Smart Traffic Light System Using Machine Learning

1st Nandana Praveen
Dep.of Electronics and Communication
Amrita Vishwa Vidhyappetham
Kollam, kerala
nandanapraveen2004@gmail.com

2nd Parvathy S
Dept. of Electronics and Communication
Amrita Vishwa Vidhyappetham
Kollam, kerala
parvathysuchith@gmail.com

2nd R Parvathy
Dept. of Electronics and Communication
Amrita Vishwa Vidhyappetham
Kollam, kerala
parvathyremesh17@gmail.com

***Abstract*—Urban traffic congestion is increasingly becoming a challenge due to increased population and car density. Traditional traffic management systems that rely on static timers fail to adapt to real-time traffic conditions and often tend to worsen congestion and inefficiencies. The paper reports the design and development of a smart traffic light system using machine learning and computer vision to optimize signal timings dynamically for improving traffic flow at intersections.**

The system makes use of two methodologies: namely, (1) the use of OpenCV to carry out edge detection and noise reduction on image processing for the counting and detection of vehicles and (2) the use of a pre-trained RetinaNet model in TensorFlow for object detection. Algorithms such as Logistic regression and decision tree-based algorithms like the Variance-Based Algorithm and Slab Division Algorithm are used to allocate a fair and efficient number of green signals across lanes. It will focus on urgent routes to balance traffic management.

It is expected that the proposed system will reduce congestion, prevent long waiting times, and increase fairness in signal time with dynamic adaptation to real-time situations. Future development involves optimizations of multi-lane optimisation, citywide synchronization, and advanced interoperability among emergency vehicles. The highly scalable and intelligent solution to mitigate urban traffic congestion with these state-of-the-art technologies is promising.

***Index Terms*—Traffic congestion, machine learning, computer vision, dynamic signal allocation, object detection, RetinaNet, traffic management.**

I. Introduction

The rapid pace of urbanization and rapidly growing number of vehicles on roads has led to increased traffic congestion. This is one of the biggest challenges modern cities have ever had regarding traffic congestion, because it disturbs the daily commutes and at the same time incurs large economic and environmental costs. This means hours wasted to commuters, fuel consumption arising from idling by the vehicles, and also causes pollution from congested traffic, among other global environmental challenges, such as climate change, and deteriorating air quality in urban areas. This also leads to stress with low productivity, which impacts the general quality of living. It operates generally utilizing a static traffic light timer regardless of the time-varying volumes of traffic characteristically varying between days or sometimes even hours of the day.

Identify applicable funding agency here. If none, delete this.

It makes use of schedules that do not adjust themselves to real situations, making an unoptimal use of the flow of traffic. For example, consider a peak hour wherein lanes happen to be occupied yet use an equal length of green signals with nearly vacated lanes resulting in wasted green signals and increased queuing delay. Similarly, during the off-peak hours, at times, vehicles will wait unnecessarily under the red lights as other roads remain empty in lanes. Lack of adaptability shows grave inadequacies in the systems under conventional systems as the former face the modern traffic challenge. The other inefficiencies that are identified in the current systems of real life include that the emergency vehicles, like ambulances, fire trucks, and police cars, typically encounter problems while passing through junctions because static systems are incapable of creating priority for such life routes.

This might eventually result in disastrous results such as delayed medical treatment or even worse, delayed responses to emergencies. Besides, during massive evacuations due to natural calamities or public gathering, the traditional systems are unable to provide adequate solutions for the rapidly increasing traffic. The difficulty in coordinating traffic lights with dynamic response to emergency cases worsens congestion and increases the propensity for accidents. For all the above issues, there is an urgent need for a smart and adaptive traffic management approach. The STMS put forward has been designed along with improvement in the capabilities of image processing as well as machine learning toward making real-time decisions with regard to dynamic adaptation of duration of traffic signals. This system analyzes real-time data taken from cameras situated in the intersections so that it can easily observe and count the number of automobiles along with the length that exist on each lane plus allocation for the appropriate green light period. This would clear the most congested lanes first, hence reducing waiting times overall and making it more efficient in terms of traffic flow. The two core methodologies used in this proposed system for vehicle detection include image processing using OpenCV and machine learning-based object detection. This method from OpenCV processes images to detect and count vehicles by noise reduction, edge detection, and comparison with reference images. The pre-trained models like RetinaNet are used as the basis of the approach to machine learning in achieving high accuracy vehicle classification and counting. Both approaches allow for robust traffic data acquisition within the system for the accomplishment of dynamic signal allocation.

The systems make use of the algorithms for dynamic allocation of the green signal durations, particularly Variance-Based Algorithm, and Slab Division Algorithm. In variance-based algorithms, the vehicle counts for each lane will have spread values to enable fairness in traffic and optimize its flow. The Slab Division Algorithm groups the counts within the various vehicles into predetermined slabs that determine green signal duration, while such algorithms will give constraints with maximum waiting time and along with emergency route conditions and one-sided traffic. This system is scalable and efficient with the solution proposed with integration of real-time adaptability along with handling emergency capabilities. Besides this ability to ameliorate daily commutes, it can greatly improve the speed of response in emergencies and evacuation.

Citywide synchronization of traffic lights, integration with Internet of Things (IoT) devices for better monitoring of traffic, and multi-lane optimization are some of the future extensions of this system. These developments will ensure the versatility of the system as it applies to a broad spectrum of traffic conditions and urban infrastructures. This paper gives a comprehensive approach towards tackling traffic congestion through intelligent traffic management. The proposed system integrates advanced technologies with dynamic algorithms for the transformation of the traffic flow management system that minimizes congestion, emissions, and delays while maximizing safety and overall urban mobility.

II. LITERATURE REVIEW

This paper [1] Smart traffic congestion reduction is one of the key activities to reduce traffic problems in urban spaces. Various studies have involved connected vehicle data and cloud security, the VIKOR approach with neural networks, as well as UAVs, for monitoring traffic. Moreover, Zigbee-based data collection for traffic, parallel computing with genetic algorithms and distributed computing, and even the prediction of congestion propagation have been studied. Advanced Traveler Information Systems (ATIS) use algorithms such as k-nearest neighbor and artificial neural networks to provide real-time traffic information to travelers. Machine learning algorithms, including linear support vector machines, data aggregation methods, fuzzy logic, and attention-based recurrent neural networks, have been found useful in traffic prediction and management. A proposed smart traffic system with roadside sensors and machine learning algorithms, such as logistic regression, random forest, and AdaBoost, was reported to be able to achieve 91% accuracy in traffic prediction. The architecture of the system includes a perception layer to collect data, a prediction layer to analyze it, and an application layer for user interface. Future research aims to enhance data security through cryptography and end-to-end encryption, while addressing limitations related to smartphone dependency and sensor costs.

This paper [2] has reviewed some intelligent traffic management systems for countering the limitations of the conventional system, which generally does not take into consideration the present scenario of traffic. Due to such a scenario, congestion, delay, and wastage of fuel occur. These systems incorporate software defined drone networks, adaptive algorithms in an emergency vehicle dispatching mechanism, reinforcement learning, fuzzy logic, calendar-based history information, Wireless Sensor Networks, and video processing techniques, for optimizing traffic flow conditions and reducing congestions in them. However, these systems are often faced with the problems of high installation and maintenance costs, dependence on real-time data, and complexity in implementation. Therefore, there is a need for a low-cost, adaptive, and easily deployable system that utilizes existing infrastructure such as surveillance cameras and readily available technologies such as cloud computing and machine learning.

This paper [3] introduces adaptive traffic light systems that can be utilized to alleviate traffic congestion, which is a major concern in cities, with machine learning that optimizes traffic flow. Traditional pre-timed traffic lights used in Lebanon lead to inefficiencies during peak hours or irregular conditions. Adaptive traffic lights employing reinforcement learning dynamically adjust signal phases according to real-time traffic conditions. Previous studies have investigated other state descriptions for reinforcement learning algorithms, such as queue lengths, discrete traffic state encoding (DTSE), and a combination of queue length and waiting time to avoid unwanted behavior, such as indefinite waiting of a single vehicle. Other studies have been aimed at multiagent models of optimizing multiple adjacent intersections but lack the coordination of a green wave achievable with pretimed systems. In the system proposed in this paper, there exists one single-agent model with queue lengths and waiting times used as input to a Q-learning algorithm to enhance isolated intersections' traffic flow. Simulation of SUMO, a piece of traffic simulation software, was used to demonstrate impressive improvements in average queue lengths and waiting times over both static traffic lights and the other proposed adaptive systems. Future research directions include extending the model to accommodate non-isolated intersections and incorporating outgoing road conditions for further optimization.

III. PROPOSED SYSTEM

The proposed system will use a machine learning algorithm to optimize control of traffic lights, helping to reduce congestion and develop better urban traffic management. The system will use real-time data on traffic, group the traffic flow patterns through clustering algorithms such as K-Means into low, moderate, and high traffic; these are further mapped to the Green, Yellow, and Red light signals, thus adapting the signal transitions to be effective and efficient. The preprocessing of the input data includes managing missing values, normalization, and outlier removal for robust model performance. Dimensionality reduction is applied through Principal Component Analysis to ensure that system computation efficiency is

Cross 1	Cross 2	Cross 3	Cross 4	Cross 5	Cross 6
105	48	30	62	31	110
97	41	32	55	42	103
76	47	44	58	40	100
98	40	39	59	43	104
87	41	47	49	35	112
80	40	35	63	34	89
92	46	39	58	26	98
80	31	21	45	29	93
91	55	35	46	26	89
75	42	26	77	32	48
48	24	15	38	19	61
61	31	17	33	16	54
60	32	23	53	19	64
59	25	22	28	21	60
42	10	14	27	17	42
55	24	24	31	16	47
38	15	26	32	23	61
42	26	18	29	25	55
47	27	15	47	14	48

Fig. 1. Data set

improved without critical feature loss in the clustering process. The clustering results are further refined and then integrated into a logistic regression model to predict the best traffic light signal for the given traffic scenario. This approach of supervised learning improves the accuracy by learning from historical traffic patterns. Further, a correlation heatmap is used to understand relationships among the traffic features, thus allowing feature selection for better model interpretability. In a later attempt to further illustrate cluster distribution and traffic light assignments, this will ensure that the predicted number of vehicles does not just sound valid but also seems believable by the traffic management people. The system is viable in the sense that the cloud-based storage system allows one to store both raw and processed data for further analysis, making it scalable. This proves that the predictions are valid as the model has used a confusion matrix and classification accuracy for its evaluation. The system decreases the idle time at intersections, reduces fuel consumption, and reduces greenhouse gas emission, therefore contributing to the development of a sustainable urban landscape. In conclusion, the intelligent traffic light system provides a feasible, scalable, and eco-friendly solution for traffic management for modern challenges in this respect.

IV. ABOUT THE DATA SET

The road traffic prediction datasets from Huawei Munich Research Center contains recorded data from 6 crosses in the urban area for the 56 days, in the form of flow timeseries, depicted the number of vehicles passing every 5 minutes for a whole day (i.e. 12 readings/h, 288 readings/day, 16128 readings / 56 days).

Total 16128 rows 6 columns.

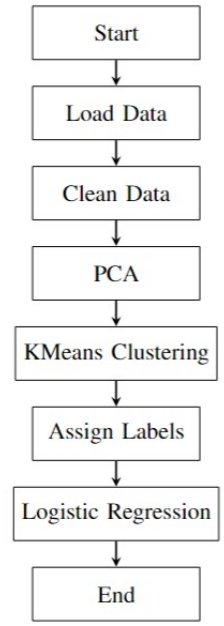

Fig. 2. flowchart

V. METHODOLOGY

The traffic dataset was downloaded by using the gdown library and loaded into a pandas DataFrame to analyze. The missing values were handled with zeros replaced by NaN to fill those gaps column-wise averages. The outliers were systematically removed by using the IQR method, hence giving a refined dataset. Moreover, normalization was carried out with MinMaxScaler in order to scale the features uniformly; extreme values were handled by keeping rows whose Z-scores are less than 3. In clustering, K-Means and DBSCAN algorithms were used to group data into clusters representing different traffic levels. The K-Means algorithm was fine-tuned by testing cluster numbers between 2 and 10, and the optimal configuration was chosen based on the Silhouette Score. Principal Component Analysis was also used to reduce dimensionality so that clusters could be visualized. The DBSCAN algorithm was used to identify density-based clusters and noise points for further traffic pattern analysis. The resulting clusters were mapped to traffic light states: Green for low traffic, Yellow for moderate traffic, and Red for high traffic.

A Logistic Regression model was developed and trained to predict traffic light states. The dataset was split into training and testing subsets, with the model showing strong accuracy during evaluation on the test set. A confusion matrix was constructed and visualized by a heatmap to evaluate the performance of the model, demonstrating its ability to classify traffic light states as Green, Yellow, or Red. This type of classification was able to relate the clustering results with actionable traffic management strategies, which makes it practically applicable.

The processed dataset with clustering results was exported as CSV files for future use and analysis. Visualization tools

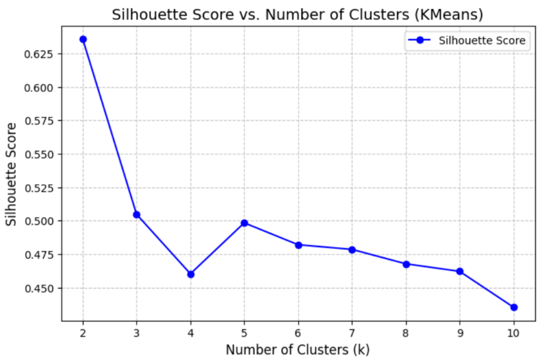

Fig. 3. Silhouette Score vs. Number of Clusters (KMeans)

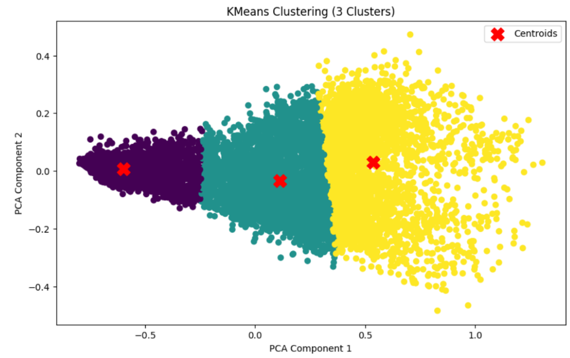

Fig. 4. KMeans Clustering (3 Clusters)

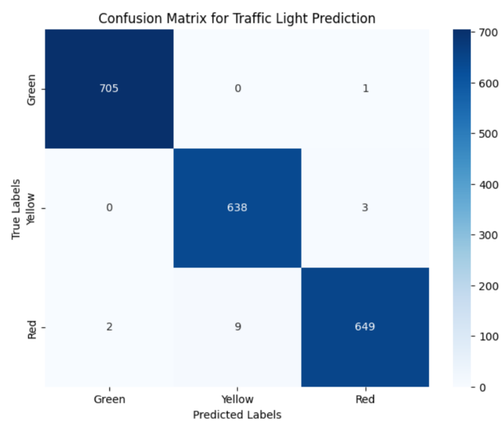

Fig. 5. Confusion Matrix for Traffic Light Prediction

such as Matplotlib and Seaborn were highly used in interpreting patterns and communicating the results. This methodology brings together clustering and classification techniques to provide a practical approach to predicting traffic light states using real-world traffic data.

VI. RESULT AND CONCLUSION

Several preprocessing steps have been done on the dataset to ensure that it is proper for meaningful analysis. The missing values were imputed by replacement with the mean of their columns, and zeros were removed to preserve data integrity. Outliers were dealt with using the IQR method, while Min-Max scaling was applied to normalize all features. This ensured uniform scaling of data, thus enhancing the performance of the subsequent clustering and classification models.

Applying PCA dimensionality reduction helped in reducing complexity of the dataset without losing all the important information so it was easier to visualize and interpret these clusters. Projection of data points onto the first two principal components made groupings that were otherwise unclear become explicit. A correlation heatmap showed strong positive correlations, making some data redundant and facilitated the selection of the most important attributes while refined the process of clustering and classification.

For clustering, K-Means was used to segment traffic patterns into three distinct clusters: low traffic (Cluster 0, Green), moderate traffic (Cluster 1, Yellow), and high traffic (Cluster 2, Red). The Silhouette Score of 43.58% indicated that the clusters were well-defined and separated. Although DBSCAN was kept in mind as it works well with noisy and non-linear data, K-Means outperformed it for the given dataset by showing better silhouette scores and interpretability. Logistic Regression was then performed for classification that gave 99.25% accuracy and confusion matrix with minimal error was found in the case of "Green" and "Red" states. These results, together with the PCA scatter plots and heatmap of the confusion matrix, established the validity of the hybrid approach in predicting and managing traffic conditions. 99.25% classification for the multi-class task was achieved using a logistic regression model for these traffic light states. Performance Class-wise: Green Light:

True Positives: 705 False Negatives: 1 False Positives: 2 Precision: 99.72% (705 out of 707) Recall: 99.86% (705 out of 706) F1 Score: 99.79% Yellow Light:

True Positives: 638 False Negative: 3 False Positives: 9 Precision: 98.61% (638 out of 647) Recall: 99.53% (638 out of 641) F1 Score: 99.07% Red Light:

True Positives: 649 False Negatives: 11 False Positives: 4 Precision: 99.39% (649 out of 653) Recall: 98.33% (649 out of 660) F1 Score: 98.86%

Error Analysis: Some interesting things come out when you analyze the confusion matrix and the errors made by the model.

Green-Red Confusion: The model misclassified three samples, one where green was annotated red, and two others which were red annotated as green. Yellow-Red Confusion: Here 12 misclassifications occur, where 3 predicted red instead of yellow and 9 predicted yellow instead of red. Green-Yellow Confusion: Surprisingly, there are no misclassifications between green and yellow. Error Distribution: The model has an overall classification of 2007 samples with 15 error cases, giving it an error rate of 0.75%. The most prevalent confusion is Yellow-Red, with 12 out of 15 errors.

In short, the logistic regression model performed Like a pro with very low errors and excellent precision, recall, and F1-scores across all light states.

The study showed quite strong performance in several crucial metrics. The KMeans clustering algorithm achieved a high silhouette score at 43.58%, assuring good quality of both separation and compactness for clusters. On the contrary, DBSCAN, enabling identification of noise and dealing well with non-linear cluster forms, had more sensitive adjustment in parameters and was hard to interpret in comparison to KMeans for this dataset. Logistic Regression, which predicts traffic lights based on the clusters. It has an accuracy level of 99.25%, in which it is reliable in classifying what traffic conditions are. For further validation, the robustness of the model shows high true positive rates about "Green" and "Red" categories with lesser misclassifications. Finally, visualization tools like PCA scatter plots and heatmaps were used to depict effectively the clustering and classification output, which supports the given methodology's validity and the interpretability of the procedure. These metrics collectively affirm the system's potential for accurate traffic analysis and prediction. In addition, preprocessing, which included feature normalization, missing values handling, and outlier removal, was also instrumental in the improvement of the quality of the dataset that enhanced the performance of the clustering and classification models. The integration of PCA with dimensionality reduction and the application of clustering techniques enabled a more effective and meaningful analysis of the complex patterns of traffic flow. These results further emphasize the applicability of this machine learning-based approach for real-time traffic management, providing a scalable and reliable solution to predict traffic light states and optimize traffic flow. This integration opens up the door to developing smarter, data-driven traffic systems that adapt in real time to varied traffic conditions.

Future Works

The potential future directions of work that may be used to build upon this study will be discussed next. Such directions include: First, the integration of extra sources of data will be considered to add accuracy and responsiveness of traffic light prediction, including data from sensor-based real-time traffic feeds or social media platforms, which can improve the traffic light prediction by providing the model with an ability to adapt to changes in patterns of traffic over time in real time.

More challenging machine learning techniques, including deep learning models or ensemble methods, would be explored in order to improve classification accuracy and possibly tackle more complicated traffic conditions. For instance, Recurrent Neural Networks or Long Short-Term Memory networks could be used for capturing temporal dependencies in the flow of traffic, making predictions for the states of traffic lights even better.

Further, fine-tuning and optimization of the models already in use, including KMeans and Logistic Regression, can be pursued to improve efficiency as well as reduce the computational cost. In fact, the application of hyperparameter search techniques like Grid Search or Random Search may lead to better models. Also, using other clustering algorithms such as hierarchical clustering or Gaussian Mixture Models (GMMs) may bring better cluster definitions and hence more robust predictions of traffic light states.

Exploring the application of reinforcement learning to dynamic traffic signal control is another promising avenue for future work. Simulating real-time traffic conditions with reinforcement learning models would continually adjust traffic light states in such a way that maximizes total traffic flow and therefore reduces congestion while optimizing urban traffic system management.

The practical applicability of the developed models would be finalized by the deployment of developed models in real-world traffic systems, which will continuously monitor and assess the results. Optimizing the system for real-world performance based on user feedback and further improvements could make it more reliable and scalable.

References

[1] Lakshna, A., et al. "Machine learning Smart Traffic Prediction and Congestion Reduction." 2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES). IEEE, 2021.

[2] Maheshwari, Rashi, Kritin Mathur, and Rohan Jagtap. "Smart Traffic Light System."

[3] Natafgi, Mohamad Belal, et al. "Smart traffic light system using machine learning." 2018 IEEE International Multidisciplinary Conference on Engineering Technology (IMCET). IEEE, 2018.