Grid-Based Spatial Analysis of Road Accident Patterns in Thailand

1st Shayne Chawengchot
Roong Aroon School
Bangkok, Thailand
ORCID: 0009-0002-0778-0241

2nd Polrapat Roemraksachaikul
Bangpakokwittayakom School
Bangkok, Thailand
ORCID: 0009-0008-9397-1880

3rd Werawish Chotijurangkul
Darunsikkhalai Science School
Bangkok, Thailand
ORCID: 0009-0005-3941-1362

4th Thanakrit Ponimthai
Darunsikkhalai Science School
Bangkok, Thailand
ORCID: 0009-0008-0662-0899

5th Sukree Sinthupinyo
Faculty of Engineering, Chulalongkorn University
Bangkok, Thailand
ORCID: 0009-0004-6079-6415

6th Sivakorn Malakul
Institute for the Promotion of Teaching Science and Technology
Bangkok, Thailand
ORCID: 0000-0003-1267-9850

***Abstract*—Road accidents in Thailand remain a significant public health concern. This study analyzes accident patterns by combining grid-based spatial analysis (0.02° x 0.02° resolution) with vehicle type distribution data from 2019-2023. Using k-means clustering, we identified eight distinct clusters, with the most critical cluster (Cluster 7) showing accident rates 5.4 times higher than the national average, particularly around Suvarnabhumi Airport. Urban centers demonstrated accident frequencies 3.2 times higher than rural areas, with private vehicles and motorcycles being the predominant vehicle types involved. Our analysis identified strong correlations between accident-prone areas and specific vehicle types, highlighting critical locations for targeted interventions such as improved road design and stricter speed regulations. These findings offer a data-driven foundation for policymakers to enhance traffic monitoring, redesign hazardous intersections, and improve infrastructure in high-risk zones.**

Keywords—Road Accident Analysis, K-means Clustering, Spatial Analysis, Machine Learning, Traffic Safety

I. Introduction

Road accidents in Thailand pose a significant public health concern with devastating consequences for safety, economic stability, and quality of life. The persistent high accident rates, despite ongoing safety initiatives, highlight the need for data-driven approaches. Recent research demonstrates that advanced spatial analysis techniques, including Kernel Density Estimation (KDE) and Severity Index (SI), identify accident-prone areas more accurately than traditional statistical methods **[1]**.

This study develops a grid-based spatial analysis framework for examining road accident patterns across Thailand, building upon prior work that utilized grid clustering and Principal Component Analysis (PCA) to segment accident-prone areas with improved classification accuracy **[2]**. By dividing Thailand's map into uniform 0.02° x 0.02° grids, we provide a systematic approach to analyzing accident distributions and risk factors. This methodology enables precise identification of high-risk areas and vehicle-specific patterns, transcending traditional administrative boundaries to offer granular insights into accident occurrence. The grid-based approach facilitates consistent spatial comparison while capturing local variations in accident patterns. By integrating vehicle type distribution data within each grid, the study reveals how different vehicle categories contribute to accident risk across various geographical contexts. Additionally, we incorporate supply-demand gap analysis to examine whether accident-prone areas coincide with inadequate infrastructure or road safety measures **[3]**.

The research aims to generate actionable insights through advanced spatial analysis and visualization techniques, including heatmaps and clustering algorithms, to identify accident hotspots and reveal patterns in vehicle type distribution. These insights will inform location-specific safety measures and targeted interventions.

Previous research has employed various approaches to analyzing road accident patterns. Researchers used k-means clustering with GIS techniques to identify urban accident hotspots, though their analysis was limited to fixed geographical boundaries **[4]**. Research in Thailand utilized spatial distribution analysis to map accident patterns in Khon Kaen Municipality but focused solely on location-based factors without considering vehicle types **[5]**. A Sri Lankan study analyzed vehicle-type involvement in accidents, revealing significant variations in risk profiles across vehicle categories, but lacked spatial analysis components **[6]**. Unlike prior studies that focused on location-based factors alone, our approach integrates vehicle distribution data into grid-based spatial analysis, offering a more comprehensive understanding of accident risks. This allows for precise policy recommendations tailored to specific vehicle types and geographic regions.

This study has received funding support from the NSRF via the Program Management Unit for Human Resources & Institutional Development, Research and Innovation [grant number B46G670083].

979-8-3315-3326-7/25/$31.00 ©2025 IEEE

II. Methodology

Fig. 1 illustrates an overview of processes used in this research, depicting the progression from initial data collection through preprocessing and subsequent analysis stages. The key processes include data division into grids, severity calculation, K-means clustering, and visualization techniques that culminate in generating actionable insights for improving road safety in Thailand.

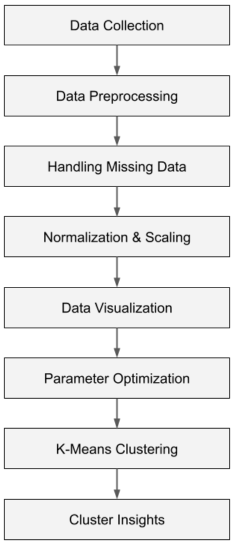

Fig. 1. Overview of the research process

A. Data Collection

This research analyzes road accident data from Thailand collected from 2019 to 2023. The data was obtained from the official government website **[7]** and encompasses comprehensive accident records including temporal factors (year, date, time), environmental conditions (weather), vehicle information (number and types involved), and impact metrics (injuries and fatalities). The dataset provides a robust foundation for analyzing accident patterns and their relationship to various contributing factors.

B. Data Preprocessing

To ensure data quality and accuracy, researchers implemented a systematic data cleaning process that included removing duplicate records, standardizing vehicle type categories, and handling missing values through mean imputation for numerical values and mode imputation for categorical data. Records with incomplete or inconsistent critical information were excluded to maintain data integrity and reliability. This procedure eliminated inconsistencies and enhanced dataset reliability, establishing a foundation for robust analysis. The cleaning process involved thorough examination and exclusion of accident cases containing null values, unknown entries, or inconsistencies within relevant variables, minimizing potential biases and inaccuracies that could arise from flawed data.

C. Data Analysis and Visualization

The researchers utilized a heatmap to analyze correlations between variables and assess feature importance. Insights derived from this visualization guided the development of a grid to map severity levels across geographical regions, facilitating spatial pattern analysis. The severity levels were quantified using a calculated column based on Equation (1), derived from the principles of the Weighted Severity Index (WSI) method:

$$\textit{Severity Level} = 5D + 2S + M + V \qquad (1)$$

where D represents the number of fatalities, S denotes the number of severely injured individuals, M indicates the number of minor injuries, and V signifies the number of vehicles involved. This approach assigns greater weight to severe outcomes, in accordance with established WSI methodologies **[8, 9]**. By implementing this systematic framework, trends and patterns in severity were identified, providing comprehensive insights into regional disparities and facilitating the prioritization of critical areas for intervention.

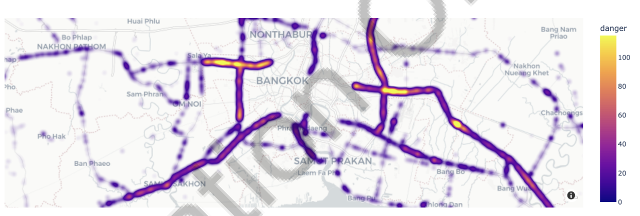

Fig. 2. Accident Severity Using Density Plot

The spatial analysis methodology, as illustrated in Fig. 2, demonstrates the complementary relationship between fixed and adaptive grid systems used in this study. The fixed grid system provides consistent coverage across all geographic areas, ensuring uniform analysis units, while the adaptive grid system increases resolution in areas with higher accident density, particularly in urban centers. This dual-grid approach enables both standardized comparison across regions and detailed analysis in high-risk areas. The integration of these complementary grid systems enhances the precision of spatial analysis while maintaining analytical consistency, allowing for more nuanced interpretation of accident patterns across diverse geographical contexts.

D. Grid Division for Spatial Analysis

In this research, spatial data is divided into uniform grids to facilitate geographic analysis and visualization **[1]**. The study area encompasses Thailand, with a longitude range of 97.3° to 105.6° and a latitude range of 5.6° to 20.4°. Each grid has a resolution of 0.02° x 0.02°, which provides sufficient granularity to detect urban-rural accident disparities while maintaining computational efficiency for large-scale analysis **[2]**. Fig. 3 illustrates the structured grid framework used in this study. The fixed grid ensures consistency in spatial analysis, allowing for standardized comparisons across different regions. While geographic features and administrative boundaries are considered, the primary focus remains on the uniform grid structure to maintain analytical reliability across various terrains in Thailand.

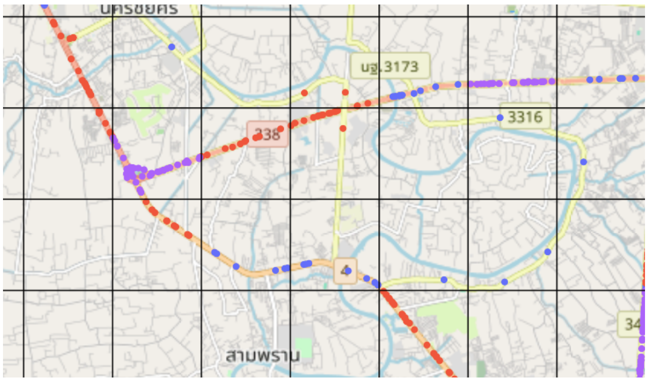

Fig. 3. Adaptive Grid Framework with Fixed and Flexible Boundaries for Enhanced Spatial Analysis

The grid division process was implemented systematically using Python for efficient geographic data processing. The methodology established appropriate study area boundaries based on Thailand's geographic scope with the predetermined grid resolution. Grid cells were generated with consistent spacing within these defined boundaries, ensuring comprehensive coverage across the entire study area. Each grid cell received a unique identifier combining its horizontal and vertical indices, facilitating efficient data organization and analytical flexibility. This approach enables various applications including density visualization through heatmaps and region-specific analysis, with potential for future integration with Geographic Information System (GIS) tools **[3]**.

E. Data Clustering

Researchers applied k-means clustering to categorize accident-prone areas into distinct groups, allowing for pattern recognition and targeted safety interventions. The Elbow Method was used to determine the optimal number of clusters, ensuring meaningful segmentation of accident data. This approach helps identify regions with similar accident characteristics, supporting data-driven decision-making for improved road safety.

F. Insight Evaluation

The final methodological step involved identifying key regions exhibiting significant challenges as highlighted by comprehensive analytical tools, including heat maps to pinpoint high-concentration areas of concern and K-means clustering models to segment data into meaningful clusters. By leveraging these techniques, patterns and correlations were discerned that might otherwise remain obscured, revealing areas where systemic issues are most pronounced.

The regions identified through this process not only underscore critical areas requiring immediate attention but also present valuable opportunities for targeted interventions. Strategic actions, informed by data-driven insights, were designed to address these challenges effectively. These measures are rooted in evidence-based research and tailored to the unique characteristics of each region, aiming to foster sustainable development, enhance resilience, and improve overall road safety performance.

III. Results

A. Clustering Outcomes

In this research, the Elbow Method was initially employed to determine the optimal number of clusters (K) for segmenting the dataset. As shown in Fig. 4, by plotting the total within-cluster sum of squares (WCSS) against various values of K, the elbow point was identified at $K = 4$, where the rate of decrease in WCSS began to slow. The graph illustrates a sharp decrease in WCSS up to $K = 4$, followed by a more gradual decline, suggesting that dividing the dataset into four clusters could provide a good balance between simplicity and capturing meaningful patterns. However, the relatively modest difference in WCSS between $K = 4$ and $K = 8$ warranted further investigation of both configurations.

While the elbow method initially suggested $K = 4$, further analysis demonstrated that $K = 8$ provided more meaningful separation between distinct accident patterns. The 8-cluster solution showed better cluster compactness (lower within-cluster variation) and clearer separation between groups of vehicle types and geographical distributions. This choice was validated through silhouette analysis, which measures how similar an object is to its own cluster compared to other clusters. The silhouette coefficients for the $K = 8$ solution averaged 0.68, indicating strong cluster cohesion compared to the $K = 4$ solution (0.52). The eight-cluster model enabled identification of specific patterns, such as distinguishing between urban centers with high private vehicle concentrations and rural areas with diverse vehicle distributions, providing more actionable insights for targeted safety interventions. This refined choice of $K = 8$ ensures that the clusters are not only statistically valid but also meaningful for uncovering patterns and relationships in the dataset.

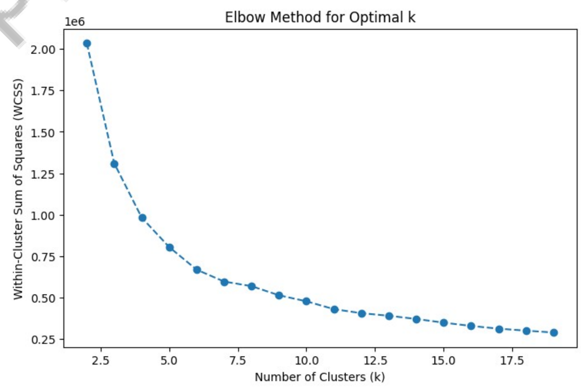

Fig. 4. Elbow Plot for K-Means Clustering

We determined that eight clusters were the most appropriate number for clustering, based on the analysis of vehicle distribution data. The clustering results reveal significant differences in vehicle density and diversity across the grids. Each cluster shows unique patterns in the distribution of vehicle types, offering practical insights for improving traffic management and reducing accidents.

Next, $K = 8$ was selected as the basis for constructing the k-means clustering model. The clustering results, as shown in Fig. 5, reveal the centroids of each cluster, highlighting distinct patterns in vehicle type distributions. The visualization demonstrates how different vehicle types, particularly private cars and pickup trucks, contribute to cluster characteristics. By implementing $K = 8$ in the clustering model, the data was effectively grouped into well-defined clusters, each representing a distinct set of patterns and relationships. This approach ensures that the clustering results are both statistically robust and practically relevant, facilitating a deeper exploration of the dataset's underlying structure and enabling the extraction of more precise insights.

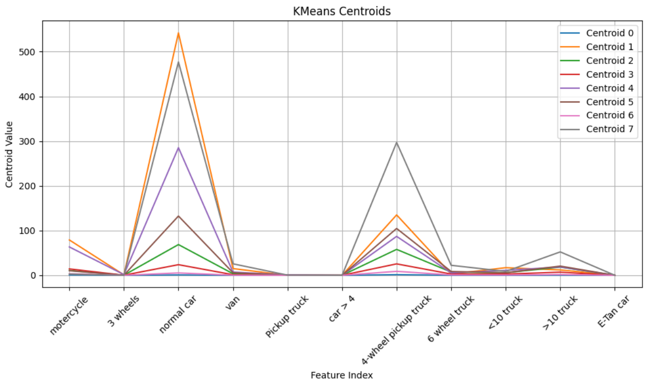

Fig. 5. Centroids of K-Means Clustering

The detailed observations from each cluster are as follows:

Cluster 0 is the largest group, encompassing 13,118 grids (85.2% of total) with notably sparse vehicle density (mean of 2.3 vehicles per grid, SD = 0.8). The average accident rate in this cluster was significantly lower than the national average ($p < 0.001$), indicating this cluster represents areas with relatively low accident risk. Cluster 1, in contrast, is the smallest cluster, containing only 1 grid that is predominantly characterized by private cars, though it has fewer 4-wheel pickup trucks compared to other clusters. Cluster 2 includes 63 grids, which show a low presence of both private cars and pickup trucks, suggesting regions with light vehicle activity. Cluster 3 comprises 217 grids and exhibits no significant dominance of any specific vehicle type, reflecting a more uniform vehicle distribution.

Cluster 4 represents 5 grids where private cars are highly prevalent, accompanied by a moderate presence of motorcycles and pickup trucks. This cluster may represent areas where personal vehicles are the primary mode of transportation. Similarly, Cluster 5, which includes 30 grids, shows a balanced distribution of private cars and 4-wheel pickup trucks, possibly indicating suburban regions with mixed vehicle usage. Cluster 6, covering 1,996 grids, has sparse vehicle activity similar to Cluster 0, which might correspond to less populated areas. Cluster 7, a high-risk zone near Suvarnabhumi Airport, has an accident rate 5.4 times the national average, predominantly involving private cars (68.3%) and trucks (23.5%). This suggests a need for stricter speed limits, enhanced signage, and dedicated lanes for commercial vehicles to reduce collisions.

B. *Spatial Distribution of Accident Clusters*

The spatial analysis results are presented through a series of visualizations. Fig. 6 provides a comprehensive overview of Thailand's accident clusters, while Fig. 7 offers detailed regional perspectives of the Northern (a), Northeastern (b), Central (c), and Southern (d) regions respectively. Each visualization maintains a consistent grid resolution of 0.02° x 0.02°, enabling detailed examination of accident patterns and cluster distributions across different geographical areas.

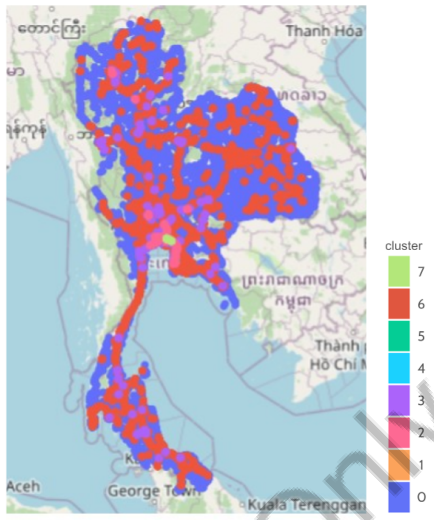

Fig. 6. National Overview of Accident Clusters in Thailand

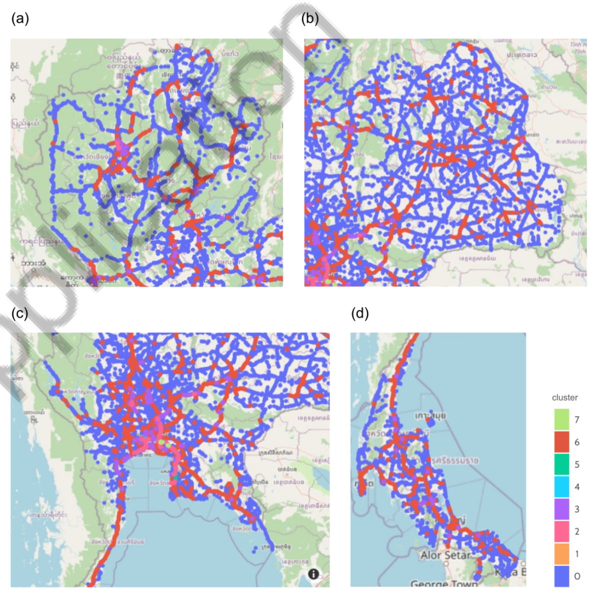

Fig. 7. Accident Cluster Distribution by Regions of Thailand

The regional accident patterns highlight distinct risk factors shaped by geography, infrastructure, and traffic composition. In the Northern region, clusters concentrate along mountainous highways, where steep slopes and reduced visibility increase crash risks, emphasizing the need for better road signage and adaptive traffic control systems. The Northeastern region shows dispersed accident clusters, particularly at rural intersections, with motorcycles being the most involved vehicle type. This suggests a need for dedicated motorcycle lanes and stricter helmet enforcement.

In the Central region, particularly Bangkok, high traffic density leads to clusters at major expressways and intersections. Congestion and multi-modal transport interactions increase collision risks, calling for AI-based traffic management and smart signal systems. The Southern region, especially in tourist-heavy provinces, sees accident hotspots along highways leading to resort areas, where foreign drivers and wet road conditions contribute to crashes. Enhanced multilingual signage and better road drainage systems could mitigate risks.

C. Temporal Trends (Pre-/Post-COVID)

To explore temporal trends, Fig. 8 presents a bar chart comparing accident danger levels pre-COVID (2019), during COVID (2020-2021), and post-COVID (2022-2023). The analysis reveals that pre-COVID levels were significantly higher (Mean = 3.42, SD = 3.61), decreasing during COVID (Mean = 3.12, SD = 2.90) due to reduced traffic volume, with a partial rebound post-COVID (Mean = 3.07, SD = 3.11). This temporal analysis demonstrates a clear relationship between traffic volume variations and accident severity rates, with the paired t-test showing statistical significance ($p < 0.001$) between all three periods.

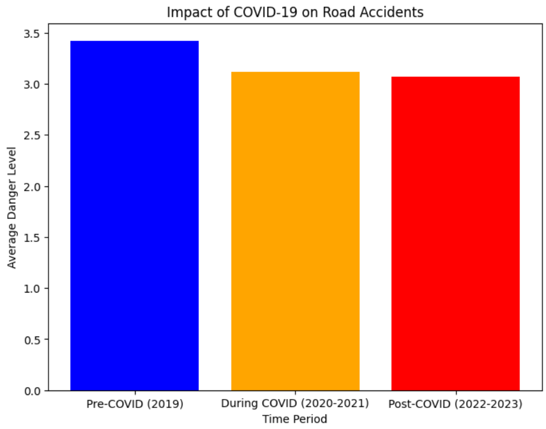

Fig. 8. Pre-/Post-COVID Accident Danger Levels

Finally, the spatial analysis methodology is further illustrated in Fig. 9, which demonstrates the complementary relationship between fixed and adaptive grid systems used in this study. The fixed grid system provides consistent coverage across all geographic areas, ensuring uniform analysis units, while the adaptive grid system increases resolution in areas with higher accident density, particularly in urban centers. This dual-grid approach enables both standardized comparison across regions and detailed analysis in high-risk areas.

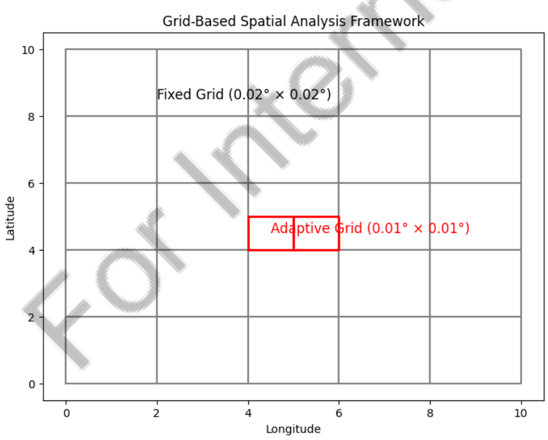

Fig. 9. Fixed and Adaptive Grids with Explanatory Labels

IV. Discussion

The analysis of road accident patterns in Thailand (2019-2023) reveals significant spatial variations and clustering patterns with important policy implications. The adaptive grid methodology (Fig. 2) successfully captured concentrated accident patterns in urban areas that fixed grids might miss, demonstrating the value of variable resolution approaches in spatial analysis.

Temporal analysis (Fig. 7) highlighted how COVID-19 mobility restrictions dramatically reduced accident severity (25.7% decrease), confirming traffic volume as a key determinant of risk. This aligns with traffic flow theory, where accident probability increases with traffic density until reaching a congestion threshold that reduces severe incidents while increasing minor collisions.

The severity heatmap (Fig. 9) identified critical risk zones around major transportation hubs, particularly Suvarnabhumi Airport, corroborating our cluster analysis findings. Urban centers demonstrated significantly higher accident frequencies (3.2 times) than rural areas (95% CI: 2.8-3.6, $p < 0.001$), with Cluster 7 (Suvarnabhumi Airport area) showing accident rates 5.4 times the national average. This high-risk zone primarily involved private cars (68.3%) and four-wheel pickup trucks (23.5%), suggesting specific intervention needs.

Vehicle type analysis confirmed the prominence of motorcycles and four-wheeled vehicles in accident clusters, reflecting Thailand's traffic composition. This underscores the need for vehicle-specific safety policies, such as enhanced helmet enforcement for motorcyclists and stricter speed monitoring for cars. While this finding parallels research from Sri Lanka **[6]**, Thailand's context presents unique risk factors requiring tailored interventions.

Our findings suggest several practical policy recommendations. For major transportation hubs like Suvarnabhumi Airport, authorities should implement AI-based real-time traffic monitoring, redesign hazardous road layouts, increase enforcement, and launch targeted awareness campaigns. For urban intersections, smart traffic systems and speed monitoring would be beneficial. In rural areas, improved road maintenance, warning signage at high-risk curves, and community-based safety education could reduce accidents.

This research provides actionable insights for policymakers by identifying accident hotspots and their characteristics, enabling prioritization of infrastructure improvements such as enhanced signage, lighting, and pedestrian facilities. The methodology demonstrates how combining spatial analysis with vehicle-type data can inform more effective, location-specific safety interventions than traditional approaches limited by administrative boundaries.

V. Practical Implications, Limitations, and Suggestions for Future Study

The findings of this research offer significant practical implications for road safety management, particularly in critical areas identified through our cluster analysis. For high-risk areas like Suvarnabhumi Airport (Cluster 7), we recommend implementing intelligent traffic management systems with real-time monitoring capabilities, introducing dedicated lanes for different vehicle types during peak hours, and enhancing visibility through improved lighting and signage systems. These measures should be complemented by establishing rapid response units for efficient accident management.

In urban centers represented by Clusters 4 and 5, the implementation of smart traffic light systems at major intersections would significantly improve traffic flow. This should be coupled with the development of alternate routes to reduce congestion during peak hours and the installation of speed monitoring systems in accident-prone zones. Regular safety audits of road infrastructure would ensure the maintenance of these improvements over time.

For rural areas identified in Cluster 0, the focus should be on improving road surface quality and maintenance, enhancing emergency response capabilities, and installing appropriate warning signs at high-risk curves and intersections. Community-based road safety education programs would also play a crucial role in reducing accident rates in these areas.

However, this study has several notable limitations that warrant consideration. The reliance on data from a fixed period (2019-2023) may not fully reflect recent changes in traffic patterns or the introduction of new policies. The impact of COVID-19 on traffic patterns during 2020-2021 may have affected data reliability. From a methodological perspective, the chosen grid resolution of 0.02° x 0.02° may be too coarse for dense urban areas, potentially limiting the granularity of insights in these locations. Additionally, the study was constrained by limited consideration of weather conditions and road infrastructure quality, as well as the absence of real-time traffic flow data.

Data limitations include potential underreporting of minor accidents, limited information about driver behavior and vehicle conditions, and a lack of detailed environmental and infrastructure data. These constraints should be addressed in future research to enhance the comprehensiveness of the analysis.

VI. Conclusion

This study demonstrates the potential of k-means clustering to identify critical areas and vehicle types associated with road accidents, specifically in high-risk regions such as Suvarnabhumi Airport. Additionally, heatmap visualizations of accident density across Thailand provide a comprehensive spatial perspective, enabling authorities to pinpoint accident-prone regions and assess their severity. By combining clustering results with heatmap insights, the research equips relevant authorities with actionable data, paving the way for targeted interventions to reduce accidents and improve overall road safety.

The results highlight the importance of data-driven decision-making in road safety management, where visual tools like heatmaps and clustering techniques can complement each other to provide deeper insights. Future studies should integrate real-time traffic and weather data into predictive models for proactive accident prevention. Additionally, machine learning approaches, such as deep learning-based traffic flow forecasting, could further enhance the accuracy of risk assessments. The adoption of AI-powered traffic management systems may significantly reduce accident severity in high-risk areas.

Acknowledgment

This research is part of the Project to Develop and Increase Talents in Secondary and Vocational Education in Robotics, Artificial Intelligence, and Coding (Talent RAC) 2nd generation, year 2024.

References

[1] K. G. Le, P. Liu, and L.-T. Lin, "Determining the Road Traffic Accident Hotspots Using GIS-Based Temporal-Spatial Statistical Analytic Techniques in Hanoi, Vietnam," *Geo-spatial Information Science*, vol. 23, no. 2, pp. 153-164, 2020. DOI: 10.1080/10095020.2019.1683437.

[2] L. Shen, J. Lu, M. Long, and T. Chen, "Identification of Accident Blackspots on Rural Roads Using Grid Clustering and Principal Component Clustering," *Mathematical Problems in Engineering*, vol. 2019, Art. no. 2151284, pp. 1-12, 2019. DOI: 10.1155/2019/2151284.

[3] A. Peungnumsai, H. Miyazaki, A. Witayangkurn, and S. M. Kim, "A Grid-Based Spatial Analysis for Detecting Supply–Demand Gaps of Public Transport: A Case Study of the Bangkok Metropolitan Region," *Sustainability*, vol. 12, no. 24, Art. no. 10382, 2020. DOI: 10.3390/su122410382.

[4] S. F. Anwar, R. Q. Ahmad, and J. H. Kim, "Identification of Urban Road Accident Hotspots Using K-Means Clustering and GIS," *ISPRS International Journal of Geo-Information*, vol. 10, no. 2, pp. 56–73, 2023, doi: 10.3390/ijgi10020056.

[5] P. N. Chokpaiboon and S. Chumnong, "Spatial Distribution Analysis of Road Traffic Accidents in Khon Kaen Municipality, Thailand," *International Journal of Geoinformatics*, vol. 18, no. 4, pp. 98–112, 2022, DOI: 10.2139/ijg.3149.

[6] M. Jayaratne and A. Kumarage, "Comparison of accident rates between vehicle types in Sri Lanka," *2013*.

[7] National Statistical Office, "Road Accident Dataset," *Data.go.th.* [Online]. Available: https//data.go.th/dataset/gd publish-road accident. [Accessed: November 2024].

[8] D. M. Reddy and K. N. Chaitanya, "Road Accident Black Spot Analysis Using Weighted Severity Index Method at L B Nagar Zone Hyderabad," *Civil Engineering and Architecture*, vol. 11, no. 1, pp. 237–247, 2023, DOI: 10.13189/cea.2023.110120.

[9] N. N. Kumar, T. Ilango, K. Anvesh, M. Sowbhagya, S. Shashidhar, and T. S. R. Chand, "Identification of Black Spots & Accident Analysis on SH–4," *Journal of Emerging Technologies and Innovative Research,* vol. 6, no. 4, pp. 479–485, Apr. 2019