

Predicting Real-Time Traffic and Congestion Hotspots in Dhaka City

Aabu Yousuf Raj
Dept. of Computer Science
BRAC University
Dhaka, Bangladesh

aabu.yousuf.raj@g.bracu.ac.bd

Shajedul Arefin
Dept. of Computer Science
BRAC University
Dhaka, Bangladesh

shajedul.arefin@g.bracu.ac.bd

Md. Roman Bin Jalal
Dept. of Computer Science and Engineering
BRAC University
Dhaka, Bangladesh

roman.bin.jalal@g.bracu.ac.bd

Abstract—Traffic congestion prediction is a critical challenge for urban management, particularly in densely populated cities like Dhaka. In this paper, we evaluate the performance of four deep learning models namely CNN, ResNet50, MobileNetV2 and EfficientNetB0 for classifying traffic congestion levels based on optical images. The dataset was manually labeled into four categories: No Traffic, Light Traffic, Moderate Traffic, and Heavy Traffic. Each model was trained on images resized to dimensions of 128x128, 256x256, and 512x512 to assess performance at varying resolutions. Among the models, EfficientNetB0 emerged as the best performer, achieving the highest accuracy, F1 score, and AUC on the ROC curve, while also maintaining efficient training times. However, a key dataset limitation is conflicting traffic scenarios in two-way roads which caused significant model confusion and limited overall performance. This paper highlights the importance of addressing dataset quality issues and proposes that minimizing these flaws could result in a dramatic improvement in model accuracy. The findings provide valuable insights into the application of AI in urban traffic management and pave the way for future advancements in this field.

Index Terms—Deep Learning, Machine Learning, CNN, ResNet50, MobileNetV2, EfficientNetB0

I. INTRODUCTION

While many steps are taken on a daily basis to cope with the problem of traffic congestion especially in busy urban cities, the results in Dhaka, Bangladesh are not desirable. Due to being a city with dense population with a huge number of vehicles and unplanned road and traffic management Dhaka consistently ranks among the most congested cities globally. Besides, some studies by analyzing Dhaka's traffic patterns while using GPS data have highlighted intra-road segment congestion and land use patterns as key contributors to the city's severe traffic challenges. However, these studies have been constrained by limited data availability and challenges in accounting for real-time dynamics. Besides, The city loses over eight million working hours each day due to traffic jams which is a notable increase from five million hours lost daily in 2017 [1]. Moreover, a study of 2019 shows that there were 4,822 accidents in Dhaka attributed to traffic congestion that resulted in 476 deaths and 5,417 injuries [1]. Furthermore, a study estimates that traffic congestion costs Dhaka's economy approximately 550 billion BDT annually which equates to about 1.5 billion BDT per day [2].

Despite this being one of the greatest problems of an entire nation resulting in significant economic losses, environmental degradation and reduced quality of life, the amount of technical research conducted centering it is not adequate especially when research on this sector with the evolution of Machine Learning is flourishing in the current context. This study addresses these gaps and along with the inclusion of machine learning techniques and provides a Machine Learning (ML) based technical prediction regarding traffic congestion severity in real time.

Unlike traditional GPS or time-series approaches, we have utilised our study with high-resolution and manually labeled image data for our target variable in order to initiate better training and evaluate our models. The dataset of optical images are being resized to multiple dimensions for experimentation evaluation during the implementation of models that include Convolutional Neural Network (CNN), ResNet50, MobileNetV2 and EfficientNetB0 to establish an efficient and scalable framework for congestion prediction.

II. LITERATURE REVIEW

A complementary study proposed a machine learning framework for real-time traffic prediction in Dhaka which focuses on short-term and long-term forecasts [3]. The study applied the Circular Model of Traffic Forecasting (CMTF) to predict congestion at intersections by using time series analysis on node information tables. Though the model was effective for day-to-day scenarios it struggled with special events like seasonal variations and weather-related disruptions which highlighted the importance of integrating external variables for improved predictions. Despite its limitations, the study forecasted the potential of time-series models in addressing traffic management challenges in densely populated urban areas.

In both of the research papers [4] [5], they used Global Positioning System (GPS) data to analyze Dhaka city's traffic pattern upon considering 13 DPZ (Detailed Planning Zone) zones proposed by RAJUK. In the first paper, they generated intra and inter-road segments and their traffic pattern for 13 different zones. They considered the effect

of road segments ratio and land use pattern on traffic congestion and determined the factors that are responsible for traffic intensity which led them to know residential and suburban areas have less inter-zone traffic intensity and intra-road segments traffic intensity is more responsible for overall traffic intensity. On the other hand, the second paper generated zone-wise clusters and analyzed the traffic pattern of each cluster by Identifying factors (e.g., land use, number of road intersections, bus routes, and social infrastructures) that may cause traffic congestion and did a regression analysis for finding the impact of factors on traffic congestion. They described the traffic pattern for different cases (whole day, effect of marketplaces, day-wise traffic variations, traffic pattern of rickshaw-free roads). Lastly, in both of them, their limitations were they had only 15 days of GPS data and stated that more data can be helpful for more accurate analysis.

A research study implemented a traffic control system using computer vision aligned with machine learning algorithms and the HERE API [6]. The users of this will get the real time update on the traffic and based on that they'll be suggested the route that has the minimum amount of traffic. Additionally, the traffic signals will also be controlled according to the suggested routes with less dynamic signal shifting than the route with heavy traffic. Therefore, the congested route's light shifting depends on the time it'll take the vehicles to get to the better route.

There have been some studies in the context of other urban cities regarding traffic congestion. For example, there has been two very similar researches to predict traffic congestion at Amman City's eighth circle [7] [8]. They used different machine learning classifiers and found FURIA as the most effective model [7] while the other research found SVM as the most effective classifier. While the results are significantly advanced they have reliance on specific data types which include GPS or time-series data and also varying focus on urban contexts that limit the application of those studies in case of Dhaka's traffic dynamics.

III. METHODOLOGY

A. Dataset Description:

The DhakaAI dataset is taken from the open source platform Kaggle. The dataset addresses the challenges of automated traffic detection in Dhaka which is a city with unique traffic dynamics characterized by only 7% of its area dedicated to roads (compared to the 25% urban standard) and over 8 million daily commuters within a 306 sq. km area. The dataset consists of 2990 optical images containing vehicles from 21 distinct classes. This variety supports multi-class vehicle detection and recognition that is essential for tackling traffic-related challenges in urban environments. Besides, it is designed to facilitate AI-driven solutions for traffic analysis in multicultural and multilingual environments by addressing complexities like the presence of various scripts and languages in traffic scenes. By providing a common problem statement,

the dataset also aims to foster collaboration among researchers and practitioners in South-East Asia while promoting the growth of an AI-focused community in the region.

B. Data Preprocessing:

The dataset consists of unique optical images for which inherently there are no redundant values, irrelevant content and null values. However, to facilitate effective training and classification, manual labeling was performed for the target variable or output feature by categorizing traffic conditions have into four target classes:

1. No Traffic: Roads are clear with minimal or no vehicles while ensuring free-flowing traffic.
2. Light Traffic: Vehicles are present but traffic moves smoothly with minimal delays.
3. Moderate Traffic: Denser traffic causes occasional slow-downs and delays particularly at intersections.
4. Heavy Traffic: Roads are heavily congested and characterized by frequent stops, significant delays and potential gridlock.

After categorizing into aforementioned classes, it is seen that the dataset is an imbalanced dataset where no traffic, light traffic, moderate traffic and heavy traffic has appeared 1, 2, 3 and 4 times respectively which is shown in the barchart [Fig-1].

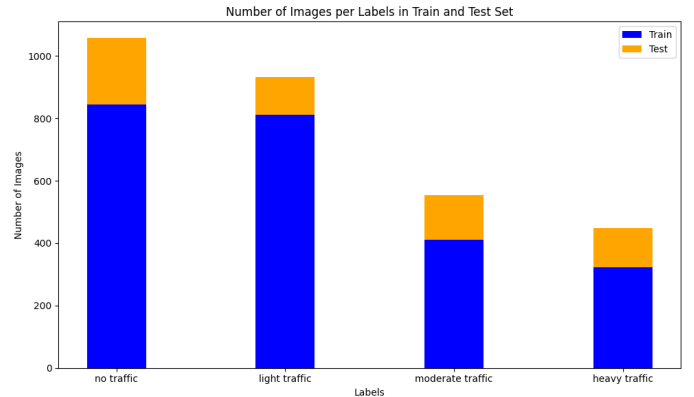


Fig. 1: Bar Chart of class distribution in target variable

Moreover, for the further model training and test, the data has been splitted into 80 and 20 percent respectively. Furthermore, each optical image has been resized to multiple dimensions (128x128, 256x256, and 512x512) for ensuring improved feature engineering during the implementation of models.

C. Learning Phase:

In this study, we have selected the models that are particularly effective for image based tasks because of their ability to learn spatial hierarchies and extract robust features from the dataset.

a) *Convolutional Neural Networks (CNNs)*: Convolutional Neural Networks (CNNs) excel in image-based tasks, automatically learning spatial patterns and

hierarchical features through convolution, pooling and fully connected layers. Their effectiveness in classification and detection makes CNNs ideal for analyzing high-resolution traffic images.

b) *ResNet50*: ResNet50 is a deep residual network with 50 layers that employs residual shortcut connections to minimize the vanishing gradient problem in deep networks. These shortcut connections allow the network to learn identity mappings and enhance feature learning without degradation in performance. The depth makes it ideal for identifying subtle variations in traffic density and vehicle arrangements in the dataset.

c) *MobileNetV2*: MobileNetV2 is a lightweight architecture optimized for mobile and embedded devices. It uses depth wise separable convolutions and inverted residual blocks to reduce computational complexity. This enables faster training and inference without significant hardware demands which makes it suitable for real-time traffic detection scenarios where computational resources may be limited.

d) *EfficientNetB0*: EfficientNetB0 is a scalable architecture which scales model network depth, width, and resolution systematically by a compound scaling method. It balances high computational efficiency with fewer parameters and less computational cost compared to traditional models. This makes it highly applicable for handling diverse traffic images in this dataset while maintaining resource productivity.

Each of these models offers unique advantages to extract meaningful features from complex traffic scenes that collectively enable a comprehensive evaluation in image classification tasks and detection methods.

IV. EXPERIMENTAL EVALUATION

Each model was implemented with optical images that have been resized to multiple dimensions (128x128, 256x256, and 512x512).

A. Convolutional Neural Networks (CNNs):

The accuracy of CNNs with optical image dimensions of 128x128, 256x256 and 512x512 are 35.83, 36.67 and 36.84 percent respectively. This model performed best with the image dimension of 512x512 and thereby the classification report for this dimension is shown below.

Metric	Precision	Recall	F1-Score	Support
heavy traffic	0.179153	0.450820	0.256410	122.000000
light traffic	0.236111	0.153153	0.185792	111.000000
moderate traffic	0.138889	0.038168	0.059880	131.000000
no traffic	0.233333	0.174129	0.199430	201.000000
accuracy	0.198230	0.198230	0.198230	0.198230
macro avg	0.196872	0.204068	0.175378	565.000000
weighted avg	0.200282	0.198230	0.176699	565.000000

TABLE I: Classification report for CNN with image size 512

B. ResNet50:

The accuracy of ResNet50 with optical image dimensions of 128x128, 256x256 and 512x512 are 33.17, 33.50 and 38.33 percent respectively. This model performed best with the image dimension of 512x512 and thereby the classification report for this dimension is shown below.

Metric	Precision	Recall	F1-Score	Support
heavy traffic	1.000000	0.000000	0.000000	122.000000
light traffic	0.186747	0.279279	0.223827	111.000000
moderate traffic	0.221918	0.618321	0.326613	131.000000
no traffic	0.352941	0.059701	0.102128	201.000000
accuracy	0.219469	0.219469	0.219469	0.219469
macro avg	0.440401	0.239325	0.163142	565.000000
weighted avg	0.429631	0.219469	0.156033	565.000000

TABLE II: Classification report for ResNet50 with image size 512.

C. MobileNetV2:

The accuracy of MobileNetV2 with optical image dimensions of 128x128, 256x256 and 512x512 are 45.50, 49.33 and 48.17 percent respectively. This model performed best with the image dimension 256x256 and thereby the classification report for this dimension is shown below.

Metric	Precision	Recall	F1-Score	Support
heavy traffic	0.400000	0.081967	0.136054	122.000000
light traffic	0.207547	0.297297	0.244444	111.000000
moderate traffic	0.240678	0.541985	0.333333	131.000000
no traffic	0.418605	0.179104	0.250871	201.000000
accuracy	0.265487	0.265487	0.265487	0.265487
macro avg	0.316707	0.275088	0.241176	565.000000
weighted avg	0.331869	0.265487	0.243936	565.000000

TABLE III: Classification report for MobileNetV2 with image size 256.

D. EfficientNetB0:

The accuracy of EfficientNetB0 with optical image dimensions of 128x128, 256x256 and 512x512 are 54.17, 54.67 and 57.84 percent respectively. This model performed best with the image dimension of 512x512 and thereby the classification report for this dimension is shown below.

Metric	Precision	Recall	F1-Score	Support
heavy traffic	1.000000	0.000000	0.000000	122.000000
light traffic	1.000000	0.000000	0.000000	111.000000
moderate traffic	1.000000	0.000000	0.000000	131.000000
no traffic	0.355752	1.000000	0.524804	201.000000
accuracy	0.355752	0.355752	0.355752	0.355752
macro avg	0.838938	0.250000	0.131201	565.000000
weighted avg	0.770807	0.355752	0.186700	565.000000

TABLE IV: Classification report for EfficientNetB0 with image size 512.

V. RESULT & ANALYSIS

For the analysis part, the measure of accuracy, f1-score, confusion matrix and ROC curve with AUC scores have been taken into consideration.

A. Accuracy Metric:

The table summarizing the model performance across different image dimensions along with the best accuracy for each model:

Model	Accuracy (128x128)	Accuracy (128x128)	Accuracy (128x128)	Best accuracy
CNN	35.83%	36.67%	36.84%	36.84%
ResNet50%	33.17%	33.50%	38.33%	38.33
MobileNetV2%	45.50%	49.33%	48.17%	49.33%
EfficientNetB0%	54.17%	54.67%	57.84%	57.84%

TABLE V: Comparison of accuracies between all model and size

For CNN, ResNet50 and EfficientB0 the accuracy increases with the dimension of image as they show the best accuracy at 512x512 image dimension but for MobileNetV2 the best accuracy is shown with image dimension of 256x256. For visualization a bar chart [Fig-2] of the models with the best accuracy is provided below.

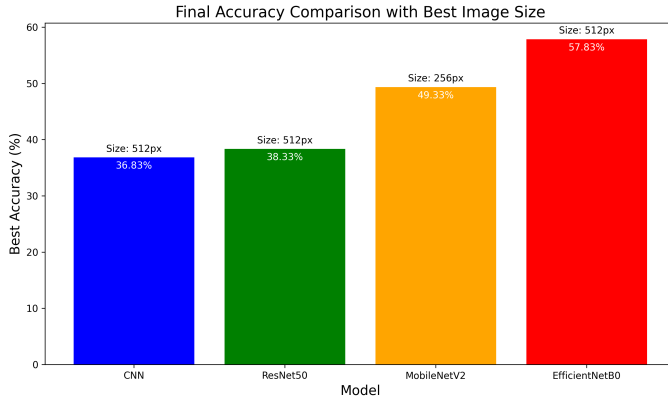


Fig. 2: Final Accuracy Comparison with Best Image Size

B. F1 score Metric:

The F1 score plot[Fig-3] indicates that EfficientNetB0 outperforms other models across all image sizes, maintaining a consistently high score above 0.9, with minimal variations. CNN shows improved performance with increasing image size up to a certain point but the F1 score stabilizes afterward. MobileNetV2 initially performs poorly but shows gradual improvement as the image size increases which suggests better adaptability to higher resolutions. In contrast, ResNet50 exhibits the weakest performance with F1 score declining significantly as the image size increases that highlights its sensitivity to image dimension changes. Overall, EfficientNetB0 is the most robust and effective model for this dataset.

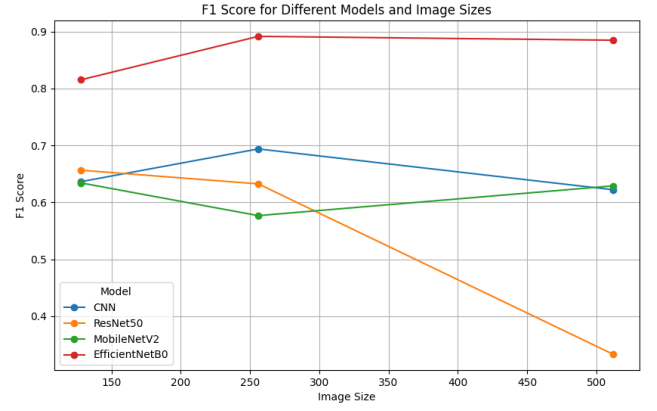


Fig. 3: F1 score for different models with different image sizes

C. Confusion Matrix Metric:

EfficientNetB0 (512px) demonstrates the best overall performance, particularly excelling in reducing misclassifications for heavy traffic and light traffic, while maintaining strong balance across all traffic conditions. ResNet50 (512px) performs exceptionally well for no traffic but struggles with heavy traffic and light traffic. CNN (512px) offers balanced performance but exhibits notable misclassifications in moderate traffic and no traffic. MobileNetV2 (256px) shows decent balance but has higher misclassifications for no traffic (class 3), making it comparatively weaker than EfficientNetB0 and ResNet50 for distinguishing between traffic conditions.

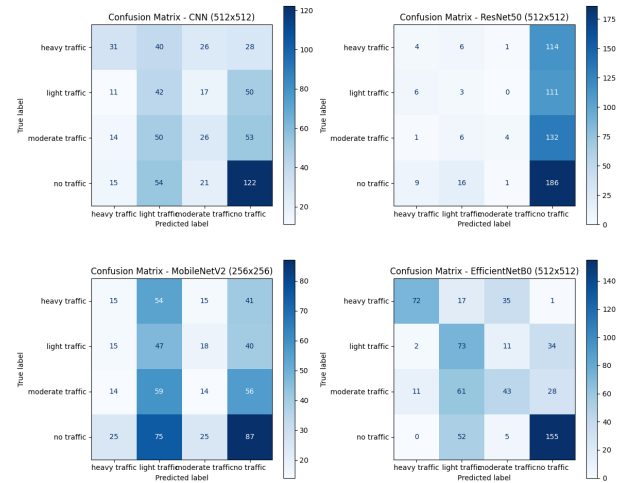


Fig. 4: Confusion Matrix of models in best performing image size

D. ROC Curve Metric:

The ROC curve [Fig-5] analysis shows that EfficientNetB0 with 512px image dimensions performs the best, achieving the highest AUC score of 0.80, followed by 0.79 at 256px and 0.75 at 128px, demonstrating consistent superiority. CNN

performs best at 128px with an AUC of 0.63, but its performance decreases with higher resolutions. ResNet50 shows relatively lower AUC scores, peaking at 0.54 for 128px and declining further at larger dimensions. MobileNetV2 performs moderately, with its best AUC at 0.52 for 128px and stable performance (0.50) across 256px and 512px. Overall, EfficientNetB0 stands out as the most effective model for this task.

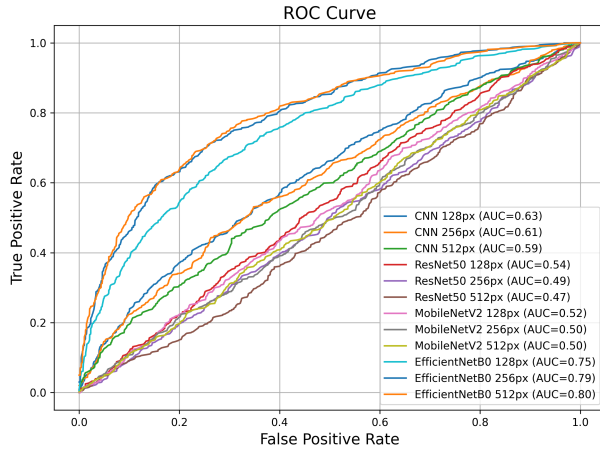


Fig. 5: ROC Curve with AUC Scores

Based on the analysis, EfficientNetB0 emerges as the best-performing model, achieving the highest accuracy (57.84%), the best F1 score (above 0.9 across image sizes) and the highest AUC (0.80) on the ROC curve. However, all models, including EfficientNetB0, face challenges due to the dataset being imbalanced and most importantly many pictures in the dataset explicitly depicts the scenario of two-way roads with conflicting traffic scenarios (e.g., heavy traffic in one lane and light traffic in the other) that causes confusion during classification of target variable, learning phase and in prediction too.

VI. CONCLUSION

In this study, we explored the performance of various deep learning models, including CNN, ResNet50, MobileNetV2, and EfficientNetB0 in order to predict traffic congestion levels in Dhaka using optical images. Among these models, EfficientNetB0 can be claimed as the best performing model achieving highest accuracy of 57.84% superior f1-score, greater success in regard to confusion matrix and best AUC on ROC curve. All the models performed poorly but the performance scores can be significantly increased if the limitation regarding two way roads can be eradicated with a dedicated dataset for similar studies. Despite the limitation, the results indicate the possibilities of AI based models to address the traffic congestion problem in Dhaka or other busy urban cities with similar studies which pave the way for further research and improvements in traffic prediction systems.

REFERENCES

- [1] Ali, Y., Rafay, M., Khan, R. D. A., Sorn, M. K., & Jiang, H. (2023). Traffic problems in Dhaka City: causes, effects, and solutions (Case study to develop a business model). *OALib*, 10(05), 1–15. <https://doi.org/10.4236/oalib.1109994>.
- [2] Ahad, K. A., Yasmeen, D., & Khan, R. R. (2023). Real time traffic congestion analyzer of Dhaka City. *IST Journal on Business & Technology*, 8(1), 17. ISSN: 2070-4135.
- [3] M. M. Chowdhury, M. Hasan, S. Safait, D. Chaki and J. Uddin, "A Traffic Congestion Forecasting Model using CMTF and Machine Learning," 2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), Kitakyushu, Japan, 2018, pp. 357-362, doi: 10.1109/ICIEV.2018.8640985.
- [4] M. M. Rahman, S. M. Ariful Hoque and M. I. Zaber, "Understanding Real Time Traffic Characteristics of Urban Zones Using GPS Data: A Computational Study on Dhaka City," 2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), Kitakyushu, Japan, 2018, pp. 514-519, doi: 10.1109/ICIEV.2018.8640961.
- [5] M. M. Rahman, M. M. Shuvo, M. I. Zaber and A. A. Ali, "Traffic Pattern Analysis from GPS Data: A Case Study of Dhaka City," 2018 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, 2018, pp. 1-6, doi: 10.1109/CONECCT.2018.8482371.
- [6] Jadhav, S., Vaghela, S., Tawde, S., Bharambe, R., & Mangalvedhe, S. (2020). Traffic Signal Management using Machine Learning Algorithm. *International Journal of Engineering Research & Technology (IJERT)*, 9(6). <https://doi.org/10.17577/IJERTV9IS060198>
- [7] Hassan, M., & Arabiat, A. (2024). An evaluation of multiple classifiers for traffic congestion prediction in Jordan. *Indonesian Journal of Electrical Engineering and Computer Science*, 36(1), 461. <https://doi.org/10.11591/ijeecs.v36.i1.pp461-468>
- [8] Arabiat, A., Hassan, M., & Almomani, O. (2024). WEKA-based machine learning for traffic congestion prediction in Amman City. *IAES International Journal of Artificial Intelligence*, 13(4), 4422. <https://doi.org/10.11591/ijai.v13.i4.pp4422-4434>