Graph Models to Forecast Road Traffic Crashes and Resource Optimization on Brazilian Federal Roads

Júlio C. F. Taveira
University of Pernambuco
Recife, Brazil
University of Highway Police
Florianópolis, Brazil
0000-0002-1355-7380

Hugo A. Amorim Neto
University of Pernambuco
Recife, Brazil
0000-0003-1669-8450

Andina A. Lerma
University of Pernambuco
Recife, Brazil
0000-0001-7904-7773

Fernando B. de Lima Neto
University of Pernambuco
Recife, Brazil
0000-0003-1200-225X

***Abstract*—Road traffic crashes present a critical public health challenge, requiring integrated predictive and operational strategies. This paper proposes an end-to-end framework for Brazilian federal highways that bridges crash risk forecasting with proactive resource allocation. By modeling the highway infrastructure as a complex undirected graph, we evaluate three GNNs architectures (GCN, GAT, and EvolveGCN) for road crash prediction. These predictions inform a multi-objective optimization layer using NSGA-II, MOPSO, and P-ACO to maximize risk coverage while minimizing response times and costs. Results show that GCN and EvolveGCN yield stable predictive performance, while optimization outcomes vary with regional network topology. Notably, the framework yields more substantial improvements in decentralized units, suggesting that automated spatial analysis can provide valuable insights into resource positioning in areas where road connectivity is more dispersed. This work validates the feasibility of a unified, data-driven pipeline that transforms statistical risk into actionable patrolling strategies, providing a decision-support tool for road safety management.**

***Index Terms*—Graph Neural Networks, Prediction, Discrete Optimization, Road Traffic Crash, Decision-Making**

I. Introduction

Several studies have investigated the use of graph neural networks (GNNs) to extract structural relationships from data and to represent/encode entities and interactions. This approach has supported research across multiple domains, including social networks [1], biological systems [2], crime analysis [3], transportation systems [4], among others.

In the context of road traffic safety, crash analysis and prediction initially relied on mathematical and statistical approaches [5]. Subsequently, modelers incorporated data-driven predictive models [6] and deep learning approaches [7]. More recently, includes works using graph neural networks to model and evaluate this social problem [8].

Road traffic crashes constitute a major public health and socioeconomic challenge worldwide. The World Health Organization (WHO) estimates that road traffic injuries impose substantial economic losses on societies, corresponding to approximately 1% to 3% of Gross Domestic Product (GDP) and, in some contexts, reaching up to 6% [9]. Beyond the economic burden, road traffic crashes disproportionately affect younger populations, they represent a leading cause of mortality among children and adolescents, resulting in significant health, social, and long-term economic impacts across communities [9].

In Brazil, this burden is particularly pronounced. The World Bank (WB) estimates that the country loses approximately US$61.3 billion annually due to traffic crashes [10]. The same report suggests that, if Brazil meets its 2020-2030 road safety target for reducing fatalities, approximately 85,000 deaths could be prevented [10].

Recent works aim to study Brazilian road traffic crashes and to predict them. Prior work includes a classification model to analyze the behavior of Brazilian federal roads [11], applied hierarchical temporal series to improve predictability as a Key Performance Indicator (KPI) [12], and hierarchical temporal series with strategic, tactical, and operational context [13]. Moura et al. [14] proposed a critical point indicator for fatal accidents, highlighting the importance of segregating and prioritizing specific segments in planning and control, with validation in Brazil and other countries. Wanker et al. [15] used a complex system approach to examine how external disruptions temporarily affect network performance, finding that most impacts are short-lived and networks tend to return to baseline conditions.

Mukhopadhyay et al. [16] presented a survey of studies on incident prediction, incident detection, resource allocation, and computer-aided dispatch for emergency response. Nippani et al. [8] evaluated existing deep learning methods for predicting crash occurrences on road networks, finding that GNNs achieved better results than other approaches. Zou et al. [17] proposed the multi-graph spatio-temporal network (MG-SNET) to accurately predict traffic accident risk, capturing semantic spatial dependencies across all regions and geographical correlations among adjacent regions. The authors also highlights three main challenges in traffic crash prediction: difficulty in capturing complex spatial correlations, neglect of differences between daily and weekly temporal patterns, and equal treatment of heterogeneous semantic and geographical spatio-temporal features [17].

This paper proposes a road traffic crash prediction model and resource allocation optimization for Brazilian federal roads, using a graph model from a network science perspective

The Brazilian science agencies CNPq, CAPES (Finance Code 001) and FACEPE funded this work.

[18]. The network use road segments as edges and these extremities as nodes. Our approach builds a time series of traffic crash counts for each segment (edge) and other features. For the prediction, we used three GNN architectures and compared them with isolated prediction approaches. The prediction generates a daily journal with incident predictions. We also proposed a resource allocation optimization to address these incidents based on the road network structure, using a three-discrete optimization approach. We compare the GNN model predictions and the optimization algorithms.

The main contribution of this work is the integration of network science for modeling, graph neural networks for forecasting, and discrete optimization to support decision-making, applied to a real-world scenario. Based on this approach, police teams may be better positioned either to prevent incidents through proactive patrolling or to respond more efficiently by being located closer to where crashes are likely to occur. This work does not analyze crash properties in depth, nor does it address human behavioral factors associated.

The code, traffic crash data, and network data are available upon request to the authors. However, the police road resources and strategy data are sensitive information, does not provided.

II. Background

In this section, we provide a brief background on GNNs, their architecture, and applications. At the same time, we discuss discrete optimization with GNN applications.

A. Graph Neural Networks

Proposed by Scarselli et al. [19], a Graph Neural Network (GNN) is a model that extends existing neural network methods for processing data in graph domains. This model implements a function $\tau(G, n)$ that maps a graph G and one of its nodes n into an m-dimensional Euclidean space. Through message-passing mechanisms, GNNs iteratively aggregate information from neighboring nodes, enabling the learning of representations that jointly encode node attributes and the relational structure.

There are several GNN architectures for solving different problems in large-scale structures. Wu et al. [20] presented a comprehensive survey proposing a new taxonomy grouping the GNN models in four large groups: recurrent GNNs (RecGNNs), convolutional GNNs (ConvGNNs), graph autoencoders (GAEs), and spatial–temporal GNNs (STGNNs). Other studies address more specific GNN architectures, such as dynamics networks [21] and networks for time series [22].

Among GNN variants, spatial GNNs model local dependencies via neighborhood connections, making them more adaptable to real-world graphs with irregular structures, such as transportation and road networks [20]. ConvGNNs define a graph convolutional operation based on nodes' spatial relationships and generalize it to large graph structures. ConvGNNs use space-based approaches, widely adopted for their scalability and effectiveness in capturing spatial interactions, such as graph convolutional networks (GCN) and graph attention networks (GAT) [20].

Other recent advances include GNNs for time series analysis, which model multivariate temporal data as spatial-temporal graphs, where each node is associated with a time series, and edges encode spatial or functional dependencies [22]. This formulation enables the joint modeling of inter-variable and inter-temporal relationships, which has proven effective in domains such as traffic and urban systems.

B. Discrete Optimization and GNNs

In the field of computer science, optimization is a search technique for finding the best solution from all feasible solutions [23]. Discrete optimization is one kind of optimization technique where we will wait for an object like an integer, permutation, or graph from a countable set [23].

Cappart et al. [24] presented a conceptual review for GNNs in combinatorial optimization. This study shows that classical optimization approaches often solve problems in isolation, ignoring real-world patterns. When using graph-structured representations, GNNs can learn connections among elements through neighborhood aggregation, thereby learning a structure representation and generalizing.

III. Methodology

In this section, we present the road structure used, the data used, the GNN variation and its representation and the optimization approaches. Figure 1 presents our proposed pipeline for tracking the problem.

A. Network Structure

Brazil is a country with a continental area (approximately 8.5 mi km^2), and its main transport mode is road-based [25]. The Brazilian federal road infrastructure comprises approximately 75,000 kilometers across all territory. The road structure was obtained from the National Road System (*Sistema Nacional Viário*, SNV) [26], which provides georeferenced data. From this dataset, we extracted information such as latitude, longitude, road pavement type, and others, collected in 2023.

In this work, we used the road network, representing it as an undirected graph $G(V, E)$, where $V = \{v_1, v_2, \dots, v_n\}$ is the set of nodes and $E = \{e_1, e_2, \dots, e_m\}$ is the set of edges linking two nodes. The nodes correspond to road intersections, junctions, state borders, significant kilometer markers, and other relevant points, while edges represent road segments connecting two nodes [18]. Figure 2 (A) presents the all Brazilian federal road network and a detailed network in a specific unit area is presented in Figure 1 (B).

B. Data

The traffic crash data were obtained from the Brazilian Federal Highway Police (*Polícia Rodoviária Federal*) [27], and are available as open data. We selected traffic crashes registered between 2022 and 2024, totaling more than 200.000 crash registers.

For each segment (edge) presented in Section III-A, we selected the crashes that occurred in that segment, creating

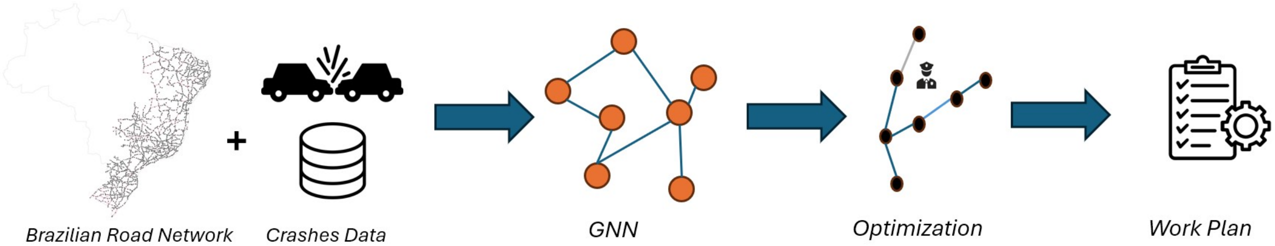

Fig. 1. Proposed pipeline from network modeling and GNN-based prediction to optimization and operational planning.

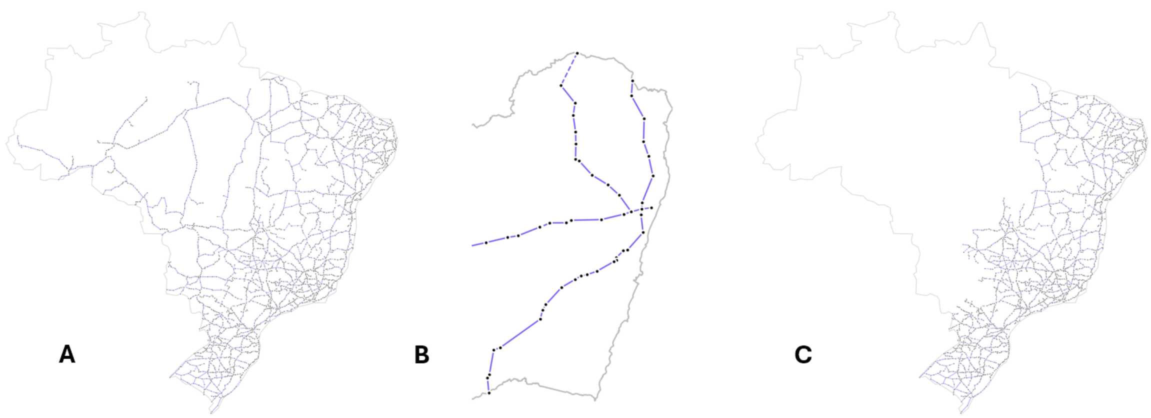

Fig. 2. Brazilian segmented network. (A) Full national network, (B) example of a specific PRF unit area, and (C) network selected for this study.

a time series of the number of crashes per day. Because road crashes are rare events [28], we used a fixed-window cumulative aggregation, where values are accumulated daily and reset at the beginning of each month [29].

C. GNNs

The structure of the data is crucial to model and select the GNN. Following the paradigm in GNNs for time series, as systematized by Jin et al. [30], we applied the system as a spatiotemporal graph in which each node contains an associated time series. We adopted spatial GNN backbones GCN and GAT to model inter-variable dependencies, combined with temporal aggregation to handle sparse and noisy event-based data. Also, we adopted EvolveGCN [31] because is Spatial GNN with temporal module (recurrence-based).

Each node in the road network graph is represented by a fixed set of static and temporal attributes describing its spatial, semantic, and temporal properties. Features are assigned to all nodes, including both city nodes and road junctions. For each node v_i, the static feature vector is defined in Equation 1.

$$\mathbf{x}_i^{\text{static}} = \left[\text{latitude}_i,\, \text{longitude}_i,\, \text{pop}_i,\, \text{is\_city}_i\right], \quad (1)$$

where latitude_i and longitude_i denote the geographical coordinates of node i. The field pop_i represents the city population, associated with the node $\text{is\_city}_i \in \{0, 1\}$, which is an indicator distinguishing a node closely from the city's town. This representation preserves the complete topological structure of the road network.

In addition to node-specific attributes, we compute a set of global temporal features that summarize the recent accident dynamics of the entire network within each training window. These features provide a shared temporal context that captures network-wide trends and short-term variability. For each temporal variable aggregated at the network level, we extract descriptive statistics over a fixed-length window. The resulting global temporal feature vector is defined in Equation 2.

$$\mathbf{x}^{\text{temp}} = \left[\mu_t,\, \sigma_t,\, x_t^{\text{last}},\, x_t^{\max},\, x_t^{\min}\right], \quad (2)$$

where μ_t and σ_t correspond to the mean and standard deviation of accident counts across all edges within the temporal window, while x_t^{last}, $x_t^{\max}$, and $x_t^{\min}$ denote the most recent, maximum, and minimum observed network-wide accident values, respectively.

The final node representation is obtained by concatenating static node-level attributes with the global temporal context, such that

$$\mathbf{x}_i = \left[\mathbf{x}_i^{\text{static}},\, \mathbf{x}^{\text{temp}}\right]. \quad (3)$$

As a result, all nodes share the same temporal information while preserving distinct spatial and semantic characteristics. Edges are used exclusively to encode the road network topology and to enable message passing between connected nodes. No temporal attributes are stored at the edge level, and the road network is modeled as an undirected graph.

D. Optimization

The optimization stage uses road traffic crash predictions to propose a proactive resource allocation plan. Given the road network graph and the predicted crash likelihood for each segment, a discrete multi-objective problem is formulated to select a set of segments to be visited by police unit, subject to operational constraints such as maximum duty time and travel time between segments. The objectives simultaneously maximize accumulated risk coverage (the sum of predicted probabilities over visited segments) while minimizing route duration and related costs (e.g., traveled distance), yielding a Pareto set of non-dominated routes that represents different trade-offs between response speed and proximity to high-risk locations.

As shown in Figure 2 (B), our objective is to plan the Brazilian Raod Police (PRF) units to move between the road segments (edges), taking into consideration the average velocity speed at the specific time and the distance reference for each segment. We established which segment the PRF units should initially be located based on the general strategy of policing for the force.

We used three algorithms (NSGA-II [32], MOPSO [33], and P-ACO [34]) and evaluated them using the same fitness function, enabling a direct comparison of optimization strategies and their potential to reduce police response time. To validate the proposal, we compared the three optimization approaches with a baseline approach that uses the PRF unit position as a reference, measuring the distance in the suggested optimization approaches relative to each PRF unit position.

IV. Experiment

This section outlines the experimental configuration and methodology used in this study.

A. Network selected

For this study, we derived the network from Taveira et al. [18], selecting 18 of the 27 states (67% of the total) with the highest number of traffic crashes (85%) concentrated in only 39% of the Brazilian territory. Our intention is to use a more compact network with more representative data. We also used the city population size from Brazilian census bureal (IBGE) data [35]. This network has 2648 nodes and 2742 edges representing the selected states roads. The used network is presented in Figure 2 (C).

B. GNN Configuration

The experimental evaluation considers different feature sets composed of static node attributes and global temporal features. Static features encode spatial properties of the road network, while global temporal features summarize recent accident values using aggregated statistics over fixed temporal windows. We evaluate models using static features only, temporal features only, and their combination in order to assess the contribution of each feature group. All features are standardized based on the training set statistics. We used 2022 and 2023 for training and 2024 for testing, over 200 epochs, and early stopping after 50 epochs without improvement.

We evaluated three architectures (GCN, GAT, and EvolveGCN) using the same feature set, as described in Table I. This selection provides a fair comparison, allowing performance differences to be attributed solely to architectural variations rather than differences in input representations. The feature type indicates whether the feature is static during training or changes over time. We applied a cyclical transform in *Day of week* and *Month of year* features. The level indicates if the feature value is a Global value or refers to the node value.

To evaluate the GNN architecture, we used global and local evaluation metrics. For global evaluation, we used Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). The local evaluation is not the main objective of the training, but we still need to assess its applicability in specific road segments (edges). For that, we used a binary classification to determine if each day the model can predict crashes. For this approach, we used the Recall and Precision metrics to capture the model's sensitivity in predicting crashes. At the same time, we measured the correlation between the number of reported crashes and aimed metrics. For each architecture tried out, results are reported as the mean and standard deviation over 30 independent runs.

C. Optimization

The optimization experiments assess whether the planned patrol routes can reduce crash response time relative to a baseline in which the vehicle remains at its usual station. For each day and station, the candidate set of road segments is defined by the station's coverage area, and travel times between segments are computed on the road network using time-dependent speeds. For the three algorithms evaluated, each one returns a Pareto set of routes for the same instance.

The segments are patrolled by the PRF team, which has a mobile unit (*Unidade Operacional* - UOP) for operational support area during the day. These UOPs are responsible for patrolling a set of segments. Our route planning has four objectives: patrol the more probable traffic-crash segment; patrol the longest extension; visit the largest number of segments with a risk of traffic crashes; and execute all tasks in the shortest possible time. We limit the restriction to the number of hours per day, not to visiting the same segment immediately after it, and to a minimum of segments per day. Another feature is that the team displacement uses a different average velocity during the hour with more jams.

To compare methods, one route is selected from each Pareto set using a fixed rule: among feasible solutions, the route with the higher accumulated risk coverage, in case of ties, the shortest total duration is used as a tie breaker. Response time is then calculated for each observed crash as the network travel time from the closest point on the selected route to the crash segment. For the baseline, response time is calculated from the station reference segment to the crash segment. The results are reported as average response time and percentage reduction relative to the baseline.

TABLE I
OVERVIEW OF THE FEATURE SET USED IN THE EXPERIMENTAL EVALUATION.

Feature	Type	Level	Description
Accumulated accidents	Temporal	Global	Monthly accumulated accident count aggregated over the entire network.
Day of week	Temporal (cyclical)	Global	Weekly temporal pattern encoded using cyclical representation
Day of month	Temporal	Global	Position of the day within the month, capturing intra-month patterns
Month of year	Temporal (cyclical)	Global	Seasonal pattern encoded using cyclical representation
Number of lanes	Structural	Global	Average number of road lanes, representing network structural capacity
Latitude	Static	Node	Geographical latitude of the node
Longitude	Static	Node	Geographical longitude of the node
Population	Static	Node	Population associated with city nodes
City indicator	Static	Node	Binary indicator distinguishing cities from road junctions

In this work, we compared ten optimization areas solution with a baseline approach, as presented in Section III-D.

V. RESULTS AND DISCUSSIONS

This section reports the most significant experimental results obtained. First, we present the GNN prediction results, showing all architectures are evaluated under identical conditions and that results are reported as the mean ± standard deviation over 30 independent runs, maintaining a just architectural comparison. After, we present the optimization results comparing the number of solutions for each optimization algorithm.

A. GNN predictions

We first evaluate global regression performance, which measures each architecture's ability to predict the aggregated number of accidents across the network. Following best practices for event-count prediction, metrics are reported in the order of MAE, RMSE, and R^2. Table II presents the global metrics results.

TABLE II
GLOBAL REGRESSION PERFORMANCE AVERAGED OVER 30 INDEPENDENT RUNS. RESULTS ARE REPORTED AS MEAN ± STANDARD DEVIATION.

Architecture	MAE ↓	RMSE ↓	R^2 ↑
GCN	**500.7 ± 44.2**	1802.7 ± 153.8	0.732 ± 0.044
EvolveGCN	507.7 ± 58.2	**1790.7 ± 154.6**	**0.735 ± 0.046**
GAT	4978.4 ± 2235.7	5822.4 ± 2758.1	-2.385 ± 3.016

The global results show that GCN and EvolveGCN reach comparable, stable performance, both substantially outperforming GAT across all regression metrics. GCN attains the lowest MAE, suggesting superior average Accuracy and consistency, while EvolveGCN achieves marginally better RMSE and R^2, showing a slightly improved handling of extreme values and explained variance.

In contrast, GAT shows instability in its results, with error magnitudes nearly one order of magnitude larger and a negative average R^2, indicating performance worse than that of a naive mean predictor. Statistical tests confirm that GCN and EvolveGCN are statistically equivalent, whereas both are significantly superior to GAT ($p < 0.001$). Statistical significance was assessed using the Wilcoxon signed-rank test at a 95% confidence level [36].

Beyond global regression, we evaluate local performance at the edge level, framing the task as binary accident detection (accident vs. no accident on a given day). This analysis can highlight how well models identify accident occurrences at specific road segments. Table III presents the global results.

TABLE III
LOCAL EDGE-LEVEL BINARY CLASSIFICATION PERFORMANCE FOR ACCIDENT DETECTION, AGGREGATED OVER ALL EDGES.

Architecture	Accuracy	Precision	Recall	F1-score
GCN	**0.916**	**0.166**	0.205	**0.183**
EvolveGCN	0.914	0.162	0.204	0.181
GAT	0.873	0.101	**0.223**	0.139

At the local level, all models face the challenge of severe class imbalance, as accident days constitute a small fraction of the total observations. Under this setting, Accuracy alone is misleading, and Precision–recall trade-offs provide a more meaningful assessment. For example, if the prediction is 0 crashes on all days, the Accuracy could exceed 0.9.

GCN achieves the best overall balance, yielding the highest Precision and F1 Score, consequently lowering false positives while preserving competitive recall. EvolveGCN presents nearly identical behavior, confirming its equivalence to GCN in conservative prediction regimes. GAT, while achieving the highest recall, suffers from systematic overprediction, resulting in substantially lower Precision and a lower F1 Score. This behavior artificially inflates recall and is especially evident in locations with low accident frequency.

Table IV presents an example of GCN result. We used the frequency range of the number of crashes in 2024, relating the number of segments and the recall statistics. We can observe the model achieving the desired quality in predicting rare events (traffic crashes) in a segmented approach for a day. For 8.3% of segments (50+ crashes) with more traffic crashes in 2024, the model could predict 24% of the crashes in a day.

TABLE IV
RECALL STATISTICS BY CRASH FREQUENCY RANGE, BASED ON 2024 PREDICTION

Number Crashes	Number Segments	Recall				
		Mean	Std	Min	Median	Max
0	804	Not measured				
1–9	774	Not measured				
10–29	741	0.07	0.06	0.00	0.07	0.33
30–49	242	0.14	0.06	0.00	0.14	0.29
50–99	169	0.24	0.07	0.06	0.25	0.40
100+	62	0.46	0.11	0.30	0.44	0.78

In summary, GCN offers the best overall balance between global and local metrics, with strong stability. EvolveGCN matches GCN and is slightly more sensitive to extreme events, but requires greater computational cost. GAT performs poorly for this configuration due to instability and excessive false positives, despite higher recall in sparse cases.

B. Optimization

The main objective of the Optimization is to stay in the segments where a crash is predicted and at the time with the highest probability of a crash. For that, we selected a GCN model to generate predictions and optimize resource allocation to either be at or stay closer to road crashes when they occur. To address this scenario, we create a daily journal containing predictions for all segments of all selected roads. Our focus is measure the top 10 sets of segments with the largest number of crashes in 2024. As these are daily predictions, we created statistics for each month, day of the week, and hour range (3-hour intervals) to complement the main probability of crashes.

In this context, we selected 10 areas with the highest number of traffic crashes on the Brazilian federal highway in 2024 to evaluate the solution's performance. Since we don't have any previous work to compare against, we compare the position suggested by the optimization with the segment of the PRF unit in the specific area (baseline), measuring the time to arrive in a traffic crash position. Table V presents the results for each PRF Unit, mean after 30 executions for each algorithm, highlighting the best algorithm, the mean value, and its statistics. The negative value indicates the optimization solution is better than the baseline solution, the optimization position reached the accident faster. Our results indicate a difference between the PRF Units. The more centralized PRF units have bad results, and the more decentralized ones have better results.

TABLE V
PERFORMANCE SUMMARY OF NSGA, MOACO, AND MOPSO FOR RESOURCE ALLOCATION ON THE EXPERIMENTAL NETWORK.

PRF Unit	Best Alg.	Mean	Std. Dev.	Min	Max
PBUOP0101	NSGA-II	-10.43	±0.98	-72.40	63.20
SCUOP0403	NSGA-II	-10.22	±1.25	-46.90	52.60
SCUOP0101	NSGA-II	-3.79	±1.73	-49.70	52.50
SPUOP0101	NSGA-II	-3.43	±0.71	-41.10	48.30
RJUOP0601	NSGA-II	2.48	±0.65	-41.20	48.20
SCUOP0102	NSGA-II	5.63	±1.23	-19.90	42.90
MGUOP0102	NSGA-II	6.55	±1.76	-63.50	67.40
PEUOP0101	NSGA-II	13.63	±0.79	-37.40	89.70
RJUOP0102	NSGA-II	31.21	±1.17	-20.80	131.45
ESUOP0201	NSGA-II	31.36	±3.44	-135.80	148.10

Beyond response time, the candidate solution can also capture other relevant characteristics, such as larger coverage areas and shorter patrol times. These results highlight the feasibility of applying the same network structure to the optimization stage following the prediction task.

The observed variance in performance across different PRF units suggests that the effectiveness of the metaheuristic optimization is intrinsically linked to the underlying spatial topology and the degree of resource centralization. While more centralized units presented challenges for the algorithms, likely due to pre-existing optimized positioning or high road density, the significant improvements in decentralized sectors demonstrate the solution's potential to enhance response logistics.

VI. CONCLUSIONS

This paper proposed an end-to-end approach for predicting road traffic crashes on Brazilian federal roads and supporting proactive patrol planning through a discrete multi-objective optimization. The federal road system was modeled as an undirected graph from SNV data, and daily crash time series were built from PRF open data. Under the same feature set and experimental protocol, GCN and EvolveGCN achieved stable and comparable global performance, while GAT displayed higher variance across runs and weaker performance. At the local edge level, class imbalance limits the value of Accuracy, even so, GCN achieved the most balanced Precision/Recall behavior, and EvolveGCN showed similar results.

Using the predicted risk as input, we formulated a route planning problem that maximizes risk coverage while minimizing travel time and related costs, and evaluated NSGA-II, MOPSO, and P-ACO with a common fitness definition. This optimization layer links prediction to actionable resource positioning, producing alternative patrol plans applied to 10 units with more traffic crashes in 2024.

An evaluation of prediction and optimization results on a road network, modeled using network science perspective, indicates that a single model can be consistently applied throughout the process. Initially, the network structure is used for feature-based prediction and subsequently supports optimization based on the predicted outcomes.

While the optimization results vary across different regions, the risk-based allocation framework offers strategic advantages that extend beyond a reduction in response times. In scenarios where the model outperforms the baseline, the proactive positioning of units in high-criticality segments functions as a potential deterrent mechanism, contributing to the prevention of traffic crashes. Even in units where response times remained stable or increased, the metaheuristic approach provides a systematic method for managing scarce resources, identifying trade-offs that aim to maximize risk coverage under logistical constraints. Furthermore, by integrating a network-science perspective, the model mitigates subjective biases and offers a standardized patrolling strategy capable of adapting to the specific topography of each federal highway unit. Consequently, the framework serves as a robust decision-support tool, enhancing operational consistency regardless of local spatial complexities.

Although this study was conducted under a specific network modeling configuration, the framework can be adapted by refining or aggregating road segments according to different operational strategies. Several avenues for future research can be explored to build upon these findings. First, integrating GNN architectures with time-varying edge weights could better capture the dynamic nature of road networks. Second, the optimization stage could be strengthened by incorporating explicit response-time objectives and constraints directly into the multi-objective formulation, thereby reducing the reliance on distance and duration as indirect proxies. Further investigations might explore network-aware problem definitions, such as candidate selection schemes based on centrality or community structure, as well as hierarchical and multi-resolution graph representations. Additionally, state-specific calibrations of coverage regions and station responsibilities could improve localized performance. Beyond these structural refinements, future extensions should address robust optimization under prediction uncertainty, potentially through ensemble methods or confidence intervals, alongside multi-vehicle coordination and dynamic repositioning based on time-dependent travel conditions. Collectively, these developments aim to enhance the consistency of performance gains across heterogeneous road networks while maintaining the interpretability of the trade-offs within the Pareto set.

Acknowledgment

The authors acknowledge Professor Ronaldo Menezes, University of Exeter, for his support and valuable contributions.

References

[1] W. Fan, Y. Ma, Q. Li, Y. He, E. Zhao, J. Tang, and D. Yin, "Graph neural networks for social recommendation," in *The world wide web conference*, 2019.

[2] S. Li, H. Hua, and S. Chen, "Graph neural networks for single-cell omics data: a review of approaches and applications," *Briefings in Bioinformatics*, 2025.

[3] W. Hassan, M. M. Cabral, T. R. Ramos, A. C. Filho, and L. G. Nonato, "Modeling and predicting crimes in the city of são paulo using graph neural networks," in *Brazilian Conference on Intelligent Systems*, Springer, 2024.

[4] J. Xue, R. Tan, J. Ma, and S. V. Ukkusuri, "Data science in transportation networks with graph neural networks: A review and outlook," *Data Science for Transportation*, 2025.

[5] D. Lord and F. Mannering, "The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives," *Transportation research part A: policy and practice*, 2010.

[6] M.-M. Chen and M.-C. Chen, "Modeling road accident severity with comparisons of logistic regression, decision tree and random forest," *Information*, 2020.

[7] F. Basso, R. Pezoa, M. Varas, and M. Villalobos, "A deep learning approach for real-time crash prediction using vehicle-by-vehicle data," *Accident Analysis & Prevention*, 2021.

[8] A. Nippani, D. Li, H. Ju, H. Koutsopoulos, and H. Zhang, "Graph neural networks for road safety modeling: Datasets and evaluations for accident analysis," *Advances in neural information processing systems*, 2023.

[9] World Health Organization, *Global status report on road safety 2023*. Geneva: World Health Organization, 2023.

[10] World Bank, "Regional poverty and inequality update: Latin america and the caribbean - october 2024," report, The World Bank, Washington, DC, 2024.

[11] O. L. de OLIVEIRA and F. B. de LIMA NETO, "Analysis of road-traffic behavior in the metropolitan area of recife-brazil," *12ª Conferência Ibérica de Sistemas e Tecnologias de Informação*, 2017.

[12] L. Melo, J. Taveira, P. Varjal, and F. B. Lima Neto, "Previsão de indicadores de resultados em sinistros de trânsito utilizando séries temporais hierárquicas, key result indicator traffic accidents forecast using hierarquical times series.," in *CISTI'2024 - 19th Iberian Conference on Information Systems*, 2024.

[13] J. C. F. Taveira, P. V. Silva de Lima, P. Varjal, and F. B. Lima Neto, "Predição multiníveis diária de sinistros de trânsito utilizando séries temporais hierárquicas, daily multi-level prediction of traffic accidents using hierarchical time series.," in *CISTI'2025 - 20th Iberian Conference on Information Systems*, 2025.

[14] D. H. Moura, "Indicador de ponto crítico de sinistro com morte: Uma contribuição para a redução de óbitos no trânsito," *ARACÊ*, vol. 7, 2025.

[15] P. Wanke, L. A. Gil-Alana, and Y. Tan, "Trends in road accidents on brazilians highways: Evidence of persistence using fractional integration," 2023.

[16] A. Mukhopadhyay, G. Pettet, S. M. Vazirizade, D. Lu, A. Jaimes, S. E. Said, H. Baroud, Y. Vorobeychik, M. Kochenderfer, and A. Dubey, "A review of incident prediction, resource allocation, and dispatch models for emergency management," *Accident Analysis & Prevention*, 2022.

[17] G. Zou, Z. Zhou, R. Weibel, Y. Li, T. Wang, Z. Liu, W. Ding, and C. Fu, "Multi-graph spatio-temporal network for traffic accident risk forecasting," *Pattern Recognition*, 2025.

[18] J. Taveira, R. Menezes, and F. Buarque, "Not all roads are created equal: Network science shows some highways matter more for brazil's connectivity," *Applied Netowrk Science*, vol. 11, 2026.

[19] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Transactions on Neural Networks*, 2009.

[20] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[21] Y. Zheng, L. Yi, and Z. Wei, "A survey of dynamic graph neural networks," *Frontiers of Computer Science*, 2025.

[22] M. Jin, H. Y. Koh, Q. Wen, D. Zambon, C. Alippi, G. I. Webb, I. King, and S. Pan, "A survey on graph neural networks for time series: Forecasting, classification, imputation, and anomaly detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[23] A. Banerjee, D. Singh, S. Sahana, and I. Nath, "Chapter 3 - impacts of metaheuristic and swarm intelligence approach in optimization," in *Cognitive Big Data Intelligence with a Metaheuristic Approach* (S. Mishra, H. K. Tripathy, P. K. Mallick, A. K. Sangaiah, and G.-S. Chae, eds.), Cognitive Data Science in Sustainable Computing, Acad. Press, 2022.

[24] Q. Cappart, D. Chételat, E. B. Khalil, A. Lodi, C. Morris, and P. Veličković, "Combinatorial optimization and reasoning with graph neural networks," *Journal of Machine Learning Research*, 2023.

[25] Empresa de Planejamento e Logística, "Plano nacional de logística 2035 - relatório executivo," technical report, Ministério da Infraestrutura, Brasília, DF, October 2021. In Portuguese.

[26] Departamento Nacional de Infraestrutura de Transporte (DNIT), "Sistema nacional viário (snv)," 2024. In Portuguese.

[27] Polícia Rodoviária Federal, "Polícia rodoviária federal – prf." https://www.gov.br/prf, 2025. Accessed: Jan. 2026.

[28] R. Prieto Curiel, H. González Ramírez, and S. R. Bishop, "A novel rare event approach to measure the randomness and concentration of road accidents," *PLOS ONE*, Aug. 2018.

[29] M. Hoffmann, L. Kotzur, D. Stolten, and M. Robinius, "A review on time series aggregation methods for energy system models," *Energies*, 2020.

[30] M. Jin, H. Y. Koh, Q. Wen, D. Zambon, C. Alippi, G. I. Webb, I. King, and S. Pan, "A survey on graph neural networks for time series: Forecasting, classification, imputation, and anomaly detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[31] A. Pareja, G. Domeniconi, J. Chen, T. Ma, T. Suzumura, H. Kanezashi, T. Kaler, T. Schardl, and C. Leiserson, "Evolvegcn: Evolving graph convolutional networks for dynamic graphs," in *Proceedings of the AAAI conference on artificial intelligence*, 2020.

[32] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: Nsga-ii," *IEEE transactions on evolutionary computation*, 2002.

[33] C. C. Coello and M. S. Lechuga, "Mopso: A proposal for multiple objective particle swarm optimization," in *Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02*, 2002.

[34] K. Doerner, W. J. Gutjahr, R. F. Hartl, C. Strauss, and C. Stummer, "Pareto ant colony optimization: A metaheuristic approach to multiobjective portfolio selection," *Annals of operations research*, 2004.

[35] Brazilian Institute of Geography and Statistics (IBGE), "Brazilian population census," 2022.

[36] W. J. Conover, *Practical Nonparametric Statistics*. Wiley, 1999.