Unraveling Urban Traffic Congestion Patterns in Bangladesh: A Hierarchical Clustering Approach with DTW

Md. Babul Hasan[1][a], Manash Sarker[1][b]
[1]*Faculty of Computer Sc. and Engr., Patuakhali Sc. and Tech. University, Dumki, Patuakhali, Bangladesh*
{*bhyean16, manash.sarker*}*@cse.pstu.ac.bd*

Keywords: Traffic patterns, seasonal components, trend components, hierarchical clustering, and dynamic time warping (DTW).

Abstract: This research presents a comprehensive study on divisional traffic analysis and clustering in Bangladesh, leveraging Google Maps and image processing techniques for traffic intensity data collection across all divisions from January 2023 to June 2023. A total of 1,39,008 snapshots were captured at 15-minute intervals, yielding a detailed traffic dataset. We conducted an in-depth analysis of the collected time series data, focusing on its decomposition into trend, seasonal, and random components (Y = T * S * R). To enhance clustering accuracy, we proposed a modification technique by dividing traffic intensity (Y) by the random fluctuations (R) to minimize random noise in the data preprocessing stage. We implemented Modified Hierarchical Clustering with Dynamic Time Warping (DTW) for clustering, demonstrating superior similarities-pattern extraction compared to traditional hierarchical clustering. Our results identified four distinct traffic clusters. This study provides insights into regional traffic behaviors and offers a robust approach to clustering traffic data, contributing to Bangladesh's more effective traffic management strategies.

1 INTRODUCTION

Bangladesh is facing rapid growth in urbanization and motorization, which combine to cause severe traffic congestion in urban areas of the country. The scenario has worsened over the last ten years due to the rapid increase in vehicles and insufficient roads to accommodate them (Mahmud et al., 2012). Traffic congestion is a critical problem for a highly populated country like Bangladesh, where it causes traffic delays, waste of time, and an increase in vehicle emissions and fuel usage, leading to environmental and health problems. Bangladesh has gradually shifted from infectious to non-communicable diseases and injuries in the past few years (TRL et al., 2004). Limited resources invested for the development of transport facilities, the rapid population growth together with limited space available for new roads, coupled with the rapid rise in transport demand, the existence of a vast number of non-motorized vehicles on roads, and the lack of application of adequate and proper traffic management schemes are producing severe transport problems in almost all the urban areas of Bangladesh (Ali et al., 2023). Urban traffic congestion is a global issue, with local characteristics that affect a city's transportation system and people's everyday lives. Understanding and detecting congestion on different roads or areas of a city is very crucial for taking initiatives to reduce traffic congestion. Identification of various congestion patterns in a city is a necessary input for traffic management policy or systems. This includes developing more advanced traffic information systems to inform drivers about road conditions, pricing initiatives, and policy-making. Yet there are few works on predicting large-scale spatiotemporal patterns, and even fewer on predicting specific abnormal events such as traffic congestion, despite the interest from transportation researchers and practitioners. Machine learning and data mining have recently become critical methodological drivers for transportation research. Yet, there is still a lack of consensus on the best methods to use in many urban transportation contexts, and few studies have rigorously evaluated a range of methods. Our research aims to fill this gap by testing various machine learning methods for spatiotemporal prediction of urban traffic congestion in Bangladesh. This paper presents a novel approach to traffic congestion analysis in Bangladesh using a hierarchical clustering method combined with Dynamic

[a] https://orcid.org/0000-0000-0000-0000
[b] https://orcid.org/0000-0000-0000-0000

Time Warping (DTW) for time-series data analysis. Our primary contributions are as follows:

- We compiled a comprehensive dataset containing over 139,000 traffic snapshots collected from all divisions of Bangladesh over six months (January 2023 to June 2023) using Google Maps and image processing techniques.
- A data modification method was proposed to enhance clustering accuracy by eliminating random noise from the traffic intensity data, thereby improving the quality of the clustering process.
- We employed a modified hierarchical clustering algorithm, using DTW as the distance metric instead of traditional Euclidean distance. This approach significantly improved the alignment of traffic patterns over time, capturing similarities between traffic patterns.

This research offers insights into regional traffic patterns and provides a framework for more effective traffic management strategies. It enables urban planners to design tailored congestion mitigation policies for different areas of Bangladesh. This study integrates advanced clustering techniques with spatiotemporal analysis, offering a nuanced understanding of traffic congestion in rapidly urbanizing contexts like Bangladesh.

2 Related Work

Urban traffic congestion has been extensively studied due to its significant impact on transportation efficiency, economic costs, and quality of life. Researchers have employed various data-driven and machine-learning methodologies to analyze and manage traffic congestion patterns, aiming to develop effective strategies for urban traffic management. Xiong introduced an innovative method using Dynamic Time Warping (DTW) to detect spatiotemporal propagation patterns of traffic congestion (Xiong et al., 2023). Analyzing fine-grained vehicle trajectory data reveals how localized congestion events can propagate across road networks, providing new insights for managing urban traffic systems. Similarly, Chen employed taxi trajectory data to model the spread of traffic congestion across neighboring road segments, offering a method for anticipating and mitigating congestion through effective traffic control measures (Chen et al., 2018) . Zang applied a self-organizing map (SOM) to cluster traffic congestion patterns based on the Traffic Performance Index (TPI) in Beijing (Zang et al., 2023). The study identified specific congestion patterns for weekdays, weekends, and holidays, providing a temporal perspective on traffic management and policy planning. Kanchanamala explored Hadoop-based hierarchical clustering for large-scale traffic data analysis, demonstrating how big data analytics can improve the scalability and efficiency of traffic monitoring and management in megacities (Kanchanamala et al., 2016). Ambühl further contributed by analyzing macroscopic fundamental diagrams (MFDs) to track urban traffic rhythms over time, providing insights into long-term traffic management strategies (Ambühl et al., 2021). Wang proposed a Spatio-Temporal Non-Negative Matrix Factorization (ST-NMF) approach to address the challenges of analyzing noisy, high-dimensional data in large-scale urban networks (Wang et al., 2021). ST-NMF enhances traffic data reconstruction and predicts future traffic states by decomposing traffic states into spatial and temporal patterns. This approach provides a robust framework for managing intelligent transportation systems through a clearer understanding of spatio-temporal traffic dynamics. Akbar conducted a comprehensive analysis of traffic speeds in 1,200 cities across 152 countries, revealing that cities in more affluent countries tend to have faster travel speeds due to their larger urban areas and more extensive road infrastructure (Akbar et al., 2023b). The study found that uncongested speed, rather than congestion reduction, is the primary driver of faster travel speeds in wealthier countries. This finding underscores the importance of infrastructure investment in improving urban mobility. Li employed a weighted K-means clustering method to analyze traffic congestion patterns in Beijing, focusing on the effects of urban policies such as vehicle license plate restrictions (Li et al., 2023). Their study illustrates the potential of big data analytics for identifying spatial and temporal congestion patterns across different city districts, contributing valuable insights for traffic management strategies. Akbar investigated traffic congestion in Indian cities using simulated trip data, finding that uncongested speed plays a more significant role than congestion in determining travel speed differences across cities (Akbar et al., 2023a) . This challenges conventional beliefs that traffic management efforts should focus primarily on reducing congestion instead of emphasizing the need for infrastructure development. In the context of Bangladesh, our study builds upon these methodologies by employing a hierarchical clustering approach combined with Dynamic Time Warping (DTW) to analyze urban traffic patterns. This research collected traffic intensity data using Google Maps data and image processing techniques across all divisions of Bangladesh, identifying four distinct traffic clusters. By enhancing the clustering accuracy with a noise reduction technique, the study provides a robust approach to understanding regional traffic behaviors. It contributes valuable insights for more effective traffic management strategies in Bangladesh.

3 Methodology

3.1 Data Collection Strategy

Traffic intensity data were collected from all divisions in Bangladesh using a Google Maps data and image processing system. The data acquisition process involved capturing traffic snapshots at 15-minute intervals, leading to a comprehensive dataset of 1,39,008 snapshots. To capture traffic conditions across all divisions in Bangladesh, we employed a systematic data collection approach using GPS-enabled imaging technology.

Figure 1: Traffic Snapshot

The process began with capturing snapshots of traffic using Google Maps to obtain geolocated images of roads under study as shown in Fig. 1. To ensure that only the relevant portions of the road were analyzed, each image was cropped to a standardized size of 940x1440 pixels, focusing on the areas most pertinent to traffic flow and intensity.

A color masking technique was then applied to isolate traffic-related elements. Red, yellow, and green were highlighted, representing varying levels of traffic intensity, while all other colors were converted to black. This step effectively emphasized traffic density and flow information in Fig. 2.

Then, the images were subsequently divided into smaller segments in Fig. 3 using a grid-based approach, splitting each image into 3,384 grid cells of

Figure 2: Masked Image

20x20 pixels each, facilitating more granular analysis.

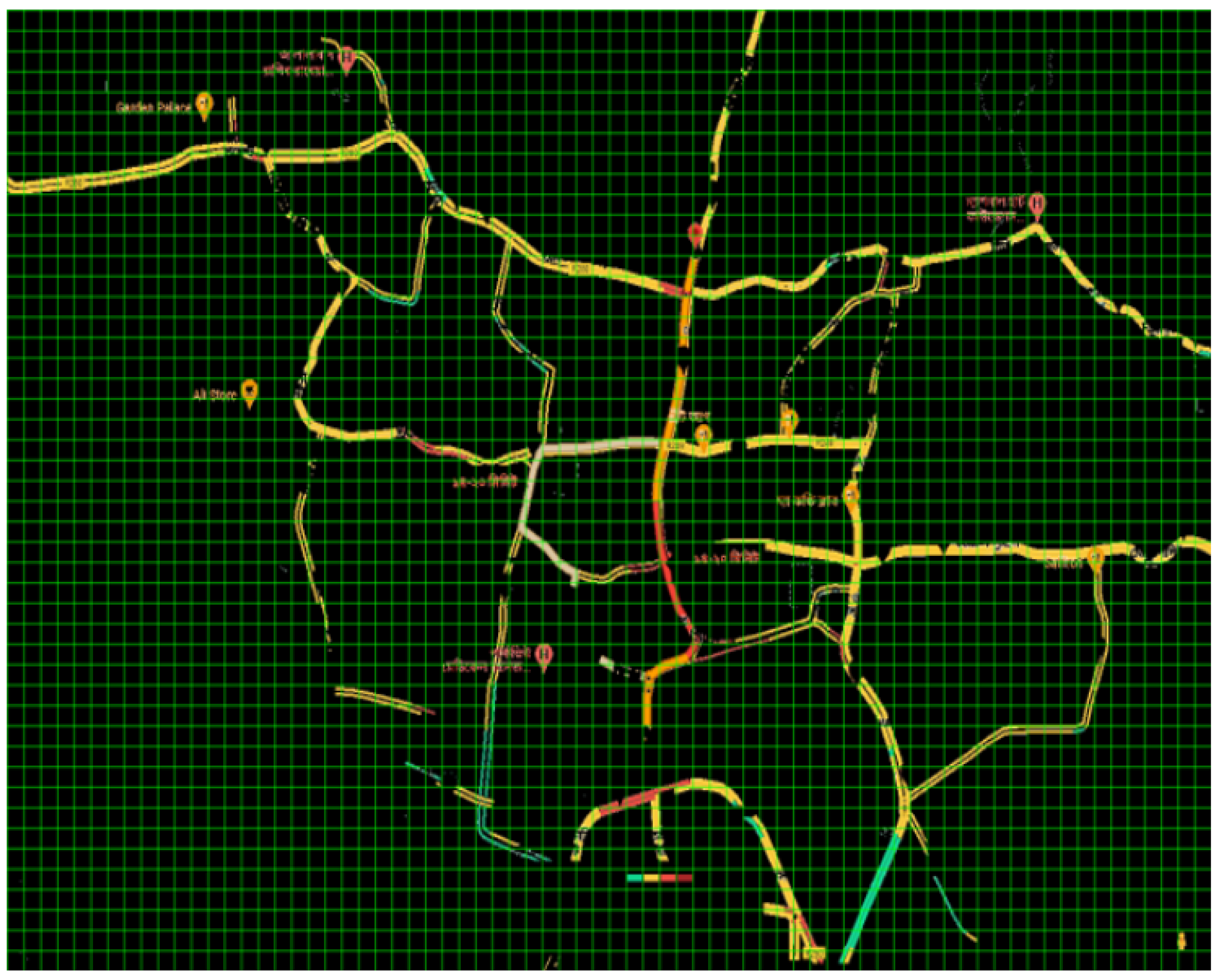

Figure 3: Image splitted to grid

Each grid cell's traffic intensity was determined based on the dominant color, assigning values Red = 3, Yellow = 2, Green = 1, and Black = 0. The overall traffic intensity of each snapshot was computed by summing the intensity values across all grid cells. This method allowed us to quantify and analyze traffic patterns systematically, creating a comprehensive dataset that accurately reflects traffic conditions across the targeted regions. This approach provides detailed temporal resolution, capturing the variability and complexity of traffic conditions across Bangladesh.

3.2 Time Series Analysis and Modification

The collected time series data were analyzed using Harvey's multiplicative formula $Y = T \times S \times R$ (Harvey,