

## A Comparative Study of Various Traffic Flow Prediction Techniques Using ML Models and Real-Time Analysis

Bhushan S. Yelure  
*Department of Information  
Technology,  
Government College of  
Engineering, Karad.*  
bhushanyelure2008@gmail.com

Siddheshwar Patil  
*Department of AI & ML,  
DYPCET, Kolhapur*  
siddheshwar.patil@gmail.com

Aditi D. Dhigare  
*Department of Information  
Technology,  
Government College of  
Engineering, Karad.*  
aditi.dhigare@gmail.com

Mrunmayee A. Solanke  
*Department of Information  
Technology,  
Government College of  
Engineering, Karad.*  
mrunmayee2308@gmail.com

Raj S. Jadhav  
*Department of Information  
Technology,  
Government College of  
Engineering, Karad.*  
rajsatishjadhav9@gmail.com

Krishna A. Gawande  
*Department of Information  
Technology,  
Government College of  
Engineering, Karad.*  
krishg4321@gmail.com

**Abstract-** Due to the rapid growth of vehicles and the expanding urban landscape, the challenge of traffic congestion is becoming increasingly prominent. There has been a considerable shift towards data-driven solutions because of the availability of massive amounts of transportation-related data and the development of machine learning algorithms.

In the proposed work, several feature sets for traffic flow are compared by using different ML methods: kNN, SVR, Decision tree, and Random Forest. The dataset used for the prediction is the traffic flow of the area between two hospitals in San Francisco. According to the result, feature extraction is more important than method selection to get accurate predictions. Selecting high quality features leads to fewer complex techniques which are easier to handle, are trustworthy and faster. Traffic flow prediction's accuracy obtained by the decision tree surpasses

the other methods. In addition to that, real-time analysis is applied to live feed captured with the help of CCTV. The live feed is then processed using the OpenCV library, the objects are detected using the YOLOv4 model which is further passed to DEEPSORT for vehicle tracking. This can further be processed by applying mathematical formulae to calculate the congestion on the road.

**Keywords** Traffic analysis, Machine Learning, Traffic prediction

### 1. Introduction

The major problem in urban areas is traffic congestion which is affecting the daily lives of millions of people in cities. Appropriate traffic flow prediction can help reduce traffic jams by enabling transport authorities to manage traffic, improve routes and reduce accidents by reducing travel time. ML techniques have shown promise

in forecasting traffic flow because of their capability to process huge amounts of dataset, identify patterns and make right predictions. In the proposed work, a study of traffic flow prediction using four popular ML algorithms: kNN, SVM, Random Forest and Decision Tree is conducted. Due to the increasing demand for accurate information, traffic forecasting has become an important research area. Conventional traffic forecasting methods depend upon historical traffic data, which often leads to inaccurate forecasts due to changes in traffic patterns and the impact of random events. On the other hand, machine learning techniques can combine real-time traffic data with other variables such as weather, temperature day of the week, and conditions to make them more accurate.

This study's objective is to contrast the execution of four widely used ML algorithms for traffic prediction: kNN, Random Forest, Decision tree, and SVM. By evaluating the accuracy and performance of these algorithms, the aim is to identify the most appropriate algorithm for traffic forecasting and provide insights to traffic officials and experts in the field.

For the study of algorithms, traffic flow of the area shown in the map between two hospitals in San Francisco is used. Data is collected from 1st January 2017 to 31st December 2017. The captured data points are:

**Day time (Zone column):** A number shows a 10-minute interval time-zone, cutting the 24 hours of a day into 144 zones.

**Weekday (CodedDay column):** It starts from Sunday to Saturday and is numbered serially from 1 to 7 respectively.

**Atmospheric Conditions (CodedWeather column):** Coded number shows the weather conditions.

**Temperature (Temperature column):** This is the average temperature during the day which is Fahrenheit.

Comparative study of ML algorithms for traffic forecasting, including kNN, Random Forest, Decision tree and SVM, has revealed that simple, naïve methods like historical averages are more effective for long-term forecasting (beyond one hour into the future). This finding is important because current traffic conditions have a greater impact on traffic in the near future, spanning hours or days.

Live video capturing for traffic analysis is a powerful tool used in transportation engineering to monitor and analyze traffic patterns. By using video cameras to capture real-time footage of vehicular traffic, transportation planners and engineers can gain valuable insights into traffic flows, congestion, and safety concerns.

Machine learning technology has gained popularity recently as cities and transportation agencies seek to optimize their transportation systems and reduce traffic congestion. With advances in video processing and analysis software, live video capturing can provide accurate and detailed data on traffic patterns, allowing planners to make informed decisions about road design, signal timing, and other key factors that impact traffic flow. Overall, live video capturing for traffic analysis is an important tool for transportation engineers, enabling them to improve the safety and efficiency of our transportation systems and ensure that our roads and highways are safe and reliable for all users. The main objective is to compare the four models of machine learning algorithms in order to obtain efficient accuracy.

Also, to study the congestion on the road using real-time tracking.

## 2. Literature Survey

Literature is collected from various research papers, and it is based on the dataset used, techniques used and limitations of the work.

The author in [1] has used a dataset from the Caltrans PeMS database. The algorithms used are a traditional SAE model trained in a greedy layer-wise fashion.

In [2], the model has been trained and tested with traffic flow data collected from the M6 freeway in the UK. The algorithm used is a stacked auto-encoder Levenberg-Marquardt (SAE-LM) architecture. The model was designed using the Taguchi method.

[3] has used the dataset of Nizampet road stretch, an urban area, by analyzing the measured data in the city of Hyderabad. Real-time traffic information prediction with ANN and SVR are applied for developing an effective and efficient traffic prediction.

The model's training and test data is from the Kaggle website for the implementation of machine learning algorithms [4]. The Technique of predicting traffic with a regression model using various libraries like Pandas, NumPy, OS, Matplotlib, Keras and Sklearn [4].

In [5], they have used a dataset from Huawei Munich Research Center. Algorithms used are MLP-NN, Gradient Boosting Regressor, Random Forest Regressor, Linear Regression and Stochastic Gradient Regressor, and two DL models based on RNNs: GRU and LSTM.

In [6], the models are trained and tested on the data that is from the area of Gonar to Jagatpura. It applies the utilization of LSTM.

In [7], the three sets of data are recorded at intervals of 15 minutes by three traffic flow detectors in different traffic environments. Algorithms used are SARIMA and FFT (Fast Traffic Flow).

## 3. Research Gaps

According to the research, all algorithms provide pros and cons. There are a wide number of parameters on which the traffic flow depends. There is no definite parameter that provides a more appropriate approach. The efficiency can be improvised when the algorithm is applied to a wider dataset.

## 4. Proposed system

Four ML algorithms have been proposed: kNN, Random Forest, Decision Tree, and SVM. The system can be trained for larger datasets. The kNN algorithm can handle complex features such as time of day, weather, etc., that can be complex and nonlinear. The SVM algorithm is used because it is effective in handling high-dimensional data and non-linear relationships between the input feature and output variables. The Decision tree algorithm classifies the data into classes, where each class contains a thousand sets. It works by recursively partitioning datasets based on different features. Random forest predicts using a combination of decision trees. The final prediction is the average or median of the predictions from all the decision trees. Further, the System has been extended to live capturing [8] and analysis using the YOLOv4 model and DEEPSORT [9]. The YOLOv4[10] model's predictions are provided to the DeepSORT model for real-time tracking.

OpenCV is required for the video's processing, presentation, and saving. To calculate the congestion, the number of vehicles passing in a given time interval is considered. The flow of the system is presented in Fig.1.

**Table 1: Taxonomy of various traffic flow prediction techniques**

Sr No.	Dataset used	Techniques/Algorithms	Limitations
[1]	Caltrans PeMS	SAE model, greedy layer	Overfitting of the model can occur when traffic flow data shows significant changes and uncertainty over time.
[2]	M6 freeway in UK	SAE-LM, Levenberg-Marquardt, Taguchi meth	The model's main flaw is that it performed poorly when the observed traffic data had a distribution that was very smooth.
[3]	Nizampet Road, Hyderabad	ANN, SVR	Less classification accuracy and classification speed
[4]	Traffic Analysis dataset from Kaggle	LSTM, ANN, RNN	The system can be improved by utilizing a variety of elements that have an impact on traffic management.
[5]	Huawei Munich Research Center	MLP-NN, Gradient Boosting Regressor, RFR, Linear Regressor, SGR, two DL models based on RNNs: GRU and LSTM	The algorithm is efficient only for smaller data.
[6]	Gonar to Jagatpura road	Long Short-Term Memory Networks (LSTM)	It is not well-suited for online learning tasks.
[7]	Traffic Monitoring Unit (TMU) Site 9956/1 Motorway Incident Detection and Automatic Signaling (MIDAS) Site 30028401 MIDAS Site 30032728	SARIMA, FFT.	SVR model is unfit for complex roads.

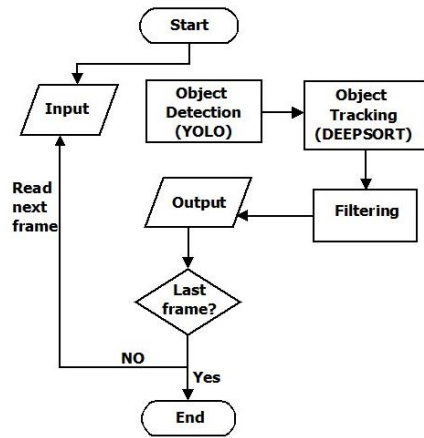


Fig 1: Real-time analysis flowchart

1. Input: The system takes in video footage as input, which is fed into the system through a camera or other source.
2. Object Detection: The video footage is then processed by the YOLOv4 [11] (You Only Look Once) deep learning algorithm to detect and classify objects in the video frames. YOLOv4 uses a convolutional neural network to identify objects based on their features and characteristics.
3. Object Tracking: Once objects have been detected, the DeepSORT [12] (Deep SORT) algorithm is used to track the objects over time. DeepSORT uses a combination of appearance and motion information to track objects and assign unique identifiers to each object in the video stream. The scenarios used for object tracking are depicted in Fig. 2.

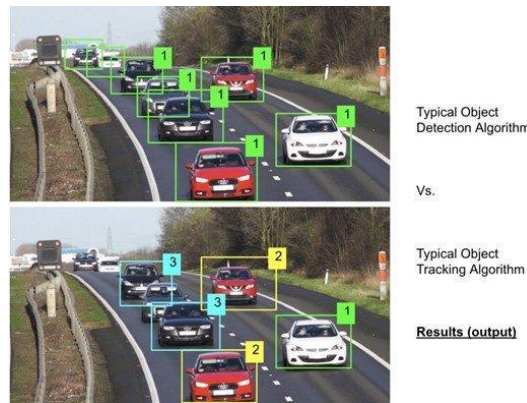


Fig 2: Object Tracking

4. Filtering: The tracking results are then passed through a filter to smooth the motion trajectories and remove any spurious or erroneous data points.
5. Output: Finally, the system outputs the tracked object data, including their unique IDs, motion trajectories, and other metadata, which can be used for further analysis or visualization.

The four metrics used to evaluate the efficiency of machine learning algorithms are presented as follows:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (1)$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (2)$$

$$F1 - score = \frac{Precision * Recall}{Precision + Recall} * 2 \quad (3)$$

$$Accuracy = \frac{True\ Positives + True\ Negatives}{True\ Positives + True\ Negatives + False\ Positives + False\ Negatives} \quad (4)$$



Four metrics are used to evaluate the real-time traffic flow analysis and those are expressed as follows:

$$\text{Average Forward Density} = \frac{\text{Vehicles In Forward Lane}}{\text{time}} \quad (5)$$

$$\text{Average Backward Density} = \frac{\text{Vehicles In Backward Lane}}{\text{time}} \quad (6)$$

$$\text{Average vehicles} = \frac{\text{Total vehicles on the lane}}{\text{total time}} \quad (7)$$

$$\text{Average stop time} = \frac{\text{Total duration of stops}}{\text{number of stops}} \quad (8)$$

## 5. Results and Discussion

The study in the paper compares several traffic flow prediction algorithms and methods. The techniques are summarized in the table. The methods and algorithms for predicting traffic flow are the SAE model, greedy layer, SAE-LM, Levenberg-Marquardt, Taguchi meth, ANN, SVR, kNN, and Random Forest [13]. The results depicted in in Fig.3 indicates that the Decision tree had the highest accuracy at 99%, followed by Random Forest at 98%, kNN at 59%, and SVM at 20%. Decision tree outperforms by 1%, 40% and 79% as compared to RF, kNN and SVM respectively.

The study also analyzed strengths and weaknesses of each algorithm, identifying kNN as ideal for noisy data and it is faster than SVM, while SVM is best suited for non-linear data. Random Forests are effective for large datasets that contain categorical data, and Decision tree combines various decisions and is faster than Random Forests, making it suitable for linear dataset.

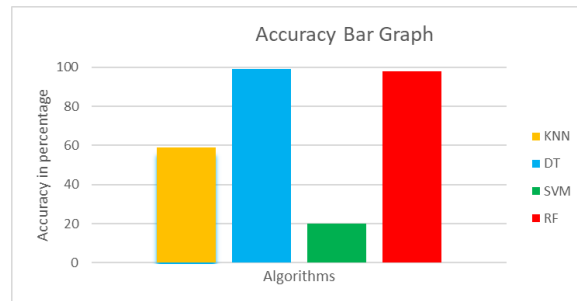
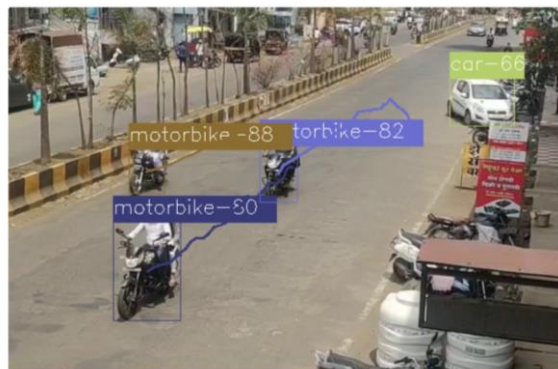


Fig 3: Accuracy Bar Graph

Additionally, the technology of live video capturing is also conducted using YOLOv4 and DEEPSORT. One of the main challenges is managing the large amounts of data that are generated by the cameras. This requires sophisticated data processing and analysis tools to extract meaningful insights from the footage.

In the real-time analysis of the traffic, real-time stats of the traffic are analyzed. This includes the average frequency of cars that traversed the intersection, the average traffic density of vehicles in the forward motion and the backward motion, and the average stop duration at the intersection. To achieve a higher FPS, the YOLOv4 model with an intake of 416x416 was employed. The potential lack of accuracy is a drawback of this option. Convert the YOLOv4 weights from the Darknet to TensorFlow weights if accuracy is the top goal.



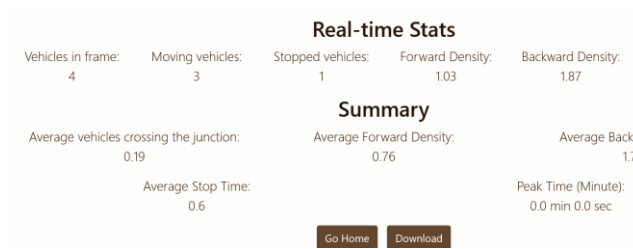


Fig 4: Real-time Tracking

## 6. Conclusion

A study was conducted to contrast the results and accuracy of four ML algorithms - kNN, Random Forests, Decision tree, and SVM - for predicting traffic flow. The results indicated that the Decision tree had the highest accuracy at 99%, followed by Random Forest at 98%, kNN at 59%, and SVM at 20%. The study emphasizes the importance of selecting the most appropriate machine learning algorithm for a specific dataset, based on various factors. Decision tree and Random Forests were found to be efficacious methods for forecasting traffic flow, and updating the data frequently and considering relevant factors can increase the precision and effectiveness of the model. In the real-time analysis of the traffic, real-time stats of the traffic are analyzed. Overall, while expanding on the discussion of live video capturing for traffic analysis may expand the size of the design, numerous ways are available to present the information in a clear, concise, and effective manner. With careful attention to language and structure, broad audiences can be informed and engaged through the communication of challenging ideas and concepts.

## References

- [1] Sanaz Shanker Sepasgozar and Samuel Pierre, "Network Traffic Prediction Model Considering Road Traffic Parameters Using Artificial Intelligence Methods in VANET", 2022
- [2] Kyungeun Lee, Moonjung Eo, Euna Jung, Yoonjin Yoon, and Wonjng Rhee, "Short-Term Traffic Prediction with Deep Neural Networks", 2021
- [3] Gaurav Meena, Deepanjali Sharma, Mehul Mahrishi "Traffic Prediction for Intelligent Transportation System using Machine Learning", 2020
- [4] B. Rajendra Kumar, Dr. Naveen Kumar Chikkakrishna, Teja Tallam, "Short-Term Predictions of Traffic Flow Characteristics using ML Techniques", 2020
- [5] Alfonso Navarro-Espinoza 1 , Oscar Roberto López-Bonilla 1 , Enrique Efrén García-Guerrero 1 , Esteban Tlelo-Cuautle 2 , Didier López-Mancilla 3 , Carlos Hernández-Mejía 4 and Everardo Inzunza-González, "Traffic Flow Prediction for Smart Traffic Lights Using Machine Learning Algorithms", 2022
- [6] Deekshetha H R 1 , Shreyas Madhav A V\* 1 , and Amit Kumar Tyagi 1,2 1 School of Computer Science Engineering, Vellore Institute of Technology, Chennai, Tamilnadu, India 2 Centre for Advanced Data Science, Vellore Institute of Technology, Chennai, Tamilnadu, India, 2022
- [7] Noor Afza Mat Razali1\* , Nuraini Shamsaimon1 , Khairul Khalil Ishak2 , Suzaimah Ramli1 , Mohd Fahmi Mohamad Amran1 and Sazali Sukardi3, "Gap, techniques and evaluation: traffic flow prediction using machine learning and deep learning", 2021

- [8] Benjamin Coifman a, David Beymer b 1 1, Philip McLauchlan b 2 2, Jitendra Malik b 3 3, “A real-time computer vision system for vehicle tracking and traffic surveillance”, 1998
- [9]Xinyu Hou; Yi Wang; Lap-Pui Chau, “Vehicle Tracking Using Deep SORT with Low Confidence Track Filtering”,2019
- [10]Fahimul Hoque Shubho; Fahim Iftekhar; Ekhfa Hossain; Shahnewaz Siddique,”Real-time traffic monitoring and traffic offense detection using YOLOv4 and OpenCV DNN”,2021
- [11]Imalie Perera, Shehan Senavirathna, Aseni Jayarathne; Shamendra Egodawela, “Vehicle Tracking based on an Improved DeepSORT Algorithm and the YOLOv4 Framework”,2021
- [12]Adson M. Santos; Carmelo J. A. Bastos-Filho; Alexandre M. A. Maciel; Estanislau Lima,”Counting Vehicle with High-Precision in Brazilian Roads Using YOLOv3 and Deep SORT”,2020
- [13] Nazirkar Reshma Ramchandra 1, Dr. C. Rajabhushanam 2, “Traffic Prediction System Using Machine Learning Algorithms”, 2021