficacy is by analysing its velocity. Xiaoxue Yang et al. used several statistical and machine learning models to assess predictive values across many scenarios, including multi-horizon forecasting, periodic values, peak times, and others. This research analysed three statistical models: ST, VAR, and ARIMA. An examination was performed on machine learning models including RNN, SVM, and MLP. The hybrid model integrates both the velocity metric and temporal data. The data consists of two separate components: the occasional component and the residual component, according to its estimation. The trigonometric method illustrates the systematic component, whereas the residual element is produced using statistical techniques or machine learning methods. The data indicate that including a constant speed component during peak hours improves prediction accuracy. The model's performance improves with time due to the significance of the periodic components. The models are evaluated using the information from the 394-highway segment in Minnesota. The MAE, RMSE, and MAPE metrics are used to assess the models' interpretations. Real-time information on traffic conditions allow drivers to adjust their travel plans or alter their routes to alleviate the effects of severe congestion [15]. Traffic forecasting is an essential element in Intelligent Transportation Systems development. The primary objective is to identify patterns and alternate pathways. Nonetheless, deep learning approaches have recently attracted increased interest from both academics and professionals in the commercial sector. Luis Romo et al. examined three machine learning models: EGB, CNN, and LSTMNN, to predict traffic speed. The writers gathered data from California's transit infrastructure. Machine learning algorithms consistently and accurately predicted future outcomes. Traffic prediction is an essential element of intelligent transport systems (ITS) that significantly influences traffic management and regulation. Precise traffic forecasting is difficult because of the geographical and temporal correlations included in the data. Xin Fu et al. proposed a prediction system that combines a neural network (NN) model with a wavelet transform methodology. The aim of this technology is to improve the precision of predictive models. The creation of high-frequency random series and low-frequency trend series necessitates the preliminary use of the wavelet transform method on the original data. Subsequently, after data processing, the coefficient values are collected. Meticulous data management strategies may re-establish spatiotemporal relationships among input values. The ARMA model was used to fit irregular high-frequency time series. Ultimately, integrate the results of these two models to get the final forecast. The data acquired from Ningbo is used to assess the suggested model. The proposed method has produced an accurate result, as shown by its development. The growth of transportation infrastructure and the number of vehicles on urban roads may be directly related to this prevalent trend. In urban areas, it presents a considerable challenge. Traffic congestion has resulted in several challenges. The main challenges in urban regions are the overconsumption of petrol, air pollution, and the expensive hourly costs borne by drivers. Ioannis Loumiotis et al. created a distinctive method for traffic forecasting using artificial neural networks (ANN) and intelligent agents. An artificial neural network (ANN) was created and implemented using the GRNN approach to evaluate the suggested technology. Motorists may make educated evaluations by using the precise forecasts offered by the traffic system, which includes the capacity to control traffic light operations to alleviate congestion. An Artificial Neural Network (ANN) is used to determine the speed of vehicles on the roadway. The results of the system demonstrate that this novel strategy produces more precise outputs with a decreased error rate. Traffic congestion is widely seen as a global problem. Taghreed Alghamdi et al. use the ARIMA model to identify the variables significantly affecting traffic congestion levels. They supplied a temporal sequence using shorter intervals for traffic data that deviates from a Gaussian distribution. This method aids decision-makers in alleviating traffic congestion via the collection and prediction of unfavourable circumstances. The authors begin by praising the characteristics and structure of the provided data collection. The R programming language may be used to prepare the specified dataset for model building. The ARIMA method is widely acknowledged for its use in hourly flow estimations in California, USA, particularly for forecasting and analytical objectives. The results obtained from the proposed model demonstrate improved effectiveness in forecasting traffic situations. Challenges in current Intelligent Transportation Systems originate from the reliance on wireless sensor networks, which are mostly used by adaptive systems to assess traffic density. It employs clustering for the same purpose. Consequently, the development of such infrastructure is essential and expensive. The rates of arrival and departure significantly influence the system's functioning in this specific arrangement. IoE devices inside the transportation system collect significant amounts of data. A robust system is necessary to monitor and evaluate this extensive data set. Google Maps sometimes delivers erroneous instructions after accidents or road maintenance. This significant amount of data is expected to provide immediate results. The system requires guidance to independently make judgements. Human judgements might cause unnecessary delays in the process. The Data-Driven Intelligent Transportation System employs video technology. It utilises image processing techniques to assess traffic density by first converting the video into a numerical picture. Nevertheless, in instances of dissent or dispute. The procedure use an algorithm to extract elements. This technique needs a considerable quantity of RAM to accommodate the video storage. Object extraction presents an additional challenge. Object extraction gets more challenging when the backdrop colour corresponds with that of the vehicle.

2. Methodology

This section presents a technique for forecasting traffic flow to mitigate accidents via the implementation of intelligent traffic systems (ITS), using the Internet of Things, machine learning, and feature selection. Figure 2 illustrates this paradigm, whereby the collecting of real-time sensor data is facilitated by the deployment of Internet of Things (IoT) devices on highways or vehicles. The dataset comprises real-time information gathered via the Internet of Things and historical traffic data. The input dataset is housed in a centralised cloud. The data undergoes preprocessing to remove extraneous influence and detect outliers. Optimal feature selection is crucial for improving accuracy and reducing the Root Mean Square Error. Particle swarm optimisation is used to find key characteristics from the input dataset. A classification model is developed using linear regression, MLP, and XGBoost techniques. Future traffic forecasts are essential for the development, execution, and enhancement of transportation systems. The management of operations and controls benefits from it. While precise economic analysis is crucial, traffic forecasting also fulfils other supplementary functions, including system design, planning, air quality assessments, and pollution management. The additional expenses associated with overdesign and under design stem from inaccuracies in traffic flow forecasts and safety evaluations. Traffic forecasting is an essential and foundational component of Intelligent Transportation Systems (ITS). It is essential to provide transportation-related information services, including traffic management and guidance. An innovative experimental analysis-based mining approach was developed to improve the accuracy of real-time traffic forecasts. The efficacy of machine learning models is heavily reliant on preprocessing activities. Preprocessing is an essential element in the development of machine learning models.

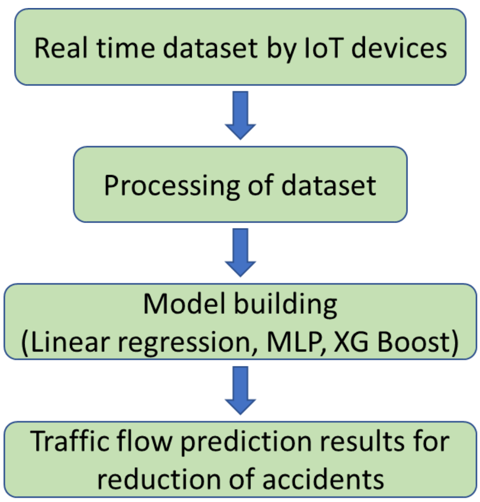

Fig. 2 Machine learning enabled traffic flow chart

The precision of the created models is also affected by the existence of noisy data. The acquired original dataset is subjected to preprocessing using several ways to eliminate extraneous data. The interpolation method was used to fill in any missing values included in the original dataset. The collected data used for developing a machine learning model demonstrates shortcomings in cleanliness and organisation. The original dataset is inadequate for developing advanced machine learning models because of its significant noise, irregular structure, and occurrence of missing values. Therefore, it is essential to guarantee the precision and integrity of the data. Preprocessing processes are crucial for improving the efficacy and precision of machine learning systems. The next phase in the preparation process is addressing the absent data within the specified dataset. Problems with machine learning models are exacerbated when datasets include missing values. Consequently, the development of machine learning models requires careful attention to sometimes overlooked factors. To tackle the issue of missing data, two main strategies are employed: eliminating the problematic row from the input file and calculating the mean value for that row or column. In machine learning preprocessing, a dataset may be partitioned into two separate subsets: a training dataset and a test dataset. The preparation duties at this stage are very demanding. In the realm of data mining, XGBoost is the most direct method for tackling classification tasks and regression challenges. This approach, like to other non-parametric algorithms, use comparisons with analogous data from the training set to assign labels to newly acquired data. Developing models before using the XG BOOST technique is superfluous, since it is a collaborative learning methodology. Comprehending the relationship between the qualities and results is not essential for the efficacy of this strategy; it is very unstructured. In instances of traffic congestion, the closest neighbor's value may duplicate prior data that precisely mirrors the present situation. The XG BOOST methodology may be used to estimate data speed and flow, using density as the target output and flow and speed values as the input variables. This method entails pinpointing input data that is both proximate and critical. XGBoost, MLP, and Linear Regression are prominent machine learning algorithms used for predictive modelling, each possessing unique attributes. XGBoost (Extreme Gradient Boosting) is a robust ensemble learning technique that employs decision trees and gradient boosting to enhance performance, delivering excellent accuracy and speed, especially for structured data. A Multilayer Perceptron (MLP) is a kind of artificial neural network designed for intricate, non-linear interactions; it has input, hidden, and output layers with linked neurones that acquire knowledge via backpropagation. Linear Regression, conversely, is an easier approach that delineates the linear connection between input variables and a target variable, making it suitable for uncomplicated prediction tasks with continuous outcomes. Collectively, these algorithms provide a range of solutions, from interpretable to extremely adaptable, for diverse data science needs. To mitigate accidents with Intelligent Transportation Systems (ITS), XGBoost, MLP, and Linear Regression may serve complimentary functions in the analysis and prediction of accident hazards. XGBoost can effectively analyse extensive transportation datasets, pinpointing critical accident predictors such as traffic density, weather conditions, and driver behaviour, while providing high accuracy and actionable insights. The Multilayer Perceptron (MLP) effectively captures non-linear interactions among intricate factors, including road geometry, real-time sensor data, and vehicle dynamics, allowing sophisticated predictions and risk evaluations. Linear Regression offers a clear methodology to comprehend the direct influence of certain variables, such as speed restrictions or traffic flow, on accident frequency. By integrating these algorithms, ITS can provide a comprehensive framework for predictive analytics, allowing authorities to execute targeted interventions, enhance traffic management, and eventually decrease accident rates efficiently.

3. Result and Discussion

This experimental setup employs various cases for model training, whereas 501 instances are used for testing the categorization model. This dataset is created by gathering data from 36 sensor points at 15-minute intervals. The specified dataset contains many essential columns, including country, city, road type, junction type, and vehicle type. This dataset has both noisy and incomplete data.

Table 1. MSE and RMSE for network algorithms used

Model	MSE	RMSE
Linear regression	0.0056	0.01
MLP	0.0029	0.009
XG Boost	0.0021	0.007

To tackle the problem of noisy or absent data, two main methods are employed: eliminating the problematic row from the input file and calculating the mean value for that row or column. Feature selection is essential for enhancing accuracy and reducing Root Mean Square Error. Particle swarm optimisation is used to identify significant characteristics from the input dataset. The initial data collection has a total of 47 properties. Thirty-one features were identified by the use of Particle Swarm Optimisation on the dataset. These attributes are chosen via p-best (personal best) and g-best (global best) methods. A criterion of 0.5 has been used to ascertain the selection of the feature. A classification model is constructed using XGBoost, MLP, and linear regression algorithms. The model is trained on 1600 cases and tested on 501 examples from the input dataset. The metrics used to assess the classification model are Accuracy, Precision, Recall, Mean Squared Error, and Root Mean Squared Error.

The XG BOOST algorithm has achieved a Mean Squared Error (MSE) rate of 0.0021 and a Root Mean Squared Error (RMSE) rate of 0.007. Fig. 3, 4 and 5 showed the accuracy, sensitivity and recall of the different network algorithms used.

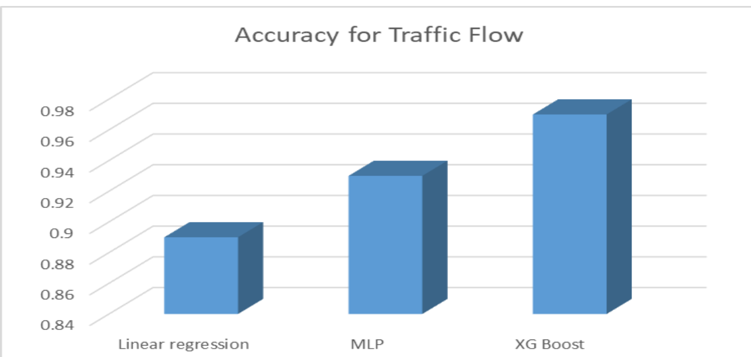

Fig. 3 Accuracy of Machine learning models (Linear regression, MLP and XG Boost) to

reduce accidents

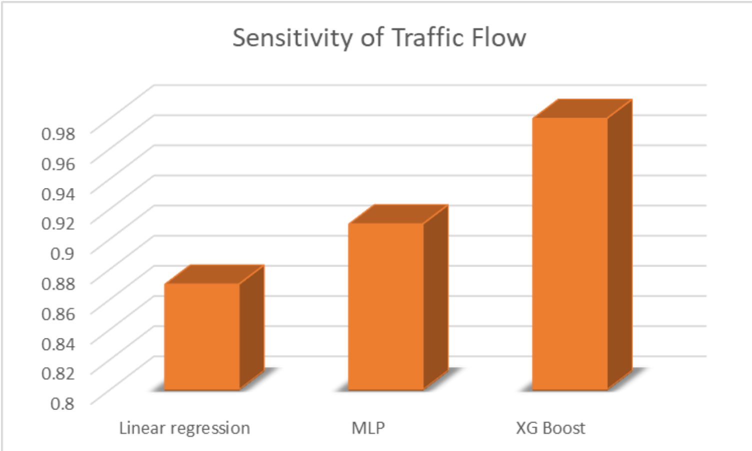

Fig. 4 Sensitivity of Machine learning models (Linear regression, MLP and XG Boost)

Accuracy Comparison of XG BOOST and other techniques are shown in Fig. 5 for Traffic Flow Prediction The XG BOOST model has a high level of accuracy, specifically 99 percent, in predicting traffic flow.

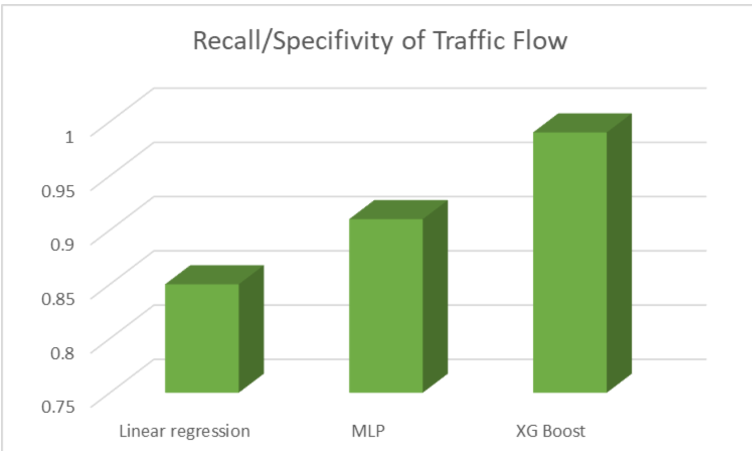

Fig. 5 Recall/Specificity of Machine learning models (Linear regression, MLP and XG Boost)

4. Conclusion

This study demonstrates the application of advanced machine learning algorithms, coupled with IoT data and feature selection techniques, to predict traffic flow and reduce accidents within Intelligent Transportation Systems. By integrating real-time and historical data, the framework effectively captures complex traffic patterns. Among the algorithms used, XGBoost outperformed others, achieving a high prediction accuracy of 98% with minimal error rates (MSE: 0.0021, RMSE: 0.007), proving its suitability for traffic forecasting. The incorporation of Particle Swarm Optimization (PSO) further enhanced model performance by selecting the most relevant features, reducing noise, and improving efficiency. The results underline the importance of leveraging IoT, machine learning, and data-driven methodologies to address traffic congestion and safety challenges. This approach offers actionable insights for transportation authorities to implement targeted interventions, optimize traffic management, and design smarter infrastructure, paving the way for safer and more efficient urban transportation systems.

References

1. Boukerche Azzedine, Yanjie Tao, Peng Sun, Artificial intelligence-based vehicular traffic flow prediction methods for supporting intelligent transportation systems, Comput. Network. 182 (2020) 107484. https://doi.org/10.1016/j.comnet.2020.10 7484.
2. Yazeed Yasin Ghadi, Tehseen Mazhar, Tamara al Shloul, Tariq Shahzad, Umair Ahmad Salaria, Arfan Ahmed, Hamam Habib, Machine Learning solution for the security of wireless sensor Network, IEEE Access 12 (2024) 12699–12719, https:// doi.org/10.1109/ACCESS.2024.3355312.
3. W. Shu, K. Cai, N.N. Xiong, A short-term traffic flow prediction model based on an improved gate recurrent unit neural network, IEEE Trans. Intell. Transp. Syst. 23 (9) (Sep. 2022) 16654–16665.
4. Yazeed Yasin Ghadi, Tehseen Mazhar, Syed Faisal Abbas Shah, Inayatul Haq, Wasim Ahmad, Khmaies Ouahada, Hamam Habib, Integration of federated learning with IoT for smart cities applications, challenges, and solutions, PeerJ Computer Science 9 (2023) e1657.
5. Moneeb Gohar, Muhammad Muzammal, Arif Ur Rahman, Smart TSS: defining transportation system behavior using big data analytics in smart cities, Sustain. Cities Soc. 41 (2018) 114–119.
6. Wen Yao, et al., On-road vehicle trajectory collection and scene-based lane change analysis: Part II, IEEE Trans. Intell. Transport. Syst. 18 (1) (2016) 206–220.
7. Wuping Xin, John Hourdos, Panos G. Michalopoulos, Vehicle Trajectory Collection and Processing Methodology and its Implementation, 2008. No. 08-2173.
8. Shuming Sun, Juan Chen, Jian Sun, Traffic congestion prediction based on GPS trajectory data, Int. J. Distributed Sens. Netw. 15 (5) (2019) 1550147719847440.
9. A. Saad, P. Yan, Robson, MDP-based connectivity and availability models for Internet of Vehicles, Internet of things 24 (Dec. 2023) 100963, https://doi.org/ 10.1016/j.iot.2023.100963, 100963.
10. Yachao Jia, Xin Wang, Intelligent traffic decision analysis system based on big data mining, IOP Conf. Ser. Mater. Sci. Eng. (2018) 15.
11. Khokale Rahul, Ghate Ashwini, Data mining for traffic prediction and analysis using big data, Int. J. Eng. Trends Technol. 48 (3) (2017) 152–156. June 2017.
12. Yuheng Zhou, Zuping Zhang, Prediction of traffic flow based on deep learning, International Journal of Advanced Computer Technology (IJACT) IX (II) (2020) 5–11. ISSN: 2319-7900.
13. Xianglong Luo, Danyang Li, Shengrui Zhang, Traffic flow prediction during the holidays based on DFT and SVR, Hindawi Journal of Sensors 2019 (2019) 1–10. Article ID 6461450.
14. S.Mehdi Hashemi, Mehrdad Almasi, Roozbeh Ebrazi, Mohsen Jahanshahi, Predicting the next state of traffic by data mining classification techniques, International Journal of Smart Electrical Engineering 1 (No.3) (2012) 180–192. Fall 2012 ISSN: 2251-9246.
15. Kovuru Sridevi, T. Ganesan, B.V.S. Samrat, S. Srihari, Traffic analysis by using random forest algorithm considering social media platforms, Int. J. Recent Technol. Eng. 7 (6S) (2019) 620–625. ISSN: 2277-3878.