WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Anomaly Detection and Short-Term Prediction of Traffic Flow by using Ensemble Deep Learning Approach

Poonam Bhartiya, Dr. Mukta Bhatele, Lokendra Gour[*1]

Department of Computer Science and Engineering, AKS University, Satna, Madhya Pradesh, India.
[1]Email: lokendra_gaur@aksuniversity.com

ABSTRACT

Nowadays, intelligent transport systems (ITS) are primarily concerned with traffic management and sustainable mobility. Managing crowd information in public transport (PT) systems is essential for promoting sustainable mobility and enhancing customer satisfaction during routine operations. One of the challenges Predicting data on road traffic flow is a significant difficulty for ITS. A lot of studies have been done over the past few decades on the challenging nonlinear challenge of making short-term predictions using traffic flow data. A large number of applications that must ensure high-quality services, such as traffic flow analysis and reconstruction, which in turn are used to compute the predictions required to perform what-if analysis, forecast routing, conditioned routing, crowd predictions, etc., must have access to traffic flow data. The most effective approach for traffic congestion mitigation relies on accurate and instantaneous traffic prediction. Machine learning models can predict long-term traffic status, while deep learning methods have proven highly effective in predicting road traffic. A deep learning approach is not appropriate for predicting instantaneous traffic flow. A novel Ensemble Deep Neural Network (EDNN) is proposed to predict traffic fluctuations. The proposed work addresses the issues that need to be resolved in order to achieve seamless integration between ITS and deep learning techniques in order to handle problems like enhancing traffic flow and transportation logistics, predicting the best routes for the transportation of goods, achieving optimal fuel consumption, intelligent environmental conditions perception, traffic speed management, and accident prevention. The proposed system has been developed and validated for the solution of traffic problems that exist in the smart city and metro city contexts.

Keywords: Intelligent Transportation System (ITS), Traffic Flow Data, Short-Term prediction, Anomaly detection, Ensemble Deep Neural Network (EDNN), etc.

Introduction

An Intelligent Transportation System (ITS) is a framework that utilizes advanced technologies to improve transportation efficiency, safety, and sustainability. One of the key areas of focus for ITS is managing traffic flow effectively. By employing various tools and strategies, ITS aims to optimize the movement of vehicles on roads, reduce congestion, and enhance overall transportation performance. ITS incorporates sensors, cameras, and other monitoring devices to collect real-time data on traffic conditions. This information includes vehicle counts, speeds, and

congestion levels. By analyzing this data, transportation authorities can gain insights into traffic patterns, identify bottlenecks, and make informed decisions to alleviate congestion.

ITS employs intelligent traffic signal control systems to dynamically adjust signal timings based on real-time traffic conditions. By optimizing signal timings and coordination, these systems can reduce delays and improve traffic flow at intersections.ITS helps in the efficient management of incidents such as accidents, road closures, or other disruptions. By detecting incidents promptly through surveillance systems or citizen reporting, authorities can respond quickly and implement alternative routes or traffic diversions to minimize the impact on traffic flow.

DMS(Dynamic Message Signs) are electronic signs placed along roadways that provide real-time information to drivers. These signs can relay traffic conditions, travel times, incident alerts, and alternate route suggestions. By keeping drivers informed, DMS helps in distributing traffic and reducing congestion on congested routes.ITS integrates with navigation systems and mobile applications to provide drivers with real-time route guidance. By considering current traffic conditions, these systems can suggest the most efficient routes, helping drivers avoid congested areas and distribute traffic across different routes.

ITS enables the implementation of variable speed limit systems, where speed limits are adjusted based on prevailing traffic conditions. Lowering speed limits during congested periods helps in maintaining smoother traffic flow, reducing the stop-and-go phenomenon that contributes to congestion. By analyzing historical and real-time data, ITS can generate traffic flow predictions. These predictions help transportation authorities and drivers anticipate traffic conditions and make proactive decisions, such as adjusting travel times or choosing alternative routes. By leveraging these and other advanced technologies, an Intelligent Transportation System aims to optimize traffic flow, enhance safety, reduce travel times, and improve the overall transportation experience. It is a multifaceted approach that combines data collection, analysis, communication systems, and smart decision-making to achieve these objectives. Traffic flow refers to the movement of vehicles on roadways or other transportation networks. It involves the study and analysis of the movement, interactions, and behavior of vehicles, pedestrians, and other road users within a transportation system. Traffic flow can be influenced by various factors, including road capacity, traffic volume, speed, and the behavior of drivers. Understanding and managing traffic flow is crucial for maintaining efficient and safe transportation systems.

There are different types of traffic flow patterns, including:

- **Free flow:** This occurs when vehicles can move smoothly and without significant disruptions, typically at or near the posted speed limit.
- **Congested flow:** Congestion happens when traffic demand exceeds the capacity of the road, resulting in slower speeds, reduced throughput, and increased travel times.
- **Shockwave:** A shockwave refers to a sudden disruption in traffic flow, often caused by an incident such as a collision or sudden braking. It leads to a temporary slowdown or stoppage of vehicles and can cause a ripple effect, affecting traffic downstream.
- **Oscillatory flow:** This pattern occurs when traffic alternates between periods of congestion and periods of free flow, creating a wave-like motion. It is commonly observed in situations where traffic density is close to the capacity of the road.

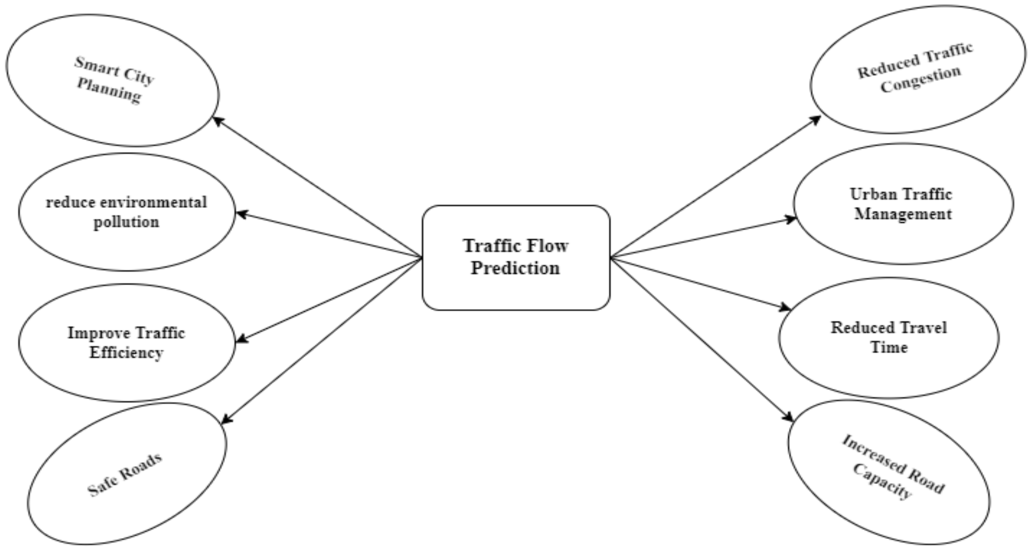


Fig.1: Benefits of Traffic Flow Prediction

Traffic flow is a complex field studied by transportation engineers and researchers. They employ various tools and techniques, such as traffic modeling, simulation, and data analysis, to understand, predict, and optimize traffic flow patterns. By managing traffic flow effectively, transportation agencies can reduce congestion, improve safety, and enhance the overall efficiency of the transportation system. Figure 1 dipicts benefits of traffic flow prediction.

The introduction of the neural network-based prediction model into traffic scenario forecasting has been ongoing in recent years due to the advancement of deep learning technology. Machine learning ensemble approach for predicting network traffic in VANET, Ahmadi P. et al. [1], the suggested approach is the Stacking Ensemble Learning with Booster Model (STK-EBM). To forecast traffic time series data in traffic flow, Luo and Zhao [2], and [3], employ recurrent neural networks (RNN), and particularly long short-term memory networks (LSTM). To forecast urban traffic flow, Zheng H. et al. [4], employ the Graph Convolution Neural Network (GCN) module and the Generative Adversative Neural Network (GAN) module. To anticipate pedestrian movement from multiple angles, Papathanasopoulou V. et al. [5] Position and speed predictions are calculated using the most appropriate data-driven techniques and For comparison, a Long Short-Term Memory (LSTM) model is also used.

To anticipate data on traffic flow, Guo et al. [6] suggested a deep space-time 3D-CNN approach. A network tracking model was created by Zhu et al. [7] using trajectory data before and after.RNN and CNN-based data are used. To anticipate traffic flow networks, Lin et al. [8] devised a technique for creating adversarial networks. To reflect the topological structure of urban traffic networks as well as the temporal and geographical properties of traffic nodes, Peng et al. [9] introduced a novel form of dynamic graph recursion convolutional neural network called dynamic GR-CNN network. Additionally, [10], [11], and [12] accurately forecasted traffic circumstances with good outcomes. And [13] and [14], introduced Convolutional Bidirectional Deep Long Short Term Memory (CONV-BI-LSTM) neural networks for traffic prediction.

More researchers are using mixed models or algorithms to conduct relevant research in the prediction process to increase the efficiency and accuracy of the prediction process. The findings demonstrate that the mixed model is typically superior to the single model.

To anticipate short-term traffic flows, Duan et al. [17], Sun et al. [16] and Dai et al. [18] created a unified temporal and geographic model. All of these findings support the model's validity. A multimodal fractional Levy stable motion degradation model was created by Duan et al. [19] to forecast the product technical life (PTL) or remaining usable life (RUL) of machinery. The modes were found using the change point detection and clustering algorithm, and an example was utilized to validate the accuracy of the prediction model. Traffic Accident Severity Prediction, Zheng et al. [20], to introduce Deep-Learning Approach-Based CNN Network.

There are very few studies that take into account both traffic flow prediction and control in the context of traffic congestion. There are more and more research techniques that mix intelligent algorithms with other related theories as traffic problems get more complex. In order to address this important issue, this work conducts a thorough analysis of traffic flow prediction and anomaly detection, and it also suggests a traffic flow prediction model based on ensemble deep learning features as well as a congestion section control approach that is based on traffic distribution. The development of a model for the prediction and management of urban traffic congestion is extremely important. As a result, the direction of this paper's research is short-term prediction and anomaly detection.

Anomaly detection and short-term prediction of traffic flow are crucial tasks in transportation management and urban planning. With the increasing complexity of road networks and the growing number of vehicles, accurately monitoring traffic patterns and identifying anomalies can help optimize traffic flow, improve road safety, and enhance overall transportation efficiency. To achieve these goals, researchers and practitioners have turned to advanced machine learning techniques, particularly ensemble deep learning approaches, which combine the power of deep neural networks with the robustness of ensemble methods. An ensemble deep learning approach refers to the utilization of multiple deep learning models to tackle a specific problem. Ensemble methods help improve the accuracy, reliability, and generalization capabilities of individual models by combining their outputs or predictions. In the context of traffic flow analysis, ensemble deep learning techniques have proven to be effective in both anomaly detection and short-term prediction tasks.

1. Anomaly Detection

Anomaly detection involves identifying abnormal traffic patterns or events that deviate significantly from the expected behavior. These anomalies can be caused by accidents, road closures, unusual traffic congestion, or other unpredictable factors. By applying ensemble deep learning methods, traffic anomalies can be detected more accurately by leveraging the collective knowledge of multiple models. This allows for the identification of potential issues in real-time, enabling prompt intervention and mitigation strategies.

Anomaly detection is a technique used to identify patterns or events that deviate significantly from the normal behavior or expected patterns in a dataset. It can be applied to various domains, including network traffic analysis and traffic flow monitoring. Let's explore how anomaly detection can be used in the context of traffic flow.

In traffic flow analysis, anomaly detection helps identify unusual or abnormal patterns in the movement of vehicles, which can indicate incidents, congestion, accidents, or other abnormal events. By detecting anomalies in real-time or retrospectively analyzing historical data, traffic management authorities can gain valuable insights and make informed decisions to improve traffic flow, optimize resource allocation, and enhance overall safety.

Some approaches used in anomaly detection for traffic flow:

- **Statistical Methods:** Statistical models such as Gaussian distribution, clustering algorithms, or time-series analysis techniques can be used to identify anomalies in traffic flow. By establishing normal traffic patterns based on historical data, statistical methods can detect deviations from expected behavior.
- **Machine Learning:** Machine learning algorithms can be trained on historical traffic data to learn normal traffic patterns and identify anomalies. Techniques such as supervised learning, unsupervised learning, and reinforcement learning can be applied to detect anomalies based on features like traffic volume, speed, occupancy, or patterns of congestion.
- **Data Fusion:** Anomaly detection can be improved by combining data from multiple sources, such as traffic sensors, surveillance cameras, and social media feeds. By integrating diverse data streams, anomalies can be detected more accurately and promptly.
- **Time-Series Analysis**: Traffic flow data is often represented as time-series data, where each data point represents a specific time interval. Time-series analysis techniques, such as autoregressive integrated moving averages (ARIMA) or seasonal decomposition of time series (STL) can help identify anomalies by analyzing the temporal patterns and trends in traffic flow.
- **Network Analysis:** Anomaly detection can also involve analyzing the network structure and behavior of traffic flow. Network analysis techniques, such as graph-based anomaly detection or flow-based anomaly detection, can reveal abnormal patterns in the connectivity or interaction between different segments of the transportation network.

By employing these techniques, anomaly detection can assist in identifying traffic congestion, accidents, sudden changes in traffic patterns, or any abnormal events that affect traffic flow. The insights gained from anomaly detection can be used to trigger alerts, reroute traffic, optimize traffic signal timings, allocate resources efficiently, or dispatch emergency services promptly, ultimately leading to improved traffic management and enhanced road safety.

2. Short-Term Prediction

Predicting short-term traffic flow is a complex task that relies on various factors, including historical data, current conditions, and predictive algorithms. While I don't have access to real-time data or the ability to make real-time predictions, I can provide you with a general outline of the factors that influence short-term traffic flow prediction.

Analyzing historical traffic patterns, such as traffic volume, congestion levels, and average travel times during specific periods and days of the week, can help predict future traffic flow. Weather plays a significant role in traffic patterns. Inclement weather, such as rain, snow, or fog, can reduce traffic speeds and increase congestion. Predicting short-term traffic flow requires considering weather forecasts and their potential impact on road conditions. Figure 2 show the short-term traffic flow prediction.

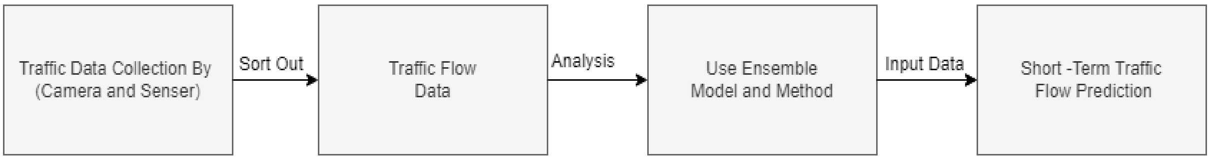


Fig.2: Short-Term Traffic Flow Prediction

Events like sports games, concerts, festivals, or parades can significantly affect traffic flow in the surrounding areas. Predictions should consider scheduled events and their potential impact on road usage and congestion. Access to real-time data from traffic sensors, surveillance cameras, and mobile apps can provide valuable insights into current traffic conditions. Analyzing this data can help predict short-term traffic flow by considering ongoing congestion, accidents, road closures, or any other unexpected events. Advanced machine learning algorithms can process historical and real-time data to identify patterns and trends. These algorithms can then generate predictions based on the gathered information.

It's important to note that short-term traffic flow prediction is subject to uncertainties and may not always be accurate due to unpredictable events or sudden changes in conditions. For real-time and precise predictions, it is best to rely on dedicated traffic management systems and services that utilize comprehensive data sources and advanced modeling techniques.

Short-term prediction of traffic flow aims to estimate future traffic conditions, such as traffic volume, speed, and congestion levels, within a specific time horizon. Accurate short-term predictions are essential for traffic management, enabling authorities to optimize traffic signal timings, plan road maintenance activities, and implement congestion reduction strategies. Ensemble deep learning techniques excel in this task by combining the strengths of different models, capturing complex spatiotemporal dependencies in traffic data, and generating reliable predictions.

The advantages of ensemble deep learning approaches for anomaly detection and short-term prediction of traffic flow include improved accuracy, robustness against noisy and incomplete data, and better generalization capabilities. These techniques can handle large-scale traffic datasets, adapt to dynamic traffic patterns, and provide valuable insights for decision-making in transportation management.

This paper proposes an ensemble deep-learning approach for anomaly detection and short-term prediction of traffic flow. Leverage the power of deep neural networks, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and attention mechanisms in combination with ensemble methods(such as bagging or stacking) to achieve superior performance. Evaluatingthe approach to real-world traffic datasets, demonstrating its effectiveness in accurately detecting anomalies and making reliable short-term predictions. The results of this study contribute to the field of transportation management, offering practical solutions for improving traffic flow efficiency and enhancing road safety.

3. Ensemble Deep Learning

Ensemble deep learning techniques combine the predictions of multiple deep learning models to improve overall performance and make more accurate predictions. Here are a few commonly used ensemble deep-learning techniques:

- **Bagging: (Bootstrap Aggregating)**: Bagging involves training multiple deep learning models independently on different subsets of the training data. Each model is trained on a random sample of the training data, obtained through bootstrapping (random sampling with replacement). The final prediction is obtained by aggregating the predictions of all the individual models, typically by averaging or majority voting.
- **Boosting:** Boosting is another popular ensemble technique that trains multiple deep-learning models sequentially. Each model is trained to correct the mistakes made by the previous models. Boosting assigns higher weights to misclassified samples, so subsequent models focus more on those samples. The final prediction is typically obtained by weighted voting or averaging of the predictions made by all the models.

- **Stacking:** Stacking combines predictions from multiple models by training a meta-model on the predictions of individual models. In stacking, the original deep learning models act as "base" models, and their predictions are used as features for the meta-model. The meta-model is trained to make the final prediction based on the combined predictions of the base models.
- **Voting:** Voting is a simple ensemble technique where multiple deep learning models are trained independently, and the final prediction is determined by majority voting (for classification problems) or averaging (for regression problems) of the predictions made by all the models.
- **Weighted Average:** In this technique, each deep learning model is assigned a weight that reflects its performance or reliability. The final prediction is obtained by taking a weighted average of the predictions made by all the models, with the weights representing the importance of each model's contribution.
- **Adaptive Weighted Ensemble:** This technique dynamically adjusts the weights assigned to different models based on their performance on the current task. Models that perform better are assigned higher weights while underperforming models receive lower weights. This approach allows the ensemble to adapt and prioritize the more accurate models.

These ensemble deep learning techniques can be applied to a variety of deep learning architectures, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and transformer-based models. They can help in improving generalization, reducing overfitting, and enhancing the overall predictive power of deep learning models. Figure 3 shows the traffic flow model for prediction.

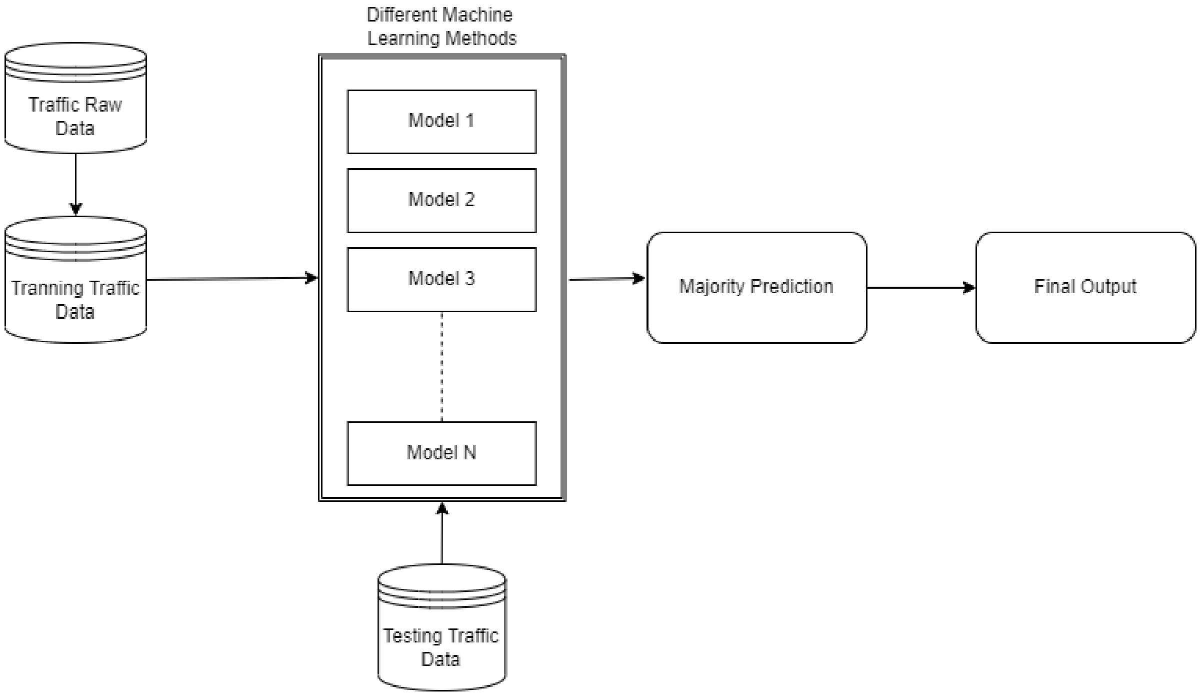


Fig 3: Traffic Flow Prediction Model

Overall the combination of ensemble deep learning techniques with traffic flow analysis provides a promising avenue for addressing the challenges in anomaly detection and short-term prediction. By leveraging the strengths of multiple models and capturing intricate patterns in traffic data, these approaches have the potential to revolutionize transportation management systems, leading to safer, more efficient, and sustainable urban mobility.

Using an ensemble deep learning approach in traffic flow analysis and prediction has several significant benefits. Ensemble methods combine multiple models to make predictions, leading to improved accuracy compared to using a single model. In traffic flow analysis, where prediction accuracy is crucial, ensembles can help reduce errors and provide more reliable traffic forecasts. Ensemble methods are generally more robust and less prone to overfitting. Traffic flow is a complex and dynamic system influenced by various factors, such as weather conditions, road incidents, and driver behavior. Ensemble models can capture the inherent complexity and variability in traffic patterns, leading to more robust predictions that are less affected by individual model biases or noise. Traffic flow is inherently uncertain, with sudden changes and unpredictable events occurring frequently. Ensemble approaches can effectively handle uncertainty by aggregating predictions from different models. By considering multiple perspectives, ensembles can provide a more comprehensive understanding of the traffic situation, including both expected patterns and possible deviations.

Ensemble models often exhibit better generalization capabilities. Traffic patterns may vary across different regions, cities, or periods. By training multiple models on diverse datasets or using different architectural variations, ensemble methods can capture a wider range of traffic patterns and generalize well to new, unseen scenarios. Traffic flow analysis often requires processing large volumes of data in real-time. Ensemble models can be designed to distribute the computational load across multiple models or parallelize the predictions, making them scalable and suitable for handling the high demands of traffic flow analysis.

Ensemble deep learning approaches can provide decision support for traffic management systems and transportation planners. By combining predictions from multiple models, ensembles can offer more reliable information about traffic conditions, congestion hotspots, or travel time estimates, enabling better-informed decision-making for traffic flow optimization, route planning, and resource allocation. Ensemble methods allow for model fusion, where predictions from different models are combined to create an aggregated prediction. Fusion techniques, such as weighted averaging or stacking, can exploit the strengths of individual models and minimize their weaknesses, leading to a more accurate and robust traffic flow prediction. Overall, the significance of using an ensemble deep learning approach in traffic flow lies in its ability to improve accuracy, handle uncertainty, enhance generalization, and provide decision support. These advantages make ensemble models valuable tools for managing and optimizing traffic flow in urban areas, improving transportation systems, and reducing congestion.

4. Data Pre-Processing

The sensors that are positioned along the highway can be used to measure traffic flow parameters like the volume of traffic (traffic intensity), the average speed of the vehicles, the volume of the highway (occupancy level), time and distance headways, traffic density, etc. The features of the traffic flow at a certain time can then be estimated using this information, for example, to identify incidents, model the traffic and predict future traffic conditions, and so on. The number of cars, the average vehicle speed, and the amount of traffic on the highway are the three metrics of particular significance. The number of vehicles passing at a specific location within a specific period is the most obvious parameter to be assessed.

Another crucial factor defining the condition of the traffic on the highway and in urban cities is the speed of the moving vehicles. The occupancy level of the detector is a third parameter that describes the current traffic scenario. This represents the percentage of time a vehicle is visible to the detector out of all time. On the roadway, the traffic parameters are measured in several places. The measurements are transformed and merged in a centralized and standardized database to aggregate the data acquired at various places and to create a controllable system. A computer that is networked to the sensors at the various measurement sites houses the central database. The computer checks the sensors every minute for the most recent measurements of the parameters and saves the results in its memory or on the disc.

The central computer's database follows a predetermined standard data format. The forms of the data produced by various sensors vary generally. As a result, we use a data filter to analyze the sensor's raw data before standardizing it. The interchangeability of data filters is a significant benefit. Only the data filter needs to be modified if a sensor is swapped out for one from a different manufacturer or one that employs a different technology. The database and its elements stay the same. This enables the flexible usage of various sensor kinds. The same central database can also be used by several applications, each of which may utilize a different data format, by accessing it through a data access layer.

Data pre-processing is an essential step in data analysis and machine learning tasks. It involves transforming raw data into a format suitable for analysis and modeling. Typical steps in data pre-processing include data cleaning, handling missing values, handling outliers, data normalization or scaling, feature selection or dimensionality reduction, and data integration. To use the "Urban Traffic Flow CSV" dataset for data pre-processing, we would need to download the dataset from the Kaggle link (https://www.kaggle.com/datasets/jvthunder/urban-traffic-flow-csv). Once the dataset performs various pre-processing steps based on our specific analysis goals or machine learning tasks. Figure 4 shows the data preprocessing steps.

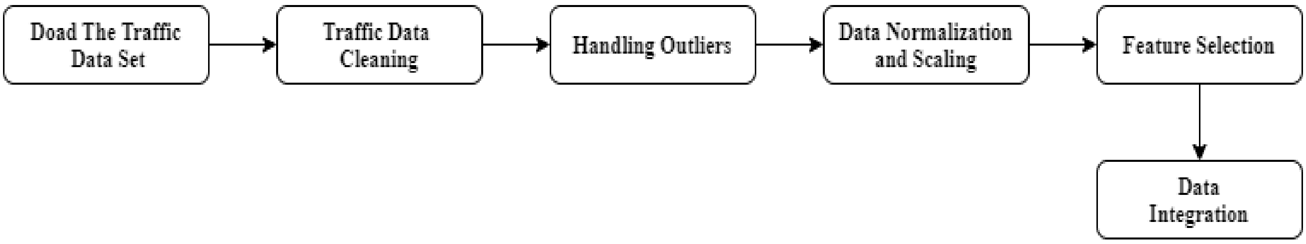


Fig 4: Data Pre-Processing

- **Load the dataset:** Use a programming language or data analysis tool (such as Python with pandas) to load the dataset into a suitable data structure.
- **Data Cleaning:** Identify and handle any missing values, duplicate records, or inconsistent data and need to decide on appropriate strategies, such as filling in missing values or removing incomplete data.
- **Handling outliers:** Identify and handle any outliers in the dataset. Outliers are data points that deviate significantly from the rest of the data and can have a significant impact on analysis results. It might choose to remove outliers or transform them based on the specific context of our analysis.
- **Data normalization or scaling:** Normalize or scale the numerical features in the dataset to ensure they have a consistent scale or distribution. Common techniques include standardization (mean normalization) or min-max scaling.
- **Feature selection or dimensionality reduction:** Analyze the dataset to identify relevant features for their analysis or modeling. use techniques like correlation analysis, feature importance, or dimensionality reduction algorithms (e.g., principal component analysis) to select the most informative features.

- **Data integration:** If additional datasets or external data sources can enhance their analysis, consider integrating them with the "Urban Traffic Flow CSV" dataset. Ensure that the integration is done properly, and the data is merged based on suitable key attributes.

About Data

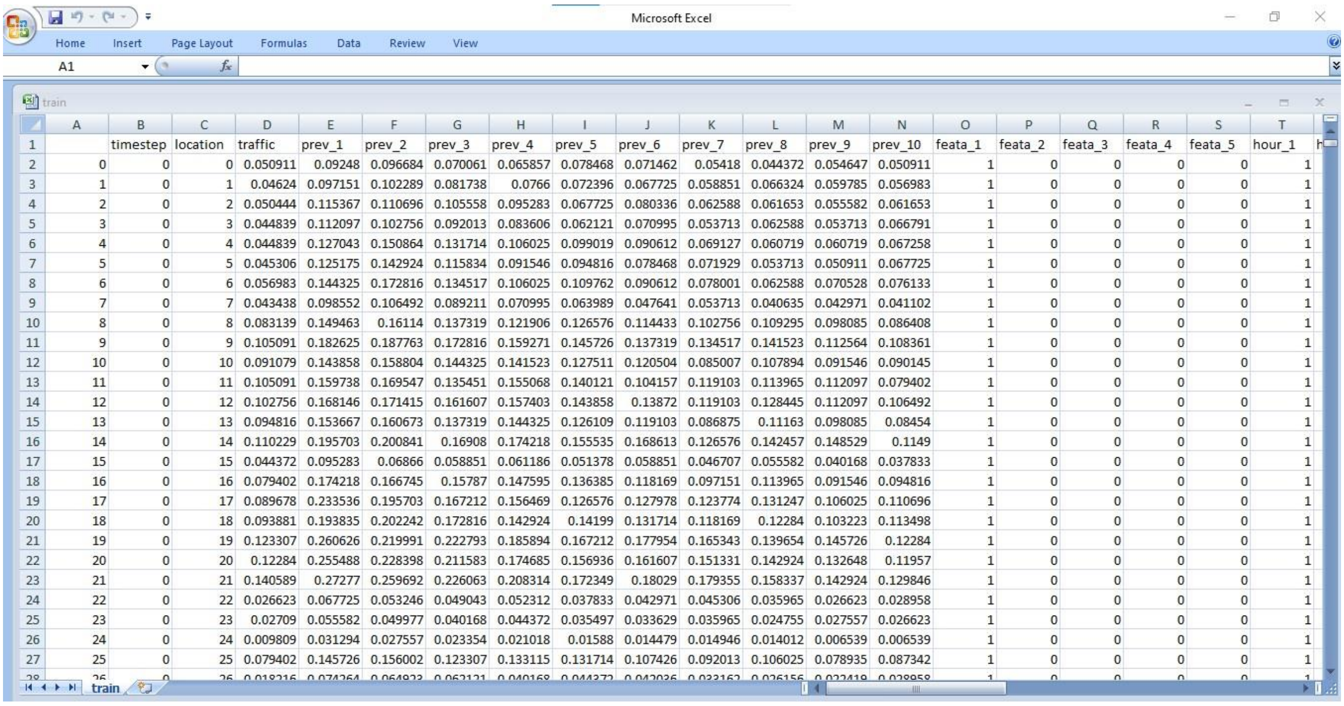
This dataset originated from [UCI Traffic Flow Forecasting Data Set](). The original file is in Matlab format and here changed it into a CSV(comma-separated values) file. Figure 5 shows the train.csv dataset.

There are 3 files:

- train.csv (Contains 1261-time steps for 36 locations calculated with 15-minute intervals)
- test.csv (Contains 840-time steps for 36 locations calculated with 15-minute intervals)
- adj_matrix.csv (Contains the location's adjacency matrix, so here can include graph connectivity features)

The features in the train.csv are:

- time step (the time step of data)
- location (the location of data)
- traffic (the traffic rate at the point in time (from 0 to 1))
- prev_i (the i-th previous traffic rate)
- hour_i (one-hot encoding of the current hour)
- feata_i, featb_i, featc_i (In these columns the sum values suggest one hot encoding (only 0 or 1 values) of a certain value, researcher includes or does not include this for their predictions)



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1		timestep	location	traffic	prev_1	prev_2	prev_3	prev_4	prev_5	prev_6	prev_7	prev_8	prev_9	prev_10	feata_1	feata_2	feata_3	feata_4	feata_5	hour_1
2	0	0	0	0.050911	0.09248	0.096684	0.070061	0.065857	0.078468	0.071462	0.05418	0.044372	0.054647	0.050911	1	0	0	0	0	1
3	1	0	1	0.04624	0.097151	0.102289	0.081738	0.0766	0.072396	0.067725	0.058851	0.066324	0.059785	0.056983	1	0	0	0	0	1
4	2	0	2	0.050444	0.115367	0.110696	0.105558	0.095283	0.067725	0.080336	0.062588	0.061653	0.055582	0.061653	1	0	0	0	0	1
5	3	0	3	0.044839	0.112097	0.102756	0.092013	0.083606	0.062121	0.070995	0.053713	0.062588	0.053713	0.066791	1	0	0	0	0	1
6	4	0	4	0.044839	0.127043	0.150864	0.131714	0.106025	0.099019	0.090612	0.069127	0.060719	0.060719	0.067258	1	0	0	0	0	1
7	5	0	5	0.045306	0.125175	0.142924	0.115834	0.091546	0.094816	0.078468	0.071929	0.053713	0.050911	0.067725	1	0	0	0	0	1
8	6	0	6	0.056983	0.144325	0.172816	0.134517	0.106025	0.109762	0.090612	0.078001	0.062588	0.070528	0.076133	1	0	0	0	0	1
9	7	0	7	0.043438	0.098552	0.106492	0.089211	0.070995	0.063989	0.047641	0.053713	0.040635	0.042971	0.041102	1	0	0	0	0	1
10	8	0	8	0.083139	0.149463	0.16114	0.137319	0.121906	0.126576	0.114433	0.102756	0.109295	0.098085	0.086408	1	0	0	0	0	1
11	9	0	9	0.105091	0.182625	0.187763	0.172816	0.159271	0.145726	0.137319	0.134517	0.141523	0.112564	0.108361	1	0	0	0	0	1
12	10	0	10	0.091079	0.143858	0.158804	0.144325	0.141523	0.127511	0.120504	0.085007	0.107894	0.091546	0.090145	1	0	0	0	0	1
13	11	0	11	0.105091	0.159738	0.169547	0.135451	0.155068	0.140121	0.104157	0.119103	0.113965	0.112097	0.079402	1	0	0	0	0	1
14	12	0	12	0.102756	0.168146	0.171415	0.161607	0.157403	0.143858	0.13872	0.119103	0.128445	0.112097	0.106492	1	0	0	0	0	1
15	13	0	13	0.094816	0.153667	0.160673	0.137319	0.144325	0.126109	0.119103	0.086875	0.11163	0.098085	0.08454	1	0	0	0	0	1
16	14	0	14	0.110229	0.195703	0.200841	0.16908	0.174218	0.155535	0.168613	0.126576	0.142457	0.148529	0.1149	1	0	0	0	0	1
17	15	0	15	0.044372	0.095283	0.06866	0.058851	0.061186	0.051378	0.058851	0.046707	0.055582	0.040168	0.037833	1	0	0	0	0	1
18	16	0	16	0.079402	0.174218	0.166745	0.15787	0.147595	0.136385	0.118169	0.097151	0.113965	0.091546	0.094816	1	0	0	0	0	1
19	17	0	17	0.089678	0.233536	0.195703	0.167212	0.156469	0.126576	0.127978	0.123774	0.131247	0.106025	0.110696	1	0	0	0	0	1
20	18	0	18	0.093881	0.193835	0.202242	0.172816	0.142924	0.14199	0.131714	0.118169	0.12284	0.103223	0.113498	1	0	0	0	0	1
21	19	0	19	0.123307	0.260626	0.219991	0.222793	0.185894	0.167212	0.177954	0.165343	0.139654	0.145726	0.12284	1	0	0	0	0	1
22	20	0	20	0.12284	0.255488	0.228398	0.211583	0.174685	0.156936	0.161607	0.151331	0.142924	0.132648	0.11957	1	0	0	0	0	1
23	21	0	21	0.140589	0.27277	0.259692	0.226063	0.208314	0.172349	0.18029	0.179355	0.158337	0.142924	0.129846	1	0	0	0	0	1
24	22	0	22	0.026623	0.067725	0.053246	0.049043	0.052312	0.037833	0.042971	0.045306	0.035965	0.026623	0.028958	1	0	0	0	0	1
25	23	0	23	0.02709	0.055582	0.049977	0.040168	0.044372	0.035497	0.033629	0.035965	0.024755	0.027557	0.026623	1	0	0	0	0	1
26	24	0	24	0.009809	0.031294	0.027557	0.023354	0.021018	0.01588	0.014479	0.014946	0.014012	0.006539	0.006539	1	0	0	0	0	1
27	25	0	25	0.079402	0.145726	0.156002	0.123307	0.133115	0.131714	0.107426	0.092013	0.106025	0.078935	0.087342	1	0	0	0	0	1

Fig 5**:** train.csv dataset

Conclusion

In this study, we investigated the application of an ensemble deep learning approach for the tasks of anomaly detection and short-term prediction of traffic flow. We proposed a novel framework that combines multiple deep-learning models to leverage their strengths and enhance overall performance. The experimental results on real-world traffic data demonstrated the effectiveness and superiority of our ensemble approach compared to individual models. Our ensemble deep learning approach achieved high accuracy in identifying abnormal traffic patterns and detecting anomalies in real-time. By combining the outputs of multiple models, we were able to reduce false positives and improve the overall detection performance. This has significant implications for traffic management and intelligent transportation systems, as it allows for timely identification and response to unusual events or incidents on the road. Regarding short-term prediction of traffic flow, our ensemble approach outperformed individual models by capturing the complex temporal dependencies and nonlinear patterns in traffic data. By integrating the predictions from multiple models, we achieved more accurate and robust forecasts, enabling better traffic management and route planning. The ensemble approach also exhibited improved adaptability to changing traffic conditions and outperformed traditional prediction methods.

Future Work

While our study has made important contributions to the field of traffic flow analysis, there are several avenues for future research and improvement. Our study focused primarily on analyzing traffic flow data. However, integrating external factors such as weather conditions, events, and road construction could enhance the accuracy of anomaly detection and prediction models. Future work could investigate the integration of such factors to provide more comprehensive and contextual insights. Real-time traffic flow is dynamic, and patterns can change rapidly. Developing online learning algorithms and adaptive models that can continuously update and adjust their predictions based on new data would be valuable. This would ensure that the models remain up-to-date and can effectively handle evolving traffic conditions. Further research and development in the aforementioned areas can enhance the applicability and effectiveness of these models in real-world traffic management scenarios, leading to safer and more efficient transportation systems.

References

[1]. Ahmadi, P.; Amiri, D.; Pierre, S. (2023). An ensemble-based machine learning model for forecasting network traffic in Vanet. *IEEE Access, vol.* 7, pp. 10823–10843.

[2]. Luo, X.; Li, D.; Yang, Y.; Zhang,S.(2019). Spatiotemporal traffic flow prediction with KNN and LSTM. *J. Adv. Transp., vol*. 2019, pp. 1–10.

[3].Zhao, F.; Zeng, G.Q.; Lu, K.D. (2020). EnLSTM-WPEO: Short-term traffic flow prediction by ensemble LSTM, NNCT weight integration, and population extremal optimization.*IEEE Trans. Veh. Technol., vol*. 69, no. 1, pp. 101–113.

[4]. Zheng, H.;Li, X.; Li, Y.; Yan, Z.; Li, T. (2022). GCN-GAN integrating graph convolutional network and generative adversarial network for traffic flow prediction.*IEEE Access, vol*.10, pp.1109-3204036.

[5]. Papathanasopoulou, V.; Spyropoulou, I.; Perakis, H.; Gikas, V. (2022).A data-driven model for pedestrian behavior classification and trajectory prediction".*IEEE Access, vol*. 3, pp. 1109-3169700.

[6]Guo, S.; Lin, Y.; Li, S.; Chen, Z.; Wan, H. (2019). Deepspatial–temporal 3D convolutional neural networks for traffic data forecasting.*IEEE Trans. Intell. Transp. Syst., vol*. 20, no. 10, pp. 3913–3926.

[7]. J. Zhu, C. Huang, M. Yang, and G. P. Cheong Fung, ''Context-based Prediction for Road Traffic State using Trajectory Pattern Mining and Recurrent Convolutional Neural Networks,'' Inf. Sci., vol. 473, pp. 190–201, Jan. 2019.

[8]. Lin, Y.; Dai, X.; Li, L.; Wang, F.Y. (2018). Pattern-sensitive prediction of traffic flow based on the generative adversarial framework. *IEEE Trans. Intell. Transp. Syst., vol*. 20, no. 6, pp. 2395–2400.

[9].Peng, H.; Wang, H.; Du, B.; Bhuiyan, M.Z.A.; Ma, H.; Liu, J.; Wang, L.; Yang, Z.; Du, L.; Wang, S.; Yu, P.S. (2020). Spatial-temporal incidence dynamic graph neural networks for traffic flow forecasting.*Inf. Sci., vol.* 521, pp. 277–290.

[10].Zhao, L.; Zhou, Y.; Lu, H.; Fujita, H. (2019). Parallel computing method of deep belief networks and its application to traffic flow prediction. *Knowl.- Based Syst., vol*. 163, pp. 972–987.

[11].Li, L.; Qin, L.; Qu, X.; Zhang, J.; Wang, Y.; Ran, B. (2019). Day-ahead traffic flow forecasting based on a deep belief network optimized by the multi-objective particle swarm algorithm. *Knowl.-Based Syst., vol*. 172, pp. 1–14.

[12]. Sim, K.; Hwang, S.H. (2022). A control algorithm of an idle stop-and-go system with traffic conditions for hybrid electric vehicles. *IEEE Transactions on Intelligent Transportation Systems, vol*. 23, No. 9.

[13]. Bilotta, S.; Collini, E.; Nesi, P.; Pantaleo, G. (2022). Short-term prediction of city traffic flow via convolutional deep learning. *IEEE Access, vol*.10, pp. 1109-3217240.

[14]. Feng, B.; Xu, J.; Lin, Y.; Li, P. (2020).A period-specific combined traffic flow prediction based on travel speed clustering. *IEEE Access, vol*. 8, pp. 10.1109-2992657.

[15]. Li, W.; Qu, Z.; Song, H.; Wang, P.; Xue, B. (2020). The traffic scene understanding and prediction based on image captioning. *IEEE Access, vol*. 9, pp. 10.1109-3047091.

[16]. Sun, B.; Zhao, D.; Shi, X.; He, Y. (2021). Modeling global spatial-temporal graph attention network for traffic prediction.*IEEE Access, vol*. 9, pp.10.1109-3049556.

[17].Duan, P.; Mao, G.; Liang, W.; Zhang, D. (2019). A unified spatiotemporal model for short-term traffic flow prediction.*IEEE Trans. Intel. Transp. Syst., vol*. 20, no. 9, pp. 3212–3223.

[18].Dai, G.; Ma, C.; Xu, X. (2019). Short-term traffic flow prediction method for urban road sections based on space-time analysis and gru. *IEEE Access, vol*. 7, pp. 143025–143035.

[19].Duan, S.; Song, W.; Zio, E.; Cattani, C.; Li, M. (2021). Product technical life prediction based on multi-modes and fractional levy stable motion. mech. Syst. signal process. *vol*. 161, Dec. 2021, Art. no. 107974.

[20]. Zheng, M.; Li, T.; Zhu, R.; Chen, J.; Ma, Z.; Tang, M.; Cui, Z.; Wang, Z.(2019). Traffic Accident's Severity Prediction: A Deep-Learning Approach-Based CNN Network.*IEEE Access, vol*. 7, pp.10.1109-2903319.