Advanced Multi-Scale Temporal Pattern Recognition for Regional Traffic Accident Prediction: An Ensemble Machine Learning Approach

Stefan ĆIRKOVIĆ, Aleksa IRIČANIN, Marija BLAGOJEVIĆ*, Srećko ĆURČIĆ

Faculty of Technical Sciences Čačak, University of Kragujevac, Svetog Save 65, 32102, Čačak, Serbia stefan.cirkovic@ftn.kg.ac.rs, aleksa.iricanin@ftn.kg.ac.rs,
marija.blagojevic@ftn.kg.ac.rs (**Corresponding author*), srecko.curcic@ftn.kg.ac.rs

Abstract: Accurate traffic accident prediction remains a key challenge in the context of improving road safety, as the traditional statistical approaches often overlook complex spatio-temporal patterns and seasonal dynamics. This study proposes an advanced system for forecasting the number of accidents across five regions of Serbia, based on multi-scale temporal modeling and ensemble machine learning methods. The analysis that was carried out relies on official data from the Road Traffic Safety Agency of the Republic of Serbia covering the 2015-2020 period, and comprising approximately 1.8 million records. Based on this dataset, 185 advanced temporal features were generated in order to capture the short-, medium-, and long-term dynamics of traffic accidents. The proposed model combines the XGBoost, LightGBM, and Random Forest algorithms through a weighted ensemble strategy, with a strict temporal validation and robust protection against data leakage. The experimental results included a mean absolute error of 10.62 ± 0.61 accidents per week per region, corresponding to a 64% and 84% reduction in the value of MAE in comparison with the Naive (Last) and Naive (Mean) baseline approaches, respectively, while achieving a R^2 value of 0.97 ± 0.01. The regional analysis confirmed the proposed system's ability to recognize risk heterogeneity with regard to traffic accidents and identify the high-risk periods, particularly in urban areas with a pronounced traffic volatility. This system provides interpretable insights and actionable recommendations, demonstrating its applicability for the Road Traffic Safety Agency and the local governments in planning preventive measures and optimizing the allocation of police and municipal resources.

Keywords: Traffic accident prediction, Ensemble learning, Temporal feature engineering, Spatio-temporal modeling, Road safety, Machine learning.

1. Introduction

Traffic accident prediction is one of the key challenges in the field of road safety, as it requires an understanding of complex spatio-temporal patterns and their impact on incident risk (Lord & Mannering, 2010; Behboudi, Moosavi & Ramnath, 2024). Traditional statistical methods and standard prediction models often rely on a limited set of indicators and simplified assumptions, which leads to the neglect of the multidimensional dynamics and interdependencies among risk factors. As a result, their predictions are often insufficiently accurate or stable for practical use in strategic planning and operational decision-making.

This study introduces a new methodological framework based on multi-scale temporal pattern recognition and ensemble machine learning techniques, aimed at improving the prediction of the number of traffic accidents at regional level in the Republic of Serbia. The developed system involves the generation of 185 advanced temporal features that capture short-term (1–6 weeks), medium-term (8–16 weeks), and long-term (26–52 weeks) traffic dynamics. Such a multi-scale architecture enables the simultaneous recognition of the immediate effects of weather conditions and local traffic disturbances, as well as seasonal cycles and structural trends.

The proposed framework further enhances the prediction accuracy through a weighted ensemble of complementary algorithms (XGBoost, LightGBM, and Random Forest), thereby jointly exploiting their strengths in detecting nonlinear patterns, efficiently processing large datasets, and maintaining robustness under noisy conditions. A special emphasis is placed on strict temporal validation and the prevention of data leakage, ensuring a realistic performance assessment and guaranteeing the model's ability to generalize under real-life conditions.

This approach facilitates the identification of high-risk periods and specific regional patterns, providing a foundation for proactive decision-making in the field of road safety.

The practical application of the developed system is particularly relevant for the Road Traffic Safety Agency, local governments, and other competent institutions, as it enables the timely planning of preventive measures, the optimization of police and municipal resources, an improved management of infrastructural interventions, and the design of targeted educational campaigns. Furthermore, the system can be applied within seasonal and regional strategies, from identifying

https://doi.org/10.24846/v35i2y202604

critical periods (e.g. winter conditions, summer tourist peaks) to distinguishing specific patterns between urban and rural areas.

The proposed model is primarily designed as a strategic planning tool based on aggregated regional indicators, rather than as a system for operational decision-making at a micro-spatial level (e.g. individual streets or intersections). The selected level of spatial aggregation and temporal resolution (in weeks) allows for the stable identification of dominant risk patterns, seasonality, and long-term trends. However, it does not capture highly local and short-term variations that are typical for large urban areas such as Belgrade, where the specific infrastructure, traffic intensity, and demographic factors lead to a higher volatility.

In this context, the results obtained by the proposed model can be used as an analytical and planning framework for national and regional institutions, as well as for municipalities, to identify high-risk periods and regions. At the municipal level, regional predictions can serve as a prioritization and guidance layer for more detailed local analyses based on finer-grained data (e.g. local traffic flows, infrastructure characteristics, or site-specific conditions). In this way, regional risk assessment can be effectively linked with the local road safety policies and interventions.

The remainder of this paper is organized as follows. Section 2 reviews the related work on traffic accident prediction and ensemble machine learning approaches. Section 3 presents the proposed methodology, including dataset preparation, feature engineering, temporal modeling, the ensemble learning framework, and the validation strategy. Further on, Section 4 discusses the experimental results and sets forth a comparative performance analysis. Finally, Section 5 concludes this paper and outlines directions for future research.

2. Related Work

Traffic accident prediction has evolved from classical statistical models to modern, data-driven architectures that explicitly capture the spatio-temporal dynamics. Early studies predominantly employed generalized linear models with negative binomial specifications, accompanied by the explicit reporting of confidence and prediction intervals - a crucial aspect for engineering applications and risk quantification across road networks (Lord & Mannering, 2010). In parallel, classical time series models such as SARIMA provide a baseline for short-term accident frequency prediction but have a limited capability to capture multi-scale patterns and structural changes (Deretić et al., 2022).

With the increasing availability of data, reviews confirm that machine learning models generally outperform classical approaches, but they also highlight methodological shortcomings -particularly an insufficiently strict temporal validation and the neglect of data leakage issues (Behboudi, Moosavi & Ramnath, 2024; Chai et al., 2024).

In the analysis of spatial “hot spots,” the ensemble techniques applied to heterogeneous spatio-temporal flows demonstrate a greater stability and a more accurate risk mapping than individual models (Ding et al., 2020). For predicting accident severity, Random Forest and Boosting methods consistently outperform decision trees, with a particular emphasis on the importance of temporal patterns and road characteristics (Umer et al., 2020; Gan et al., 2020; Muktar & Fono, 2024). A similar conclusion was drawn by a recent conference paper on ensemble learning for traffic accident severity classification, focusing on the engineering pipeline (preprocessing–learning–evaluation) (Yu & Chen, 2025), as well as by research linking accident severity assessment to traffic flow impact using ensemble approaches (Ara & Hashemi, 2021).

In addition to frequency and severity, accident duration is a critical operational metric. A heterogeneous ensemble for accident duration prediction demonstrates that combining different modeling paradigms (and feature sets) yields measurable improvements over individual models, particularly in the presence of heterogeneous causal patterns and imbalanced distributions (Zhao & Deng, 2022). Regarding frequency, neural network approaches have also been explored: one study shows that the application of artificial neural networks provides certain improvements, noting that stability largely depends on carefully defined temporal windows and appropriate regularization (Gatarić et al., 2023).

Although not primarily developed for accident prediction, traffic flow forecasting architectures have motivated the development of multi-scale models encompassing short-, medium-, and long-term cycles. T-GCN, by combining graph convolutional layers with GRU units, effectively models seasonality and road network dynamics (Zhao et al., 2020), while DCRNN integrates diffusion convolutions with a recurrent encoder–decoder for the precise tracking of state propagation (Li et al., 2018). In recent years, these approaches have increasingly been applied to accident analysis, utilizing lags, temporal windows, and graph-based aggregations.

Heterogeneous stacking frameworks, which explicitly combine parametric and machine learning–based models, consistently achieve a superior out-of-sample performance in comparison with individual paradigms, reducing model-specific biases and improving calibration (Ahmad, Wali & Khattak, 2023). In line with this, this study integrates statistical rigor (strict temporal splitting, control of data leakage, clearly defined prediction metrics) with multi-scale, leak-safe temporal features and a conservatively regularized ensemble, evaluated through a multi-period, temporally separated analysis - thereby addressing the gaps identified in previous studies (Behboudi, Moosavi & Ramnath, 2024; Chai et al., 2024).

While the study by Ramani & Selvaraj (2014) focuses on predicting traffic accident severity using feature selection methods and classical classifiers, the present work addresses the prediction of accident frequency across regions using multi-scale temporal modeling and ensemble machine learning techniques.

In addition to methodological contributions, operational support systems have also been developed, including applications for the visualization and cluster analysis of traffic accidents in Serbia, demonstrating the practical value of spatial analyses for safety strategies (Ćirković et al., 2025).

Unlike earlier studies that focused primarily on large urban regions or relied on individual models (GLM, SARIMA, ANN) or ensembles without strict temporal validation, this study shows that the proposed methodological framework is applicable in smaller spatial contexts with a pronounced seasonality and heterogeneity, such as the road network in Serbia. This confirms the flexibility and broader applicability of the proposed model while addressing common issues such as data leakage, a limited evaluation across temporal horizons, and poor interpretability.

3. Methodology

To predict the number of traffic accidents across the territory of the Republic of Serbia, an advanced spatio-temporal model was developed based on the principles of *multi-scale temporal pattern recognition* and *ensemble machine learning approaches*. The core challenge in this domain lies in the fact that accident counts depend not only on the current conditions but also on the complex history of previous incidents, seasonal cycles, and regional specificities, which together form multi-layered temporal patterns spanning from one to fifty-two weeks.

The proposed methodological framework comprises several interconnected phases: spatio-temporal data organization, the construction of multi-scale temporal features across six distinct analysis categories, the implementation of an ensemble learning framework with four complementary algorithms, and a robust statistical performance evaluation through bootstrap validation and comprehensive baseline benchmarking.

This integrated approach enables the development of an interpretable and efficient traffic accident prediction system that demonstrates an outstanding performance and a high degree of generalization.

Figure 1 illustrates a flow diagram representing the key components of the proposed methodology: from the initial dataset of 1.8 million records, through the *feature engineering* process that generates 185 advanced temporal features, to the ensemble learning phase with four algorithms, and finally to validation based on multiple testing scenarios.

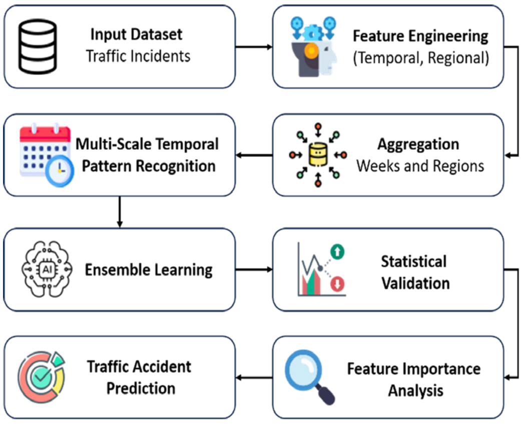

Figure 1. Overview of the methodological framework

3.1 Dataset Description

The official traffic accident dataset for Serbia for the period 2015–2020, provided by the Road Traffic Safety Agency (Road Traffic Safety Agency of the Republic of Serbia, 2025), was used in this study (~1.8 million records with dozens of spatial, temporal, and contextual attributes). The experiments were conducted in Python, and the complete code, including the data preparation pipeline, ensemble implementation, and validation procedures, is publicly available on GitHub (Ćirković, Iričanin & Blagojević, 2025). Prior to modeling, initial data cleaning (the removal of extreme and missing values using domain knowledge) and the selection of the most relevant features were performed (Table 1).

Table 1. Feature Selection

Feature	Description
DATE	Date of the accident
HOUR	Hour of the day when the accident occurred (0–23)
DAY_OF_WEEK	Day of the week (0–6)
WEEK, YEAR, MONTH	Derived temporal labels (aggregation by week and month)
X_COORD, Y_COORD	Geographic coordinates of the accident (longitude and latitude)
REGION	Region categorization: Belgrade, Vojvodina, Central Serbia, South, East
WEATHER_ CONDITIONS	Weather conditions at the time of the accident
LIGHTING	Lighting conditions (day, night, dusk, etc.)
ROAD_SURFACE_ CONDITIONS	Road surface condition (dry, wet, icy, etc.)
ROAD_ CHARACTERISTICS	Road characteristics (curve, incline, intersection, tunnel etc.)

The dataset also contains attributes that were not directly used (e.g. information on vehicles, participants, outcomes), while the presented subset of features was selected based on their relevance for the spatio-temporal prediction of accident counts. Since the raw data is inadequate for sophisticated models and the numerical representation of time and space loses its semantic meaning (e.g. December and January are consecutive), comprehensive feature engineering was performed to transform them in order to make them suitable for machine learning and the discovery of hidden patterns.

3.2 Feature Engineering

To enhance the predictive power of the model and enable insights into spatio-temporal accident patterns, comprehensive feature engineering was performed. This process included generating cyclical temporal transformations as well as the spatial categorization of regions based on geographic coordinates, allowing for the aggregation of accidents across the regions of Serbia.

3.2.1 Cyclical Temporal Variables

The temporal variables in the dataset (e.g. hour of the day, day of the week, month of the year) have a cyclical nature. In their linear numerical representation, the information about adjacency is lost (e.g. 23:00 and 0:00, or December and January), which can lead to misinterpretation by the model.

To overcome this issue, sine and cosine transformations were applied to preserve the cyclical continuity.

Mathematical formulation:

For each temporal variable x, two cyclical components are defined:

$$x_{sin} = sin\left(2\pi \times \frac{x}{P}\right) \tag{1}$$

$$x_{\cos} = cos\left(2\pi \times \frac{x}{P}\right) \tag{2}$$

where:

- x is the value of the temporal variable (e.g. month = 3);
- P is the period of the observed variable (24 for hours, 7 for days of the week, 12 for months, 52 for weeks in a year);
- x_{sin}, x_{cos} are transformed values that preserve the cyclical nature.

In this way, the model is capable of capturing seasonal and periodic patterns, such as high-risk traffic intervals (morning and evening rush hours, nighttime hours, weekends). In parallel with the cyclical transformations, a set of indicator variables was introduced to explicitly mark the known high-risk periods (e.g. MORNING_RUSH for hours 7–9, FRIDAY_EVENING for weekend hours, WINTER_PERIOD for December–March). This enhances model interpretability and accelerates the learning process.

3.2.2 Spatial Categorization of Regions and Aggregation

The geographic location of an accident (longitude and latitude) provides important spatial information, but its direct use in models often reduces their performance due to the continuous nature of the data and technical incompatibility with classical algorithms. To overcome this issue, discretization was introduced through a heuristic function, AssignRegion (Ćirković, Iričanin & Blagojević, 2025), which maps each accident to one of five predefined regions of Serbia (Belgrade, Vojvodina, Central Serbia, South, East). This categorization enables a fast and efficient aggregation of accidents without the need for GIS tools and provides a stable basis for spatio-temporal modeling.

Although this spatial discretization reduces fine local detail, the use of a limited number of regions represents a deliberate trade-off between spatial resolution and the statistical stability of the time series. More detailed spatial divisions would introduce higher noise and unstable temporal patterns, especially in areas with a lower traffic intensity. Since the goal of this study is regional-level temporal prediction rather than micro-location risk identification, the chosen aggregation level enables a more reliable learning of the seasonal and structural patterns.

To create a coherent spatio-temporal data structure suitable for prediction, traffic accidents were aggregated by taking weeks as time units and according to the predefined geographic regions. The temporal dimension includes YEAR and WEEK, providing a fixed granularity of 52 units per year, which enables a stable segmentation, facilitates time series analysis, and supports the implementation of predictive models in practical settings. The spatial dimension, REGION, allows for capturing the heterogeneity of accidents across different parts of Serbia.

Based on this aggregation, the dependent variable was defined as the number of traffic accidents in a given region r during week t, which can be formally expressed as:

$$y_{r,t} = N_{r,t} \tag{3}$$

where $N_{r,t}$ represents the number of accidents in region r during week t.

This weekly aggregation reduces variance and mitigates random fluctuations, resulting in a greater stability and an improved generalization of the predictive model.

In addition to spatial structuring and cyclic time transformations, recognizing the complex temporal dynamics that influences the occurrence of traffic accidents is of crucial importance. Simply converting dates into numeric labels is not sufficient – it is necessary to capture short-term oscillations, seasonal transitions, and long-term cycles that span the entire year. For this reason, a multi-layered system for constructing temporal features was developed, enabling the model to recognize patterns across different time horizons.

3.2.3 Temporal Feature Construction

One of the key challenges in traffic accident prediction lies in the fact that accidents do not occur in isolation, but rather within complex temporal patterns. Driver behavior, seasonal influences, climatic factors, and infrastructure characteristics operate across different time horizons. Therefore, the number of accidents cannot be accurately predicted based solely on "yesterday" or "last week"; instead, it is necessary to model a rich spectrum of temporal dynamics. This subsection presents the development of a multi-scale architecture of temporal features that enables the model to simultaneously capture short-term oscillations, seasonal transitions, and long-term cycles.

The selected temporal windows are not arbitrary but motivated by well-known traffic safety patterns and the need to ensure time series stability. Short-term windows (1–6 weeks) allow the model to capture immediate and short-lived effects, such as weather conditions, holidays, roadworks, and sudden changes in traffic flow. Medium-term windows (8–16 weeks) correspond to seasonal transition periods and enable the detection of gradual changes and risk trends. Long-term windows (26–52 weeks) reflect annual seasonal cycles and allow the comparison with historical patterns from the same periods of the previous years. This multi-scale approach enables the model to reliably and interpretably learn short-term fluctuations, seasonal transitions, and long-term structural patterns.

3.2.3.1 Short-term Dynamics

Short-term features (covering 1–6 weeks back) enable the model to recognize the immediate effects of recent events. For instance, a sudden increase in the number of accidents in the past few weeks may indicate specific weather conditions (snow, rain) or infrastructure issues (roadworks, increased local traffic). If the model relied solely on data from the current week, such effects would remain "invisible." By introducing lagged features, the model gains the ability to learn how recent fluctuations propagate into future outcomes.

In this way, short-term dynamics acts as the system's memory, recording what occurred in previous weeks and helping the model understand whether the current fluctuations are merely temporary oscillations or part of a broader trend.

To this aim, lagged features were constructed by shifting the time series by L weeks, where:

$$L \in \{1,2,3,4,5,6\} \tag{4}$$

Formally, for each region r and time period t:

$$Lag_{r,t}^{(L)} = y_{r,t-L} \tag{5}$$

where $y_{r,t}$ epresents the number of accidents in region r during week t.

These indicators capture short-term effects, such as a sudden increase in the number of accidents due to adverse weather conditions or seasonal traffic surges (e.g. during summer travel periods).

3.2.3.2 Medium-term Dynamics

While short-term indicators capture the immediate fluctuations, medium-term features (covering 8–16 weeks) provide insights into transitional phases, i.e. changes occurring across seasons. For example, the shift from winter to spring may bring improved driving conditions but also an increased traffic intensity. If the model relied only on short-term values, it might interpret such changes as random increases or decreases in the number of accidents. However, by comparing the average values across different periods (e.g. the last month versus the preceding three months), the model recognizes a gradual trend rather than a random oscillation. In other words, medium-term dynamics helps detect the "momentum," indicating whether the number of accidents is rising or falling continuously. This further stabilizes the model and enables it to predict broader trends rather than merely short-term spikes. To quantify these effects, momentum indicators were introduced, measuring the relative change between two time horizons:

$$Momentum_{r,t}^{(L1,L2)} = \frac{Lag_{r,t}^{(L1)} - Lag_{r,t}^{(L2)}}{Lag_{r,t}^{(L2)} + \varepsilon} \tag{6}$$

where $L1$ and $L2$ represent different lag lenghts ($L1 < L2$), and ∈ is a small stabilization parameter ($\varepsilon = 10^{-8}$) to prevent division by zero.

For example, the $Momentum^{(4,12)}$ indicator measures the difference between the monthly average and the three-month trend, allowing the model to detect whether accident counts are increasing or declining.

Linear trend coefficients were calculated over temporal windows of 8–16 weeks, enabling the detection of gradual changes in risk patterns.

3.2.3.3 Long-term Dynamics

Long-term features (covering 26–52 weeks) represent perhaps the most important component, as they capture the seasonal and structural patterns that repeat from year to year. For example, accident counts tend to increase during winter months due to adverse weather conditions, while in the summer months they rise because of tourism-related traffic and an increased travel volume. Indicators that compare the number of accidents in the same week across different years (so-called year-over-year changes) are particularly significant. This allows the model to distinguish between typical seasonal patterns and deep structural changes in the traffic system.

In order to formalize long-term effects, a Year-over-Year (YoY) change indicator was constructed:

$$YoY = \frac{y_{r,t-1} - y_{r,t-52}}{y_{r,t-52} + \varepsilon} \tag{7}$$

This indicator quantifies whether the number of accidents in a given week has significantly deviated from the corresponding week of the previous year, thereby capturing structural changes in traffic patterns.

3.2.4 Stabilization and Data Leakage Prevention

One of the key challenges when working with time series is the risk of data leakage, i.e. the inadvertent transfer of information from the future into the past. If the model had access to future information during training, it could achieve an artificially high performance but would fail completely in real-world conditions. To prevent this, all temporal features were generated strictly from past data (e.g.: $t-1$, $t-2$, $t-3$), while moving averages were computed exclusively based on past values:

$$MA_{r,t}^{(\omega)} = \frac{1}{\omega}\sum_{i=1}^{\omega} y_{r,t-i} \tag{8}$$

where ω denotes the size of the temporal window (in weeks) used for computing the moving average based exclusively on past observations.

Buffer periods were inserted between the training and test segments to prevent the "spillover" of seasonal patterns from one period to another. This approach ensures full methodological rigor and guarantees that the obtained results genuinely reflect the model's ability to generalize in real-world conditions.

3.3 Ensemble Learning Framework

One of the central challenges in predictive modeling is that no single algorithm can fully capture the complexity of the data. Linear models describe simple relationships well, while nonlinear models detect complex interactions but often suffer from overfitting. An ensemble approach combines the strengths of different algorithms, reducing variance and bias, and providing more stable predictions.

3.3.1 Model and Hyperparameter Selection

For the ensemble approach, four complementary algorithms were selected to achieve stability and an improved generalization. XGBoost was used as the primary model due to its efficiency in capturing complex nonlinear relationships, with a variant with a more conservative regularization to reduce the risk of overfitting. LightGBM contributed to the ensemble's speed and diversity, while Random Forest provided robustness when working with heterogeneous data.

Hyperparameter tuning was explicitly performed for all ensemble components using strict time-series cross-validation. For the XGBoost and LightGBM models, the learning rate, maximum tree depth, subsampling ratio, column sampling ratio, number of estimators, and regularization parameters (L1 and L2) were adjusted to balance the predictive accuracy and temporal stability while reducing the overfitting risk. A more conservative XGBoost variant with stronger regularization and shallower trees was additionally included to improve robustness. For the Random Forest model, the number of trees, maximum tree depth, the minimum samples per split and per leaf, and the feature subsampling strategy were tuned to control variance and enhance generalization. All hyperparameter choices were guided by the validation performance under time-series cross-validation rather than default settings.

3.3.2 Combination Strategy

The results of individual models were integrated through a weighted aggregation of predictions. The ensemble weighting strategy was heuristic and intentionally conservative. Ensemble weights were assigned based on the comparative performance and stability of the individual models observed across multiple temporally separated validation periods. This approach avoids overfitting and excessive sensitivity to specific time segments, which is particularly important in the context of non-stationary traffic accident data. During the experimental phase, a limited set of reasonable weighting configurations were empirically evaluated and compared, with emphasis on the consistency of performance rather than point-wise optimization. Based on these evaluations, the most stable and reliable configuration of weights was selected (0.4, 0.3, 0.2, and 0.1) allowing the main XGBoost model to have a dominant contribution, the additional XGBoost variant to provide conservative adjustment, LightGBM to enhance efficiency, and Random Forest to ensure robustness. A fully optimized weighting scheme was intentionally avoided to reduce the risk of overfitting to specific time segments and to preserve temporal robustness. Specifically, no continuous optimization, exhaustive grid search, or meta-learning approach (e.g. stacking) was applied to determine the ensemble weights, as such procedures could inadvertently fit to particular temporal segments and compromise generalization under non-stationary conditions. The final prediction was defined by the following linear combination:

$$\hat{y} = 0.4 \cdot \hat{y}_{XGB-main} + 0.3 \cdot \hat{y}_{XGB-cons} + 0.2 \cdot \hat{y}_{LGBM} + 0.1 \cdot \hat{y}_{RF} \tag{9}$$

3.4 Validation Methodology

In this study, a multi-layered validation strategy was implemented, combining temporal data splitting, composite evaluation metrics, bootstrap testing, and the regional disaggregation of performance.

3.4.1 Temporal Evaluation Across Multiple Periods

Unlike the conventional approaches where data is randomly split into training and test sets, this study employed temporal validation. Since the dataset represents a time series, random partitioning would lead to "data leakage" from the future into the past. Instead, testing was conducted across four temporally distinct segments, each covering a three-month period. This strategy enabled the verification of whether the model could maintain a stable performance across different time horizons and seasonal cycles, rather than being overfitted to a single data segment.

Formally, $T = \{t_1, t_2, ..., t_n\}$ shall denote the set of weeks within the observed period. For each test segment, the following were defined:

$$Train = \{t_1, ..., t_k\} \tag{10}$$

$$Test = \{t_{k+1}, ..., t_m\} \tag{11}$$

A buffer period of several weeks was introduced between the training and testing phase to eliminate the indirect contamination of the test set through lagged features.

3.4.2 Bootstrap Validation

To ensure statistical robustness, the bootstrap technique was applied. More than 1,000 resampled subsets were generated from the original test set, and all the evaluation metrics were recalculated based on these subsets. This procedure allows not only the estimation of the average performance values but also their standard deviations, providing a more precise insight into the model's stability. For example, the results are reported in the following format:

$$MAE = 154.8 \pm 26.9, \qquad R^2 = 0.913 \pm 0.042$$

Such a representation clearly demonstrates that the model's high performance was not randomly achieved, being a consistent property of its structure.

The following section presents the experimental results, including a comparison between the ensemble and individual models, as well as a detailed interpretation of its predictive accuracy across different temporal and regional contexts.

4. Results and Discussion

The implemented model was tested over four independent periods (July–October 2020), with strict protection against data leakage and conservative ensemble parameters. The model achieved a high accuracy and stability under realistic temporal validation with buffer zones that effectively eliminated information spillover from future data. Potential data leakage was controlled through minimal lag windows of one week, the removal of highly correlated and redundant features, and the application of strict temporal splits and model regularization.

The obtained results show that LightGBM achieved the lowest mean absolute error (30.37 ± 20.76), while Random Forest exhibited the highest MAE (35.09 ± 22.21). The primary XGBoost model reached a score of 31.98 ± 21.02, whereas its conservative variant slightly exceeded it with a score of 32.18 ± 21.65, indicating consistency with regard to the performance of the two approaches. These findings represent a solid foundation for further model comparison and optimization.

The performance results reported above correspond to the individual base models evaluated independently. All the subsequent evaluation metrics presented in this section (MAE, R^2, the relative error, and Hit Rate) refer exclusively to the final ensemble model, unless explicitly stated otherwise.

Given the high predictive performance of the ensemble model, particular attention was paid to the potential impact of temporal autocorrelation on the target variable. To mitigate the risk of an inflated performance due to trivial persistence effects, all temporal features were constructed using a minimum lag of one week, combined with strict temporal splits and buffer zones between training and test periods. This design ensures that the reported results reflect a genuine predictive capability rather than a mere exploitation of short-term autocorrelation.

Figure 2 depicts the histogram of the Mean Absolute Error (MAE) across four independent test periods (P1–P4) within the timeframe from July to October 2020. The Y-axis represents the MAE values, while the X-axis represents the observed periods. The lowest error was recorded in period P4 (MAE = 9.9), indicating the highest predictive accuracy. In contrast, P3 exhibited the highest error (MAE = 11.3), suggesting a higher data complexity or variability. The average MAE across all periods was 10.6, with a difference of 1.4 between the minimum and maximum values, confirming the model's stable performance throughout the entire observed interval.

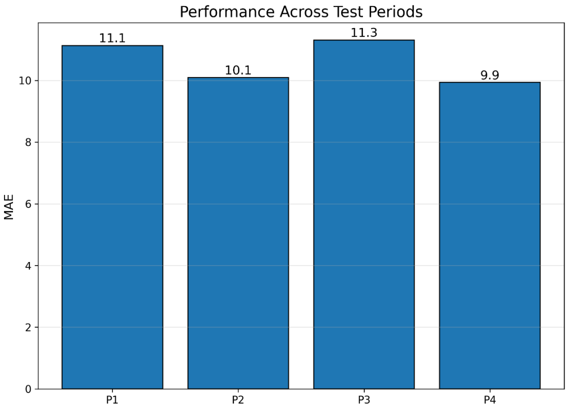

Figure 2. MAE of the model across four test periods (P1–P4), July–October 2020

Further on, Figure 3 illustrates the comparison between the actual and predicted values for the number of traffic accidents across all test periods. Each point represents a pair of actual and model-predicted values, while the dashed line (y = x) marks a perfect alignment. The shaded area indicates a ±20% interval around the ideal line, serving as a threshold for practical accuracy. The results reveal an exceptionally high degree of agreement, confirmed by a correlation coefficient r = 0.985. Most predictions fall within the ±20% zone, clearly demonstrating that the proposed model achieved both a high precision and stability across different periods.

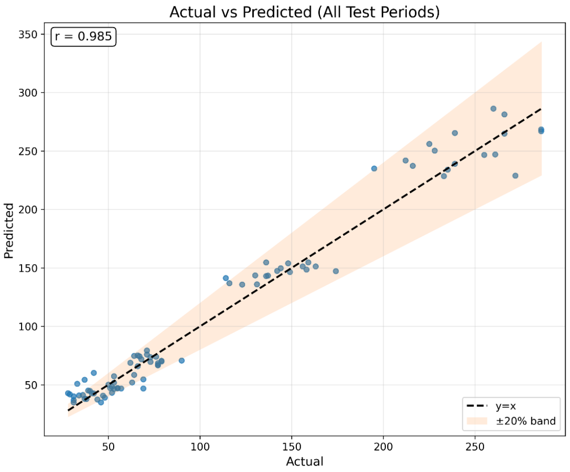

Figure 3. Actual vs. Predicted values for the number of traffic accidents (All Test Periods)

The proposed model achieved a MAE of 10.62 ± 0.61 and a R^2 of 0.967 ± 0.01, explaining over 96% of the variance in the data. The relative error amounted to 12.66% ± 1.1%, while the 20% Hit Rate reached 84.8% ± 3.6%. The best performance was recorded in September and October ($R^2 \approx 0.97–0.98$), whereas a slight decline was observed in July and August ($R^2 \approx 0.95$) due to seasonal fluctuations.

The predictions of the Mean Absolute Error by region, aggregated over all test periods, are shown in Figure 4. A considerable heterogeneity is evident: Belgrade recorded the highest error (≈32.5), followed by the South (≈30.2) and Vojvodina (≈26.4), while Central Serbia (≈13.5) and especially the East (≈10.4) obtained a much lower error. This pattern is expected because MAE is an absolute, scale-dependent metric; therefore, the regions with higher baseline accident volumes may naturally exhibit a higher MAE even when the relative prediction accuracy is similar. In urban settings (e.g. Belgrade), a higher error can additionally arise from a higher non-stationarity and unmodeled exogenous shocks (e.g. major events, roadworks, tourism peaks), which are partly smoothed yet also obscured under weekly aggregation. Moreover, regional partitioning can mix heterogeneous sub-areas and mobility regimes within the same series, further increasing the prediction difficulty. The absence of exposure variables (e.g. traffic intensity or vehicle-kilometers traveled) limits the model's ability to distinguish true changes regarding the risk of accidents due to traffic volume-driven fluctuations, contributing to the regional differences in MAE. These effects reflect structural data characteristics and aggregation-related limitations rather than a systematic regional bias of the proposed model.

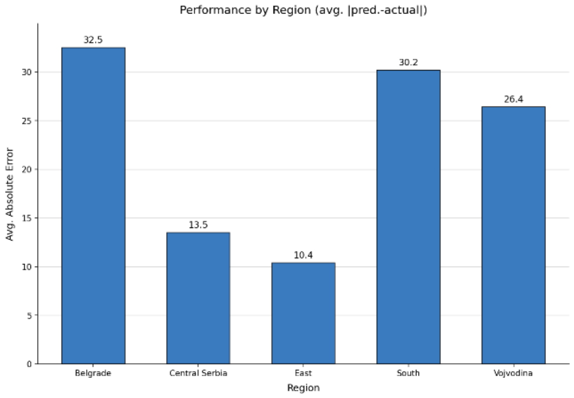

Figure 4. Regional MAE predictions

In comparison with the Naive baseline approaches, the ensemble model achieved a substantially better performance, as illustrated in Figure 5. The Mean Absolute Error (MAE) was reduced by approximately 64% relative to the Naive (Last) persistence baseline (29.1 - 10.6) and by approximately 84% relative to the Naive (Mean) baseline (66.8 - 10.6). These baseline models represent standard reference points in practical forecasting tasks and rely solely on simple assumptions of temporal persistence or global averaging. The observed improvements indicate that the proposed system does not depend only on the short-term repetition of the previous values, but it successfully learns stable temporal and structural patterns present in the data.

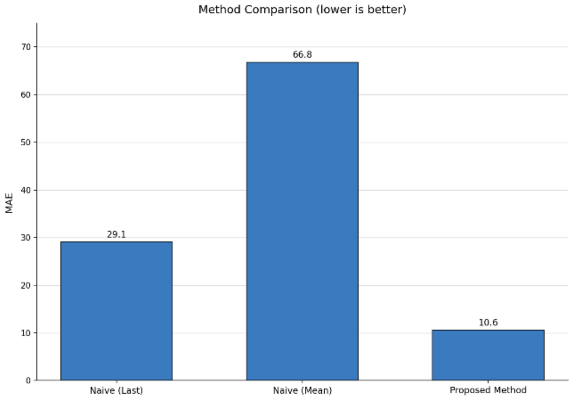

Figure 5. Comparison with Naive Baseline Models

The following section presents the concluding remarks derived from the presented findings, highlighting the scientific and practical value of the proposed system.

5. Conclusion

The proposed multi-scale ensemble for predicting the number of traffic accidents demonstrated a high accuracy and stability, achieving a MAE of 10.62 ± 0.61, a R^2 of 0.967 ± 0.011, a relative error of 12.7% ± 1.1%, and a 20% Hit Rate of 84.8% ± 3.6%. These results confirm that this system explains more than 96% of the data variance, while reducing the Mean Absolute Error by 64% and 84%, respectively in comparison with the Naive baseline approaches. This demonstrates the operational value of the developed methodological framework, which integrates strict data leakage protection, multi-scale temporal modeling, and ensemble algorithms. This model predicts the total number of accidents accurately and it also recognizes regional and seasonal specificities, making it suitable for developing targeted traffic safety strategies.

Feature importance analysis reveals the dominance of leak-safe lagged features, while calendar and seasonal variables enhance data interpretability and fine-tuning precision. At the same time, minor region-specific performance disparities (consistent over-/underprediction in some regions) were observed, suggesting that a simple post-hoc local calibration could further reduce residual errors. Limitations include the reliance on historical data aggregated at a weekly level, heuristic regional mapping, and the absence of real-time streams (traffic, meteorological conditions, roadworks), which reduce the system's ability to detect regime shifts in risk patterns at an early stage.

Future research directions could involve introducing probabilistic predictions with uncertainty intervals (e.g., quantile boosting, conformal inference), enriching the model with exogenous and granular real-time data, developing graph-based spatio-temporal architectures (ST-GNN) for modeling inter-regional interactions, and integrating online recalibration procedures to mitigate the concept drift. A particularly promising implementation could be carried out in the context of autonomous vehicles, where accident prediction models could support real-time decision-making systems, safety scenario evaluation, and an adaptive strategy design for enhancing traffic safety within the future intelligent transportation systems.

Acknowledgements

This study was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, and these results were obtained under the Grant No. 451-03-136/2025-03/ 200132 of the University of Kragujevac - Faculty of Technical Sciences Čačak.

REFERENCES

Ahmad, N., Wali, B. & Khattak, A.J. (2023) Heterogeneous ensemble learning for enhanced crash forecasts – A frequentist and machine learning based stacking framework. *Journal of Safety Research*. 84, 418–434. https://doi.org/10.1016/j.jsr.2022.12.005.

Ara, Z. & Hashemi, M. (2021) Identifying the severity of road accident impact on traffic flow by ensemble model. In: *Proceedings of the 2021 IEEE 22nd International Conference on Information Reuse and Integration for Data Science (IRI)*, *10-12 August 2021, Las Vegas, USA*. New York, USA, IEEE. pp. 115–122. https://doi.org/10.1109/iri51335.2021.00022.

Behboudi, N., Moosavi, S. & Ramnath, R. (2024) *Recent advances in traffic accident analysis and prediction: A comprehensive review of machine learning techniques*. [Preprint] https://arxiv.org/abs/2406.13968 [Accessed 30th September 2025].

Chai, A.B.Z., Lau, B.T., Tee, M.K.T. et al. (2024) Enhancing road safety with machine learning: Current advances and future directions in accident prediction using non-visual data. *Engineering Applications of Artificial Intelligence*. 137(A), Art. ID 109086. https://doi.org/10.1016/j.engappai.2024.109086.

Ćirković, S., Blagojević, M., Ilić, M. et al. (2025) Software Solution for Visualization, Clustering, and Filtering of Traffic Accidents. In: *Proceedings of the LXIX Conference ETRAN*, *9–12 June 2025*, *Čačak, Serbia*. pp. 1–4. https://doi.org/10.69994/69E25048.

Ćirković, S., Iričanin, A. & Blagojević, M. (2025) *Advanced multi-scale temporal pattern recognition for regional traffic accident prediction. GitHub repository*. https://github.com/cirkovicstefan123/Advanced-Multi-Scale-Temporal-Pattern-Recognition-for-Regional-Traffic-Accident-Prediction [Accessed 30th September 2025].

Deretić, N., Stanimirović, D., Awadh, M.A. et al. (2022) SARIMA modelling approach for forecasting of traffic accidents. *Sustainability*. 14(8), Art. ID 4403. https://doi.org/10.3390/su14084403.

Ding, W., Xia, Y., Wang, Z. et al. (2020) An ensemble-learning method for potential traffic hotspots detection on heterogeneous spatio-temporal data in highway domain. *Journal of Cloud Computing*. 9, Art. ID 170. https://doi.org/10.1186/s13677-020-00170-1.

Gan, J., Li, L., Zhang, D. et al. (2020) An alternative method for traffic accident severity prediction: Using deep forests algorithm. *Journal of Advanced Transportation*. 2020, Art. ID 1257627. https://doi.org/10.1155/2020/1257627.

Gatarić, D., Ruškić, N., Aleksić, B. et al. (2023) Predicting road traffic accidents - Artificial neural network approach. *Algorithms*. 16(5), Art. ID 257. https://doi.org/10.3390/a16050257.

Li, Y., Yu, R., Shahabi, C. et al. (2018) Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In: *Sixth International Conference on Learning Representations (ICLR)*, *30 April - 3 May 2018, Vancouver, Canada*.

Lord, D. & Mannering, F. (2010) The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice*. 44(5), 291–305. https://doi.org/10.1016/j.tra.2010.02.001.

Muktar, B. & Fono, V. (2024) Toward safer roads: Predicting the severity of traffic accidents in Montreal using machine learning. *Electronics*. 13(15), Art. ID 3036. https://doi.org/10.3390/electronics13153036.

Ramani, R.G. & Selvaraj, S. (2014) A pragmatic approach for refined feature selection for the prediction of road accident severity. *Studies in Informatics and Control*. 23(1), 41–50. https://doi.org/10.24846/v23i1y201405.

Road Traffic Safety Agency of the Republic of Serbia. (2025) *Homepage*. https://www.abs.gov.rs/rsc [Accessed 30th July 2025].

Umer, M., Sadiq, S., Ishaq, A. et al. (2020) *Comparison analysis of tree based and ensembled regression algorithms for traffic accident severity prediction*. [Preprint] https://arxiv.org/abs/2010.14921 [Accessed 30th September 2025].

Yu, Y. & Chen, Z. (2025) Ensemble learning-based prediction of road traffic accident severity. In: *Proceedings of the International Conference on Smart Transportation and City Engineering (STCE 2024)*, *6-8 December 2024*, *Chongqing, China* (*Proceedings of SPIE*, vol. 13575). Bellingham, WA, USA, SPIE. p. 13. https://doi.org/10.1117/12.3060632.

Zhao, L., Song, Y., Zhang, C. et al. (2020) T-GCN: A temporal graph convolutional network for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems*. 21(9), 3848–3858. https://doi.org/10.1109/tits.2019.2935152.

Zhao, Y. & Deng, W. (2022) Prediction in traffic accident duration based on heterogeneous ensemble learning. *Applied Artificial Intelligence*. 36(1), Art. ID 2018643. https://doi.org/10.1080/08839514.2021.2018643.