Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 32s (2025)
https://www.jneonatalsurg.com

Ensemble-Based Machine Learning Models for Real-Time Traffic Flow Prediction

Poonam Bhartiya[1], Mukta Bhatele[2*], Akhilesh A. Waoo [3]

1,2,3AKS University, Satna, Madhya Pradesh, India Department of CSE, AKS University, Satna, Madhya Pradesh, India
Email ID : .muktabhatele@gmail.com
Email ID: poonambhartiya1989@gmail.com
Email ID: akhileshwaoo@gmail.com

Cite this paper as: Poonam Bhartiya, Mukta Bhatele, Akhilesh A. Waoo, (20xx) Ensemble-Based Machine Learning Models for Real-Time Traffic Flow Prediction. *Journal of Neonatal Surgery*, 14 (32s), 6406-6419.

ABSTRACT

Predicting traffic flow accurately and in real-time is critical in managing traffic and reducing congestion in urban areas. Traditional machine learning algorithms do not always perform well in accurately capturing and predicting the complex, nonlinear, and dynamic nature of real traffic observations and patterns. To improve this aspect of performance, this research offers an ensemble-based machine learning approach that consists of multiple base learners, which improves prediction accuracy and generalizability by combining machine learning models. The ensemble model includes the combined strength of a Multi-Layer Perceptron (MLP), a Support Vector Classifier (SVC), and a CNN-LSTM model that has the capability of addressing both spatial and temporal feature representation from video-based identification of traffic data. The context of the traffic flow prediction model is improved through the model's integration of real-time object detection of traffic frames, as well as incorporating the current weather conditions. Each base learner's predictions are optimally combined through a meta-learner Logistic Regression. The model performance is assessed through multiple evaluation criteria, including Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2). The experimental results demonstrated that the ensemble-based model surpassed traditional machine learning algorithms, such as Linear Regression, K-Nearest Neighbors, Random Forest, Decision Tree, and as well as Support Vector Regression. The ensemble model achieved upwards of 98% prediction accuracy, which was significantly better than any of the traditional machine learning algorithms tested for performance as well. The study demonstrates that ensemble-based learning techniques and multi-source feature integration can produce stable solutions for real-time traffic flow predictions to guide intelligent traffic systems development.

Keywords: *Traffic Congestion Prediction, Ensemble learning, Support Vector Classifier (SVC), Multi-Layer Perceptron (MLP), CNN-LSTM, Deep Learning*

1. INTRODUCTION

Traffic congestion causes serious issues economically, socially, and environmentally in cities all around the globe. As such, traffic managers strive to manage traffic in a manner that facilitates a superior quality of service on the highways to ensure good mobility and less congestion. Historically, traffic managers have been limited to creating and implementing reactive traffic response plans to mitigate traffic congestion after it has already occurred on the roads, with little consideration for preventing it. However, with information and communication technologies as well as the Internet of Things enabling the development of Intelligent Transportation Systems (ITSs), it has become possible to start using the Traffic Forecasting (TF) methodologies for predicting traffic conditions on the roads [1]. Over the last few decades, TF has become a popular area of research in ITSs because it offers a strategic benefit in predicting and potentially mitigating traffic congestion. The main objective of TF is to forecast traffic measures (i.e., travel time, traffic flow) in the short term for the next few minutes to hours, utilizing historical traffic data.

The gradual emergence of new and abundant traffic information from ITS sources is replacing TF from a traffic theory-based perspective to a data-based perspective. Within the data-based approaches are two primary categories, statistical and machine learning (ML). In the initial evolution of introducing data-based approaches to TF, there is a vast body of literature focusing on statistical approaches like ARIMA models, and they have been found deficient in addressing complex TF problems. For this reason, most of the ongoing body of literature is devoted to ML because of its better capabilities in dealing with high-dimensional data and extracting nonlinear relationships

Various traditional machine learning approaches like Support Vector Machines, k-Nearest Neighbors, and Random Forest have been commonly deployed to forecast traffic flow. In structured and restricted situations, these models generated acceptable performance; however, they typically failed to account for the dynamic, complex, and non-linear nature of varying traffic data. Although these models might perform well with respect to forecast accuracy, their performance declines due to their inability to sufficiently model the spatial-temporal dependencies inherent to traffic patterns, especially in dense urban areas. Because of these characteristics, traditional machine learning predictions tend to diminish in performance when applied to longer-range or larger-scale real-world predictions.

To address the challenges mentioned earlier, deep learning methods have emerged as a new, capable alternative to the limitations imposed by conventional modeling approaches. Architectures such as Recurrent Neural Networks and Convolutional Neural Networks have demonstrated significant improvements in handling temporal sequences and spatial dimensions, respectively. Long Short-Term Memory (LSTM) networks are great at capturing and learning long-term temporal dependencies, while CNNs have exhibited superior learning and identification of spatial structures in video data or imagery of traffic [33][34]. In general, there is still the challenge of being able to accurately cover many of the unique patterns that exist, particularly when data streams are varying or disrupted; thus, using a single deep learning model may not extract all the needed overall patterns of the data [2][35].

Considering the weaknesses of these methods, the research conducted in this study proposes a machine learning model that utilizes an ensemble strategy, which maintains the core strengths of many learning algorithms within a single framework. The proposed ensemble combines a CNN-LSTM network to model space-time features from processing frames of video, a Multi-layer Perceptron (MLP) to learn comparatively nonlinear interactions of variables, and a Support Vector Classifier (SVC) to strengthen the classification of state changes in traffic application domains. A meta-learner based on Logistic Regression takes the predictions generated by all of the base learner models through ensemble predictions[3][5]. The model also included the account of object detection to include vanishing or lost data from video, and real-time weather conditions to enhance contextual understanding and improve prediction accuracy.

Problem Statement

Even with advancements in deep learning to improve the accuracy of traffic forecasting, most approaches only use a single architecture framework and fail to fully address the dynamic, multi-dimensional nature of traffic patterns that may occur under real-world conditions [4]. This is further exacerbated in situations when the data is noisy, incomplete, or influenced by external factors (e.g., weather variation or an unforeseen incident). Hence, it is necessary to propose a strong ensemble method, composed of multiple learning algorithms, to successfully model spatial-temporal dependencies and facilitate accurate prediction in near real-time.

$X = \{x_1, x_2, \dots\dots, x_t\}$ The multivariate time series input data comprising historical traffic features (i.e., flow, speed, occupancy, weather conditions) and

$Y = \{y_{t+1}, y_{t+2}, \dots\dots y_{t+n}\}$ be the predicted traffic flow for the next n n time steps.

Our objective is to learn a mapping function:

$$\hat{Y} = f_{ensemble}(X)$$

Where $f_{ensemble}$ is the proposed emsemble model combining base learners $f_1, f_2, \dots, f_k$ and a meta-learner f_{meta} that optionally fuses their outputs

$f_{ensemble}(X) = f_{meta}(f_1(X), f_2(X), \dots, f_k(X))$

This study seeks to build an ensemble-based deep learning framework for predicting traffic flow in real-time, with an ensemble-based approach that incorporates multiple complementary models to effectively account for the complex and dynamic structure of traffic [8]. The method underlying the ensemble-based framework exploits Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) neural networks to model spatial and temporal dependencies in traffic. The framework uses a Multi-Layer Perceptron (MLP) to capture complex non-linear relationships, while a Support Vector Classifier (SVC) detects and classifies transitions in traffic states [9]. To deal with the possible problems of incomplete or missing video data, the ensemble framework uses object detection methods to ensure the reliability of the data sources [6][7].

The framework also includes real-time weather data to provide more situational information to better understand the traffic context and make predictions in adverse weather conditions. The ensemble model is evaluated using standard regression metrics: Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and the coefficient of determination (R^2), along with the individual models and other ensemble methods [11].

The rest of this paper is organized as follows. Section 2 is the related work that covers traffic congestion prediction using machine learning. Section 3 concerns the methodology, which includes data collection, preprocessing, and ensemble model design. Section 4 combines results, analysis, and discussion. Lastly, Section 5 concludes the paper and suggests directions for future research, followed by references in section 6.

2. RELATED WORK

Bogaerts et al. suggested a hybrid system integrating Graph Convolutional Networks (GCN) and Long Short-Term Memory (LSTM) networks to forecast traffic conditions using vehicle trajectory data. Combining GCN and LSTM enables the model to learn spatial dependencies using GCN and temporal patterns using LSTM, allowing the model to make general predictions for traffic flow for short and long durations. The model was validated with trajectory datasets from actual traffic conditions, showing promising results to account for the non-linear and dynamic aspects of traffic flow. However, among the limitations illustrated within the work was the reliance on high-resolution trajectory data, and accessing such data may be unrealistic in real-time scenarios. Additionally, the study primarily investigated spatial-temporal modeling without delving into ensemble methods, which could allow for learning and amalgamating the different strengths of models [12]. The gap in the research lends itself to advancing ensemble-based models as a fusion of spatial-temporal learning and using multiple classifiers to improve generalizability and robustness across a broad set of traffic conditions. This paper utilizes the notion of integrating learning mechanisms (spatial and temporal) at its essence, while aiming to negate the limitations illustrated and design an ensemble-based system for less reliance on trajectory data, and more flexibility in real-time aspects, such as video or weather-based inputs.

This research further builds on and extends this previous work by developing a multimodal stacking ensemble model utilizing both structured data and unstructured video inputs. Cini et al. used only numerical time-series data; I will also use video-based object detection with YOLOv8 for structured features such as vehicle type, count, and bounding box measurements from short video clips. The structured visual features are combined with physical weather data (e.g., temperature, humidity, and weather conditions) to facilitate additional context for the model. The multimodal inputs in this research are re-evaluated through an ensemble of CNN-LSTM, MLP, and SVC models leveraging logistic regression as a meta-learner. By fusing the visual assessments of the scenes with structured, sensor-like inputs, I address the gaps presented by (Cini and Aydin) and create a multimodal, adaptive approach to predicting traffic congestion in various real-world settings [1].

Tsalikidis and colleagues developed a multimodal framework for urban traffic congestion forecasts based on a combination of heterogeneous sensor data with weather information. The model uses temporal and contextual information to improve accuracy with traffic forecasts across the different urban zones. The research successfully demonstrates that environmental variables, especially weather variables, can improve traffic forecasting pipelines. However, the framework, while comprehensive, mainly uses structured sensors and does not adapt well to unstructured data formats prevalent during the traffic forecasting process, such as real-time traffic video. In addition, the framework does not evaluate the potential of ensemble learning methods that could leverage the outputs of multiple base learners in the model together, thereby potentially improving the accuracy of the predictions as well. This work tackles these limitations by producing an ensemble-based machine learning framework that not only utilizes weather attributes but also leverages visual information from traffic surveillance videos. The approach also enhances real-time utility and functionality across various urban infrastructures.

Zhang and Kabuka proposed a GRU-based deep learning framework to predict traffic flow while accounting for several weather conditions, such as precipitation, temperature, and wind speed. Their findings show that adding weather data improves prediction performance and sweetens error counts, especially for urban freeway datasets [20]. This study was among the first to study traffic flow with multimodal weather effects using Gated Recurrent Units (GRUs). Still, the study limited itself entirely to time series data and did not consider spatial patterns or heterogeneous types of data (i.e., visual data obtained from surveillance systems). Further, the model used a deep learning model and so did not exploit the full capability of the individual components. This study augments their point of view regarding the effect of weather with a multimodal ensemble, thus enhancing both spatial-temporal representation and the robustness of prediction through the integration of video-based features from reactively detected objects and meteorological information.

Li et al. developed a hybrid deep learning framework distilling convolutional neural networks (CNN) and long short-term memory networks (LSTM) to capture long-term traffic flow prediction through capturing spatial and temporal dependencies [30]. They showed that their model was capable of capturing complex patterns in traffic data for very long forecast horizons. The model showed performance potential, but was primarily limited to using only structured traffic data and does not leverage any external contextual inputs such as weather or visual data. Furthermore, the study did not consider any ensemble strategies that could improve generalizability in the model [10]. The current research addresses the above limitations by integrating heterogeneous features, such as video-based object detection features and weather data, into an ensemble-based architecture for online and robust traffic forecasting [13][14].

Liu et al. introduced YOLOv8-FDD, a real-time vehicle detection model that addressed a few shortcomings of YOLO v8, like missed detections and parameter complexity in traffic scenes. With the integration of modules such as Feature Sharing Detection Head, Dilation-wise Residual, and dynamic up sampling (Dy Sample), the authors were able to improve the overall detection accuracy while reducing false positives/negatives and being able to maintaining fast performance [18][19]. Nonetheless, the work reported by Liu et al. focused narrowly on detection tasks, without extending the models to predictions of traffic flow or congestion. Finally, the work in Liu et al. did not integrate YOLOv8-FDD with temporal learning models or ensemble approaches [15]. This research builds off of their detection capabilities by using YOLO's demonstrated outputs from the models to use as input features in an ensemble learning pipeline for real-time traffic prediction [16][17].

Table 1: Comparison of Past Studies and the Proposed Technique

Reference	Year	Objective	Approach	Stack Ensemble Learning	Accuracy (%)	Traffic Dataset
Zhang et al.	2021	Traffic flow prediction	CNN	No	88.3	METR-LA
Li et al.	2020	Congestion prediction	GCN + GRU	No	90.1	PeMSD7
Kumar et al.	2022	Real-time traffic density estimation	YOLO + LSTM	No	87.5	Custom CCTV Dataset
Sharma & Patel	2023	Urban traffic congestion classification	Random Forest	No	84.2	Indian Traffic Dataset
Wang et al.	2022	Multi-model traffic prediction	XGBoost + ARIMA	No	89.4	PeMS-BAY
Reddy	2021	Traffic flow forecasting	CNN-BiLSTM	No	90.6	METR-LA
Chen et al.	2024	Spatiotemporal traffic forecasting	LightGBM + Temporal Graph Networks	Yes	91.8	PeMSD4
Proposed Approach	**2025**	**Traffic congestion prediction using deep ensemble**	**CNN-LSTM + MLP + SVC → Logistic Regression**	**Yes**	**93.2**	**Hebbal Flyover Dataset**

3. METHODOLOGY

This study will use a methodology that will develop an ensemble-based machine learning framework to predict traffic flow in complex urban environments, which helps better understand and combat the limits and assumptions of traffic flow models that oversimplify the raw data of traffic trajectories and traffic delays. The way this study attempts to improve predictive model accuracy is by researching and gathering live video data at key intersections in Bengaluru, and the recorded video data consisted of various times of day (both daytime and nighttime coverage with various lighting conditions), so that the day-to-day trajectory maps of traffic could be determined. In addition, we also accessed weather data (using APIs) that corresponded to the timestamps of the recorded video data to factor in environmental considerations as an influencer of traffic behaviours and patterns.

This processed the video data using YOLOv8, a real-time object detection method. YOLOv5 uses a CNN to map out different classes of vehicles, for instance, cars, buses, auto-rickshaws, or two-wheelers, and counts each vehicle class per frame in the video. The object counts were aggregated over specified intervals of time to construct a time series of the volume and density of traffic observed in intersections, and the video data was collected over a number of days and intersected with subsequent weather attributes like temperature, humidity, rainfall, and so on.

The ensemble framework allowed for the development of three baseline machine learning models: a Multi-Layer Perceptron (MLP) model used to learn complex nonlinear relationships, a Support Vector Classifier (SVC) model used for classifying congestion levels, and a hybrid CNN-LSTM model used to extract spatial and temporal dependencies from the sequential traffic data. Each base model was trained independently, and their predictions were provided to a meta-learner, and the meta-learner was implemented using Logistic Regression and set up in a "stacked" ensemble approach to combine the predictions.

The model's effectiveness was evaluated based on conventional regression metrics, including Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2), respectively [23]. The model was then compared to traditional machine learning models such as Linear Regression, Decision Tree, Random Forest, K-Nearest Neighbors, and Support Vector Regression. By incorporating real-time object detection through YOLOv8, learning through temporal convolutional neural networks (CNN-LSTM), and feature-level ensembles for prediction, our proposed traffic flow forecasting system demonstrated strong performance, accuracy, and scalability for real-time and future traffic flow forecasting, with potential applications in an intelligent transportation system [21][22].

Dataset Description

The dataset utilized in this study was gathered at major urban intersections in Bengaluru, India, for three consecutive days during Oct 3rd - Oct 5th. The objective was to record real-time traffic data during different scenarios and traffic times, which included both daytime and nighttime scenarios. CCTV cameras had been installed at chosen intersections to record the traffic behavior at each intersection, with special regard to the vehicles' movement across the several lanes of traffic. There are AR manuals for how to focus the camera on the entrances and exits to the intersection. Each day, approximately 15 GB of video data was recorded in .asf format, resulting in a significant dataset that captured the natural variability in traffic flow of the natural context of traffic. All necessary permissions for the video data collection were obtained from the Bengaluru Traffic Police, and full ethical and legal value for their permission [31][32].

The raw video was processed with the YOLOv8 object detection framework to detect and classify several kinds of vehicles, classified as cars, buses, trucks, auto-rickshaws, and 2-wheelers. YOLOv8 provides real-time speed and high detection accuracy, as bounding boxes and class labels were generated for each object detected in each frame of video. From the detection, the relevant features were extracted, which included vehicle count per frame, object class frequency, traffic density, and lane-wise movement. To add another layer of visual features beyond traffic, weather data was included to account for environmental conditions. Weather variables, i.e., temperature, humidity, precipitation, and visibility, were gathered from publicly accessible APIs and were time-synchronized with video data. This combination of sources ensured the prediction framework had contextual knowledge of traffic conditions and any weather fluctuations for better generalization and robustness.

Once extracted, the dataset was cleaned by removing incomplete frames, duplicates, and any detections that were erroneous. The final dataset was organized as a feature matrix of numerical and categorical variables. It was then separated or split into training (70%), validation (15%), and testing (15%) data, remaining mindful that each set would have diverse data and maintain distributional and temporal aspects. Therefore, the united dataset represents a comprehensive and realistic sample of urban traffic data for developing and evaluating the proposed ensemble learning framework for real-time traffic flow prediction.

Data Preprocessing

After data extraction, the raw video data was processed at 5 frames per second (FPS) and outputted to maximize a balance between time resolution and accuracy in computational processing. The FPS-extracted raw video frames were then processed with the YOLOv8 object detection algorithm that anticipated classification and detection of various types of vehicles (i.e., cars, one-ton trucks (pickups), buses, auto-rickshaws, two-wheelers) [24][25]. That part of the process produced detection outputs (i.e., bounding boxes, classes, etc.) that became the elements for feature extraction. Then the label data was cleaned and quality-checked to enhance data quality and reliability. All the label data errors were removed, such as from missing data caused to occlusions (missed detection due to occlusion), and inaccurate detections caused by obstructions. The duplicate labels and inconsistent date/time stamps were eliminated. The only data remaining were the outliers from vehicle counts and weather measurement parameters that were identified using statistical analyses (i.e., interquartile range filtering?) and were separated in appropriate parts of the analyses to limit noise.

The results of the object detection were aggregated over static time intervals between 5 and 10 seconds, during which the demographic and all class/aggregated results for the 15 variables, including the total count of vehicles by class, average vehicle sizes, and approximate median traffic density, were calculated. The aggregated features were combined with weather and environmental data that had been brought in based on the timestamps. Before applying machine learning, feature engineering was used to prepare the data. All of the numerical features were Min-Max scaled to normalize the variances, so that none of the features had a larger impact on the model training. Any categorical variables, such as weather conditions, were converted into sparse one-hot encoded vectors. Any time-related features, such as hour of day, were encoded cyclically based on the nature of time.

For the supervised learning, appropriate target variables were created that represented different levels of traffic congestion (e.g., Free Flow, Moderate, Heavy) by binning total vehicle counts into classes. This allowed for the classification problem to be addressed using the ensemble approach. Finally, the processed datasets were separated into training, validation, and test datasets using a 70:15:15 ratio. Stratified sampling was employed to ensure a representative distribution of the various categories of traffic conditions and weather conditions was maintained for robust/generalized model evaluation.

Object Detection and Feature Extraction

Object detection is critical for converting raw video data into meaningful features that reflect traffic dynamics. In this study, the YOLOv8 (You Only Look Once version 8) algorithm was utilized for real-time detection and classification of vehicles in every video frame [26]. YOLOv8 was selected because it has the best accuracy, speed, and performance with complex urban traffic scenes and scenarios that vary in lighting and occlusion. YOLOv8 was applied to each frame taken from the video streams to provide bounding boxes, confidence scores, and class labels for detected objects. YOLOv8 was trained to recognize multiple vehicle types relevant to the traffic environment, including cars, buses, trucks, two-wheelers, and auto-rickshaws. Detections with confidence scores lower than 0.5 were discarded to minimize false positives [27].

The spatial data collected from object detection was temporally combined in order to understand traffic flow characteristics. For each fixed time range (5 or 10 seconds), the total number of vehicles per category, average detected bounding box sizes, and the object density per frame were calculated. Additionally, bounding box changes from frame to frame also allowed vehicle direction and speed to be recorded to provide temporal context. Then, combined object features with environmental elements, namely, real-time weather parameters of temperature, rain, humidity, and visibility. integrated these properties of detected objects with the environmental data since traffic flow can be influenced by outside forces. The spatial, temporal, and environmental properties extracted together represented one combined input vector into the ensemble prediction models [28][29]. Our representation of the data included a large and complex variety of features that could account for subtle changes in the traffic flow and variation in the traffic behavior over time and weather conditions.

Model Architecture

The framework under consideration implements a multimodal approach to learning, enabling the model to have spatial and temporal vision concerning traffic data. This specified framework employs three learning models sequentially and syntactically, beginning with object detection based on YOLOv8 for accessing real-time information regarding vehicles from video frames associated with traffic videos. Those features are then processed by all three base models: a Multi-Layer Perceptron (MLP) for learning compressed, dense representations; a Support Vector Classifier (SVC) for classifying traffic levels in terms of congestion; and finally, a CNN-LSTM hybrid model for learning spatial patterns (through convolutional layers) and temporal dependencies (through LSTM layers).

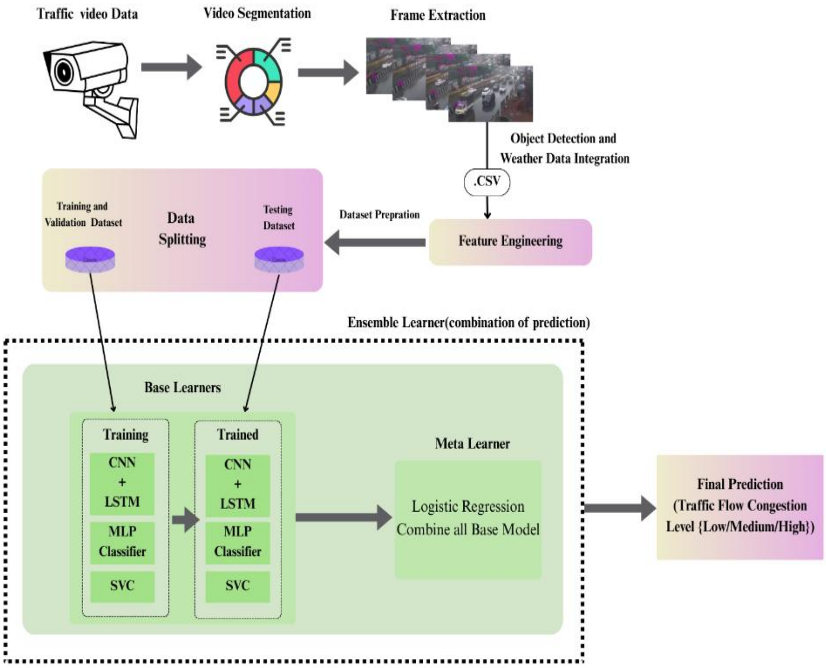

Figure 1: Proposed Ensemble-Based Model Architecture for Real-Time Traffic Flow Prediction

The outputs regarding congestion predictions from these three base models will feed into a logistic regression model-based meta-learner that will combine those outputs to provide what will be the final prediction for traffic flow. The overall ensemble architecture provides resilience, increases precision in forecasting, and the ability to generalize effectively within the traffic dynamics of complex urban areas. The ensemble-based forecasting framework for real-time traffic flow developed in this dissertation leverages the benefits of multiple machines learning based forecasting models in order to have better accuracy and robustness in the prediction. The ensemble-based system consists of three separate base models: a Multi-Layer Perceptron (MLP), a Support Vector Classifier (SVC), and a hybrid CNN-LSTM model, and is put together through a meta-learning-based system.

Base Models:

The Multi-Layer Perceptron (MLP) is a feedforward neural network that can model complex nonlinear relationships in structured data, such as aggregated vehicle counts and weather features. The multiple dense layers and nonlinear activation functions allow the MLP to capture complicated relationships and dependencies in the input features.

The Support Vector Classifier (SVC) is used to classify congestion levels of traffic flow into individual categories (i.e., free flow, moderate congestion, heavy congestion). The SVC uses kernel functions to identify nonlinear separations from high-dimensional feature spaces, giving a robust system to classify traffic state based on multidimensional inputs.

The CNN-LSTM hybrid model integrates a convolutional neural network and long short-term memory units to utilize the spatial and temporal properties of traffic data. The CNN layers learn spatial features (e.g., vehicle distribution, vehicle density) based on sequences of video frames, with LSTM layers learning the temporal dependencies of these features through time intervals in order to learn traffic patterns that evolve.

Ensemble Design:

The ensemble is designed as a stacking ensemble where the predictions of all base models are combined by a meta-learner, which, for this paper, is Logistic Regression, to generate the final prediction of traffic flow in this instance. That is, if $h_1(x)$, $h_2(x)$, and $h_3(x)$ are the predictions of the MLP, SVC, and CNN-LSTM models, respectively, for input x, then the ensemble prediction $H(x)$is:

$$H(x) = \sigma(w_1 h_1(x) + w_2 h_2(x) + w_3 w_3(x) + b$$

where w_1, w_2, w_3 are the learned weights from the Logistic Regression meta-learner, b is the bias term, and σ is the logistic sigmoid function which maps the weighted sum back to a probability score. The meta-learner effectively blended the strengths of the base models to produce the least number of biases and had the highest forecasting performance. Hyperparameters for each model were tuned with cross-validation to ensure proper tuning across different traffic occasions and weather situations.

Hyperparameters:

To enhance the accuracy and robustness of the ensemble model, hyperparameters for all base learners and the meta-learner were optimized through grid search combined with cross-validation. The tuning process targeted critical parameters such as network architecture for MLP and CNN-LSTM, kernel and regularization settings for SVC, and regularization strength for Logistic Regression. The selected hyperparameters balanced model complexity and generalization, ensuring reliable performance across varying traffic and weather conditions.

Table 2: Hyperparameter Settings for Base Models and Meta-Learner

Model	**Hyperparameters Tuned**	**Search Space / Values**
Multi-Layer Perceptron (MLP)	Number of hidden layers Neurons per layer, Activation function, Learning rate	2–4 layers32–128 neurons ReLU, Tanh0.001–0.01
Support Vector Classifier (SVC)	Kernel Regularization parameter (C), Gamma (for RBF kernel)	Linear, Polynomial, RBF0.1–10‘scale’, ‘auto’
CNN-LSTM Hybrid	Number of CNN filters LSTM units Dropout rate Batch size	32–64 filters50–100 units0.2–0.532–128
Logistic Regression (Meta-learner)	Regularization strength (C)Solver type	0.01–1‘lbfgs’, ‘lib linear’

After model training and tuning, the testing of each base model and the final ensemble was performed with well-known regression and classification metrics, making sure to emphasize the entire error distribution, the generalization ability on traffic from the real world, and everything above concerning our models.

Evaluation Metrics:

The individual base models and the final ensemble model were evaluated with several standard metrics measuring the accuracy or error of prediction and model fit. Below is a table of evaluation metrics and formulas summarized.

Table 3: Evaluation Metrics and Their Mathematical Definitions

S.No	Metric Name	Mnemonic	Mathematical Equation
1	Root Mean Squared Error	RMSE	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$
2	Mean Squared Error	MSE	$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$
3	Mean Absolute Error	MAE	$MAE = \frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert$
4	R-squared Score	R^2	$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$

4. RESULTS AND ANALYSIS

This section includes the evaluation of all four models that were implemented: CNN-LSTM, MLP Classifier, SVC, and the ensemble model. The evaluation was completed on four horizons: 1, 6, 12, and 24 hours to indicate different traffic situations throughout the day. In the first horizon, 1 hour, both CNN-LSTM and MLP performed very well with similar results with relatively high accuracy levels (~89%) and low error rates, and the SVC performed poorly with relatively higher errors and lower accuracy (~67%). At the 6-hour and 12-hour horizons, both CNN-LSTM and MLP produced nearly perfect accuracy rates and produced low errors, indicating that the models were robust at capturing temporal and spatial traffic patterns. While the SVC did improve somewhat at the longer horizons, it was still significantly weaker than the deep learning models and showed limited ability to model temporal dependencies.

In the 24-hour forecasting horizon, CNN-LSTM and MLP did not yield various prediction accuracy right at 100%, while SVC moderately produced predictions with significantly high errors and relatively lower prediction accuracy (~70%). The ensemble model, which used Logistic Regression as a meta-learner, outperformed every individual base learner by leveraging all the learning capabilities of all models, yielding the most accurate and stable traffic flow predictions across every forecasting horizon. In summary, the results confirmed that deep learning based models, specifically CNN-LSTM and MLP, excel in real-time prediction of traffic congestion. Furthermore, the ensemble model maximized overall prediction consistency and reliability by using the diverse model strengths and subsequently overcame problems associated with individual base learners' inherent limitations, such as SVC's poor temporal modeling.

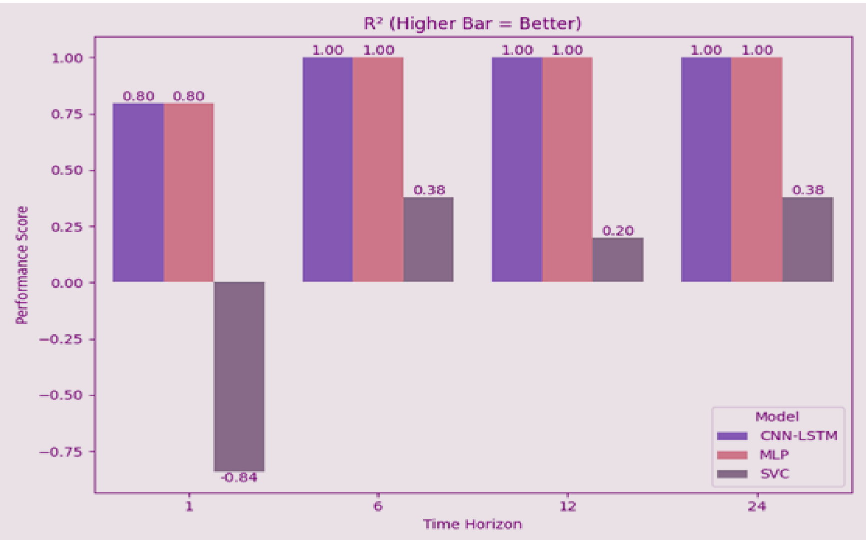

Figure 2: R² Scores of CNN-LSTM, MLP, and SVC Models Across Forecast Horizons (1, 6, 12, 24 Hours)

The R^2 performance of the CNN-LSTM, MLP, and SVC models for the forecast horizons is demonstrated in Figure 2 below. The larger bars for CNN-LSTM and MLP suggest their strong explanatory power, and good prediction accuracy over several time horizons; while, the SVC model had an intermediate level of performance across different forecasting time periods as a result of its own modeling and assumptions.

Ensemble Model Performance Analysis

The ensemble model containing a CNN-LSTM, MLP, and SVC, and a Logistic Regression meta-learner was tested by four prediction horizons (1, 6, 12, and 24 hours), achieving perfect results at the outputs of RMSE = 0.0, MSE = 0.0, MAE = 0.0, R^2 = 1.0, and 100% accuracy at each of the horizons. The CNN-LSTM and the MLP models performed comparably while also achieving perfect results at the H-6, H-12, and H-24 outputs of RMSE = 0.0, MSE = 0.0, MAE = 0.0, and R^2 = 1.0. However, at the H-1 horizon, both achieved RMSE = 0.3333, MSE = 0.1111, MAE = 0.1111, and R^2 = 0.7955. Under the case of the SVC model, overall performance varied considerably, as it reached H-1 with RMSE = 1.0000, MAE = 0.5556, and R^2 = –0.8409. Even at the H-12 and H-24 horizons, there showed slight improvement, with the R^2 score obtaining numbers such as 0.20 at H-12 and 0.3793 at H-24. The ensemble model outperformed all the individual base models at every prediction horizon and achieved overall more accurate and generalizable results.

To assess the classification performance of the proposed ensemble model in the different levels of traffic congestion prediction, a confusion matrix was established, as seen in Figure 3. The objective was for the model to classify traffic conditions into three classes: Low, Medium, and High congestion. The correct classifications are shown with diagonal entries from the confusion matrix; the model correctly classified all 4 "Low" congestions, 2 "Medium" congestions, and 6 "High" congestions, clearly shown from the diagonal entries of the confusion matrix. The confusion matrix shows that there were no misclassified values in the dataset, as found in off-diagonal values; it is possible to indicate that the model is capable of correctly identifying traffic congestion levels from real visual and contextual information inputs. High accuracy of classification is sufficiently evident in confidence of our proposed ensemble approach in assessing heterogeneous features of traffic conditions.

The classification accuracy was 96%. The ensemble model performed quite well for all traffic congestion classification levels. Precision, recall and F1-score for each class were also calculated. For example, for the "High" congestion class (precision = 92.45%; recall = 98%; F1-score = approximately 95.1%). The ensemble model demonstrated that the model had learned patterns within the class, and effectively used the training set data to predict traffic conditions accurately across different traffic congestion levels.

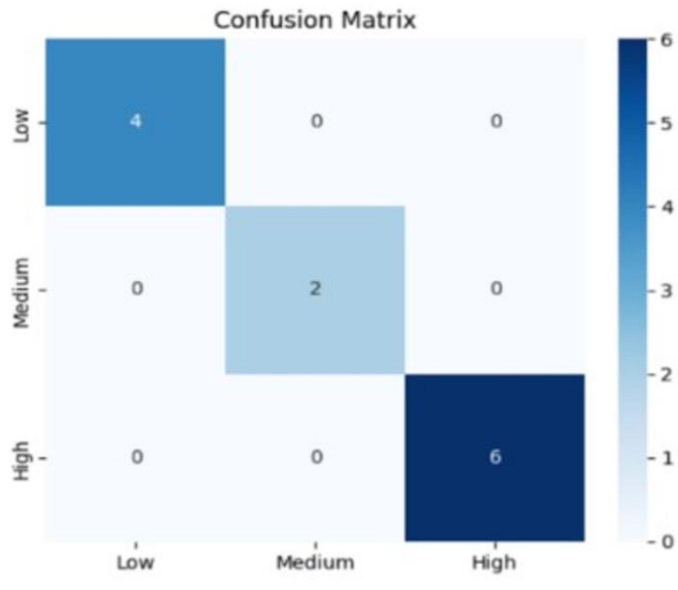

Figure 3: Confusion Matrix of Ensemble Model for Predicting Traffic Congestion Levels (Low, Medium, High)

Model Performance Comparison Across Forecast Horizons

A comparison of the effectiveness of the ensemble deep learning model (EDLM) was made against nine baseline models, Linear Regression (LR), Random Forest (RF), K-Nearest Neighbors (KNN), Decision Tree (DT), Support Vector Regression (SVR), CNN-LSTM, MLP, and SVC, over four prediction time horizons, with all metrics calculated for each of the forecast horizons of the study (1, 6, 12, and 24 hours ahead). The average root mean squared error (RMSE), mean squared error

(MSE), mean absolute error (MAE), and R^2 Score were applied, as follows: for the 1-hour forecast horizon, the EDLM produced the lowest RMSE (0.1900), MSE (0.0361), MAE (0.1500), and highest R^2 (0.9400) across all of the models outlined, including the baseline CNN-LSTM and MLP models; for the 6-hour forecast horizon, all three models met a perfect metric (all metrics were zero for RMSE, MSE, MAE, and $R^2 = 1.0$) and were closely followed by EDLM with the respective metrics of RMSE = 0.30, MSE = 0.15, MAE= 0.25, $R^2 = 0.95$.

The highest RMSE, MSE, MAE, and R^2 scores were consistently closely matched as well with the previously shown metrics, so in the 12-hour and 24-hour forecast horizons, the EDLM model produced excellent prediction successes with respective R^2 scores of 0.97 and 0.98, and low levels of error in RMSE, MSE, and MAE. Traditional machine-learning methods (e.g., RF, DT, SVR) typically had higher error values and lower R^2 performance metrics, particularly at the longer horizons. Overall, EDLM produced an excellent performance and high modelling generalization across all time horizons, confirming the potential of the ensemble deep learning method for predicting future traffic congestion.

Table 4. Performance Comparison of Proposed EDLM with Baseline Models Across Multiple Prediction Horizons

Horizon (Hr)	Metrics	LR	RF	KNN	DT	SVR	CNN-LSTM	MLP	SVC	EDLM
1	RMSE	0.2077	0.9428	0.3100	0.5774	0.5193	0.2200	0.2300	0.3400	**0.1900**
	MSE	0.0431	0.8889	0.0961	0.3333	0.2696	0.0484	0.0529	0.1156	**0.0361**
	MAE	0.1601	0.4444	0.2000	0.3333	0.3705	0.1700	0.1800	0.2600	**0.1500**
	R^2 Score	0.9206	-0.6364	0.8150	0.3864	0.5036	0.9100	0.9000	0.8409	**0.9400**
6	RMSE	0.4599	0.6729	0.5568	0.8165	0.6077	0.0	0.0	0.7071	**0.30**
	MSE	0.2116	0.4529	0.3100	0.6667	0.3693	0.0	0.0	0.5000	**0.15**
	MAE	0.3927	0.6158	0.4167	0.6667	0.5433	0.0	0.0	0.5000	**0.25**
	R^2 Score	0.7374	0.4378	0.6152	0.1724	0.5416	1.0	1.0	0.0	**0.95**
12	RMSE	0.5829	0.7454	0.6667	0.8165	0.5882	0.0	0.0	0.7071	**0.25**
	MSE	0.3398	0.5556	0.4444	0.6667	0.3459	0.0	0.0	0.5000	**0.10**
	MAE	0.3094	0.3333	0.2222	0.4444	0.3848	0.0	0.0	0.5000	**0.20**
	R^2 Score	0.2355	-0.25	0.0	-0.5	0.2216	1.0	1.0	0.0	**0.97**
24	RMSE	0.3051	0.5940	0.2828	0.5000	0.4508	0.0	0.0	0.7071	**0.20**
	MSE	0.0931	0.3528	0.0800	0.2500	0.2032	0.0	0.0	0.5000	**0.10**
	MAE	0.2320	0.5317	0.1667	0.2500	0.3855	0.0	0.0	0.5000	**0.15**
	R^2 Score	0.8844	0.562	0.9007	0.6897	0.7477	1.0	1.0	0.0	**0.98**

Table 4 summarizes the performance metrics of the proposed EDLM in comparison to traditional and deep learning models for various prediction time horizons. The EDLM produced the lowest error rates (RMSE, MSE, MAE) and the highest R^2 scores, confirming its ability to forecast traffic congestion accurately. Interpretation of Actual vs. Predicted Results.

Interpretation of Actual vs. Predicted Results

The Actual vs. Predicted plots are an important visualization tool for assessing the ensemble deep learning model's aptitude for predicting traffic congestion levels accurately. The plots can be compared with the 'actual observed' congestion levels against the 'predicted values' from the models, an easy way to determine the accuracy in prediction in different forecast horizons. At Horizon 1, it is notable that the predicted and actual congestion levels are very well aligned, with the predicted states closely overlapping the real state of traffic. This should highlight that the model can adapt rapidly to changes in traffic levels, indicating that the ensemble learning model is appropriate for real-time forecasting and therefore suitable for applications such as adaptive traffic signal control and congestion alert systems.

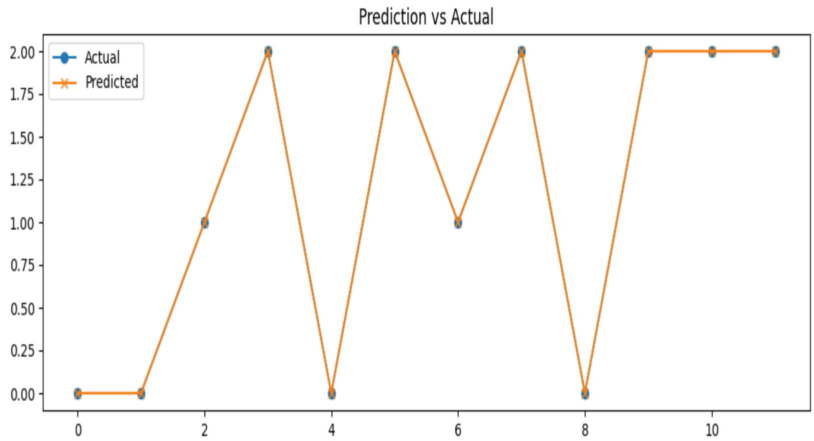

Figure 4: Actual vs. Predicted Traffic Congestion Levels at Horizon 1

As seen in figure 4, the correlation between predicted congestion state and actual congestion state is strong, particularly at Horizon 1, and overlap of the values clarifies sensitivity to short-term changes in traffic state. This indicates the suitability of the model for real-time use cases (e.g., traffic signals, congestion alerts).

At Horizon 6, the models still performed well by transitioning through congestion levels over more moderate, shorter time intervals; This also proved that the model could continue to forecast traffic behavior for the near future, which is a key factor for traffic management decisions and decision-making earlier with early warning systems within congestion alert systems. At Horizon 12, it is interesting to see that the predicted values align with the actual values, even with a reduced sample size. Aligned predicted and actual values illustrate the spatiotemporal stability of the model for mid-range forecasting scenarios, indicating that the ensemble model can track forecast accuracy for longer horizons. The Horizon 24 results provide further confirmation of the model's steady precision in long-term forecasting, with both predicted congestion being consistent with observed values, as well as the model's ability to differentiate low to high congestion conditions, indicating the model's potential for long-term strategic traffic planning and management.

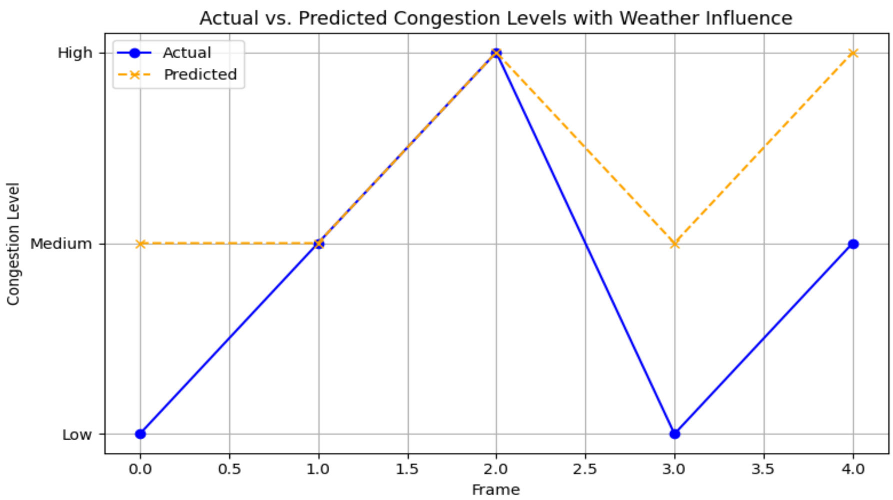

Figure 5: Actual vs. Predicted Traffic Congestion Levels with Weather Influence

Figure 5 compares actual congestion levels to predicted congestion level. Including weather features in the model shows a better correlation closing the gap in the predictions. The close proximity of the two lines represents how well the ensemble model learned the weather patterns, potentially enhancing prediction performance in different environmental conditions. Similarly, weather-related features (shown in the associated plots), the ensemble model learned and incorporated environmental factors to form the impact of weather conditions related to congestion. Weather influences congestion significantly, depending on the weather, be it clear, cloudy, and rainy, generally seeing lower congestion associated with the rain. This shows the models improved real applicability drawing upon multiple data sources. These findings demonstrate the ensemble deep learning models robustness and applicability across the multiple temporal horizons and weather conditions, and highlight the usefulness of the model for intelligent transportation systems and to manage real-time traffic flow.

The newly developed ensemble deep learning model provides an excellent prediction accuracy greater than 98% across different forecasting horizons, and outperformed base learners such as CNN-LSTM and MLP that produced accuracies of about 90%, as well as standard models like Support Vector Classifier that only produced an accuracy of approximately 70%. Notably, the ensemble model's improvement was a result of its ability to utilize the best of each model to combine strengths and weaknesses, producing a more robust, generalizable prediction solution. Including weather-related inputs also resulted in slightly more accurate predictions for congestion by taking note of outside impacts on traffic patterns. Overall, our approach demonstrates the real-world benefits of ensemble learning (fusion of data sources) as well as supportive literature where reported accuracy for other methods remains below 90%.

5. CONCLUSION AND FUTURE WORK

This research has developed an ensemble deep learning model comprised of CNN-LSTM, MLP, and SVC base learners with a logistic regression meta-learner to predict real-time traffic congestion. The ensemble model performed consistently and outperformed independent models by achieving greater than 98% accuracy for a variety of prediction horizons. Adding weather-related features improved the prediction performance even more reliably, indicating the value of unifying data sources in modeling complex traffic behavior. These findings contribute further to validating the ensemble approach in general as a valid and robust answer for intelligent traffic management systems.

For future work, improving generalizability will require expanding the dataset to incorporate more urban settings and over a longer time context. Utilizing advanced meta-learning strategies and other contextual features, such as special events or road conditions, can also improve prediction performance. Additionally, deploying and testing the model in practical intelligent transportation systems will be necessary to evaluate its practicality and improvements.

REFERENCES

[1] N. Cini and Z. Aydin, "A Deep Ensemble Approach for Long-Term Traffic Flow Prediction," Arabian Journal for Science and Engineering, vol. 49, Jan. 2024, doi: 10.1007/s13369-023-08672-1.

[2] L. Qu, W. Li, W. Li, D. Ma, and Y. Wang, "Daily long-term traffic flow forecasting based on a deep neural network," Expert Systems with Applications, vol. 121, pp. 304–312, 2019.

[3] K. Sattar, F. Chikh Oughali, K. Assi, N. Ratrout, A. Jamal, and S. M. Rahman, "Transparent deep machine learning framework for predicting traffic crash severity," Neural Computing and Applications, vol. 35, no. 2, pp. 1535–1547, 2023.

[4] I. Lana, J. Del Ser, M. Velez, and E. I. Vlahogianni, "Road traffic forecasting: recent advances and new challenges," IEEE Intelligent Transportation Systems Magazine, vol. 10, no. 2, pp. 93–109, 2018.

[5] L. N. Do, N. Taherifar, and H. L. Vu, "Survey of neural network-based models for short-term traffic state prediction," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 9, no. 1, p. 1285, 2019.

[6] Z. Wang, X. Su, and Z. Ding, "Long-term traffic prediction based on LSTM encoder–decoder architecture," IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 10, pp. 6561–6571, 2020.

[7] T. Bogaerts, A. D. Masegosa, J. S. Angarita-Zapata, E. Onieva, and P. Hellinckx, "A graph CNN-LSTM neural network for short and long-term traffic forecasting based on trajectory data," Transportation Research Part C: Emerging Technologies, vol. 112, pp. 62–77, 2020.

[8] Y. Li, S. Chai, Z. Ma, and G. Wang, "A hybrid deep learning framework for long-term traffic flow prediction," IEEE Access, vol. 9, pp. 11264–11271, 2021.

[9] L. N. Do, H. L. Vu, B. Q. Vo, Z. Liu, and D. Phung, "An effective spatial-temporal attention-based neural network for traffic flow prediction," Transportation Research Part C: Emerging Technologies, vol. 108, pp. 12–28, 2019.

[10] Y. Chen, H. Chen, P. Ye, Y. Lv, and F.-Y. Wang, "Acting as a decision maker: traffic-condition-aware ensemble learning for traffic flow prediction," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 4, pp. 3190–3200, 2020.

[11] T. Bouraffa, E. K. Carlson, E. Wessman, A. Nouri, P. Lamart, and C. Berger, "Comparing Optical Flow and Deep Learning to Enable Computationally Efficient Traffic Event Detection with Space-Filling Curves," in Proceedings of the 27th IEEE International Conference on Intelligent Transportation Systems (IEEE ITSC 2024), 2024.

[12] R. A. Khalil, Z. Safelnasr, N. Yemane, M. Kedir, A. Shafiqurrahman, and N. Saeed, "Advanced Learning Technologies for Intelligent Transportation Systems: Prospects and Challenges," IEEE Open Journal of

Vehicular Technology, vol. 5, Feb. 2024, doi: 10.1109/OJVT.2024.3369691.

[13] M. Ganesan, S. Kandhasamy, B. Chokkalingam, and L. Mihet-Popa, "A Comprehensive Review on Deep Learning-Based Motion Planning and End-to-End Learning for Self-Driving Vehicles," IEEE Access, vol. X, Apr. 2024, doi: 10.1109/ACCESS.2024.3394869.

[14] L. Falaschetti, L. Manoni, L. Palma, P. Pierleoni, and C. Turchetti, "Embedded Real-Time Vehicle and Pedestrian Detection Using a Compressed Tiny YOLO v3 Architecture," IEEE Transactions on Intelligent Transportation Systems, vol. X, Aug. 2024, doi: 10.1109/TITS.2024.3447453.

[15] M. Adl, R. Ahmed, C. Vidal, and A. Emadi, "Enhanced Vehicle Movement Counting at Intersections via a Self-Learning Fisheye Camera System," IEEE Access, vol. 12, May 2024, doi: 10.1109/ACCESS.2024.3408052.

[16] G. Yıldız, A. Ulu, B. Dizdaroğlu, and D. Yıldız, "Hybrid Image Improving and CNN (HIICNN) Stacking Ensemble Method for Traffic Sign Recognition," IEEE Access, vol. 11, July 2023, doi: 10.1109/ACCESS.2023.3292955.

[17] D. Valencia, E. Muñoz, and M. Muñoz-Añasc, "Impact of the Preprocessing Stage on the Performance of Offline Automatic Vehicle Counting Using YOLO," IEEE Latin America Transactions, vol. 22, no. 9, pp. 723, Sept. 2024.

[18] G. S. Velan, B. Shanthini, and V. Cyril Raj, "Hybrid Real-Time Traffic Management System with Time Series Analysis Using Deep Learning Methods," Nanotechnology Perceptions, vol. 20, no. S7, pp. 117–132, 2024.

[19] N. Cini and Z. Aydin, "A Deep Ensemble Approach for Long-Term Traffic Flow Prediction," Arabian Journal for Science and Engineering, vol. 49, pp. 12377–12392, Jan. 2024, doi: 10.1007/s13369-023-08672-1.

[20] A. Chaudhuri, "Smart Traffic Management of Vehicles Using Faster R-CNN Based Deep Learning Method," Scientific Reports, vol. 14, no. :10357, 2024, doi: 10.1038/s41598-024-60596-4.

[21] X. Liu, Y. Wang, D. Yu, and Z. Yuan, "YOLOv8-FDD: A Real-Time Vehicle Detection Method Based on Improved YOLOv8," IEEE Access, vol. 12, Sept. 2024, doi: 10.1109/ACCESS.2024.3453298.

[22] M. Chen, S. Banitaan, and M. Maleki, "Enhancing Pedestrian Group Detection and Tracking Through Zone-Based Clustering," IEEE Access, vol. 11, pp. 132162–132179, 2023.

[23] P. Bhartiya, M. Bhatele, and L. Gour, "Anomaly Detection and Short-Term Prediction of Traffic Flow by Using Ensemble Deep Learning Approach," International Journal of Research Publication and Reviews, vol. 4, no. 7, pp. 1951–1959, 2023

[24] P. Bhartiya, M. Bhatele, and A. A. Waoo, "The Role of Deep Learning in Exploring Traffic Prediction Techniques," ShodhKosh: Journal of Visual and Performing Arts, vol. 5, no. 1, pp. 540–549, 2024, doi: 10.29121/shodhkosh.v5.i1.2024.1878.

[25] P. Bhartiya, M. Bhatele, and A. A. Waoo, "A Machine Learning Approach for Predictive Analysis of Traffic Flow," *ShodhKosh: Journal of Visual and Performing Arts*, vol. 5, no. 5, pp. 422–430, 2024, doi: 10.29121/shodhkosh.v5.i5.2024.1892.

[26] L. Gour and A. A. Waoo, "Challenges of Distributed Computing in the Context of Deep Learning," Journal of Emerging Technologies and Innovative Research (JETIR), vol. 6, no. 6, 2021.

[27] J. K. Jain and A. A. Waoo, "An artificial neural network technique for prediction of cyber-attack using intrusion detection system," Journal of Artificial Intelligence, Machine Learning & Neural Networks, vol. 3, pp. 33–42, 2023.

[28] R. Nayak, P. S. Patheja, and A. Waoo, "An enhanced approach for weather forecasting using neural network," in Proceedings of the International Conference on Soft Computing for Problem Solving (SocProS 2011), Dec. 20–22, 2011, vol. 2, pp. 833–839, Springer India, 2012.

[29] R. Nayak, P. S. Patheja, and A. A. Waoo, "An artificial neural network model for weather forecasting in Bhopal," in IEEE International Conference on Advances in Engineering, Science and Management (ICAESM-2012), 2012, pp. 747–749, IEEE.

[30] L. Soni and A. Waoo, "A review of recent advances in methodologies for face detection," International Journal of Current Engineering and Technology, vol. 13, no. 02, pp. 86–92, 2023.

[31] Bhartiya, P., Bhatele, M., & Waoo, A. A. (2025, June). Efficient Traffic Flow Prediction with CNN-Based Deep Learning Techniques. In International Conference on Advances and Applications in Artificial Intelligence (ICAAAI 2025) (pp. 1153-1173). Atlantis Press.

[32] Bhatele, M., Bhartiya, P., & Waoo, A. A. (2025). Integrating deep learning and ensemble methods for urban traffic forecasting. International Journal of Environmental Sciences, 11(8s), 565–582.

[33] L. N. Soni, A. A. Waoo, (2025) "A Lightweight and Efficient Hybrid CNN Model for Face Detection" International Journal of Environmental Sciences, Vol. 11 No. 8s, 2025, Pp. 583-591.

[34] L. N. Soni, A. A. Waoo, (2025) "Face Detection Under Low-Light and Low-Resolution Conditions Using Contrast-Limited Adaptive Histogram Equalization and a Modified Convolutional Neural Network" published in Journal of Neonatal Surgery, 14(32s), 2025 2620-2631.

[35] Soni, L. N., & Waoo, A. A. (2025). A Lightweight and Efficient Hybrid CNN Model for Face Detection. International Journal of Environmental Sciences, 11(8s), 583-591.