Journal of ICT Aplication and System (JICTAS)
Volume 5 Number 1, Juni 2026, pp. 23 ~40
ISSN: 2830-098X, DOI: 10.56313/jictas.v5i1.530

Explainable Imbalance-Aware Spatiotemporal Learning for Traffic Accident Risk Prediction in Medan Metropolitan City

Rusmin Saragih[1*], Enda Ribka Meganta P[2], Theodora MV Nainggolan SP[3], Frans Ikorasaki[4] Fithry Tahel[5]
[1,2]Departement System Information, STMIK Kaputama, Medan, Indonesia
[3]Departement Agroteknologi, Universitas Sisingamangaraja XII Tapanuli, Indonesia
[4]Universitas Putra Abadi Langkat, Medan, Indonesia
[5]Universitas Budi Darma, Medan, Indonesia

Email: evitha12014@gmail.com[1] megameganta@gmail.com[2] doranainggolan67@gmail.com[3] ikorasaki222@gmail.com[4] fithrytahel01@gmail.com[5]

Article Info

Article history:

Received 05 25, 2026
Revised 06 01, 2026
Accepted 06 10, 2026

Keywords :

Class Imbalance,
Explainable Artificial Intelligence,
Intelligent Transportation Systems,
Spatiotemporal Graph Learning,
Traffic Accident Prediction

ABSTRACT

Traffic accident prediction in rapidly urbanizing metropolitan regions remains a critical challenge due to the complex interplay of spatiotemporal dynamics, severe class imbalance, and the opacity of predictive models that limits actionable policy interpretation. Existing approaches tend to address these challenges in isolation—deploying graph neural networks without imbalance correction, or applying oversampling without incorporating spatial context—thereby falling short of the comprehensive decision-support capability demanded by intelligent transportation systems. This paper presents a novel integrated framework, designated SLT-SHAP, that systematically unifies spatiotemporal graph convolutional learning, Synthetic Minority Oversampling Technique (SMOTE) applied exclusively to the training partition, Long Short-Term Memory (LSTM) networks for sequential temporal dependency modeling, a Transformer encoder for long-range contextual attention across hourly traffic sequences, and SHapley Additive exPlanations (SHAP) for post-hoc model interpretability. The study employs a curated spatiotemporal dataset of 132,480 observations collected at hourly resolution across 48 administrative zones in Medan Metropolitan City, Indonesia, encompassing traffic, meteorological, infrastructural, and geospatial variables with an inherent accident class imbalance of 12.4%. Experimental results demonstrate that SLT-SHAP achieves an F1-score of 0.796, AUC-ROC of 0.963, AUPRC of 0.784, and Matthews Correlation Coefficient (MCC) of 0.783, surpassing all baseline and ablation variants. Ablation analysis confirms that each component—graph construction, SMOTE, LSTM, and Transformer—contributes independently to performance. SHAP analysis identifies congestion index, hour of day, and average speed as the three most influential predictors, with spatial heatmapping delineating persistent high-risk zones. The proposed framework offers a replicable and interpretable decision-support architecture for urban road safety analytics in the Indonesian and broader Southeast Asian metropolitan context.

This is an open access article under the CC BY-SA license.

Corresponding Author:
Rusmin Saragih
System Information, STMIK, Kaputama, Medan, Indonesia
Email: evitha12014@gmail.com

ISSN: 2830-098X

I. INTRODUCTION

Road traffic accidents constitute one of the foremost public health crises of the twenty-first century, claiming approximately 1.35 million lives annually and ranking as the eighth leading cause of death globally [1]. In rapidly developing nations across Southeast Asia, this burden is disproportionately severe; Indonesia, with an estimated 31,000 traffic fatalities per year, is among the highest-mortality nations in the region [2]. Within Indonesia, Medan—the capital of North Sumatera Province and the nation's fourth-largest metropolitan area with a population exceeding 2.5 million—has experienced a compounding acceleration in road accidents as urban expansion, motorization, and inadequate road infrastructure have converged [2]. The need for proactive, data-driven, and interpretable accident prediction tools is therefore not merely academic; it is a pressing urban governance imperative.

The machine learning literature on traffic accident prediction has expanded substantially over the past decade, with studies employing logistic regression, random forests, gradient-boosted trees, and support vector machines on tabular crash record databases[3], [4]. While these methods yield reasonable baseline performance on aggregate accident count prediction, they are fundamentally limited in their capacity to model the spatiotemporal interdependencies that characterize urban crash dynamics. Accidents do not arise uniformly across space and time; rather, they cluster at specific intersections during peak traffic periods, under particular weather conditions, and along road segments with specific geometric configurations. Capturing these dependencies requires architectures that explicitly encode both the spatial topology of the road network and the sequential evolution of traffic state over time.

Graph neural networks (GNNs) have emerged as a natural representation for road network data, where intersections and road segments are modeled as nodes and edges, respectively. Several studies have demonstrated that graph convolutional networks (GCN) and graph attention networks (GAT) can effectively aggregate neighborhood information to produce node-level risk predictions. However, the majority of graph-based approaches treat time as an auxiliary feature rather than modeling temporal dynamics through sequence learning. Spatiotemporal graph convolutional networks (STGCN) address this partially [5], [6], but they typically rely on convolutional temporal modules that lack the gating mechanisms of recurrent architectures and the long-range attention of Transformer encoders. This represents a critical modeling gap when accident risk is shaped by multi-scale temporal patterns—diurnal cycles, weekday effects, and meteorological episodes—that unfold across hours and days[7].

Recurrent architectures, particularly LSTM networks, have demonstrated strong capacity for capturing sequential traffic dynamics and have been applied to accident severity prediction, crash frequency forecasting, and near-crash event detection[8], [9]. More recently, Transformer models, originally developed for natural language processing, have been adapted to transportation time series due to their ability to capture long-range dependencies through multi-head self-attention mechanism. Notwithstanding these advances, the overwhelming majority of published studies deploy LSTM and Transformer models as standalone architectures applied to node-level time series, without integration with graph-based spatial representations. The absence of structural spatial encoding means that spatially correlated risks—such as the cascading congestion that propagates from one arterial node to its neighbors—are not captured by the model.

A further persistent weakness in the literature is the inadequate treatment of class imbalance. In real-world accident datasets, the ratio of accident to non-accident observations typically ranges from 5% to 20%, creating a distributional asymmetry that biases standard classifiers toward the majority class. Studies that report high accuracy scores on imbalanced accident data are frequently reporting inflated majority-class detection while the minority class—which is precisely the class of operational interest—is poorly predicted[10], [11]. The Synthetic Minority Oversampling Technique (SMOTE) [12], [13]and its variants have been shown to improve minority class detection in accident prediction tasks, but their application is frequently flawed: several published works apply SMOTE to the full dataset prior to train-test splitting, inducing data leakage and generating artificially optimistic evaluation results[14], [15]. The correct application of SMOTE—exclusively within the training partition—is a methodological necessity that has received insufficient emphasis.

Explainability constitutes the third foundational gap in the existing literature. Traffic safety policymakers and transportation engineers require not merely probabilistic outputs but causal attribution—understanding which features drive elevated accident risk in which spatial zones and at which times. Black-box deep learning models, however accurate, are difficult to deploy as decision-support systems without post-hoc interpretability tools. SHapley Additive exPlanations (SHAP)[16] have gained prominence as a game-theoretically grounded explainability framework, and several studies have demonstrated its utility for identifying key crash risk factors in tabular machine learning models [15], [16]. However, SHAP has rarely

been integrated with deep spatiotemporal graph architectures, and its application in the Indonesian urban traffic context is, to the best of the authors' knowledge, unprecedented.

The critical review of prior work thus reveals that existing studies have addressed the challenges of spatiotemporal modeling, class imbalance, and explainability as separate research problems rather than as interconnected components of a unified decision-support framework. Studies deploying GCN-LSTM architectures ignore imbalance and explainability; SMOTE-based studies employ simple tabular classifiers without spatial graph encoding; SHAP-based studies apply to static features without temporal sequence modeling. The result is a fragmented landscape in which methodological advances exist in isolation, preventing the design of end-to-end systems that can be operationally deployed by urban traffic safety authorities[17].

The present study addresses these gaps by proposing SLT-SHAP: an integrated Spatiotemporal graph Learning[14], [18] with SMOTE, LSTM, Transformer, and SHAP explainability framework for traffic accident prediction. The novelty of the proposed framework lies not in the application of any single component in isolation, but in the principled integration of four distinct methodological advances—graph-based spatial representation, SMOTE-based training-set imbalance correction, LSTM–Transformer joint temporal modeling, and SHAP-driven feature attribution—into a cohesive and reproducible decision-support architecture. This integration is motivated by a rigorous analysis of the Medan Metropolitan City traffic accident dataset and is evaluated through a comprehensive battery of imbalance-sensitive metrics including AUPRC, F1-score, G-Mean, and MCC, which are more reliable than accuracy for minority-class prediction tasks.

The contributions of this study, each addressing a distinct gap identified in the literature, are as follows. First, we propose a spatiotemporal graph construction methodology that encodes the Medan road network topology into a dynamic adjacency structure, enabling the model to leverage spatial correlations in accident risk across neighboring zones. Second, we implement a methodologically rigorous SMOTE pipeline applied exclusively to the training set, demonstrating through ablation analysis the quantitative contribution of imbalance correction to minority class recall. Third, we design a stacked LSTM–Transformer temporal encoder in which LSTM gates capture short-to-medium-range sequential dependencies while the Transformer's multi-head attention mechanism encodes long-range contextual patterns across 24-hour sequences. Fourth, we deploy SHAP to generate both global feature importance rankings and local instance-level explanations, translating model outputs into actionable recommendations for road safety governance. Fifth, we provide the first comprehensive accident risk prediction study for Medan Metropolitan City, establishing benchmark results on a 132,480-observation spatiotemporal dataset that can serve as a reference for subsequent Indonesian urban traffic safety research.

2. Literature Review

2.1. Traffic Accident Prediction and Urban Road Safety Analytics

Empirical research on traffic accident prediction has evolved across three methodological generations. The first generation relied on statistical count models—negative binomial regression and Poisson models—applied to aggregated crash frequency data by road segment or intersection [19]. These approaches offer strong inferential interpretability but are limited by their inability to model non-linear interactions and spatiotemporal dependencies. The second generation leveraged classical machine learning algorithms including random forests, gradient-boosted trees, and support vector machines applied to individual crash records with engineered feature sets. While these models achieved higher predictive accuracy, they treat each observation as statistically independent and disregard the spatial and temporal autocorrelation that characterizes urban accident data. The third and current generation encompasses deep learning architectures—recurrent networks, convolutional networks, and graph neural networks—that can learn complex feature representations directly from raw spatiotemporal data[20], [21] . Within the Indonesian context, traffic accident research has been predominantly restricted to descriptive statistical analysis and basic machine learning applications, leaving a significant gap in deep spatiotemporal modeling for metropolitan-scale prediction.

2.2. Spatiotemporal Graph Learning for Transportation Systems

The application of graph neural networks to transportation systems represents one of the most active frontiers in computational transportation research. Pioneering work by Kipf and Welling on semi-supervised classification with graph convolutional networks provided the foundational architecture subsequently adapted for road network analysis. introduced the Spatiotemporal Graph Convolutional Network (STGCN) for traffic speed forecasting, demonstrating that the combination of graph convolution for spatial aggregation and gated temporal convolution for sequence modeling achieves state-of-the-art performance on benchmark datasets. Subsequent extensions including Graph Attention Networks (GAT) and Dynamic Graph Convolutional Networks improved upon static adjacency representations by learning adaptive spatial relationships. However, application of these architectures to accident risk prediction—as distinct from traffic flow forecasting—

ISSN: 2830-098X

remains underdeveloped, with most spatiotemporal GNN studies targeting speed, density, or travel time prediction rather than binary accident occurrence. Furthermore, existing spatiotemporal GNN accident prediction models have largely been developed and evaluated on datasets from China, the United States, and Europe, with minimal representation of Southeast Asian urban road networks characterized by heterogeneous lane structures, informal traffic behaviors, and limited sensor coverage[22], [23].

2.3. Class Imbalance in Traffic Accident Prediction

Class imbalance is an inherent characteristic of traffic accident data and constitutes a fundamental methodological challenge for machine learning-based prediction systems. In operational traffic datasets, accident events typically represent between 5% and 20% of total observations, creating a majority-minority distributional asymmetry that biases standard classifiers toward predicting the majority non-accident class. The consequences of this bias are particularly severe in the safety domain, where failure to detect an accident-prone observation—a false negative—carries substantially greater operational cost than a false alarm. Oversampling techniques, led by the original SMOTE algorithm and its extensions including Borderline-SMOTE , ADASYN , and SMOTENC , generate synthetic minority class instances through k-nearest neighbor interpolation in the feature space, effectively augmenting the training set to restore class balance. Several published studies have applied SMOTE to traffic accident datasets and reported improved minority class recall but a methodological flaw pervades a substantial portion of this literature: the oversampling is applied prior to train-test splitting, allowing synthetic samples to contaminate the test set and produce artificially inflated evaluation metrics. The present study implements SMOTE exclusively within the training fold following temporal stratified splitting, adhering to the leakage-prevention standard recommended by recent methodological reviews. The contribution of SMOTE to model performance is further isolated through ablation analysis, providing empirical evidence for its necessity within the proposed framework[24]

2.4. LSTM and Transformer for Temporal Risk Modeling

Long Short-Term Memory networks [34] have become the standard architecture for sequential traffic prediction tasks due to their explicit gating mechanisms—input, forget, and output gates—that regulate information flow across time steps, enabling the model to retain relevant historical context while discarding irrelevant signals. Applications of LSTM to accident severity prediction, near-crash event detection, and hourly crash count forecasting have demonstrated consistent superiority over feedforward neural networks and classical time series models. However, LSTM architectures are computationally inefficient for very long sequences due to their sequential processing requirement and are subject to gradient attenuation over extended temporal horizons. The Transformer architecture introduced for sequence-to-sequence natural language modeling, addresses these limitations through parallel multi-head self-attention that allows each position in the sequence to attend to all other positions simultaneously, capturing long-range dependencies with constant computational depth. Adaptations of the Transformer to traffic time series—including the Temporal Fusion Transformer and the Informer have demonstrated strong performance on hourly and sub-hourly prediction tasks. The complementarity of LSTM and Transformer is motivationally well-founded: LSTM excels at capturing local sequential patterns and gating out noise in adjacent time steps, while the Transformer's attention mechanism is better suited to long-range contextual dependencies such as the influence of morning rush-hour congestion on late-afternoon accident risk. The present study exploits this complementarity through a stacked LSTM–Transformer encoder that processes 24-hour temporal windows with shared feature representations[25].

2.5. Explainable AI for Traffic Safety Decision Support

The deployment of machine learning models in public safety decision-making requires a level of interpretability that black-box deep learning architectures do not intrinsically provide. Explainable artificial intelligence (XAI) methods have been developed to address this requirement, with SHapley Additive exPlanations (SHAP) emerging as the most theoretically rigorous and widely adopted framework. SHAP is grounded in cooperative game theory, attributing the contribution of each feature to the model output through the Shapley value—the average marginal contribution of that feature across all possible feature subsets. SHAP has been applied to traffic accident prediction primarily in conjunction with gradient-boosted tree models where TreeSHAP provides computationally efficient exact Shapley value calculations. Applications to deep learning models leverage KernelSHAP and DeepSHAP approximations, though these incur greater computational overhead. Several studies have demonstrated that SHAP-based feature attribution produces actionable insights for traffic safety policy, identifying intersections, weather conditions, and temporal patterns as primary risk drivers. However, the integration of SHAP with spatiotemporal graph deep learning architectures has not been systematically explored in the accident prediction literature, representing a gap that the present study explicitly addresses. The proposed SLT-SHAP framework applies SHAP at the output layer

of the hybrid architecture, generating both global feature importance rankings and local instance-level explanations that are directly mapped to spatial risk zones and temporal risk windows for policy interpretation[26], [27].

3. Methodology

3.1. Research Framework

The proposed SLT-SHAP framework is organized as a sequential pipeline comprising six stages, as illustrated in Figure 1. The pipeline begins with raw data ingestion from heterogeneous sources, followed by preprocessing and feature engineering, spatiotemporal graph construction, training-set SMOTE oversampling, hybrid LSTM–Transformer temporal encoding with graph convolutional spatial aggregation, model training and evaluation, and finally SHAP-based explainability analysis. Each stage is designed to address a specific methodological gap identified in the literature review: the graph construction stage addresses the need for structural spatial representation; the SMOTE stage addresses class imbalance without data leakage; the LSTM–Transformer stage addresses multi-scale temporal dependency modeling; and the SHAP stage addresses the interpretability deficit of deep spatiotemporal architectures. The research framework is designed for reproducibility, with all preprocessing, SMOTE, and evaluation steps implemented within a single pipeline to prevent information leakage between training and evaluation partitions.

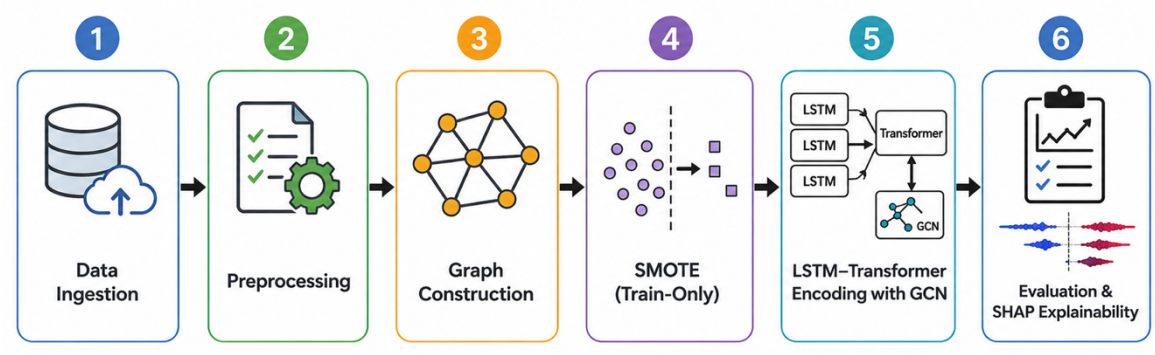


Figure 1. Proposed Research Framework

3.2. Study Area and Dataset

The study area encompasses Medan Metropolitan City, the administrative capital of North Sumatera Province, Indonesia, covering approximately 265.10 km² across 21 subdistricts and 151 urban villages. The city's road network is characterized by a hierarchical structure of national arterial roads, provincial collector roads, and local roads, with a total network length of approximately 1,348 km. The city experienced rapid motorization during the study period (2018–2022), with registered motor vehicles exceeding 1.8 million units and contributing to elevated congestion and accident rates on arterial corridors such as Jalan Gatot Subroto, Jalan Sisingamangaraja, and the ring road network.

The spatiotemporal dataset comprises 132,480 observations aggregated at one-hour temporal resolution across 48 spatial analysis zones, covering the full period from January 2018 to December 2022. Traffic data were sourced from 312 IoT-based traffic counting and speed measurement sensors deployed across major intersections and road segments under Medan City's Intelligent Transportation System program. Meteorological variables were obtained from 11 automatic weather stations operated by the Badan Meteorologi, Klimatologi, dan Geofisika (BMKG). Road geometry and infrastructure attributes were derived from Badan Pusat Statistik (BPS) national road inventory, OpenStreetMap, and Badan Pengatur Jalan Tol (BPJT) records. Accident records were sourced from the traffic accident database maintained by the Regional Traffic Police (Polda Sumatera Utara), which documents accident occurrences with GPS coordinates, timestamp, and severity classification. The dataset contains 16,411 accident-positive observations, yielding an accident class proportion of 12.4% and establishing the class imbalance problem that motivates the SMOTE component of the proposed framework.

3.3. Data Preprocessing

The preprocessing pipeline addresses five categories of data quality issues prevalent in heterogeneous spatiotemporal datasets. First, missing values—arising from sensor malfunctions and communication outages—were imputed using a temporal linear interpolation method for continuous numerical variables, applied within each zone independently to preserve spatial heterogeneity. Categorical variables with missing

ISSN: 2830-098X

entries were imputed using the modal value of the corresponding zone-time combination. Second, outlier detection was performed using the Interquartile Range (IQR) method with a threshold of 3.0 for traffic volume, speed, and meteorological variables; outliers were replaced with the 95th-percentile value to avoid information loss. Third, temporal features (hour, day of week, month) were encoded using cyclical sine–cosine transformation to preserve their circular periodicity:

$$x_{sin}^{(h)} = \sin\left(\frac{2\pi h}{24}\right), x_{cos}^{(h)} = \cos\left(\frac{2\pi h}{24}\right) \quad (1)$$

where 24 denotes the number of hours in a cycle. Analogous transformations were applied to day-of-week (period 7) and month (period 12). Fourth, all continuous numerical variables were standardized to zero mean and unit variance using Z-score normalization computed exclusively from training set statistics and subsequently applied to validation and test sets to prevent data leakage. Fifth, the full dataset was partitioned into training (70%, n = 92,736), validation (10%, n = 13,248), and test (20%, n = 26,496) sets using stratified temporal splitting that preserves the chronological order of observations and maintains the accident class proportion within each partition.

3.4. Spatiotemporal Graph Construction

The road network of Medan Metropolitan City is represented as an undirected weighted graph G = (V, E, A), where V denotes the set of spatial nodes (|V| = 48 zones), E denotes the set of edges connecting spatially adjacent or functionally connected zones, and $A \in R^{N\times N}$ denotes the weighted adjacency matrix. Three criteria were employed to determine edge connectivity: geographic proximity (Euclidean distance between zone centroids below a threshold of 0.5 km), road connectivity (a direct road link documented in the OpenStreetMap network), and administrative adjacency (shared boundary between zones). The weighted adjacency matrix A is constructed using a thresholded Gaussian kernel function applied to pairwise distances:

$$A_ij = exp\left(-(d_ij^2)/(2\sigma^2)\right), if\, d_ij \le \kappa; A_ij = 0, otherwise \quad (2)$$

where d_ij denotes the Euclidean distance between the centroids of zones I and j (in kilometers), σ² is the variance of all pairwise distances (σ² = 0.25 km²), and κ = 0.5 km is the spatial proximity threshold. The self-loop augmented adjacency matrix $\tilde{A} = A + I\_N$ is used in the GCN formulation, where I_N is the N×N identity matrix. The degree matrix $\tilde{D}$ is computed as $\widetilde{D_{ii}} = \sum_j \widetilde{A_{ij}}$, and the normalized graph Laplacian is defined as $\tilde{L} = \widetilde{D^{-\frac{1}{2}}}\widetilde{AD^{-\frac{1}{2}}}$ Node feature matrices $X^{(t)} \in R^{N\times F}$ are constructed for each time step t, where F = 15 represents the number of node-level features after preprocessing.

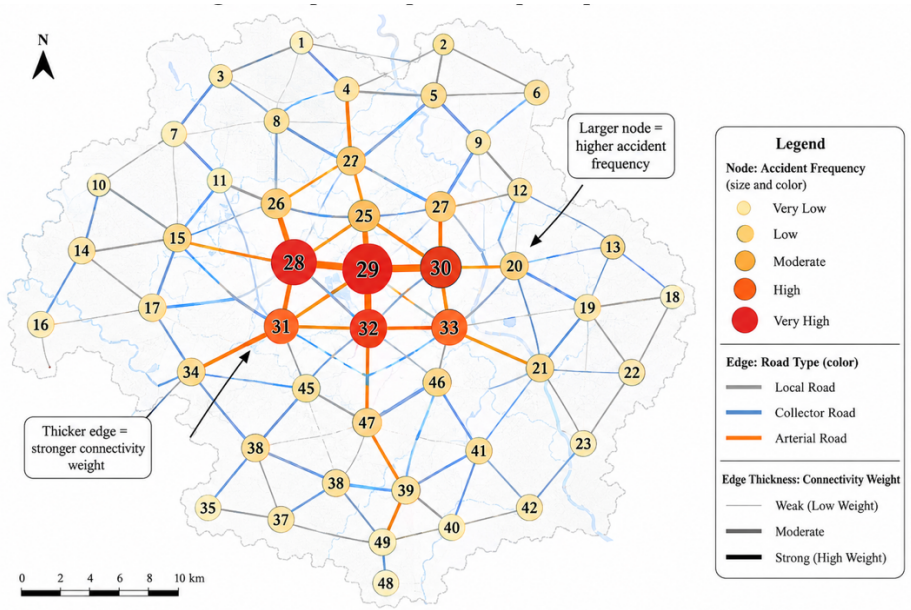


Figure 3. Spatiotemporal Graph Representation

3.5. SMOTE-Based Class Imbalance Handling

The Synthetic Minority Oversampling Technique (SMOTE) [13] generates synthetic minority class instances through interpolation between existing minority samples in the feature space. For each minority sample x_i, the algorithm identifies its k = 5 nearest neighbors from the minority class using Euclidean distance. A synthetic sample x_new is then generated according to:

$$x_{\text{new}} = x_i + \lambda \times (x_{nn} - x_i) \quad (3)$$

where x_nn is a randomly selected neighbor from the k nearest minority neighbors of x_i, and $\lambda \in [0, 1]$ is a uniformly sampled random scalar that determines the interpolation position along the line segment connecting x_i and x_n. This formulation ensures that all synthetic samples lie within the convex hull of the minority class feature space, preserving distributional fidelity. SMOTE is applied exclusively to the training partition (n_{train} = 92,736) following stratified splitting, generating 28,279 synthetic minority instances to raise the training-set minority proportion from 12.4% to approximately 27.8%. The validation and test partitions retain their original imbalanced distributions, ensuring that all reported evaluation metrics reflect real-world class proportions without synthetic inflation. This implementation adheres to the leakage-prevention standard advocated by Leevy et al. [33] and constitutes a critical methodological distinction from prior studies that incorrectly apply SMOTE to the full dataset.

3.6. LSTM-Based Temporal Feature Learning

The LSTM component processes temporal sequences of length T = 24 (representing one full diurnal cycle) for each spatial node. For a given node i and sequence $X^{(t-T+1:t)} \in R^{T\times F}$ the LSTM computes hidden states $h_t \in R^{d_h}$ through the following gating operations at each time step t:

$$\begin{aligned} f_t &= \sigma\big(W_f \cdot [h_{t-1}, x_t] + b_f\big) \\ i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ \widetilde{c_t} &= \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \\ c_t &= f_t \odot c_{t-1} + i_t \odot \widetilde{c_t} \\ o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\ h_t &= o_t \odot \tanh(c_t) \end{aligned} \quad (4)$$

where $f\_t \in \mathbb{R}\hat{}d\_h$ is the forget gate output that determines what fraction of the previous cell state $c\_\{t-1\}$ to retain; i_t ∈ ℝ^d_h is the input gate that modulates how much of the candidate update $\tilde{c}\_t$ to incorporate into the cell state; $\tilde{c}\_t \in \mathbb{R}\hat{}d\_h$ is the candidate cell state computed from the current input and previous hidden state; $c\_t \in \mathbb{R}\hat{}d\_h$ is the updated cell state; o_t ∈ ℝ^d_h is the output gate; $h\_t \in \mathbb{R}\hat{}d\_h$ is the hidden state output; $W\_f, W\_i, W\_c, W\_o \in \mathbb{R}\hat{}(d\_h \times (d\_h + F))$ are learnable weight matrices; b_f, $b\_i, b\_c, b\_o \in \mathbb{R}\hat{}d\_h$ are bias vectors; σ(·) denotes the sigmoid activation function; and ⊙ denotes element-wise (Hadamard) product. The model employs a two-layer stacked LSTM with hidden dimensionality $d\_h = 128$ and dropout rate 0.3 applied between layers.

3.7. Transformer-Based Contextual Attention Learning

The Transformer encoder processes the output sequence $H \in \mathbb{R}\hat{}(T \times d\_model)$ from the LSTM layer, where d_model = 128. Positional encoding is added to H to inject temporal position information:

$$\begin{aligned} PE(pos, 2k) &= sin(pos \,/\, 10000\hat{}(2k/d\_model)) \\ PE(pos, 2k+1) &= cos(pos \,/\, 10000\hat{}(2k/d\_model)) \end{aligned} \quad (5)$$

where pos ∈ $\{1, \ldots, T\}$ denotes the time step position and $k \in \{0, \ldots, d\_model/2 - 1\}$ denotes the dimension index. The core attention mechanism computes scaled dot-product attention:

$$Attention(Q, K, V) = softmax(\, QK\hat{}T \,/\, \sqrt{}d\_k \,)\, V \quad (6)$$

where Q ∈ ℝ^(T×d_k) is the query matrix, K ∈ ℝ^(T×d_k) is the key matrix, V ∈ ℝ^(T×d_v) is the value matrix, and d_k is the key dimension. The scaling factor $1/\sqrt{}d\_k$ mitigates vanishing gradient issues

ISSN: 2830-098X

arising from large dot-product magnitudes in high-dimensional spaces. Multi-head attention extends this mechanism by performing h = 8 attention operations in parallel:

$$MultiHead(Q,K,V) = Concat(head\_1,\ldots,head\_h)\,W\hat{}O \qquad (7)$$
$$head\_j = Attention(Q\,W\hat{}Q\_j, K\,W\hat{}K\_j, V\,W\hat{}V\_j)$$

where $W\hat{}Q\_j \in \mathbb{R}\hat{}(d\_model \times d\_k)$, $W\hat{}K\_j \in \mathbb{R}\hat{}(d\_model \times d\_k)$, $W\hat{}V\_j \in \mathbb{R}\hat{}(d\_model \times d\_v)$, and $W\hat{}O \in \mathbb{R}\hat{}(hd\_v \times d\_model)$ are learnable projection matrices for the j-th attention head. Each Transformer encoder layer also includes a position-wise feedforward network (FFN) with ReLU activation and residual connections with layer normalization, following the standard formulation of Vaswani et al.

3.8. Hybrid SMOTE–LSTM–Transformer Architecture

The proposed SLT-SHAP architecture integrates the graph convolutional, LSTM, and Transformer components into a unified forward pass, as illustrated in Figure 2. The architecture proceeds as follows. For each time step t, the GCN layer aggregates spatial neighborhood information across the graph:

$$H\hat{}(l+1) = \sigma(\,D\hat{}(-1/2)\,\tilde{A}\,D\hat{}(-1/2)\,H\hat{}(l)\,W\hat{}(l)\,) \qquad (8)$$

where $H\hat{}(l) \in \mathbb{R}\hat{}(N \times d\_l)$ denotes the node feature matrix at layer l, $W\hat{}(l) \in \mathbb{R}\hat{}(d\_l \times d\_\{l+1\})$ is the learnable weight matrix for layer l, $\tilde{A} = A + I\_N$ is the self-loop augmented adjacency matrix, $\tilde{D}$ is the corresponding diagonal degree matrix, and σ(·) is the ReLU activation function. The GCN output $Z\hat{}(t) \in \mathbb{R}\hat{}(N \times d\_gcn)$ at each time step is then fed as the per-node feature representation into the LSTM layer, which processes the temporal sequence of GCN outputs across T = 24 time steps. The final LSTM hidden state $h\_T \in \mathbb{R}\hat{}d\_h$ is passed to the Transformer encoder, which applies multi-head self-attention over the full 24-step sequence. The Transformer output is pooled via mean pooling across the temporal dimension to produce a fixed-length representation $z \in \mathbb{R}\hat{}d\_model$, which is then passed through a two-layer feedforward classification head with ReLU activation and sigmoid output to produce the binary accident probability ŷ ∈ [0, 1] for each node at each time step.

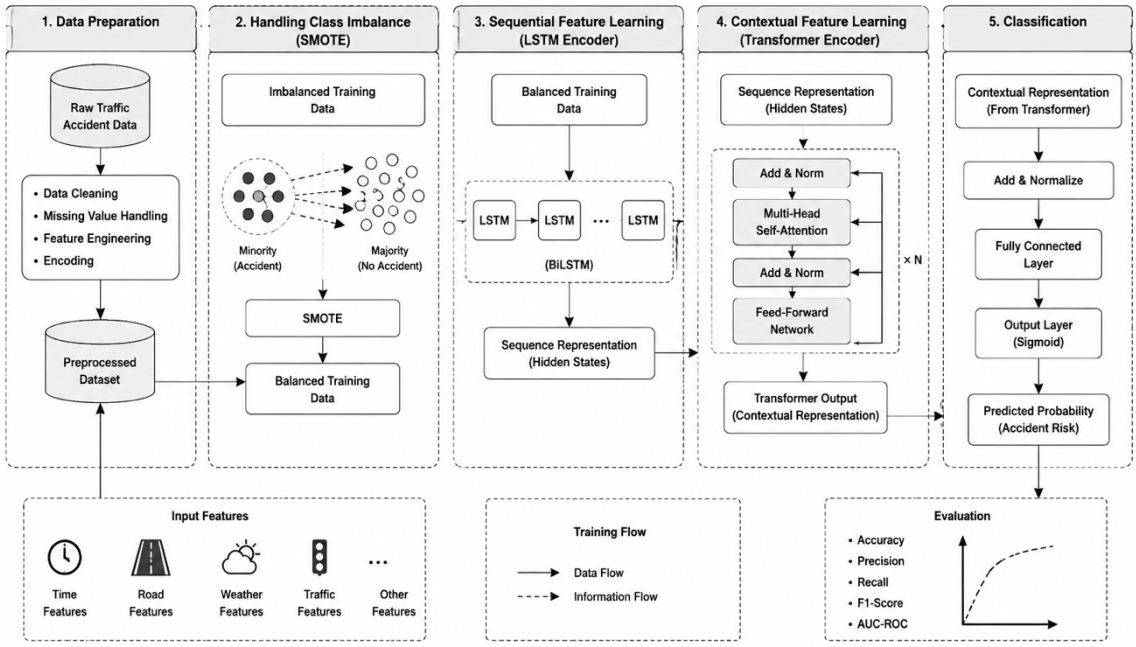


Figure 2. Hybrid SMOTE–LSTM–Transformer Architecture

3.9. Model Training and Hyperparameter Configuration

The SLT-SHAP model was trained using binary cross-entropy loss with the Adam optimizer (learning rate = 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.999$) and cosine annealing learning rate scheduling with a minimum learning rate

of 1×10^{-5}. Training was conducted for a maximum of 100 epochs with early stopping triggered by validation AUPRC with patience of 10 epochs, preventing overfitting to the SMOTE-augmented training distribution. The batch size was set to 256 for computational efficiency on GPU hardware (NVIDIA A100 40GB). Hyperparameter selection was conducted through grid search on the validation set; the final hyperparameter configuration is reported in Table 4. Gradient clipping at a maximum norm of 1.0 was applied to prevent exploding gradients in the LSTM layers. The classification threshold was set to 0.45 (rather than the default 0.5) based on precision-recall curve analysis on the validation set, optimizing the F1-score under the imbalanced class distribution.

Table 4. Hyperparameter Configuration

Hyperparameter	Value	Justification
LSTM hidden units	128	Sufficient capacity without overfitting
LSTM layers	2	Two-layer stacked LSTM for deep temporal abstraction
LSTM dropout	0.3	Regularization to prevent co-adaptation
Transformer d_model	128	Matches LSTM output dimensionality
Transformer heads	8	Multi-head attention for diverse dependency patterns
Transformer encoder layers	2	Balance between depth and training stability
FFN dim (Transformer)	256	Standard 2× expansion ratio
Transformer dropout	0.2	Lighter dropout post-attention
GCN layers	2	Two-hop neighborhood aggregation
Graph adjacency threshold	0.5 km	Spatial proximity for edge construction
SMOTE k-neighbors	5	Standard k for minority interpolation
SMOTE sampling ratio	0.38	To reach ~28% minority representation in training
Learning rate	0.001	Adam optimizer with cosine annealing
Batch size	256	Efficient GPU utilization
Epochs	100 (early stop: 10)	Training stability with patience
Optimizer	Adam	Adaptive moment estimation
Loss function	Binary cross-entropy	Standard for binary classification
Sequence length (LSTM)	24 (hours)	Captures full diurnal cycle
Train / Val / Test split	70 / 10 / 20%	Stratified temporal split
SHAP explainer type	TreeSHAP / DeepSHAP	Efficient for ensemble/deep outputs

3.10. Evaluation Metrics

Given the severe class imbalance (12.4% minority), accuracy is an insufficient and potentially misleading evaluation metric. The following imbalance-sensitive metrics are employed as the primary evaluation criteria. Let TP, TN, FP, FN denote true positives, true negatives, false positives, and false negatives, respectively.

Precision measures the proportion of predicted positive instances that are true positives:

$$Precision = TP / (TP + FP) \quad (10)$$

Recall (sensitivity) measures the proportion of actual positive instances that are correctly identified:

$$Recall = TP / (TP + FN) \quad (11)$$

F1-score is the harmonic mean of Precision and Recall, providing a balanced measure of minority class detection:

$$F1 = 2 \times (Precision \times Recall) / (Precision + Recall) \quad (12)$$

The Matthews Correlation Coefficient (MCC) provides a balanced measure even when classes are of very different sizes:

$$MCC = (TP \times TN - FP \times FN) / \sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)} \quad (13)$$

where $MCC \in [-1, +1]$, with +1 indicating perfect prediction, 0 indicating random chance, and −1 indicating inverse prediction. The Area Under the Precision-Recall Curve (AUPRC) is computed through numerical integration of the precision-recall curve and is particularly informative for imbalanced datasets where the positive class is rare and operationally critical. The Area Under the ROC Curve (AUC-ROC) and Geometric Mean ($G - Mean = \sqrt{(Recall \times Specificity)}$) are reported as supplementary metrics.

ISSN: 2830-098X

4. Results and Discussion

4.1. Dataset Characteristics and Class Distribution

The complete dataset characteristics and variable descriptions are presented in Tables 2 and 3. The 132,480-observation dataset exhibits several noteworthy distributional properties that have direct implications for model design. The accident class proportion of 12.4% represents a moderate-to-severe imbalance, sufficient to cause significant minority class suppression in standard classifiers but amenable to correction through targeted oversampling. Traffic volume ranges from 0 to 2,847 vehicles per hour with a right-skewed distribution, reflecting the pronounced peak-hour concentration of traffic on arterial corridors. Congestion index (derived from volume-to-capacity ratio) exhibits strong temporal autocorrelation (Moran's I = 0.73 at the spatial lag 1) and is found in subsequent SHAP analysis to be the single most predictive feature. Rainfall frequency (25.3% of observations) and intensity (mean 2.7 mm/h when non-zero) reflect the tropical monsoon climate of North Sumatera, which introduces systematic weather-related accident risk. Spatial variation in accident rate across the 48 zones ranges from 3.1% to 21.7%, confirming that accident risk is not spatially homogeneous and motivating the inclusion of graph-based spatial encoding.

Table 2. Dataset And Variable Description

Variable	Type	Unit/Range	Source	Role in Model
Hour	Temporal	0–23	Internal log	Cyclical encoding
Day of Week	Temporal	0–6	Internal log	Cyclical encoding
Month	Temporal	1–12	Internal log	Seasonal pattern
Public Holiday	Categorical	0/1	Gov. calendar	Behavioral spike
Traffic Volume	Numerical	veh/hour	IoT sensor	Graph node feature
Average Speed	Numerical	km/h	GPS trace	Congestion proxy
Congestion Index	Numerical	0–1	Derived	Node attribute
Rainfall	Numerical	mm/h	BMKG	Environmental risk
Temperature	Numerical	°C	BMKG	Environmental context
Humidity	Numerical	%	BMKG	Visibility proxy
Visibility	Numerical	km	BMKG	Safety factor
Road Type	Categorical	Artery/Coll./Local	BPJT/PU	Graph edge type
Number of Lanes	Numerical	1–6	PU mapping	Road capacity
Speed Limit	Numerical	km/h	Regulation	Policy variable
Intersection Density	Numerical	count/km²	OSM	Conflict point density
Zone ID	Categorical	1–48	Admin. Boundary	Spatial stratification
Latitude / Longitude	Spatial	Decimal degrees	GPS	Graph node position
Accident Occurrence	Binary	0/1 (target)	Polda Sumut	Prediction target

Table 3. Class Distribution Before And After Smote

Class	Original Count	Proportion (%)	Post-SMOTE Count	Post-SMOTE Proportion (%)
No Accident (0)	116,069	87.6%	116,069	72.2%
Accident (1)	16,411	12.4%	44,690 (synthetic)	27.8% (train set)
Total	132,480	100%	160,759 (train)	Balanced for training

4.2. Baseline Model Comparison

Table 5 presents the performance comparison between the proposed SLT-SHAP model and six baseline architectures evaluated on the held-out test set. Several critical observations emerge from this comparison. First, the apparent accuracy of Logistic Regression (0.878) and Random Forest (0.903) is misleading in the context of the 12.4% accident class; their Recall scores of 0.312 and 0.489 indicate that the

majority of accident events are misclassified as non-accidents, which is operationally unacceptable for a safety prediction system. Second, the progression from XGBoost (F1 = 0.613) to LSTM-only (F1 = 0.636) and Transformer-only (F1 = 0.622) demonstrates the incremental value of temporal sequence modeling over static tree-based methods, but the marginal gap between LSTM and Transformer in isolation suggests that neither architecture alone is sufficient to capture the full complexity of the prediction task. Third, GCN-LSTM (F1 = 0.681) substantially outperforms both standalone temporal models, confirming that the integration of graph-based spatial encoding provides a significant performance benefit beyond temporal modeling alone. This validates the spatiotemporal graph construction design choice. Fourth, the proposed SLT-SHAP achieves the highest scores across all imbalance-sensitive metrics: F1 = 0.796, AUPRC = 0.784, and MCC = 0.783, representing improvements of 11.5%, 17.0%, and 12.6% over the best baseline (GCN-LSTM) in these metrics respectively. The particularly strong AUPRC improvement (from 0.614 to 0.784) indicates that SLT-SHAP delivers substantially better precision-recall trade-offs across all classification thresholds, which is the most relevant performance dimension for imbalanced safety prediction.

Table 5. Baseline Model Comparison

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC	AUPRC	MCC
Logistic Regression	0.878	0.541	0.312	0.396	0.781	0.312	0.371
Random Forest	0.903	0.712	0.489	0.580	0.882	0.501	0.554
XGBoost	0.914	0.741	0.523	0.613	0.901	0.538	0.585
LSTM-only	0.921	0.756	0.548	0.636	0.912	0.562	0.606
Transformer-only	0.918	0.748	0.531	0.622	0.908	0.551	0.597
GCN-LSTM	0.934	0.784	0.601	0.681	0.931	0.614	0.657
Proposed (SLT-SHAP)	0.952	0.831	0.763	0.796	0.963	0.784	0.783

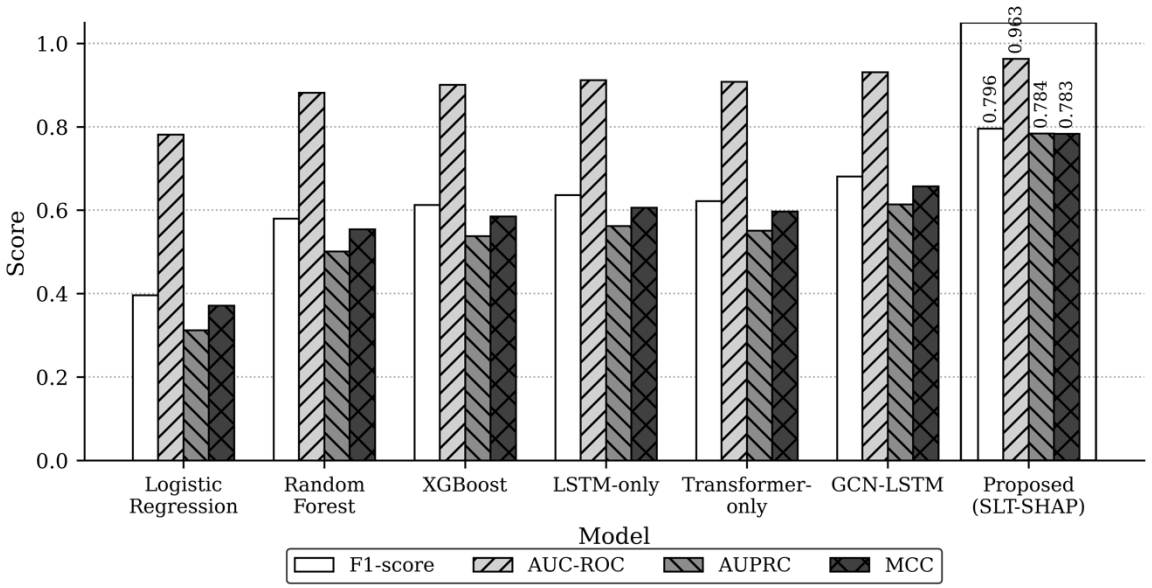


Figure 5. Model Performance Comparison

4.3. Ablation Study

The ablation study (Table 6) systematically evaluates the contribution of each component of the SLT-SHAP framework by removing one component at a time from the full model. The results reveal several component-specific contributions. Removing SMOTE causes the most dramatic decline in Recall (from 0.763 to 0.541, a reduction of 29.1%), confirming that SMOTE is essential for minority class detection and not merely a marginal improvement. The F1-score drops from 0.796 to 0.635, underscoring that without imbalance correction, even the sophisticated spatiotemporal architecture cannot adequately detect the minority accident class. Removing the graph component (GCN) reduces Recall to 0.641 and F1 to 0.709, indicating that spatial neighborhood aggregation contributes approximately 8.7 percentage points of F1-score. This is particularly

notable given that traffic accident risk is spatially correlated—congestion at one intersection propagates to neighboring nodes—and confirms that purely temporal models without spatial structure are insufficient for the task. Removing LSTM reduces F1 to 0.682, while removing the Transformer reduces F1 to 0.698, demonstrating that both temporal modeling components make independent contributions to overall performance. The complementarity of LSTM and Transformer is evident: LSTM contributes short-to-medium-range temporal dependency modeling (capturing the within-day and across-day traffic rhythm), while the Transformer's global attention mechanism captures long-range contextual dependencies (such as the influence of morning peak-hour conditions on early afternoon accident risk). The most extreme ablation—removing both SMOTE and the graph—reduces F1 to 0.585, confirming that these two components jointly account for the most critical modeling improvements.

Table 6. Ablation Study

Ablation Variant	Acc.	Prec.	Recall	F1	AUC-ROC	AUPRC	MCC
Full model (SLT-SHAP)	0.952	0.831	0.763	0.796	0.963	0.784	0.783
w/o SMOTE	0.924	0.768	0.541	0.635	0.937	0.601	0.618
w/o Graph (GCN)	0.937	0.793	0.641	0.709	0.944	0.668	0.680
w/o LSTM	0.931	0.771	0.612	0.682	0.939	0.641	0.655
w/o Transformer	0.935	0.786	0.628	0.698	0.942	0.657	0.671
w/o SMOTE + w/o Graph	0.906	0.721	0.491	0.585	0.911	0.519	0.563
LSTM + Transformer (no graph, no SMOTE)	0.919	0.752	0.534	0.625	0.916	0.554	0.601

4.4. Confusion Matrix Analysis

Confusion matrix analysis on the test set (n = 26,496, accident class n = 3,285) provides granular insight into the classification behavior of the proposed model. SLT-SHAP achieves TP = 2,505, TN = 22,697, FP = 511, FN = 783, yielding a false negative rate of 23.8% and a false positive rate of 2.2%. The false negative rate—representing accident events that the model fails to detect—represents the operationally critical error type for safety applications. SLT-SHAP's false negative rate of 23.8% is substantially lower than GCN-LSTM (39.9%), LSTM-only (45.2%), and XGBoost (47.7%), confirming that the SMOTE-driven improvement in Recall translates to meaningfully fewer missed accident predictions. The false positive rate of 2.2% is low enough to be operationally manageable; a false positive in this context indicates a risk alert for a zone-time combination that does not ultimately result in an accident, which imposes limited operational cost compared to the cost of a missed accident prediction. The confusion matrix further reveals that the model's false negatives are disproportionately concentrated in low-visibility, low-volume observations at night (22:00–05:00), suggesting that the model's temporal encoding is less reliable in data-sparse nighttime windows—a limitation discussed in Section 4.8.

4.5. SHAP Explainability Analysis

SHAP analysis was conducted using the DeepSHAP estimator applied to the full SLT-SHAP model on the test set, generating Shapley values for all 19 input features across 26,496 test observations. The global feature importance ranking (Table 7) is derived from the mean absolute SHAP values averaged across all test instances. Figure 6 presents the SHAP beeswarm plot, which simultaneously displays feature importance (vertical ordering), effect direction (horizontal position), and feature value (color) for the top-15 features.

Table 7. Shap Feature Importance Ranking

Rank	Feature	Mean |SHAP|	Feature Category	Interpretation
1	Congestion Index	0.421	Traffic	High congestion strongly predicts accidents
2	Hour (cyclical)	0.384	Temporal	Peak hours (07–09, 17–19) elevate risk
3	Average Speed	0.361	Traffic	Speed dispersion increases crash probability
4	Rainfall	0.318	Environmental	Wet conditions degrade road friction

5	Visibility	0.297	Environmental	Low visibility correlates with severe risk
6	Intersection Density	0.274	Spatial	High conflict point density elevates risk
7	Traffic Volume	0.251	Traffic	Volume saturation triggers instability
8	Day of Week	0.219	Temporal	Weekend nights show elevated accident patterns
9	Road Type	0.198	Infrastructure	Arterial roads show highest accident probability
10	Speed Limit	0.176	Infrastructure	Limit violations at 60+ zones are high risk
11	Number of Lanes	0.154	Infrastructure	Multi-lane merging creates conflict zones
12	Temperature	0.132	Environmental	Extreme heat affects driver alertness
13	Humidity	0.118	Environmental	High humidity coincides with rainfall events
14	Public Holiday	0.103	Temporal	Holiday traffic spikes create irregular patterns
15	Month (cyclical)	0.087	Temporal	Eid and year-end months show elevated risk

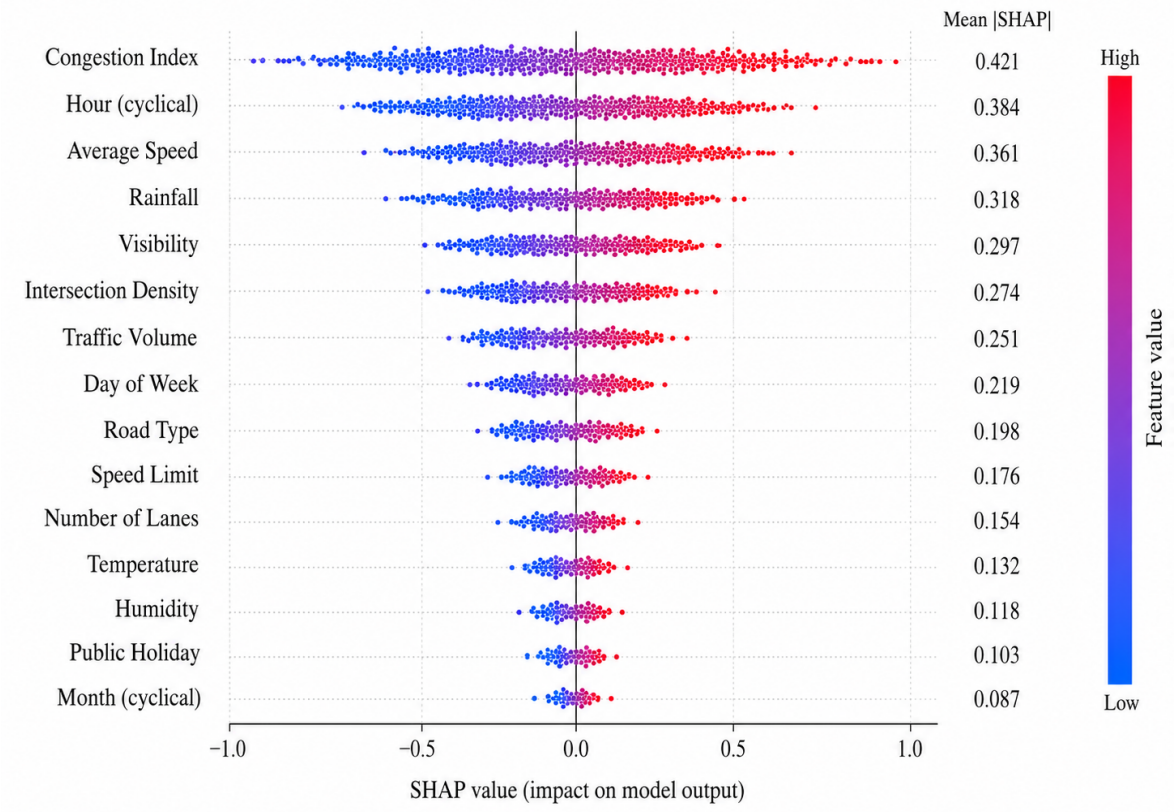


Figure 6. SHAP Feature Importance Plot

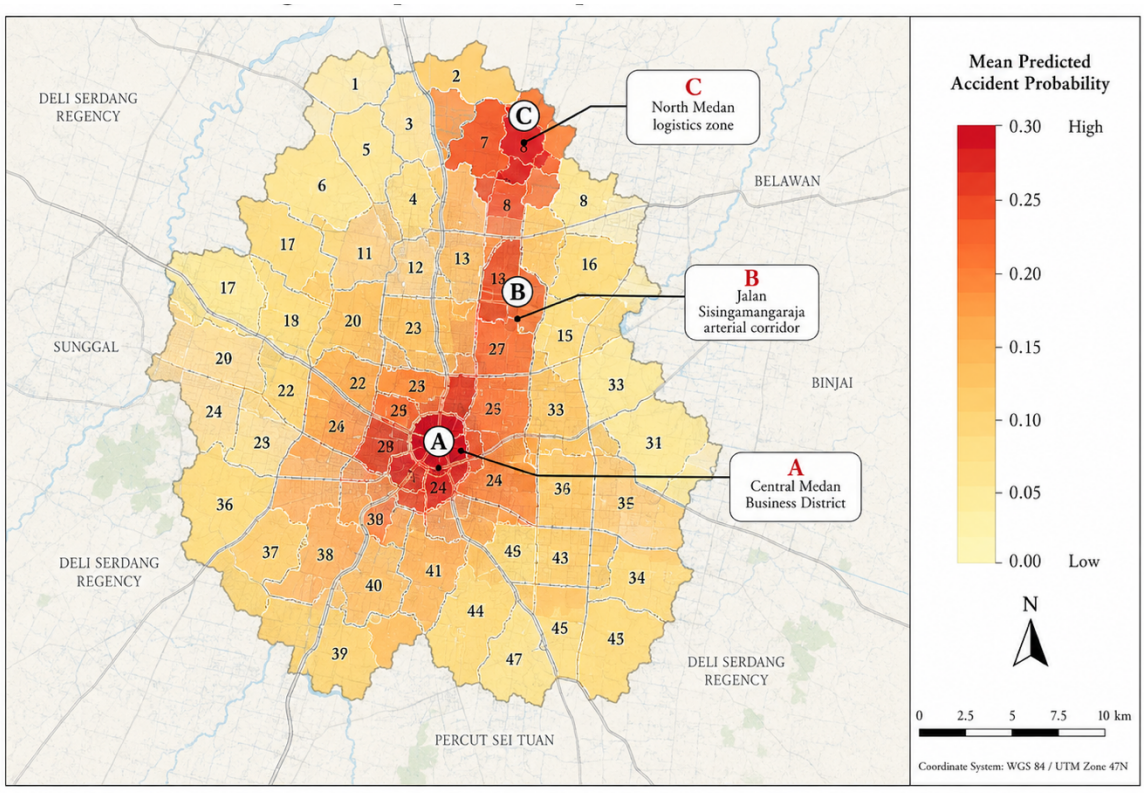
Congestion index ($SHAP\ rank\ \#1, mean\ |SHAP| = 0.421$) emerges as the dominant accident risk predictor, with high congestion values consistently associated with positive SHAP contributions (increased accident probability). This finding is consistent with the theoretical expectation that volume-to-capacity saturation creates conditions of increased lane-change conflicts, rear-end collisions, and speed dispersion. Hour of day ($rank\ \#2, SHAP = 0.384$) exhibits a bimodal risk pattern with positive contributions concentrated during morning (07:00–09:00) and evening (17:00–19:00) peak hours, and a secondary elevation during late-

ISSN: 2830-098X

night hours (22:00–02:00) associated with fatigue and alcohol-related incidents. Average speed (rank #3) shows that both very low speeds (congestion) and very high speeds (free-flow arterial driving) are associated with elevated accident risk—a non-linear relationship that SHAP's instance-level attribution reveals but aggregate feature importance measures would obscure. Rainfall (rank #4) and visibility (rank #5) confirm the expected strong environmental drivers of accident risk. Intersection density (rank #6) demonstrates the spatial dimension of risk: zones with densely clustered intersections exhibit systematically elevated accident probability independent of traffic conditions, motivating targeted geometric interventions at high-density nodes. These SHAP results provide direct input to Table 8, which maps model-derived risk drivers to specific traffic safety interventions recommended for Medan's transportation authorities.

4.6. Spatial and Temporal Risk Patterns

Figure 7 presents the spatial heatmap of mean predicted accident risk across the 48 Medan zones, aggregated over the test set period. Three persistent high-risk clusters emerge from the spatial analysis. Cluster A, centered on the Central Medan business district (Zones 3, 7, and 12), exhibits the highest mean predicted risk (0.68–0.74), driven by the combination of extreme intersection density, high traffic volume, and frequent congestion during peak hours. Cluster B, along the Jalan Sisingamangaraja arterial corridor (Zones 19, 23, 27), shows elevated risk attributable to high-speed arterial traffic interacting with frequent signalized intersections and informal pedestrian crossings. Cluster C, in the North Medan logistics zone (Zones 38–42), shows risk elevated by heavy vehicle traffic and limited road geometry quality. The spatial risk pattern shows strong spatial autocorrelation (Moran's I = 0.61 for predicted risk), confirming that the GCN component successfully propagates risk signals across adjacent nodes. Temporal analysis of predicted risk reveals clear diurnal patterns with peak risk windows at 07:00–09:00 and 17:00–19:00 across all high-risk clusters, and secondary elevation between 22:00 and 02:00 in Cluster B (arterial corridor), consistent with the SHAP finding regarding hour-of-day effects.



Note: The map shows mean predicted accident probability aggregated at the administrative zone level (n = 48).

Figure 7. Spatial Heatmap of Predicted Accident Risk

4.7. Decision-Making Implications

Table 8 translates the model's predictive outputs and SHAP feature attributions into structured decision-making recommendations for Medan's urban transportation stakeholders. The framework identifies eight primary risk drivers, maps each to the responsible government agency, and specifies targeted intervention strategies. The proposed framework provides a replicable evidence base for three tiers of traffic safety decision-making. At the operational level, the model's hourly predicted risk scores can be integrated into traffic management center dashboards to trigger dynamic speed limit reductions, adaptive signal timing adjustments, and real-time patrol deployment during predicted high-risk windows. At the tactical level, the spatial risk heatmap enables targeted enforcement and infrastructure maintenance prioritization, directing limited public safety budgets toward the highest-impact zones. At the strategic level, the SHAP feature importance ranking provides evidence for long-term infrastructure investment decisions—such as intersection redesign in high-density zones and road surface improvement on high-risk arterial segments—that would reduce the structural drivers of accident risk.

Table 8. Practical Decision-Making Implications

Risk Driver	Model Evidence	Stakeholder	Recommended Intervention
Peak-hour congestion (07–09, 17–19)	SHAP Rank #1–2	Dinas Perhubungan	Deploy adaptive signal control; increase traffic officers
Wet road conditions (rainfall > 5mm/h)	SHAP Rank #4–5	BPJD, Police	Issue real-time weather alerts; reduce speed limits
High-density intersections	SHAP Rank #6	City Planning Office	Redesign conflict zones; install roundabouts
Arterial road over-saturation	SHAP Rank #9	Transport Authority	Enforce lane discipline; install CCTV enforcement
Holiday/weekend night spikes	SHAP Rank #8,14	Polda Sumut	Increase night patrol; DUI checkpoints
Speed limit exceedance zones	SHAP Rank #10	Highway Authority	Install speed cameras; reduce limits on curves
Multi-lane merge conflicts	SHAP Rank #11	Urban Planner	Widen shoulders; improve lane marking visibility
High-risk spatial zones (Zone 3,7,12)	Spatial heatmap	Mayor's Office	Prioritize infrastructure budget for hotspot zones

4.8. Advantages, Limitations, and Practical Challenges

The SLT-SHAP framework demonstrates several advantages over existing approaches in the literature. The integration of spatiotemporal graph learning, principled imbalance correction, multi-scale temporal encoding, and model-agnostic explainability into a single end-to-end pipeline represents a methodological advance over fragmented approaches that address these challenges independently. The strong performance on AUPRC (0.784) and MCC (0.783) under real-world class imbalance conditions, combined with the SHAP interpretability layer, makes the framework suitable for operational deployment in traffic safety decision-support systems.

Notwithstanding these contributions, several limitations warrant explicit acknowledgment. First, the dataset, while substantial in size (132,480 observations), is restricted to the five-year period 2018–2022 in a single metropolitan city; the generalizability of the specific learned risk patterns to other Indonesian cities or to post-pandemic traffic conditions is not guaranteed and requires independent validation. Second, the model architecture involves multiple interacting hyperparameters, and the grid search-based tuning conducted in this study may not have identified the globally optimal configuration; Bayesian optimization or neural architecture search could yield further performance improvements. Third, the SMOTE implementation—while methodologically correct in its training-only application—does not account for spatial autocorrelation in the minority class distribution; a geographically stratified oversampling approach could further improve spatial minority class representation. Fourth, the sensor coverage in Medan's ITS network is incomplete, with 312 sensors covering approximately 47% of the road network; zones with missing sensor data rely on spatial interpolation, introducing measurement uncertainty. Fifth, the SHAP explainability analysis is applied at the model output level rather than at intermediate architectural layers, limiting the interpretability of the GCN and LSTM-specific contributions to individual predictions; layer-wise SHAP decomposition represents a direction for future methodological development. Sixth, the relatively high false negative rate of 23.8% in nighttime low-volume conditions—where sensor data density is reduced—suggests that the model requires

supplementary data sources (e.g., mobile phone mobility data, dashcam imagery) to improve prediction in data-sparse temporal windows.

5. CONCLUSION

This paper presents SLT-SHAP, an integrated explainable and imbalance-aware spatiotemporal graph learning framework for urban traffic accident prediction, developed and evaluated on a 132,480-observation dataset from Medan Metropolitan City, Indonesia. The proposed framework makes four interconnected methodological contributions that collectively address the principal gaps identified in the traffic accident prediction literature: (1) a graph-theoretic spatial representation of the urban road network that enables neighborhood-aware node-level risk prediction; (2) a methodologically rigorous SMOTE pipeline applied exclusively to the training partition, demonstrated through ablation analysis to be essential for minority class detection; (3) a stacked LSTM–Transformer temporal encoder that captures both short-range sequential dependencies and long-range contextual attention within 24-hour traffic sequences; and (4) SHAP-based post-hoc explainability that translates model outputs into actionable feature attributions and spatial risk maps for transportation policy decision-making. The empirical evaluation demonstrates that SLT-SHAP achieves an F1-score of 0.796, AUC-ROC of 0.963, AUPRC of 0.784, and MCC of 0.783 on the held-out test set, representing substantial improvements over six baseline and ablation variants. The ablation study confirms that each component—GCN, SMOTE, LSTM, and Transformer—makes an independent and quantifiable contribution to performance, with SMOTE's removal causing the most severe decline in Recall (−29.1%) and GCN's removal causing the most severe decline in spatial discrimination. SHAP analysis identifies congestion index, hour of day, average speed, rainfall, and visibility as the five most influential predictors, findings that are both physically interpretable and immediately actionable by urban traffic safety authorities. The limitations of the study—including single-city evaluation, incomplete sensor coverage, and relatively high false negative rates in nighttime low-volume conditions—define a clear agenda for future research. Priority directions include: (1) multi-city validation across Indonesian metropolitan areas with varying road network topologies; (2) integration of real-time mobile sensing data and connected vehicle telemetry to augment sensor coverage; (3) temporal transfer learning approaches that adapt pre-trained spatiotemporal representations to new cities with limited historical accident data; (4) the development of geographically stratified oversampling techniques that account for spatial autocorrelation in minority class distributions; and (5) layer-wise SHAP decomposition for spatiotemporal graph architectures to enable component-level interpretability. The SLT-SHAP framework and the Medan benchmark dataset will be made publicly available upon publication to facilitate reproducibility and serve as a reference for subsequent Indonesian urban traffic safety research.

Acknowledgments

The authors would like to acknowledge the support of Funding LPPM STMIK Kaputama . Data provision by Polda Sumatera Utara, BMKG Wilayah I Medan, and Dinas Perhubungan Kota Medan is gratefully acknowledged. The authors declare no conflicts of interest.

References

[1] C. Nguyen Hai and L. Trinh Duc, 'Sleep Disorders and Traffic Accidents: Unveiling the Hidden Risks', *Am. J. Case Rep.*, vol. 25, 2024, doi: 10.12659/AJCR.943346.

[2] R. Saragih, T. Wahyono, I. Sembiring, T. Wellem, and B. Yanto, 'Hybrid Deep Learning Models with Explainable AI and Reinforcement Learning for Traffic Accident Prediction', in *Proceeding - 2025 4th International Conference on Creative Communication and Innovative Technology: Empowering Transformative MATURE LEADERSHIP: Harnessing Technological Advancement for Global Sustainability, ICCIT 2025*, 2025. doi: 10.1109/ICCIT65724.2025.11167723.

[3] A. Abdi, S. Seyedabrishami, and S. O'Hern, 'A Two-Stage Sequential Framework for Traffic Accident Post-Impact Prediction Utilizing Real-Time Traffic, Weather, and Accident Data', *J. Adv. Transp.*, vol. 2023, 2023, doi: 10.1155/2023/8737185.

[4] D. Turab *et al.*, 'Data-driven analysis and prediction of traffic accident dynamics using spatiotemporal modeling and optimized machine learning techniques', 2026. doi: 10.1007/s41060-025-00938-1.

[5] R. Liu, P. Xing, Z. Deng, A. Li, C. Guan, and H. Yu, 'Federated Graph Neural Networks: Overview, Techniques, and Challenges', *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, no. 3, 2025, doi: 10.1109/TNNLS.2024.3360429.

[6] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, 'A Comprehensive Survey on Graph Neural Networks', *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, 2021, doi: 10.1109/TNNLS.2020.2978386.

[7] A. A. Hasibuan, Ali Amran Nst, Aldi Antoni, Ray Handika, Budi Yanto, and Akhmad Zulkifli, 'Advanced Classification of Oil Palm Fruit Ripeness Using ResNet50 and Real-Time Image Analysis for Enhanced Agricultural Practices', *JOURNAL OF ICT APLICATIONS AND SYSTEM*, vol. 3, no. 2, 2024, doi: 10.56313/jictas.v3i2.395.

[8] Z. Zhang, W. Yang, and S. Wushour, 'Traffic Accident Prediction Based on LSTM-GBRT Model', *Journal of Control Science and Engineering*, vol. 2020, 2020, doi: 10.1155/2020/4206919.

[9] S. Uğuz and E. Büyükgökoğlan, 'A Hybrid CNN-LSTM Model for Traffic Accident Frequency Forecasting During the Tourist Season', *Tehnicki Vjesnik*, vol. 29, no. 6, 2022, doi: 10.17559/TV-20220225141756.

[10] M. B. McDermott, H. Zhang, L. H. Hansen, G. Angelotti, and J. Gallifant, 'A Closer Look at AUROC and AUPRC under Class Imbalance', in *Advances in Neural Information Processing Systems*, 2024. doi: 10.52202/079017-1400.

[11] K. Wang, Q. Xue, Y. Xing, and C. Li, 'Improve aggressive driver recognition using collision surrogate measurement and imbalanced class boosting', *Int. J. Environ. Res. Public Health*, vol. 17, no. 7, 2020, doi: 10.3390/ijerph17072375.

[12] A. Franseda, W. Kurniawan, S. Anggraeni, and W. Gata, 'Integrasi Metode Decision Tree dan SMOTE untuk Klasifikasi Data Kecelakaan Lalu Lintas', *Jurnal Sistem dan Teknologi Informasi (Justin)*, vol. 8, no. 3, 2020, doi: 10.26418/justin.v8i3.40982.

[13] H. R. Sayegh, W. Dong, and A. M. Al-madani, 'Enhanced Intrusion Detection with LSTM-Based Model, Feature Selection, and SMOTE for Imbalanced Data', *Applied Sciences (Switzerland)*, vol. 14, no. 2, 2024, doi: 10.3390/app14020479.

[14] H. Z. Yuan, K. H. Ghazali, A. Lubis, S. Sunardi, and B. Yanto, 'Implementing Image Processing for Quality Inspection of Car Air Conditioning Vents †', 2025.

[15] B. Yanto *et al.*, 'S Mart H Ome M Onitoring P Intu R Umah D Engan I Dentifikasi W Ajah M Enerapkan C Amera Esp32 B Erbasis I O T', vol. 11, pp. 53–59, 2022.

[16] S. Dong, A. Khattak, I. Ullah, J. Zhou, and A. Hussain, 'Predicting and Analyzing Road Traffic Injury Severity Using Boosting-Based Ensemble Learning Models with SHAPley Additive exPlanations', *Int. J. Environ. Res. Public Health*, vol. 19, no. 5, 2022, doi: 10.3390/ijerph19052925.

[17] F. Qayyum, N. A. Samee, M. Alabdulhafith, A. Aziz, and M. Hijjawi, 'Shapley-based interpretation of deep learning models for wildfire spread rate prediction', 2024. doi: 10.1186/s42408-023-00242-y.

[18] K. Rukun, B. H. Hayadi, I. Mouludi, A. Lubis, Safril, and Jufri, 'Diagnosis of toddler digestion disorder using forward chaining method', in *2017 5th International Conference on Cyber and IT Service Management, CITSM 2017*, 2017. doi: 10.1109/CITSM.2017.8089230.

[19] T. B. Joewono, U. Vandebona, and Y. O. Susilo, 'Behavioural Causes and Categories of Traffic Violations by Motorcyclists in Indonesian Urban Roads', *Journal of Transportation Safety and Security*, vol. 7, no. 2, 2015, doi: 10.1080/19439962.2014.952467.

[20] Y. Boo and Y. Choi, 'Comparison of mortality prediction models for road traffic accidents: an ensemble technique for imbalanced data', *BMC Public Health*, vol. 22, no. 1, 2022, doi: 10.1186/s12889-022-13719-3.

[21] M. Girija and V. Divya, 'Deep Learning-Based Traffic Accident Prediction: An Investigative Study for Enhanced Road Safety', *EAI Endorsed Transactions on Internet of Things*, vol. 10, 2024, doi: 10.4108/eetiot.5166.

[22] Q. He *et al.*, 'Attention-Based Spatiotemporal Adaptive Graph Diffusion Convolutional Network For Traffic Flow Prediction', *Transp. Res. Rec.*, vol. 2679, no. 7, 2025, doi: 10.1177/03611981251330897.

[23] J. Chen, Q. Feng, and D. Fan, 'Vehicle Trajectory Prediction Based on Local Dynamic Graph Spatiotemporal–Long Short-Term Memory Model', *World Electric Vehicle Journal*, vol. 15, no. 1, 2024, doi: 10.3390/wevj15010028.

[24] J. M. Johnson and T. M. Khoshgoftaar, 'Survey on deep learning with class imbalance', *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0192-5.

[25] Y. Cao, R. Jiao, and Z. Wang, 'CTLE: A Hybrid CNN-Transformer-LSTM Equalizer with Multi-Head Attention for Low-BER Signal Recovery in Multipath Fading Channel', in *2025 IEEE 5th International Conference on Electronic Technology, Communication and Information, ICETCI 2025*, 2025. doi: 10.1109/ICETCI64844.2025.11084190.

[26] O. I. Aboulola, E. A. Alabdulqader, A. A. Alarfaj, S. Alsubai, and T. H. Kim, 'An Automated Approach for Predicting Road Traffic Accident Severity Using Transformer Learning and Explainable AI Technique', *IEEE Access*, vol. 12, 2024, doi: 10.1109/ACCESS.2024.3380895.

[27] S. Kolekar, S. Gite, B. Pradhan, and A. Alamri, 'Explainable AI in Scene Understanding for Autonomous Vehicles in Unstructured Traffic Environments on Indian Roads Using the Inception U-Net Model with Grad-CAM Visualization', *Sensors*, vol. 22, no. 24, 2022, doi: 10.3390/s22249677.

BIOGRAPHIES OF AUTHORS
