Emergency Response Optimization for Traffic Accidents in Washington, D.C.

Sambit Sahoo
ssahoo@umd.edu
University of Maryland - College Park
USA

Abstract

Rapid emergency response time is critical in mitigating the severity of traffic accidents, reducing fatalities, and ensuring timely aid for injured individuals. We utilized machine learning models to analyze high-risk areas for accidents and strategically place ambulances. The framework was designed to improve emergency response time, reduce fatalities, and enhance the efficiency of resource distribution, particularly in Washington, D.C. This study demonstrates how machine learning and geospatial analysis can be applied to dynamically adjust emergency response strategies. Spatial analysis techniques, such as Grid-Based Density Analysis and Kernel Density Estimation, were used to generate visualizations to highlight high-risk areas requiring readily available medical attention. K-Means Clustering was employed to identify optimal ambulance placements based on the accident distribution patterns. To enhance realism, Network Analysis with Dijkstra's algorithm was applied to incorporate road constraints and estimated travel times to offer a more dynamic model for ambulance routing. We evaluated the success of our framework by comparing ambulance arrival times, average, median, and maximum, as well as the percentage of accidents reached within predefined response time thresholds, between the current and proposed ambulance configurations. By minimizing this response time, the framework aims to ensure the quickest possible aid for all traffic accidents, prioritizing efficiency in emergency interventions. Our results revealed a discrepancy between current ambulance locations and areas with the highest need, suggesting opportunities for improvement for resource allocation. This research highlights the potential of geographic information science and data-driven methods to enhance public safety infrastructure and demonstrates adaptability to other urban environments facing similar challenges.

1 Introduction

Road traffic accidents are a significant threat to public safety and health in the United States. Each year, there are approximately 6 million crashes and over 40,000 casualties, along with hundreds of thousands of injuries. Traffic-related injuries are a leading cause of death in the United States. Motor vehicle crashes are especially devastating for individuals between the ages of 1 and 54, where they remain the leading cause of death [8]. In Washington, D.C., traffic fatalities have been rising in more recent years despite efforts to reduce accidents through infrastructure improvements and stricter traffic regulations. Although the United States accounts for a significant portion of global traffic-related fatalities, urban centers such as Washington, D.C., face the compounded challenges of dense populations, complex traffic patterns, and aging infrastructure. These factors make efficient emergency response time even more critical.

Reducing the consequences of traffic accidents depends on how quickly emergency services can reach the scene and provide medical assistance. In urban environments like Washington, D.C., traffic density fluctuates throughout the day, leading to delays and unpredictability in emergency routing. These delays significantly impact public safety. Despite established emergency response protocols, current systems often suffer from inefficiencies in dispatch coordination and routing, leading to suboptimal arrival times. A data-driven approach that accounts for various factors that affect response time can enable rapid and effective emergency interventions. In this study, we develop a data-driven framework to optimize ambulance placement and routing by employing geospatial analysis and machine learning techniques.

While data-driven methods can improve emergency response times, prior research has explored several strategies for optimizing emergency response systems. For example, studies have emphasized machine learning techniques to evaluate the placement of healthcare facilities [10, 15]. Additionally, there are studies on resource allocation during natural disasters, which typically involve ambulances. Although the circumstances are different, they provide valuable insights on optimizing emergency response under time-sensitive conditions [7]. Some researchers apply machine learning techniques like Random Forest for predicting demand [1], while others have used clustering methods, such as hierarchical clustering, to group accident hotspots [12]. While these methods offer important insights, our approach enhances clustering by incorporating geospatial analysis and machine learning to create a dynamic, adaptable model for identifying high-need zones in real time.

In this study, we propose a data-driven framework to optimize emergency response times in Washington, D.C. by placing the ambulances in high-risk areas. Using historical traffic accident data, we apply geospatial analysis and machine learning techniques to identify high-risk zones. First, we implement Nearest Neighbor Analysis to determine the spatial distribution of the traffic accidents. The results of this influenced our methodology of proposing a cluster-based approach to improve the placement of ambulances around Washington, D.C. Our approach integrates spatial analysis methods, such as Grid-Based Density Analysis and Kernel Density Estimation, to visualize accident hotspots. By analyzing past accidents, emergency service locations, and coverage areas, we identify bottlenecks and frequent accident zones to redistribute teams and optimize travel paths to incident sites. Then, K-Means Clustering is used to determine optimal ambulance placements. Lastly, we incorporate Network Analysis with Dijkstra's algorithm to simulate

realistic routing based on the city's road network for the current and proposed configurations. We compare the performance of these setups by evaluating the response times and coverage. For response times, we examine the average, median, and maximum response times. For coverage, we compute the percentage of accidents that can be reached within 1, 2, 3, 5, and 10 minutes. This integrated approach offers actionable insights for first responders and policymakers, ultimately enhancing public safety in Washington, D.C.

2 Related Work

Optimizing response time is crucial since delays can significantly impact people's lives and emergency response system performance. Researchers have proposed various approaches to improve response time. These strategies include identifying common trends in car accident data, using geospatial techniques for ambulance placement, optimizing ambulance routing, forecasting accidents, and maximizing ambulance coverage.

2.1 Identification and Clustering

Dai [4] investigated pedestrian crash records from the early 2000s in the urban region of Atlanta, Georgia. Using spatial clustering, the author examined the factors that contributed to the clusters. Rather than just looking at the traffic volumes, the study explored personal and environmental factors. For example, the personal factors included sex, age, and intoxication levels of drivers and pedestrians. The environmental factors included weather, light, and surface conditions. This study presented preliminary results of factors that affect accidents. In our study, we examined time of day and seasonal patterns, but explored other techniques to help injured people in the aftermath of crashes.

This paper [12] introduced the generation of accident groups, then identified the locations to place ambulances in the center of the group of accidents. The authors developed a double standard model to maximize the coverage of a limited number of ambulances over the accident clusters. Although we do not incorporate this algorithm into the work, it is important to note their methodology of limited ambulances for many accident clusters and allocating the resources efficiently. In our study, we are limited to utilizing a total of 43 public ambulances for the large region of Washington, D.C.

2.2 Resource Allocation

Yunus and Abdulkarim [15] introduced utilizing Nearest Neighbor Analysis to calculate the distribution pattern of emergency healthcare facilities, ambulances, and road traffic crash incident places. Additionally, they used Network Analysis to calculate the shortest and closest route between ambulances and road traffic crashes in terms of time and distance. However, the authors did not modify the placement of the ambulances. In our study, we similarly utilize Nearest Neighbor Analysis for high-level analysis and Network Analysis for computing the travel time of the ambulances. Unlike their approach, we propose a method to optimize the placement of the ambulances, building upon their work.

This paper [13] examined a query model that allocates emergency resources using real-time traffic and accident rate data. This improved the efficiency of response operations and reduced the impact of traffic accidents. In contrast, our approach utilizes machine learning algorithms to understand the high-risk areas of traffic accidents to give a stronger data-driven approach to allocating ambulances.

This study [1] introduced predicting the demand for emergency services without historical demand data and utilizing machine learning models to predict travel time. They used a random forest model and found a 43.3% to 64.2% improvement in prediction accuracy over baseline approaches. Our approach utilizes multiple machine learning techniques to compare our algorithms to the current system. With Network Analysis, we can compare the travel times of the ambulances to a given traffic accident.

The authors of this paper [7] introduced an integer linear programming model for the dynamic allocation of emergency services using linear constraints. The primary objective was to prove that their heuristic consistently improves the efficacy of emergency services. This paper is more related to natural disaster events, but it uses a similar methodology in optimizing response to an emergency. In our approach, we use different algorithms but prove that there are methods to enhance emergency services' efficiency.

2.3 Forecasting and Coverage

In their study, the authors [11] examined the importance of incident forecasting to understand the future demand of emergency resources for a given area to proactively allocate resources to the community. Additionally, they used policies to create a general mapping from states of the environment to actions that should be taken to allocate and dispatch resources. In our approach, we follow similar ideologies by optimizing the placement of ambulances in the region with K-Means Clustering. We incorporate different machine learning algorithms, like Nearest Neighbor Analysis.

This study [10] introduced multiple models, such as the Location Set Covering Problem and the Maximal Covering Location Problem. These algorithms focus on selecting facility locations to cover demand points and maximize coverage within a specified time. Our approach uses a similar methodology by computing the demand points and maximizing coverage, but utilizes a variety of techniques to do so. Also, we use a realistic approach of moving ambulances to be in a better position to arrive on the scene quickly.

3 Preliminaries

In order to optimize emergency response strategies, we utilized a combination of spatial analysis techniques and machine learning algorithms. Each algorithm contributes uniquely to understanding the distribution of traffic accidents and improving the placement and routing of ambulances. The following algorithms form the core of our analytical framework:

- **Nearest Network Analysis**: Understand the spatial distribution of traffic accidents.
- **K-Means Clustering**: Identify potential ambulance locations based on traffic accident clusters.
- **Grid-Based Density Analysis**: Visualize traffic accident frequency in a uniform grid format.
- **Kernel Density Estimation**: Identify traffic accident clusters using a smooth version of grid analysis to highlight high-risk areas.

- **Network Analysis with Dijkstra's Algorithm**: Model realistic ambulance travel times. Evaluate the dispatch performance by using a road network by selecting the best candidate locations from K-Means Clustering.

3.1 Nearest Neighbor Analysis

Nearest Neighbor Analysis (NNA) [15] is a spatial statistics technique that analyzes the distribution pattern of point events across a geographic area. This technique helps determine whether accident locations are clustered, randomly distributed, or uniformly distributed. The observed mean distance is denoted as:

$$O = \frac{\sum_{i=1}^{n} d_i}{n}$$

where d_i is the distance from point i to its nearest neighboring accident and n is the total number of accidents.

Additionally, the expected mean distance is calculated on the assumption that the accidents were randomly distributed. The expected mean distance is denoted as:

$$E = 0.5 \times \sqrt{\frac{A}{n}}$$

where A is the total study area (in km^2) and n is the total number of accidents.

Lastly, the Nearest Neighbor Index (NNI) is the ratio between observed and expected mean distance, calculated using the formula:

$$NNI = O/E$$

The resulting NNI value indicates whether accident locations exhibit clustering, randomness, or uniformity in their spatial distribution.

- NNI < 1: Indicates clustering of accidents.
- NNI = 1: Suggests a random spatial distribution.
- NNI > 1: Suggests a uniform distribution.

3.2 K-Means Clustering

K-Means Clustering [14], an unsupervised machine learning algorithm, identifies spatial patterns in traffic accidents. The algorithm begins by selecting k random points as initial cluster centers, which are iteratively optimized to represent the best potential ambulance locations. Each accident is assigned to the closest centroid, forming k groups. The centroid of each cluster is then recalculated as the mean position of all positions within the cluster, and this process repeats until the clusters stabilize. One metric that can be used to evaluate the quality of the clustering is the silhouette score. This metric quantifies how well each point fits within its assigned cluster compared to the other clusters. The score ranges from -1 to 1, where a score close to 1 indicates well-matched clustering for its cluster and poorly matched clustering to neighboring clusters. A score of 0 indicates that the clusters are overlapping. A score closer to -1 suggests potential misclassifications of points to a cluster.

This approach is not nearly as computationally expensive as approaches that utilize geospatial techniques. K-Means Clustering offers a practical balance between scalability and effectiveness, enabling rapid identification of high-priority zones without significant computational overhead.

There are a few limitations to this approach. K-Means Clustering calculates the Euclidean distance between two points, the straight-line distance between points, rather than accounting for the actual road network. Furthermore, K-Means Clustering assumes that the clusters are isotropic and of similar size, which may not hold in the real-world traffic accidents distribution. Therefore, this could lead to skewed centroid placements, especially if the region has complex road layouts.

Algorithm 1 K-Means Clustering

Require: Coordinates C, number of clusters (ambulances) k
Randomly select k coordinates from C as initial centroids
repeat
  for each coordinate c in C **do**
    Compute Euclidean distance from c to each centroid
    Assign c to the nearest centroid
  end for
  for each cluster **do**
    Update centroid: Compute mean of assigned points
  end for
until centroids do not change significantly
return centroids

3.3 Grid-Based Density Analysis

Grid-Based Density Analysis [2] divides a geographic area into equal-sized cells and counts the number of traffic accidents in each cell. Unlike K-Means Clustering, it does not require a predefined number of clusters. Additionally, this approach is better at detecting local anomalies that may not be apparent with K-Means Clustering. Since K-Means is designed to assign every point to a cluster, it can obscure outliers and localized variations. The choice of grid cell size significantly affects the resolution of the analysis: smaller cells yield finer granularity, but may introduce noise, while larger cells smooth the data but may obscure localized hotspots.

Algorithm 2 Grid-Based Density Analysis

Require: Coordinates C, grid cell size s, density threshold t
Divide spatial region into a uniform grid of cells of size s
Initialize an empty map `GridDensity`
for each coordinate c in C **do**
  Determine the grid cell g containing c
  Increment `GridDensity[`g`]` by 1
end for
for each cell g in `GridDensity` **do**
  if `GridDensity[`g`]` $\geq t$ **then**
    Mark cell g as a dense region
  else
    Mark cell g as sparse or discard
  end if
end for
*Optional: merge adjacent dense cells into clusters
return List of dense cells or merged regions

3.4 Kernel Density Estimation

Kernel Density Estimation (KDE) [3] is primarily utilized for identifying where accidents are most concentrated without artificial

cluster boundaries. In contrast to K-Means Clustering, it does not require assumptions about fixed groupings and equal-sized regions. While similar to Grid-Based Density Analysis, the output is continuous rather than discrete. It returns a smooth kernel function over each point and these values are aggregated to generate a smooth surface to highlight the densely packed regions. KDE places a smooth kernel, typically Gaussian, over each data point and sums the overlapping values to generate a continuous surface. This algorithm is effective at revealing spatial patterns in dense urban environments. KDE's output is a valuable visualization to understand relative clusters throughout the region.

Algorithm 3 Kernel Density Estimation

Require: Coordinates C, bandwidth parameter h
1: **for** each coordinate c in C **do**
2: Compute the kernel function $K\left(\frac{c-c_i}{h}\right)$
3: Sum the kernel values to get $\hat{f}(c)$:
4: $\hat{f}(c) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{c-c_i}{h}\right)$
5: **end for**
6: **return** $\hat{f}(c)$ for all coordinates c

3.5 Network Analysis with Dijkstra's Algorithm

While the previous algorithms are useful for understanding the clustering of traffic accidents, this algorithm realistically models ambulance travel from Point A to Point B.

Network Analysis [9] provides a more realistic model of urban movement by incorporating road connectivity, directionality, and estimated travel times. In this model, the road network is represented as a directed, weighted graph, where intersections are nodes and roads are edges, each weighted by estimated travel time. In order to find the most efficient ambulance routes, a shortest path algorithm is used, Dijkstra's Algorithm. Dijkstra's Algorithm finds the shortest path from a starting node to all other nodes in a weighted graph by iteratively selecting the node with the lowest known cost and updating its neighboring nodes. This allows for rapid and realistic estimation of emergency response times.

Algorithm 4 Network Analysis with Dijkstra's

Require: Dataset D, Ambulance Placements A
1: Load Road Network
2: **for** each edge in the network **do**
3: Add travel time as edge weight based on road length and speed limit
4: **end for**
5: Select nearest ambulance in A for each accident location
6: **for** each ambulance node **do**
7: Calculate travel times to all accident nodes using Dijkstra's
8: Store ambulance that is closest
9: Store estimated time to arrive on scene
10: **end for**

4 Method

Our approach aimed to optimize emergency response times in Washington, D.C., using geospatial analysis and machine learning models. After preprocessing the data, we utilized traffic accident data to visualize high-risk regions for traffic accidents and optimize the placement of ambulances. Then, we computed the travel time for ambulances to arrive at the accident scenes. Afterward, we compared metrics of the response time for the ambulances between the current setup and our proposed design. Additionally, we provided statistics about ambulance coverage within specific response time thresholds.

Washington, D.C. spans approximately 176.99 square kilometers and operates 43 public ambulances at any given time. These ambulances are stationed at local fire stations. In addition, 25 private ambulances provided by AMR are available between 7 AM and 1 AM and are dynamically deployed across the city based on real-time demand and specific emergencies. However, this analysis excludes private ambulances, which are not consistently dispatched to traffic-related incidents. While incorporating private ambulances would reduce average response times, they are typically reserved for a broader range of emergency medical calls beyond traffic accidents.

4.1 Dataset Overview

Two datasets were used for our analysis. First, the traffic accident data in Washington, D.C. was provided by Open Data D.C. [5] and includes historical records of traffic accidents and emergencies. This dataset contains information for each traffic accident from 1900 to February 2025. There are a total of 323,555 reported accidents. However, only 358 records fall between 1900 and 2007, making them insufficient for analysis. As a result, we limited our analysis to the 323,197 accidents recorded from 2008 to 2025.

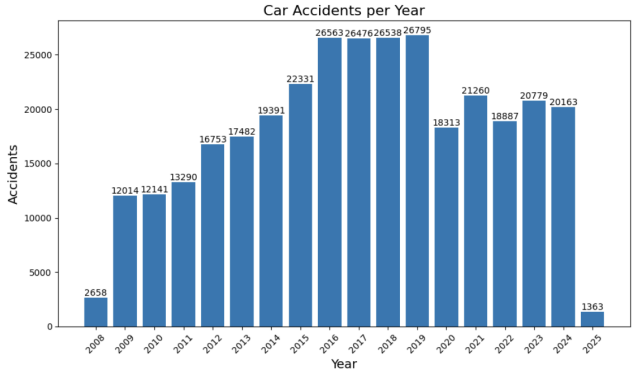

Figure 1: Bar chart of car accident counts per year in Washington, D.C. from January 2008 to February 2025.

This dataset contains a total of 66 attributes. The most prominent ones include the report date, latitude, longitude, address, and several fields related to injuries and fatalities. We cleaned the data by removing incomplete records. Then, we normalized the timestamp, making it easier to find records using the time and date. This preprocessing step enables more effective temporal analysis and supports the extraction of features, such as time of day, day of week, and seasonality, which may correlate with accident frequency and severity.

To ensure relevance and computational feasibility, we focused on recent data subsets for specific analyses: Nearest Neighbor Analysis

used data from 2022 to 2024, while clustering methods were applied to the period from 2020 to 2024.

Additionally, we utilized an Open Data D.C. dataset [6] that details information about the locations of fire stations, which are where the ambulances are located. Here is the current ambulance setup across Washington, D.C.:

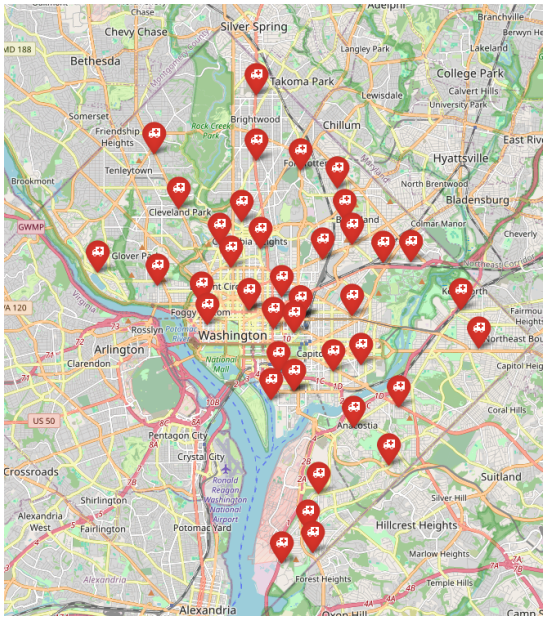

Figure 2: Ambulance placements of the current configuration based on the fire station locations.

4.2 Approach

After preprocessing, we used Nearest Neighbor Analysis to determine the spatial distribution of traffic accident locations. This method identified high-risk clusters where incidents occur frequently. With this information, we had a better understanding of whether the accidents are typically clustered, uniformly distributed, or randomly distributed. In order to calculate the sum of distances for the observed mean distance, we computed the pairwise distance matrix using the Haversine distance, the great-circle distance between two geographic coordinates on Earth's surface. This ensures accuracy when analyzing geographic coordinates.

Next, we utilized Grid-Based Density Analysis and Kernel Density Estimation to visualize accident hotspots and better understand their spatial distribution and identify areas that have high quantities of accidents. Grid-Based Density Analysis partitioned the data into uniform grid cells, facilitating the identification of localized patterns and variations that would have been overlooked in the aggregated data. The study area was divided into a grid of square cells approximately 1.02 km in size. Each cell was assigned a density value representing the number of accidents that occurred within its boundaries. KDE was beneficial for identifying high-level trends by smoothing the data and producing a continuous surface of accident density. Grid-Based Density Analysis emphasized discrete differences, while KDE highlighted broader zones of concentrated risk.

After, we utilized K-Means Clustering to determine the placement of our ambulances within Washington, D.C. Due to the large region size of Washington, D.C. and the high quantity of traffic accidents, K-Means Clustering offered a practical balance between scalability and effectiveness, enabling rapid identification of high-priority zones without significant computational overhead. Using these findings, we proposed improved ambulance locations based on proximity to accident clusters.

Finally, we constructed a road network graph of Washington, D.C., where nodes represent intersections and edges represent road segments. Edge weights are calculated based on distance and estimated travel time. We utilize Network Analysis with Dijkstra's algorithm to simulate ambulance routing, prioritizing time efficiency over geometric distance.

4.3 Baseline

We evaluated our approach by comparing the travel times from existing ambulance stations with those from our proposed placements based on Network Analysis. For each historical accident, we computed the shortest estimated response time from both current and optimized ambulance locations using Dijkstra's algorithm on the road network graph.

This comparison enabled us to assess whether the proposed placements lead to statistically significant reductions in response time and improvement in coverage within time thresholds. By aligning ambulance coverage with accident clusters, we can enhance emergency service accessibility and improve patient outcomes across Washington, D.C. We reported statistical results to analyze the difference in average, median, and maximum response times. Additionally, we calculated the percentage of accidents reached within specific time intervals: 1, 2, 3, 4, 5, and 10 minutes.

5 Results

5.1 Preliminary Nearest Neighbor Analysis

We conducted Nearest Neighbor Analysis (NNA) to assess the spatial distribution of traffic accidents across Washington, D.C. from 2022 to 2024. The goal was to determine whether accident locations exhibit a random, dispersed, or clustered pattern.

For each quarter, we calculated the Observed Mean Distance between accident points, the Expected Mean Distance assuming a random distribution, and the Nearest Neighbor Index (NNI), the ratio between the two.

Table 1 shows the quarterly results. Across all 12 periods, the Nearest Neighbor Index remained below 1.0, consistently falling between the range of 0.49 and 0.52 across all 3-month spans. These low values suggested that these traffic accidents exhibit significant spatial clustering throughout the city and are not randomly distributed.

The least clustering occurred in 2024 Q2 with an NNI score of 0.4942. This period also had one of the highest accident counts, indicating a potential link between volume and spatial concentration. In contrast, the highest NNI was observed in 2024 Q1 with an NNI score of 0.5173. This suggested relatively less dense clustering during that period. Both 2024 Q1 and Q4 showed slightly more

Quarter	# Accidents	Obs Dist	Exp Dist	NNI
2022 Q1	4476	0.0511	0.0994	0.5138
2022 Q2	5025	0.0471	0.0938	0.5015
2022 Q3	4717	0.0494	0.0969	0.5103
2022 Q4	4669	0.0491	0.0974	0.5048
2023 Q1	4877	0.0482	0.0953	0.5058
2023 Q2	5313	0.0461	0.0913	0.5054
2023 Q3	5531	0.0448	0.0894	0.5009
2023 Q4	5058	0.0468	0.0935	0.5006
2024 Q1	4751	0.0499	0.0965	0.5173
2024 Q2	5454	0.0445	0.0901	0.4942
2024 Q3	5083	0.0470	0.0933	0.504
2024 Q4	4875	0.0490	0.0953	0.5142

Table 1: Quarterly Nearest Neighbor Index (NNI) results for traffic accidents in Washington, D.C. from 2022 to 2024, including the number of reported accidents.

spatial dispersion compared to other quarters, while mid-year quarters generally had denser clustering. The patterns from 2022 and 2023 were relatively stable, with NNIs consistently hovering around 0.50–0.51. This consistency suggested the presence of persistent high-risk zones that may benefit from targeted resource allocation.

Overall, segmenting the data into quarters allowed us to detect seasonal fluctuations in clustering intensity and supported the development of more dynamic and responsive strategies for optimizing emergency service placement.

5.2 Visualizations of the Traffic Accidents

Our next step involved applying density estimation techniques to reveal underlying spatial patterns in the accident data, utilizing Grid-Based Density Analysis and Kernel Density Estimation. We overlayed the results from K-Means Clustering, which is discussed later, in the figures in this section to provide a comparative view of how different methods capture spatial concentration and identify hotspots. This comparison was useful in evaluating whether the ambulance placements suggested by K-Means Clustering were consistent with high-density accident areas identified by the other methods, assessing the validity of K-Means-based deployment strategies.

5.2.1 Grid-Based Density Analysis. We investigated the traffic accidents by implementing a grid-based approach, overlaying a uniform spatial grid over Washington, D.C. For each cell, we counted the number of accidents occurring within that region. This approach allowed us to detect micro-clusters and high-density zones. This was useful later for identifying if there were high-risk areas that did not conform to the centroid-based clustering in K-Means Clustering. Additionally, we created a visualization in the form of a heatmap, labeled Figure 3.

Figure 4 revealed a higher concentration of accidents in downtown Washington, D.C. This pattern was reflected in the corresponding ambulance placements, which were more densely distributed in that area to account for the elevated risk. This further illustrated the clustering of accidents and the necessity of strategically placing ambulances to address demand.

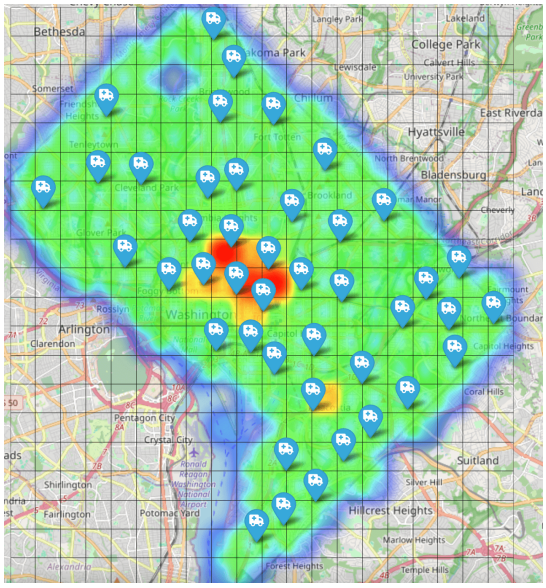

Figure 3: Grid-Based Density Analysis with ambulances from K-Means Clustering.

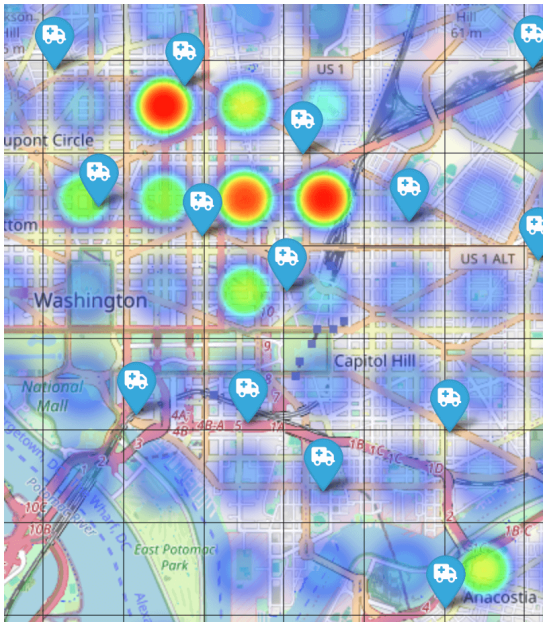

Figure 4: Zoomed-in Grid-Based Density Analysis from major high-risk area in downtown Washington, D.C. with ambulances from K-Means Clustering.

Across all cells, a total of 99,392 accidents were recorded. The maximum cell density reached 2,375 accidents, indicating areas with recurrent accidents. The average cell density was 248.48 accidents during the 2020 to 2024 period. This analysis highlighted specific urban zones where accident rates were consistently high, which can serve as key candidates for increased emergency coverage.

5.2.2 Kernel Density Estimation. We applied Kernel Density Estimation to visualize the continuous spatial distribution of traffic accidents across Washington, D.C. to complement the clustering and grid-based approaches. KDE produces a smoothed surface that highlights regions with higher accident concentrations without being constrained by predefined grid boundaries or cluster assignments.

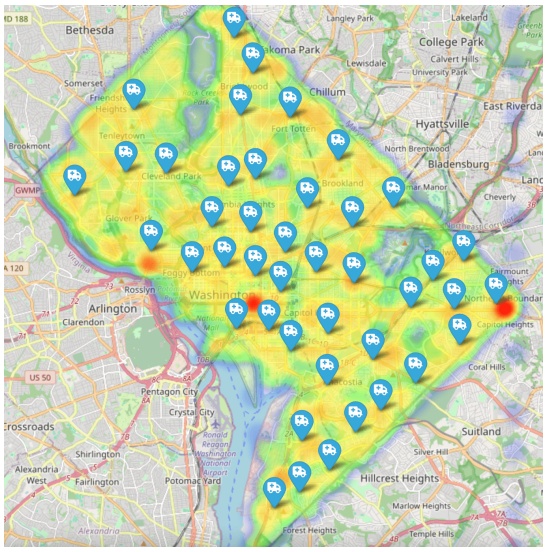

Figure 5: Kernel Density Estimation with ambulances from K-Means Clustering

The KDE heatmap revealed clear density peaks in the downtown core and surrounding high-traffic areas, aligning with findings from both K-Means and grid-based density analysis. Additionally, this approach highlighted high-risk areas that were not apparent with Grid-Based Density Analysis. In the eastern corner of Washington, D.C., we observed a high concentration of traffic accidents. Its high density made it a candidate for prioritized ambulance placement using K-Means Clustering. This method was particularly effective in identifying subtle variations in accident density that may not have been apparent through discrete clustering techniques.

KDE yielded some interesting metrics. The model had a mean density of 0.001685 and a standard deviation of 0.02109. These metrics indicate a highly skewed distribution, where a few areas exhibit extremely high accident density while the majority of the city has comparatively low values. This reinforces the notion of spatial clustering and further justifies the need for targeted emergency response strategies in high-risk zones.

5.3 Ambulance Placement using K-Means Clustering

We used K-Means Clustering to identify high-risk areas where traffic accidents tended to concentrate. This unsupervised machine learning algorithm was used to divide the accident data from 2020 to 2024 into 43 regions, matching the number of public ambulances currently deployed in Washington, D.C. The clustering was based solely on accident locations, using the latitude and longitude coordinates, to approximate optimal ambulance coverage zones.

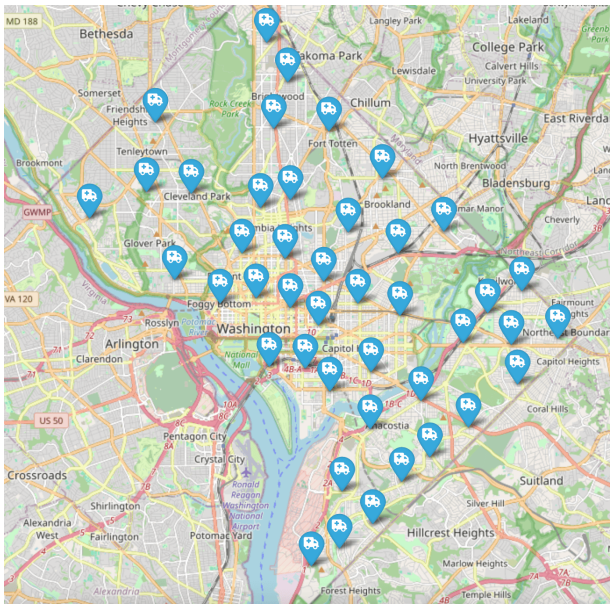

Figure 6: Ambulance placements based on results from K-Means Clustering

The clustering algorithm returned an average distance of 0.01 km from any accident point to its nearest ambulance location, with a maximum distance of 0.03 km. This suggested a good spatial fit between cluster centroids and accident locations. The silhouette score of 0.399 indicates a moderate level of separation between clusters, reflecting meaningful spatial divisions while acknowledging some overlap near region boundaries.

The number of accidents per region varied significantly. Regions 11, 12, and 28 handled over 4,000 accidents each. Other regions had fewer than 1,000 accidents, such as Regions 20, 22, and 34, highlighting imbalances in accident volume that could inform dynamic resource reallocation strategies.

This approach provided a data-driven foundation for assessing whether current ambulance deployment aligns with historical accident concentrations and where adjustments could reduce emergency response times. This approach determined our ambulance placements used in the subsequent Network Analysis.

5.4 Network Analysis

Lastly, we employed a network-based approach to better align ambulance deployment with real-world travel constraints. This method modeled the actual road network of Washington, D.C. and incorporated travel time as the primary metric for evaluating emergency response efficiency.

We began by projecting all accident locations onto the nearest road network nodes. To account for travel-based accessibility rather than simple Euclidean distance, we calculated shortest travel-time paths across the road graph using Dijkstra's algorithm. Using our results from K-Means Clustering, we initialized the placement of 43

ambulances for the proposed ambulance locations. From the second dataset, we initialized the placement of the ambulances from the baseline setup.

We compared our proposed ambulance locations against the baseline:

Metric	Baseline	Optimized
Average response time (minutes)	1.45	**1.08**
Median response time (minutes)	1.36	**1.18**
Maximum response time (minutes)	**5.86**	6.05
Coverage within 1 minute (%)	31.4	**43.1**
Coverage within 2 minutes (%)	79.1	**90.7**
Coverage within 3 minutes (%)	94.5	**98.0**
Coverage within 4 minutes (%)	**99.5**	99.2
Coverage within 5 minutes (%)	99.9	99.9
Coverage within 10 minutes (%)	100.0	100.0

Table 2: Emergency response time statistics, comparing baseline and optimized ambulance placements

The results from Network Analysis, provided in Table 2, indicated that the proposed network-optimized placement significantly improves early-stage response times. Ambulances are able to arrive at the scene within 1 minute for approximately 43.1% of accidents, compared to 31.4% under the current setup. Within 2 minutes, the optimized configuration also achieves over 90% coverage, over 10% better than the current setup. This highlighted its superior ability to deliver rapid response in high-density accident zones. For coverage within 3, 4, 5, and 10 minutes, the results are fairly comparable. For a given traffic accident, the ambulance response time for the optimized approach was faster by about 22.2 seconds. Furthermore, the median response time was better, faster than the baseline by 10.8 seconds.

However, the current setup is better than the proposed plan in terms of maximum response time. It would take about 6.05 minutes for the optimized setup, 11.4 seconds lower compared to the current deployment. This suggests that while our optimization improves average and early-stage responsiveness, it may leave some less-accessible regions with slightly longer wait times.

Based on the results, there are some trade-offs between the two approaches. While the optimized approach achieves faster response times for the vast majority of accidents and increases coverage in critical high-density urban cores, the current deployment is more effective for arriving at more isolated areas faster.

Overall, Network Analysis highlighted the contrast in the two possible configurations. For the proposed setup, we determined that the ambulances would arrive at most accidents rapidly, but accidents that are far from the center of the cluster centroids would take longer for the ambulances to get to. For the current setup, it is slower on average, but gets to edge cases slightly faster. There are trade-offs between the two approaches, both efficient in their own way.

6 Discussion

The observations from this study contribute to the broader discussion of optimizing emergency response times through data-driven spatial analysis.

First, we identified that the traffic accidents showed signs of being clustered. Then, we identified traffic accident hotspots and clusters of higher risk areas. Grid-Based Density Analysis and Kernel Density Estimation were utilized to ensure that high-risk areas were not overlooked. As a result, we applied machine learning techniques such as K-Means Clustering to offer a data-driven method for optimizing ambulance placement to reduce emergency response times in high-risk areas. Furthermore, Network Analysis with Dijkstra's introduced a dynamic component to ambulance deployment by incorporating road network constraints and travel time estimates. These results aligned with existing literature that emphasized the importance of utilizing data for crucial decisions in emergency services. This is particularly important in urban areas where traffic congestion and population density present logistical challenges. The discrepancy between current ambulance locations and our optimized placements suggests room for resource allocation improvement, potentially a hybrid approach. Overall, this study highlights the potential of GIS-based analytics and machine learning techniques in optimizing public safety services, demonstrating an approach that can be adapted and expanded to other cities to improve emergency response times.

7 Future Work

Future work on this topic can address several limitations identified in this study. One major constraint is the limited computational power available and performing large-scale Network Analysis due to the extensive accident data and complex routing graphs. With greater computing resources, simulations could be conducted to optimize ambulance placement.

Additionally, future work could expand the scope beyond traffic accidents. Emergency medical services respond to a variety of incidents and access to a more informed dataset could enable a general-purpose optimization. However, broader implementations present challenges, especially in situations where ambulances cannot provide aid until police secure the scene.

Another direction for future work is incorporating predicative modeling to deal with real-time traffic, weather, and other conditions. This would enable more dynamic dispatching and routing decisions that adapt to the current conditions. This would also facilitate situations where an ambulance is occupied with another emergency and the remaining ambulances must temporarily cover more area.

8 Conclusion

This paper presents a data-driven framework for optimizing emergency response times for traffic accidents in Washington, D.C. By combining geospatial analysis with machine learning techniques, we developed models to identify accident hotspots and optimize ambulance placement and routing. Our use of Nearest Neighbor Analysis, Grid-Based Density Analysis, Kernel Density Estimation,

K-Means Clustering, and Network Analysis with Dijkstra's Algorithm enabled a comprehensive understanding of spatial and temporal trends in emergency incidents. Nearest Neighbor Analysis provided information that the traffic accidents were more clustered than randomly or uniformly distributed. Grid-Based Density Analysis and Kernel Density Estimation provided insight into specific regions that had higher concentrations of traffic accidents. We were able to optimize the ambulance placements with K-Means Clustering. Lastly, Network Analysis provided a road network graph to calculate the estimated time for ambulances to get to accident scenes, allowing us to compare the baseline and proposed approaches.

The quantitative evaluation from Network Analysis showed that our optimized deployment significantly reduced average and median response times, and increased early-stage coverage, particularly within high-density urban zones. However, the maximum response time slightly increased, suggesting that while our approach improves overall efficiency, it may require refinement to address accessibility in low-density or isolated regions. A potential solution is to utilize a hybrid strategy, using data-driven optimization as a baseline while selectively positioning a few ambulances to cover outlier cases.

This integrated approach provides a framework that can be adapted to other urban areas facing similar challenges. Ultimately, our study highlights the value of combining geospatial analysis with machine learning to strengthen public safety infrastructure and emergency response resource allocation and strategy.

References

[1] Justin J. Boutilier and Timothy C. Y. Chan. 2019. Ambulance Emergency Response Optimization in Developing Countries. arXiv:1801.05402 [math.OC] https://arxiv.org/abs/1801.05402

[2] Daniel Brown, Arialdis Japa, and Yong Shi. 2019. A Fast Density-Grid Based Clustering Method. In *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*. 0048–0054. https://doi.org/10.1109/CCWC.2019.8666548

[3] Yen-Chi Chen. 2017. A Tutorial on Kernel Density Estimation and Recent Advances. *Biostatistics & Epidemiology* 1, 1 (2017), 161–187. https://doi.org/10.1080/24709360.2017.1396742

[4] Dajun Dai. 2012. Identifying clusters and risk factors of injuries in pedestrian-vehicle crashes in a GIS environment. *Journal of Transport Geography* 24 (2012), 206–214. https://doi.org/10.1016/j.jtrangeo.2012.02.005

[5] Open Data DC. 2025. Crashes in DC. https://opendata.dc.gov/datasets/crashes-in-dc/about Accessed: 2025-02-04.

[6] Open Data DC. 2025. Fire Stations. https://opendata.dc.gov/datasets/DCGIS::fire-stations/about Accessed: 2025-02-04.

[7] Paraskevi S. Georgiadou, Ioannis A. Papazoglou, Chris T. Kiranoudis, and Nikolaos C. Markatos. 2010. Multi-objective evolutionary emergency response optimization for major accidents. *Journal of Hazardous Materials* 178, 1-3 (2010), 792–803. https://doi.org/10.1016/j.jhazmat.2010.02.010

[8] Insurance Institute for Highway Safety (IIHS). 2024. *Yearly Snapshot - Fatality Statistics*. https://www.iihs.org/topics/fatality-statistics/detail/yearly-snapshot Accessed: 2025-04-16.

[9] Ismail Rakip Karas and Sait Demir. 2011. Dijkstra algorithm interactive training software development for network analysis applications in GIS. *Energy Education Science and Technology Part A: Energy Science and Research* 28, 1 (2011), 445–452.

[10] X. Li, Z. Zhao, X. Zhu, et al. 2011. Covering models and optimization techniques for emergency response facility location and planning: a review. *Mathematical Methods of Operations Research* 74, 3 (2011), 281–310. https://doi.org/10.1007/s00186-011-0363-4

[11] Geoffrey Pettet, Hunter Baxter, Sayyed Mohsen Vazirizade, Hemant Purohit, Meiyi Ma, Ayan Mukhopadhyay, and Abhishek Dubey. 2022. Designing Decision Support Systems for Emergency Response: Challenges and Opportunities. *CPS-ER*.

[12] Mohammad Javad Abdolhosseini Qomi, Mahdi Zare, and Mohammad Reza Delavar. 2020. Optimizing Ambulance Locations for Coverage Enhancement of Accident Sites in South Delhi. *Transportation Research Procedia* 48 (2020), 280–289. https://doi.org/10.1016/j.dib.2020.105438

[13] Hamid R. Sayarshad. 2022. Designing an intelligent emergency response system to minimize the impacts of traffic incidents: a new approximation queuing model. *International Journal of Urban Sciences* 26 (2022), 691–709. https://doi.org/10.1080/12265934.2022.2044890

[14] Kristina P. Sinaga and Miin-Shen Yang. 2020. Unsupervised K-Means Clustering Algorithm. *IEEE Access* 8 (2020), 80716–80727. https://doi.org/10.1109/ACCESS.2020.2988796

[15] Sulaiman Yunus and Ishaq Aliyu Abdulkarim. 2022. Road Traffic Crashes and Emergency Response Optimization: A Geo-Spatial Analysis Using Closest Facility and Location-Allocation Methods. *Geomatics, Natural Hazards and Risk* 13, 1 (2022), 1535–1555. https://doi.org/10.1080/19475705.2022.2086829