Introduction

Law enforcement agencies in Sri Lanka grapple with formidable challenges in effectively addressing accidents and crimes. Historically, traditional approaches to case analysis and investigation have proven to be labour-intensive, error-prone, and lacking predictive capabilities. These conventional methods often lead to delayed justice, while the escalating volume of data overwhelms human analytical capacities. Moreover, the absence of proactive predictive tools hinders law enforcement's ability to anticipate and prevent accidents and crimes. The pressing need for a more advanced, data-driven solution to enhance public safety and optimize law enforcement efforts becomes evident in the face of these challenges.

In response to these issues, this research project, titled "The Future of Crime Prevention: Police Case Analysis Using Machine Learning," endeavours to revolutionize how law enforcement agencies tackle the multifaceted issues of crime and accidents. This integrated system comprises four main components, each with its distinct objectives.

Accidents are unpredictable events that can have devastating consequences, making it paramount to understand their patterns and causes. The "Accident Case Analysis" component aims to empower law enforcement agencies by predicting accident percentages, fatal percentages, and accident causes for the next three years within specified divisions. This component serves as a solution to the labor-intensive and time-consuming nature of traditional accident case analysis. By leveraging machine learning techniques, it forecasts future accident trends, identifies causes, and enables in-depth analyses. It empowers law enforcement with actionable insights to proactively address safety concerns and allocate resources effectively, even predicting peak accident occurrence times.

Crime analysis has always been pivotal for law enforcement, primarily relying on statistical methodologies. This research recognizes the potential of integrating advanced machine learning techniques to enable proactive crime prediction and prevention. The core objective of our research is to create a robust system that empowers users to gain insights into crime dynamics by providing essential inputs: the year, month, and location of interest. Through predictive capabilities, it identifies the highest occurring crimes, demographics of affected individuals, vehicle details, and

information regarding stolen objects. This system offers predictions of crime rates for the next five years and generates graphical representations of crime statistics, harnessing the power of machine learning to address the pressing need for accurate and proactive crime analysis.

Additionally, the field of crime analysis has evolved significantly, shifting from manual data entry and analysis to technology-driven solutions. The case document classification using text analysis system is proposed to eliminate manual case searches, expedite investigations, reduce errors, and enhance decision-making by extracting crucial information from unstructured text data. This solution promises substantial benefits, including reduced case backlog and improved efficiency and accuracy in the criminal justice system.

Furthermore, the economic crisis in Sri Lanka has led to a drastic increase in criminal cases, with a specific focus on crimes against women. These crimes have a severe impact on women's well-being, often going unreported due to stigma, fear, and mistrust in the legal system. The "Crimes Against Women Analysis" component proposes a solution by clustering crimes against women based on characteristics such as location, type, and year. This component aims to develop a crime forecasting model, enabling law enforcement agencies to take preventive measures and ensure women's safety.

By integrating machine learning algorithms into police case analysis, this research project aims to overcome the challenges faced by traditional methods. It provides law enforcement agencies with advanced analytical tools, predictive capabilities, and efficient systems to allocate resources effectively, prevent accidents and crimes, and ultimately enhance public safety in Sri Lanka. In the subsequent sections of this report, we will delve into each component's methodology, data analysis, and findings, shedding light on their potential to transform law enforcement and public safety.

Literature survey

Numerous studies have extensively explored approaches in police accident case analysis, crime analysis, text classification, and crimes against women. Li, Wu, and Peng [1] developed a traffic accident prediction model using spiking neural networks (SNNs) and convolutional neural networks (CNNs). This research addresses the critical need for accurate post-impact prediction in traffic accident management, utilizing SNNs to capture spatial and temporal features effectively, potentially improving prediction accuracy over traditional methods. V. Prasannakumar, H. Vijith, R. Charutha, and N. Geetha [2] conducted research on road accident analysis using geo-information technology to identify accident hotspots and employed spatial statistics, including Moran's I and Getis-Ord Gi* statistics, within a GIS-based framework to assess spatial clustering, aiding traffic management and safety decisions. It aimed to assess the spatial and temporal patterns of accidents in the Thiruvananthapuram city area, providing insights into accident hotspots and cold spots. This research contributed to understanding the distribution and variations of road accidents, which can inform traffic management strategies and accident reduction efforts, emphasizing the significance of geospatial and temporal data analysis in accident research and prevention.

In the study conducted by Manzoor et al. [3], the focus was on predicting the severity of road accidents, a critical concern given the numerous factors involved. The authors employed an ensemble model called RFCNN and compared its performance with various base learner models, including tree-based ensemble models (RF, AC, ETC, GBM) and an ensemble of regression algorithms (Voting classifier, LR+SGD). Notably, Random Forest (RF) identified 20 significant features from the dataset, and these features were used for the experiments. The results demonstrated that RFCNN outperformed other models in terms of accuracy, achieving an impressive accuracy rate of 99.1%, while also excelling in precision, recall, and F-score. The study emphasized the importance of identifying significant features and highlighted the potential for improving accident severity prediction while reducing data collection costs. Senanayake and Joshi [4] developed the Road Accident Pattern Miner (RAP miner), an innovative system using a hybrid learning algorithm and the Self-Expiring Association (SEA) algorithm for association rule mining. This system achieved an impressive accuracy rate of 92.75%, surpassing other algorithms' maximum accuracy of 92.25%. Integrating SEA with Case-Based Reasoning (CBR) significantly

enhanced the RAP miner's performance, reducing processing time by a remarkable 67%. This advancement holds promise for real-time accident prediction based on location and severity, offering significant contributions to road safety. Yang, Han, and Chen's study [5], the prediction of traffic accident severity was achieved using a Random Forest algorithm. Their model exhibited a remarkable accuracy rate of 80%, outperforming other machine learning models. Key accident characteristics, such as location, collision pattern, road information, and speed limits, were employed to enhance prediction accuracy, offering valuable insights for traffic safety management and accident prevention.

A noteworthy study by Neil Shah, Nandish Bhagat, and Manan Shah [6] has spotlighted the potential of machine learning and computer vision in crime forecasting. Their research emphasizes the significance of harnessing data streams such as facial recognition, number plate recognition, augmented and mixed realities, location determination, and object identification to predict and preempt criminal activities. The proposal to employ motive as a critical criterion for assessing the nature of a crime and comprehensive categorization for efficient analysis signifies the depth of their approach. Steven Walczak's work [7] has made strides in the application of neural network models for predicting specific crime types based on temporal and spatial information. His findings reveal that neural network models can predict crime types with remarkable accuracy, offering valuable insights for police decision-making in crime prevention strategies.

Karabo Jenga, Cagatay Catal, and Gorkem Kar [8] have explored the versatility of machine learning and data mining in crime prediction and prevention. Their research underscores the adaptability of these technologies in addressing the complex dynamics of criminal activities. Furthermore, the work of Suhong Kim, Param Joshi, Paraminder Singh Kalsi, and Pooya Taheri [9] highlights the significance of decision trees in crime analysis. Using a dataset of over 560,000 records to predict crimes in Vancouver, their study achieved notable accuracy rates ranging from 39% to 44%, depending on crime categories and timing. This research underscores the utility of decision trees in enhancing crime prediction accuracy. The literature on crime analysis using machine learning is diverse, covering domains like e-government, serial crime patterns, theft, cyber-crimes, social crimes, and text mining.

The background survey on crime analysis and classification using machine learning techniques is extensive and diverse, with a focus on various aspects of crime detection, prediction, and

prevention. The studies mentioned in the references provide valuable insights into the use of machine learning algorithms for crime analysis, classification, and prediction, covering different domains, such as e-government, serial criminal patterns, theft crimes, cyber-crimes, social crimes, and text mining for crime analysis. Chih-Hao Ku et al. [10] propose a decision support system for automated crime report analysis and classification in e-government. They leverage machine learning techniques to automatically classify crime reports into predefined categories, which can assist law enforcement agencies in efficiently analyzing and managing crime data.Dahbur and Muscarello [11] present a classification system for serial criminal patterns using artificial intelligence and law. They propose a rule-based expert system that uses decision trees and statistical techniques to identify patterns in serial crimes, which can aid in profiling and predicting serial criminal behavior. Ghankutkar et al. [12] propose a machine learning model for analyzing crime news. They utilize natural language processing techniques and machine learning algorithms to classify crime news articles into categories such as robbery, murder, and fraud, to aid in crime analysis and prediction. Qi [13] proposes a text classification approach for theft crimes based on the TF-IDF (Term Frequency-Inverse Document Frequency) technique and the XGBoost machine learning model. The proposed method effectively classifies theft crimes into different categories, such as pickpocketing, burglary, and shoplifting, based on text data extracted from crime reports. Alruily et al. [14] focus on crime type document classification from an Arabic corpus. They propose a machine learning-based approach that uses features such as keywords, text statistics, and machine learning classifiers to automatically classify Arabic crime documents into different crime types, such as theft, assault, and fraud, which can assist in crime analysis in Arabic-speaking regions.

In the context of mitigating crimes against women and children, Rokonuzzaman Reza [15] introduced an innovative approach leveraging machine learning techniques. Their system analyzed diverse data sources, including social media and news articles, to detect patterns of oppression. Impressively, this system achieved an exceptional accuracy rate of 99%. It not only identifies such incidents but also offers real-time monitoring and intervention capabilities, ensuring swift responses to address these issues promptly, thus contributing significantly to creating a safer environment for women and children. Additionally, Adderley [16] conducted a comprehensive examination of crime predictive systems with a focus on data analysis and machine learning algorithms. Their research aimed to forecast criminal activities, taking into account the

effectiveness, ethical considerations, and potential biases within these systems. While specific accuracy rates may vary, their work underscores the importance of considering these critical factors in crime prediction models.

In a different approach, Islam et al. [17] introduced the "Joy 109" app, which empowers users to send distress messages containing GPS locations, audio recordings, and photos to relevant authorities. This innovative tool enhances response times and situational awareness, although a specific accuracy rate is not mentioned. Furthermore, Kiani et al. [18] proposed a crime prediction analysis approach involving clustering algorithms. This method identifies patterns in crime data and forecasts criminal activities within specific areas. While accuracy rates are not explicitly mentioned, the use of clustering algorithms highlights a data-driven strategy for crime prediction. Lastly, Karmakar et al. [19] introduced the "SafeBand" wearable device, complemented by mobile apps for victim assistance and police support. This wearable allows for location tracking and emergency messaging, providing an added layer of security. Specific accuracy rates for this device are not provided in the context. These studies underscore the remarkable potential of machine learning, data analysis, and innovative technologies in the fields of traffic accident prediction, crime analysis, document classification and crimes against women.

Research Gap

In the realm of traffic accident prediction and analysis, there is a substantial research gap, particularly when it comes to combining the prediction of accident percentages for future years, causative factors, and in-depth accident case analysis into a unified system. While existing studies have made significant progress in predicting accident locations and causes, the holistic approach proposed here remains unexplored. The novelty lies in the integration of these critical facets, providing invaluable insights for long-term traffic safety planning, targeted prevention strategies, and a more comprehensive understanding of accident dynamics. This unified system represents a significant research gap, as no such comprehensive approach currently exists in traffic safety management. Similarly, in the field of crime case analysis, the absence of an all-encompassing and user-friendly system that fully utilizes the Decision Tree model for predicting future crime patterns stands out as a conspicuous research gap. Existing research often takes a fragmented approach, focusing on singular aspects of crime analysis, leaving a void in the development of a comprehensive crime analysis system. This research aims to bridge that gap by offering a platform that not only predicts crime patterns but also provides insights into demographics, crime types, patterns involving stolen objects, and long-term crime rate predictions. This holistic perspective on crime patterns is currently lacking, making the proposed system a significant contribution to the field.

Additionally, within the domain of automated text analysis systems for decision support in criminal investigations, there is a notable research gap, especially in the Sri Lankan context. Existing studies often draw from international research, which may not directly apply to the unique challenges faced by the Sri Lankan law enforcement and court system. This knowledge gap impedes the development of a tailored automated text analysis system customized for Sri Lanka, hindering effective decision-making in criminal investigations. Addressing this gap through empirical research on Sri Lanka's specific requirements is essential. Moreover, in the context of preventing violence against women and children, existing studies primarily focus on victim support, analysis, and prediction, with limited attention to preventive measures. The absence of a standardized method for clustering crimes against women compounds the research gap, making it difficult to compare results across studies. The proposed system, encompassing crime prediction, location-based precautions, and data analysis, seeks to fill this gap by offering

a comprehensive approach to predicting and preventing crimes against women, providing valuable support for government and law enforcement agencies.

Overall, the research gap lies in the absence of a unified system that combines these diverse components, spanning accident case analysis, crime analysis, case document classification, and crimes against women analysis. The proposed system aims to address this gap comprehensively, offering a novel and integrated approach that has the potential to revolutionize safety management and criminal justice.