https://doi.org/10.59865/abacj.2024.70

LEVERAGING MACHINE LEARNING TO ENHANCE ROAD SAFETY: A SOCIAL MARKETING APPROACH

Wasinee Noonpakdee[1], Manit Satitsamitpong[2,*], Prarawan Senachai[3], and Kittipong Napontun[4]

Abstract

Road safety remains a critical concern worldwide. This research aims to investigate the use of Explainable AI (XAI) techniques, particularly SHAP (Shapley Additive Explanations), to identify key factors influencing road accident severity and create a social marketing campaign encouraging people to change their behavior in relation to road safety. Several machine learning models were developed using data from Thailand's Ministry of Transport. The results show that the Light Gradient Boosting Machine (LGBM) model achieved the highest accuracy of 0.85 and an F1 score of 0.83. SHAP analysis revealed that the most significant contributing factors were the number of motorcycle involvements, road code, and the total number of vehicles and people involved in the accident. A practical framework for promoting sustainable road safety was proposed, focusing on raising awareness, delivering emotionally impactful communication, and fostering immediate behavioral change. This research provides valuable insights for strategic road safety initiatives and demonstrates the effectiveness of integrating machine learning with XAI. The findings can guide government authorities, policymakers, insurance companies, and social marketing planners in improving road safety.

Keywords: Road accidents, Explainable AI, XAI, SHAP, Severity Prediction, Social Marketing, Insurance Industry

[1] Asst. Prof. Dr. Wasinee Noonpakdee (First Author) is currently working as a lecturer in the Master of Digital Business Transformation Program at Thammasat University, Thailand. She received her B.E. degree in Electrical Engineering from Chulalongkorn University, Thailand, M.ISM. degree in Information Systems Management from Carnegie Mellon University, USA, and D.Sc. degree in Information and Telecommunications from the Graduate School of Global Information and Telecommunication Studies (GITS), Waseda University, Japan. Her research interests include optical wireless communications, business intelligence, data analysis, data visualization, machine learning and data mining. Email: wasinee@citu.tu.ac.th

[2,*] Dr. Manit Satitsamitpong (Corresponding Author) is a lecturer in the Master of Digital Business Transformation Program at Thammasat University, Thailand. He holds a bachelor's degree in Computer Science and Engineering from the University of Pennsylvania, an MBA from the University of Illinois, and an MS in Information Technology Management from Georgia Institute of Technology. He received his doctoral degree in Global Information and Telecommunication Studies at Waseda University, Japan. His research interests are in economics and policy issues on digital technology utilization in developing nations as well as in business and technical issues on the impact of digital technology within an organizational context. Email: manit@citu.tu.ac.th

[3] Asst. Prof. Dr. Prarawan Senachai (Essentially Intellectual Contributor) is a lecturer in the Department of Marketing at the Faculty of Business Administration and Accountancy at Khon Kaen University in Thailand . She obtained her doctoral degree in Marketing Communication from the Faculty of Arts and Design at the University of Canberra, Australia. Her research interests include Communications & Media, Customer Relationship Management, Service Marketing, and research related to marketing. Email: prarse@kku.ac.th

[4] Kittipong Napontun (Co-Author) is currently an assistant researcher in the Consumer Insights in Sports or Service-Related Business Research Unit at Chulalongkorn University. He is also pursuing a master's degree in Branding and Marketing at Chulalongkorn Business School. Email: Kittipong.n51@gmail.com

1. INTRODUCTION

Each year, road accidents cause millions of fatalities, with significant socio-economic consequences. According to the Global Status Report on Road Safety (2023), there were an estimated 1.19 million road traffic deaths worldwide in 2021, a slight reduction from the previous decade. Despite this modest improvement, road traffic injuries remain the leading cause of death among children and young people aged 5 to 29 years (World Health Organization, 2023). Regarding Thailand, the Global Status Report on Road Safety (2023) revealed that the country had a road traffic death rate of 25.4 per 100,000 population in 2021, which remains one of the highest in Asia and among upper-middle-income nations (World Health Organization, 2023).

A number of factors, including road conditions, characteristics of vehicles, human behavior, and environmental factors can be driving causes for road accidents (Ahmed et al., 2023; Viswanath et al., 2021). The broader impacts of road accidents extend beyond fatalities, as they impose substantial financial and socio-economic burdens on families, communities businesses and governments. The insurance industry is one of the most significantly impacted business sectors, as road accidents directly influence their financial risks and operational costs (Kiptoo et al., 2021) . In order to mitigate the effects of road accidents, These include updating road safety policies, many countries have implemented measures such as adopting new technologies, and improving traffic law enforcement mechanisms (World Health Organization, 2023). Accurate analysis and understanding of the factors contributing to road accidents is essential for designing safer roads and developing more effective policies. For example, the integration of systems science into social marketing design (Flaherty et al., 2020), can lead to the development of a social marketing plan which promotes voluntary behavioral changes within the target audience that benefit the entire population (Truong, 2014).

Some research has focused on the integration of systems science, developing predictive models to address road accident complexities (Isler et al., 2024; Viswanath et al., 2021). Recently, technological advancements have meant that machine Learning (ML) approaches have become increasingly valuable in analyzing road accident data. It is generally recognized that predictive modeling has been useful in identifying accident severity and contributing factors. However, many machine learning and deep learning models lack transparency. This “black box” nature raises concerns which could reduce stakeholder trust (Ersöz et al., 2022; Kumar et al., 2024). To address this issue, some researchers have developed predictive models using various techniques or applied Explainable AI (XAI) methods, such as SHAP, to reveal the most influential features in accident prediction. These approaches improve predictive accuracy and facilitate identification of the main factors influencing the predictive model. Contributing factors are typically categorized into road characteristics, vehicle conditions, human behavior, and speed limits (Ahmed et al., 2023). By improving understanding of how different factors influence accident prediction, researchers can aid the development of more effective strategies for accident prevention (Aboulola et al., 2024).

Significant progress has been made in enhancing road safety through strategic initiatives, including road audits, stricter enforcement of vehicle safety standards, and improvements in emergency care systems. Regarding accident prediction, Ratanawimon and Tanawongsuwan (2024) investigated several machine-learning algorithms to predict the severity of road traffic accidents in Thailand, focusing on the Extreme Gradient Boosting (XGBoost) algorithm. Their study compared XGBoost’s performance against other algorithms such as Random Forest, Bagging, Decision Tree, and Multilayer Perceptron, concluding that XGBoost outperformed the others, especially in predicting fatal accidents. While previous models have achieved strong results, there remains a need for more explainable models to deepen our understanding of the factors driving road accident severity in Thailand.

Within the field of social marketing, behavioral change is known to occur within a complex social context, making it essential to assess the impacts on multiple stakeholders in order to fully evaluate the effectiveness of any intervention (Buyucek et al., 2016). Unfortunately, studies on certain key stakeholders are missing (Troy et al., 2015), especially in the insurance industry, as road accidents directly influence the financial risks and operational costs within this industry sector (Kiptoo et al., 2021). Focusing solely on the target audience in an evaluation provides a limited perspective and may not adequately reveal how social marketing interventions can be improved to optimize behavioral outcomes (Buyucek et al., 2016). By incorporating a broader stakeholder assessment can better ensure sustained engagement and long-term success, thereby enhancing the longevity of social marketing interventions for reducing road accidents in Thailand.

This paper aims to explore the role of explainable machine learning models for predicting road accident severity in Thailand, focusing on identifying the most significant contributing factors and enhancing model transparency. Understanding these factors will enable the creation of more targeted interventions, ultimately reducing the impact of road traffic accidents in Thailand, while facilitating multilevel, systemic change to transform societies (Flaherty et al., 2020).

2. LITERATURE REVIEW

2.1 Road Accidents and Machine Learning

The fatalities and injuries caused by road traffic accidents continue to pose a significant global health and development challenge. In 2021, there were an estimated 1.19 million road traffic deaths worldwide. While this is a step in the right direction, marking a 5% decrease from 2010, despite substantial increases in vehicle numbers, road networks, and population growth (World Health Organization, 2023), it is clear that further measures must be taken to reduce this number further. When considering the consequences of road accidents it is necessary to emphasize the growing societal impact and the urgent need for effective preventive measures to reduce or eliminate these incidents (Venkatesh Raja et al., 2023).

In response to this challenge, numerous studies have explored the potential role of machine learning and predictive models in mitigating road traffic accidents, and reducing the associated number of deaths and injuries (Berhanu et al., 2024; Obasi & Benson, 2023; Shaik et al., 2021). Several studies have identified important factors contributing to traffic accidents, such as vehicle operation and time of day. For example, Obasi and Benson (2023) achieved an 87% accuracy rate with Random Forest and Logistic Regression models when analyzing UK traffic data from 2005 to 2014. Chen et al. (2024) explored accident duration prediction through a multimodal deep learning architecture that combined structured and text data for improved accuracy. Additionally, innovative methodologies have been employed, including the Apriori algorithm and Support Vector Machines for improvement of road designs in Bangalore (Viswanath et al., 2021), as well as fuzzy comprehensive evaluation techniques to address data uncertainty in predictions of freeway accidents (Wang et al., 2024). Certain prior studies have also focused on specific challenges, such as Isler et al. (2024) who developed predictive models for crash occurrences based on factors relating to the urban traffic network in São Paulo, while Berhanu et al. (2024) combined Random Forest with spatial network analysis to predict crash likelihood with 78% accuracy. Furthermore, Zhao et al. (2019) improved intelligent transportation systems by developing the AdaBoost-SO algorithm to manage class imbalance, and Zhankaziev et al. (2022) explored real-time vehicle collision prediction methodologies using transport detector data. These studies imply the value of machine learning (ML) in analyzing road accident data, with many studies framing accident prediction as a classification problem (Ahmed et al., 2023; Obasi & Benson, 2023).

2.2 Social Marketing and the Insurance Industry

Social marketing has gained significant attention across various disciplines globally (Truong & Hall, 2013), primarily for encouraging voluntary behavioral changes which benefit the greater good (Truong, 2014). Social marketing moves beyond addressing downstream, midstream, or upstream audiences and interventions in isolation, instead focusing on a coordinated, multilevel, and systemic change. This approach is able to transform communities, organizations, or societies, and the broader world (Flaherty et al., 2020). Social marketing emphasizes that behavioral change is closely linked to social and systemic transformations, as we are all interconnected within mutually dependent communities (Flaherty et al., 2020). To achieve such transformations, social marketing integrates the strengths of diverse methodologies, including qualitative approaches such as content, sentiment, and narrative analysis, and quantitative methods such as structural equation modeling (Bryant et al., 2000). This includes marketing research, product or program development, incentives, and facilitation (Fox & Kotler, 1980). The effectiveness of social marketing has become evident through various case studies, including in the context of transportation (Cooper, 2007; Fox & Kotler, 1980).

The insurance industry is one business sector which has been significantly impacted by road accidents. This is due to the fact that such incidents not only affect the financial risks of insurance companies, but also increase operational costs and claims management expenses (Spilbergs et al., 2022). In the context of Thailand, which has a high rate of road traffic fatalities (World Health Organization, 2023), the industry must implement clear measures to manage the risks posed by road accidents more effectively. Research over recent years has highlighted the crucial role of social marketing in changing consumer behavior, particularly in industries related to road safety, such as the insurance sector (Roger et al., 2023). In this context, social marketing focuses on raising awareness, changing attitudes, and encouraging behavioral changes among a target audience, specifically promoting safer driving practices to reduce the risk of accidents (Barrie et al., 2011). Use of social marketing in the insurance industry could change driver behavior which would in turn positively affect both the industry and society as a whole (Plant et al., 2017). Developing social marketing campaigns which emphasize safe driving could reduce the number of insurance claims, lower operational costs, and foster positive relationships between consumers and insurance companies. Implementing such an approach in Thailand is particularly crucial, given the high rate of road accidents (Elias, 2021), ultimately helping Thai society to become a better place.

2.3 The Knowledge, Attitude, and Practice (KAP) Theory and Social Marketing

The Knowledge, Attitude, and Practice (KAP) theory is a critical conceptual framework for researching and evaluating human behavior, particularly in the context of health and social development (Muleme et al., 2017; Zheng et al., 2021). This theory emphasizes the importance of understanding the relationships between individuals' or groups' knowledge, attitudes, and practices, to develop effective strategies for behavioral change (Valente et al., 1998). In social marketing, KAP theory is employed to design and evaluate programs aimed at changing the behavior of a specific target audience, not only by providing information or raising awareness but also by adjusting attitudes and encouraging desired practices (Wang et al., 2009). Traditionally, social marketing has focused on downstream interventions, such as providing knowledge to individuals; however, this targeted approach alone has been found to be insufficient for creating sustainable change (Crawshaw, 2014). Therefore, applying KAP theory alongside interventions which encompass upstream, midstream, and downstream target groups is essential for effective change.

Interventions implemented at multiple levels mean actions that target individuals and encompass whole communities, organizations, or public policies (Muleme et al., 2017). Coordinating efforts across these levels can help to establish an environment that supports behavioral change, such as through policy reform, creating infrastructure conducive to desired behaviors, or reinforcing community support (Valente et al., 1998; Wang et al., 2009). This approach aligns with the systems change concept, which considers that individual behaviors are influenced by multiple factors at both personal and societal levels (Matsumoto, 2007). Additionally, coordination between levels enhances the likelihood of achieving sustainable, widespread change. Social marketing which utilizes KAP theory in this way, not only focuses on individual behavioral change, but also fosters systemic transformation, which can have long-term impacts on the overall well-being of society (Senachai et al., 2022; Wang et al., 2009).

2.4 Explainable AI (XAI)

Modern machine learning algorithms, especially deep learning models, often achieve exceptional performance but lack transparency in their predictions. However, understanding the rationale behind a model's decisions is essential for building trust among users and stake-holders as well as ensuring that the model operates securely and reliably. As a result, researchers are increasingly focused on developing techniques for improved explainability to make these models more interpretable and their decisions more understandable (Javeed et al., 2024).

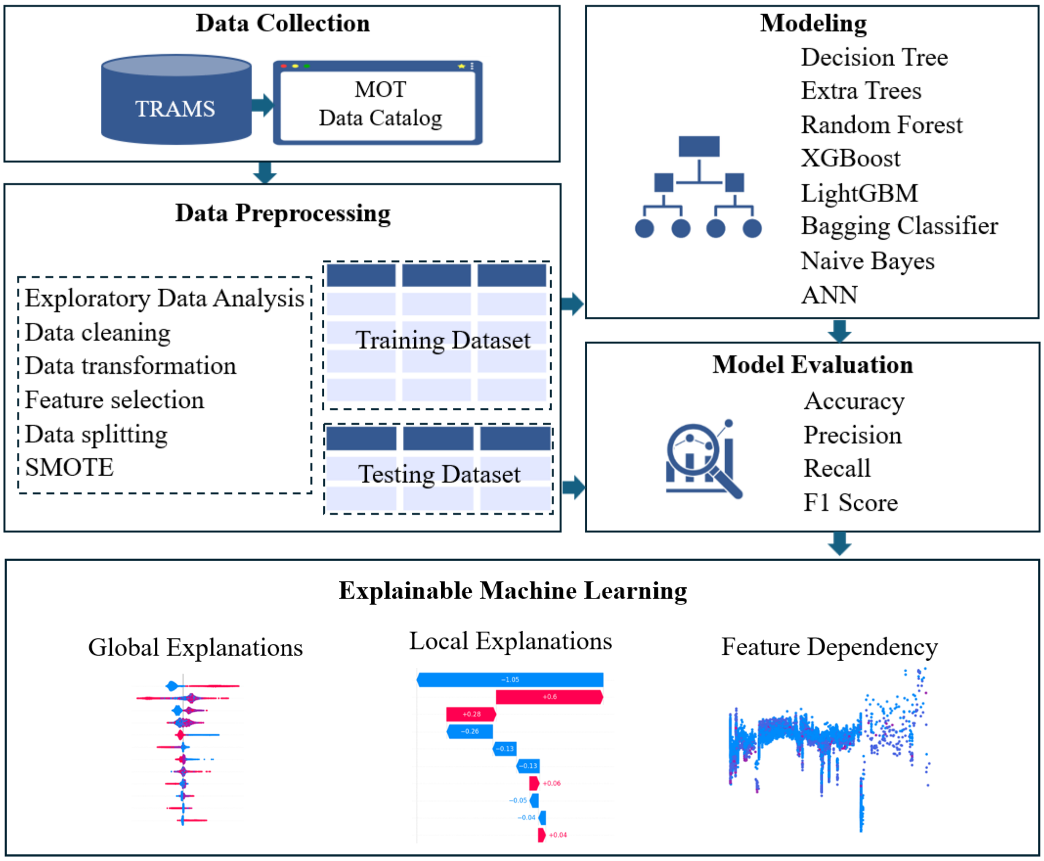
Explainable AI (XAI) has been developed to handle challenges such as transparency, interpretability, and explainability, in complex machine learning models (Kumar et al., 2024). Shapley Additive Explanations (SHAP) is one of the most popular XAI techniques used for model explainability. SHAP leverages Shapley values, which represent the average marginal contribution of a feature across all possible coalitions, providing accurate and consistent explanations (Ahmed et al., 2023). SHAP values quantify the contribution of each feature to a model's output by considering all possible feature combinations in order to provide both local and global explanations of model predictions (Kok et al., 2023). Regarding local interpretability, SHAP analyzes how individual features influence specific predictions, offering deeper insights than traditional global feature importance techniques. SHAP can enhance the transparency of ensemble models through visual and intuitive representations, such as force visualizations, ultimately making complex results accessible to non-technical users (Sahlaoui et al., 2021).

SHAP has proven to be an essential tool for explaining machine learning models applied to road accident data, assisting in interpretation of results, and identification of contributing factors, ultimately informing the development of data-driven interventions for improving road safety. Many studies have employed SHAP to improve model explainability. For instance, Aboulola et al. (2024) utilized SHAP with transfer learning using MobileNet to create a predictive model for traffic accident severity with a US dataset covering the period 2016 to 2021. Similarly, Ahmed et al. (2023) analyzed accident severity with ensemble machine learning models applied to data from New Zealand (2016-2020). Here SHAP was employed to interpret the Random Forest model, which achieved the highest accuracy at 81.45%. The study identified road category and vehicle count as significant contributors to injury severity. Meanwhile, Sufian et al. (2024) combined SHAP with traditional statistical methods, including logistic regression and econometric models, to study accident severity in the UK. Here, the integration of SHAP allowed for the interpretation of influential factors, such as the location of drivers and casualties, thus supporting targeted, evidence-based road safety policies. Additionally, Pei et al. (2024) developed a CNN-BiLSTM-Attention model to predict road traffic accident risks, utilizing DeepSHAP for their result interpretation. This innovative approach allowed for the identification of key factors influencing accident severity, improving model performance, and minimizing iteration time, without compromising accuracy.

3. RESEARCH METHODOLOGY

As illustrated in Figure 1, the research process consists of five stages: Data Collection, Data Preprocessing, Modeling, Model Evaluation, and Explainable Machine Learning.

Figure 1 Research Process



3.1 Data Collection

This research utilized a dataset obtained from Thailand's Ministry of Transport (Ministry of Transport, 2024), specifically from the MOT Data Catalog, which connects to the Transport Accident Management System (TRAMS). The dataset includes road traffic accidents that occurred on national highways, rural roads, and expressways in Thailand from January 1, 2020, to August 31, 2024. It comprises 101,948 records and contains 36 attributes.

3.2 Data Preprocessing

For data preprocessing, Exploratory Data Analysis (EDA) was first conducted to analyze and summarize the dataset's structure and uncover patterns. Next, the data cleaning process involves several key steps: removing duplicate records to prevent bias in model training, eliminating outliers, correcting inconsistencies in categorical values, and cross-checking the data with external systems such as the Transport Accident Management System (TRAMS) to verify accuracy and completeness.

During data transformation, a new target variable, "Severity," was created by categorizing the data into two classes: "Serious" and "Non-Serious". Records with one or more fatalities or serious injuries were labeled as "Serious," while those with only minor injuries or no injuries were classified as "Non-Serious". Additionally, the original "Presumed Cause" attribute was split into three new attributes: "Driver Factors," "Vehicle Issues," and "Road Conditions". Label Encoding was then applied to convert categorical data into numerical values, assigning integer values to each category.

For feature selection, relevant features identified from previous studies (Ahmed et al., 2023; Isler et al., 2024; Obasi & Benson, 2023; Viswanath et al., 2021) were selected to ensure that only valuable predictors were included. This approach reduces the dimensionality of the dataset and minimizes the risk of overfitting. A total of 23 features were selected as independent variables for analysis, covering aspects such as the date, time, conditions during the accident, Road Code (which indicates specific roads, e.g., Motorway Route 7 and National Highway No. 12), accident location, characteristics, weather conditions, road conditions, and the number of various vehicle types (e.g., motorcycles, private cars, trucks). These features were used to predict the dependent variable (class or label), severity, categorizing accidents as "Non-Serious" (0) or "Serious" (1).

The dataset was then split into training and testing sets using an 80:20 ratio. The dataset included 19,793 "Serious" accidents and 82,155 "Non-Serious" accidents, creating a significant class imbalance. To address this, the Synthetic Minority Over-sampling Technique (SMOTE) was employed. SMOTE generates synthetic samples for the minority class—in this case, the "serious" accidents—by interpolating between existing examples.

3.3 Modeling

In this step, the training dataset was used to train several machine learning models for severity prediction, including Decision Tree, Extra Trees, Random Forest, XGBoost, Light Gradient Boosting Machine (LGBM), Bagging Classifier (Bootstrap aggregating), Naive Bayes, and Artificial Neural Networks (ANN).

3.4 Model Evaluation

The performance of the models was assessed using several metrics, including accuracy, precision, recall, and F1 Score. Accuracy indicates the proportion of correct classifications made by the classifier. Precision measures how many of the predicted positive cases are actually positive. Recall measures how well the classifier correctly identifies all actual positive cases. The F1 Score provides a balance between precision and recall. The formulas for calculating these metrics use the following definitions: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) (Obasi & Benson, 2023).

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1\ Score = 2 \text{ x } \frac{Presition\ x\ Recall}{Presision + Recall} \tag{4}$$

3.5 Explainable Machine Learning

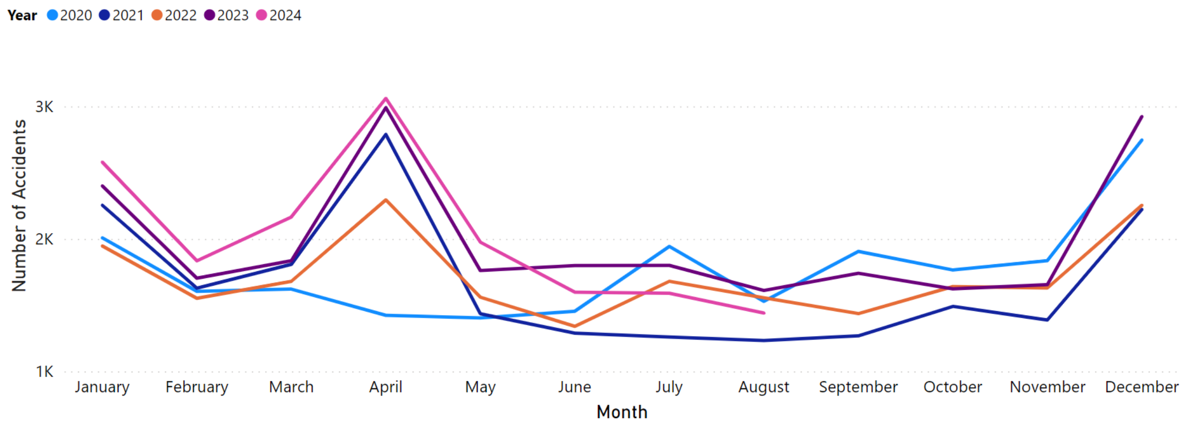
For explainable machine learning, SHAP was utilized including global explanations, local explanations, and feature dependencies. This helped identify key features influencing accident severity predictions.

4. RESULTS

The Exploratory Data Analysis (EDA) shows the monthly accident trends from January 2020 to August 2024, as illustrated in Figure 2. The data reveals similar patterns for each year, with a significant increase in accidents during April, which aligns with the Thai New Year (Songkran Festival) and around New Year celebrations. However, in April 2020, there was a sharp decline in accidents due to the nationwide lockdown imposed by the COVID-19 pandemic. The lockdown reduced travel and public movement, resulting in a significant drop in accident rates during that period.

Due to this unusual event, the analysis was conducted on two groups of datasets to assess its impact. The first group includes the year 2020 (data from January 1, 2020, to August 31, 2024), which reflects the pandemic's influence on accident patterns, while the second group excludes 2020 (data from January 1, 2021, to August 31, 2024), offering a more typical trend unaffected by the pandemic. A comparison between these two datasets was then performed to determine which group provides more consistent insights into accident trends and to identify any potential performance differences in predictive analysis.

Figure 2 Number of Accidents



4.1 Results of Model Evaluation

This section presents the results of the model evaluation, comparing datasets that include and exclude the impact of the pandemic (year 2020) and evaluating the effect of applying SMOTE.

Table 1 Results of Dataset that Includes the year 2020 Without SMOTE

Model	Accuracy	Precision	Recall	F1 Score
LGBM	0.843306	0.825703	0.843306	0.823890
XGBoost	0.840106	0.822054	0.840106	0.823123
Random Forest	0.834679	0.814343	0.834679	0.815683
Extra Trees	0.828347	0.807430	0.828347	0.811239
Bagging Classifier	0.826955	0.804453	0.826955	0.807853
Naive Bayes	0.801489	0.774586	0.801489	0.783085
ANN	0.787364	0.780482	0.787364	0.783699
Decision Tree	0.772892	0.776007	0.772892	0.774412

Table 1 represents the model performance when the dataset includes data from the year 2020, without applying the SMOTE technique to address class imbalance. The LGBM model demonstrates the best performance in terms of accuracy (0.8433) and F1 score (0.8239), followed by XGBoost and Random Forest.

Table 2 shows how applying SMOTE to balance the classes impacts model performance when the dataset still includes 2020 data. Both XGBoost and LGBM achieve similar results, leading with an accuracy of 0.8321, but the scores are slightly lower compared to the same dataset without SMOTE (Table 1). The Naive Bayes model suffers considerably yielding the lowest accuracy (0.5019).

Table 2 Results of the Dataset Including the Year 2020 with SMOTE

Model	Accuracy	Precision	Recall	F1 Score
XGBoost	0.832104	0.820663	0.832104	0.824847
LGBM	0.832104	0.820777	0.832104	0.824940
Random Forest	0.825911	0.811247	0.825911	0.816216
Extra Trees	0.818606	0.804229	0.818606	0.809541
Bagging Classifier	0.812135	0.795429	0.812135	0.801523
ANN	0.785138	0.751169	0.785138	0.762959
Decision Tree	0.764542	0.776350	0.764542	0.769970
Naive Bayes	0.501879	0.737318	0.501879	0.550671

Table 3 illustrates the performance of models trained on a dataset excluding the year 2020, without SMOTE. LGBM again demonstrates the highest accuracy (0.8510), with slightly better performance across all metrics compared to the dataset including 2020 (Table 1).

Table 3 Results of the Dataset Excluding the Year 2020 Without SMOTE

Model	Accuracy	Precision	Recall	F1 Score
LGBM	0.851018	0.833135	0.851018	0.831804
XGBoost	0.848135	0.829752	0.848135	0.830830
Random Forest	0.840975	0.819670	0.840975	0.821753
Bagging Classifier	0.838092	0.816915	0.838092	0.820527
Extra Trees	0.836511	0.814785	0.836511	0.818686
ANN	0.821259	0.780804	0.821259	0.782538
Naive Bayes	0.810193	0.782260	0.810193	0.791279
Decision Tree	0.779503	0.782762	0.779503	0.781096

Table 4 examines the dataset excluding 2020 but with SMOTE applied. The accuracy of the models slightly decreases compared to Table 3 (without SMOTE), with LGBM still achieving the highest accuracy (0.834).

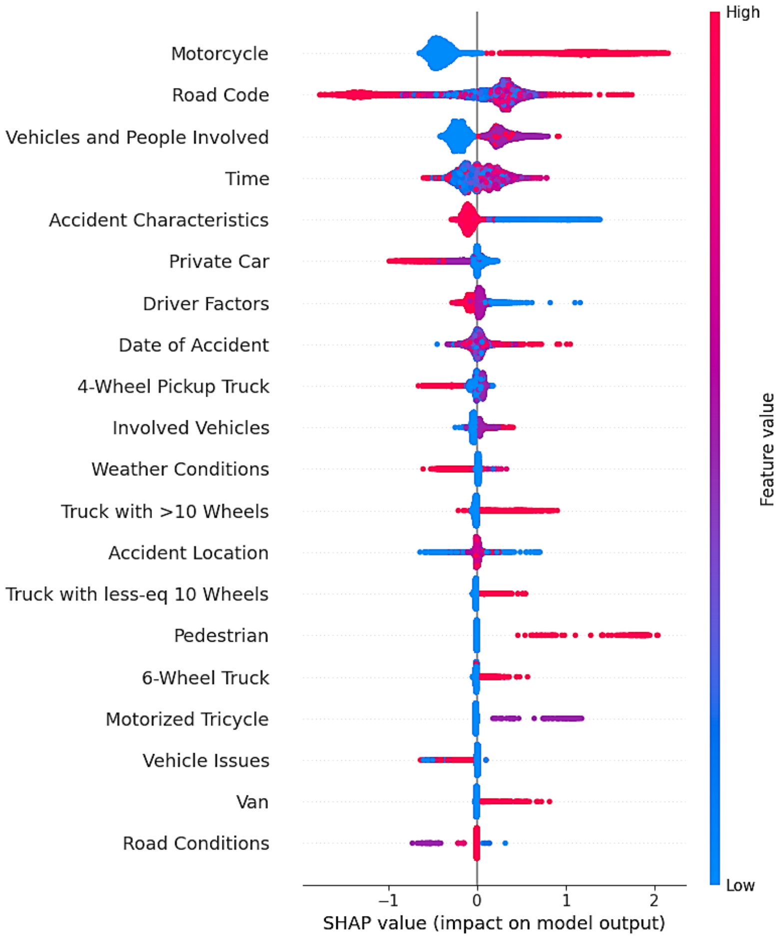
Table 4 Results of the Dataset Excluding the Year 2020 with SMOTE

Model	Accuracy	Precision	Recall	F1 Score
LGBM	0.834000	0.820646	0.834000	0.825521
XGBoost	0.832884	0.819054	0.832884	0.824071
Random Forest	0.828141	0.812847	0.828141	0.818379
Extra Trees	0.821631	0.806619	0.821631	0.812398
Bagging Classifier	0.820050	0.802100	0.820050	0.808609
ANN	0.790663	0.751596	0.790663	0.765638
Decision Tree	0.764903	0.777214	0.764903	0.770618
Naive Bayes	0.507579	0.752769	0.507579	0.560496

4.2 Results of Explainable Models

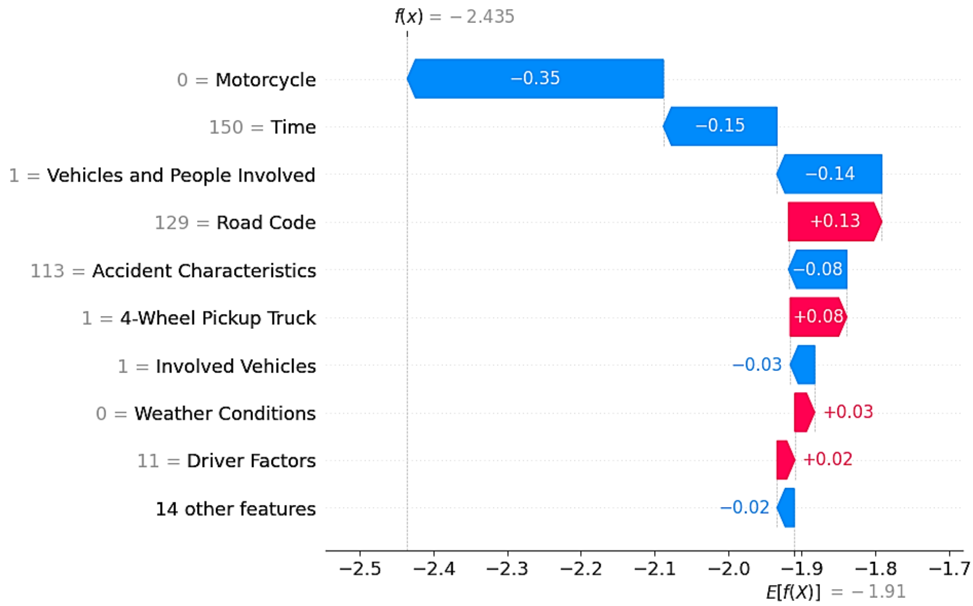
Based on the model evaluation, the optimal choice is the LGBM model without SMOTE and with the exclusion of anomalous data from 2020. Therefore, selecting this model for SHAP analysis would be advantageous, as it is more likely to provide valuable insights into feature importance and model behavior.

Figure 3 LGBM Summary Plot

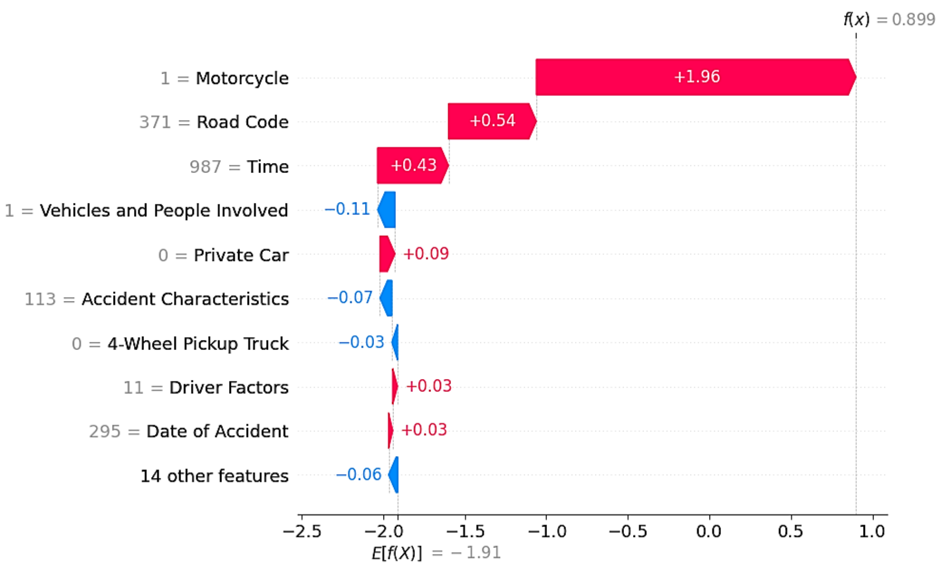


For global explanations, the SHAP summary plot in Figure 3 provides an overview of how various features contribute to the LGBM model's predictions. According to Figure 3, the most influential features are Motorcycle, Road Code, and Vehicles and People Involved, respectively. For Motorcycles, higher values (shown in red) correspond to higher SHAP values, indicating that a greater presence of motorcycles in an accident increases the likelihood of a severe outcome. For Road Code, the mix of blue and red dots suggests that different road codes (representing specific roads where accidents occur) have varying impacts on accident severity; some roads are more prone to serious accidents, resulting in positive SHAP values. Regarding Vehicles and People Involved, red dots signify that a higher number of vehicles and people involved in an accident shifts the prediction towards higher severity, while blue dots, indicating fewer vehicles and people, are linked to lower severity predictions.

Figure 4 Individual SHAP Waterfall Plot for Observations of LGBM for (a) Instant Index = 1000 (b) Instant Index = 4000

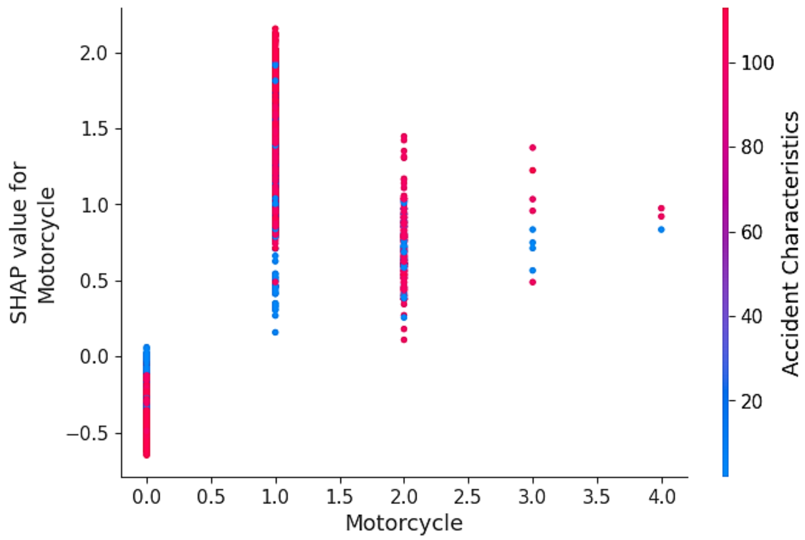


(a) instant index = 1000

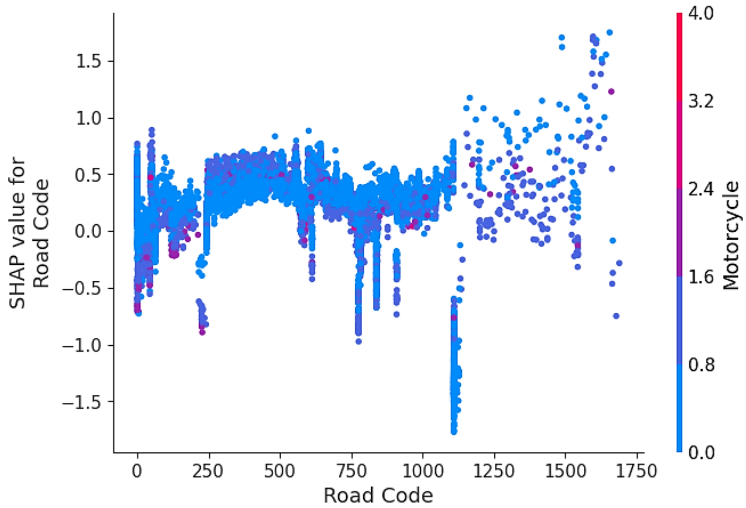


(b) instant index = 4000

Figure 5 Feature Dependency Analysis Using SHAP LGBM (a) Motorcycle (b) Road Code



(a)



(b)

Figure 4 shows the SHAP waterfall plots, which illustrate how specific features contribute to the prediction outcome for two instances (indexed as 1000 and 4000) in the dataset. For Instant Index = 1000 (Figure 4 (a)), certain features such as Motorcycle (= 0) have a negative SHAP value (-0.35 for Motorcycle), indicating a decrease in the prediction towards the non-serious class (0). In contrast, features such as Road Code, with a SHAP value of +0.13, slightly increase the likelihood of the prediction leaning toward class 1 (serious). Overall, the plot depicts how each feature influences the prediction for this specific instance, with a final predicted value of -2.435, indicating a non-serious accident outcome. For Instant Index = 4000 (Figure 4 (b)), the feature Motorcycle (= 1) has a large positive SHAP value of +1.96, making it the most significant factor pushing the prediction towards class 1 (serious). The final SHAP-based prediction value of +0.899 suggests a higher probability of classifying this instance as a serious accident.

Figure 5 presents the SHAP Feature Dependency Analysis, offering a deeper look into how specific features (Motorcycle and Road Code) interact with the model's predictions throughout the dataset. For the Motorcycle Dependency Plot (Figure 5 (a)), as the "Motorcycle" feature value increases, the SHAP value also tends to increase, indicating a higher likelihood of predicting a serious accident when more motorcycles are involved. The clustering of red points (high accident characteristics) alongside high SHAP values suggests that certain accident conditions combined with motorcycle involvement contribute significantly to predicting serious outcomes. For the Road Code Dependency Plot (Figure 5 (b)), the SHAP values vary with different "Road Code" values, suggesting that certain road locations are more strongly associated with severe accidents. The colors indicate different levels of "Motorcycle" involvement, showing how the combination of road codes and motorcycle presence affects severity predictions.

5. DISCUSSION

Findings from the Exploratory Data Analysis (EDA) indicate seasonal trends, with an increase in accidents during the Thai New Year (Songkran Festival) and the general New Year period. However, 2020 presented an anomaly due to the COVID-19 pandemic, which resulted in nationwide lockdowns and curfews from March 26, 2020, to May 31, 2020 (Wetchayont, 2021). This anomaly highlights the importance of considering external factors, which can impact predictive accuracy.

When addressing class imbalance with SMOTE, the study's performance metrics show that models trained without SMOTE achieve better accuracy and F1 scores. These results contradict the findings of Ahmed et al. (2023) who observed improved model accuracy with SMOTE. The discrepancy may be because the synthetic samples generated by SMOTE do not fully capture the patterns of the minority class. This aligns with the work of Elreedy et al. (2024) who found that synthetic data generated by SMOTE may not accurately reflect the true distribution of the minority class, which can affect classification performance. Additionally, SMOTE can lead to overfitting, as the synthetic samples are based on existing minority class samples. These new samples may not generalize well, resulting in biased models that perform poorly on unseen data Alkhawaldeh et al. (2023). In terms of accuracy, LGBM outperforms other models, followed by XGBoost and Random Forest. These findings are consistent with Feng et al. (2024), who reported that LGBM demonstrated the best generalization performance in predicting marine accident severity.

In comparison to the study conducted by Ratanawimon and Tanawongsuwan (2024), which utilized the same dataset source but focused on data from 2019 to 2022 to predict the Severity Level (Fatal and Injured) of Road Traffic Accidents in Thailand, our results demonstrate superior performance. Their study reported that the XGBoost model achieved a precision of 78%, recall of 57%, F1 score of 66%, and balanced accuracy of 77%. In contrast, in this study, the LGBM model achieved a precision of 83%, recall of 85%, F1 score of 83%, and accuracy of 85%. This improvement may be attributed to differences in the severity classification; while their study categorized severity levels into Fatal and Injured, our model distinguishes between Serious and Non-Serious accidents. Additionally, they used class weighting to address data imbalance, whereas we applied SMOTE. Furthermore, our dataset excludes data from 2020, a year that was significantly impacted by the COVID-19 pandemic, which may have introduced anomalies that affected model performance.

Regarding model explainability, the SHAP summary plot indicates the top three significant contributing features in the LGBM model's predictions: the number of motorcycles involved, road code, and the total number of vehicles and people involved in the accident. When comparing our findings with global traffic studies, our results are consistent with previous research, such as Pei et al. (2024), who used DeepSHAP to analyze road traffic accident risks from UK and US datasets. They found that the number of vehicles was a key factor in predicting fatal and serious accidents. Similarly, Ahmed et al. (2023), who evaluated machine learning (ML) models to predict road accident severity based on the latest NZ road accident dataset, reported that the number of vehicles involved was among the top contributing factors using SHAP. However, our study differs in that the number of motorcycles involved is the most significant factor in predicting accident severity. This discrepancy could be due to differences in the datasets. While these studies used data from the US, UK, and New Zealand, our analysis is based on data from Thailand, where motorcycle usage is significantly higher, resulting in different traffic patterns and accident risks.

In addition to predictive modeling, the practical implications of these findings extend to the insurance industry, particularly through the application of social marketing strategies. As road accidents significantly impact insurance claims and operational costs, insurance companies can leverage this data to develop targeted marketing campaigns aimed at promoting safer driving behaviors (Tapp et al., 2023). Specifically, focusing on reducing motorcycle accidents, which represent a significant proportion of traffic incidents in Thailand, can be a key priority. Social marketing campaigns highlighting the benefits of safer driving and offering incentives for accident prevention could help insurers reduce the frequency and severity of claims while fostering positive societal change (Diegelmann et al., 2020). This strategic combination of predictive modeling and social marketing creates an opportunity for the insurance industry to actively contribute to road safety initiatives while simultaneously managing financial risks more effectively.

6. CONCLUSION

This research investigates explainable machine learning models for predicting road accident severity in Thailand. Based on the results, SHAP analysis indicated that the most influential features were the number of motorcycles, the road code, and the number of vehicles and people involved. These findings hold significant practical implications, especially for sectors such as the insurance industry. Accurately predicting road accident severity provides valuable insights for insurance companies aiming to manage risk more effectively. Insurers can enhance their risk assessment strategies by integrating these predictive models into their operations. Furthermore, through targeted social marketing campaigns, policymakers and insurers can actively promote safer driving behaviors, particularly among motorcyclists, who were identified as the most significant contributors to severe accidents in Thailand. This approach reduces claims and operational costs and contributes to societal well-being by lowering accident rates.

6.1 Theoretical Implications

This study contributes to the exploration of road safety and machine learning by utilizing explainable models that provide valuable insights for building stakeholder trust. The results also emphasize the necessity for careful data selection and preprocessing, including the consideration of anomalies introduced by external factors such as the COVID-19 pandemic, which enhances the theoretical investigation of data quality in predictive modeling. Furthermore, this study expands the notion of social marketing, demonstrating that integrating systems science approaches, such as machine learning, into social marketing can strengthen a campaign’s effectiveness and help develop more impactful practices.

6.2 Practical Implications

Integrating Knowledge, Attitude, and Practice (KAP) theory into social marketing strategies is pivotal for advancing sustainable road safety. This framework focuses on enhancing awareness, delivering emotionally impactful communication, and fostering immediate behavioral change. By utilizing insights from SHAP, social marketing strategies can be further refined to target the most influential factors contributing to road accidents, such as high-risk zones, specific types of vehicles such as motorcycles, or peak accident times. These insights enable tailored communication campaigns and interventions that address the root causes of risky behaviors effectively. Moreover, implementing multi-level interventions that extend beyond individuals to encompass communities, organizations, and public policies from upstream to midstream and downstream can cultivate a supportive environment for lasting behavioral modifications.

Step One: Enhancing Awareness and Knowledge

In the first stage, it is recommended to prioritize awareness-raising among motorcycle riders, who significantly contribute to severe accidents. At the midstream and downstream levels, policymakers and insurance companies could develop campaigns to raise awareness of the risks associated with motorcycle use (Campbell et al., 2022; Siebert et al., 2021). Utilizing predictive models to pinpoint high-risk zones and accident-prone roads (Road Code) would also allow for more targeted dissemination of safety information and alerts to drivers in these areas (Hu et al., 2020). Additionally, communication tools such as advertising can effectively broaden awareness, mainly through digital media and social channels (Davis et al., 2016). For instance, well-curated TrueView ads on YouTube can attract attention through relevant, informative, and engaging content (Napontun & Senachai, 2023). Another effective strategy is collaborating with nonprofit organizations to conduct social marketing campaigns, which can drive change more effectively, especially at the midstream level (Luca et al., 2016).

At the upstream level, the government could lead national awareness campaigns, using mass media to convey the risks associated with motorcycle use and high-risk roads (Campbell et al., 2022; Siebert et al., 2021). In parallel, enforcing strict legal measures, such as mandatory helmet use and clearly defined penalties for non-compliance, underscores the importance of road safety (Siebert et al., 2021).

Step Two: Modifying Attitudes

The second stage focuses on emotionally resonant campaigns aimed at shifting attitudes among motorcycle riders, particularly at the midstream and downstream levels. Insurance companies should leverage emotionally compelling storytelling, as research shows this approach is often more effective in shaping public perceptions than relying solely on statistical data (Betsch et al., 2011). For instance, sharing accident survivor stories or accounts from affected families can encourage empathy and awareness regarding the broader impact of unsafe driving (von Beesten & Bresges, 2022). To further support these attitude shifts, continuous governmental campaigns play an essential role. For instance, regularly sending SMS or social media reminders reinforces the importance of safety compliance and cultivates a sense of shared responsibility, helping to establish safe driving as a societal norm (Campbell et al., 2022). Moreover, the government should emphasize infrastructure improvements as a key factor in reducing accident severity. Using the Road Code variable to identify accident-prone areas allows for targeted visual campaigns that effectively highlight the risks associated with high-incident roads (Hu et al., 2020).

Step Three: Encouraging Behavioral Change

The final stage focuses on interventions that prompt immediate behavioral responses or enforce compliance. At the midstream and downstream levels, insurance companies could introduce incentives or measures that encourage prompt behavioral shifts, such as offering insurance discounts to drivers who adhere to traffic laws. Campaigns that leverage real-time data to share information on traffic conditions, high-risk areas, and safety warnings help drivers better understand and respond to situational risks (Ali et al., 2021; Arvin et al., 2021; Kashevnik et al., 2020; Ma et al., 2022; Omerustaoglu et al., 2020). By integrating data from traffic cameras, sensors, and GPS, authorities can monitor vehicle density and identify periods of increased accident risk (Lee et al., 2022; Perafan-Villota et al., 2022). This data can be used to create alerts that are specifically customized to the most influential SHAP features, such as alerts for high-density areas or motorcycle-related risks. Disseminating this data through social marketing channels, such as mobile apps, social media, and digital billboards, provides drivers with timely information about high-risk locations and peak accident times, encouraging safer choices, alternative routes, or adjusted travel schedules to avoid congested areas (Ali et al., 2021). Insurance companies could further support these efforts by embedding machine-learning insights into navigation systems or smartphone applications, offering real-time alerts for high-density or historically risky areas as part of a personalized social marketing strategy that prompts drivers to take safer routes (Kashevnik et al., 2020).

At the upstream level, the government could implement immediate regulations, such as limiting vehicle numbers, managing pedestrian traffic in high-risk zones, enforcing speed limits on roads with high incident rates, and applying strict penalties for non-compliance (Charyk Stewart et al., 2021; Shams et al., 2022). Additionally, the government could emphasize infrastructure improvements, such as installing traffic signals, enhancing road conditions, and placing warning signs in hazardous areas. Insights from SHAP regarding the Road Code could guide prioritization of these interventions, to more effectively reduce accident severity (Diegelmann et al., 2020).

7. LIMITATIONS AND FUTURE RESEARCH

This research has some limitations. First, the reliance on historical data from Thailand may limit the generalizability of the findings to other regions or countries. Second, the dataset was restricted to specific features and years. Third, the study focused solely on employing the SMOTE technique to address data imbalance issues.

Future research should consider using more extensive datasets over longer time frames to identify emerging trends. Comparisons of dates and times should be conducted, such as examining the days and times when accidents occur most frequently, as well as the patterns between normal periods and festival or holiday periods. Moreover, additional factors should be collected for further investigation, such as the age and gender of those involved and whether accidents are more common when traveling alone or in groups. Furthermore, road codes should be thoroughly examined to gain a deeper understanding of which codes contribute most significantly to accidents. Additionally, exploring other techniques for handling imbalanced data, such as ADASYN (Adaptive Synthetic Sampling) or under-sampling methods, may result in more dynamic and adaptive predictive models. Lastly, conducting in-depth interviews could help confirm the results of Explainable AI (XAI), ensuring that a social marketing campaign can be significantly enhanced for maximum effectiveness.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS:

During the preparation of this work the authors used ChatGPT in order to check grammar and improve readability of the manuscript. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

REFERENCES

Aboulola, O. I., Alabdulqader, E. A., AlArfaj, A. A., Alsubai, S., & Kim, T. H. (2024). An Automated Approach for Predicting Road Traffic Accident Severity Using Transformer Learning and Explainable AI Technique. *IEEE Access*, *12*, 61062-61072. https://doi.org/10.1109/ACCESS.2024.3380895

Ahmed, S., Hossain, M. A., Ray, S. K., Bhuiyan, M. M. I., & Sabuj, S. R. (2023). A study on road accident prediction and contributing factors using explainable machine learning models: analysis and performance. *Transportation Research Interdisciplinary Perspectives*, *19*, 100814. https://doi.org/https://doi.org/10.1016/j.trip.2023.100814

Ali, F., Ali, A., Imran, M., Naqvi, R. A., Siddiqi, M. H., & Kwak, K.-S. (2021). Traffic accident detection and condition analysis based on social networking data. *Accident Analysis & Prevention*, *151*, 105973. https://doi.org/https://doi.org/10.1016/j.aap.2021.105973

Alkhawaldeh, I. M., Albalkhi, I., & Naswhan, A. J. (2023). Challenges and limitations of synthetic minority oversampling techniques in machine learning. *World J Methodol*, *13*(5), 373-378. https://doi.org/10.5662/wjm.v13.i5.373

Arvin, R., Khattak, A. J., & Qi, H. (2021). Safety critical event prediction through unified analysis of driver and vehicle volatilities: Application of deep learning methods. *Accident Analysis & Prevention*, *151*, 105949. https://doi.org/https://doi.org/10.1016/j.aap.2020.105949

Barrie, L. R., Jones, S. C., & Wiese, E. (2011). "At Least I'M Not Drink-Driving": Formative Research for a Social Marketing Campaign to Reduce Drug-Driving among Young Drivers. *Australasian Marketing Journal*, *19*(1), 71-75. https://doi.org/10.1016/j.ausmj.2010.11.010

Berhanu, Y., Schröder, D., Wodajo, B. T., & Alemayehu, E. (2024). Machine learning for predictions of road traffic accidents and spatial network analysis for safe routing on accident and congestion-prone road networks. *Results in Engineering*, *23*, 102737. https://doi.org/https://doi.org/10.1016/j.rineng.2024.102737

Betsch, C., Ulshöfer, C., Renkewitz, F., & Betsch, T. (2011). The Influence of Narrative v. Statistical Information on Perceiving Vaccination Risks. *Medical Decision Making*, *31*(5), 742-753. https://doi.org/10.1177/0272989x11400419

Bryant, C. A., Forthofer, M. S., Brown, K. R. M., Landis, D. C., & McDermott, R. J. (2000). Community-based prevention marketing: the next steps in disseminating behavior change. *American Journal of Health Behavior*, *24*(1), 61-68.

Buyucek, N., Kubacki, K., Rundle-Thiele, S., & Pang, B. (2016). A systematic review of stakeholder involvement in social marketing interventions. *Australasian Marketing Journal*, *24*(1), 8-19.

Campbell, B., Heitner, J., Amos Mwelelo, P., Fogel, A., Mujumdar, V., Adams, L. V., Boniface, R., & Su, Y. (2022). Impact of SMS text messaging reminders on helmet use among motorcycle drivers in Dar es Salaam, Tanzania: randomized controlled trial. *Journal of Medical Internet Research*, *24*(4), e27387.

Charyk Stewart, T., Edwards, J., Penney, A., Gilliland, J., Clark, A., Haidar, T., Batey, B., Pfeffer, A., Fraser, D. D., Merritt, N. H., & Parry, N. G. (2021). Evaluation of a population health strategy to reduce distracted driving: Examining all "Es" of injury prevention. *Journal of Trauma and Acute Care Surgery*, *90*(3), 535-543. https://doi.org/10.1097/ta.0000000000002948

Chen, J., Tao, W., Jing, Z., Wang, P., & Jin, Y. (2024). Traffic accident duration prediction using multi-mode data and ensemble deep learning. *Heliyon*, *10*(4). https://doi.org/10.1016/j.heliyon.2024.e25957

Cooper, C. (2007). Successfully Changing Individual Travel Behavior:Applying Community-

Based Social Marketing to Travel Choice. *Transportation Research Record*, *2021*(1), 89-99. https://doi.org/10.3141/2021-11

Crawshaw, P. (2014). Changing Behaviours, Improving Outcomes? Governing Healthy Lifestyles Through Social Marketing. *Sociology Compass*, *8*(9), 1127-1139. https://doi.org/https://doi.org/10.1111/soc4.12196

Davis, K. C., Shafer, P. R., Rodes, R., Kim, A., Hansen, H., Patel, D., Coln, C., & Beistle, D. (2016). Does Digital Video Advertising Increase Population-Level Reach of Multimedia Campaigns? Evidence From the 2013 Tips From Former Smokers Campaign. *J Med Internet Res*, *18*(9), e235. https://doi.org/10.2196/jmir.5683

Diegelmann, S., Ninaus, K., & Terlutter, R. (2020). Distracted driving prevention: an analysis of recent UK campaigns. *Journal of Social Marketing*, *10*(2), 243-264. https://doi.org/10.1108/JSOCM-07-2019-0105

Elias, W. (2021). The effectiveness of different incentive programs to encourage safe driving. *Sustainability*, *13*(6), 3398.

Elreedy, D., Atiya, A. F., & Kamalov, F. (2024). A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning. *Machine Learning*, *113*(7), 4903-4923. https://doi.org/10.1007/s10994-022-06296-4

Feng, Y., Wang, X., Chen, Q., Yang, Z., Wang, J., Li, H., Xia, G., & Liu, Z. (2024). Prediction of the severity of marine accidents using improved machine learning. *Transportation Research Part E: Logistics and Transportation Review*, *188*, 103647. https://doi.org/https://doi.org/10.1016/j.tre.2024.103647

Flaherty, T., Domegan, C., Duane, S., Brychkov, D., & Anand, M. (2020). Systems social marketing and macro-social marketing: A systematic review. *Social Marketing Quarterly*, *26*(2), 146-166.

Fox, K. F. A., & Kotler, P. (1980). The Marketing of Social Causes: The First 10 Years. *Journal of Marketing*, *44*(4), 24-33. https://doi.org/10.1177/002224298004400404

Hu, J., Huang, M.-C., & Yu, X. (2020). Efficient mapping of crash risk at intersections with connected vehicle data and deep learning models. *Accident Analysis & Prevention*, *144*, 105665.

Kok, İ., Okay, F. Y., Özgecan, M., & Özdemir, S. (2023). Explainable Artificial Intelligence (XAI) for Internet of Things: A Survey. *IEEE Internet of Things Journal*, *10*(16), 14764-14779. https://doi.org/10.1109/JIOT.2023.3287678

Isler, C. A., Huang, Y., & de Melo, L. E. A. (2024). Developing accident frequency prediction models for urban roads: A case study in São Paulo, Brazil. *IATSS Research*, *48*(3), 378-392. https://doi.org/https://doi.org/10.1016/j.iatssr.2024.07.002

Javeed, D., Gao, T., Kumar, P., & Jolfaei, A. (2024). An Explainable and Resilient Intrusion Detection System for Industry 5.0. *IEEE Transactions on Consumer Electronics*, *70*(1), 1342-1350. https://doi.org/10.1109/TCE.2023.3283704

Kashevnik, A., Lashkov, I., & Gurtov, A. (2020). Methodology and Mobile Application for Driver Behavior Analysis and Accident Prevention. *IEEE Transactions on Intelligent Transportation Systems*, *21*(6), 2427-2436. https://doi.org/10.1109/TITS.2019.2918328

Kiptoo, I. K., Kariuki, S. N., & Ocharo, K. N. (2021). Risk management and financial performance of insurance firms in Kenya. *Cogent Business & Management*, *8*(1), 1997246. https://doi.org/10.1080/23311975.2021.1997246

Kumar, R., Javeed, D., Aljuhani, A., Jolfaei, A., Kumar, P., & Islam, A. K. M. N. (2024). Blockchain-Based Authentication and Explainable AI for Securing Consumer IoT Applications. *IEEE Transactions on Consumer Electronics*, *70*(1), 1145-1154. https://doi.org/10.1109/TCE.2023.3320157

Lee, H., Lee, J., & Chung, Y. (2022). Traffic density estimation using vehicle sensor data.

Journal of Intelligent Transportation Systems, *26*(6), 675-689. https://doi.org/10.1080/15472450.2021.1966626

Luca, N. R., Hibbert, S., & McDonald, R. (2016). Midstream value creation in social marketing. *Journal of Marketing Management*, *32*(11-12), 1145-1173. https://doi.org/10.1080/0267257X.2016.1190777

Ma, Y., Xie, Z., Chen, S., Wu, Y., & Qiao, F. (2022). Real-Time Driving Behavior Identification Based on Multi-Source Data Fusion. *International Journal of Environmental Research and Public Health*, *19*(1), 348. https://www.mdpi.com/1660-4601/19/1/348

Matsumoto, D. (2007). Culture, Context, and Behavior. *Journal of Personality*, *75*(6), 1285-1320. https://doi.org/https://doi.org/10.1111/j.1467-6494.2007.00476.x

Ministry of Transport. (2024). *Accidents on Ministry of Transport's Road Network*. https://datagov.mot.go.th/dataset/roadaccident

Muleme, J., Kankya, C., Ssempebwa, J. C., Mazeri, S., & Muwonge, A. (2017). A framework for integrating qualitative and quantitative data in knowledge, attitude, and practice studies: a case study of pesticide usage in eastern Uganda. *Frontiers in Public Health*, *5*, 318.

Napontun, K., & Senachai, P. (2023). Identifying factors influencing consumers not to skip TrueView advertising on YouTube. *ABAC Journal*, *43*(1), 85-102. https://doi.org/10.14456/abacj.2023.6

Obasi, I. C., & Benson, C. (2023). Evaluating the effectiveness of machine learning techniques in forecasting the severity of traffic accidents. *Heliyon*, *9*(8), e18812. https://doi.org/https://doi.org/10.1016/j.heliyon.2023.e18812

Omerustaoglu, F., Sakar, C. O., & Kar, G. (2020). Distracted driver detection by combining in-vehicle and image data using deep learning. *Applied Soft Computing*, *96*, 106657. https://doi.org/https://doi.org/10.1016/j.asoc.2020.106657

Pei, Y., Wen, Y., & Pan, S. (2024). Road Traffic Accident Risk Prediction and Key Factor Identification Framework Based on Explainable Deep Learning. *IEEE Access*, *PP*, 1-1. https://doi.org/10.1109/ACCESS.2024.3451522

Perafan-Villota, J. C., Mondragon, O. H., & Mayor-Toro, W. M. (2022). Fast and Precise: Parallel Processing of Vehicle Traffic Videos Using Big Data Analytics. *IEEE Transactions on Intelligent Transportation Systems*, *23*(8), 12064-12073. https://doi.org/10.1109/TITS.2021.3109625

Plant, B. R. C., Irwin, J. D., & Chekaluk, E. (2017). The effects of anti-speeding advertisements on the simulated driving behaviour of young drivers. *Accident Analysis & Prevention*, *100*, 65-74. https://doi.org/https://doi.org/10.1016/j.aap.2017.01.003

Ratanawimon, S., & Tanawongsuwan, P. (2024, August 19 - 21, 2024). Prediction of Severity Level of Road Traffic Accident in Thailand using Machine Learning. *Proceedings of the 10th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS'24)*, Barcelona, Spain.

Roger, A., Dourgoudian, M., Mergey, V., Laplanche, D., Ecarnot, F., & Sanchez, S. (2023). Effectiveness of prevention interventions using social marketing methods on behavioural change in the general population: a systematic review of the literature. *International Journal of Environmental Research and Public Health*, *20*(5), 4576.

Sahlaoui, H., Alaoui, E. A. A., Nayyar, A., Agoujil, S., & Jaber, M. M. (2021). Predicting and Interpreting Student Performance Using Ensemble Models and Shapley Additive Explanations. *IEEE Access*, *9*, 152688-152703. https://doi.org/10.1109/ACCESS.2021.3124270

Senachai, P., Julsrigival, J., & Sann, R. (2022). Social Marketing Strategy to Promote Traditional Thai Medicines during COVID-19: KAP and DoI Two-Step Theory

Application Process. *International Journal of Environmental Research and Public Health*, *19*(14), 8416.

Shaik, M. E., Islam, M. M., & Hossain, Q. S. (2021). A review on neural network techniques for the prediction of road traffic accident severity. *Asian Transport Studies*, *7*, 100040. https://doi.org/https://doi.org/10.1016/j.eastsj.2021.100040

Shams, M., Maleki, M., Shariatinia, S., Omidimorad, A., Zakeri, H., Fallah Zavareh, M., Hamelmann, C., Mooren, L., Shuey, R., & Ranjbar, M. (2022). Insights for speed management among Iranian drivers: a social marketing formative research study. *Journal of Injury and Violence Research*, *14*(3), 199-208. https://doi.org/10.5249/jivr.v14i3.1690

Siebert, F. W., Hellmann, L., Pant, P. R., Lin, H., & Trimpop, R. (2021). Disparity of motorcycle helmet use in Nepal – Weak law enforcement or riders' reluctance? *Transportation Research Part F: Traffic Psychology and Behaviour*, *79*, 72-83. https://doi.org/https://doi.org/10.1016/j.trf.2021.04.005

Spilbergs, A., Fomins, A., & Krastiņš, M. (2022). Multivariate Modelling of Motor Third Party Liability Insurance Claims. *European Journal of Business Science and Technology*, *8*(1), 5-18.

Sufian, M. A., Varadarajan, J., & Niu, M. (2024). Enhancing prediction and analysis of UK road traffic accident severity using AI: Integration of machine learning, econometric techniques, and time series forecasting in public health research. *Heliyon*, *10*(7), e28547. https://doi.org/https://doi.org/10.1016/j.heliyon.2024.e28547

Tapp, A., Ursachi, G. M., & Campsall, D. (2023). Exploring the relationship between car brands and risky driving. *Journal of Social Marketing*, *13*(4), 554-571. https://doi.org/10.1108/JSOCM-04-2023-0074

Troy, D. M., Maynard, O. M., Hickman, M., Attwood, A. S., & Munafò, M. R. (2015). The effect of glass shape on alcohol consumption in a naturalistic setting: a feasibility study. *Pilot and Feasibility Studies*, *1*(1), 27. https://doi.org/10.1186/s40814-015-0022-2

Truong, V. D. (2014). Social Marketing:A Systematic Review of Research 1998–2012. *Social Marketing Quarterly*, *20*(1), 15-34. https://doi.org/10.1177/1524500413517666

Truong, V. D., & Hall, C. M. (2013). Social Marketing and Tourism:What Is the Evidence? *Social Marketing Quarterly*, *19*(2), 110-135. https://doi.org/10.1177/1524500413484452

Valente, T. W., Paredes, P., & Poppe, P. R. (1998). Matching the Message to the Process: The Relative Ordering of Knowledge, Attitudes, and Practices in Behavior Change Research. *Human Communication Research*, *24*(3), 366-385. https://doi.org/10.1111/j.1468-2958.1998.tb00421.x

Venkatesh Raja, K., Siddharth, R., Yuvaraj, S., & Ramesh Kumar, K. A. (2023). An Artificial Intelligence based automated case-based reasoning (CBR) system for severity investigation and root-cause analysis of road accidents – Comparative analysis with the predictions of ChatGPT. *Journal of Engineering Research*. https://doi.org/https://doi.org/10.1016/j.jer.2023.09.019

Viswanath, D., P, K., N, R., & B, R. (2021, 8-10 April 2021). A Road Accident Prediction Model Using Data Mining Techniques. *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*,

von Beesten, S., & Bresges, A. (2022). Effectiveness of road safety prevention in schools. *Frontiers in Psychology*, *13*, 1046403.

Wang, B., Zhan, S., Sun, J., & Lee, L. (2009). Social mobilization and social marketing to promote NaFeEDTA-fortified soya sauce in an iron-deficient population through a public–private partnership. *Public Health Nutrition*, *12*(10), 1751-1759. https://doi.org/10.1017/S136898000800431X

Wang, G., Li, J., Shen, L., Ding, S., Shi, Z., & Zuo, F. (2024). Towards efficient and accurate prediction of freeway accident severity using two-level fuzzy comprehensive evaluation. *Heliyon*, *10*(16), e36396. https://doi.org/https://doi.org/10.1016/j.heliyon.2024.e36396

Wetchayont, P. (2021). Investigation on the Impacts of COVID-19 Lockdown and Influencing Factors on Air Quality in Greater Bangkok, Thailand. *Advances in Meteorology*, *2021*, 1-11. https://doi.org/10.1155/2021/6697707

World Health Organization. (2023, 13 December 2023). *Global status report on road safety 2023*. https://www.who.int/teams/social-determinants-of-health/safety-and-mobility/global-status-report-on-road-safety-2023

Zhankaziev, S. V., Zamytskih, A. V., Vorobyev, A. I., Gavrilyuk, M. V., & Pletnev, M. G. (2022, 10-11 Nov. 2022). Predicting Traffic Accidents Using the Conflict Coefficient. *2022 Intelligent Technologies and Electronic Devices in Vehicle and Road Transport Complex (TIRVED)*

Zhao, H., Yu, H., Li, D., Mao, T., & Zhu, H. (2019). Vehicle Accident Risk Prediction Based on AdaBoost-SO in VANETs. *IEEE Access*, *7*, 14549-14557. https://doi.org/10.1109/ACCESS.2019.2894176

Zheng, S., Zhao, L., Ju, N., Hua, T., Zhang, S., & Liao, S. (2021). Relationship between oral health-related knowledge, attitudes, practice, self-rated oral health and oral health-related quality of life among Chinese college students: a structural equation modeling approach. *BMC Oral Health*, *21*(1), 99. https://doi.org/10.1186/s12903-021-01419-0