2025 Engineering for Palestine Conference (ENG4PAL)
PPU, Hebron, Palestine, September 29-30, 2025

Integrating Large Language Models for Traffic Crash Classification in Data-Constrained Regions

Huthaifa I. Ashqar, Ahmed Jaber, Khaled Al-Sahili, Mujahid I. Ashqer, and Fady MA Hassouna

Abstract— **Traffic crashes remain a critical public safety issue in developing regions, where poor data quality, inconsistent reporting, and inadequate classification frameworks hinder effective analysis and policy action. This study presents a novel methodology for enhancing traffic crash classification using advanced large language models in a data-constrained environment. We employ OpenAI's GPT-4o to reclassify 48,815 crash records from Palestine (2019–2022), converting unstructured narratives into a structured taxonomy aligned with international standards. Using few-shot learning, GPT-4o achieved over 92% classification accuracy, revealing dominant causes such as failure to yield (42.64%) and driver negligence (34.76%), while identifying 11.55% of cases as pedestrian-related. To explore predictive potential, we integrate this refined dataset into a ConvLSTM (Convolutional Long Short-Term Memory) model, capturing spatio-temporal dependencies to forecast crash frequencies across regions and time intervals. This integration of large language model-driven reclassification with deep spatio-temporal learning not only improves the reliability of crash datasets but also uncovers actionable insights for targeted road safety interventions. Comparative analysis with existing literature highlights the novelty of applying LLMs in tandem with spatio-temporal models for traffic analysis in low-resource settings. The findings carry significant implications for data-driven policymaking, infrastructure design, and urban traffic safety strategies.**

I. Introduction

Traffic crashes remain a critical issue worldwide, particularly in developing countries where the lack of structured and detailed traffic crash data exacerbates the challenge of addressing road safety effectively [1]. In many of these regions, crash records are often unstructured, containing vague classifications and inconsistent reasoning codes, making it difficult to analyze the root causes and patterns of crashes. A better understanding of the contributing factors to crashes is vital for implementing effective traffic safety interventions, improving road design, and guiding policymaking. However, to achieve this understanding, existing data must be cleaned, restructured, and categorized in a way that is both systematic and actionable [2], [3].

Traffic crashes in Palestine represent a significant public safety concern, with data indicating a steady increase in the number of road traffic crashes over recent years [4], [5], [6]. Our data showed that between 2019 to 2022, there were more than 21,000 reported injuries and 880 fatalities. These statistics underscore the urgent need for comprehensive analysis and intervention to enhance road safety. However, the effectiveness of such efforts is often compromised by inconsistent traffic crash data and the absence of detailed classification, which hinder robust analysis and informed policymaking [7], [8].

The complexity of traffic crashes is further compounded by their spatio-temporal nature. Crashes are influenced by a range of factors, such as weather conditions, time of day, and the number of vehicles involved; all of which vary across space and time. To accurately model these dynamics, advanced techniques are required, capable of capturing both the spatial and temporal dependencies in crash data. Recent advancements in machine learning, particularly recurrent neural networks (RNNs) with convolutional structures such as ConvLSTM, offer promising solutions for spatio-temporal prediction in traffic studies [9], [10].

This study aims to address these challenges by employing two main contributions. First, we applied GPT-4o as a smart data cleaner and annotator, turning noisy, low-quality data into structured, globally comparable inputs using in-context learning in low-resource Natural Language Processing (NLP) settings. Second, the study employs ConvLSTM networks to predict and interpret the causes of crashes by modeling the spatio-temporal dependencies inherent in traffic crashes. These contributions are expected to offer valuable insights for traffic safety planning, interventions, road design improvements, and evidence-based policy development.

II. Literature Review

Traffic safety is a critical concern worldwide, and numerous studies have focused on understanding the factors contributing to road crashes, especially in developing regions where data is often unstructured or inconsistent. Various machine learning models have been applied to analyze traffic crash data to predict outcomes, identify key factors, and develop safety interventions. For instance, Ghandour et al. [11] aimed to identify key factors contributing to fatal road injuries in Lebanon by using a hybrid ensemble machine learning model. Their model, which combined sequential minimal optimization and decision trees, analyzed 8,482 crash incidents. The study identified several significant variables, such as crash type, injury severity, spatial cluster-ID, and crash

H. I. Ashqar is with the Department of AI and Data Science, Arab American University, Palestine, and AI Program, Columbia University, New York, NY, USA; (e-mail: Huthaifa.ashqar@aaup.edu)

A. Jaber is with the Department of Transport Technology and Economics, Budapest University of Technology and Economics, Hungary, and Association of Palestinian Local Authorities, Ramallah, P300, Palestine; (e-mail: ahjaber6@edu.bme.hu)

K. Alsahili is with the Civil and Architectural Engineering Department, An-Najah National University, Palestine; (e-mail: alsahili@najah.edu)

M. I. Ashqer is with the Computational and Industrial Engineering, North Carolina A&T State University, Greensboro, NC, USA; (e-mail: miashqer@ncat.edu)

F. Hassouna is with the Civil and Architectural Engineering Department, An-Najah National University, Palestine; (e-mail: fady.h@najah.edu)

978-9950-8599-0-6

time, as contributing to fatal crashes. These insights were intended to guide policymakers in developing targeted safety programs. Similarly, Zhang et al. [12] employed an ensemble machine learning framework to predict crash frequency and found that models like random forest and extremely randomized trees performed better in predicting crash rates. The study highlighted important crash factors and provided recommendations for reducing crashes.

Rahman et al. [13] developed decision tree regression models to predict pedestrian and bicycle crashes in Florida using statewide traffic analysis data. The study revealed that traffic volume, roadway characteristics, and sociodemographic factors were significant predictors. Akin et al. [14] focused on driver errors along a major highway in Saudi Arabia, applying binomial logistic regression and comparing it with random forest and k-nearest neighbor models. They identified key factors, such as road type and traffic volume, which increased the likelihood of driver error-related crashes.

Al-Mistarehi et al. [15] integrated machine learning with geographic information systems (GIS) to predict crash severity and identify high-risk areas. Their study found that Random Forest outperformed other algorithms in predicting crash severity, and GIS integration provided valuable insights for road safety interventions. Abdel-Aty et al. [16] focused on angle crashes at unsignalized intersections, applying machine learning techniques to predict the likelihood of such crashes based on various factors, including road geometry and weather conditions. The goal was to inform safety improvements at these intersections. Mirzahossein et al. [17] developed statistical and machine learning models to predict road traffic crash severity on rural roads. Their study emphasized the role of inattentiveness and vehicle-motorcycle crashes, identifying factors that could inform targeted interventions. Komol et al. [18] applied machine learning classification techniques to model injury severity among vulnerable road users, such as pedestrians, bicyclists, and motorcyclists. Their research highlighted the significant role of road type, lighting conditions, and vehicle type in influencing injury severity, underscoring the need for tailored safety measures for these at-risk groups.

Recent studies have highlighted how road design and urban form influence the severity of bike crashes. In Budapest, Jaber and Csonka [19] used GWR analysis to show that crash severity was higher near crossings, commercial areas, and public transport stops, while green spaces and suburban traffic signals were linked to lower severity. Similarly, a binary regression of over 14,000 cyclist injuries found that one-way roads, signalized intersections, and urban areas were associated with more severe outcomes, emphasizing the role of infrastructure in shaping cycling safety [20].

While these studies make significant contributions to the field of traffic safety through predictive modeling and machine learning techniques, there remains a gap in the systematic structuring and preprocessing of unstructured data in developing countries. Furthermore, while most studies focus on crash prediction and severity analysis, few integrate spatio-temporal models to capture the complexities of crash dynamics over time and space.

This study differentiates itself by addressing two key challenges. First, it tackles the issue of unstructured and inconsistent data by using GPT-4o to preprocess and classify crash data systematically. This approach resolves inconsistencies and introduces a new classification scheme based on a detailed evaluation of crash characteristics. Second, it employs ConvLSTM, a spatio-temporal model, to predict and interpret the causes of crashes, incorporating both spatial and temporal dependencies in the crash data. This methodology provides a more understanding of the factors contributing to crashes, particularly in regions like Palestine, where data quality and structure are significant hurdles.

III. Methodology

The methodology employed in this study consists of two primary phases: data preprocessing and spatio-temporal crash prediction. First, recognizing that the original crash dataset was poorly structured with inconsistent classifications, we developed a preprocessing pipeline using fine-tuned GPT-4o model. This automated system systematically analyzed each crash record individually, accurately identifying the vehicles involved and reclassifying the causes based on a detailed evaluation of the crash characteristics and original reasoning codes [21]. This step was critical because traditional datasets from developing regions often suffer from generalization and ambiguity, hindering accurate analysis. By introducing a more refined classification scheme, we laid a stronger foundation for subsequent predictive modeling and ensured that the resulting insights would be both reliable and actionable for traffic safety interventions.

Following the reclassification, we employed a ConvLSTM model to capture the spatio-temporal dynamics of crash occurrences and predict underlying causes. ConvLSTM was selected over traditional machine learning models because it integrates both convolutional structures (to handle spatial features) and recurrent architectures (to handle temporal sequences), making it particularly suitable for modeling crash data that varies across space and time. This dual focus allows the model to interpret complex interactions between environmental factors, temporal patterns, and crash causes. The preprocessing innovation and the advanced predictive model distinguish our methodology from previous works, which will enable more precise and practical recommendations for road design, traffic policies, and safety strategies.

A. Dataset

The dataset used in this study contains 48,815 crash records collected from traffic authorities in Palestine from 2019 to 2022. These records include several variables, such as weather conditions, time, number of vehicles involved, road conditions, and crash type. However, the raw dataset presented several challenges in terms of data quality, including inconsistent classification, incomplete information, and general terms used for crash causes. Crash distribution by the day of the week and by the month are shown in Figure 1 and Figure 2, respectively.

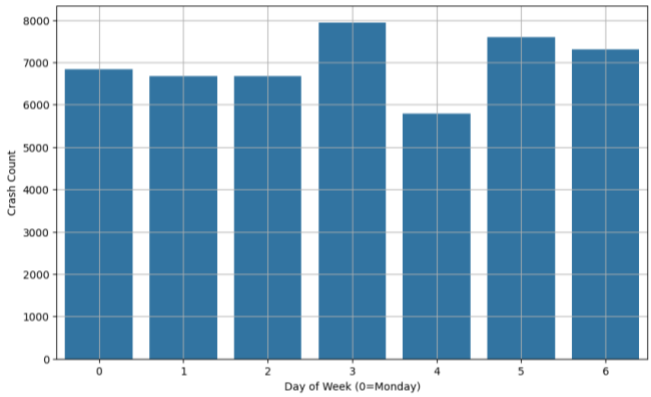

Figure 1. Crash Distribution by the Day of the Week.

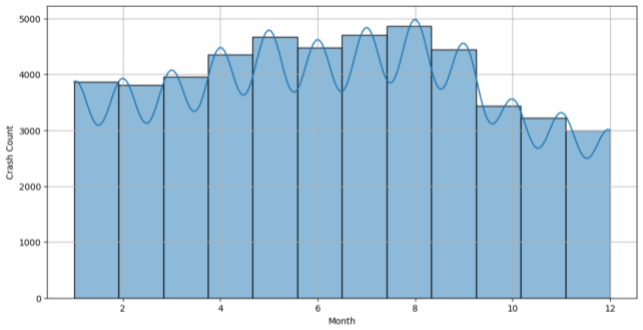

Figure 2. Crash Distribution by the Month.

B. GPT-4o-Based Preprocessing and Reclassification

To address the aforementioned issues, the study applied GPT-4o-based preprocessing to restructure and clean the dataset [22], [23], [24], [25], [26]. The preprocessing pipeline focused on the following tasks:

• Identification and Removal of Inconsistencies: GPT-4o was used to analyze and identify discrepancies in the classification of crashes, standardizing terms and ensuring that the crash data could be reliably compared.

• Crash and Reason Classification: A new classification scheme was developed to categorize crashes based on specific characteristics, such as crash type, injury severity, weather conditions, and temporal aspects (e.g., time of day). This classification process aimed to provide a more detailed and actionable understanding of the factors contributing to crashes.

Once the data was cleaned and organized, the next step involved classifying the crashes into different categories, based on both pre-existing and newly defined attributes. Key features identified for classification included:

• Vehicle Types Involved: The specific vehicles involved in the crash (e.g., cars, trucks, motorcycles) were identified and categorized.

• Weather and Temporal Features: Weather conditions, such as rain, fog, or clear skies, and time-related features, such as the hour of the day or day of the week, were also integrated as crucial factors influencing crash outcomes.

• Road Conditions: Factors such as road type, lighting conditions, and the presence of traffic signals were considered.

Figure 3 shows the multi-phase methodology that this study adopted, which combines expert-guided data curation with large language model (LLM)-assisted classification to restructure and reclassify crash data for improved accuracy and analytic value [27], [28], [29]. The original dataset consisted of 48,815 records, each containing a wide range of features, including structured variables (e.g., date, time, number of vehicles, location) and unstructured narrative fields such as "traffic_case_reason." Upon preliminary inspection, the dataset exhibited significant inconsistencies, redundancy, and misalignment between the narrative descriptions and categorical labels. These issues limited the dataset's reliability for downstream analytics or policy-relevant insights.

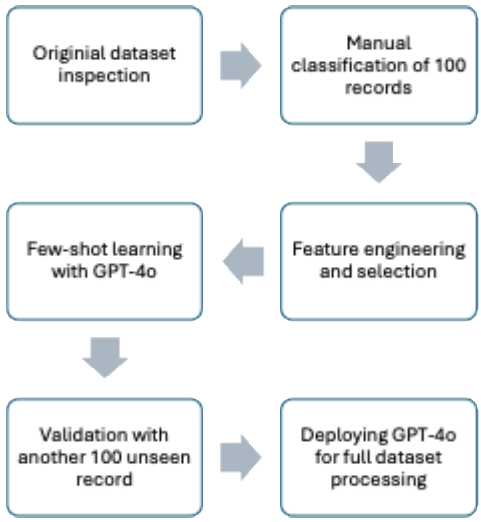

Figure 3. Flowchart of GPT-4o-Based Reclassification of Crash Data.

To address this, we first conducted a comprehensive manual audit of 100 randomly selected records. This subset was chosen based on stratified sampling to ensure it represented a wide range of crash types, locations, and contextual diversity. Through this manual review, we identified key discrepancies between the narrative and the labeled crash type, along with irrelevant or sparsely populated features that diluted the signal in the data. We reclassified each of the 100 records using a schema inspired by internationally recognized crash classification frameworks, allowing us to define a new feature: Global Crash Classification. Simultaneously, we extracted a new variable, Reclassified Reason, to encode the underlying cause of the crash in a standardized and policy-actionable way.

After feature pruning and refinement, the resulting dataset retained only the most relevant variables, such as id, case_number, year, date, time, full_location, vehicle_count, related_parties_count, and indicators for party role (Driver, Passenger, Witness), weather conditions (e.g., Rain, Snow), vehicle type (e.g., Bicycle), age, and gender. This curated feature set preserved contextual richness while eliminating noise, preparing the data for efficient processing by an AI model.

To automate the reclassification across the entire dataset, we applied few-shot in-context learning using OpenAI's ChatGPT-4o model. The 100 manually audited and labeled

records were used as exemplars to condition the model's understanding of how to map narratives and associated fields to the new classification standards. The model was then tested on a second random batch of 100 records to verify alignment with human expectations. Results showed approximately 94% accuracy, judged by manual reviewers. Following this validation, ChatGPT-4o was deployed across the entire dataset. A third round of validation was conducted on an additional 100 randomly sampled outputs, yielding an accuracy of 92%, confirming the model's generalizability.

The dataset was then split for downstream analysis, with the new classifications used as target variables for training and evaluation. By combining human expertise, standardized crash classification principles, and LLM capabilities, this methodology ensured both semantic correctness and scalability in the crash data reclassification process. The pseudocode for the GPT-4o-based method of reclassification crash data is shown in Table I.

TABLE I. PSEUDOCODE FOR GPT-4O-BASED RECLASSIFICATION OF CRASH DATA.

Pseudocode for GPT-4o-Based Reclassification of Crash Data

Step 1: Load and clean dataset (D)
D_raw ← load_dataset("crash_data.csv")
D_clean ← clean_and_standardize(D_raw)

Step 2: Select representative sample for manual classification
D_sample_train ← select_representative_sample(D_clean, size=100)

Step 3: Manually annotate new labels (Global Crash Classification and Reclassified Reason)
for record in D_sample_train:
  record["Global_Crash_Classification"] ← manual_classification(record)
  record["Reclassified_Reason"] ← manual_reason_identification(record)

Step 4: Define prompt structure for few-shot learning
prompt_examples ← []
for record in D_sample_train:

prompt_examples.append(format_as_few_shot_example(record))

Step 5: Apply GPT-4o to a test set for validation
D_sample_test ← select_random_sample(D_clean, size=100, exclude=D_sample_train)
correct_predictions ← 0

for record in D_sample_test:
  input_prompt ← build_prompt(prompt_examples, record)
  prediction ← GPT4o.predict(input_prompt)

  if prediction["classification"] == manual_classification(record) and \
    prediction["reason"] == manual_reason_identification(record):
    correct_predictions += 1

accuracy_test ← correct_predictions / length(D_sample_test)

Step 6: Apply GPT-4o to full dataset
D_full_reclassified ← []

for record in D_clean:
  input_prompt ← build_prompt(prompt_examples, record)
  prediction ← GPT4o.predict(input_prompt)

  record["Global_Crash_Classification"] ← prediction["classification"]
  record["Reclassified_Reason"] ← prediction["reason"]
  D_full_reclassified.append(record)

Step 7: Final validation on another random subset
D_sample_final ← select_random_sample(D_full_reclassified, size=100)
accuracy_final ← validate_predictions(D_sample_final, manual_ground_truth)

Output the final reclassified dataset
save_dataset(D_full_reclassified, "reclassified_crash_data.csv")

C. Spatio-Temporal Prediction Using ConvLSTM

For predicting and analyzing the causes of crashes, we applied a ConvLSTM model, a type of recurrent neural network (RNN) with convolutional layers, to capture both spatial and temporal dependencies in the crash data. This method allows the model to consider the sequential nature of traffic crashes, as well as the spatial distribution of crashes across various regions [9], [10], [30]. The ConvLSTM architecture consists of Convolutional layers, where these layers capture spatial patterns, helping the model to identify regional clusters or high-risk zones where crashes are more frequent, and LSTM layers, where these layers process the temporal relationships between crashes, considering factors like time of day and seasonality, to predict crash trends over time.

Given an input sequence of spatial grids $\{\mathbf{X}_t\}_{t=1}^{T}$, where $\mathbf{X}_t \in \mathbb{R}^{h\times w\times c}$ represents the crash-related features (e.g., frequency, weather, classification) at time t, the ConvLSTM updates its hidden and cell states using the following equations [31]:

$$\boldsymbol{i}_t = \sigma(\boldsymbol{W}_{xi} * \mathbf{X}_t + \boldsymbol{W}_{hi} * \mathbf{H}_{t-1} + \boldsymbol{b}_i)$$
$$\boldsymbol{f}_t = \sigma(\boldsymbol{W}_{xf} * \mathbf{X}_t + \boldsymbol{W}_{hf} * \mathbf{H}_{t-1} + \boldsymbol{b}_f)$$
$$\boldsymbol{C}_t = \boldsymbol{f}_t \odot \boldsymbol{C}_{t-1} + \boldsymbol{i}_t \odot \tanh(\boldsymbol{W}_{xc} * \mathbf{X}_t + \boldsymbol{W}_{hc} * \mathbf{H}_{t-1} + \boldsymbol{b}_c)$$
$$\boldsymbol{o}_t = \sigma(\boldsymbol{W}_{xo} * \mathbf{X}_t + \boldsymbol{W}_{ho} * \mathbf{H}_{t-1} + \boldsymbol{b}_o)$$
$$\mathbf{H}_t = \boldsymbol{o}_t \odot \tanh(\boldsymbol{C}_t)$$

where: σ is the sigmoid activation function, $*$ denotes the convolution operator, $\odot$ denotes the Hadamard (element-wise) product, $\boldsymbol{i}_t, \boldsymbol{f}_t, \boldsymbol{o}_t$ are the input, forget, and output gates, and $\boldsymbol{C}_t$ and $\boldsymbol{H}_t$are the cell state and hidden state at time t. This formulation allows the model to not only remember temporal sequences but also learn localized spatial features through convolutional filters, making it particularly powerful for predicting traffic crash risks across both time and space [31].

The ConvLSTM model was trained using a subset of the dataset, and the outcomes were validated using a holdout testing set to assess the accuracy and reliability of the predictions. The model aimed to predict the likelihood of different types of crashes occurring under varying conditions, helping to uncover hidden patterns and trends in traffic crashes.

The ConvLSTM architecture, summarized in Table II, begins with a ConvLSTM2D layer consisting of 32 filters, each capturing spatio-temporal patterns over 11-time steps. This layer captures localized temporal dynamics by convolving over the time series of spatial inputs. It is followed by a Batch Normalization layer to stabilize and accelerate learning. The 3D output is then flattened and passed through a Dense layer with 64 units, enabling the extraction of high-level temporal features. A final Dense layer with a single neuron is used to produce the output, which represents the predicted crash frequency (or count) for a given spatial region and time step.

This modeling approach allows for both short-term and long-term pattern discovery, making it a powerful tool for predicting traffic crash occurrences. The integration of reclassified crash types and causes as model features further enhances its forecasting accuracy and interpretability.

TABLE II. CONVLSTM ARCHITECTURE.

Layer Type	Output Shape	Parameters
ConvLSTM2D	(Samples, 11, 1, 32)	4,352
BatchNormalization	(Samples, 11, 1, 32)	128
Flatten	(Samples, 352)	0
Dense	(Samples, 64)	22,592
Dense	(Samples, 1)	65

IV. ANALYSIS AND RESULTS

A. *GPT-4o-Based Preprocessing and Reclassification Results*

To improve the quality and interpretability of traffic crash data, we applied a structured few-shot learning methodology using GPT-4o to reclassify both the crash types and the underlying reasons. The original dataset consisted of 48,815 traffic crash records with numerous issues, including ambiguous or inconsistent labels, redundant features, and vague narrative descriptions. A pilot subset of 100 representative records was manually annotated according to globally recognized standards of crash classification and causation, establishing a clean and reliable benchmark. These annotated examples served as the input for GPT-4o's few-shot learning process. Following an initial training phase, the model was validated on a second random set of 100 records, yielding an accuracy of 94%. After reclassifying the entire dataset, a third validation on another random sample showed 92% accuracy, confirming the method's robustness and scalability.

Crash Type Reclassification Results: Table III shows the transformation from the original crash classification system, which was ambiguous and overlapping, into a globally aligned classification. The reclassification resulted in a more interpretable and granular dataset. For instance, the most common original categories were "collision" (41.93%) and "property damages" (40.60%), which offered little differentiation regarding the crash dynamics or vehicle types involved. Post-reclassification, the data was distributed across more meaningful and actionable categories, which reveals insights previously obscured by generic labeling. Multi-Vehicle Collisions became the largest category, now comprising 39.47% of crashes. Single-Vehicle Crashes were clarified and accounted for 22.28% of cases, highlighting the significance of isolated driver errors or environmental hazards. Heavy Vehicle Collisions represented 7.94%, offering new perspectives on the role of trucks and other large vehicles. Minor but meaningful categories such as Collision with Animal Cart (0.05%) and Special Purpose Vehicle Collision (0.04%) were also recovered, pointing to previously buried crash modalities. A category for Public Transport Crashes (0.75%) was created, enabling better policy targeting for commuter safety. These new classes align with international traffic safety norms and help differentiate risk factors by vehicle type and crash topology; something the original dataset lacked.

TABLE III. ORIGINAL VS. GPT-4O RECLASSIFIED CRASH TYPE DISTRIBUTION.

Original Classification	**#**	**%**
Crash with an Animal	80	0.16%
Collision	20,467	41.93%
General	336	0.69%
Hit by a Vehicle	4,532	9.28%
Property Damages	19,819	40.60%
Rollover	3,581	7.34%
Total	**48,815**	**100%**
Reclassified (Global Classification)	**#**	**%**
Collision with Animal Cart	22	0.05%
Collision with Bicycle	1,324	2.71%
Heavy Vehicle Collision	3,876	7.94%
Multi-Vehicle Collision	19,265	39.47%
Public Transport Crash	364	0.75%
Single-Vehicle Crash	10,874	22.28%
Special Purpose Vehicle Collision	20	0.04%
Unspecified or Miscellaneous	13,070	26.77%
Total	**48,815**	**100%**

Crash Causes Reclassification Results: Table IV shows the distribution of reclassified crash causes. The results highlight key behavioral and structural contributors to crashes, enabling better policy focus and preventative strategies. The reclassification of crash causes also produced sharper, evidence-based insights. In the original dataset, causes were poorly documented or embedded in narrative formats. After reclassification, the leading contributing factors were dramatically changed. Failure to Yield/Stop accounted for 42.64%, making it the dominant contributing cause to crashes. Driver Negligence followed with 34.76%, capturing broad behavioral failings such as distraction or inattention. Pedestrian-Related Issues emerged as a significant third category, constituting 11.55% of crashes—information critical for urban planning and pedestrian infrastructure. Less frequent but vital causes like Speeding and Reckless Driving (7.28%) and Overtaking Errors (0.83%) provided nuanced understanding of high-risk driving behaviors. Notably, Environmental Hazards (0.01%) and Unintentional Outcomes (0.22%) were extremely rare, suggesting either underreporting or low incidence in the dataset. This prompts further exploration of how environmental data is captured and integrated.

TABLE IV. RECLASSIFIED CAUSES FOR TRAFFIC CRASHES

Reclassified Cause	#	%
Failure to Yield/Stop	20,813	42.64%
Driver Negligence	16,968	34.76%
Pedestrian-Related Issues	5,638	11.55%
Speeding and Reckless Driving	3,553	7.28%
Parking/Load Issues	1,321	2.71%
Overtaking Errors	407	0.83%
Unintentional Outcomes	108	0.22%
Environmental Hazards	7	0.01%
Total	48,815	100%

Method Effectiveness and Implications for Data Collection: The high classification accuracy and the clarity of the resulting taxonomy reflect the effectiveness of using GPT-4o with few-shot learning for structured reclassification tasks. The method not only cleaned and standardized the data but also uncovered latent patterns that were otherwise obscured. It demonstrated the utility of large language models in processing semi-structured and narrative-heavy datasets, particularly when traditional supervised learning approaches would require larger annotated corpora and extensive feature engineering.

Beyond methodological contributions, this process exposed critical insights into the state of traffic crash reporting. The dominance of vague or non-standardized labels in the original dataset hindered actionable policy formulation. The success of reclassification points to a need for more structured data entry at the point of reporting, including clearer taxonomies for crash types, involved parties, and causative factors. Additionally, the significant presence of pedestrian and public transport-related crashes calls for targeted safety interventions.

B. ConvLSTM Spatio-Temporal Prediction Results

To better understand the structural organization and potential separability of traffic crash types prior to deep learning, we visualized the input space of the ConvLSTM model using t-distributed Stochastic Neighbor Embedding (t-SNE) [32]. The ConvLSTM model requires inputs in a 5D tensor format $[samples, time\ steps, rows, cols, channels]$. In this study, the reshaped input took the form of $(n, 1, f, 1,1)$, where each crash sample was represented as a time-invariant single frame of spatially flattened features. To perform the visualization the high-dimensional ConvLSTM input was flattened to a 2D matrix and t-SNE was applied to project the feature vectors into a 2D space with $n_{components} = 2$ and $perplexity = 30$. Each point in the resulting plot was colored according to its reclassified crash category as determined by the GPT-4o-based classification step. The result of this visualization is shown in Figure 4. This approach enables interpretable examination of the distribution of different crash types within the ConvLSTM feature space prior to model training. The separation between clusters suggests that the reclassification by GPT-4o aligned well with the underlying feature structure, which supports both the reliability of the classification model and the potential predictive power of the ConvLSTM model. This visualization provides an intuitive view into the quality and organization of the input features and gives early evidence of how crash typologies may be learnable via deep learning approaches.

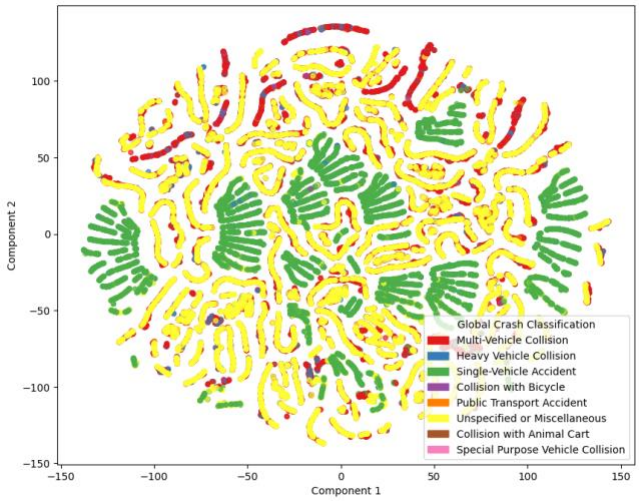

Figure 4. Two-dimensional t-SNE Projection of the ConvLSTM Input Features.

Each point in Figure 4 represents a traffic crash case, and colors indicate the reclassified crash type based on the GPT-4o model. The emergent clusters suggest that different crash types possess distinguishable input patterns, validating both the input representation and the reclassification process.

Figure 5 presents the loss curves observed during the training and validation phases of the ConvLSTM model for spatio-temporal crash prediction. The model was trained for 30 epochs using a mean squared error (MSE) loss function. Results illustrate the progression of both training and validation loss across epochs. Initially, both curves exhibit a steep decline, indicating rapid learning and effective weight updates during the early training phase. From epoch 5 onward, the training and validation losses closely follow each other, with minimal divergence. This tight coupling between training and validation performance indicates a low risk of overfitting, suggesting that the model generalizes well to unseen data. Minor fluctuations in the validation curve—such as the slight increase around epochs 20 and 27—are normal and do not signify performance degradation. The steady decrease and stabilization of the loss curves confirm the model's training stability and convergence, reinforcing its suitability for the spatio-temporal prediction of crashes.

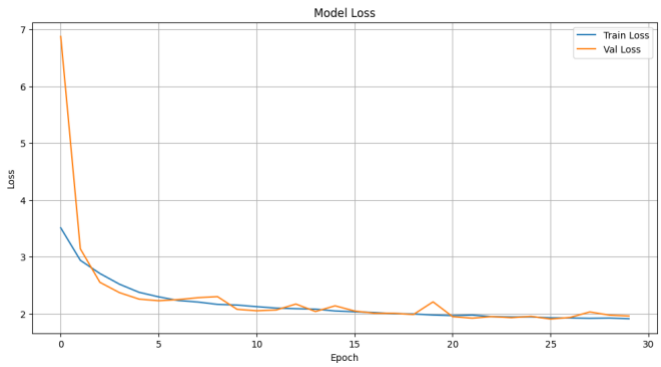

Figure 5. Training and Validation Loss Curves of the ConvLSTM Model Over 30 Epochs. Both curves converge smoothly, with minimal overfitting, suggesting effective learning and generalization to unseen crash data.

V. Discussion and Comparison

While numerous studies have explored traffic crash classification and prediction using machine learning methods, the findings of this study provide distinct insights due to its dual contribution: smart data reclassification using GPT-4o and spatio-temporal crash prediction using ConvLSTM.

In comparison to Ghandour et al. [11], who used hybrid ensemble models to analyze fatal crash factors in Lebanon, our study goes beyond traditional classification by restructuring unstructured narrative data and introducing internationally aligned categories. Their work identified crash time and spatial cluster as major predictors, which aligns partially with our ConvLSTM results, where temporal and spatial dependencies strongly influenced crash frequency.

Rahman et al. [13] and Komol et al. [18] focused on vulnerable road users, pedestrians and cyclists, and highlighted the influence of roadway characteristics and lighting conditions. Similarly, our reclassification using GPT-4o revealed that pedestrian-related crashes accounted for 11.55% of all incidents, which is an insight that was obscured in the original dataset due to vague and overlapping categories. This significant proportion of pedestrian-related incidents underscores the urgent need for targeted infrastructure interventions, particularly in densely populated urban areas.

Akin et al. [14] emphasized driver behavior and errors as leading causes of crashes, using supervised machine learning models in the Saudi context. Our findings align with this, revealing driver negligence (34.76%) and failure to yield/stop (42.64%) as the top two causes of traffic crashes after reclassification. This reinforces the critical need for behavioral interventions, road-user education, and stricter enforcement in similar low-resource settings.

Importantly, few studies have attempted to combine data cleaning with deep learning. Our application of GPT-4o enabled the identification of nuanced crash types such as public transport collisions (0.75%), special purpose vehicle collisions (0.04%), and animal cart collisions (0.05%), categories virtually absent in earlier literature due to dataset limitations. Additionally, ConvLSTM enabled the prediction of crash likelihood by capturing the underlying sequence of events across both space and time, a methodological strength largely missing from the machine learning studies reviewed.

In summary, the findings of this study bridge critical gaps in the current literature by enhancing crash classification granularity and enabling temporally aware predictions. This combination provides a more complete understanding of crash dynamics in developing regions and offers a scalable framework for AI-based traffic safety policy formulation.

VI. Conclusion

This study presents a dual-layered framework for traffic crash analysis and prediction in a developing region context. By using the GPT-4o model for automated data cleaning, reclassification, and annotation, we successfully transformed a noisy, inconsistent crash dataset into a structured and interpretable corpus aligned with international standards. This automated yet accurate method achieved over 92% classification accuracy, enabling detailed insights into crash typologies and causes, thus identifying dominant patterns like failure to yield and driver negligence.

Moreover, the integration of ConvLSTM for spatio-temporal crash prediction demonstrated the ability of deep learning models to learn from complex traffic datasets, capturing both spatial clustering and temporal sequences of crash events. The t-SNE visualization confirmed the underlying structural coherence of GPT-4o-based classifications, while training curves validated the ConvLSTM model's generalizability.

These findings highlight the practical implications for both policymaking and technology. Policymakers can now rely on AI-enhanced crash categorization to design more targeted and effective road safety interventions, especially in resource-limited settings. Future work could integrate real-time traffic feeds and environmental sensors to further refine prediction accuracy. This framework opens pathways for scalable, automated crash analytics applicable to a wide range of low-resource traffic data environments.

References

[1] R. Fisa, M. Musukuma, M. Sampa, P. Musonda, and T. Young, "Effects of interventions for preventing road traffic crashes: an overview of systematic reviews," BMC Public Health, vol. 22, no. 1, p. 513, 2022.

[2] H. Bhuiyan et al., "Crash severity analysis and risk factors identification based on an alternate data source: a case study of developing country," Sci Rep, vol. 12, no. 1, p. 21243, 2022.

[3] A. Sajid Hasan, M. Jalayer, E. Heitmann, and J. Weiss, "Distracted driving crashes: A review on data collection, analysis, and crash prevention methods," Transp Res Rec, vol. 2676, no. 8, pp. 423–434, 2022.

[4] A. S. B. Ali et al., "Factors contributing to road traffic accidents in the Gaza Strip a comprehensive analysis," Sci Rep, vol. 14, no. 1, p. 31198, 2024.

[5] A. Jaber and K. Al-Sahili, "Severity of Pedestrian Crashes in Developing Countries," SAE Int J Transp Saf, vol. 11, no. 3, pp. 307–320, 2023.

[6] F. M. A. Hassouna, S. Abu-Eisheh, and K. Al-Sahili, "Analysis and modeling of road crash trends in Palestine," Arab J Sci Eng, vol. 45, pp. 8515–8527, 2020.

[7] Y. Sarraj, "Developing Road Accidents Recording System in Palestine," 2016.

[8] H. I. Ashqar, T. I. Alhadidi, M. Elhenawy, and S. Jaradat, "Factors affecting crash severity in Roundabouts: A comprehensive analysis in the Jordanian context," Transportation Engineering, vol. 17, p. 100261, 2024, doi: https://doi.org/10.1016/j.treng.2024.100261.

[9] S. R. Vinta, P. Rajarajeswari, M. V. Kumar, and G. S. C. Kumar, "BConvLSTM: a deep learning-based technique for severity prediction of a traffic crash," International Journal of Crashworthiness, vol. 29, no. 6, pp. 1051–1061, 2024.

[10] M. T. Kashifi, M. Al-Turki, and A. W. Sharify, "Deep hybrid learning framework for spatiotemporal crash prediction using big traffic data," International journal of transportation science and technology, vol. 12, no. 3, pp. 793–808, 2023.

[11] A. J. Ghandour, H. Hammoud, and S. Al-Hajj, "Analyzing factors associated with fatal road crashes: a machine learning approach," Int J Environ Res Public Health, vol. 17, no. 11, p. 4111, 2020.

[12] X. Zhang, S. T. Waller, and P. Jiang, "An ensemble machine learning-based modeling framework for analysis of traffic crash frequency," Computer-Aided Civil and Infrastructure Engineering, vol. 35, no. 3, pp. 258–276, 2020.

[13] M. S. Rahman, M. Abdel-Aty, S. Hasan, and Q. Cai, "Applying machine learning approaches to analyze the vulnerable road-users' crashes at statewide traffic analysis zones," J Safety Res, vol. 70, pp. 275–288, 2019.

[14] D. Akin, V. P. Sisiopiku, A. H. Alateah, A. O. Almonbhi, M. M. H. Al-Tholaia, and K. A. A. Al-Sodani, “Identifying Causes of Traffic Crashes Associated with Driver Behavior Using Supervised Machine Learning Methods: Case of Highway 15 in Saudi Arabia,” Sustainability, vol. 14, no. 24, p. 16654, 2022.
[15] B. Al-Mistarehi, A. H. Alomari, R. Imam, and M. Mashaqba, “Using machine learning models to forecast severity level of traffic crashes by R Studio and ArcGIS,” Front Built Environ, vol. 8, p. 860805, 2022.
[16] M. Abdel-Aty and K. Haleem, “Analyzing angle crashes at unsignalized intersections using machine learning techniques,” Accid Anal Prev, vol. 43, no. 1, pp. 461–470, 2011.
[17] H. Mirzahossein, M. Sashurpour, S. M. Hosseinian, and V. N. M. Gilani, “Presentation of machine learning methods to determine the most important factors affecting road traffic accidents on rural roads,” Frontiers of Structural and Civil Engineering, vol. 16, no. 5, pp. 657–666, 2022.
[18] M. M. R. Komol, M. M. Hasan, M. Elhenawy, S. Yasmin, M. Masoud, and A. Rakotonirainy, “Crash severity analysis of vulnerable road users using machine learning,” PLoS One, vol. 16, no. 8, p. e0255828, 2021.
[19] A. Jaber and B. Csonka, “Towards a sustainable and safe future: mapping bike accidents in urbanized context,” Safety, vol. 9, no. 3, p. 60, 2023.
[20] A. Jaber, J. Juhász, and B. Csonka, “An analysis of factors affecting the severity of cycling crashes using binary regression model,” Sustainability, vol. 13, no. 12, p. 6945, 2021.
[21] M. A. Tami, M. Elhenawy, and H. I. Ashqar, “Multimodal Large Language Models for Enhanced Traffic Safety: A Comprehensive Review and Future Trends,” arXiv preprint arXiv:2504.16134, 2025.
[22] S. Jaradat, M. Elhenawy, H. I. Ashqar, A. Paz, and R. Nayak, “Leveraging Deep Learning and Multimodal Large Language Models for Near-Miss Detection Using Crowdsourced Videos,” IEEE Open Journal of the Computer Society, 2025.
[23] S. Jaradat, M. Elhenawy, R. Nayak, A. Paz, H. I. Ashqar, and S. Glaser, “Multimodal Data Fusion for Tabular and Textual Data: Zero-Shot, Few-Shot, and Fine-Tuning of Generative Pre-Trained Transformer Models,” AI, vol. 6, no. 4, p. 72, 2025.
[24] C. Arteaga and J. Park, “A large language model framework to uncover underreporting in traffic crashes,” J Safety Res, vol. 92, pp. 1–13, 2025.
[25] P. Mudgal, B. Arbab, and S. S. Kumar, “CrashEventLLM: Predicting System Crashes with Large Language Models,” in 2024 International Conference on Information Technology and Computing (ICITCOM), IEEE, 2024, pp. 72–76.
[26] S. Jaradat, T. I. Alhadidi, H. I. Ashqar, A. Hossain, and M. Elhenawy, “Investigating patterns of freeway crashes in Jordan: Findings from a text mining approach,” Results in Engineering, vol. 26, p. 104413, 2025.
[27] M. A. Tami, H. I. Ashqar, M. Elhenawy, S. Glaser, and A. Rakotonirainy, “Using Multimodal Large Language Models (MLLMs) for Automated Detection of Traffic Safety-Critical Events,” Vehicles, vol. 6, no. 3, pp. 1571–1590, 2024.
[28] Z. Liu et al., “Testing the limits: Unusual text inputs generation for mobile app crash detection with large language model,” in Proceedings of the IEEE/ACM 46th international conference on software engineering, 2024, pp. 1–12.
[29] M. A. Tami, M. Elhenawy, and H. I. Ashqar, “HazardNet: A Small-Scale Vision Language Model for Real-Time Traffic Safety Detection at Edge Devices,” arXiv preprint arXiv:2502.20572, 2025.
[30] Z. Hu, J. Zhou, K. Huang, and E. Zhang, “A data-driven approach for traffic crash prediction: A case study in Ningbo, China,” International Journal of Intelligent Transportation Systems Research, vol. 20, no. 2, pp. 508–518, 2022.
[31] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W. Woo, “Convolutional LSTM network: A machine learning approach for precipitation nowcasting,” Adv Neural Inf Process Syst, vol. 28, 2015.
[32] G. E. Hinton and S. Roweis, “Stochastic neighbor embedding,” Adv Neural Inf Process Syst, vol. 15, 2002.