



Classification of crime-risk areas by using the urban environment and satellite images

KRIANGKRAI MANOCHAI

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR MASTER DEGREE OF SCIENCE
IN GEOINFORMATICS
FACULTY OF GEOINFORMATICS
BURAPHA UNIVERSITY

2025

COPYRIGHT OF BURAPHA UNIVERSITY

การจำแนกพื้นที่เสี่ยงต่ออาชญากรรมโดยใช้สิ่งแวดล้อมเมืองและภาพถ่ายดาวเทียม



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต

สาขาวิชาภูมิสารสนเทศศาสตร์

คณะภูมิสารสนเทศศาสตร์ มหาวิทยาลัยบูรพา

2568

ลิขสิทธิ์เป็นของมหาวิทยาลัยบูรพา

Classification of crime-risk areas by using the urban environment and satellite images



KRIANGKRAI MANOCHAI

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR MASTER DEGREE OF SCIENCE
IN GEOINFORMATICS
FACULTY OF GEOINFORMATICS
BURAPHA UNIVERSITY
2025
COPYRIGHT OF BURAPHA UNIVERSITY

The Thesis of Kriangkrai Manochai has been approved by the examining committee to be partial fulfillment of the requirements for the Master Degree of Science in Geoinformatics of Burapha University

Advisory Committee

Examining Committee

Principal advisor

.....
(Professor Dr. Zhenfeng Shao)

..... Principal
examiner
(Professor Dr. Timo Balz)

Co-advisor

.....
(Assistant Professor Dr. Phattraporn
Soytong)

..... Member
(Professor Dr. Milan Konecny)

..... Member
(Associate Professor Dr. Lei Wang)

..... Member
(Associate Professor Dr. Lei Shi)

..... Member
(Associate Professor Dr. Ming Li)

..... Dean of the Faculty of Humanities and
Social Sciences
(Associate Professor Dr. Suchada Pongkittiwiwiboon)

This Thesis has been approved by Graduate School Burapha University to be partial fulfillment of the requirements for the Master Degree of Science in Geoinformatics of Burapha University

..... Dean of Graduate School
(Associate Professor Dr. Witawat Jangiam)

65910030: MAJOR: GEOINFORMATICS; M.Sc. (GEOINFORMATICS)

KEYWORDS: crime risk, remote sensing data, machine learning, urban environment, Random Forest, Extra trees, lightGBM

KRIANGKRAI MANOCHAI : CLASSIFICATION OF CRIME-RISK AREAS BY USING THE URBAN ENVIRONMENT AND SATELLITE IMAGES.
ADVISORY COMMITTEE: ZHENFENG SHAO, Ph.D. PHATTRAPORN SOYTONG, Ph.D. 2025.

The tourism industry is very important to Thailand's economy, accounting for 11% of the nation's GDP. The COVID-19 pandemic has impacted the worldwide tourism landscape, providing challenges for the industry's recovery. This study focuses on enhancing public safety across Bangkok, particularly within the tourism sector, which is a leading global destination and an important hub for Thailand's tourism. Despite Bangkok's popularity, safety remains a sensitive issue, with crime impacting both tourists and residents. This study aims to classify crime risk areas in Bangkok by integrating diverse datasets, including satellite imagery, government records, and open-source data. These findings provide guidance for improving urban safety and developing public policy.

Remote Sensing (RS) and Geographic Information Systems (GIS) are important instruments in crime risk analysis, providing for assessing spatial and environmental factors of criminal activity. Through the integration of satellite imagery and urban datasets, including Points of Interest (POI), population density, and greenspace area, RS and GIS provide the efficiency mapping and classification of high-risk areas. These technologies help with the identification of crime hotspots and contribute to the development of predictive models that enhance crime prevention strategies and urban planning.

This research adopts a data-driven approach using a crime risk record and urban environment factors influencing the crime rate by considering covering human activity, socio-economics, and infrastructure. The factors were extracted from government agencies, satellite imagery, and open data. A crime risk record extracted from the Metropolitan Police Bureau (MPB), Bangkok Metropolitan Administration (BMA), and civil complaints with label high, medium, and low risk, respectively, was

assessed by five experts and defined as train data. Urban environment factors are extracted into four categories consisting of:

- 1). Points of Interest (POI) provide the functional characteristics of urban areas as a hub of economic and social activities.
- 2). Greenspace refers to tree cover, parks, and ground cover. Urban areas that are well-designed and well-maintained would reduce property and violent crime rates, but poorly maintained green spaces could attract criminal activities.
- 3). Demographics from the LandScan project reflect the real situation of population density, which refers to the area influx of people would increase opportunities for property crimes.
- 4). Nighttime lighting from the Suomi National Polar-Orbiting Partnership (SNPP) refers to the effects of increased lighting on crime rates, which are more pronounced in urban areas with higher nighttime business activity.

The four categories could be broken down into 14 factors as input data represents in raster format preparing for processing. The three machine learning methods (Extra Trees, Random Forest, and Light Gradient Boosting Machine) were applied to classify and predict crime risk areas across Bangkok. Train data, divide it into training (80%) and testing (20%), and use K-Fold Cross Validation to evaluate the performance. Model evaluating the accuracy from Precision, Recall, F1Score, ROC-AUC, and confusion matrix.

The result shows Extra Tree (ET) model outperforms other with accuracy of 97.53%. The results also reveal Business and Industrial (P8), Administration (P5), and Catering and Recreation (P2) are the most important factors that influence model accuracy, and NDVI (P13) has the least impact on the model. While analysis into specific groups of factors by the average score shows demographics is highest, followed by nighttime lights, POI, and greenspace.

The classification of crime-risk areas reveals that most areas across Bangkok are classified as low-risk, accounting for 94.89% of the total area. However, the small percentage of high and medium crime risk, accounting for 2.59% and 2.52%, respectively, are very sensitive to investigations that are thorough. The investigation shows Urban and built-up areas are the largest proportion of land use

types in the total of high and medium crime risk areas, which emphasizes the correlation between urbanization, population density, and increased crime risk. Additionally, this study also analyzes specifics into eight districts where crime against tourists occurs. These districts, which attract a high influx of visitors because of famous attractions, shopping places, and nightlife, are especially vulnerable to crimes. The study reveals Pathum Wan, Watthana, and Khlong Toei districts are the highest proportion from the total area of high- and medium-risk areas in eight districts that aligns with reporting crime incidents against tourists in eight districts in 2022.

Bangkok is one of Southeast Asia's most densely populated megacities, and mass transportation plays a very important role in serving people. The Bangkok metro system, consisting of the BTS Skytrain, MRT, and Airport Rail Link, provides efficient and accessible mass transit. It facilitates mobility across commercial, residential, and tourist zone connectivity. As such, this study also analyzes specific metro stations. The classification results show that of the 150 metro stations in Bangkok within a station's 100-meter radius, 34 are in high crime risk areas and 9 are in medium crime risk areas. Furthermore, out of 25 metro stations located in districts with recorded crimes against tourists, 11 stations are classified as having a high crime risk level, while 1 station falls under the medium crime risk. These findings emphasize the relationship between metro station locations and crime risks, particularly in tourist density areas. Utilize it to develop a suitable map of the metro system for tourists and residents.

Finally, the findings indicate that areas with high POI density, population density, nighttime light, and low greenspace are more prone to crime that has an impact on safety, tourism, urbanization, social issues, transportation, and economic activity. Ensuring public safety is important for urban development, sustaining tourism, and economics. Data-driven crime assessments are helpful policymakers in enhancing public safety, promoting economic growth, and ensuring sustainable urban development.

ACKNOWLEDGEMENTS

During my master's degree studies, I am deeply grateful to my supervisor, Prof. Zhenfeng Shao, for his invaluable guidance, advice, expertise, and experience, which provided insight and direction for my dissertation work. I would also like to thank Asst. Prof. Phattharaphorn Soithong, my co-advisor, for her generous support and valuable knowledge, which have greatly aided me in both my academic pursuits and personal life.

I would like to thank the Thai Government for providing scholarships, the Bangkok Metropolitan Administration, and the Royal Thai Police for providing valuable data, which was essential for training the model used in my research. Mr. Oran Viriyincy to allow me to utilize Bangkok Transit Map on his project. Special thanks to Captain Phongphat Japhichom, my classmate, for his assistance with the technical methods of this research.

I am also grateful to my senior, Ms. Sunantha Ousaha, for her invaluable advice. Additionally, I want to thank my fellow SCGI#5 classmates for their constant support and assistance over the past two years, making my studies enjoyable and fulfilling.

I extend my gratitude to the lecturers from the Faculty of Geo-Informatics at Burapha University, Thailand, and the State Key Information Engineering Laboratory for Surveying, Mapping, and Remote Sensing (LIESMARS) in China. I also appreciate the Geo-Informatics and Space Technology Development Agency of Thailand (GISTDA) for providing me with the opportunity to continue my studies. I am thankful to the commander of the Ministry of Defence for allowing me this opportunity to enhance my experience.

Finally, I would like to express my deepest gratitude to my parents for their unwavering encouragement and pure love, as well as to everyone behind the scenes. It would have been impossible for me to graduate without the help of these people.

Kriangkrai Manochai

TABLE OF CONTENTS

	Page
ABSTRACT.....	D
ACKNOWLEDGEMENTS.....	G
TABLE OF CONTENTS.....	H
List of Tables.....	L
List of figures.....	M
CHAPTER 1	1
INTRODUCTION	1
1.1 Background	1
1.2 Problem Statement	1
1.3 Research questions	4
1.4 Research objectives	5
1.5 Dissertation structure.....	6
CHAPTER 2	8
LITTERATURE REVIEW	8
2.1 Background of This Study.....	8
2.2 Urban environment.....	8
2.3 Kernel density estimation (KDE)	10
2.4 Tourism and Remote sensing.....	10
2.5 Tourism and safety issues	11
2.6 Crime and urban environment.....	11
2.7 Crime and machine learning approaches.....	12
2.7.1 Machine learning.	12
2.7.2 Random Forest (RF).....	13
2.7.3 Light Gradient Boosting Model (LightGBM)	14
2.7.4 Extra Trees (ET)	14

2.8 Summaries related work with machine learning approaches	15
CHAPTER 3	17
METHODOLOGY	17
3.1 Study Area	17
3.2 Concept framework	18
3.3 Data Set and Pre-processing.....	19
3.3.1 Data training (Crime risk points).....	19
3.3.2 Data input (Urban environment factors).....	24
3.4 Classification process	35
3.4.1 Machine learning method.....	35
3.4.2 Accuracy assessment	37
3.4.3 Analyze the crime data and create a suitable area map for locals and tourists.	39
3.5 Tools and Instrumentation	39
CHAPTER 4	40
RESULTS AND VALIDATION	40
4.1 Analysis of crime risk in Bangkok (Data training).....	40
4.1.1 Crime statistics	41
4.1.2 Crime risk points	47
4.2 Analysis of Urban environment in Bangkok (Data input).....	49
4.2.1 P1-Retail (62,406 points)	50
4.2.2 P2-Catering and Recreation (36,300 points)	51
4.2.3 P3-Tour and Hotel Service (4,939 points).....	51
4.2.4 P4-Healthcare (10,801 points).....	52
4.2.5 P5-Administration (13,107 points).....	53
4.2.6 P6- Financial Services (25,273 points).....	54
4.2.7 P7- Education (7,279 points).....	55
4.2.8 P8- Business and Industrial Production (55,400 points)	56
4.2.9 P9-Transportation (8,258 points).....	57

4.2.10 P10-Road network	58
4.2.11 P11-Residential (24,373 points)	59
4.2.12 P12- Population density	60
4.2.13 P13- Greenspace area (NDVI).....	61
4.2.14 P14- Nighttime light.....	62
4.3 Comparing the Performance of Classification ML Models.....	63
4.3.1 Buffer analysis.....	63
4.3.2 Model training and hyper-parameter optimization.....	64
4.3.3 Features importance of parameters.....	65
4.3.4 Evaluating ML model performance with a confusion matrix.....	68
4.3.5 Evaluating ML model performance with ROC-AUC.....	69
4.3.6 Evaluating ML model performance with PRC-AUC	71
4.3.7 Evaluating ML model performance with accuracy	73
4.4 Classification of risky area map in Bangkok.....	74
4.4.1 The classification of risky areas across Bangkok.....	74
4.4.2 The district where crime statistics against tourists	76
4.4.3 The influence of land use to crime risk area.....	82
4.4.4 The influence of nighttime light to crime risk area	85
4.4.5 The influence of population density to crime risk area	86
4.5 Sample of risky area in Bangkok from the best model	87
4.6 A comprehensive map along Bangkok's metro lines	91
4.7 Provide information to the policymaker.....	108
4.8 Summary of Experiment and Result	109
CHAPTER 5	112
CONCLUSION AND FUTURE WORK	112
5.1 Conclusion.....	112
5.2 Recommendations and Future work.....	114
REFERENCES	2
BIOGRAPHY	11



List of Tables

	Page
Table 1 Summaries related work.....	15
Table 2 List of three classes of risk-prone spots.....	23
Table 3 List of POI from the internal system.....	26
Table 4 List of 14 parameters defined as input data.....	34
Table 5 Confusion matrix.....	38
Table 6 Tools and Instrument.....	39
Table 7 Summary parameter of ML Models.....	65
Table 8 ML Models measurement.....	74
Table 9 Classification of risky area ratios by zonal statistics.....	76
Table 10 The ratio of crime risk areas across eight districts.....	79
Table 11 The ratio of high-medium crime risk areas across eight districts.....	80
Table 12 The ratio of crime risk areas to each land use class.....	83
Table 13 Ratio of crime risk level to metro station.....	95

List of figures

	Page
Figure 1 Diagram show the structure of dissertation.	7
Figure 2 Study area Bangkok Thailand.....	17
Figure 3 Research Framework.	18
Figure 4 Shows risk-prone spots served by MPB.....	20
Figure 5 BKK risk map in Bangkok shows risk-prone spots served by BMA.....	21
Figure 6 Civil complaints from Traffy fondue website.	22
Figure 7 Risky point with buffer (a) 10 m (b) 20 m (c) 30 m (d) 40 m (e) 50 m (d) 100 m in Bangkok area.	24
Figure 8 Kernel density of POI (a) retail (b) Catering and recreation (c) Tour and hotel service (d) Healthcare (e) Administration (f) Financial service (g) Education (h) Business and industrial production (i) Transportation (j) Road network (k) Residential in Bangkok.	29
Figure 9 LandScan Global shows population density in Bangkok, Thailand.....	30
Figure 10 Raster file of popular density in Bangkok.	31
Figure 11 Extract urban greenspace from Sentinel-2 L2A by NDVI.	32
Figure 12 Extract nighttime light from SNPP by product VNP46A4.....	33
Figure 13 The statistics of crime against person in the Bangkok 2022.	42
Figure 14 The statistics property of crimes in Bangkok 2022.	44
Figure 15 The map displays the crime statistics of Bangkok in 2022.	45
Figure 16 The map displays statistics of tourist criminal cases in Bangkok in 2022. .	47
Figure 17 Crime risk spot count within each data source in the Bangkok area.....	48
Figure 18 POI counts for each type within the Bangkok area.	49
Figure 19 The density of retail with high-medium crime risk points.	50
Figure 20 The density of catering and recreation with high-medium crime risk points.	51
Figure 21 The density of tour and hotel service with high-medium crime risk points.	52

Figure 22 The density of healthcare with high-medium crime risk points.	53
Figure 23 The density of administration with high-medium crime risk points.	54
Figure 24 The density of financial services with high-medium crime risk points.....	55
Figure 25 The density of education with high-medium crime risk points.	56
Figure 26 The density of business and industrial production with high-medium crime risk points.....	57
Figure 27 The density of transportation with high-medium crime risk points.	58
Figure 28 The density of road network with high-medium crime risk points.	59
Figure 29 The density of residential with high-medium crime risk points.....	60
Figure 30 The density of population density with high-medium crime risk points.....	61
Figure 31 The density of greenspace with high-medium crime risk points.....	62
Figure 32 The density of nighttime light with high-medium crime risk points.....	63
Figure 33 Model accuracy with different buffer.	64
Figure 34 Feature importance of ET model.	66
Figure 35 Group average feature importance of ET model.	66
Figure 36 Feature importance of RF model.	67
Figure 37 Group average feature importance of RF model.	67
Figure 38 Feature importance of LGBM model.	68
Figure 39 Group average feature importance of LGBM model.	68
Figure 40 Confusion matrix of (a) extra tree (b) random forest (c) lightGBM.	69
Figure 41 ROC AUC of ET Model.	70
Figure 42 ROC AUC of RF Model.	70
Figure 43 ROC AUC of LGBM Model.	71
Figure 44 PRC-AUC of ET Model.	72
Figure 45 PRC-AUC of RF Model.	72
Figure 46 PRC-AUC of LGBM Model.	73
Figure 47 Classification map of a risky area in Bangkok from (a) ET model (b) RF mode (c) LGBM model.....	75
Figure 48 The map shows the districts where crime statistics against tourists were found.	77

Figure 49 A map shows the risky areas of districts were crime statistics against tourists.....	78
Figure 50 A classification risky area ratio of districts where crime statistics against tourists.....	80
Figure 51 The ratio classification area on high- and medium-risk classes to the total area of each class in districts were crime statistics against tourists.	81
Figure 52 Percentage ratio of land use within the study area in the year 2023.	82
Figure 53 A classification map shows Bangkok's risky area overlayed on land use. ..	83
Figure 54 The ratio of crime risk area classification to each land use class.	84
Figure 55 The ratio of each land use class to all high crime risk areas.	85
Figure 56 The ratio of each land use class to all medium crime risk area.	85
Figure 57 The classification of crime risk area over nighttime light.	86
Figure 58 The classification of crime risk area over population density.	87
Figure 59 Example areas a) a high-risk class in Bangkok and b) a sample of a risky area.	89
Figure 60 Example area a) a medium-risk class in Bangkok and b) A sample of a risky area.	91
Figure 61 Metro system and risky area in Bangkok.	92
Figure 62 Statistic of metro stations are in a risky area.	93
Figure 63 The ratio of crime risk areas to each metro station.	94
Figure 64 A map of metro stations located in Bangkok's districts where crime statistics against tourists are present.	107
Figure 65 A comprehensive map of metro stations with a risk level.	108

CHAPTER 1

INTRODUCTION

1.1 Background

Tourism is one of the primary engines driving Thailand's economy, accounting for 11% of the country's GDP [1]. The COVID-19 pandemic has altered the global tourism landscape. Recovering the tourism sector to be competitive and preparing the ecosystem, especially safety-first travel, to enable a sustainable transition is therefore very important and build back better. Safety and security are important in the tourism industry because tourists are often victims of crime. Due to many factors, including personal behavior, not being familiar with the area, and being noticed in the crowd, Since the 1950s, advances in transportation technologies, an increase in personal income, and more leisure time have all contributed to the growth of mass tourism. While the relationship between tourism and safety has existed since industry and academic interest began in the 1970s [2], it has gained momentum in the last few decades. However, empirical research is limited due to reluctance to raise safety issues and difficulty accessing data. This study aims to deepen our understanding of tourism safety and fill gaps in the literature. This chapter describes the research problem, objectives, research questions, expected result, the scope of the study, and the definition of terms.

1.2 Problem Statement

Thailand is one of the most popular tourism destinations in the world, especially the capital city of Bangkok. In 2022, Bangkok was able to take first place in the Business Traveler Asia-Pacific Awards 2022 in the category "Best Leisure City in the Asia-Pacific" for the 6th year in a row and entered Ranked 3rd in the category Best Business Cities in Asia. Tourists' evolving behavior and digital technology are driving the creation of new tourism models. Tourist behavior is more like a free-independent traveler (FIT) on location services. Digital platforms play a vital role in providing travel information, aiding in planning with details on services, hours, and important data. However, choosing a holiday destination can be challenging due to various factors, including budget, time, and safety considerations. Safety is a crucial

aspect to prioritize, whether traveling alone or with companions. Being aware of potential risks at different destinations helps ensure preparedness for managing common challenges during the trip.

In the year 2022 [3], there was a substantial increase in global international visitor arrivals, almost doubling in comparison to the previous year. This recovery can be attributed to the recovery after the substantial fall experienced because of the COVID-19 pandemic. Europe had a substantial surge in the number of incoming tourists, with approximately 595 million arrivals being recorded. Despite this notable yearly rise, the figures seen in Europe were lower than the pre-pandemic levels.

The potential recovery of the worldwide tourism industry in 2023 The data reveals that international visitor arrivals in the first quarter of 2023 have reached around 80% of the levels observed prior to the pandemic, corresponding to a significant recovery. The rebound can be attributed to the lifting of travel restrictions, the worldwide distribution of vaccines, and the accumulation of suppressed consumer demand. The resurgence of tourism is being led by Europe and the Americas, where visitor numbers are approaching or surpassing levels seen prior to the pandemic in many locations. The projected forecast for the remaining duration of 2023 [4] is characterized by a favorable view, as it is anticipated that the ongoing recovery will continue to grow. Nevertheless, it is worth acknowledging the fact that there are obstacles such as economic uncertainty and geopolitical conflicts.

The national tourism strategy places significant emphasis on the development of the entire tourism system to preserve the country's position as a foremost global tourism destination. Emphasize high-quality visitors [5]. Develop tourism diversity to satisfy the needs of visitors. and concentrate on developing tourism in areas with potential while preserving the country's distinctive characteristics. Persist in generating revenue through the tourism industry. Developing a foundation for tourism that prioritizes quality and international standards requires establishing travelers' trust in their safety. Ensuring safety through legal enforcement, prohibiting the exploitation of visitors, and prioritizing sustainable tourism development.

Urban transportation, especially metro stations, plays an important role as key sites for interaction within the expanding tourism framework of bustling cities. These stations serve as accessibility gateways to cultural, historical, and entertainment

centers, offering them to both tourists and residents. In the competitive landscape of urban tourism, it is important to understand the spatial relevance of metro stations related to tourist sites.

World Economic and Forum [6] report Thailand's ranking in the Travel and Tourism Development Index 2024 out of 119 countries Thailand is 47th dropped from 41st (from 2019), Thailand has an overall score of 4.12 (full score of 7), with Thailand's Safety and Security score at 4.87, lower than the ASEAN average of 5.85 for policy scores that are conducive to travel (Policy and Enabling Condition). Thailand scored only 4.48. While the regional average is 4.76, the category in which Thailand scores higher than before is the infrastructure and service category and the travel and tourism resource category. However, Thailand's environmental sustainability score is the lowest in the ASEAN region, at 4.13 points.

Safety is a big issue that Bangkok is still facing. Safe Cities Index 2021 [7] is a survey from the Economist Intelligence Unit, a global research unit on safe cities, Bangkok received an overall score of 60.2, which is lower than the average of 66.1, ranking 43rd out of 60 cities. In addition, based on INRIX [8] Bangkok still faces the highest traffic delay based on the Global Traffic Scorecard 2022, which rated Bangkok the 32nd most congested of all cities and the 2nd congestion rank in Asia, with 67 hours on average that Bangkok drivers spent stuck in traffic. The development of mass transport, especially metro systems, plays an important role in improving the quality of life in urban mobility. Because it is related to the movement of local people and tourists with large passengers, it is convenient, safe, fast, and time controllable.

Bangkok's 20-year development plan (2018–2027), according to the Bangkok as a Safe City strategy, has developed digital information of risk areas (BKK Risk Map) in Bangkok for public access to risk information related to life and property crime for use in making decisions about living in the city. the whole trip. Choosing a residential zone. Crime risk points come from surveying the area with officials and collecting information from relevant agencies. But nowadays, remote sensing technology has been widely applied to the study of the relationship between crime incident locations and the urban environment for predicting and defining risky points through machine learning methods.

This study focuses on the need for a sustainable and safety-first approach to tourism development, emphasizing the importance of understanding and developing tourism to ensure safety and promote sustainable tourism development. The classification of risky areas for safety in Bangkok, especially around metro stations, which are important for urban transportation and tourism, by utilizing remote sensing technology and machine learning methods to predict and define risky area, the study aims to identify risk areas from risk-prone spots that trend to be crime-prone areas by analyzing the relationship between influence factors and satellite imagery. Through training data like the Bangkok risk point, the police risk point, and civil complaints, it does this by using various factors comprehensive urban environment such as the Point of Interest (POI), urban greenspace (NDVI) from Sentinel-2, nighttime light from Suomi National Polar-Orbiting Partnership's (SNPP), and demographics (population density) from LanScan. Train data, divide it into training (80) and testing (20), and process in three machine learning models consisting of extra tree (ET), a random forest (RF), and a light gradient boosting model (LightGBM) are used in this method to process and check the training data with k-fold validation. In terms of accuracy assessment, use accuracy metrics like the Precision-Recall, F1 score, ROC-AUC, and confusion matrix to compare models and select the one that outperforms with high accuracy. The final output from the best model will be a map of suitable areas, improving the safety and travel experience for tourists and residents along Bangkok's metro lines and relevant agencies.

1.3 Research questions

The scientific challenge of identifying factors such as Point of Interest (POI) density, population density, nighttime light intensity, and green space requires the integration of multi-source data, including satellite imagery and government records, while also addressing inconsistencies and variations in data formats and quality. The relationship between these factors and crime risks is highly context-dependent, varying across urban landscapes, from central business districts to suburban areas. Developing a machine learning classification model has complexity due to the evaluation and comparison of several methods (e.g., Extra Trees, Random Forest, and LightGBM) with accuracy and interpreting results with high- and medium-risk zones.

Selecting input variables (features), which include diverse data such as spatial attributes, demographic patterns, and environmental factors. Ensuring the model accurately predicts crime-risk areas provides information for policymakers and urban planners to improve safety in Bangkok.

What is the impact of different urban environment factors and satellite imagery features on Bangkok's crime-risk classification models' accuracy?

How can we accurately represent the spatial distribution of high, medium, and low crime-risk areas along Bangkok's metro lines on a comprehensive map to tourists and residents?

How can we use the classified crime risk data to improve urban safety, offer tourism recommendations, and enhance public safety and tourist experiences in Bangkok?

1.4 Research objectives

Tourism is an important economic activity on a global scale, with the quality of transportation systems playing a crucial role in improving tourist experiences inside urban areas [9]. Furthermore, there exists an interdependence between transport and tourism, wherein each domain exerts an effect on the other [10]. The metro system offers higher comfort levels compared to the bus system. This is critical, as it highlights the importance of comfort influencing public transport preferences and the potential of new technologies to enhance public transport services. The comprehensive approach to enhancing urban safety and promoting sustainable tourism in Bangkok involves a strategic mix of technology and data analysis. The objectives aim to utilize machine learning and satellite imagery for urban planning and safety enhancements.

To develop and train a machine learning model for accurate classification of crime-risk areas in Bangkok using the urban environments and satellite imagery factors.

To develop a comprehensive map that identifies high, medium, and low crime-risk areas of Bangkok's metro lines for tourists and locals.

To provide recommendations for urban safety improvements and tourism guidance based on classified data.

1.5 Dissertation structure

This dissertation includes 5 chapters shown in Figure 1. In the following chapters will introduce a theoretical and scientific introduction, data and methods, and conclusion.

Chapter 1 presents an overview of the tourism trend both globally and in Thailand, especially in the Bangkok metropolitan area. The important and sustainable tourism development in Thailand emphasizes safety-first travel to build back better. This chapter also includes the research question and research objectives of this study.

Chapter 2 provides a literature review that relates to public safety and crime risk assessment areas with integrated machine learning techniques and factors consisting of urban environment and remote sensing data, emphasized satellite imagery, and crime risk area classification.

Chapter 3 describes the details of the research methodology of this study, which includes study areas, datasets, software, methodology workflow, and details of the processes.

Chapter 4 describes the results from the study, including analyzing factors, comparing machine learning models, and creating a comprehensive map of crime risk areas across Bangkok.

Chapter 5 conclusion shows the findings and suggestions.

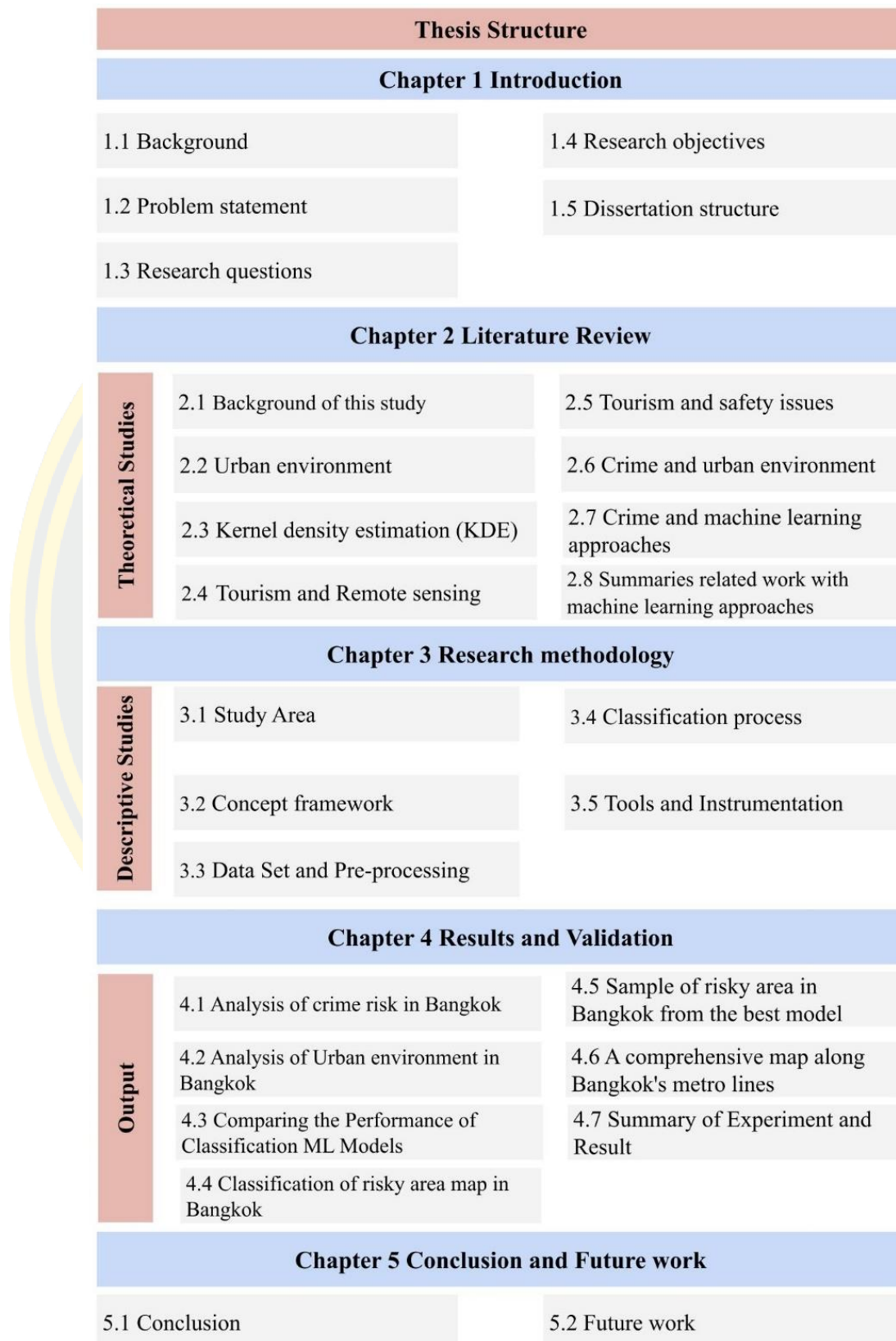


Figure 1 Diagram show the structure of dissertation.

CHAPTER 2

LITTERATURE REVIEW

2.1 Background of This Study

Bangkok, the capital city of Thailand, is a special administrative region under the Bangkok Metropolitan Administration (BMA) [11, 12]. Bangkok has experienced rapid urbanization and has the highest population density in Thailand. However, a lack of proper urban planning has led to a disorganized urban environment and insufficient infrastructure. Urbanization is strongly correlated with social crime [13]. For instance, rapid urbanization, which involves the design of buildings and infrastructure for housing and transportation, as well as the accessibility and layout of outdoor spaces, can impact crime rates. According to Bangkok, it is ranked the 30th most dangerous city for tourists compared to 60 international cities with a risk score of 53.34 [14]. The criminal and health security categories had the highest risk scores, scoring 39. The infrastructure risk was assessed at a score of 22, the digital security risk at 17, and the personal security risk at 5. Although local authorities have improved safety, including CCTV installation and increased police presence, there is a certain lack of confidence from the public [15].

2.2 Urban environment

The urban environment is a feature of cities [16]. It consists of the built environment in and around a city, such as buildings, roads, transport networks, shops, markets, greenspace, and other public infrastructure. These locations serve as hubs for economic, cultural, and social activities. The urban built environment influences people's behavior [17]: the effect of the urban built environment is that a lack of planning and the negligence of the main principles of crime prevention through environmental design (PCPTED) with physical and social disorders tend to have more crimes. Crime incidences are related to built environment, socioeconomic, and demographic factors [18] Land use is important for predicting crime rates. The crime rates tend to be higher in areas with high-density residential developments and commercial property, and lower in areas characterized by industry, parks, and schools. Any change in land use may lead to other changes in nearby areas. For instance, the

development of a new community facility may change the routine activities of the community involved and create crime opportunities for the offenders.

Routine Activity Theory (RAT) is a criminological theory that explains how the convergence of motivated offenders, suitable targets, and the absence of capable guardians can lead to criminal events [19]. Motivated Offenders refers to these individuals having a desire to engage in criminal activities, influenced by economic, psychological, or situational variables. Suitable targets refer to the characteristics of a target, such as its visibility, accessibility, or perceived value, that affect its attraction to offenders. Capable guardianship includes law enforcement as well as situational factors such as lighting, community awareness, and the implementation of security measures. RAT emphasizes the importance of routine behaviors in influencing the patterns of crime. Routine activities indicate the routine behaviors individuals engage in during their daily lives, going to work, school, or social engagements. RAT proposes that changes in daily routines and the spatial and temporal environment of criminal occurrences might affect crime rates.

Crime Pattern Theory (CPT) provides a framework for understanding the spatial and temporal distribution of criminal activities. It integrates individual routine behaviors with the built and social environments to explain the temporal and spatial likelihood of criminal occurrences. Crimes are not randomly distributed, but the pattern is influenced by the everyday activities and routines of offenders. Offenders' familiarity with areas due to their daily routines, making them more prone to commit crimes in known and frequented areas, as this familiarity reduces their chance of arrest. This theory is built on the rational choice perspective and the routine activity approach, which examine the decision-making processes of criminals and the routine activities of potential victims.

Point of Interest (POI) represents the multidimensional of the urban feature such as restaurants, parks, and stores, which led to urban vitality [20] and attraction in urban areas. The distribution and different types of POIs can indicate the regional centrality of different urban regions [21], including the functional characteristics of urban areas. The area with a high density of POI is usually defined as hubs for economic and social activities [22] because the increased human activity and social interactions tend to be more central and accessible. For example, a high density of

educational institutions and libraries might indicate an educational hub, while a high density of cafes and coworking spaces could be a business hub. Understanding the characteristics of POIs can be helpful in urban planning, identifying underserved areas, and planning for future growth. In addition, a variety of POIs influence travel patterns within the city and attract more visitors, affecting traffic flow and public transportation usage.

2.3 Kernel density estimation (KDE)

Kernel Density Estimation (KDE) is a nonparametric method [23] that estimates the spatial distribution of data points without making any prior assumptions about the distribution. The technique assigns higher weights to data points closer to the center of the calculation area and lower weights to those further away, resulting in a weighted average density of all points in the study area. KDE can be used to visualize crime hotspots or building density [24] by generating a continuous, smooth surface map on a grid. The influence of environmental factors on property crime using potential risk factors [25] comes from human activity (population density and night lighting), socio-economic factors (road network density, distance to police stations, distance to main roads, distance to bus stops), and infrastructure (average house price, entertainment density, hotel density, and shop density) and the night light data. The Average Nearest Neighbor and Kernel Density Estimation were utilized to inspect the heterogeneity of the spatial distribution of crime and used Bayesian regression and geo-detector to jointly explore the driving factors of property crime. The result showed that shop density, hotel density, entertainment density, and house price were the four dominant environmental drivers of property crime.

2.4 Tourism and Remote sensing

Remote sensing is an important tool that can improve the tourism industry by providing detailed satellite imagery and GIS data for tourism management. Multiple Criteria Decision Analysis (MCDA) models [26] are applied for ecotourism suitability analysis. And using GIS and AHP by collecting remote sensing and crowdsourced data. Landcover type is one parameter that helps to understand the general characteristics of the surface and using the CART (Classification and

Regression Trees) classifier in the Google Earth Engine (GEE) platform and then all raster data layers, and their classes were overlaid to develop a final suitability map.

2.5 Tourism and safety issues

The safety issues related to tourism impact the utilization of tourism services in Bangkok [27]. From foreign tourists A representative sample of foreign travelers visiting various important tourist attractions There were 400 individuals in Bangkok. The study revealed that safety significantly influenced both the decisions made before travelling and the level of satisfaction experienced by tourists during and after their trip. Foreign tourists' satisfaction with tourism activities in Bangkok is generally moderate. Life safety is an essential aspect of ensuring safety during travel. Ensuring the safety of properties and protecting public health foreign tourists' utilization of tourism services in Bangkok is impacted.

2.6 Crime and urban environment

Remote sensing was used to study the relationship between crime and urban layout [28] and reveal that areas with high homicide rates tend to have high local variance and less general homogeneity in line with the broken windows theory and CPTED, in terms of that metropolitan areas with more different and disordered layouts tend to have a higher rate of homicide. Patterns of criminal activity [29] were studied using mapping and spatial analysis techniques to examine the analyzed crimes, including robbery, theft, motor vehicle theft, and homicides. The empirical proof of a non-random spatial concentration for all crimes. The occurrence of crimes increases as the distance from the closest police station grows, and police stations have a deterrent effect on crime. The relationship between street environmental data and property crime [30]. With statistical analysis, the results show that the street environment and design have a significant impact on the incidence of crime. The relationships between various elements, such as the frequency of road use and the function of buildings along the road. The placement of furniture and roadblocks and crime rates; the environmental factors influencing theft are banks and ATMs; and the factors influencing robbery are restaurants and schools.

An analysis of machine learning algorithms for predicting crime hotspots [31] by using the historical data of public property crime in high population density communities. The study found the built environment data of points of interest (POIs) and urban road network density are input into the LSTM model as covariates have better prediction than the original model that is based on historical crime data alone. A literature review using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist [32] of 45 papers was selected from 30,000 papers covering research in urban areas (not rural or suburban) that involved green space area to crime data and revealed the presence of parks and other green spaces reduces urban crime.

A statistical technique known as multilevel modeling is used to investigate and predict the relationship between crime risk and green spaces [33]. Green Space operates by the value for NDVI, and combined with potential factors linked to crime, such as demographic, socioeconomic, and climate variables, the study found that census block groups with more greenspaces had a lower risk of both property and violent crime. While using remote sensing data from Sentinel-2 imagery and crime data [34] in a supervised Random Forest classification model revealed the increase in total green space will decrease in violent crime and property crime.

The effect of ambient light level on crimes that influence lighting on crime and the impact of darkness on crime location [35], which comprises commercial, education, government, healthcare, hotel, leisure, open space, residence, retail, street, transportation, and vehicle. robbery will increase during darkness, and darkness leads to an increase in arson and curfew loitering offenses. A pilot study to investigate the increased lighting would be effective in reducing serious crime [36], temporary streetlights that were randomly allocated to high-priority areas of crime rates and revealed more lighting can reduce nighttime outdoor index crimes, resulting in a decline in arrests, indicating the effectiveness of street lighting in deterring crime.

2.7 Crime and machine learning approaches

2.7.1 Machine learning.

Machine learning (ML) methods for crime classification and prediction were applied for much research by utilizing ML algorithms to analyze historical crime data.

ML models can handle large datasets with various crime types and influencing factors. Review over 150 articles [37] to explore various machine learning and deep learning algorithms applied to crime prediction. The study reveals distribution of crime-related articles among Science Direct and IEEE databases found that 67% of articles have used ML and 21% have used DL. In terms of the article's technique's classes, 63% of articles have focused on the classification task, which is the prime focus, utilizing various ML and DL-related techniques. While a systematic literature review (SLR) methodology for machine learning-based crime prediction [38], which focused on 68 selected machine learning papers from five highly reliable and trustworthy scientific databases: the ACM Digital Library, IEEE Xplore, Springer Link, Science Direct, and Scopus. The study defined all data using eight quality assessment questions, including any papers scoring above 4.5. The study discovered the use of five types of datasets: actual crime records, traffic data, location data, visual data, and text data. The study identified 14 different machine-learning algorithms. The most used algorithms were artificial neural networks, random forests, and KNN algorithms. The evaluation metrics used more than five times are accuracy, root mean square error (RMSE), mean square error (MSE), mean absolute error (MAE), precision, recall, area under the ROC curve (AUC), F1 score, hit rate, and R2 score. Law enforcement agencies can improve urban safety by utilizing this methodology to support proactive policing, reduce crime rates, create comprehensive crime-risk maps, and provide safety recommendations for locals and tourists.

2.7.2 Random Forest (RF)

Random forest (RF) is an ensemble learning method commonly used for classification and regression functions. RF is a tree-based approach that combines multiple decision trees to effectively handle imbalanced data and reduce overfitting. The result of the model gathering output from all the trees as a single result to improve classification and prediction accuracy. From this point, RF makes a robust and reliable crime classification and prediction of crime hotspots and trends [39, 40]. RF can improve crime prediction at micro places. According to the study, it focuses on robbery and use Crime generators [41] are larger business retailers that include department stores, hardware stores, grocery stores, pharmacies, and dollar and craft stores. Smaller food and clothing stores include bakeries, herbal supplement stores,

second-hand stores, and clothing stores. Eating and drinking places (aka restaurants and bars) include movie theaters, amusement parks, and recreation services. The last category includes gyms, bowling alleys, museums, and hair salons to generate long-term crime forecasts. The study reveals RF's outperform predictive accuracy and effectively debunked the idea that machine learning is a black-box process by demonstrating how interpretable summaries can clarify the reasons behind the identification of high-risk locations. Furthermore, it revealed the non-linear and spatially varying characteristics of factors that influence crime.

2.7.3 Light Gradient Boosting Model (LightGBM)

LightGBM (Light Gradient Boosting Machine) is a gradient-boosting framework that improves efficiency, scalability, and accuracy [40]. LightGBM effectively handles large datasets and high-dimensional data that is suitable for crime classification and prediction. Due to the outstanding features of speed and high accuracy that make LightGBM popular with researchers. The relationship between crime incidence and urban environment [42] using big data such as smart civil complaint data, street image data, and points-of-interest (POI) including accommodation, restaurants, retail stores, bars, leisure facilities, offices, and residential facilities. The built environment variables related to crime in the city, such as CCTV, building age, number of bus stops, and distance to a subway station. Meanwhile, the variables of subway stations and bus stops represent the accessibility of public transportation. The crime incidence data processed on LightGBM showed the best predictive power that confirms a nonlinear relationship, identifying inflection points and interaction effects of urban environmental variables. LightGBM was utilized in the classification of Large Scale Land Use and Land Cover [43] with the study area being a mixed-use landscape functional urban region by using Sentinel-2 satellite images to compare three machine learning techniques that reveal that LightGBM are more accurate and perform around 25% quicker on average.

2.7.4 Extra Trees (ET)

Extra Trees (Extremely Randomized Trees) is a tree-based support classification and regression problem [44]. The Extra Trees (ET) algorithm fits each decision tree on the entire training dataset and selects split points entirely at random that ensures each tree is built using all available data, enhancing the model's

robustness. Bagging and random forest methods built each decision tree from a bootstrap sample of the training dataset and used a greedy algorithm to select the optimal split point based on criteria. In terms of crime study ET was utilized to analyze the 6 categories of urban indicators on crime risk [45], including geography, economy, education, housing, urbanization, and population, that revealed the three main factors influencing crime occurrences are the residential land area, the number of mobile telephone customers, and the employed population. While crime studies based on characteristics of the residential population from traditional census data can improve [46] models' quality by extending more features with spatial and spatiotemporal data to describe the ambient population.

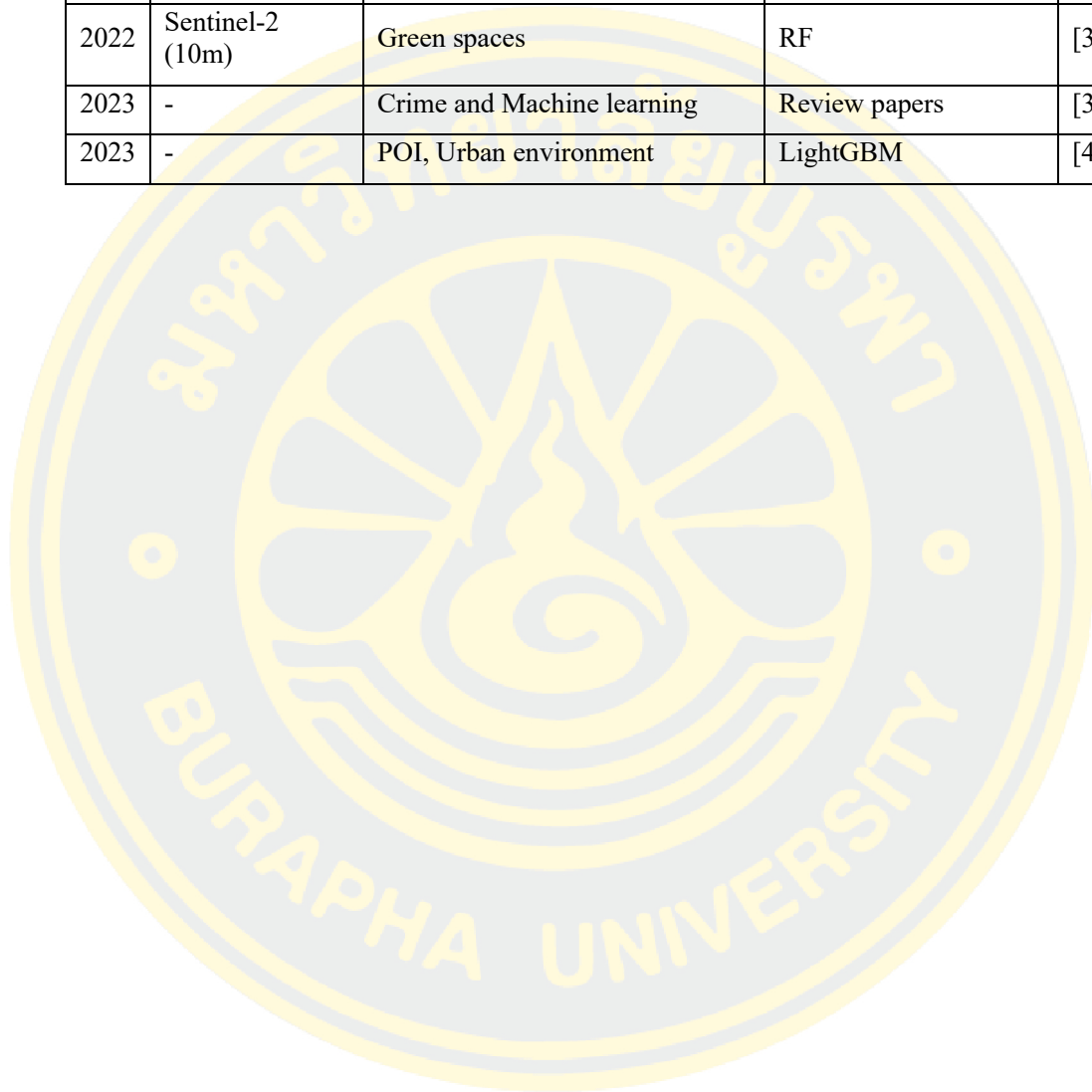
2.8 Summaries related work with machine learning approaches

This table summarizes studies that analyze the relationship between urban factors and crime risk using satellite data. Studies used different satellite images to examine factors like urban layout, green spaces, crime incidents, and points of interest (POI). Researchers employed various methods such as Extra Trees, LightGBM, Random Forest (RF), and spatial or statistical analysis. The related research mainly focuses on urban metrics, crime generators, and environmental factors such as green spaces to predict or analyze crime risk. The ML model was applied for studying in various domains, especially crime analysis, as shown in Table 1. The finding is useful to understand the method and factors that influence the model.

Table 1 Summaries related work.

Year	Satellite	Factors	Method	Ref.
2014	Quickbird 0.6 m	Urban layout	Maximum likelihood algorithm in ENVI	[28]
2019	-	Green space area	Literature review	[32]
2020	-	POI, crime incident	LSTM	[31]
2020	Sentinel-2	LCLU	LightGBM	[43]
2020	-	Urban metric	ET	[45]
2021	-	Police stations, crime incident	Spatial analysis	[29]
2021	-	Street environment, crime incident	Statistical analysis	[30]

Year	Satellite	Factors	Method	Ref.
2021	-	Crime generators and demographic factors	RF	[41]
2022	Landsat 8 (30m)	Green spaces	Pearson's	[33]
2022	Sentinel-2 (10m)	Green spaces	RF	[34]
2023	-	Crime and Machine learning	Review papers	[37]
2023	-	POI, Urban environment	LightGBM	[42]



CHAPTER 3

METHODOLOGY

This chapter provides the details of the research methodology of this study which includes study areas, datasets, software, methodology workflow, and details of the processes.

3.1 Study Area

As the spatial scope of this study, Bangkok is a metropolitan city in Thailand and is an area of high population density, as of 2023 population is now estimated at 11 million, the city covers 1,568.7 square kilometers in central Thailand's [47]. The latitude area is 13.45 degrees north, longitude 100.28 degrees east. Bangkok's 50 districts and 160 sub-districts are administrative subdivisions under the Bangkok Metropolitan Administration's (BMA) control [48]. Bangkok includes land uses such as residential areas, commercial areas, and industrial areas, representing various urban environments shown in Figure 2.

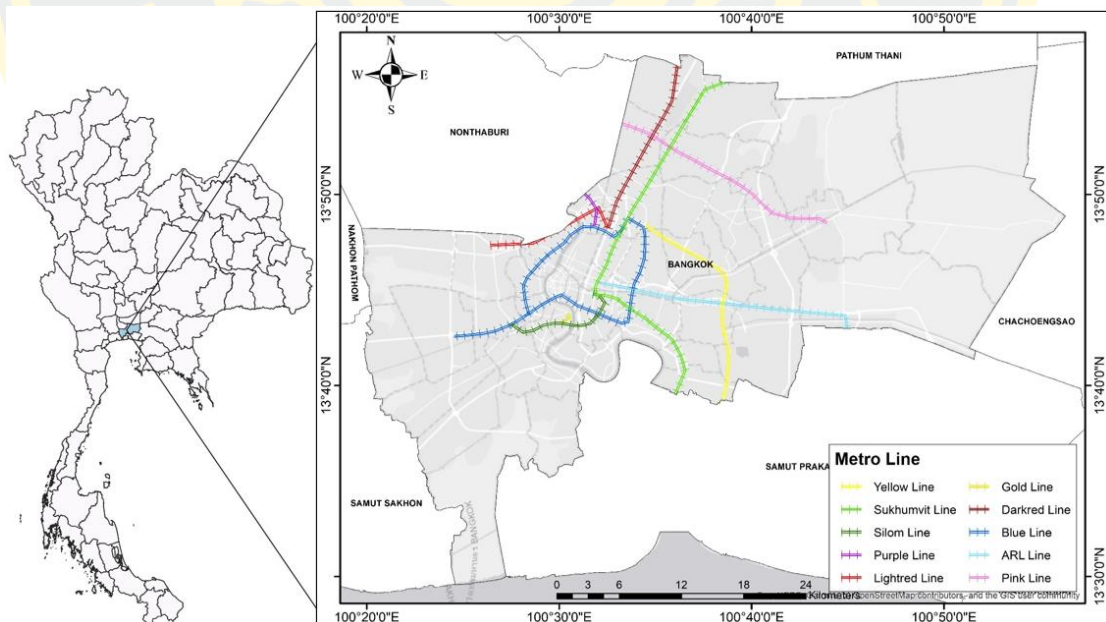


Figure 2 Study area Bangkok Thailand.

3.2 Concept framework

The general idea of this study is shown in Figure 3. The process is divided into two parts are follow. The first part is a part of pre-processing it utilizes a crime risk point dataset including Bangkok risk point, police risk point, and civil complaint as data train and extracting urban environment data comprised of satellite images from Sentinel-2 and Suomi National Polar-Orbiting Partnership's (SNPP), TripAdvisor, Google Maps, open data, LandScan website, and internal system defined as data input.

The second part is a classification process, this part use three different machine learning models to compare and choose the best model by splitting the training data into train data (80%) and test data (20%), and 10-fold cross-validation is used to prevent overfitting. After that, various types of machine learning and statistical models were analyzed and evaluated, and the model with highest accuracy was chosen.

Finally, the result image from the best model will be used to overlay the metro line in Bangkok to create a suitable area map for tourists. The study's findings can be applied to developing methods for risk area assessment, urban planning, and the making of policies that are focused on an understanding of the environments surrounding a risky area.

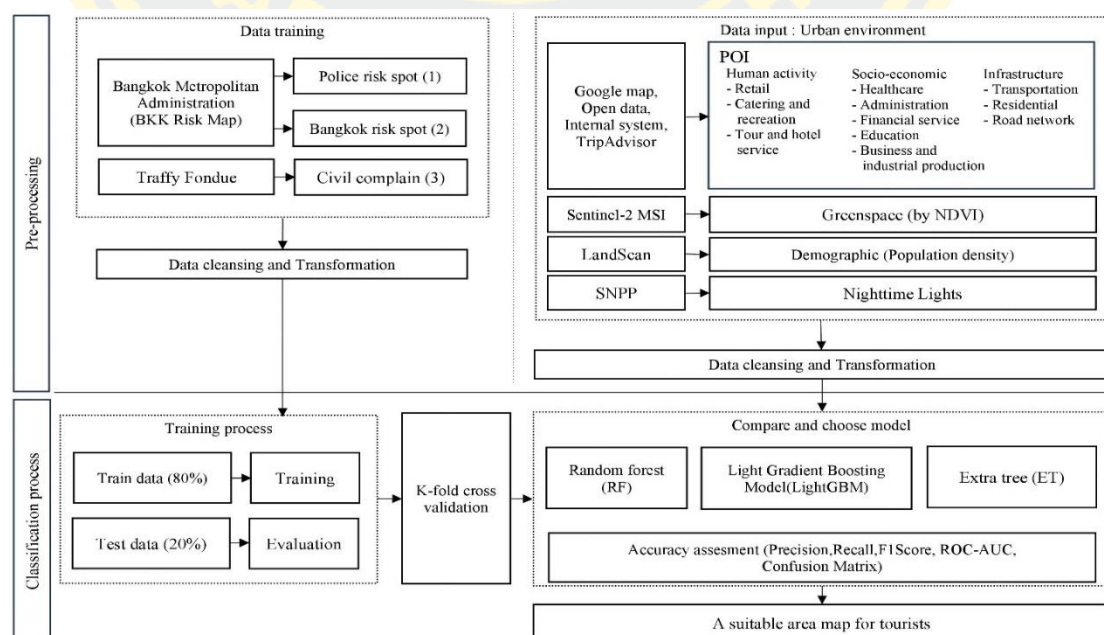


Figure 3 Research Framework.

3.3 Data Set and Pre-processing

The physical characteristics of a city are an important factor that contributes to urban crime. Areas with high crime rates are often those that attract many activities [49] related to accessibility, which in turn benefit from the theory of movement economy. Clusters of buildings or various commercial components, densely packed, attract people from different areas. This results in densely populated, bustling and chaotic areas. This has an impact on the population's social well-being. The study of crime problems in terms of physical aspects leads to management and urban planning that minimize crime.

Bangkok, the capital and largest city of Thailand is a city with mixed-use developments comprised of residential, commercial, and public spaces in an urban environment. This will include high-rise buildings, bustling streets, shops, transportation network, and public areas, all within the context of Bangkok's.

As Bangkok is a large city and the center of the country, it is also a famous tourist city that is popular with tourists around the world. As a result, Bangkok still has many social problems, especially the safety of life and property that affect the livelihoods of citizens and tourists. Responsible government agencies such as the BMA and the Metropolitan Police Bureau have therefore collaborated to survey and reveal crime risk points or crime-prone places in Bangkok to the public through the official website, which is open data that everyone can access. It is not only beneficial for the people. But it is also important information useful for research studies to develop new methods for assessing risk areas. This research has therefore collected data on risk points and spatial factors that relate to the urban environment, including POI, built environment, road network, and demographics, including the application of information from high-resolution satellite images to extract information on deserted areas and dark areas that result in the area where there is a risk. All the above information is divided into two groups: data training and data input, as follows.

3.3.1 Data training (Crime risk points)

In Bangkok, most crimes across the city occurred at night, pedestrian overcrossings were found to be the most at-risk among crime-prone areas in public areas, while others were small alleys, roads, and public parks, respectively [50]. The BKK Risk Map is a digital database on risky locations in Bangkok, providing

risk-prone spots that tend to be crime-prone areas. For risk management, the Bangkok Metropolitan Administration (BMA) gathers data from various agencies, both within and outside its jurisdiction, covering various aspects. The BMA aims to create a holistic approach to disaster management, involving data collection on various factors like dark spots, crime rates, accidents, and flooding. This information will be used to develop the BKK Risk Map websites, which serve to raise community awareness and preparedness [51].

Police risk points, Risk management requires complete and holistic data on circumstances and environments in each area, which must be fed from several agencies both within and outside the supervision of the BMA, such as the Metropolitan Police Bureau (MPB), which also provides 336 risk points shown in Figure 4 on the BKK Risk Map.

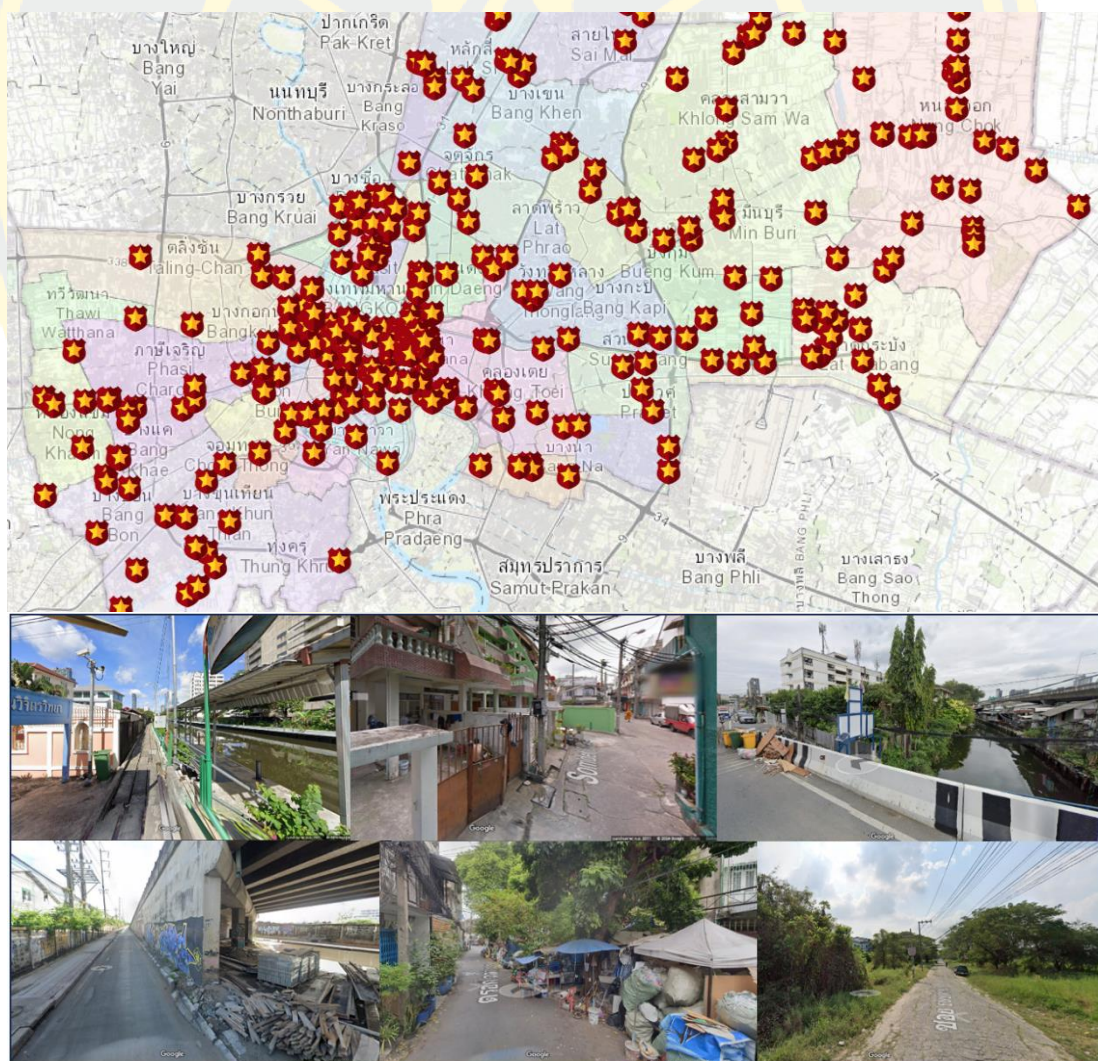


Figure 4 Shows risk-prone spots served by MPB.

Bangkok risk points come from the city law enforcement officer's exploration of the area and accumulate a database of relevant information, such as dark locations and crime rates in 2022, to analyze the collected data to address the issues, improve infrastructure to prevent threats, or support risk response. The BKK Risk Map provides information about the risk-prone spots and crime-prone areas that pose a risk to the body, life, and property of people in Bangkok. It provides details about the position, characteristics of the location, and risk/danger characteristics of 347 risk points, as shown in Figure 5.



Figure 5 BKK risk map in Bangkok shows risk-prone spots served by BMA.

Civil complaints, Brightly-lit Bangkok is one of the governors of Bangkok to create more friendly environments that are safe for nighttime travel. BMA officials and Bangkok residents are both capable of monitoring the readiness and effectiveness

of public lighting systems. Residents can promptly submit reports through an application, thereby ensuring adequate lighting and safety. insufficient lighting problems. Civil complaints are spatial data because the reporting location can be studied, and the urban environment can be assessed. This study filters only the dark-spot safety categories of civil complaint data. In 2022, 6,507 reports appeared in Bangkok shown in Figure 6.

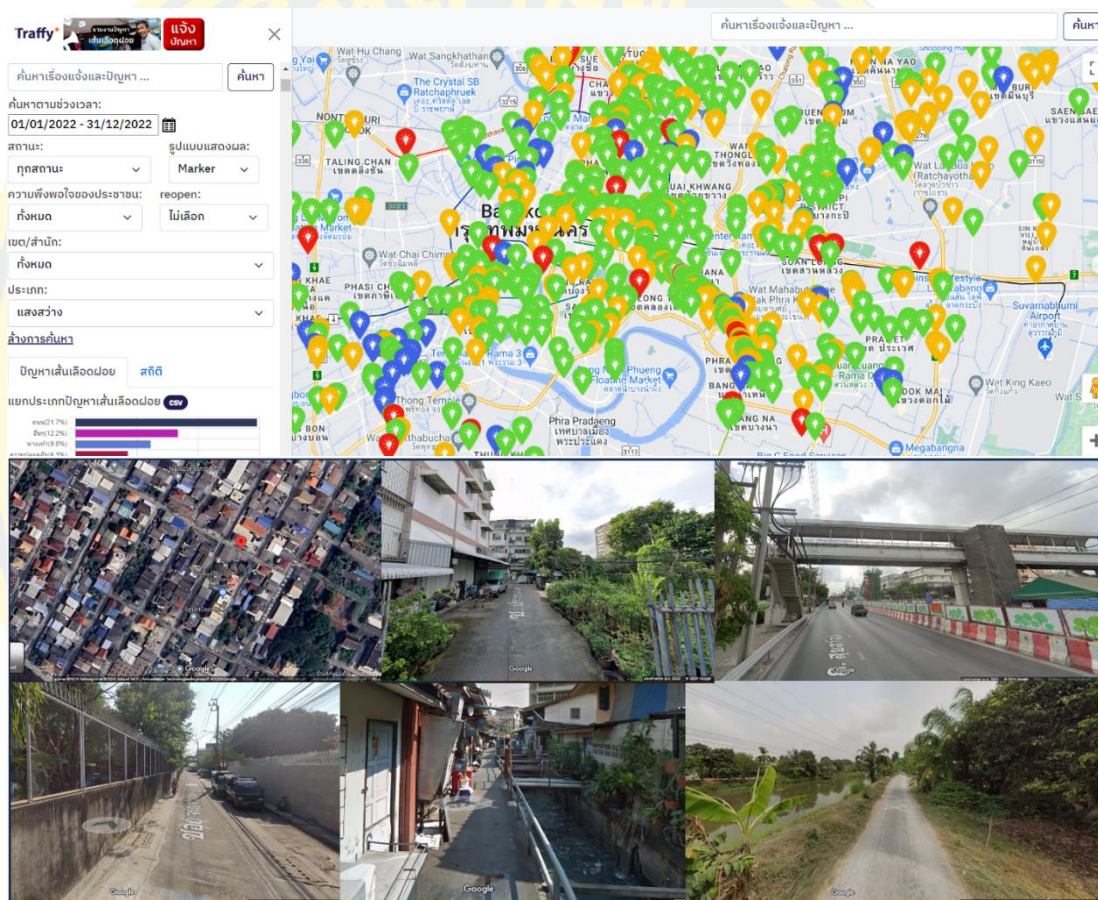


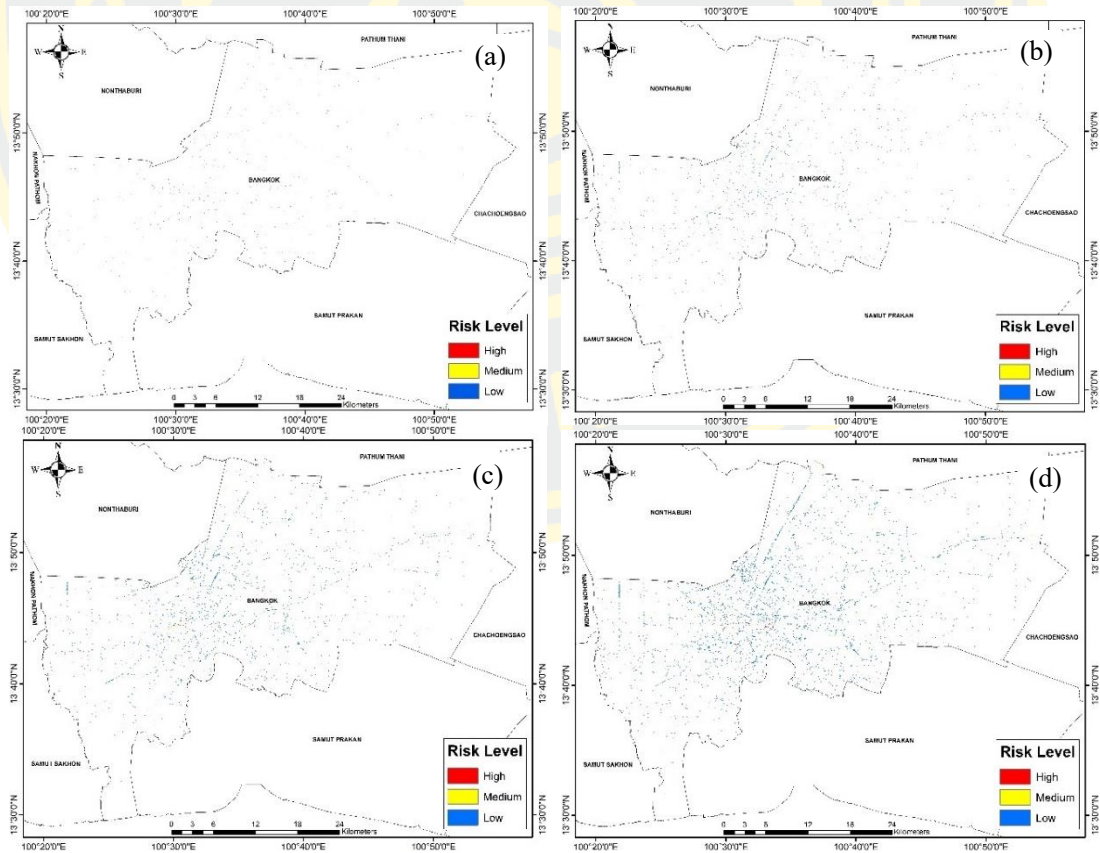
Figure 6 Civil complaints from Traffy fondue website.

Three different classes are labelled from assessments by five experts in Table 2 that relate to police risk points referring to the areas identified as high-risk areas based on crime data; Bangkok risk points refer to the locations that were identified based on urban-specific risks, environmental, demographic, or infrastructure factors; and the civil complaint risk point datasets refer to the areas identified by public complaints, indicating perceived safety issues or other community concerns.

Table 2 List of three classes of risk-prone spots.

Class (Label)	Class value	Data set	Point
High	1	Police risk point	336
Medium	2	Bangkok risk point	347
Low	3	Civil complaint	6,507

A risk dataset can be utilized to represent crime risk points using buffer tools. For different buffer distances of each class (10, 20, 30, 40, 50, and 100 meters) surrounding each risk spot, this step is repeated as shown in Figure 7. Machine learning models can be used to find the optimal buffer size that maximizes the model's performance by comparing its accuracy over a range of buffer distances. Using this method not only helps us understand the spatial distribution of risk areas, but it also improves our predictive models for urban planning and crime prevention.



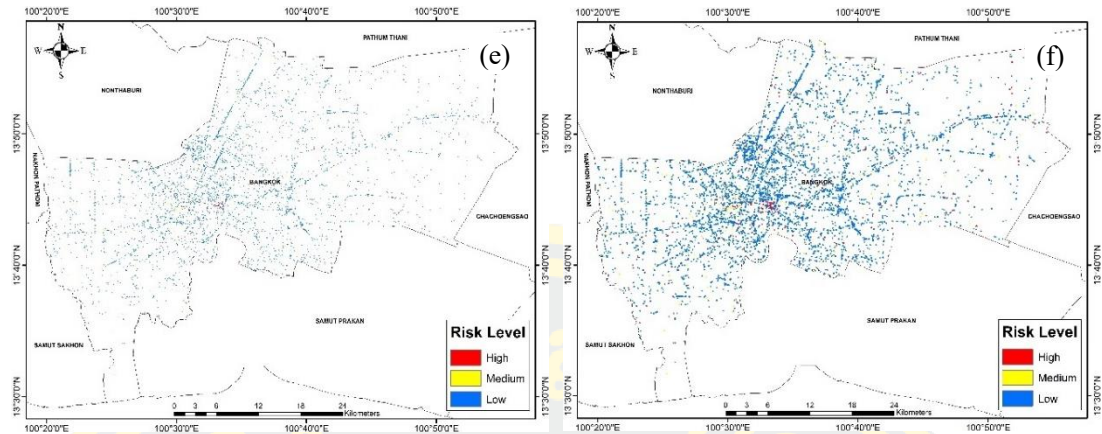


Figure 7 Risky point with buffer (a) 10 m (b) 20 m (c) 30 m (d) 40 m (e) 50 m (d) 100 m in Bangkok area.

3.3.2 Data input (Urban environment factors)

Urban environment is a feature of the city that can identify the main functions of the city and is also related to urban land use. The major urban function types within a city [52] can be categorized as a Point of Interest (POI) and usually include residential, retail, catering and recreation, financial service, education, healthcare, administration and other public service, tour and hotel service, and business and industrial production. The POI data were collected from four primary sources: TripAdvisor, Google Maps, Open data, and Internal system.

TripAdvisor, a widely recognized online review platform in the tourism sector, is known for its role in understanding user perceptions, including tourist perspectives on transportation modes. It offers details on travelers' experiences at tourist attractions, encompassing ratings, reviews, and more [53]. TripAdvisor is recognized in research for its extensive data, credibility, and universal acceptance among the research community [54]. Many studies explore consumer assessments using TripAdvisor as an essential information source. This study chose the Bangkok area because, as an international metropolis with a lot of famous, it attracts consumers from all over the world, which reduces concern about culture and area. It collects data from TripAdvisor (Tourist Attractions, Hotels, and Restaurants), which likely includes data relating to geographical locations and POI in Bangkok. The data extraction and compilation of attractions from TripAdvisor were part of the data collection process using the Python script.

Web scraping is the process of gathering and processing data from the

internet, and Python has created powerful tools to help with this. Because the internet is a massive source of information, web scraping is vital for gathering and analyzing data from websites in sectors such as data science, business intelligence, and investigative reporting [55]. To web scrape every URL and extract information from TripAdvisor with searchable information in Bangkok, a Python script was developed using Microsoft Visual Studio Code (VS Code) version 1.83.0 software, an open source.

The first step was to identify the URL of the attraction by filtering it with "Bangkok, Thailand." Second, the web pages were scraped, and the script used BeautifulSoup, a Python library version 3.11.1, to retrieve data from HTML. Each attraction web page collected included a title, the number of reviews, a rating from 1 to 5, duration, detail, the area, neighborhood, attraction category, location, and popular mention. Third, the results were exported into an individual CSV file. Finally, cleansing the data from csv file preparing for next step.

Google Maps is a web-based mapping service by Google, offering maps, satellite imagery, and real-time traffic data. It provides directions for various transportation modes and a street view feature. Google Maps is a versatile location-based tool for navigation and finding places of interest [56].

This study was conducted in the Bangkok metropolitan area of Thailand. In the year 2023, the existing mass transit system in Bangkok consists of four distinct types of metro systems: 1) the elevated BTS system, commonly referred to as the Skytrain, which opened in 1999; 2) the underground MRT (Metropolitan Rapid Transit) is a mass rapid transit system operated by Bangkok Expressway and Metro (BEM), which opened in 2004; 3) the Airport Rail Link (ARL) operated by the State Railway of Thailand (SRT), which opened in 2010; and 4) the SRT Red Lines are two commuter lines (SRT Dark Red Line and SRT Light Red Line) that are operated by the State Railway of Thailand, which opened in 2021. While there is no physical track connection between the systems, the metro stations were collected from Google Maps and the official website in June 2023 [57].

Internal system, this study utilizes built environment as a factor in the processing model. The important role of physical characteristics of the built environment and social factors induce or inhibit violent crime and property crime [58,

59]. POI is one of the urban environmental variables that has a relationship with crime incidence [42]. Potential risk factors come from human activity, socio-economic factors, and infrastructure [36]. Therefore, the POI was extracted from all data source defined as 11 parameters shown in table 3.

Table 3 List of POI from the internal system.

Category	POI parameters	Detail	Amount
Human Activity	(1) Retail	Retail store	34,522
		Supermarket	770
		Convenient Store	4,870
		Bookstore	2,407
		Material Store	6,670
		Automobile	13,167
	(2) Catering and recreation	Fast food	790
		Coffee shop	7,110
		Theatre	66
		Restaurant	7,160
		Recreation	3,150
		Nightclub	2,704
		Spa and beauty	15,320
	(3) Tour and hotel service	Hotel	3,976
		Tour	963
Socio-economic	(4) Healthcare	HealthCare	10,801
	(5) Administration	Government	5,828
		Religious	6,763
		Police	282
		CCTV	234
	(6) Financial service	Financial	25,273
	(7) Education	School	7,279
	(8) Business and industrial production	Company	47,931
		Office building	2,444
		Factory	4,900
Infrastructure	(9) Transportation	Farm	125
		Gas station	1,081
		Parking	559
		Metro stations	151
		Bus stop	6,290
		Train stations	41
		Pier	124
	(10) Road network	Road network	-
	(11) Residential	Residential	24,373

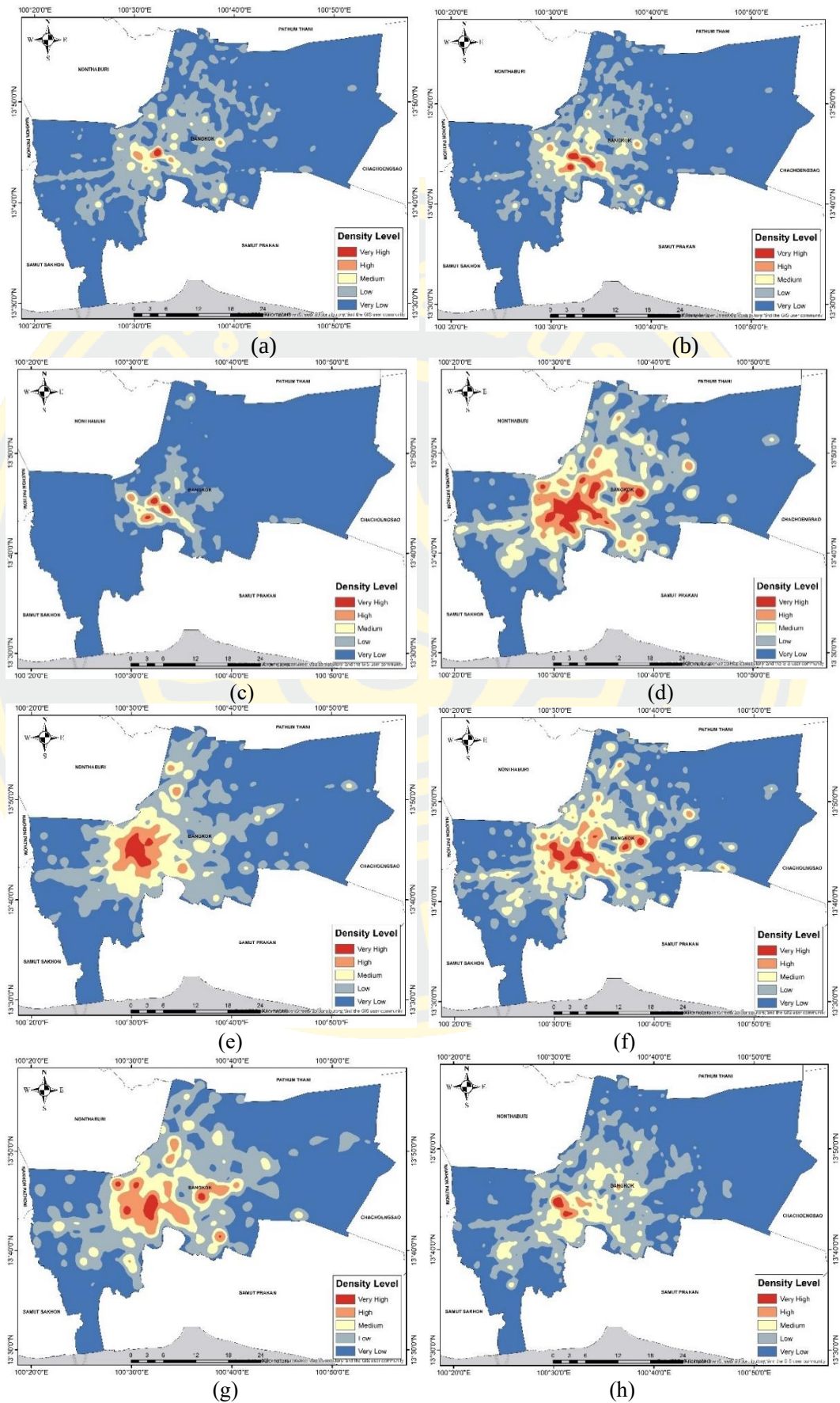
The human activity category covers retail, catering and recreation, tour and hotel services. Area with high retail activity areas could be considered risky due to higher foot traffic and potential for crimes like theft. Retail density can be a significant predictor of crime rates. Catering and Recreation is an area with a lot of

catering and recreational facilities that might be deemed risky because they attract large crowds, increasing the likelihood of accidents or crimes. Tour and hotel service with tourist-heavy areas can be high-risk due to potential crimes targeting tourists and higher population density.

The socio-economic category covers healthcare, administration, financial service, education, business and industrial production. Healthcare facilities might be seen as less risky unless considering factors like biological hazards or emergency situations. However, the accessibility of healthcare services can influence the overall risk profile of an area. Administrative regions may have varying risk levels depending on the presence of government buildings which could be targets for protests or attacks. Financial districts may carry risks related to economic crime such as fraud or theft. Educational institutions might have lower physical risks but could face issues like data breaches or on-campus violence. Business and industrial production areas can be at risk due to workplace accidents, environmental hazards, or economic disruptions affecting production.

Infrastructure category covers transportation, road network, residential. Transportation hubs are often considered risky due to the high volume of people and goods moving through them, which can lead to accidents or criminal activities. Residential areas' risk levels depend on factors like crime rates, building safety standards, and access to emergency services. Road network is more extensive the road network is, the higher the risk for traffic accidents and congestion-related issues. In addition, road density is related with urban expansion and population growth [60].

The POI data are the eleven datasets that can be utilized to represent density using Kernel Density Estimation (KDE). All data quantify the density of POI by creating a smooth surface over each dataset using KDE as shown in Figure 8.



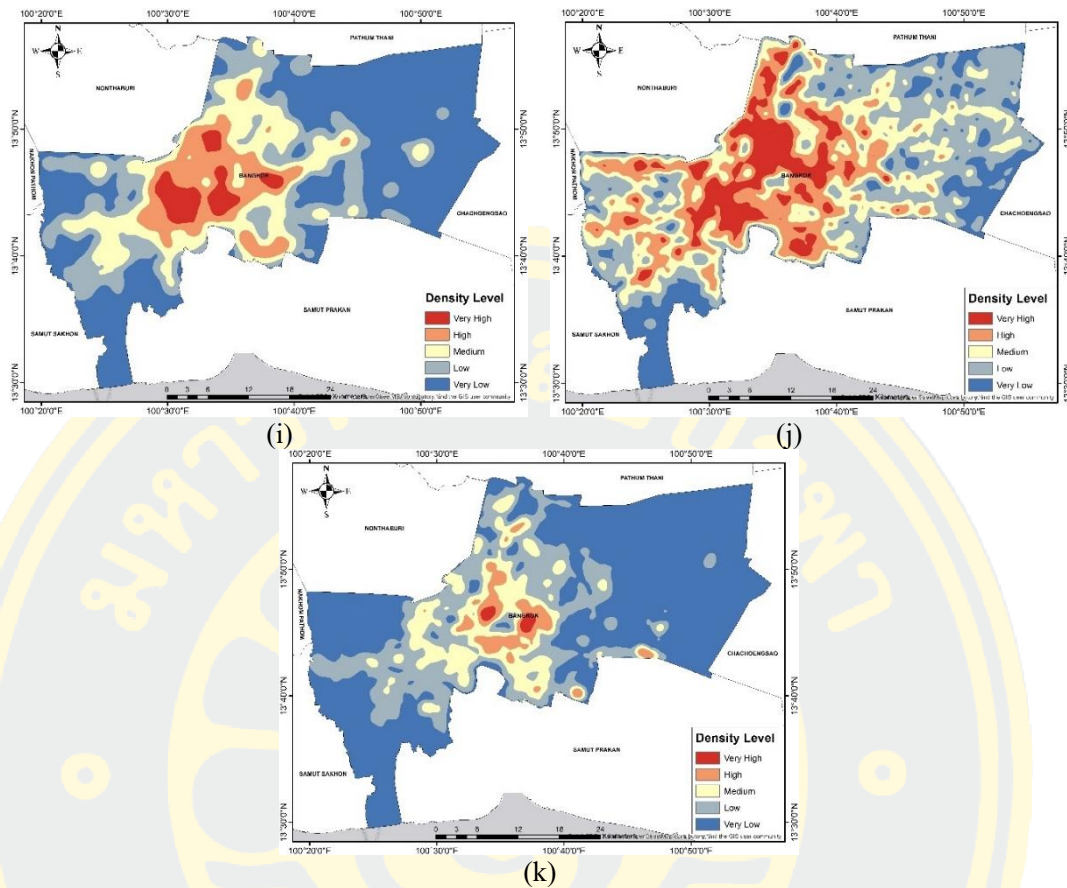


Figure 8 Kernel density of POI (a) retail (b) Catering and recreation (c) Tour and hotel service (d) Healthcare (e) Administration (f) Financial service (g) Education (h) Business and industrial production (i) Transportation (j) Road network (k) Residential in Bangkok.

Demographically, Bangkok has a total area of about 1,568.7 square kilometers and a total population of about 9.06 million people as of 2022 [61], and the population statistics from the Bureau of Registration Administration (BORA) report a household registration population of 5.49 million [62], suggesting that approximately half of the population is domestic migrants or non-local population, including the latent population of about 2.75 million people [63] and others such as tourists and laborers. The LandScan Program, started in 1997 at Oak Ridge National Laboratory (ORNL), was designed to improve population estimates for assessing the impact of natural and manmade disasters. It aimed to create more accurate population distributions, considering where people are throughout the day, not just at their residential locations. The program produced high-resolution data for use in evaluating events on various geographic scales. Initially launched as LandScan Global in 1998, it has been updated annually to provide

more precise and timely population data for disaster management. LandScan Global, which uses geospatial science, remote sensing technology, and machine learning algorithms to represent the geographical distribution of population at a one-kilometer resolution over an average 24-hour period, collected population density from the LandScan program. Provide the most accurate and reliable data, which includes daytime and nighttime population rather than just a residential location [64]. The data obtained by downloading files in TIFF format covers the entire world shown in Figure 9.

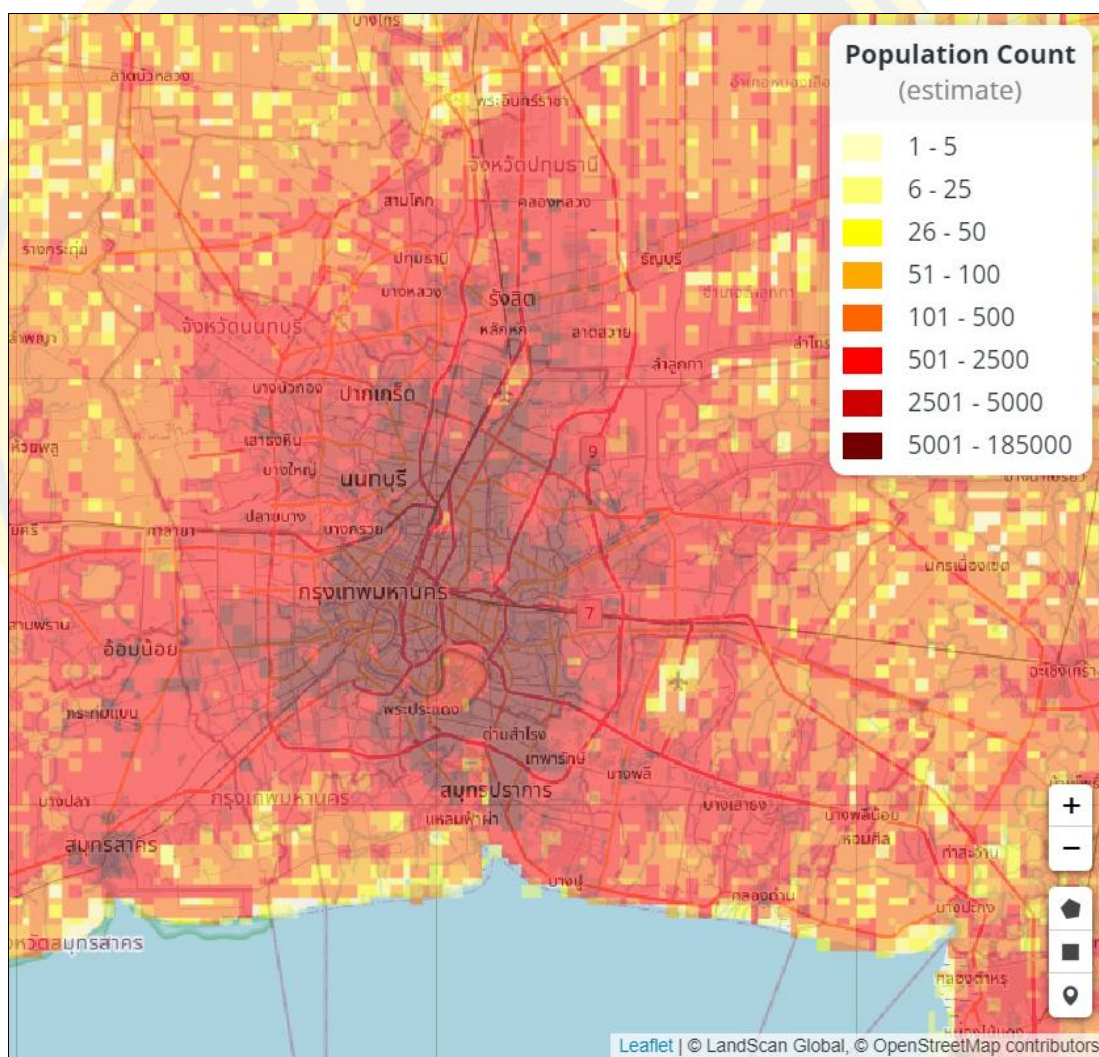


Figure 9 LandScan Global shows population density in Bangkok, Thailand.

Bilinear interpolation is a technique that determines the value of a new cell by computing a weighted average based on the distances to the four nearest input cell centers. This method is particularly beneficial for handling continuous data, as it

inherently smooths the data during the interpolation process. To extract population density in the Bangkok area from the LandScan raster file by using the resample tool in ArcMap, resampling the spatial resolution to 10 meters is shown in Figure 10.

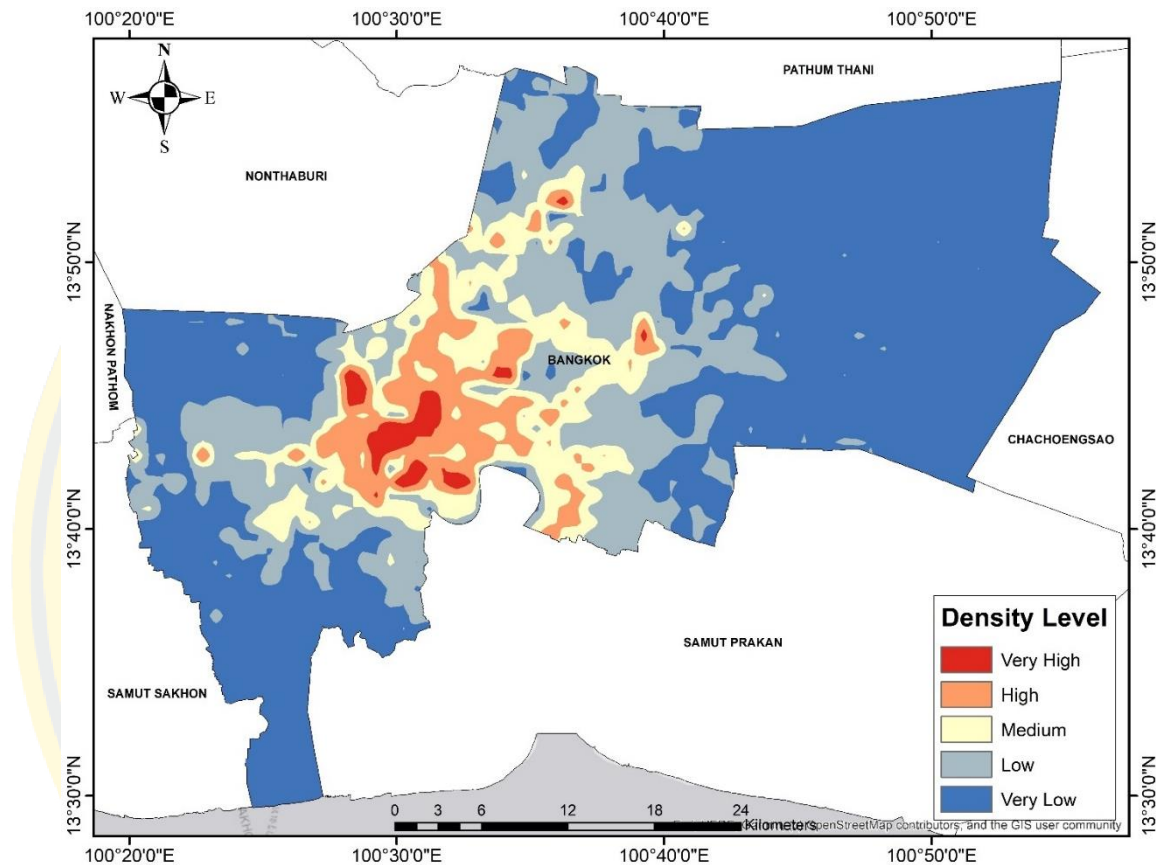


Figure 10 Raster file of popular density in Bangkok.

Green space, Sentinel-2 is a part of the Copernicus program, operated by the European Space Agency (ESA). The mission provides global coverage of the Earth's land surface every 5 days. Sentinel-2 L2A Multispectral Imagery (MSI) data as it is already atmospherically, it's delivering resolution optical images with 10 m, 20 m, and 60 m spatial resolutions, depending on the spectral band [65]. For the classification project, Sentinel-2 stands out among all publicly accessible satellite images because of its higher spatial resolution of 10 m [43]. The image scene selection from Sentinel-2 L2A of Bangkok area acquired on period 1-31 January, 2022 with less than 1% cloud coverage.

Urban greenspace (UGS) is a key component of the city landscape and consists of a variety of ecosystems, usually referred to as any type of plant-covered

environment (public or private), that improves the quality of life in cities. However, one potential negative impact of green space is crime, which is of particular concern within many communities. The Normalised Difference Vegetation Index (NDVI) is commonly used in research exploring the associations between UGS and human behavior and health by assessing all green vegetation through remotely sensed satellite imagery include all aspects of green infrastructure within the built environment [33]. Figure 11 shows the data that has been extracted and downloaded using Google Earth Engine (GEE).

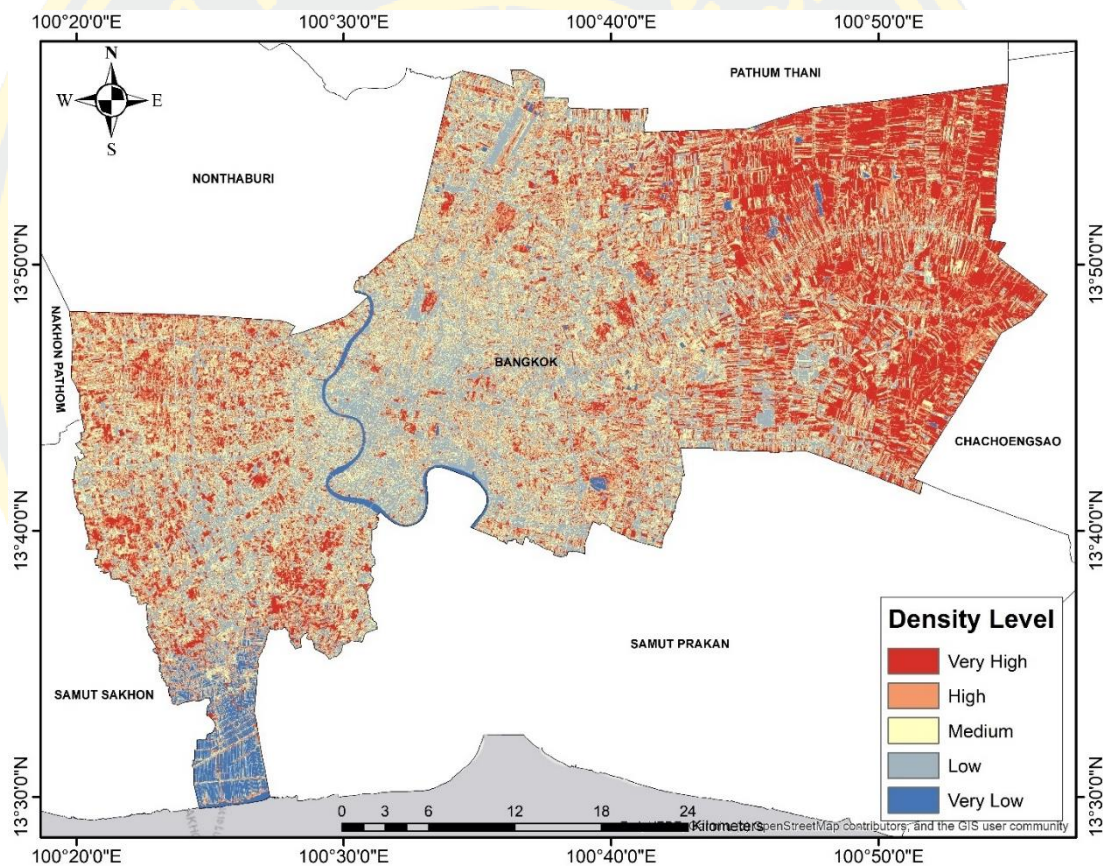


Figure 11 Extract urban greenspace from Sentinel-2 L2A by NDVI.

The Normalised Difference Vegetation Index (NDVI) quantifies vegetation by measuring the difference between near-infrared light (which vegetation reflects strongly) and red light (which vegetation absorbs). The value of NDVI ranges from -1 to +1. The lack of green leaves results in a value of zero, whereas a value near +1 (0.8–0.9) indicates the highest density of green leaves possible. Nevertheless, the land cover category does not have a specific threshold for NDVI values. The equation for

computing the NDVI is as follows:

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)}$$

NIR – reflection in the near-infrared spectrum

RED – reflection in the red range of the spectrum

The Suomi National Polar-Orbiting Partnership's (SNPP) Visible Infrared Imaging Radiometer Suite (VIIRS), which supports a Day-Night Band (DNB) sensor that provides daily global measurements of nocturnal visible and near-infrared (NIR) light, provides the nighttime light data. This enables the creation of better nighttime products. The Black Marble suite's fourth nighttime light (NTL) product is VNP46A4. It offers yearly composites of corrected NTL radiance and includes 28 layers of various data. The product records began on January 1, 2012. The spatial resolution of the nighttime light image data is 500 m × 500 m as shown in Figure 12 [66].

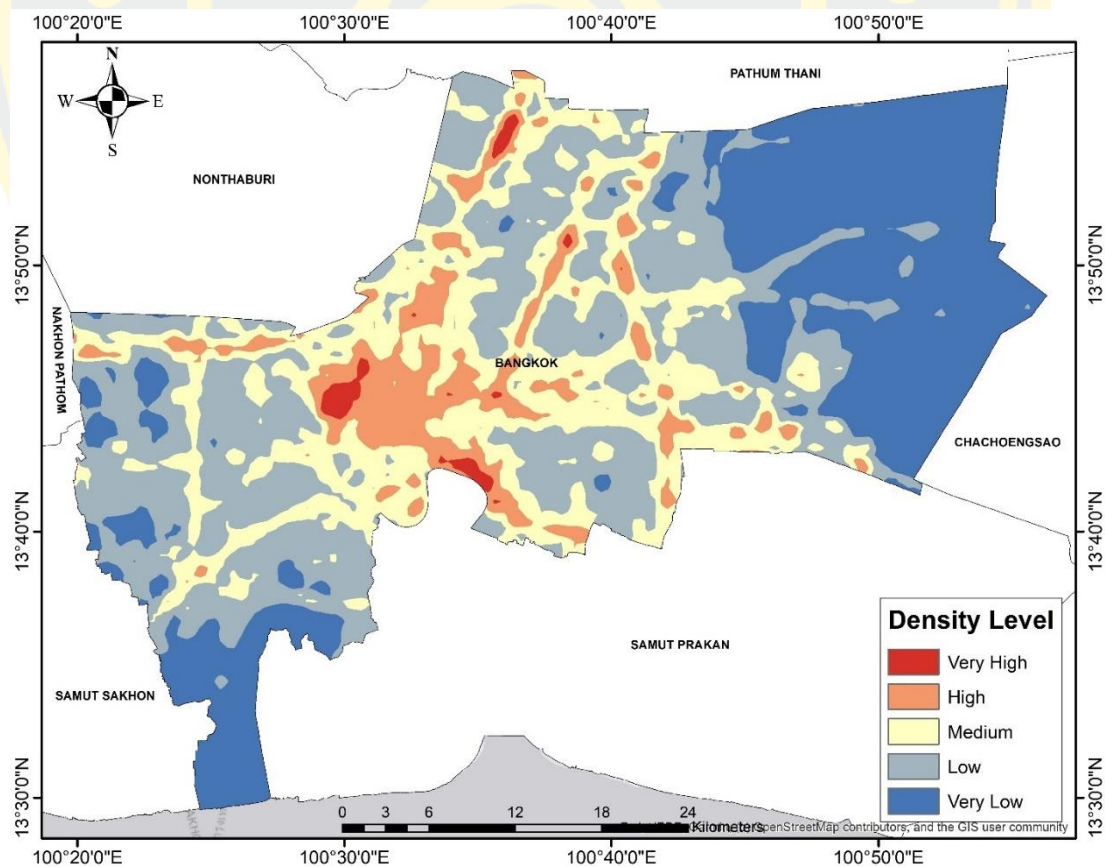


Figure 12 Extract nighttime light from SNPP by product VNP46A4.

A list of parameters shown in Table 4 that are extracted and classified from the urban environment consist of points of interest (POI), demographics, and satellite image data. POI refers to specific points that might be interesting for users, often related to business, tourism, or services. This group includes the parameters P1 through P11, which cover a variety of sectors, from retail to residential. Satellite imagery can collect various types of data through P12 to P14. By analyzing satellite imagery, we can detect population density, which reflects the distribution and intensity of people's full potential activity space during the day and night, not just a residential location. The NDVI (Normalized Difference Vegetation Index) is a measure of vegetation health and density calculated using the red and near-infrared light reflected by plants. Satellite images collect nighttime light data to assess human activity and urbanization. The factors derived from satellite images of different resolutions will be processed with resampling techniques to ensure the same resolution of 10 m before being input into the model. Each parameter type provides a variety of perspectives on the characteristics of a geographical area. When combined, these factors provide a comprehensive understanding of the area's environment and human activities.

Table 4 List of 14 parameters defined as input data.

Code	Parameter name	Type	Resolution	Year	Source
P1	Retail	POI	-	2022	Google map, TripAdvisor, BKK Risk Maps, Internal System
P2	Catering and recreation	POI	-	2022	
P3	Tour and hotel service	POI	-	2022	
P4	Healthcare	POI	-	2022	
P5	Administration	POI	-	2022	
P6	Financial service	POI	-	2022	
P7	Education	POI	-	2022	
P8	Business and industrial production	POI	-	2022	
P9	Transportation	POI	-	2022	
P10	Road network	POI	-	2022	
P11	Residential	POI	-	2022	
P12	Population density	demographics	900m	2022	LandScan

Code	Parameter name	Type	Resolution	Year	Source
P13	NDVI	Satellite image	10m	2022	Sentinel-2 L2A
P14	Nighttime light	Satellite image	500m	2022	SNPP

3.4 Classification process

In the classification process of machine learning, the important steps are preparing data, constructing models, and evaluating results. The process begins with collecting and preparing a dataset. Choose a machine learning method based on the data features and requirements. The dataset is divided into training and testing sets. The training set provides the model to create relationships between the input attributes and the target classes. Model performance can be improved by adjusting hyperparameters in the training process. Model accuracy is evaluated by using testing sets through evaluation metrics consisting of accuracy, precision, recall, and F1-score.

3.4.1 Machine learning method

Classification is a technique in supervised machine learning that involves the process of categorizing data or objects into predefined classes or categories based on their features or attributes. Machine learning methods used in the study included three machines to learn to classify risk areas consisting of Random Forest (RF), Light Gradient Boosting Model (LightGBM), and Extra tree (ET).

Random forests (RF) are an effective machine learning algorithm. develop by creating multiple decision trees and combining them into an ensemble. The algorithm is based on the idea that a combination of bootstrap aggregated classifiers performs better than a single classifier [67]. In this case, bootstrapping refers to the process of parameterizing each individual tree by using a randomly selected subset of observations from the training dataset [43].

The Light Gradient Boosting Model (LightGBM) is a gradient boosting framework using decision tree algorithms, developed by Microsoft. LightGBM is fast and efficient training, especially with large datasets, and is mainly used for ranking and classification tasks [68]. The framework introduces two key innovations:

gradient-based one-side sampling (GOSS) and exclusive feature bundling (EFB), which speed up the training process of conventional GBDT by up to over 20 times while achieving almost the same accuracy. which enhance efficiency by reducing the data and features used in computations without significantly compromising accuracy [69].

Extra Tree is a supervised learning method, also known as Extremely Randomized Trees, is an ensemble method that combines multiple decision trees to create a more robust model like Random Forest. This method can be considered an extension of the commonly used random forest algorithm. Its design aims to reduce overfitting, like the random forest [70]. The main difference between the two algorithms is the selection feature for each decision tree. The ET includes extra randomness while creating trees from the entire dataset and trains each base estimator using a random subset of features to split the node. In contrast to a random forest, which selects a random subset of features for each tree and chooses the best feature for each node split, it does not use a random selection of features and their related values for node splitting.

Machine learning model training process begins with initial libraries by importing necessary Python libraries that provide functions for data manipulation, machine learning algorithms, and data scaling.

GDAL, a geographic data abstraction library, implements the data ingestion step by reading data features or independent variables from input data in raster image format as data X and training data as data y, which comprises the labels or dependent variables we aim to predict.

Data normalization is a process that standardizes the feature data to a uniform scale while keeping distinctions in the value ranges. This is an important process because it ensures that each feature still influences the model, avoiding any single feature from dominating due to its scale.

Imbalanced handling utilizes the under-sampler technique used to get balance in the dataset. It reduces the proportion of samples set in the major class with the minority class, resulting in a balanced distribution in the dataset that is important to the optimal performance of the classification model.

Data preparation and splitting is the process of splitting the dataset after stacking input data (data X) and training data (data y), into split training and testing

sets. The `train_test_split` function can be processed by partitioning the dataset, allocating 80% of the data for training and 20% for testing. The aim of this process is to evaluate the model's effectiveness with previously unexplored data.

Cross-validation is used to split the data into k subsets. and 10-fold cross-validation were used in the model. During the cross-validation iteration, nine subsets are assigned for training and one subset for testing. This process is repeated 10 times, with each subset being utilized as the test set once. It improves evaluating the accuracy of the model to generalize to a separate dataset.

Hyperparameter optimization utilizes `RandomizedSearchCV` and `GridSearchCV` algorithm that makes a random search over hyperparameters to find the best parameters that improve the performance of the model.

Model training is the process of adjusting a machine learning algorithm's weights using the provided data and set hyperparameters. The model learns patterns from the training data, allowing it to create predictions for new data.

3.4.2 Accuracy assessment

Accuracy assessment involves evaluating the accuracy of different machine learning models to determine which areas across Bangkok are safe and suitable for locals and tourists. Scikit-learn [71] is an open-source machine learning library that supports supervised and unsupervised learning. Additionally, it offers a variety of tools for model fitting, data preprocessing, model selection, model evaluation, and providing many other utilities. classification accuracy using the scikit-learn metrics API with functions Precision, Recall, F1Score, ROC-AUC, and confusion matrix for comparing different models to choose the one that performs the best with high accuracy.

The F1 score is the harmonic mean of the balance between precision and recall, used to measure model accuracy on a dataset. Precision comes from the number of true positive results divided by the number of all positive results, including those not identified correctly. Recall comes from the number of true positive results divided by the number of samples that should have been identified as positive [72, 73], which are calculated by the following equations:

$$\text{Precision} = \frac{\text{True}}{(\text{True} - \text{Positive} + \text{False} - \text{Positive})}$$

$$\text{Recall} = \frac{\text{True} - \text{Positive}}{(\text{True} - \text{Positive} + \text{False} - \text{Positive})}$$

$$\text{F1 - score} = 2 * \frac{(\text{Recall} * \text{Precision})}{(\text{Recall} + \text{Precision})}$$

The ROC (Receiver Operating Characteristic) curve is a graph representation of evaluation metrics of a classification model performance with includes all the possible decision thresholds. This curve plots the true positive rate (TPR) and the false positive rate (FPR). AUC (Area Under the ROC Curve) measures the entire area underneath the ROC curve. AUC ranges in value from 0 to 1, where a value of 0 indicates a perfectly inaccurate test and a value of 1 reflects a perfectly accurate test [74].

A confusion matrix is a tabular representation used to evaluate the accuracy of a classification model. The matrix is two-dimensional, with one dimension representing the true class of an object and the other dimension representing the class that the classifier assigned [75]. It provides the total number of predictions made by the classifier that were true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), as indicated in Table 5.

Table 5 Confusion matrix.

		Assigned class	
		Positive	Negative
Actual class	Positive	TP	FN
	Negative	FP	TN

The following four are the number of measures of classification outcomes.

$$\text{Specificity} = \text{True negative rate} = \frac{TN}{(TN + FP)}$$

$$\text{Sensitivity} = \text{True positive rate} = \text{Recall} = \frac{TP}{(TP + FN)}$$

$$\text{Positive predictive value} = \text{Precision} = \frac{TP}{(TP + FP)}$$

$$\text{Negative predictive value} = \frac{TN}{(TN + FN)}$$

3.4.3 Analyze the crime data and create a suitable area map for locals and tourists.

This study also analyzes crime situations with crime statistics for supporting the classification result from the best model to emphasize the result from the model in line with the official report before utilizing the result to create a comprehensive map. Once the classified map has been created, an accuracy assessment is performed to evaluate the best model. The result from the best model will be processed in ArcMap software to create a suitable area map for locals and tourists by overlaying metro lines and providing this data to policymakers and relevant agencies.

Metro stations in Bangkok commonly length about 150 meters; to cover the station area based on the station length, this study chose a radius of 100 meters to identify each station's area. Zonal statistics as Table is a function in ArcGIS that was utilized to calculate the metro stations in crime risk area by the value based on crime risk level from the classification process throughout the TIFF file from the best model. Each station will identify a high, medium, and low risk level depending on the value that is found in the area with a relative priority of high, medium, and low. Finally, create a risk level sign as a circle covering a metro station, where the red circle represents high-risk, the blue circle represents medium-risk, and no circle represents low-risk on the metro map.

3.5 Tools and Instrumentation

Several tools and instruments are applied for data collection, processing, and analysis. Google Earth Engine (GEE), ArcMap, and Quantum GIS (QGIS) are the primary tools utilized in this research as shown in Table 6.

Table 6 Tools and Instrument.

Tools and Instrumentation	Usability
Microsoft Visual Studio Code (VS Code)	Scrab POI data from TripAdvisor and Processing machine learning model and accuracy assessment
Google Earth Engine (GEE)	Extract Greenspace by NDVI from Sentinel-2 L2A
ArcMap	Preprocessing
QGIS	Preprocessing

CHAPTER 4

RESULTS AND VALIDATION

The results of this thesis include a comprehensive classification of risk areas utilizing urban environment data and satellite imagery through three machine learning models: Extra Trees (ET), Random Forest (RF), and Light Gradient Boosting Machine (LightGBM). These models will be evaluated based on standard metrics such as F1 score, ROC-AUC, and confusion matrix to choose the one that performs the best with high accuracy.

The integration of urban environment with satellite imagery is expected to help understand urban risk dynamics by focusing on the areas of potential concern based on human activity, socioeconomics, infrastructure, and geographical features. The study aims to reveal the key features and indicators of risk from the combined dataset, offering insights into the spatial distribution of risk areas.

Finally, the study's findings are expected to provide valuable information for policymakers and relevant agencies, suggesting ways to utilize the classified risk areas for urban planning and public safety enhancements.

4.1 Analysis of crime risk in Bangkok (Data training)

The analysis of crime risk in Bangkok relates to multi-factors that cover the perception and actual incidence of crime, which includes crime awareness among residents, the effectiveness of crime prevention measures, safety perceptions of foreign tourists, and environmental factors. The Metropolitan Police Bureau (MPB) in Bangkok faces an increase in crime rates from the economic difficulties after the COVID-19 pandemic [50], especially robbery and fraud. While having a substantial police force, officers are overwhelmed with various duties, including crime suppression, traffic management, and riot control. The tourism sector is an important sector in Bangkok and is very sensitive to crime [76]. Crimes at tourist destinations not only have an immediate negative impact on tourism demand but also have prolonged effects by the influence of word-of-mouth recommendations that influence tourists to consider choosing their destination choices by life-threatening crimes, property crimes, and violent crimes. Urban crime in Bangkok is also closely linked to

environmental factors [77]. In selected slum communities, environmental improvements have been shown to reduce and prevent urban crimes. Enhancements in community consciousness, ownership, responsibility, cohesion, and self-defense mechanisms contribute to a safer environment, thereby reducing crime rates. Furthermore, the fear of crime among Bangkok residents varies by gender [78], with females exhibiting higher vulnerability and risk perception. Factors such as perceived social disorganization, experience of victimization, and satisfaction with police measures indirectly influence the fear of crime. These insights are crucial for developing community crime prevention measures and reducing the overall fear of crime.

4.1.1 Crime statistics

Crime statistics in Bangkok, especially crime against persons and property crime, reflect a problem in a city that has an impact on society and the urban environment. Crime is a big issue in Bangkok that affects daily life [79], including tourism, government administration, and community well-being. The increasing crime rates need a comprehensive understanding of crime patterns for effective planning and prevention strategies. Crimes against persons in Bangkok 2022 that are collected from the Royal Thai Police and provided by the national statistical office (NSO) that show in Figure 13 revealed assault as the most common incident of violence, followed by other violent crimes such as rape, attempted murder, and murder. Assault is related to a range of actions, from threats or attempts to cause physical harm that leave the victim permanently disabled. Assault was recorded in 1,759 cases that emphasizes the seriousness and frequency of assaults in Bangkok and that physical violence is a primary issue in personal safety concerns. Rape ranks second, with 151 cases. Although it is less than assault, rape remains an important issue in society and indicates ongoing issues related to gender-based violence and the safety of vulnerable groups. This was followed by 126 cases of attempted murder, thus representing an alarming incidence of violent crime where a person must undertake a direct action towards the act of murdering and possess a specific intent of killing the life of that person. The statistics show that although murder is not highly frequent, the threat of such extreme violence remains important. Murder was recorded in 83 cases. Although lower than other violent crimes, the occurrence of murderous behavior emphasizes the

need for continued efforts in law enforcement and preventive measures. Negligent homicide was recorded in 38 cases, reflecting deaths caused by negligence rather than intentional violence. Although it is a small portion of total crimes, it indicates issues in public safety, possibly relating to accidents or irresponsible behavior leading to fatalities. The other category was recorded in 206 cases, consisting of infrequent or uncategorized crimes against persons. This emphasizes the various threats to personal safety in the metropolis.

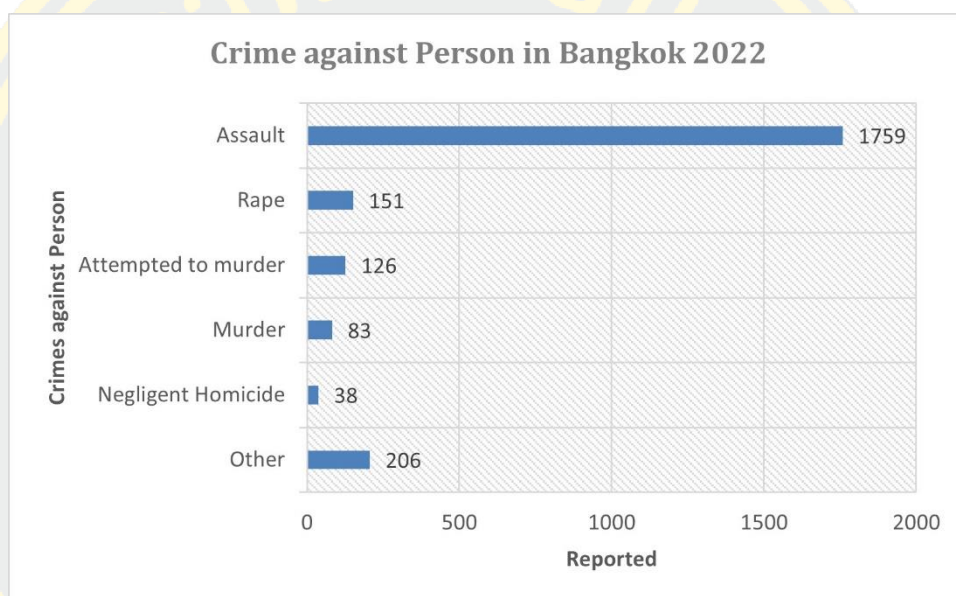


Figure 13 The statistics of crime against person in the Bangkok 2022.

The distribution of property crimes in Bangkok in 2022 shown in Figure 14 revealed a concentration of offenses in theft and fraudulent activities, which together are the major proportion of reported incidents. Theft is the most frequently recorded property crime in Bangkok, with 3,365 cases. The difference between theft and other crimes emphasizes an important issue that necessitates enhanced prevention measures prior in Bangkok. Fraudulent activities are ranking second with 2,050 cases. This high revealed underscore the need to strong focus on cybercrime and financial fraud prevention strategies. This was followed by 946 cases of misappropriation, which reflect crimes where property is used unlawfully but without the explicit intent to steal.

Although statistics are less than theft and fraud, misappropriation continues to provide a significant issue. There were 557 cases of mischief, which can include

acts of vandalism or deliberate damage to property. This may suggest a need for better urban management and law enforcement to prevent such behavior. Receiving stolen property was recorded in 153 cases. The focus on an underlying network for illicit trade is the production or distribution of a good or service that is considered illegal. This calls for disrupting the flow of stolen property. Motorcycle theft was recorded in 143 cases, which reflects a specific trend in motor vehicle theft that is impacting everyday commuters. Snatching was recorded in 123 cases, thus representing the common practice of forcefully taking possessions from victims in public areas. Robbery was recorded in 102 cases, which commonly related to violence or the threat of violence.

Although the statistics are lower, the nature of this crime makes it a sensitive concern for public safety. And some of the property crimes that represent less frequent but serious offenses are 29 cases of extortion, 21 cases of car theft, 21 cases of arson, 16 cases of gang robbery, and 1 case of kidnapping. Their lower incidence does not lessen their social impact, especially kidnapping, which, though rare, can result in serious consequences. The other category consists of 520 cases defined as uncategorized in property crimes. This suggests a diverse range of criminal activities that fall outside the main categories. The difference between high-frequency crimes (theft, fraud) and lower-frequency ones (e.g., kidnapping, gang robbery) suggests targeted, specific measures might be most effective when focused on these prevalent issues. Furthermore, misappropriation and mischief, while lower in statistics, still influence the overall crime, indicating the necessity for crime prevention strategies in the short and long terms.

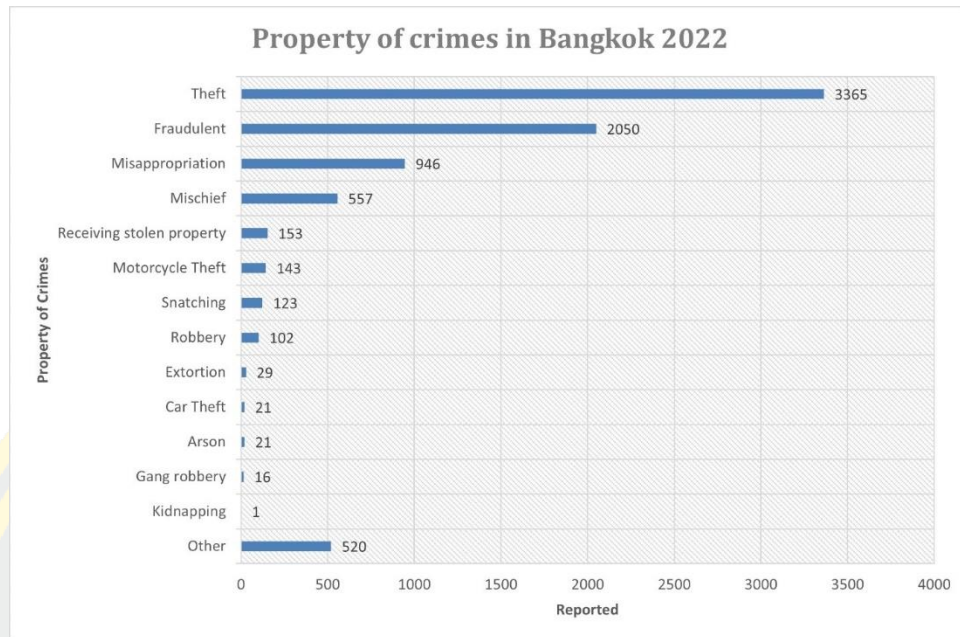


Figure 14 The statistics property of crimes in Bangkok 2022.

Figure 15 represents the map of crime statistics for various districts of Bangkok in 2022. Every district is colored to show different crime rates, categorized into five intervals: 18-75 (dark blue), 76-106 (blue), 107-156 (light blue), 157-216 (dark orange), and 217-319 (dark red). Every district is marked with a district code that is related to its name, while the color shading indicates the density of crimes. Dark shades of red show higher crime rates, while dark blues show lower crime rates. This spatial distribution reflects crime patterns throughout Bangkok, with areas such as district 1015-Phra Khanong, 1016-Khlong Toei, 1018-Suan Luang, 1019-Bang Khen, 1023-Lat Phrao, 1039-Watthana, and 1042-Sai Mai showing the highest crime rates (217-319 case), while others, such as 1025-Nong Chok, have lower crime rates (18-75 case). These statistics emphasize the need to improve urban safety and reduce crime rates in Bangkok.

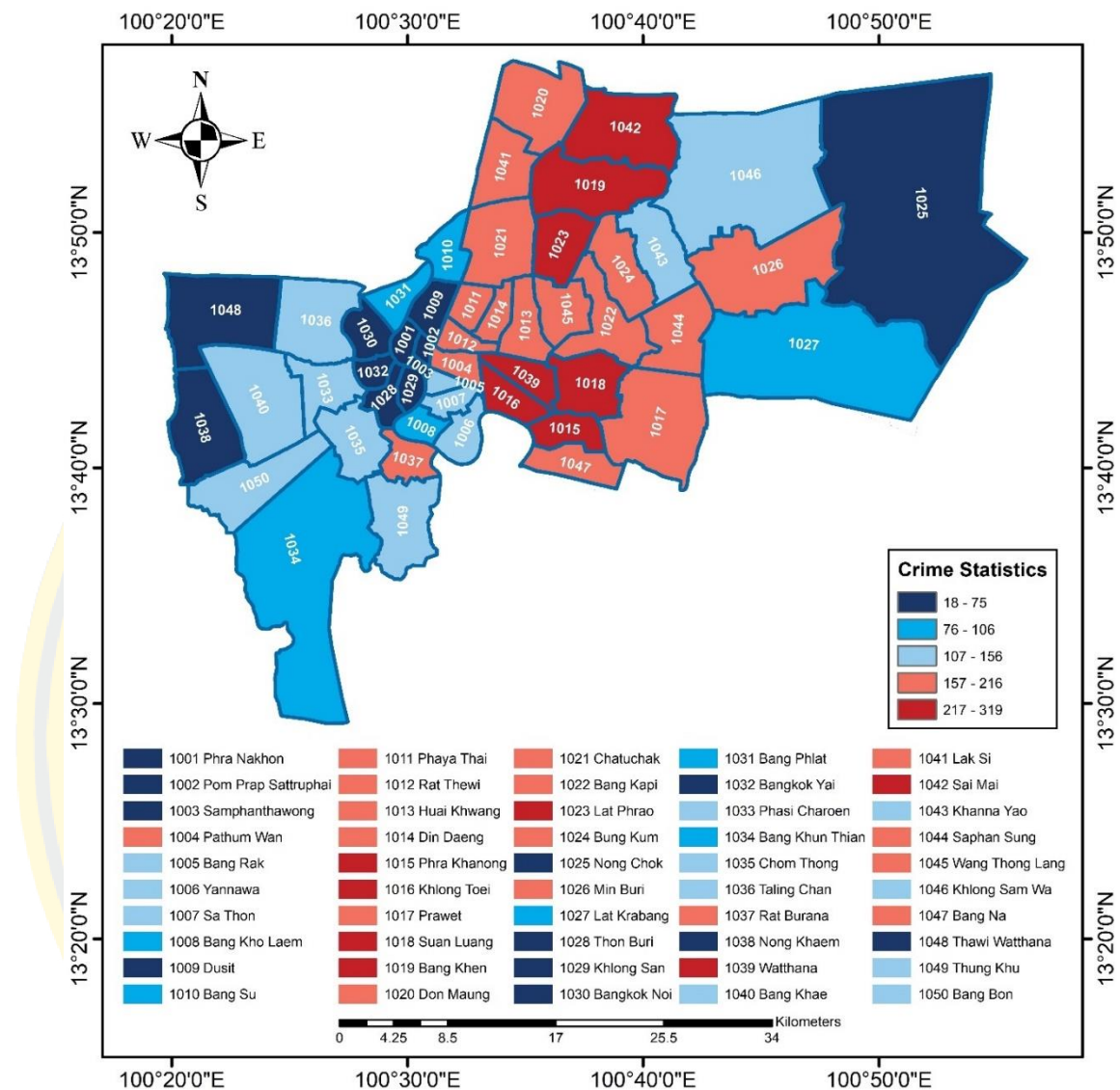


Figure 15 The map displays the crime statistics of Bangkok in 2022.

Based on figure 16, the map displays the tourist criminal cases in eight districts of Bangkok in 2022; every district is colored to show different crime rates, categorized into five intervals: Dark blue areas (0 cases) cover the majority of Bangkok, suggesting these areas have no recorded tourist criminal cases. Sky blue areas (1–2 cases) consist of districts such as Samphanthawong (1003), Yan Nawa (1006), and Bang Kho Laem (1008), which indicates low crime rates. Samphanthawong is home to Chinatown and attracts a more niche group of visitors, especially family-friendly tourism activities that are interested in culture and cuisine rather than nightlife, which might reduce opportunities for crimes like theft or

harassment. Yan Nawa and Bang Kho Laem are more residential and have less nightlife or major tourist attractions that make this area less that could attract opportunistic criminals against tourists. Light blue areas (3–6 cases) consist of districts such as Bang Rak (1005), which indicates moderate crime rates. These zones are popular tourist spots such as the Patpong Night Market and various luxury hotels along the Chao Phraya River that attract a high volume of tourists.

As a commercial center with many businesses and financial institutions, it leads to high daily foot traffic and transportation via BTS Skytrain stations, which attract a considerable influx of people and may serve as hotspots for crime. Orange areas (7–17 cases) consist of districts such as Sa Thon (1007) and Khlong Toei (1016), which indicates a higher crime risk. Khlong Toei is one of Bangkok's largest slum areas. Sa Thon is a business district with many hotels, embassies, and tourist attractions. Red areas (18–22 cases) consist of districts of Pathum Wan (1004) and Watthana (1039), which indicates the highest risk zones; both districts are popular tourist destinations, attracting large numbers of visitors. Pathum Wan is the major shopping centers like Siam Paragon and MBK Center and usually refers to the Lifestyle and College Neighborhood Zone, while Watthana consists of areas such as Sukhumvit Road, known for its nightlife and entertainment. Tourists might have increased vulnerability after having drunk or while traveling in an unfamiliar area, especially during late-night hours. The high concentration of tourists makes these areas attractive targets for opportunistic crimes, including pickpocketing, bag snatching, and scams. Common scams include overcharging for services, fake goods, and various forms of street hustling. The crowded urban environment with streets and transportation facilitates criminal activity and complicates law enforcement's ability to effectively monitor all areas. Economic inequality may also lead to increased crime rates in these areas, as people might turn to crime for their livelihood.

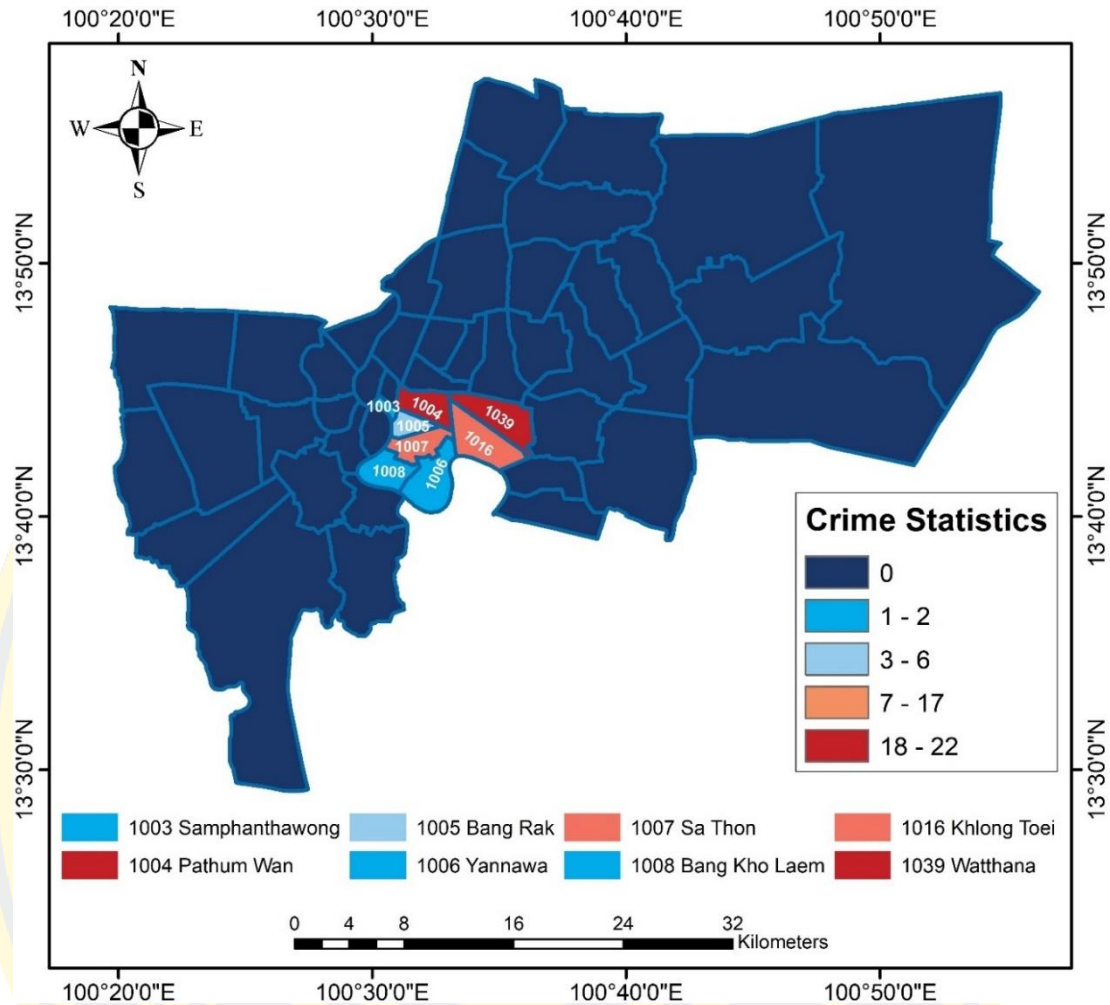


Figure 16 The map displays statistics of tourist criminal cases in Bangkok in 2022.

4.1.2 Crime risk points

The machine learning model accurately classifies crimes and categorizes districts based on crime rates and utilizing spatial data, leading to better predictive accuracy compared to traditional methods [41, 80, 81]. The distribution of risky areas in Bangkok is represented by the risk spot, which is labeled by risk level. The red graph represents 336 points of Police Risk Spot (High) are considered high-risk by law enforcement in Bangkok. However, the BMA Risk Spot (medium) has 347 points represented in yellow, and the Civil Complaint (low) has 6,507 points represented in blue, both are moderate and low risk areas, respectively shown in Figure 17. Defined by government agencies and civilian reports. The differences in risk spot statistics from various data sources may be inferred from different risk assessment criteria and a relationship between resident's risk perceptions and the assessments of the

Metropolitan Police Bureau and Bangkok Metropolitan Administration. In conclusion, no location in Bangkok is completely safe from crime.

Based on a list of crime-risk areas in the city in 2022 [50] by the City Law Enforcement Department of Bangkok Metropolitan Administration, every district contains vulnerable areas for crimes and the figure shows the important role of differences in data sources in the assessment of urban risks and the importance of cooperation between residents and the government to enhance public safety. While the statistics reveal imbalanced data, imbalance can lead to a bias for the majority class, resulting in underestimation of critical high-risk areas. Predictive models trained on such data would tend to have high accuracy for low-risk predictions but less performance in identifying medium- and high-risk areas, potentially leading to false negatives and omitting serious threats. To solve this imbalance, under-sampling can serve as an effective method, which the majority class (low-risk) reduces to get a more balanced with the minority classes (medium- and high-risk). This method reduces the predominance of low-risk data, helping the model to more efficiently learn and generalize patterns related to higher-risk crime areas. An appropriate method of under sampling while maintaining the integrity of the minority class patterns is important for improving model performance in predicting high- and medium-risk areas.

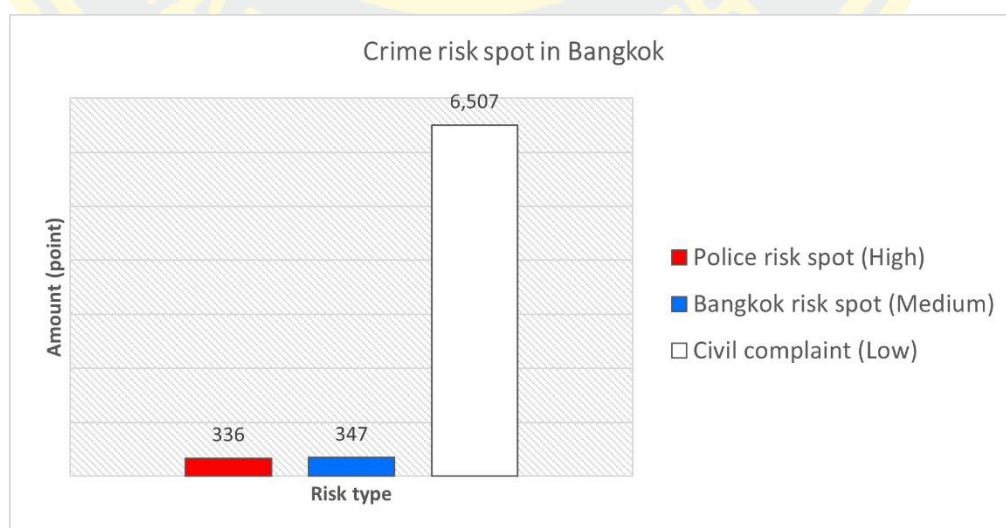


Figure 17 Crime risk spot count within each data source in the Bangkok area.

4.2 Analysis of Urban environment in Bangkok (Data input)

The urban environment of Bangkok is influenced by a mix of traditional and contemporary [82]. Urban rapid growth has been specific by extension and densification, with built-up areas having expanded rapidly since the 1980s. This growth has led to various urban problems, especially safety issues. Crime is one of the important issues that has affected the perception of society, like the fear of crime in Bangkok being higher among females [78], with many factors such as gender, social disorganization, and victimization experiences influencing this fear indirectly through perceived victimization risk, police crime prevention strategies satisfaction, and participation in crime prevention campaigns. High-risk areas for a young person who has committed a criminal offense in Bangkok [83] are slums, poorly lit alleys, entertainment venues, and 24-hour convenience stores, which are related to crimes such as drug use, property crimes, assault, and motorcycle drag racing.

The density of points of interest (POIs) in an urban area can influence crime risk in many ways. The density of POIs in a particular area can affect crime rates. Analysis of POI data can help create models that predict crime risk in urban areas. Figure 18 shows the quantity graph of POIs in Bangkok used in this study, with the retail category having the highest number of POIs at 62,406 points, while transportation has 4,939, the smallest number.

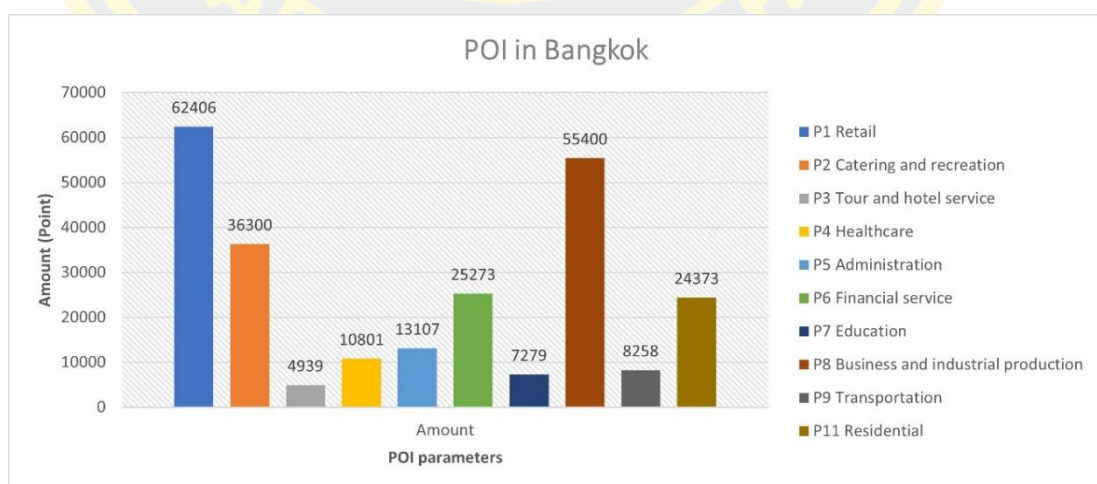


Figure 18 POI counts for each type within the Bangkok area.

4.2.1 P1-Retail (62,406 points)

Bangkok is famous for its variety and bustling retail environments, including a combination of luxury malls, lively markets, and unique shopping areas. Each of these areas provides a unique shopping experience, representing Bangkok's bustling retail environment. The relationship between retail areas and crime risk in urban environments is complex. There is much research on the impact of retail environments on crime rates and perceptions of the safety of residents. Increased retail land use is related to a higher perceived crime risk among residents. Figure 19 shows the distribution of crime risk points in retail areas. Retail areas usually have high foot traffic and valuable goods that attract large crowds of customers, in particular tourists. Which can attract criminal activities such as theft, burglary, and violent crimes, especially in urban centers [84]. Violent crime affects the retail sector [85, 86], due to high crime rates leading to a decrease in the number of retail merchants and job opportunities, increasing business costs, especially security costs, and reducing potential consumers. This is a reason for the influx of strangers that retail destinations attract, which locals may perceive as threatening safety [87, 88].

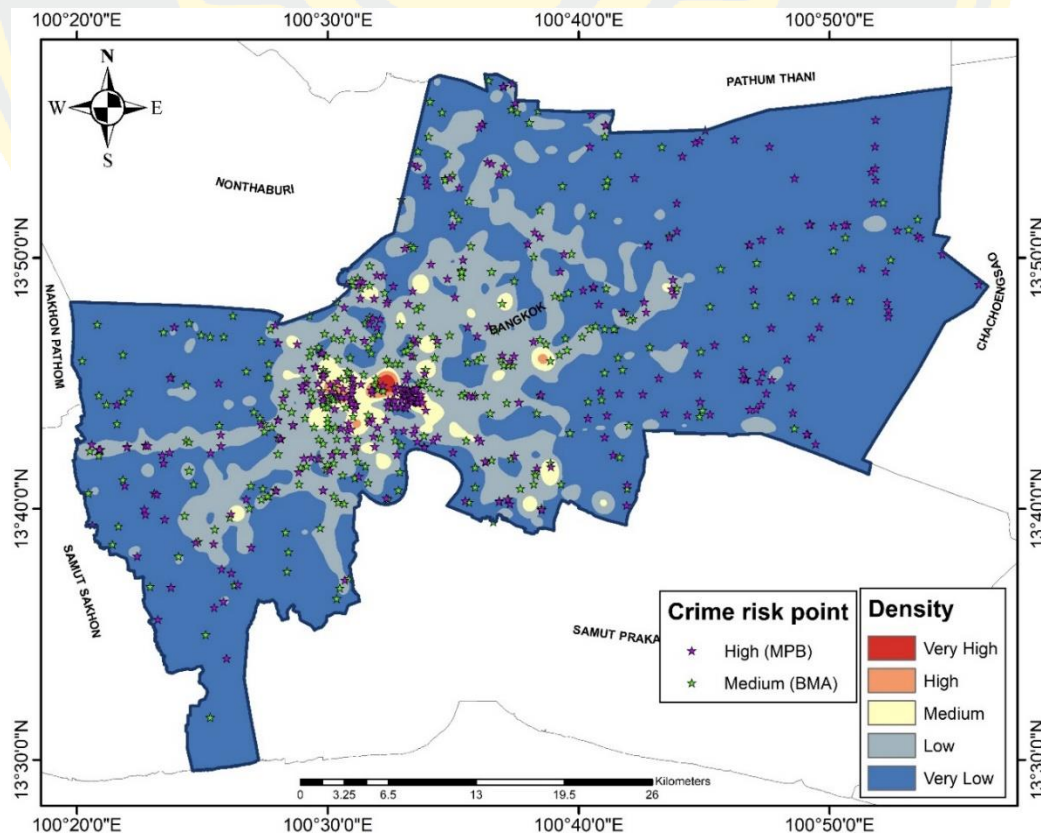


Figure 19 The density of retail with high-medium crime risk points.

4.2.2 P2-Catering and Recreation (36,300 points)

Criminal activity and fear of crime are frequent in recreational areas, such as urban parks, restaurants, cafes, and nightlife venues. Areas with increased visitor inflows for recreation significantly raise the number of crimes committed by both visitors and residents. Figure 20 shows the distribution of crime risk points in recreational areas, recreational visits contribute to high crime rates, followed by shopping visits. Frequently, crime prevention strategies include trash and graffiti removal, improved signage, and enhanced lighting. Infrequently used strategies include the assignment of uniformed officers, CCTV call boxes, and park watch groups [89, 90].

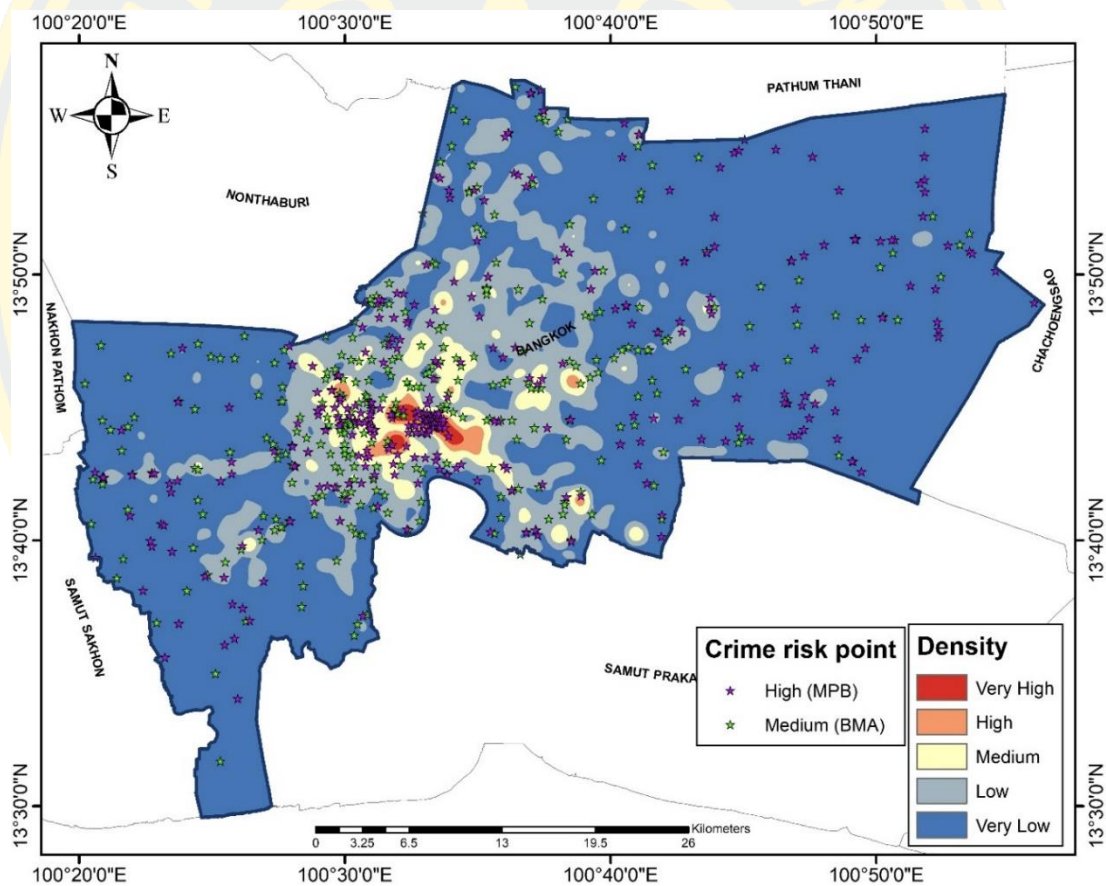


Figure 20 The density of catering and recreation with high-medium crime risk points.

4.2.3 P3-Tour and Hotel Service (4,939 points)

Hotels and tour services are related to the tourism industry, and tourists are mostly customers. Figure 21 shows the distribution of crime risk points in hotels and tour services areas. The characteristics of the hotel environment, such as the level of

security, location, and high-value targets, could influence crimes and make hotels attractive to criminal activities [91-93]. International and local tourists usually face victimization related to property crimes, including theft, burglary, and vehicle break-ins. While personal crimes are usually reported by locals more than internationals [94]. Crime is an important factor that influences tourists to make decisions about where to stay and visit, requiring cooperation between tourism stakeholders and high-technology tools for crime detection and prevention.

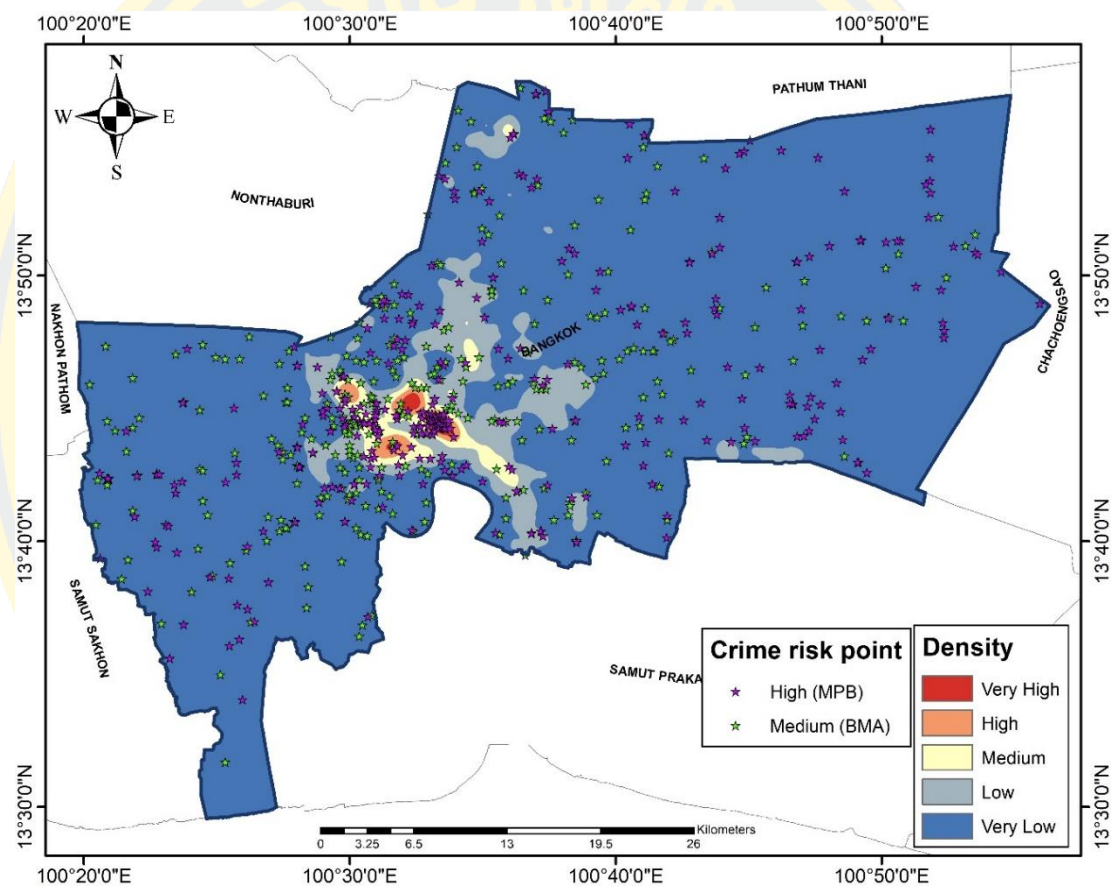


Figure 21 The density of tour and hotel service with high-medium crime risk points.

4.2.4 P4-Healthcare (10,801 points)

The healthcare environment in Bangkok presents a wide range of medical services with international standards, which make it an important healthcare center in the region. Healthcare facilities are open and accessible, which leads to vulnerability to violence and homicides [95]. Figure 22 shows the distribution of crime risk points in healthcare areas. Healthcare workers are at risk of victimization in the workplace, which includes both physical and psychological violence, which is growing in

Thailand [96]. Most violent incidents are perpetrated by patients and their relatives. Mental health services could improve community safety by increasing the number of mental healthcare offices within a county, which could reduce crime rates [97]. Healthcare facilities are not typically a high-risk crime area for tourists. However, the medical service industry in Thailand has the potential to provide high-quality hospitality and medical services [98] that can attract global medical tourists that may be more vulnerable than general tourists and can make them targets for crimes. Healthcare facilities could improve safety and security [99] by implementing Crime Prevention through Environmental Design (CPTED) through safety features such as access control, CCTV, ultrasonic occupancy sensors, and duress alarms.

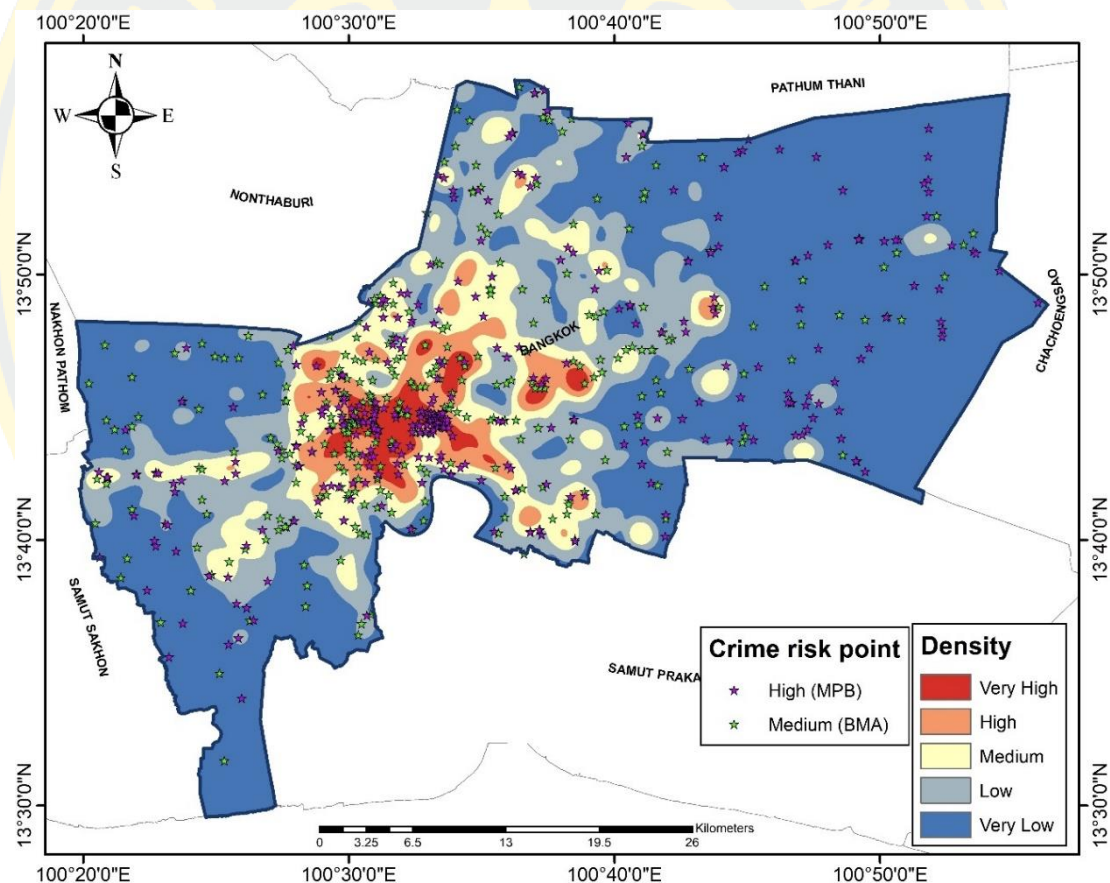


Figure 22 The density of healthcare with high-medium crime risk points.

4.2.5 P5-Administration (13,107 points)

Administrative areas often refer to governmental districts that are more focused on local governance. Government districts can increase crime risks due to their high population density, which can attract theft. Public events increase

opportunities for pickpocketing and violence. Economic inequality in surrounding areas may lead to petty crimes. Figure 23 shows the distribution of crime risk points in administration areas. The likelihood of criminal activity in these areas is often lowered, especially during the daytime. The spatial distribution of police stations could influence the pattern of crime incidence in cities [100], with crime incidence increasing with increasing distance from police stations and decreasing within a two-kilometer radius. The integration between CCTV and proactive police monitoring can reduce property crime in car parks and public transport areas [101].

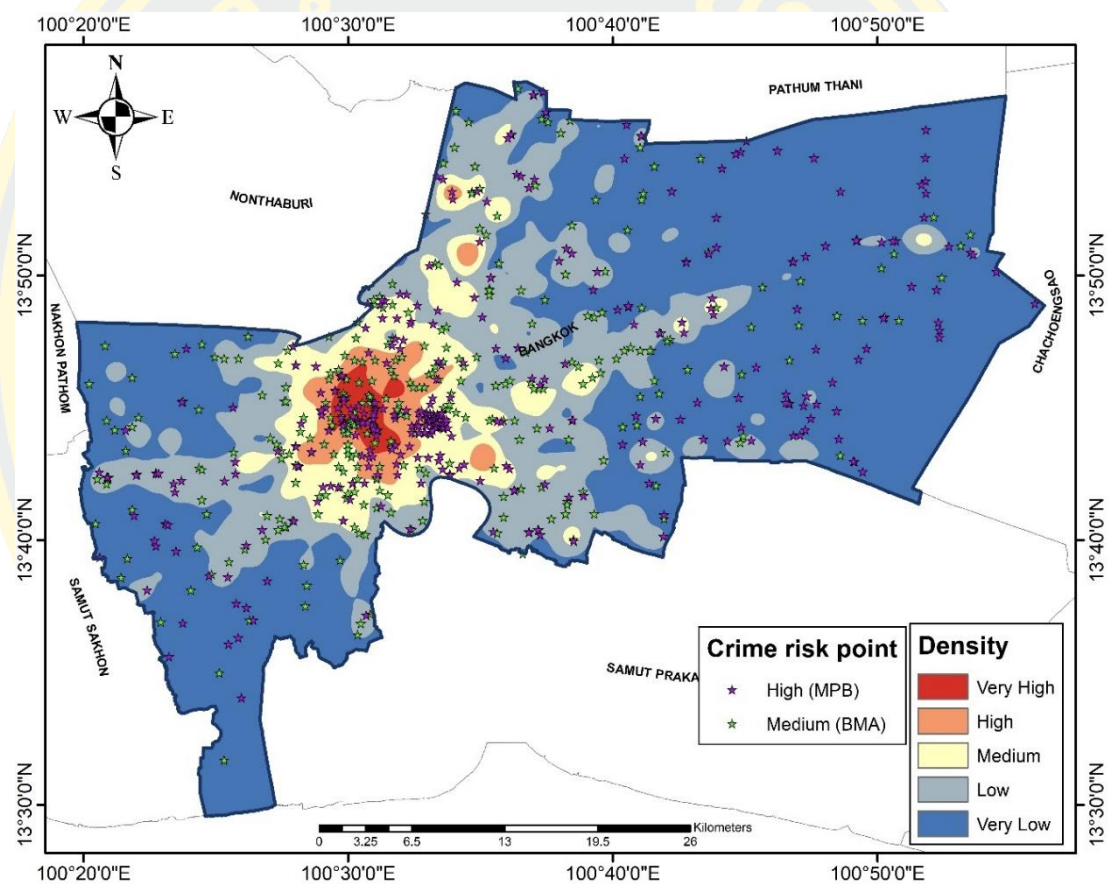


Figure 23 The density of administration with high-medium crime risk points.

4.2.6 P6- Financial Services (25,273 points)

Figure 24 shows the distribution of crime risk points in financial services areas. Financial institutions are often attracted to criminal activities [102] such as bank robberies, cybercrimes, fraud, ATM theft, or financial scams. Tourists carrying out cash exchange or utilizing ATMs in these areas may be vulnerable to targeting. Location of ATM can impact perceived safety with many factors like visibility,

lighting, access control, target hardening, surrounding land use, and zoning [103]. The quality of waiting areas for ATM installations also influenced users' fear of crime.

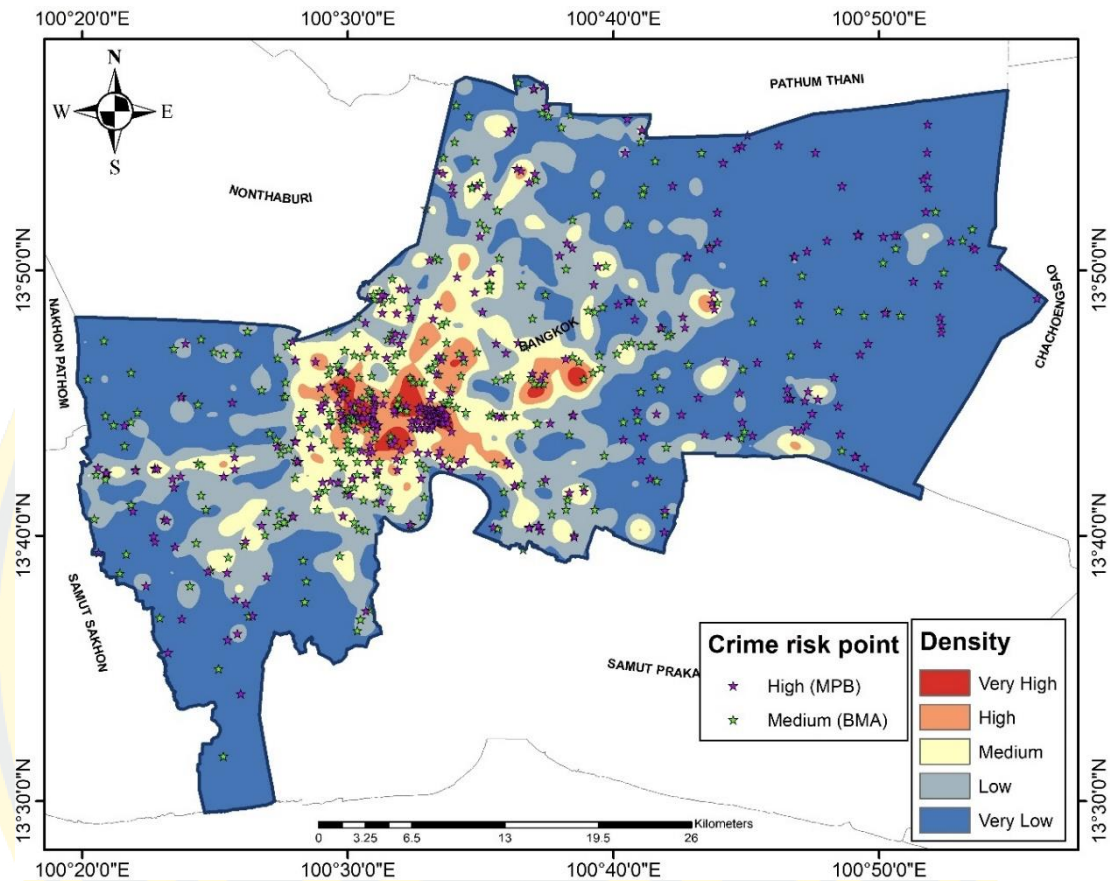


Figure 24 The density of financial services with high-medium crime risk points.

4.2.7 P7- Education (7,279 points)

Figure 25 shows the distribution of crime risk points in education areas. Terrorist attacks on educational institutions are not often but important target types [104]. Schools, colleges, and universities often lack security measures and can be perceived as relatively vulnerable, leading to high potential for mass casualties with common methods including bombings, armed assaults, and kidnappings. An important number of students fear serious harm or death might occur from student behavior within their schools [105]. Bullying and humiliation can lead some students to express retaliation violently. The major crime in institutions of higher education is often related to sexual assault, harassment, and stalking [106].

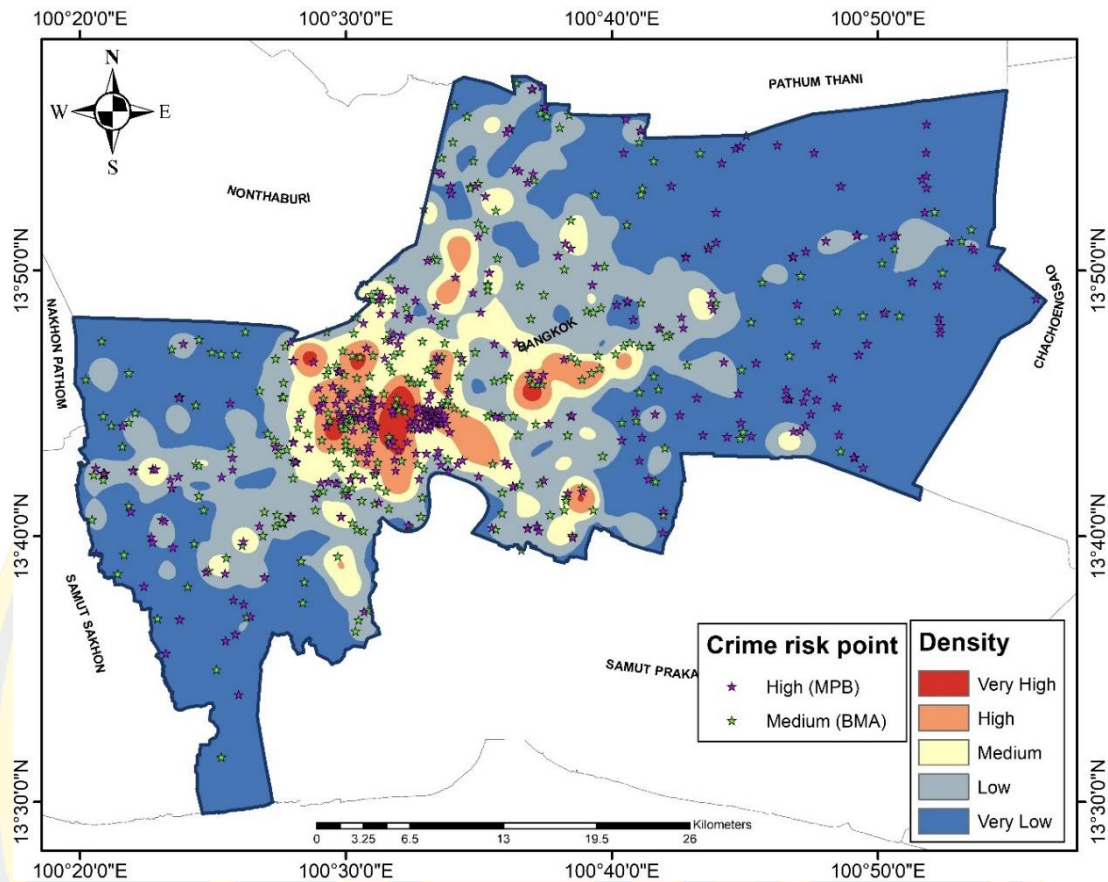


Figure 25 The density of education with high-medium crime risk points.

4.2.8 P8- Business and Industrial Production (55,400 points)

Businesses located in areas with high traffic and high commercial property have more experience with criminal activity [107]. Figure 26 shows the distribution of crime risk points in business and industrial production areas. This indicates that bustling environments and busy businesses can lead to increased crime rates. Commercial land use leads to more street crime in walkable neighborhoods [108]. However, this effect is mitigated by high population density, making dense mixed-use areas safer than typical residential areas.

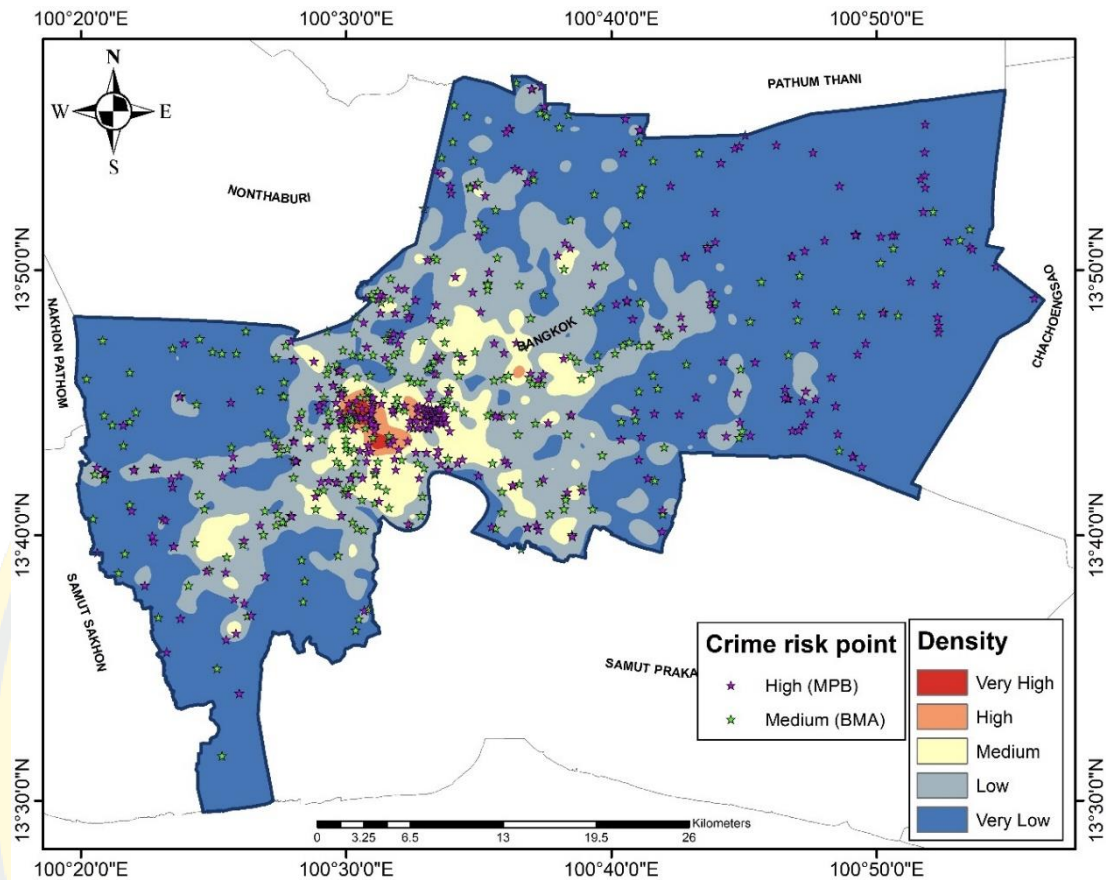


Figure 26 The density of business and industrial production with high-medium crime risk points.

4.2.9 P9-Transportation (8,258 points)

Crime in public transport covers different types of transport, such as buses, subways, trains, and taxis. Transit environments are public spaces related to people's everyday lives that contribute to high traffic volumes of passengers [109]. Figure 27 shows the distribution of crime risk points in transportation areas. Transportation hubs are attractive targets for criminal activities like smuggling, vandalism, and theft, influenced by environmental, social, and geopolitical situations that affect the system itself, employees, or passengers.

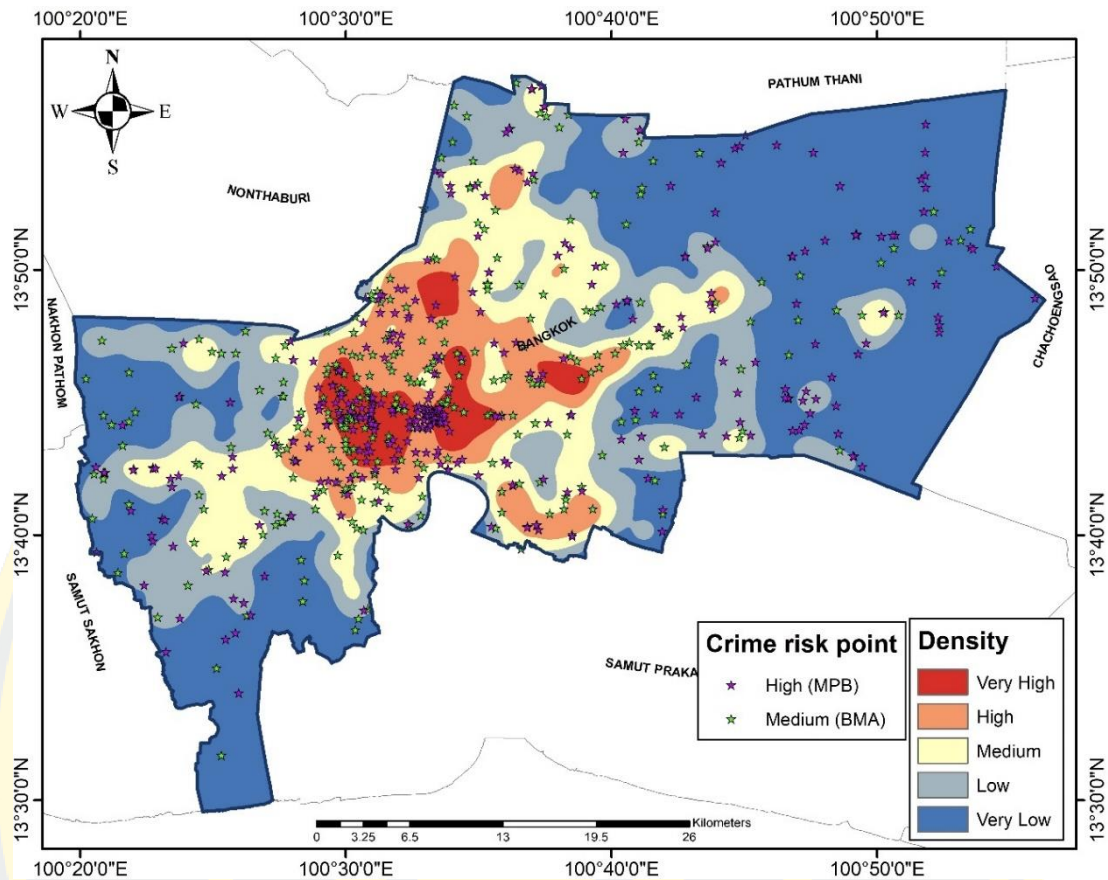


Figure 27 The density of transportation with high-medium crime risk points.

4.2.10 P10-Road network

Road networks influence crime by offering convenient access and escape routes, which lead to facilitating theft and criminal activity. Main roads and intersections attract higher crime rates due to heavy traffic and economic activity. Efficient road networks can facilitate the transportation of illegal goods such as drugs or stolen property. Poorly lit or unmonitored roads increase vulnerability to crimes such as muggings. Increased road network connectivity correlates with lower rates of criminal lethality, indicating that improved connectivity might decrease the severity of crimes [110, 111]. Offenders frequently engage in criminal acts along pathways that connect with major criminal attractors or activity nodes. Figure 28 shows the distribution of crime risk points in road network areas. The routes from criminals' residences to locations of crime are influenced by the road network, with some segments being traversed more often. This indicates that road networks have influenced the spatial distribution of crimes and the selection of criminal targets.

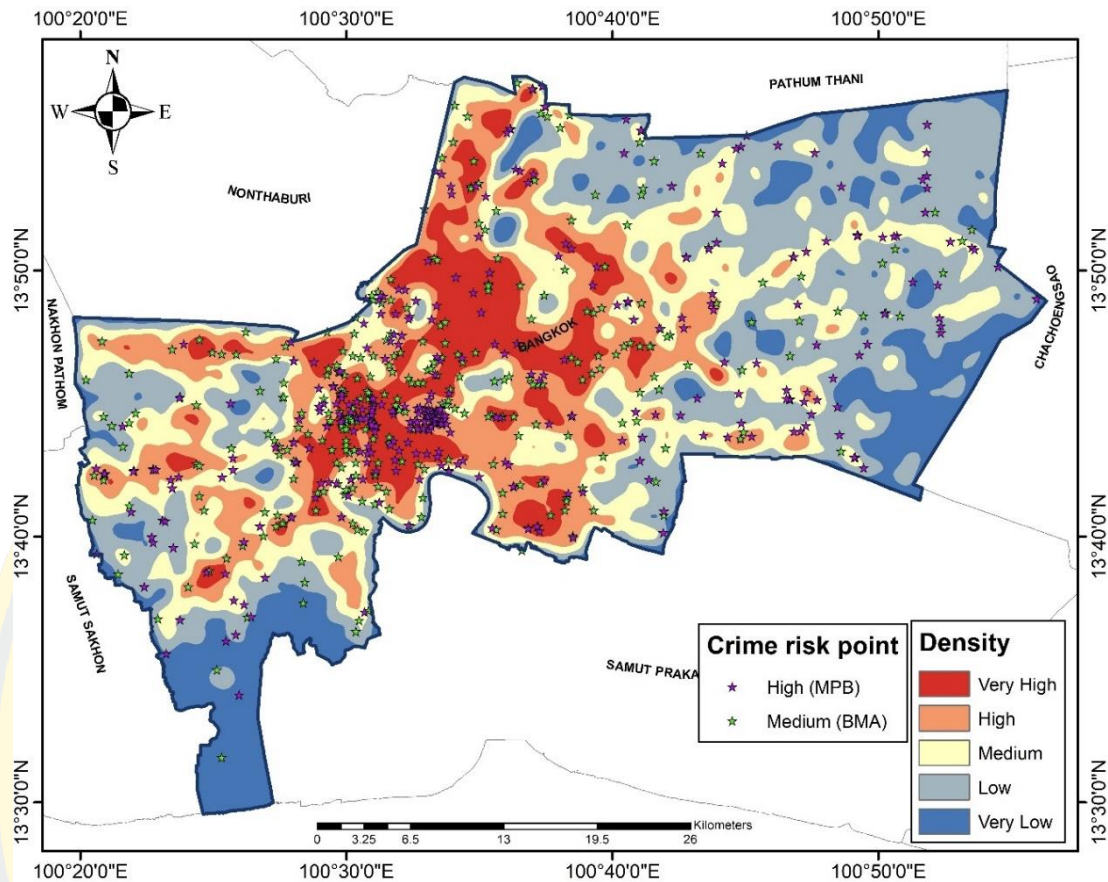


Figure 28 The density of road network with high-medium crime risk points.

4.2.11 P11-Residential (24,373 points)

Residential areas and crime risk interact through various factors. High population density and poverty are often linked with increased crime rates, whereas wealthy or well-organized areas usually deal with lower rates. Urban areas tend to have higher crime rates compared to suburban or rural areas due to density and economic inequality. Strong community cohesion and proper urban design, like proper lighting and surveillance, can reduce crime. The spatial awareness having an influence on the decision-making process of offenders is more likely to perpetrate crimes in their current and previous residence locales [112]. Figure 29 shows the distribution of crime risk points in residential areas. Socio-economic and social problems are strongly related to crime rates in residential areas, while the areas with high ethnic heterogeneity and a high percentage of single-family residences are more likely to be targeted for burglaries [113]. The built environment, including factors like street density, bus-stop density, and land use diversity, influences residential crime rates.

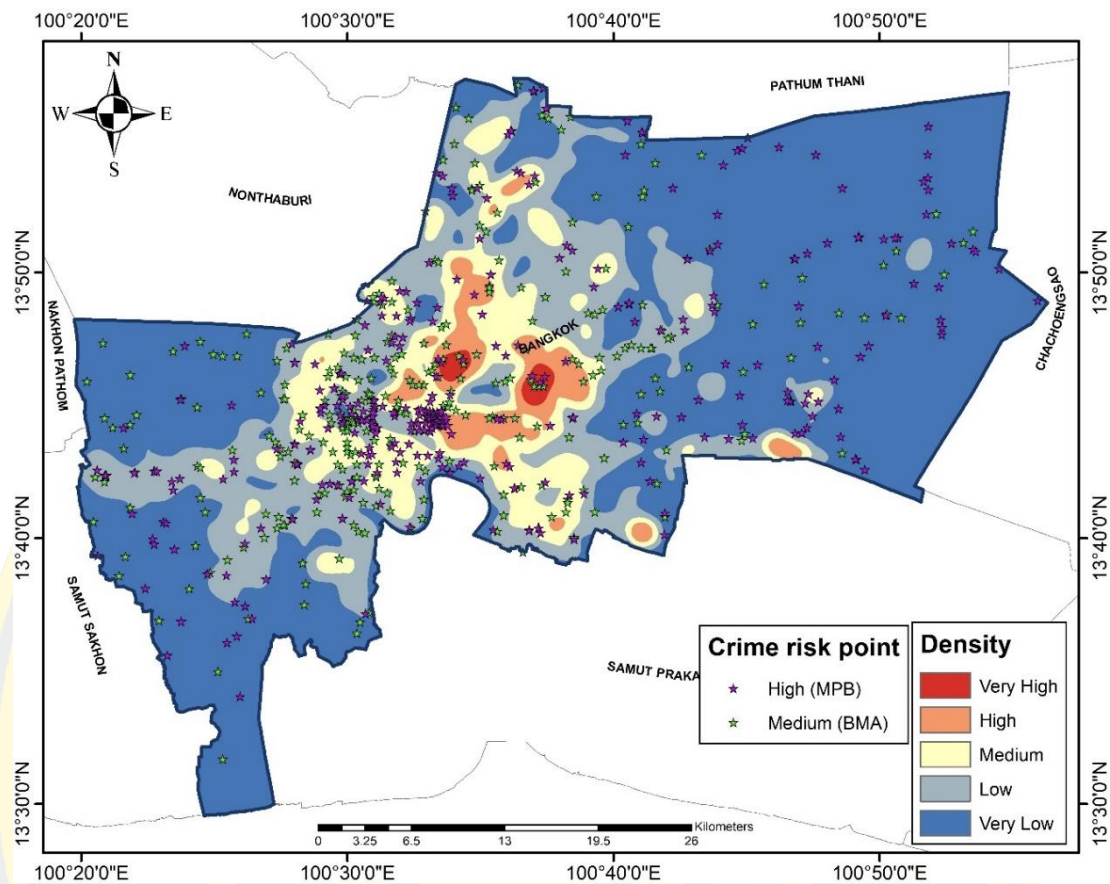


Figure 29 The density of residential with high-medium crime risk points.

4.2.12 P12- Population density

High population density correlates with increased opportunities for property crimes, in line with traditional criminological theories [114]. Figure 30 shows the distribution of crime risk points in population density areas. The effects of population density on crime can be nonlinear [115], with micro-density (local population density) having stronger effects on crime rates than macro-density (regional population density). Routine activity theory [116] suggests that increased human presence can both increase and decrease criminal activity, depending on the specific context and type of area (e.g., shopping, school, work areas).

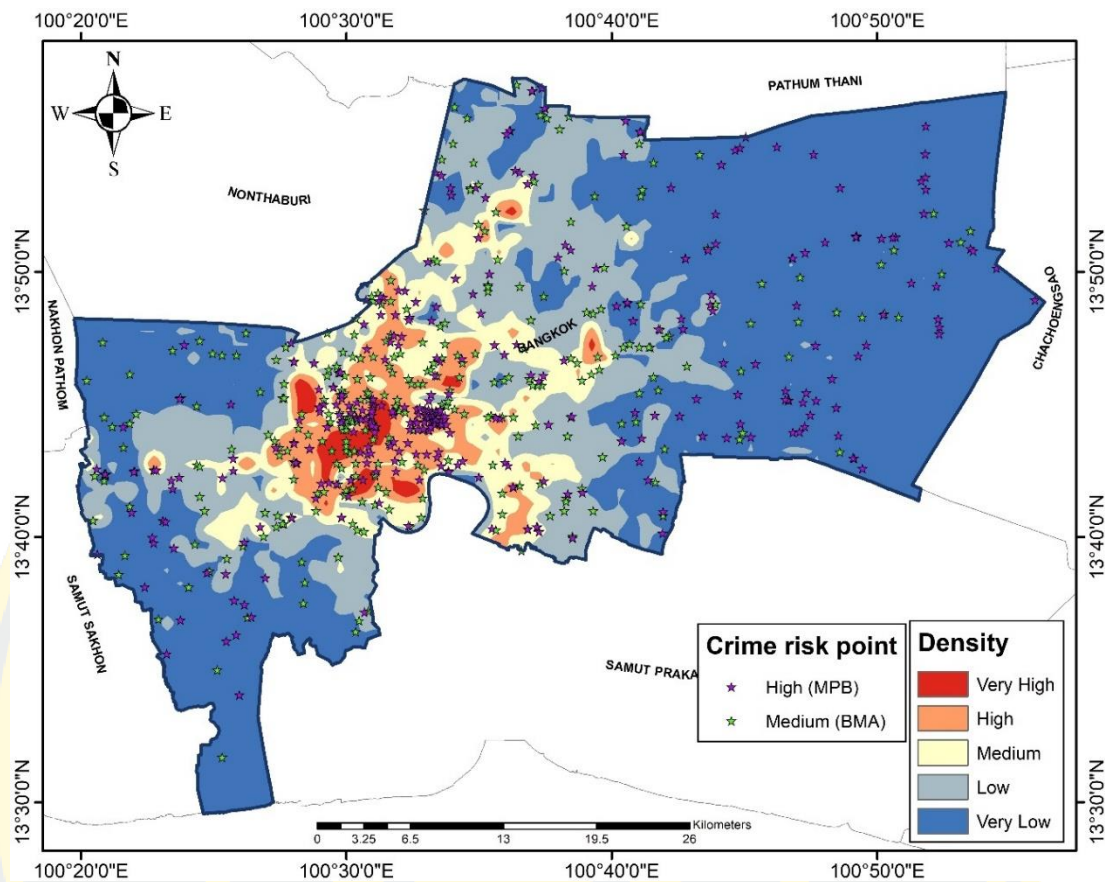


Figure 30 The density of population density with high-medium crime risk points.

4.2.13 P13- Greenspace area (NDVI)

The general green spaces refer to tree cover, parks, and ground cover. Figure 31 shows the distribution of crime risk points in greenspace areas. The correlation between green spaces and crime indicates green spaces in urban areas that are well-designed and well-maintained would support social interaction, recreation, community collaboration, reduced stress, and improved mental health that could also reduce property and violent crime rates, but poorly maintained green spaces could attract criminal activities by increasing vandalism, drug-related offences, and loitering [32, 117-119]. While the influenced socioeconomic context of the surrounding neighborhoods can also impact the impact of green spaces on crime, in areas with high levels of resource deprivation and physical disorder, parks may function as criminal marketplaces rather than providing social benefits. Facilities in green spaces and the socioeconomic context of surrounding neighborhoods have influenced the frequency and timing of criminal activities.

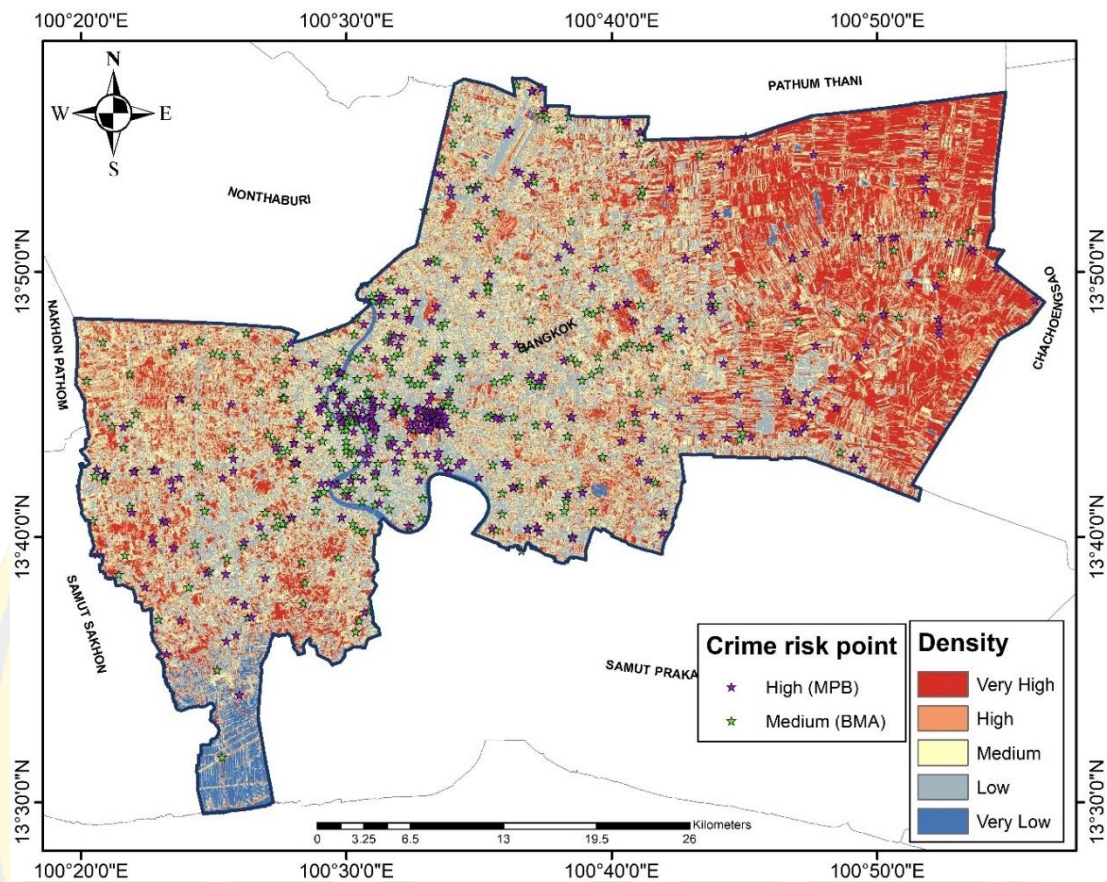


Figure 31 The density of greenspace with high-medium crime risk points.

4.2.14 P14- Nighttime light

The effects of increased lighting on crime rates are more pronounced in urban areas with higher nighttime business activity [120, 121]. Figure 32 shows the distribution of crime risk points in nighttime light areas. Street lighting can reduce nighttime crimes and arrests in public spaces, not more than daytime crimes. Indicate that community pride and informal social control may be more effective than increased surveillance and deterrence.

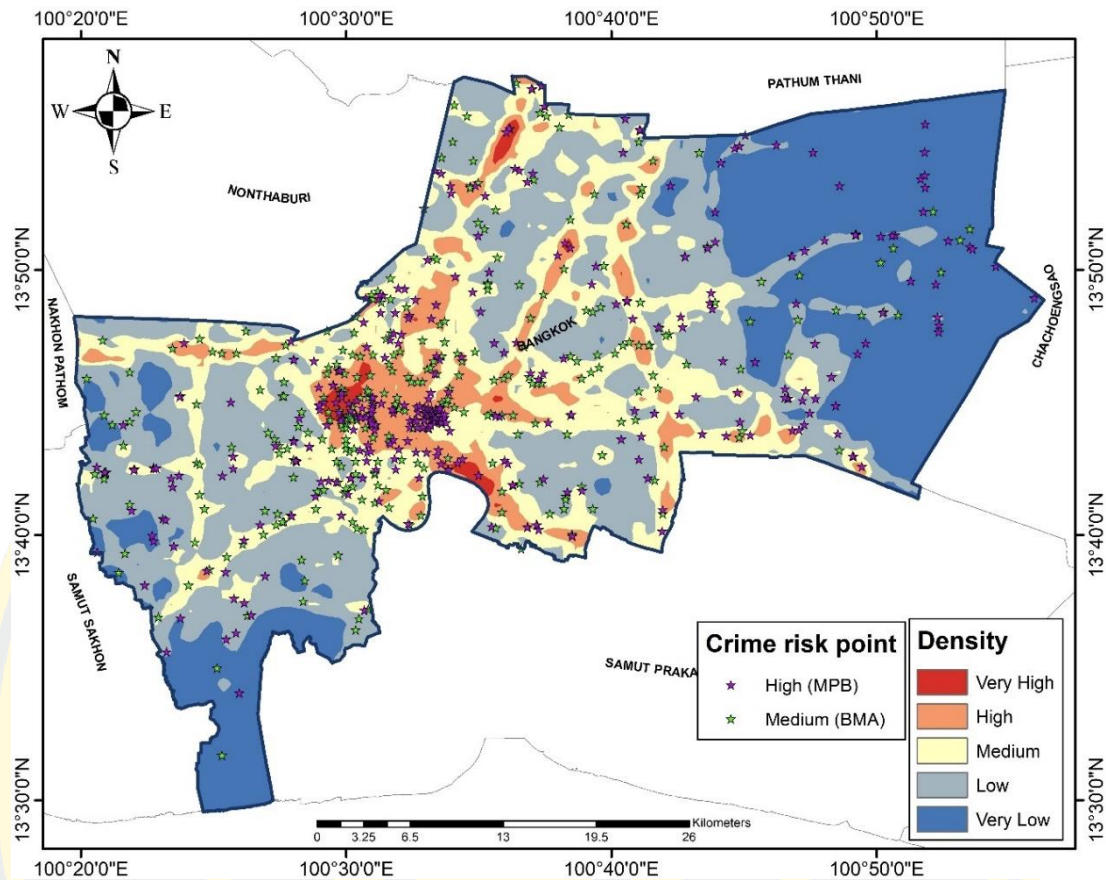


Figure 32 The density of nighttime light with high-medium crime risk points.

4.3 Comparing the Performance of Classification ML Models

4.3.1 Buffer analysis

The crime density analysis needs to determine that the radius value is not too large, around 10–100 meters, because when analyzed, correct prevention and corrective actions can be taken at the point in the area or radius of the area. A crime has occurred directly [122]. Therefore, this research was conducted to compare the area radius of crime risk points in different distances from 10, 20, 30, 40, 50, and 100 meters to compare the accuracy with the machine learning model that gives high accuracy, and the comparison results shown in Figure 33 show that the buffer distance of 20 meters is the distance where the accuracy is stable, this research chose a buffer distance of 20 meters.

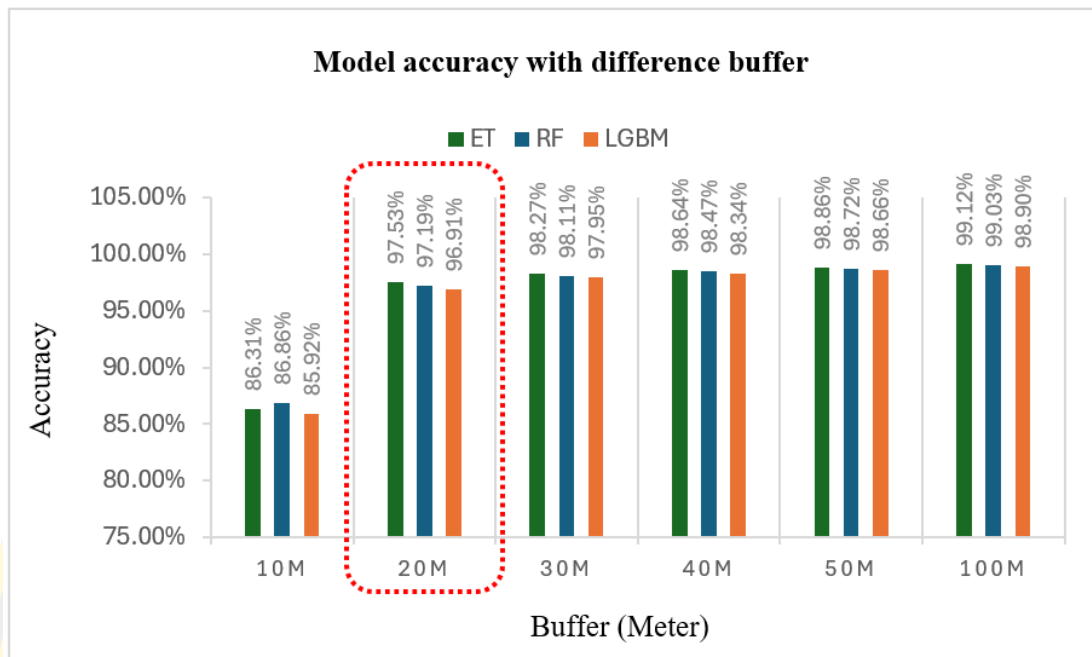


Figure 33 Model accuracy with different buffer.

4.3.2 Model training and hyper-parameter optimization

A machine learning method was used to analyze risky areas in Bangkok by using data on crime risk points from the Metropolitan Police Bureau, BMA, and civil complaints. The study also includes point-of-interest (POI) data and satellite images. To predict areas that tend to be at risk of crime occurring, three machine learning models will be processed and compared: extra tree (ET), random forest (RF), and light gradient boosting machine (LGBM). Evaluating accuracy, precision, recall, and F1 score to find the optimal performance of models.

Building a machine learning model in a Jupyter Notebook environment starts with loading raster images, which is the first step in data preparation. (preprocessing) The model is created using two types of data: training data (data Y) and input data (data X). The training set data is divided into two parts: train 80% and validate 20%. Splitting the data this way allows us to easily evaluate the performance of the model. It also helps prevent overfitting, a situation where a model performs poorly on the training data. But it doesn't work well with new data. A 10-fold cross-validation method, which divides the data into 10 subsets and trains the model 10 times, using each subset as the test set and the remaining subsets as the training set, prevents overfitting. Find the average of the model's performance from all 10 times.

Customizing hyperparameters Model performance is an important part of building a machine-learning model. This can be done by experimenting and evaluating validated data or by using random search automated methods, for comparing model performance with parameters as shown in Table 7.

Table 7 Summary parameter of ML Models.

Characteristics	Descriptions
Model	Extra tree, Random Forest, LightGBM
K-fold	K = 10
Parameters	RF (n_estimators = 744) LGBM (n_estimators = 400, learning_rate = 0.2, num_leaves = 40) ET (estimators = 100, min_samples_leaf = 1, min_samples_split = 5)
Dependent variable	Crime risk point (high, medium, low)
Independent variable	P1 - P14

4.3.3 Features importance of parameters

Feature selection is a decision-making process to select features that is important to handling high-dimensional data in machine learning methods [123], as it influences efficiency by reducing computation time, enhancing accuracy, and improving model interpretability., also known as variables or attributes, to use in the model. Is an important step in building an ML model. The feature importance function is a common technique for interpreting classification machine learning models [124] to assess the score of each input feature. The feature that has a higher score has a greater effect on the model's ability to predict a certain variable.

Finding feature importance will make the model more accurately reflect the real factors in the prediction, reflecting the risks of both quantitative variables and important qualitative variables, as shown in Figure 34, which is the feature importance of the ET model that revealed the Business and Industrial (P8) factor had the highest impact on the prediction, accounting for 8.06%. There are 13 out of 14 factors with no significant difference in affecting the model, except for factor NDVI (P13), which had the least impact on the model, accounting for 0.02%. While the

group average feature importance of the ET model shows demographics is highest, followed by nighttime lights, POI, and greenspace shown in Figure 35.

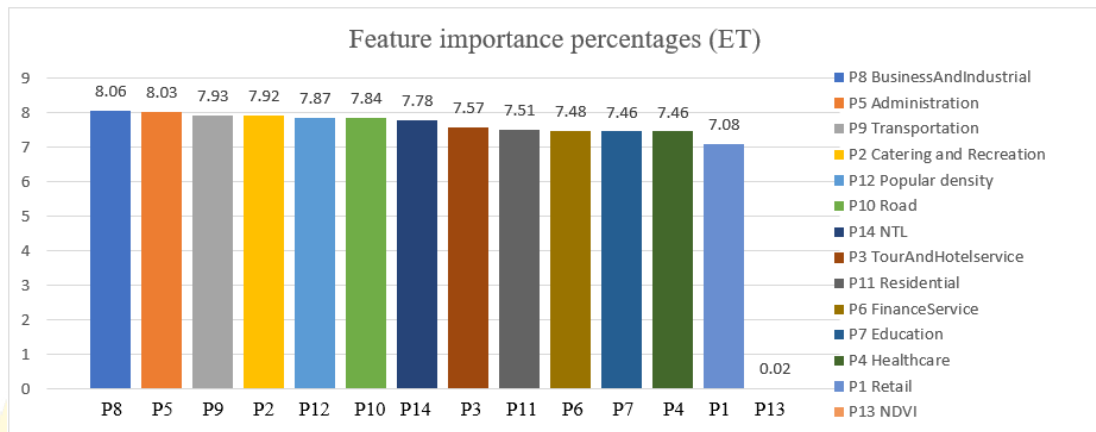


Figure 34 Feature importance of ET model.

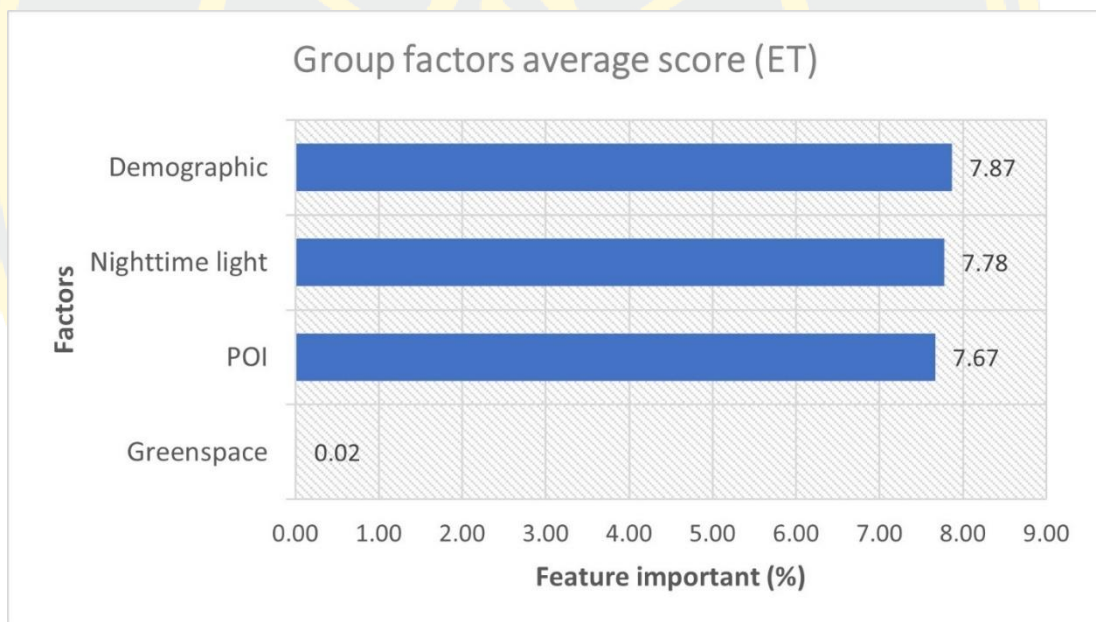


Figure 35 Group average feature importance of ET model.

The feature importance of the RF model as shown in Figure 36. The business and industrial (P8) factor had the highest impact on the prediction, accounting for 8.85%. There are 13 out of 14 factors with no significant difference in affecting the model, except for factor P13 NDVI, which had the least impact on the model, accounting for 0.01%. While the group average feature importance of the RF model shows demographics is highest, followed by POI, nighttime lights, and greenspace shown in Figure 37.

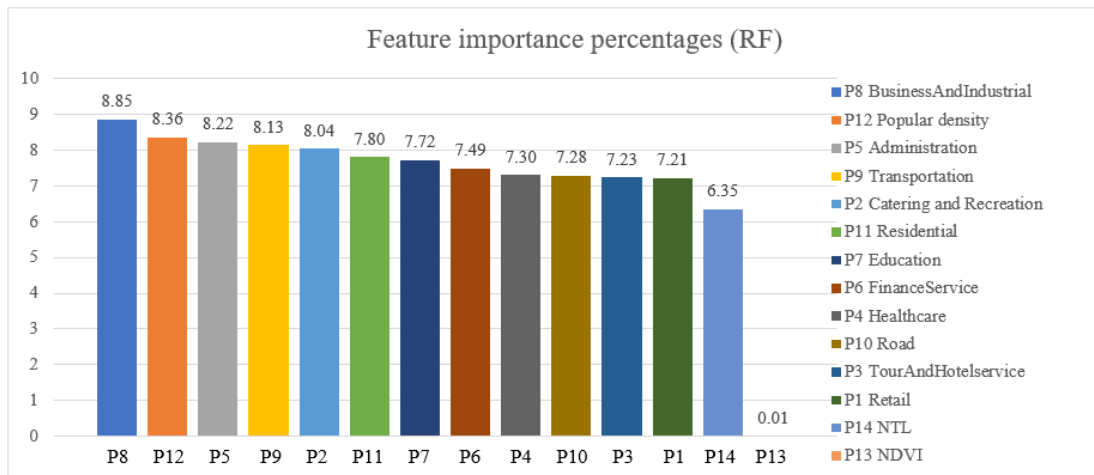


Figure 36 Feature importance of RF model.

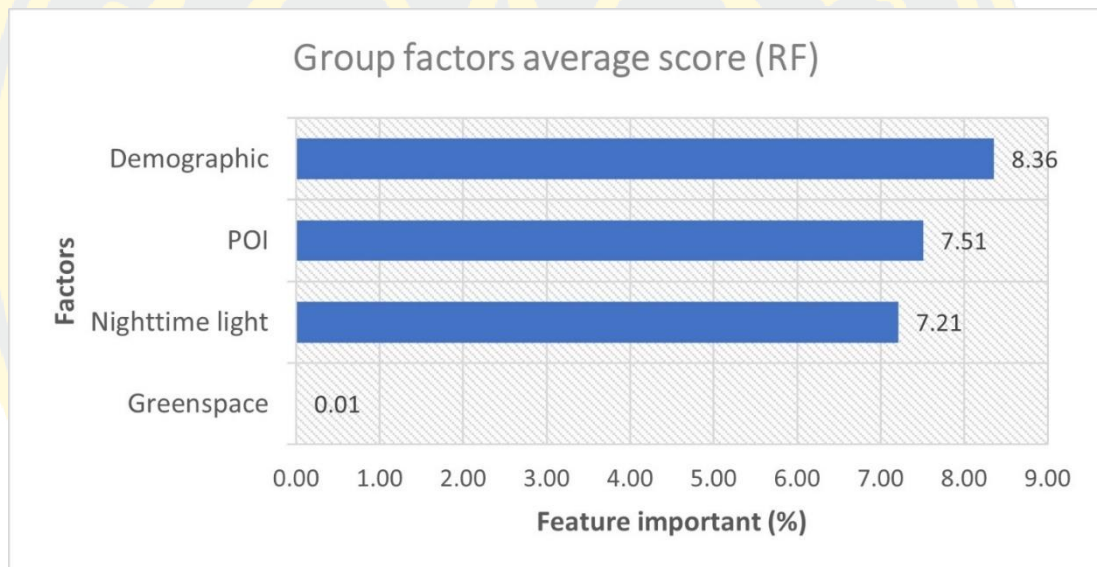


Figure 37 Group average feature importance of RF model.

The feature importance of the LGBM model as shown in Figure 38. The road network (P10) factor had the highest impact on the prediction, accounting for 8.14%, while the NDVI (P13) factor, which had the least impact on the model, accounted for 4.78%. However, with 14 factors, there is no significant difference affecting the model. While the group average feature importance of the ET model shows demographics is highest, followed by nighttime lights, POI, and greenspace shown in Figure 39.

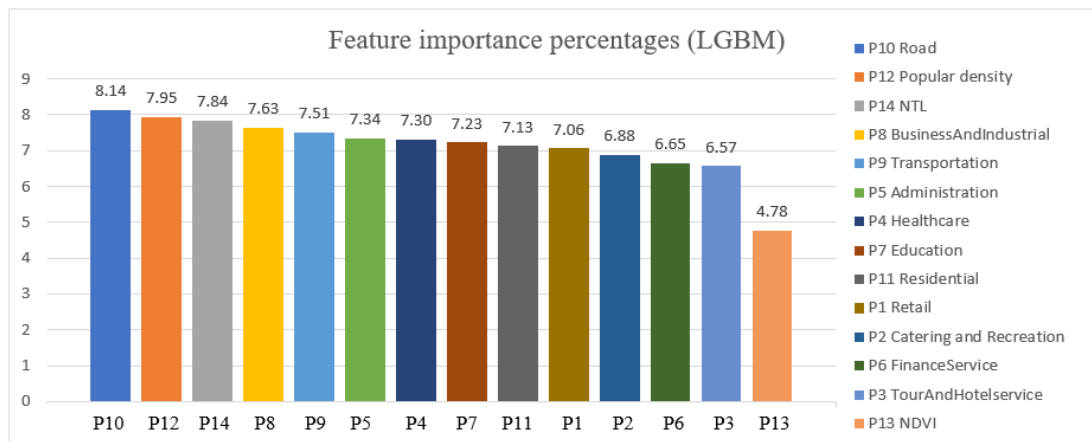


Figure 38 Feature importance of LGBM model.

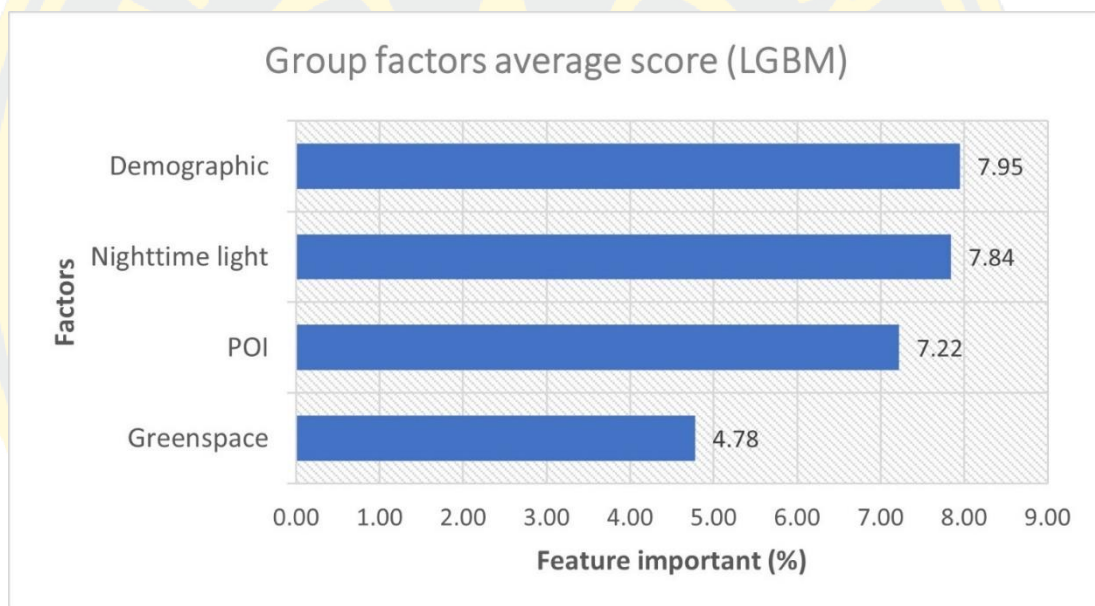


Figure 39 Group average feature importance of LGBM model.

4.3.4 Evaluating ML model performance with a confusion matrix

A confusion matrix is a summarized table of the number of correct and incorrect predictions used to check the performance of classification models and understand the model's performance. Figure 40 shows a square matrix comparing the three machine learning model predictions with actual values. The diagonal square matrix, from top left to bottom right, represents the number of correct predictions for each class. The off-diagonal square matrix shows the number of incorrect predictions.

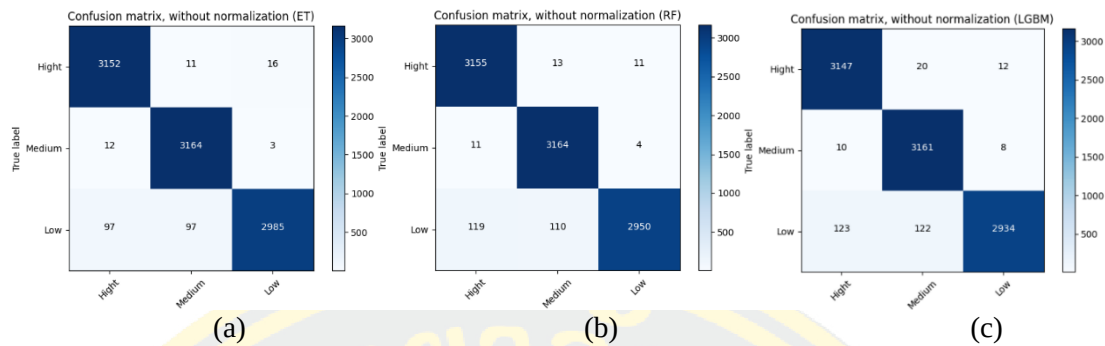


Figure 40 Confusion matrix of (a) extra tree (b) random forest (c) lightGBM.

4.3.5 Evaluating ML model performance with ROC-AUC

The Receiver Operating Characteristic (ROC) is a graphical method used to evaluate the effectiveness of a classifier. Classifier performance is characterized by a pair of statistics known as the true positive rate and the false positive rate. The data is represented on a two-dimensional graph, where the x-axis represents the false positive rate, and the y-axis represents the true positive rate. The generated figure can be used to assess the relative efficacy of several classifiers as well as to ascertain whether a classifier outperforms random guessing. The ROC graph measures the performance of True Positive Rate and False Positive Rate. The ROC AUC value is a measure of the area under ROC curve, where the value of a well-performing model gives an ROC AUC value close to 1.

ET Model: The ROC graph and ROC AUC values can be shown in Figure 41. The ET model provides the highest ROC AUC values for all 3 classes (high, medium, and low) at ROC AUC 100%.

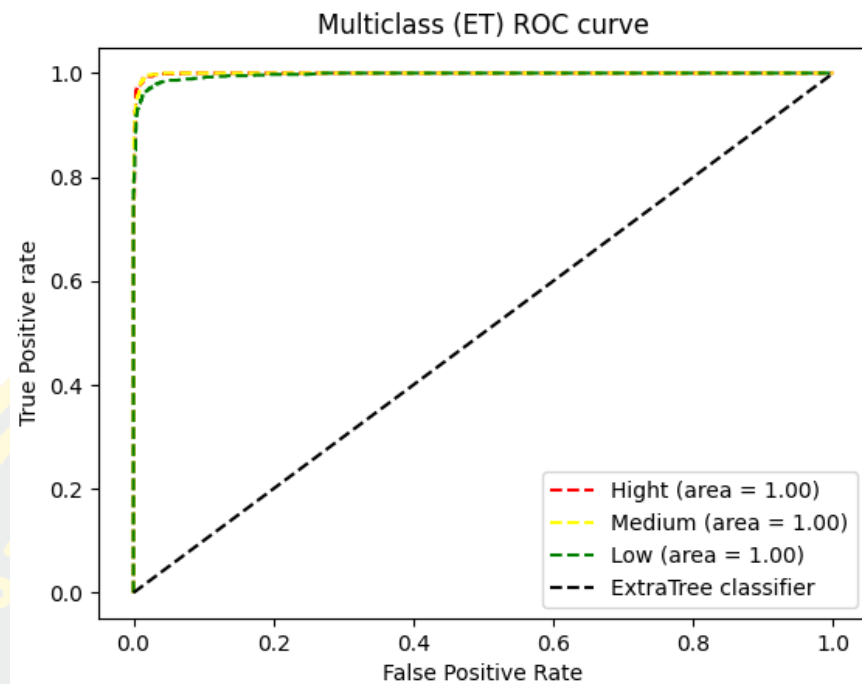


Figure 41 ROC AUC of ET Model.

RF Model: The ROC graph and ROC AUC values can be shown in Figure 42. The RF model provides the highest ROC AUC values for all 3 classes (high, medium, and low) at ROC AUC 100%.

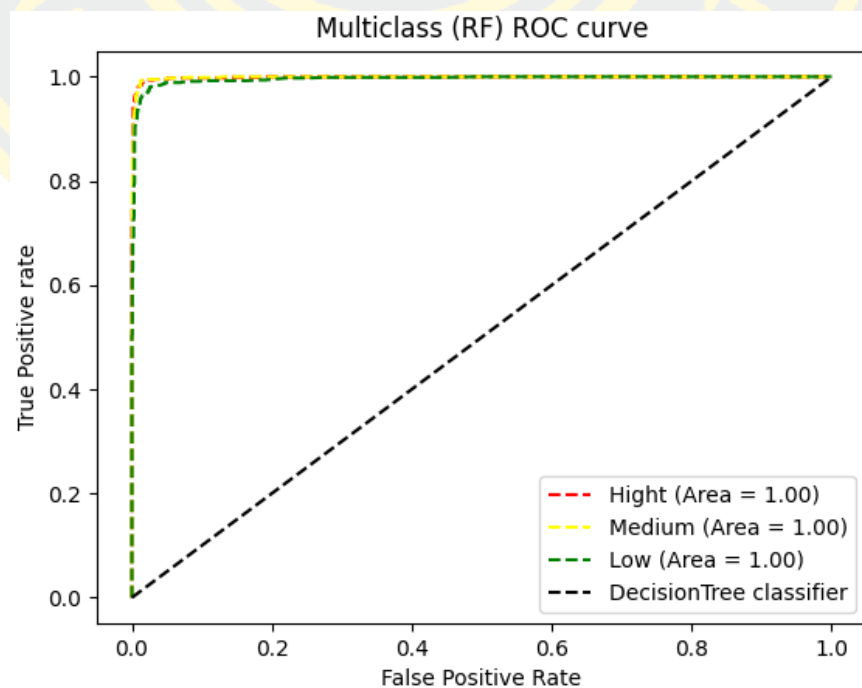


Figure 42 ROC AUC of RF Model.

LGBM Model: The ROC graph and ROC AUC values can be shown in Figure 43. The LGBM model provides the highest ROC AUC values for all 3 classes (high, medium, and low) at ROC AUC 100%.

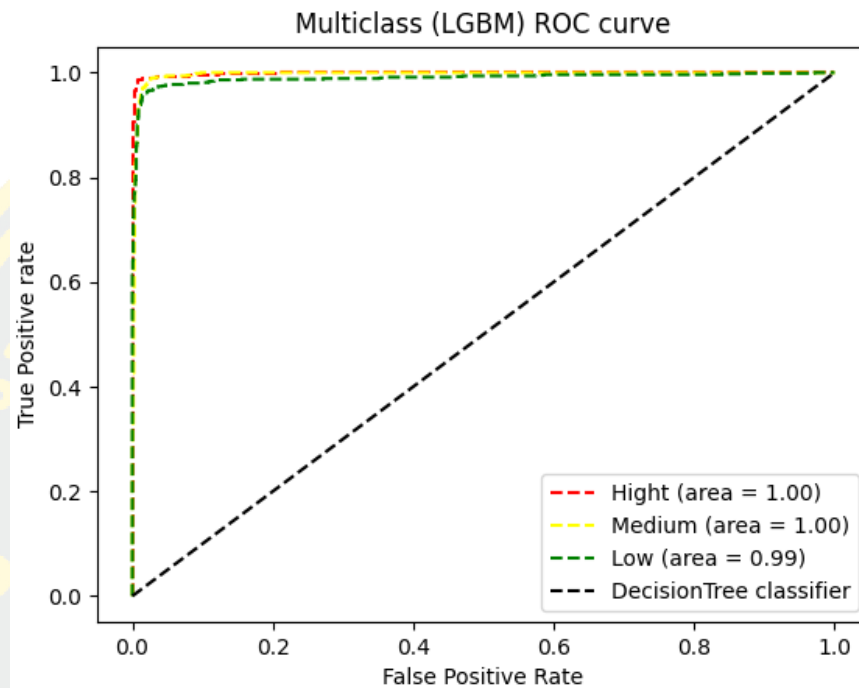


Figure 43 ROC AUC of LGBM Model.

4.3.6 Evaluating ML model performance with PRC-AUC

The precision-recall curve (PRC) is widely used for evaluating the performance of classification. The curve shows a relationship between the positive prediction value of a classifier and its true positive rate and often provides a useful alternative to the well-known receiver operating characteristic (ROC).

The PRC-AUC is a quantitative statistic that represents the extent of the area under the Precision-Recall Curve (PRC). PRC-AUC values close to 1 indicate highly effective model performance.

ET Model: The PRC graph and PRC-AUC values can be shown in Figure 44. The ET model gives the highest PRC-AUC values: class high is represented in the blue line with 100%, class medium is represented in the orange line with 100%, and class low is represented in the green line with a lower ROC than others with 99%, while overall average precision of 99% is represented in the red line.

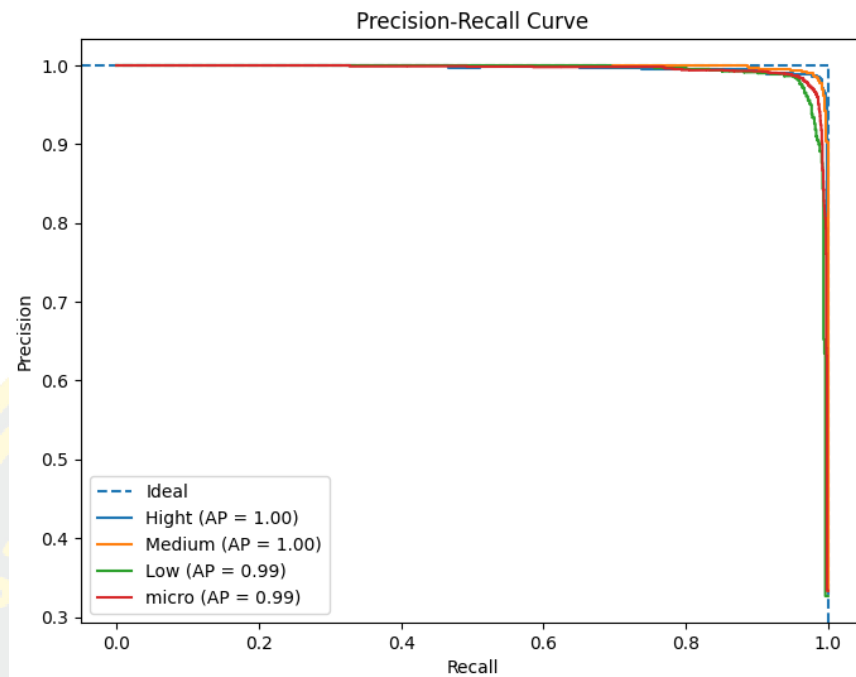


Figure 44 PRC-AUC of ET Model.

RF Model: The PRC graph and PRC-AUC values can be shown in Figure 45. The RF model provides the highest PRC-AUC values for all 3 classes (high, medium, and low), with a PRC-AUC of 100%.

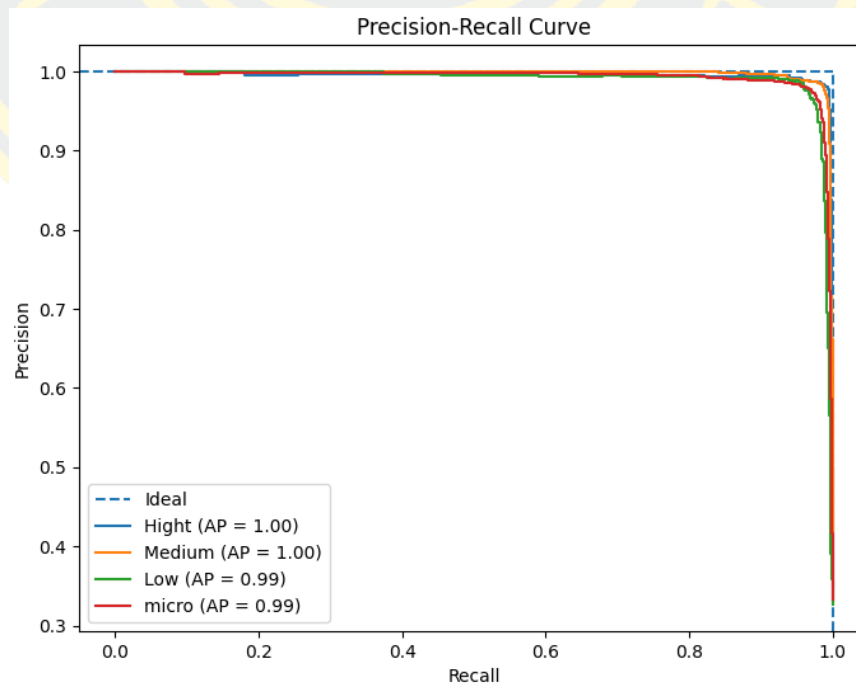


Figure 45 PRC-AUC of RF Model.

LGBM Model: The PRC graph and PRC-AUC values can be shown in Figure 46. The LGBM model provides the highest PRC-AUC values for all 3 classes (high, medium, and low), with a PRC-AUC of 100%.

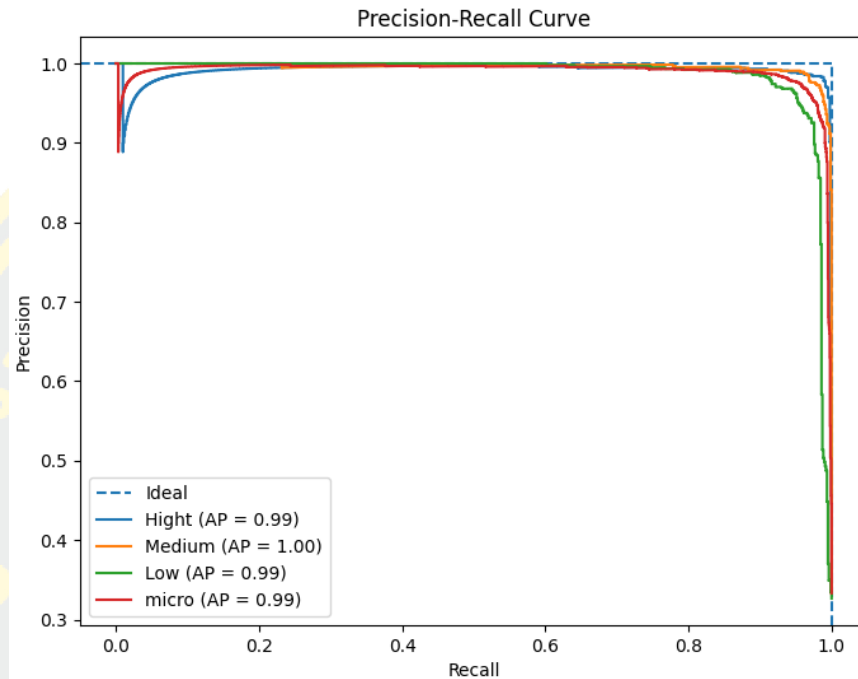


Figure 46 PRC-AUC of LGBM Model.

4.3.7 Evaluating ML model performance with accuracy

The three machine learning models compare performance by using evaluation metrics like accuracy, precision, recall, and F1 score on the train and test datasets shown as Table 8. The measurements considered in both the train and test are The ET model has a train score of 99.89 in accuracy, precision, recall, and F1Score, while the test scores are 97.53, 97.58, 97.53, and 97.51, respectively. The RF model has a train score of 99.96 in accuracy, precision, recall, and F1Score, while the test scores are 97.19, 97.27, 97.19, and 97.17, respectively. The LGBM has a train score of 99.96 in accuracy, precision, recall, and F1Score, while the test scores are 96.91, 96.99, 96.91, and 96.88, respectively.

In conclusion, the results reveal the ET model has outperformed other models with a training accuracy of 99.89% and a test accuracy of 97.53%. The ET model performs slightly better than the RF model in classification risky areas in Bangkok. While the LGBM model is lower than the other models.

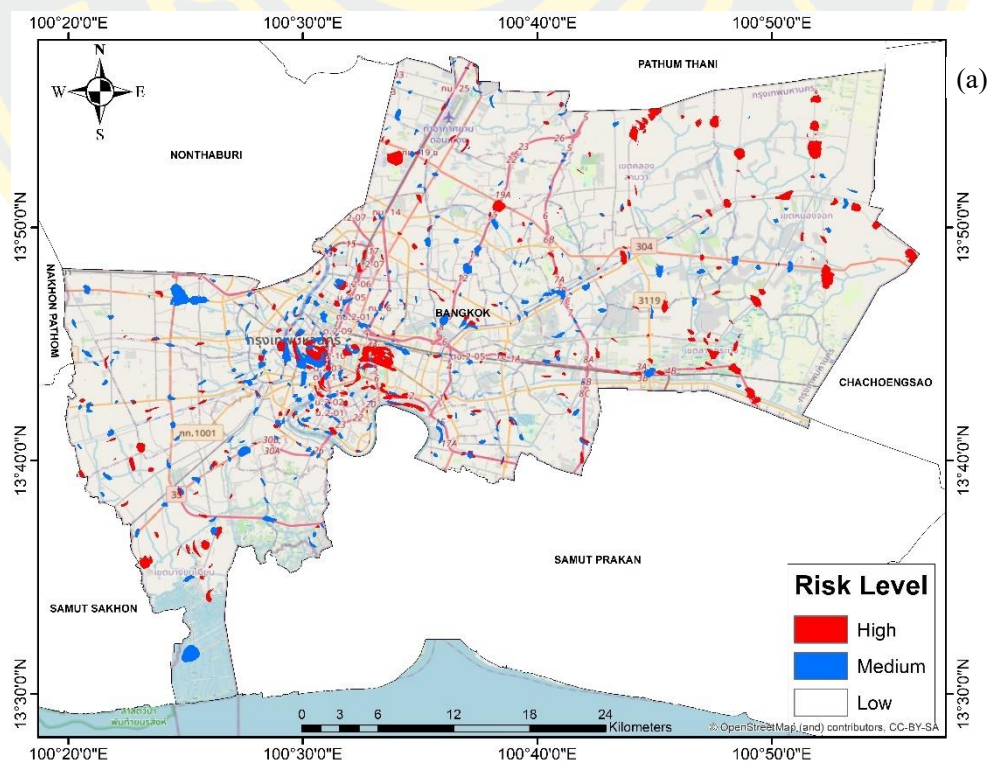
Table 8 ML Models measurement.

ML Models	Accuracy (%)		Precision (%)		Recall (%)		F1Score (%)	
	Train	Test	Train	Test	Train	Test	Train	Test
ET	99.89	97.53	99.89	97.58	99.89	97.53	99.89	97.51
RF	99.96	97.19	99.96	97.27	99.96	97.19	99.96	97.17
LGBM	99.96	96.91	99.96	96.99	99.96	96.91	99.96	96.88

4.4 Classification of risky area map in Bangkok

4.4.1 The classification of risky areas across Bangkok

Classification is a method in machine learning that uses predictive algorithms to identify the class or category of a certain data point within a dataset. Final maps from class prediction for the study area from three ML models are visually interpreted based on the three risky class in colors, including high risky areas are red color, medium risky area are blue color, and low risky area are transparent color as shown in Figure 47.



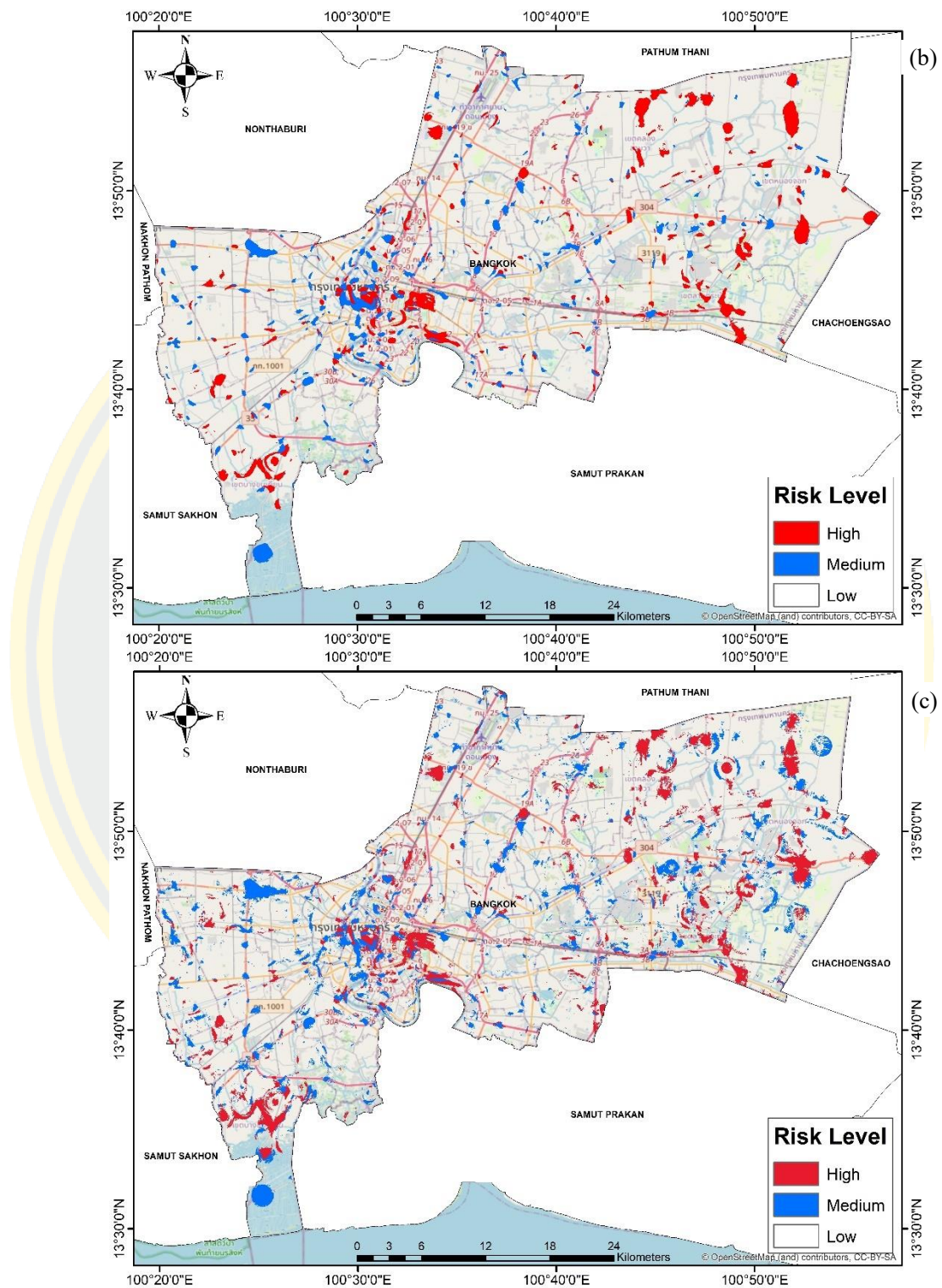


Figure 47 Classification map of a risky area in Bangkok from (a) ET model (b) RF mode (c) LGBM model.

Zonal statistics are calculated from the values of raster cells within zones (also known as classes or regions) defined by another raster or vector dataset. Zonal

statistics summarize the values of a group of cells by class value. Zonal statistics are useful for comparison between zones or cells in a zone. That is useful for monitoring trends over time across various zones or geographical regions. Table 9 is shown the summarize each of the classification risky areas in Bangkok from three ML models by using raster layer zonal statistics in QGIS. The ET model has ratios of risky area class of 2.59%, 2.52%, and 94.89% in class high, medium, and low risk respectively. The RF model has ratios of risky area class of 4.16%, 3.07%, and 92.77% in class high, medium, and low risk respectively. The LGBM model has ratios of risky area class of 5.51%, 6.01%, and 88.48% in class high, medium, and low risk respectively.

Table 9 Classification of risky area ratios by zonal statistics.

Model	Risky area class ratios (%)		
	High	Medium	Low
ET	2.59	2.52	94.89
RF	4.16	3.07	92.77
LGBM	5.51	6.01	88.48

4.4.2 The district where crime statistics against tourists

Bangkok is generally safe for tourists, but caution is advised in certain areas where scams, petty crimes, and overcharging can occur. To stay safe, use reputable transportation, avoid unlicensed establishments, and remain vigilant in crowded places. Bangkok has 50 districts, but eight districts found crimes against tourists based on the official report shown in Figure 48. The map displays the colors on the map representing different ranges of crime statistics. Blue-shaded (light and dark) areas have reported low crime. The light blue represents districts with 1-2 incidents in Bang Rak (1005), while the darker blue indicates 3-6 incidents in Samphanthawong (1003), Yannawa (1006), and Bang Kho Laem (1008). The orange-shaded areas are Sa Thon (1007) and Khlong Toei (1018) have moderate crime reports, which represent districts with 7–17 incidents. The red area is Pathum wan (1004) and Watthana (1039) districts have reported the highest number of incidents, ranging from 18 to 22.

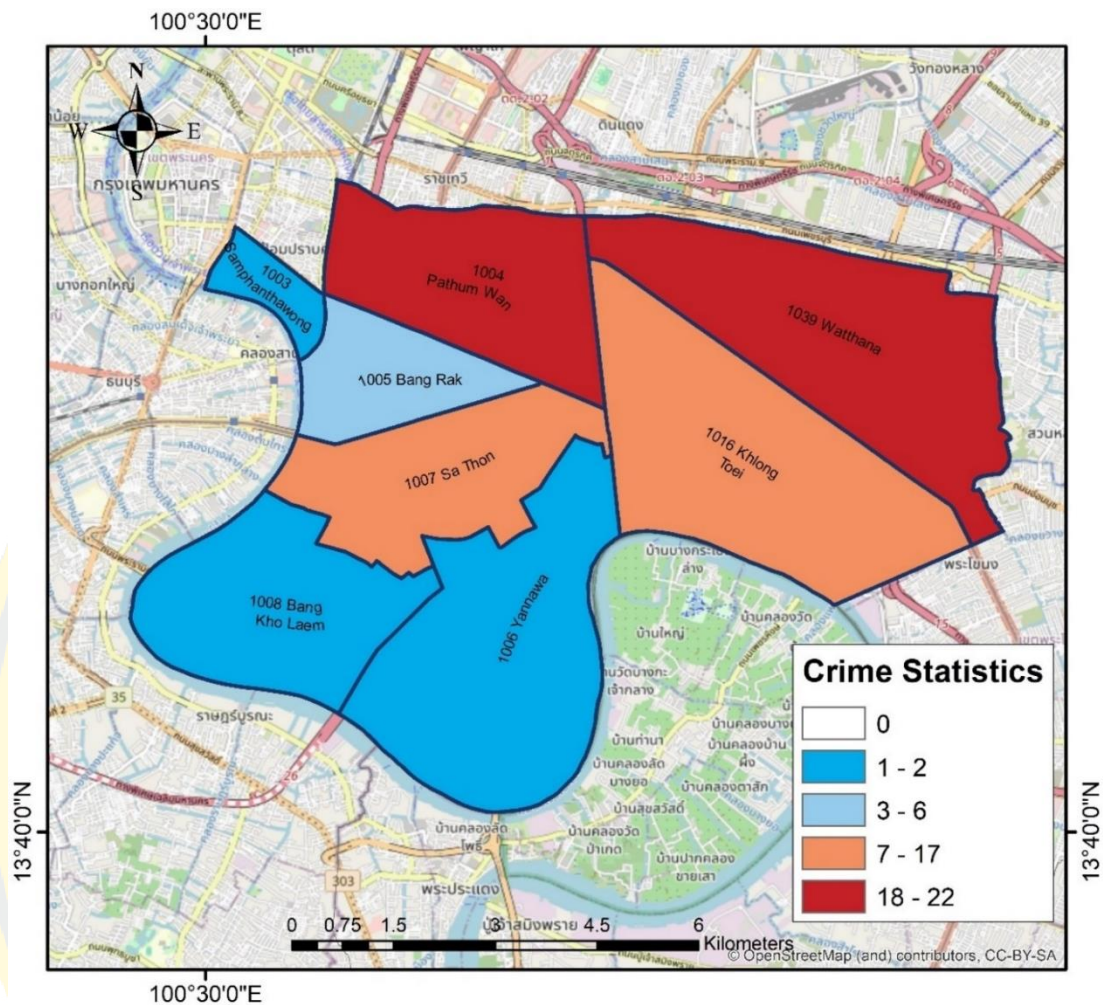


Figure 48 The map shows the districts where crime statistics against tourists were found.

Classification crime risk area in eight districts that found crime statistics against tourists shown in Figure 49. The map shows the distribution of high- and medium-risk areas across districts such as Pathum Wan, Watthana, Khlong Toei, and Bang Rak. Red areas represent high-risk zones, where crimes against tourists are more frequent. These are areas tourists should be more cautious of because there is a greater likelihood of criminal activity occurring. Blue areas represent medium-risk zones of crime for watching. While white areas indicate low-risk zones with fewer crime incidents reported, that trend is safer for residents and tourists.

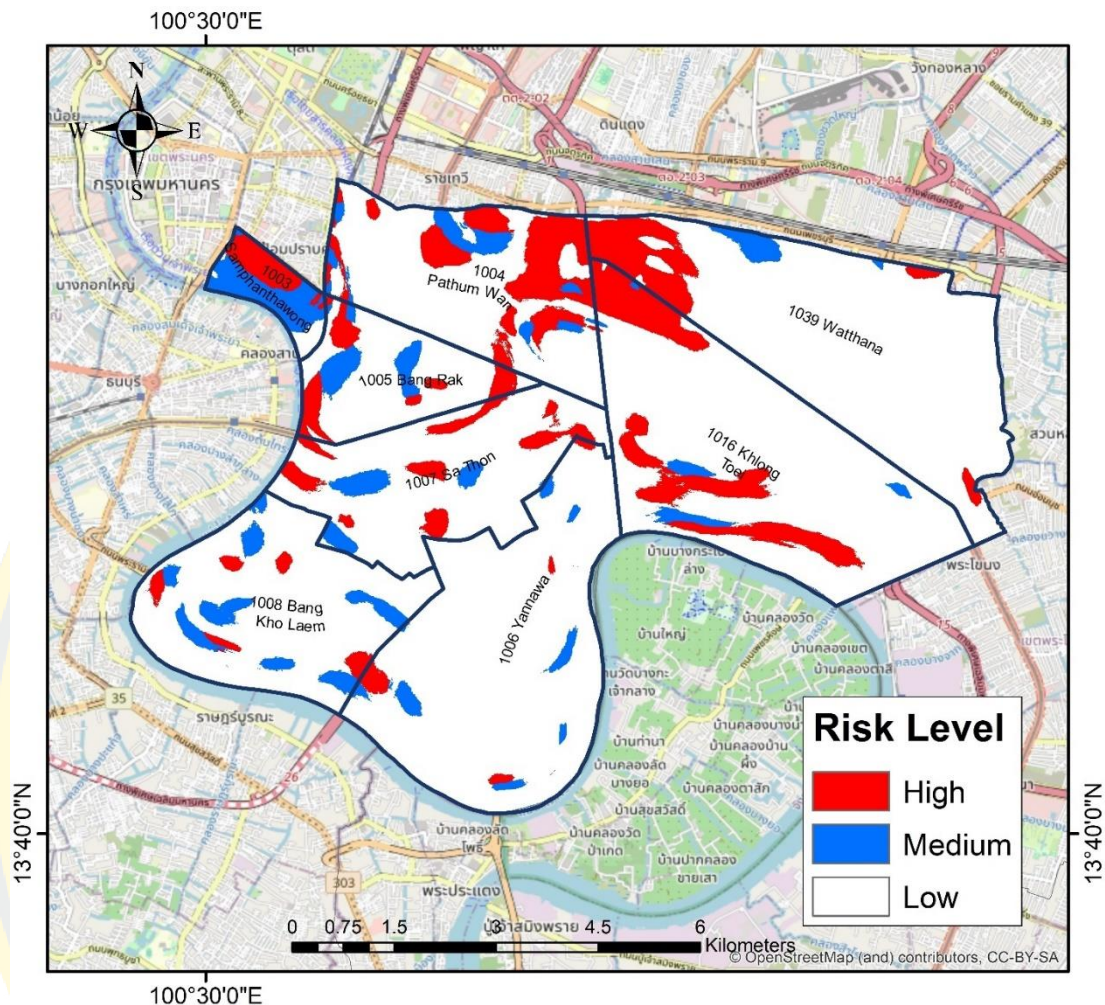


Figure 49 A map shows the risky areas of districts were crime statistics against tourists.

The statistics presented in Table 10 provide a detailed breakdown of the crime risk classification across eight districts in Bangkok, specifically focusing on the risk against tourists. The table classified the areas based on their level of crime risk as high, medium, or low. These categories are represented along the X-axis, with each district listed on the Y-axis. Figure 50 further revealed the percentage distribution of areas within each district that are classified under these three risk levels and shows different districts in Bangkok are perceived in terms of safety for tourists.

1003 Samphanthawong: Has an important portion of high-risk areas (33.15%), with medium risk at 12.57% and low risk at 54.28%.

1004 Pathum Wan: Has the majority area as low risk (66.86%), while medium risk is at 25.88% and only 7.26% is high risk.

1016 Khlong Toei: Has the majority area as low risk (80.16%), with medium risk at 17.37% and a small high-risk portion of 2.47%.

1005 Bang Rak: Has the majority area as medium risk (15.82%), with high risk at 12.38% and low risk at 71.80%.

1007 Sathon: Has the majority area as low risk (82.05%), with medium risk at 11.64% and high risk at only 6.30%.

1039 Watthana: Shows the low percentage of high-risk areas at 2.24%, with medium risk at 10.24% and most areas classified as low risk (87.52%).

1008 Bang Kho Laem: Has the majority area as low risk at 84.28%, while medium and high risks account for 11.50% and 4.23%, respectively.

1006 Yannawa: Shows the district with the highest proportion of low-risk areas (94.64%), with medium and high-risk areas making up only a small proportion of 3.52% and 1.84%, respectively.

Table 10 The ratio of crime risk areas across eight districts.

Code	District name	Crime-risk level		
		High (%)	Medium (%)	Low (%)
1003	Samphanthawong	33.15	54.28	12.57
1004	Pathum Wan	25.88	7.26	66.86
1016	Khlong Toei	17.37	2.47	80.16
1005	Bang Rak	15.82	12.38	71.80
1007	Sa Thon	11.64	6.30	82.05
1039	Watthana	10.24	2.24	87.52
1008	Bang Kho Laem	4.23	11.50	84.28
1006	Yannawa	1.84	3.52	94.64

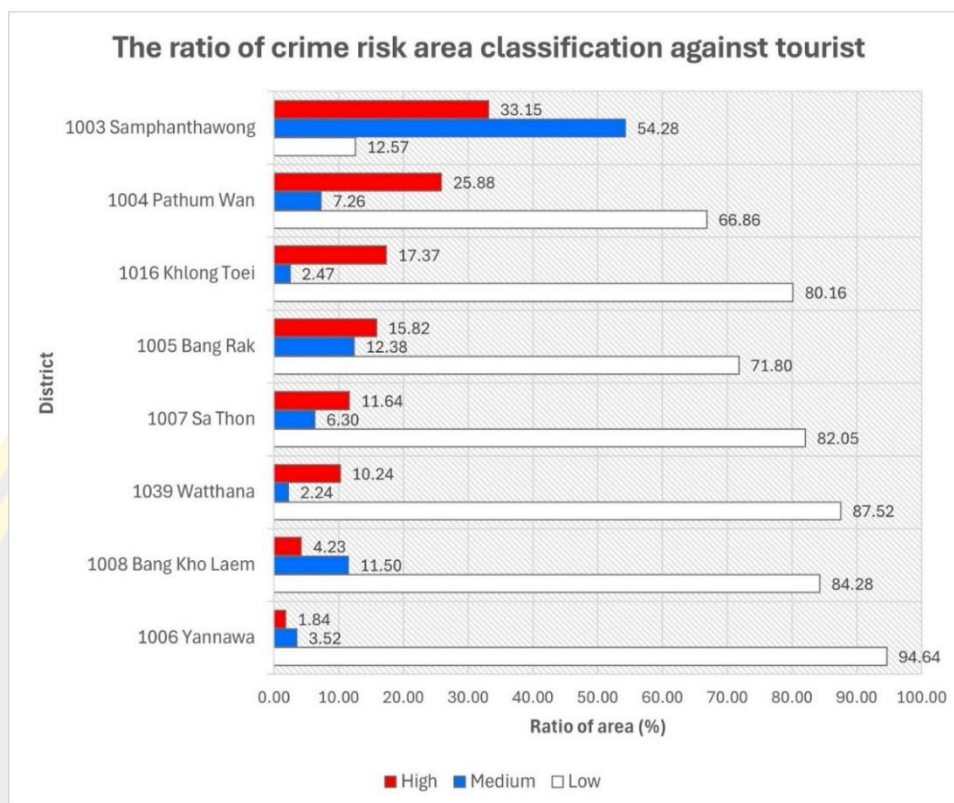


Figure 50 A classification risky area ratio of districts where crime statistics against tourists.

The ratio of high and medium crime risk areas across districts where crime statistics against tourists are present in Bangkok to the total area of each risk class is shown in Table 11. Klong Toei has the highest ratio of high crime-risk level, followed by Pathum Wan, Watthana, Sa Thon, Bang Rak, Samphanthawong, Bang Kho Laem, and Yannawa has the least ratio. While Bang Kho Laem has the highest ratio of medium crime-risk level, Samphanthawong, Pathum Wan, Bang Rak, Sa Thon, Yannawa, Klong Toei, and Watthana have the lowest ratios.

Table 11 The ratio of high-medium crime risk areas across eight districts.

Code	District name	Crime-risk level	
		High (%)	Medium (%)
1016	Khlong Toei	27.87	7.58
1004	Pathum Wan	25.24	13.55
1039	Watthana	16.05	6.71
1007	Sa Thon	10.50	10.87

Code	District name	Crime-risk level	
		High (%)	Medium (%)
1005	Bang Rak	7.70	11.52
1003	Samphanthawong	5.68	17.80
1008	Bang Kho Laem	4.20	21.84
1006	Yannawa	2.76	10.12

The ratio of high-risk areas to total in Figure 51 reveal Khlong Toei (27.87%) has the highest proportion of high-risk crime areas, followed by Pathum Wan (25.24%) and Watthana (16.05%), while Sa Thon (10.50%) and Bang Rak (7.70%) show a moderately high-risk proportion. Districts such as Bang Kho Laem (21.84%) and Samphanthawong (17.80%) have the highest proportion of medium-risk crime areas to the total area. This information emphasizes this statistic ratio in line with crime statistics against tourists in 2022 to concern for public and tourist safety.

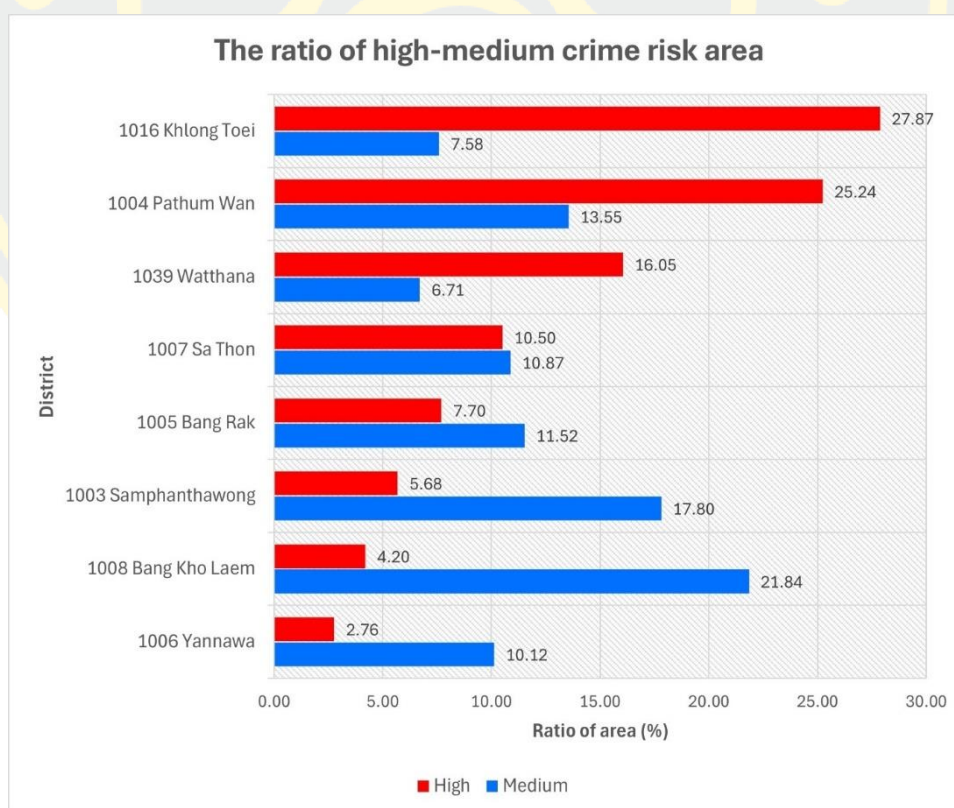


Figure 51 The ratio classification area on high- and medium-risk classes to the total area of each class in districts were crime statistics against tourists.

4.4.3 The influence of land use to crime risk area

This pie chart shown in Figure 52 represents the percentage ratio of land use within a specific study area for the year 2023. The chart breaks into five categories. Urban and built-up land (67.75%) is the largest, shown in brown, and represents most of the area being used for urban development and infrastructure. Agricultural land (23.34%) represented in green is the second-largest land use category, revealing a portion of the land is still utilized for farming and agriculture. Miscellaneous land (5.81%) represented in dark blue indicated the cover of various other forms of land use. Water bodies (2.95%), such as lakes, rivers, or reservoirs within the study area, are represented in pink. The forest land, the smallest portion accounting for 0.16% of the total land use, is represented in yellow.

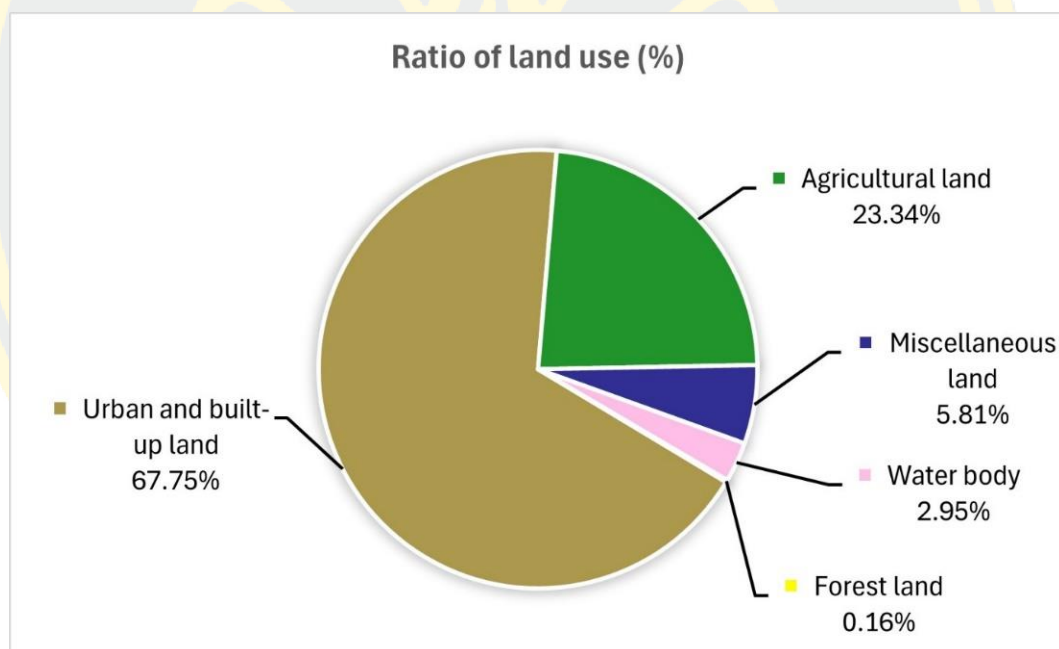


Figure 52 Percentage ratio of land use within the study area in the year 2023.

The map in Figure 53 represents the spatial distribution of crime risk in relation to land use patterns within Bangkok. High-risk areas are often located in densely urbanized regions, while rural and agricultural zones have lower crime risk levels. This spatial correlation between land use and crime risk is very useful to urban planners and policymakers in developing targeted crime prevention strategies. The overlay of land use and risk data emphasizes the importance of urban planning in mitigating potential risk factors in developed areas.

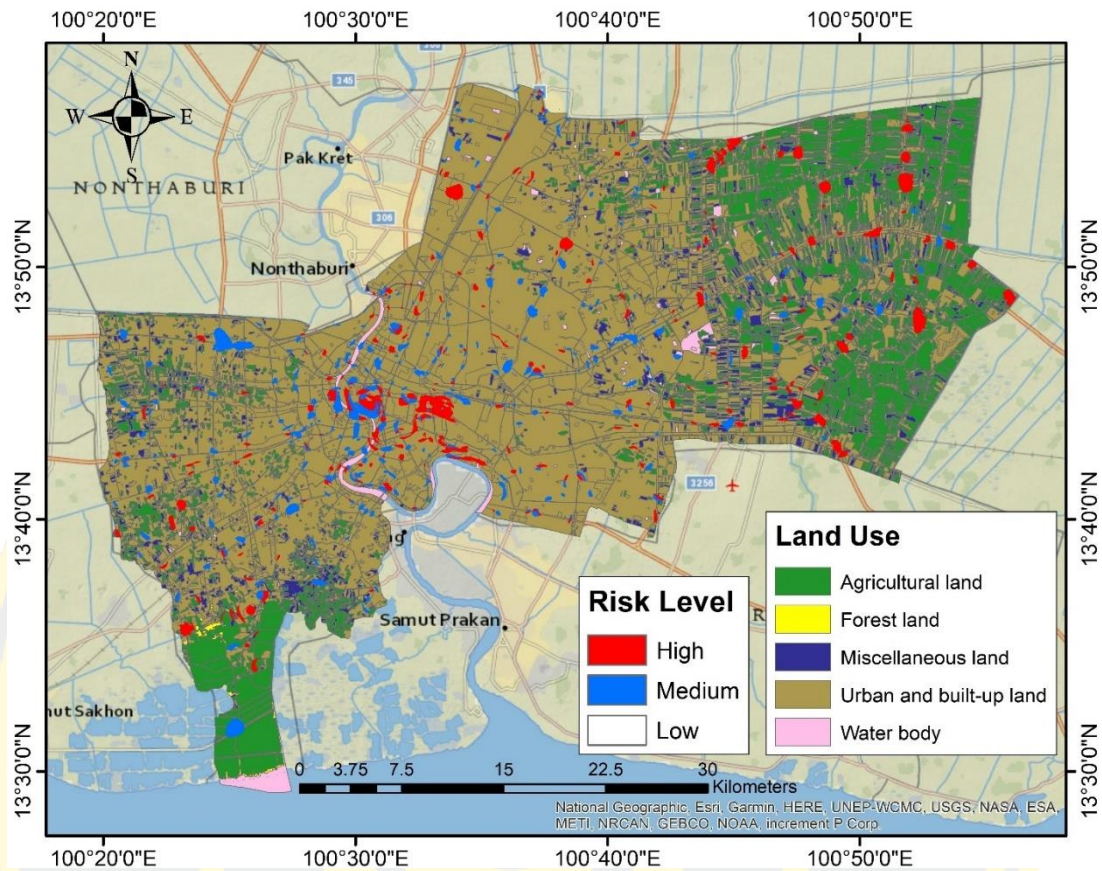


Figure 53 A classification map shows Bangkok's risky area overlaid on land use.

The chart shown in Figure 54 represents crime risk across five distinct land use types: urban and built-up land, water bodies, miscellaneous land, forest land, and agricultural land. The analysis reveals that most areas within each land use category are classified as low-risk, accounting for over ninety percent of the total area shown in Table 12. However, despite the relatively small proportions of high- and medium-risk areas, these zones warrant significant attention, particularly because Bangkok is the capital city of Thailand and the primary gateway to the country. Understanding the spatial distribution of crime risk within these higher-risk areas is important for informed urban planning and public safety strategies in this key metropolitan region.

Table 12 The ratio of crime risk areas to each land use class.

Land use type	Crime-risk level		
	High (%)	Medium (%)	Low (%)
Urban and built-up land	2.99	3.07	93.94
Water body	2.66	4.42	92.92

Land use type	Crime-risk level		
	High (%)	Medium (%)	Low (%)
Miscellaneous land	2.39	1.72	95.89
Forest land	2.02	0.71	97.27
Agricultural land	1.56	0.81	97.63

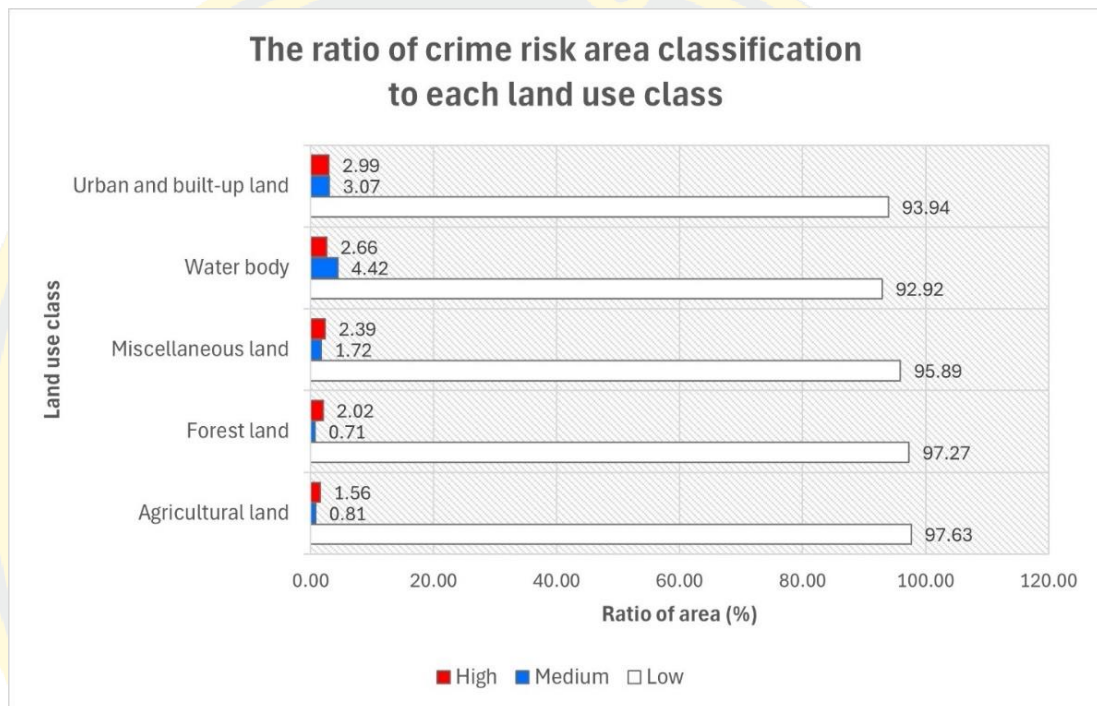


Figure 54 The ratio of crime risk area classification to each land use class.

The chart shown in Figure 55 represents urban and built-up areas, which have the highest proportion, accounting for 77.89% of the total of high-crime risk areas. This indicates a strong relationship between urbanization, population density, and increased crime risk. Agricultural land and miscellaneous land use, contributing a smaller portion, account for 13.96% and 5.33%, respectively. While the proportion of forest land at 0.12% and water bodies at 2.70% indicated these more natural, less densely populated areas are related to reduced crime rates.

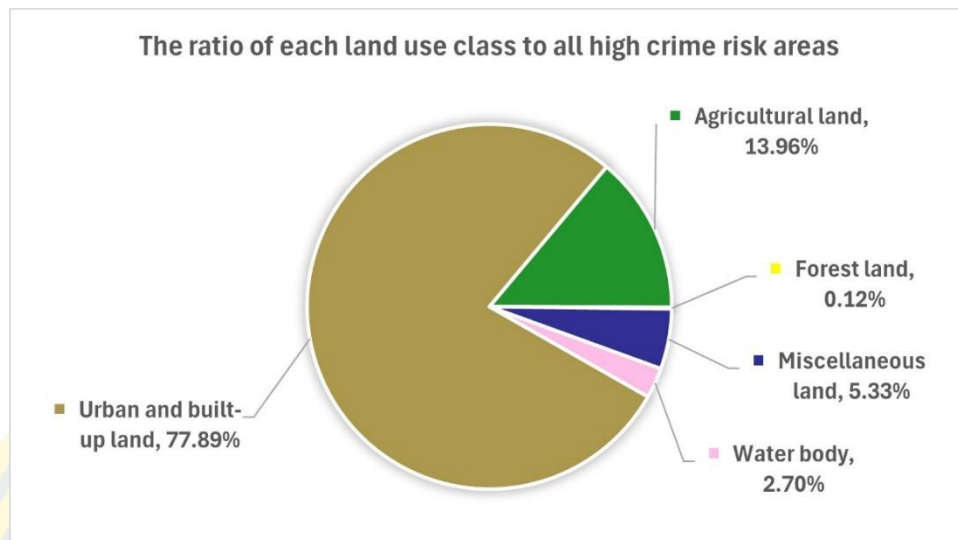


Figure 55 The ratio of each land use class to all high crime risk areas.

The chart shown in Figure 56 represents a ratio of medium crime risk across different land use classes. Urban and built-up areas account for the largest portion of medium-risk, 83.67%, in line with a proportion of high-risk. Agricultural land follows as the second largest, at 7.59%, while water bodies and miscellaneous land contribute 4.68% and 4.02%, respectively. Forest land remains largely unaffected, contributing only 0.04%.

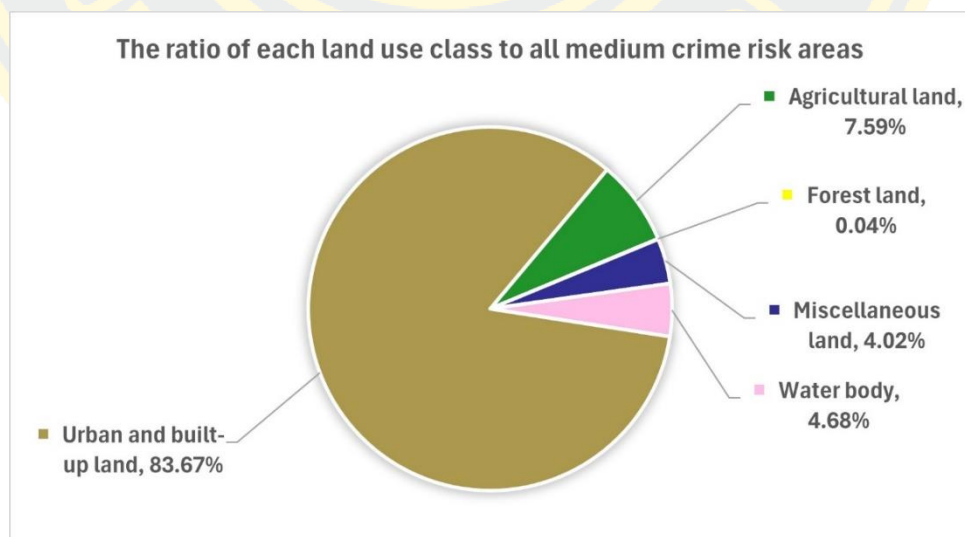


Figure 56 The ratio of each land use class to all medium crime risk area.

4.4.4 The influence of nighttime light to crime risk area

The map in Figure 57 shows the relationship between nighttime activity,

as indicated by nighttime light intensity, and crime risk levels. Areas with higher light intensity are related to increased crime risk, likely driven by high human activity and urbanization. On the other hand, poorly lit areas may not necessarily experience a decrease in crime; rather, poor lighting may increase opportunities for criminal activity due to a lack of surveillance. Furthermore, areas with proper lighting management can effectively prevent crime [120, 121]. This classification provides insights into the spatial distribution of crime risk across Bangkok that is related to nighttime light.

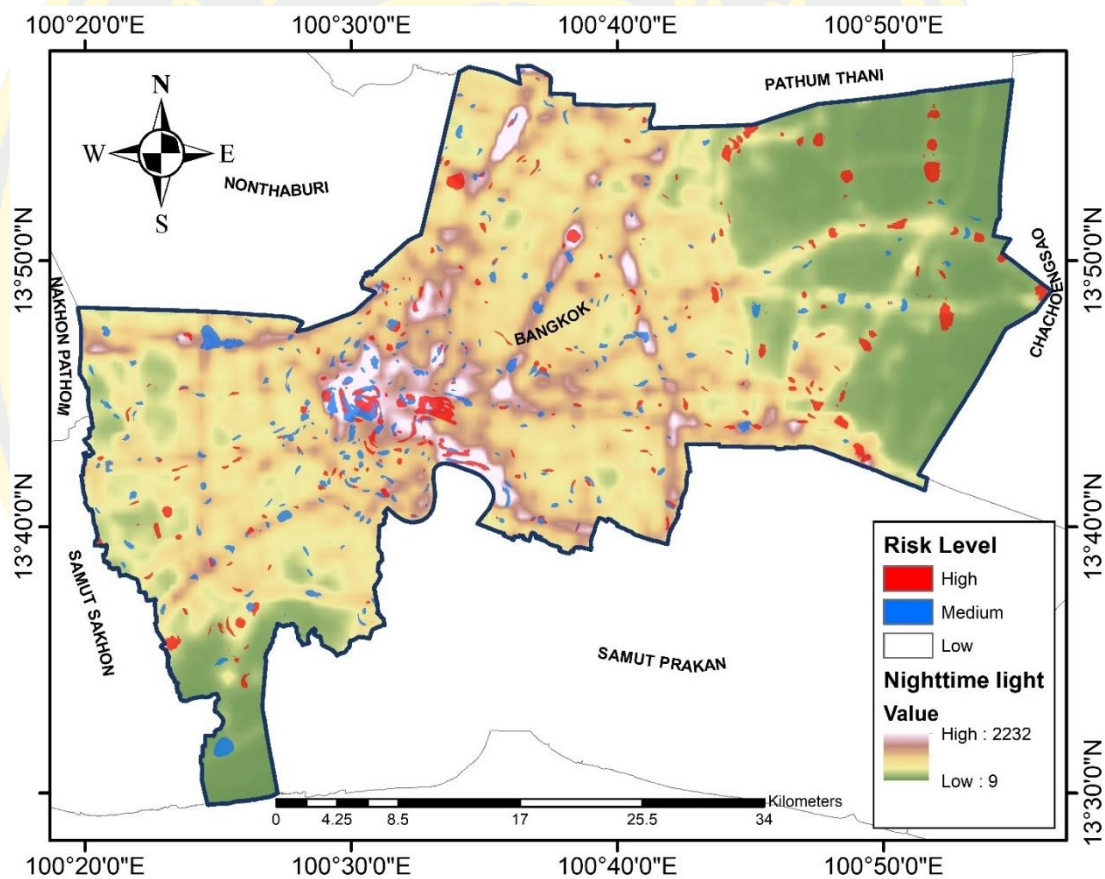


Figure 57 The classification of crime risk area over nighttime light.

4.4.5 The influence of population density to crime risk area

The map in Figure 58 shows population density, sorted from low (green) to high (dark green), revealing the influence of population concentration on crime risk. Areas with higher population density, especially in central Bangkok, tend to increase crime risks, while areas with lower population density show decreased vulnerability.

In Bangkok's densely populated environment, crime can be analyzed through routine activity theory, as criminal activity occurs when three elements converge: a motivated offender, a suitable target, and the lack of capable guardianship. High population density increases the opportunity of finding suitable targets while lacking public surveillance. Moreover, areas with high population density often experience more interpersonal interactions, which can lead to opportunities for criminal activity. Population density is also related to factors such as access to public services, income distribution, and urban planning.

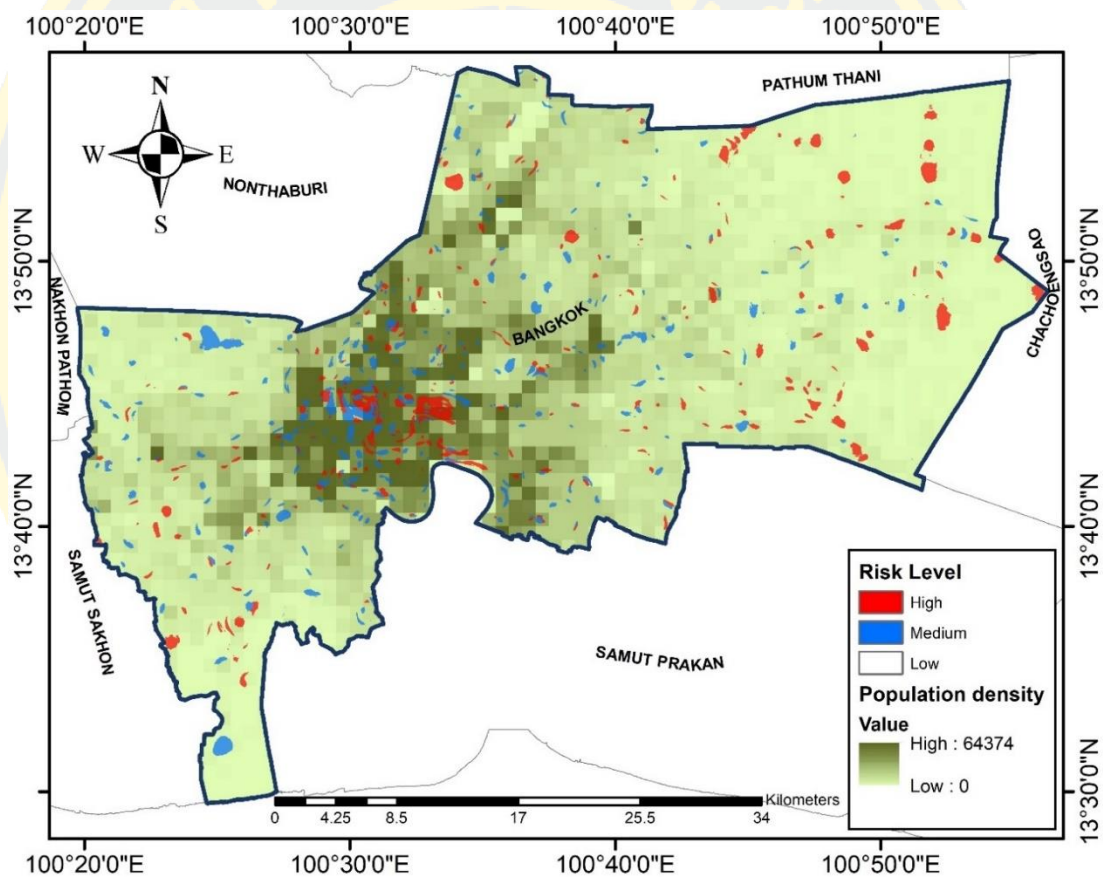
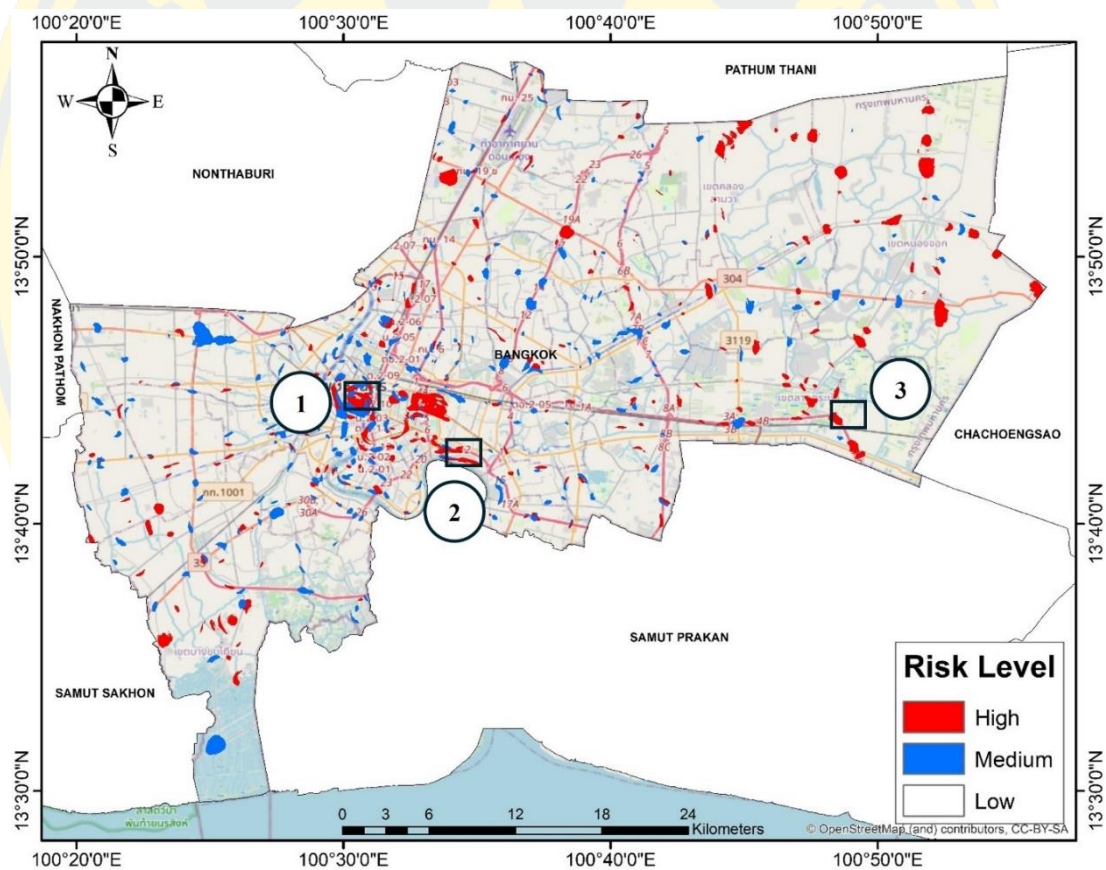


Figure 58 The classification of crime risk area over population density.

4.5 Sample of risky area in Bangkok from the best model

Google Street View (GSV) is a geospatial technology in Google Maps and Google Earth that can observe and characterize geographical landscapes. It is a useful tool to investigate urban environments, helping to virtually and remotely audit environmental features. and importance of studying the relationship between crime

and environmental factors. The GSV provides an alternative way to evaluate crime [58]. In this study, the ET model outperforms other models, with a test accuracy of 97.53%. The classification map shows that high-risk accounts for 2.59% of the Bangkok area. Exploration of the risky area from class prediction with GSV, as shown in some examples in Figure 59. The result from GSV represents an overview of the area and a specific area. These areas are characterized by dilapidated environments, conspiracy, and crowded communities. The road is narrow and difficult to travel on. Including remote areas and desolate areas with no light. These factors all contribute to the occurrence of crime.



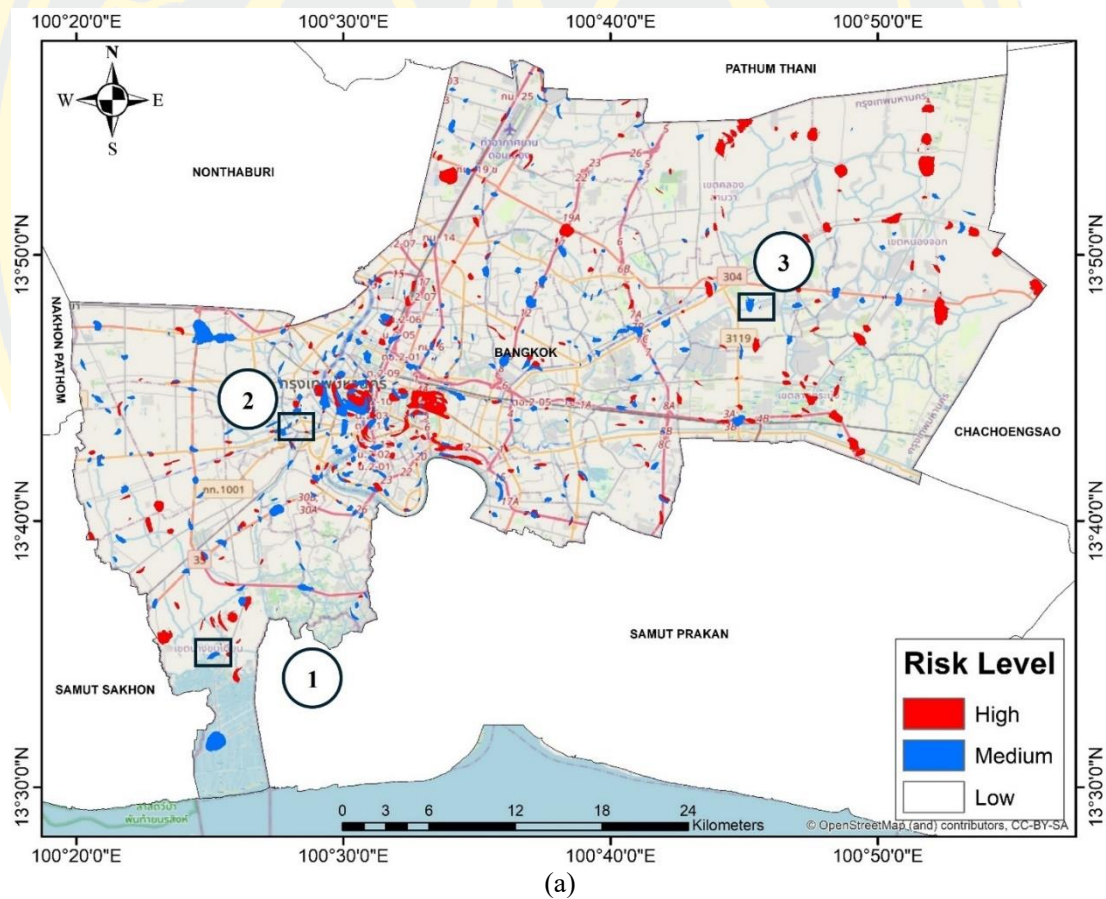
(a)



Figure 59 Example areas a) a high-risk class in Bangkok and b) a sample of a risky area.

The classification map reveals that medium-risk areas comprise 2.52% of the total Bangkok area. This assessment is based on class predictions that indicate areas with moderate levels of risk, which were further explored using Google Street View (GSV) imagery, as shown in several examples in Figure 60. The GSV analysis provides both an overview of the entire medium-risk area and a closer look at specific

locations within these areas. Then the images, it is evident that the medium-risk zones are characterized by dilapidated environments which may include poorly maintained infrastructure, abandoned buildings, or rundown neighborhoods. These areas tend to have crowded communities, which can increase vulnerability to crime due to the higher likelihood of social tension and conflicts in densely populated regions. In addition to these characteristics, remote and desolate areas are also identified within the medium-risk classification. These areas are often marked by lack of lighting, which significantly contributes to the occurrence of crime. Poor visibility and reduced surveillance in poorly lit or isolated areas make them prime locations for criminal activity, as offenders are less likely to be detected.



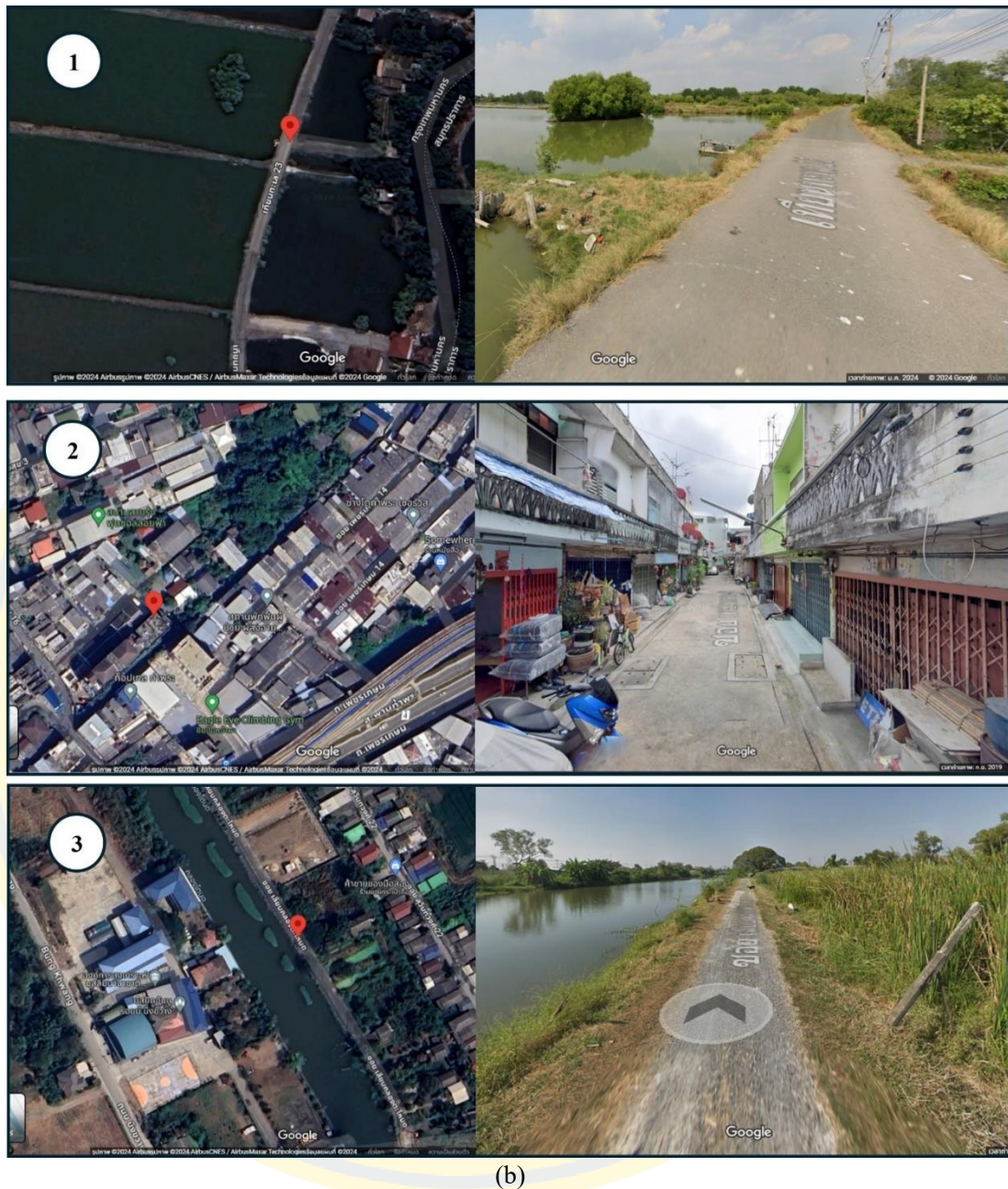


Figure 60 Example area a) a medium-risk class in Bangkok and b) A sample of a risky area.

4.6 A comprehensive map along Bangkok's metro lines

The Metro system in Bangkok has 10 lines with 150 stations available for service. The layer of Metro network and risky areas was overlayed on the map of Bangkok in ArcMap, as shown in Figure 61.

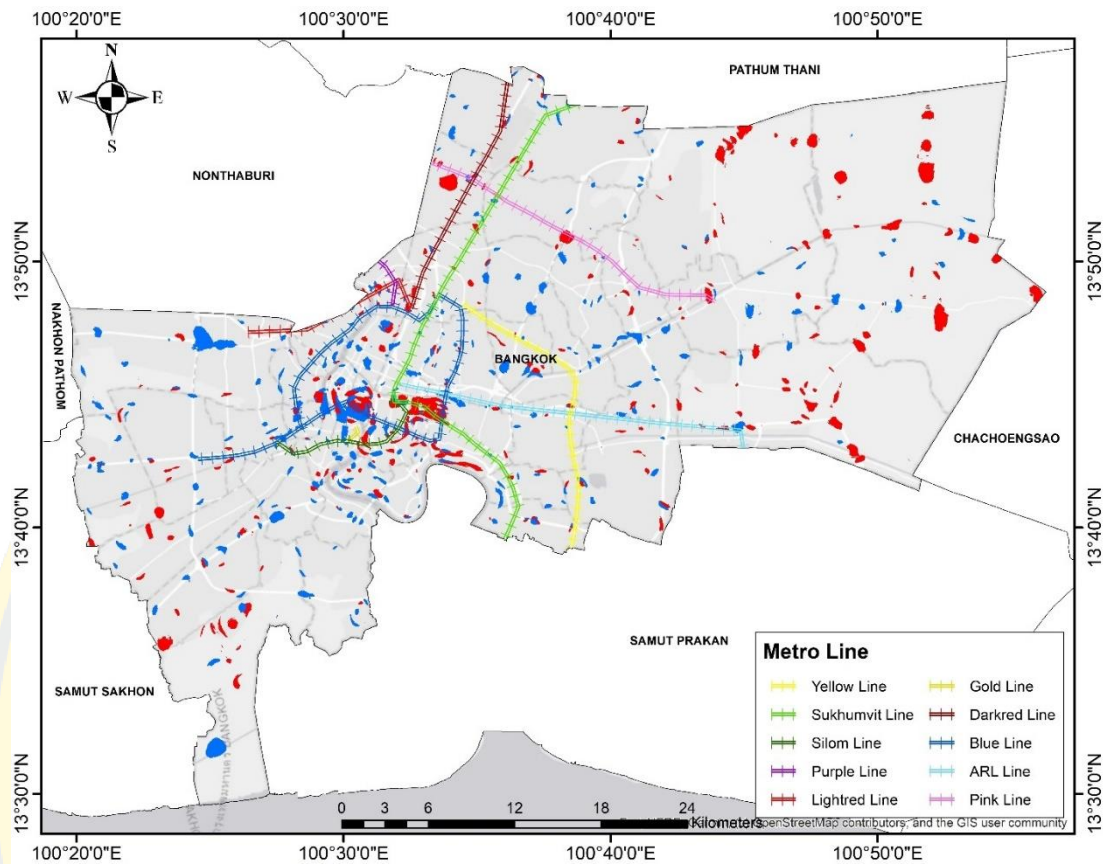


Figure 61 Metro system and risky area in Bangkok.

The Metro stations in Bangkok are approximately 150 meters in length. To cover the stations, this research uses a buffer of 100 meters to define the area of each station to check the risk level from the prediction class of the ET model using the function zonal statistics as a table in ArcMap to find the risk level pixels value within the radius of the station area. We determine each station's risk level by assigning a priority of high, medium, and low risk by considering the risk classes (1=high, 2=medium, 3=low) found in the buffer area of each station. Figure 62 presents the results of the station risk level analysis. The analysis revealed that 34 stations were in high-risk areas accounting for 22.67%, 9 in medium-risk areas accounting for 6%, and 107 in low-risk areas accounting for 71.33%.

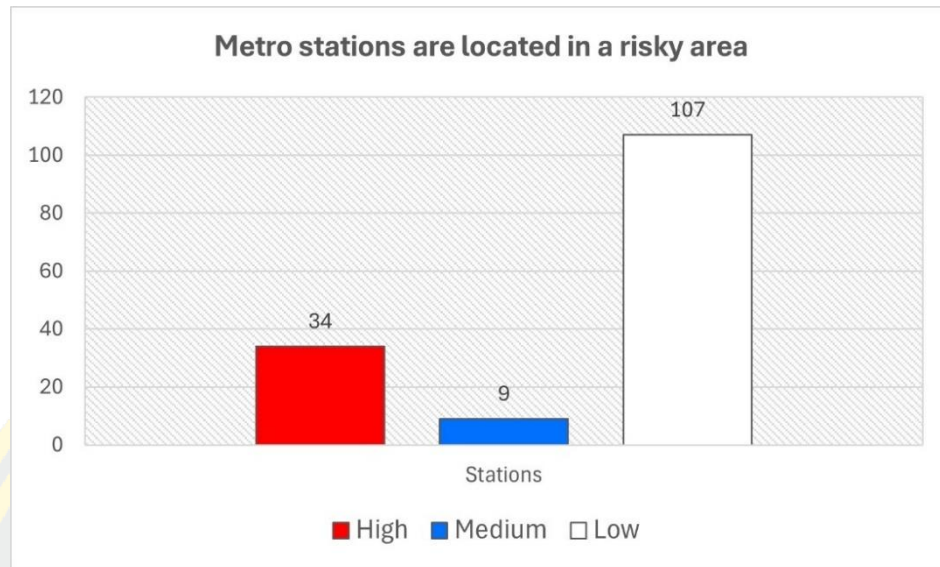


Figure 62 Statistic of metro stations are in a risky area.

Crime is an important problem that directly affects the safety of life and property. The ongoing impact on emotions, anxiety, and uncertainty in life. The physical characteristics of the city are one factor that influences urban crime. The social disorganization theory, which explains this, refers to the social conditions of communities that undergo constant changes due to the growth of both the community and its residents, as well as urbanization. This causes a wide circulation of people in the area. This leads to the weakening of community social control mechanisms and an increase in urban crime. Group conflict theory shows that the zone in transition, a zone supporting business and industrial growth that extends from the central business district, is the city's highest crime scene [122]. The statistics on crime in Bangkok reveal that the inner-city area is the most crime-ridden area in the city. The study's findings indicate that 34 Metro stations are in high-risk areas, while 9 Metro stations are situated in medium-risk areas, as shown in Figure 63, both of which are primarily located in economic centers. The stations surrounded by a crime-risk area are denoted as red, indicating a high-risk level, blue, indicating a medium-risk level, and white is at low risk. Furthermore, the study identifies 107 stations that are not categorized as high or medium risk. These stations are in areas that have low crime risk or in areas in which crime risk is not an important factor. These findings suggest that while certain Metro stations in economic centers face crime risks, a substantial portion of stations are in safer or lower-risk areas.

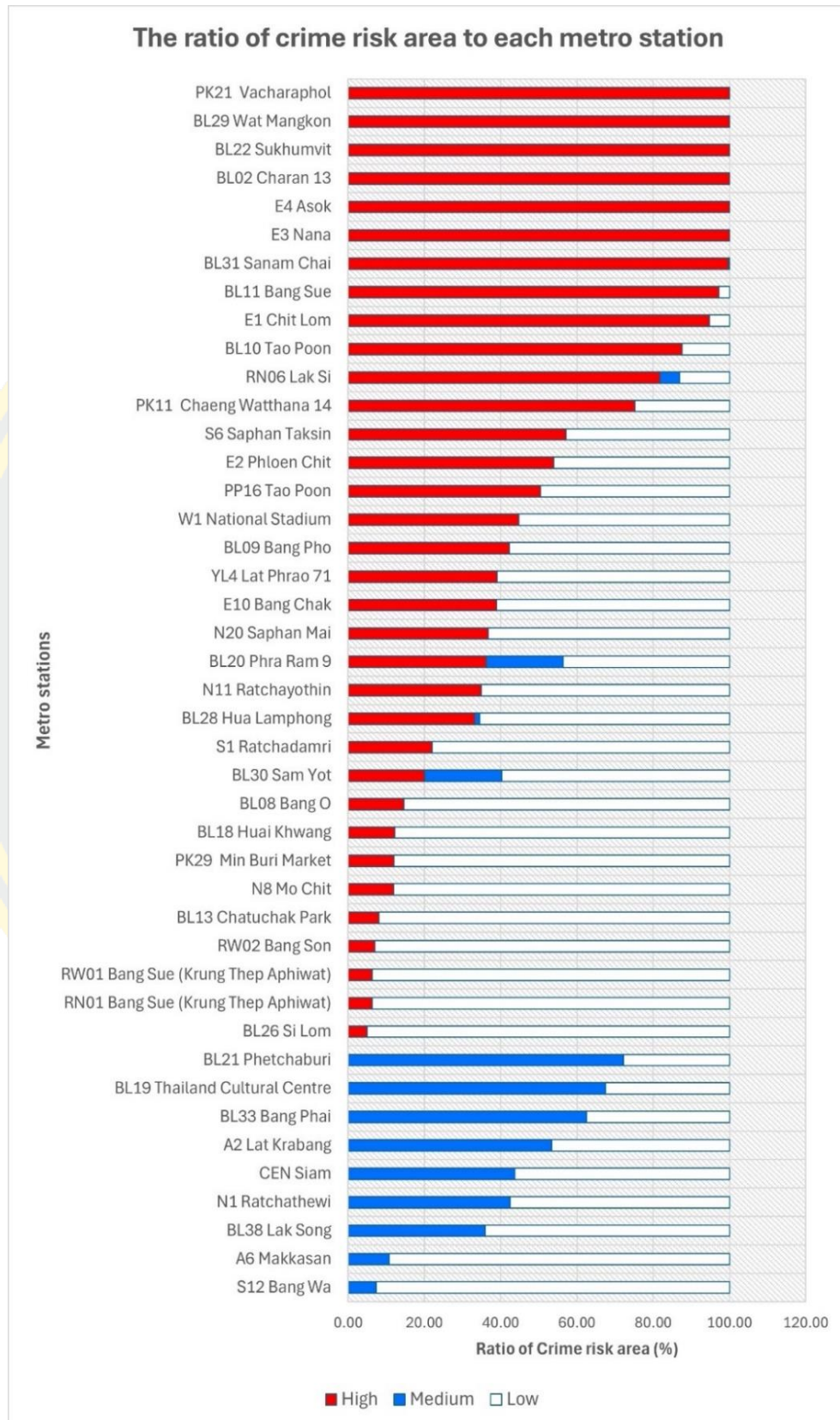
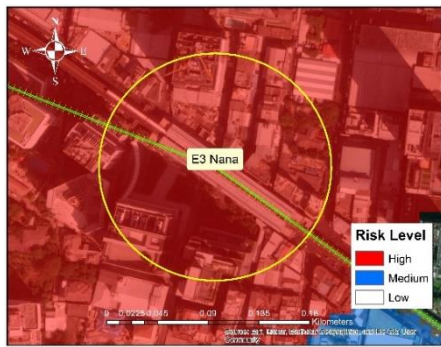


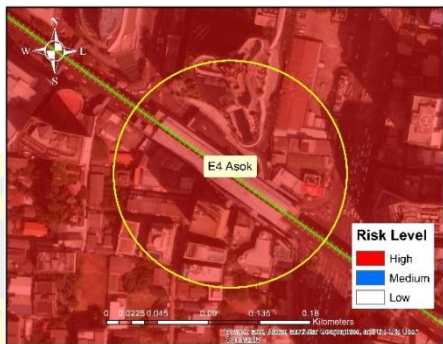
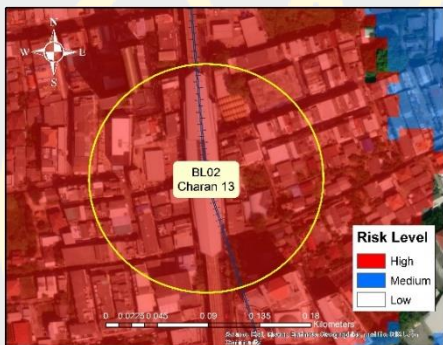
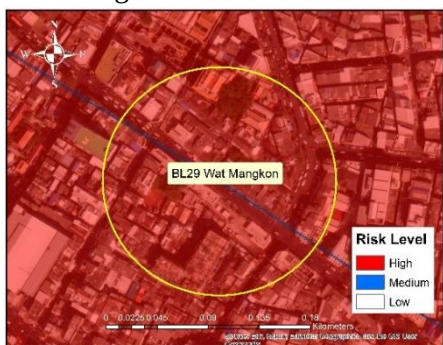
Figure 63 The ratio of crime risk areas to each metro station.

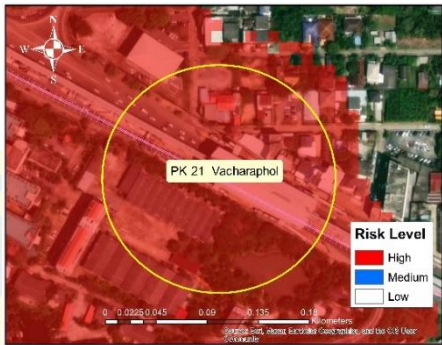
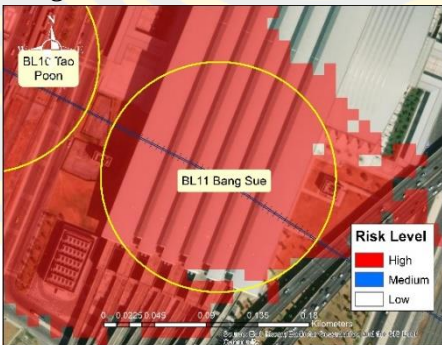
Table 13 presents metro stations and the ratios of crime risk levels labelled as high, medium, and low. Stations classified under the High Crime Risk category, where the high percentage ratio reaches 100%, indicate that all recorded incidents fall within the high-risk category. This analysis perceives Nana (E3), Asok (E4), Charan 13 (BL02), Sukhumvit (BL22), Wat Mangkon (BL29), and Vacharaphol (PK21) as stations with increased crime risks. As the proportion of high-risk incidents decreases, stations that have more balanced distributions between risk levels reflect medium and low crime risk. For instance, Bang O (BL08), where only 14.60% of incidents fall under the high-risk category and 85.40% are classified as low risk, is considered a relatively safer station.

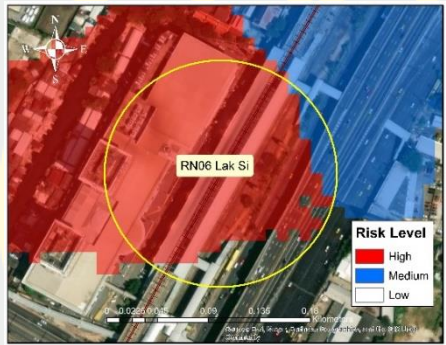
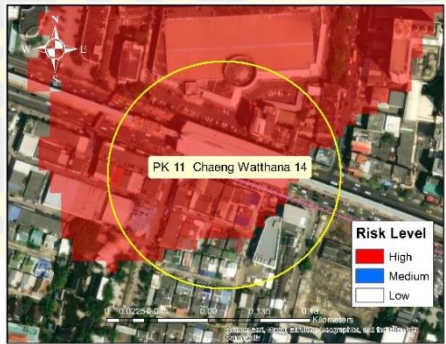
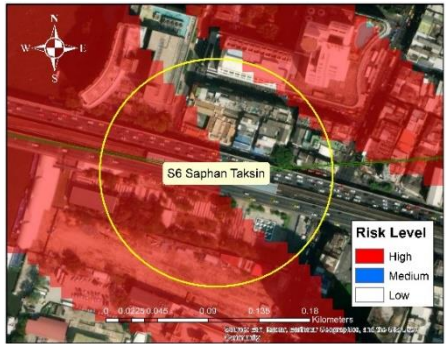
The stations display mixed risk levels across high, medium, and low categories. For instance, 36.22% of incidents at Phra Ram 9 (BL20) station are classified as high risk and 43.65% as low risk, indicating a diverse risk. Similarly, Sam Yot (BL30) also shows variability in crime risk distribution. Stations categorized as low crime risk show low high-risk percentages and high low-risk percentages, indicating lower overall crime rates. Examples include Bang Sue (Krung Thep Aphiwat) (RN01 & RW01), Bang Son (RW02), and Bang Wa (S12), where low-risk incidents comprise over 90% of the total. The stations such as Phetchaburi (BL21), Thailand Cultural Centre (BL19), and Lat Krabang (A2) report no high-risk incidents, further indicating areas with minimal crime exposure.

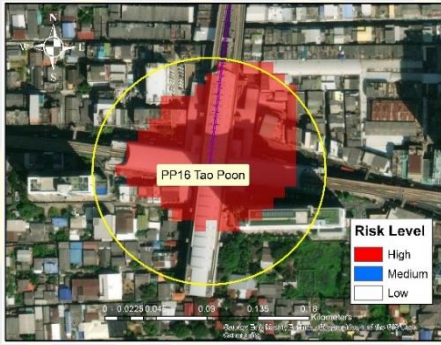
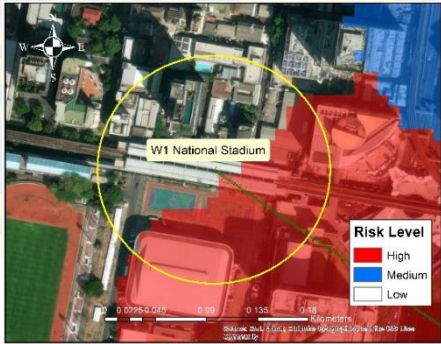
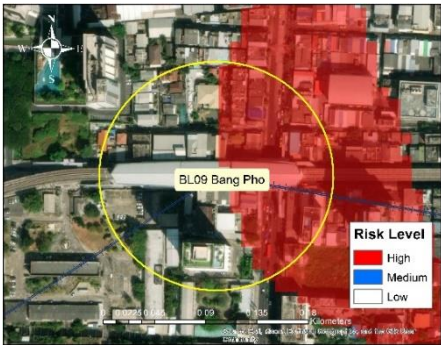
Table 13 Ratio of crime risk level to metro station.

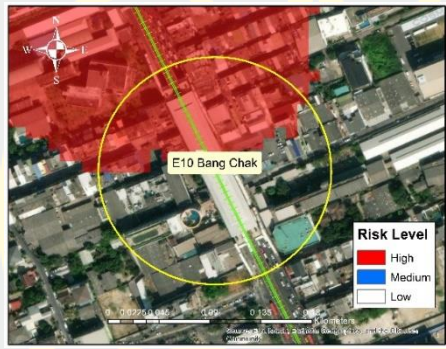
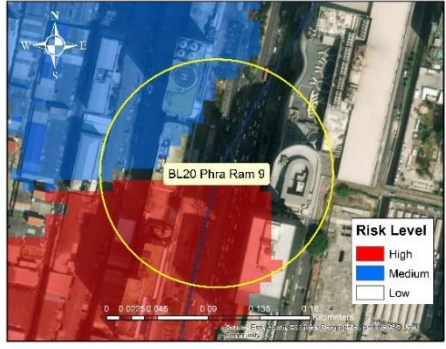
No.	Code	Station name	Ratio of risk level (%)		
			High (%)	Medium (%)	Low (%)
1	E3	Nana 	100.00	-	-

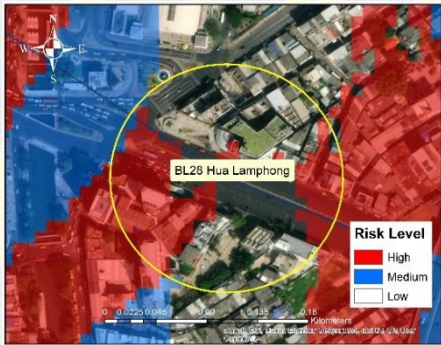
No.	Code	Station name	Ratio of risk level (%)		
			High (%)	Medium (%)	Low (%)
2	E4	Asok	100.00	-	-
					
3	BL02	Charan 13	100.00	-	-
					
4	BL22	Sukhumvit	100.00	-	-
					
5	BL29	Wat Mangkon	100.00	-	-
					

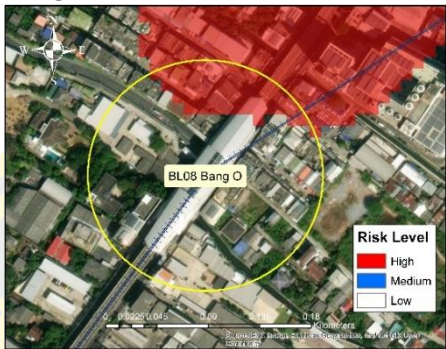
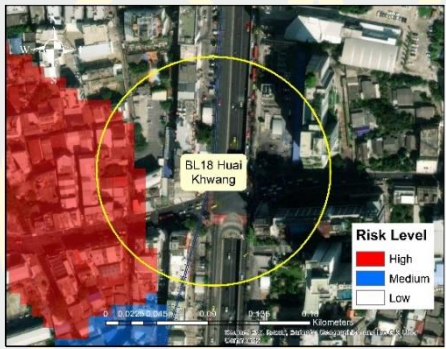
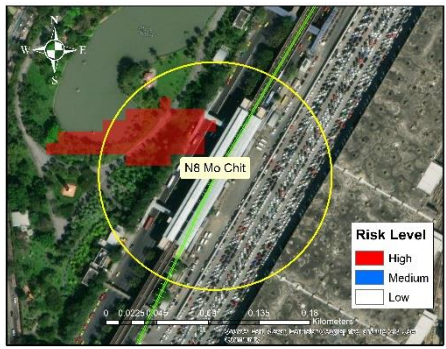
No.	Code	Station name	Ratio of risk level (%)		
			High (%)	Medium (%)	Low (%)
6	PK21	Vacharaphol	100.00	-	-
					
7	BL31	Sanam Chai	99.69	-	0.31
					
8	BL11	Bang Sue	97.23	-	2.77
					
9	E1	Chit Lom	94.69	-	5.31
					

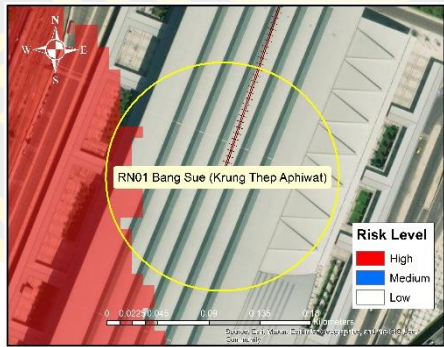
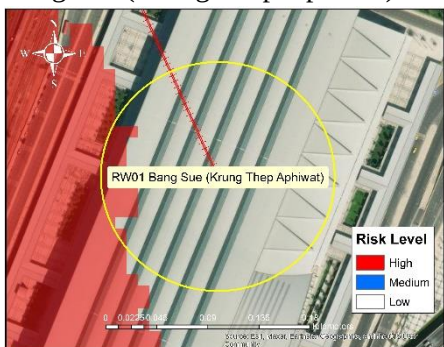
No.	Code	Station name	Ratio of risk level (%)		
			High (%)	Medium (%)	Low (%)
10	BL10	Tao Poon	87.50	-	12.50
					
11	RN06	Lak Si	81.71	5.18	13.11
					
12	PK11	Chaeng Watthana 14	75.15	-	24.85
					
13	S6	Saphan Taksin	57.10	-	42.90
					

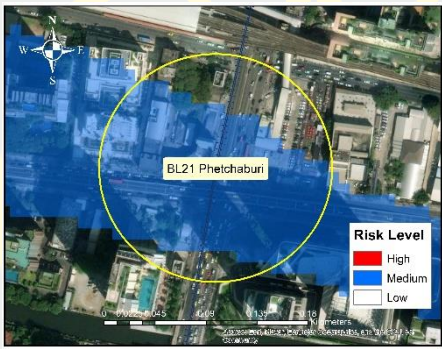
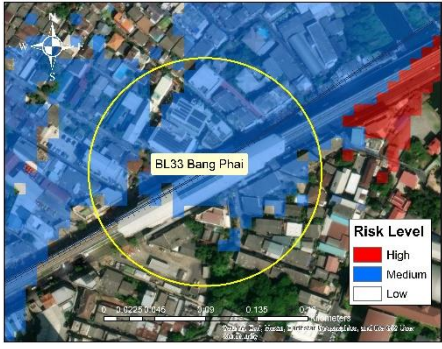
No.	Code	Station name	Ratio of risk level (%)		
			High (%)	Medium (%)	Low (%)
14	E2	Phloen Chit	53.89	-	46.11
					
15	PP16	Tao Poon	50.46	-	49.54
					
16	W1	National Stadium	44.79	-	55.21
					
17	BL09	Bang Pho	42.28	-	57.72
					

No.	Code	Station name	Ratio of risk level (%)		
			High (%)	Medium (%)	Low (%)
18	YL4	Lat Phrao 71	39.06	-	60.94
					
19	E10	Bang Chak	38.89	-	61.11
					
20	N20	Saphan Mai	36.78	-	63.22
					
21	BL20	Phra Ram 9	36.22	20.12	43.65
					

No.	Code	Station name	Ratio of risk level (%)		
			High (%)	Medium (%)	Low (%)
22	N11	Ratchayothin	34.86	-	65.14
					
23	BL28	Hua Lamphong	33.23	1.24	65.53
					
24	S1	Ratchadamri	21.95	-	78.05
					
25	BL30	Sam Yot	20.00	20.31	59.69
					

No.	Code	Station name	Ratio of risk level (%)		
			High (%)	Medium (%)	Low (%)
26	BL08	Bang O	14.60	-	85.40
					
27	BL18	Huai Khwang	12.27	-	87.73
					
28	PK29	Min Buri Market	11.96	-	88.04
					
29	N8	Mo Chit	11.93	-	88.07
					

No.	Code	Station name	Ratio of risk level (%)		
			High (%)	Medium (%)	Low (%)
30	BL13	Chatuchak Park	8.06	-	91.94
					
31	RW02	Bang Son	7.03	-	92.97
					
32	RN01	Bang Sue (Krung Thep Aphiwat)	6.27	-	93.73
					
33	RW01	Bang Sue (Krung Thep Aphiwat)	6.27	-	93.73
					

No.	Code	Station name	Ratio of risk level (%)		
			High (%)	Medium (%)	Low (%)
34	BL26	Si Lom	4.92	-	95.08
					
35	BL21	Phetchaburi	-	72.14	27.86
					
36	BL19	Thailand Cultural Centre	-	67.48	32.52
					
37	BL33	Bang Phai	-	62.46	37.54
					

No.	Code	Station name	Ratio of risk level (%)		
			High (%)	Medium (%)	Low (%)
38	A2	Lat Krabang	-	53.35	46.65
					
39	CEN	Siam	-	43.65	56.35
					
40	N1	Ratchathewi	-	42.41	57.59
					
41	BL38	Lak Song	-	35.91	64.09
					

No.	Code	Station name	Ratio of risk level (%)		
			High (%)	Medium (%)	Low (%)
42	A6	Makkasan	-	10.74	89.26
					
43	S12	Bang Wa	-	7.34	92.66
					

The map in Figure 64 shows the locations of metro stations located in the area with crime against tourists. The yellow circles represent the locations of 25 metro stations across eight districts, which are important transit points for both residents and tourists. The analysis found 11 stations found high crime risk level in radius 100 meter. The stations consist of Wat Mangkon (BL29), Sukhumvit (BL22), Asok (E4), Nana (E3), Chit Lom (E1), Saphan Taksin (S6), Phloen Chit (E2), National Stadium (W1), Hua Lamphong (BL28), Ratchadamri (S1), Si Lom (BL26), and CEN. Siam is a station located in a medium crime risk level. Although these areas are economic and tourism hotspots, they also have safety issues. Bangkok shows a moderate crime rate, characterized by pervasive non-confrontational robberies like pickpocketing, particularly in places visited by tourists. The metro stations are important for Bangkok's economic and tourism attraction. However, their importance requires enhanced security measures to ensure the safety of both residents and visitors.

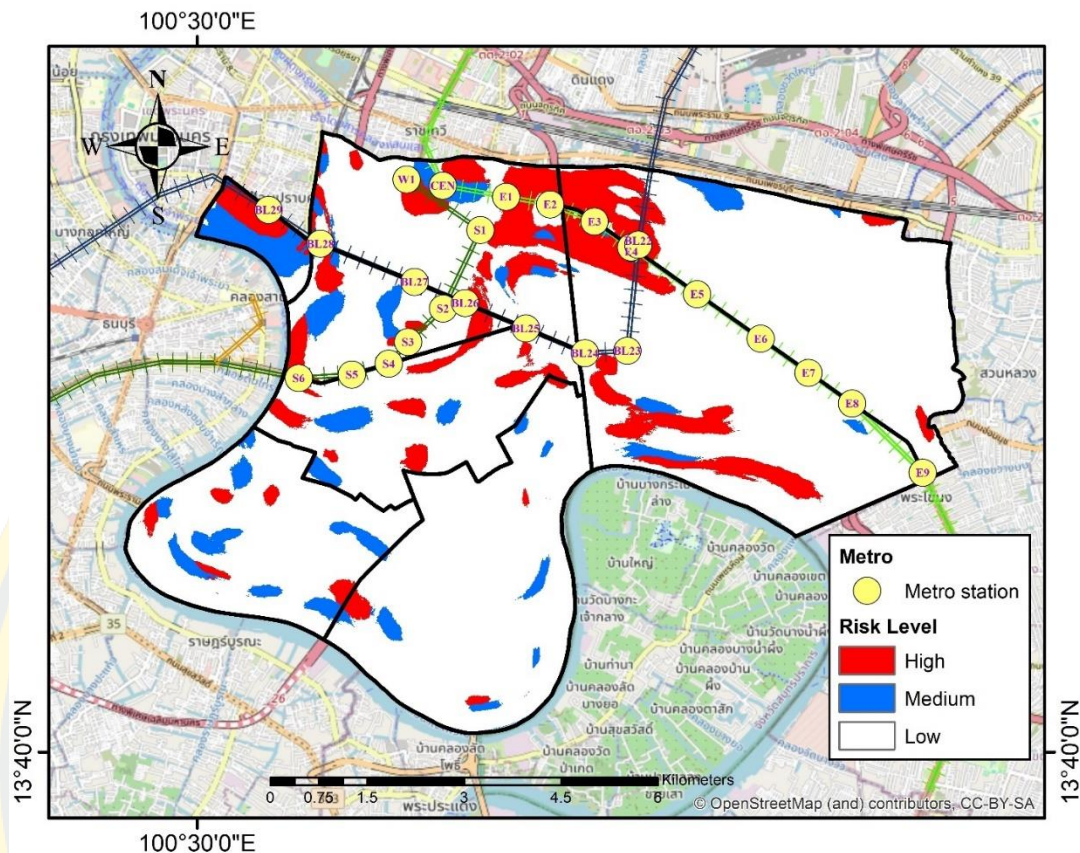


Figure 64 A map of metro stations located in Bangkok's districts where crime statistics against tourists are present.

The comprehensive map in Figure 65 shows the utilize classification crime risk area from ET model and Bangkok Metropolitan Rapid Network [125] revealed the metro stations located in crime-risk areas, with 34 stations found high-risk represented in the red circle station and 9 stations found medium-risk represented in the blue circle station. This information can be provided to policymakers or authorities for improving law enforcement or security measures at stations identified as high-risk to reduce the opportunity for criminal activity. For residents who often use the metro system, this map is an important tool to raise awareness of safety concerns. Understanding which stations are related to higher crime risks may influence commuting, such as adjusting travel times or choosing alternative stations with lower risk levels. Tourists who are unfamiliar with the city can use the map as a safety guide. Visitors can make more informed decisions about which metro stations to avoid or prepare themselves for dealing with unexpected situations when they must visit risky areas, especially during late hours or when travelling alone.

4.7 Provide information to the policymaker

108

(CCTV), in these high-risk areas allows for real-time crime monitoring and supports post-event forensic investigation. In addition, improvements in urban design that include crime prevention strategies such as enhancing lighting in poorly lit zones, ensuring clear sightlines, and reducing physical barriers can reduce the opportunity for criminal activity. The findings emphasize the need for urban design initiatives that integrate accessibility and security. Implementing pedestrian-only zones in places with high commercial or recreational activity, along with centralized security measures, can reduce vulnerabilities while maintaining urban operation. When implemented carefully, these solutions can convert high-risk urban areas into safer, more resilient places for residents and visitors.

4.8 Summary of Experiment and Result

In conclusion, the classification of crime risk area across Bangkok by using three machine learning models (Extra Tree, Random Forest, and LightGBM) with integration of three public data sources of crime risky points 2022 provided by government agencies (MPB and BMA) and multifactor consider cover human activity, socio-economic, and infrastructure total 14 variables consist of POI from open data, greenspace (NDVI) from sentinel-2 MSI, demographic (population density) from the LandScan project, and nighttime light from SNPP. The purpose of this study is to analyze and investigate the area related to crime-prone places. The study focuses on both the overall district and specific areas, such as the eight districts that have crime statistics against tourists and utilizes classified data to construct a comprehensive map of the mass rapid transit system. The findings from this study are summarized as follows.

Crime statistics in Bangkok, especially crime against persons and property crime, reflect a problem in a city that has an impact on society and the urban environment. According to a report of the national statistical office (NSO), crime statistics in Bangkok 2022 show crime against persons has 2947 reported, revealing assault as the most common incident was recorded in 1,759 cases. Property crimes have 10,350 reported cases, revealing theft is the most frequently recorded property crime in Bangkok, with 3,365 cases. However, the spatial distribution reflects crime patterns throughout Bangkok, with areas such as district 1015-Phra Khanong, 1016-

Khlong Toei, 1018-Suan Luang, 1019-Bang Khen, 1023-Lat Phrao, 1039-Watthana, and 1042-Sai Mai showing the highest crime rates (217-319 cases), while others, such as 1025-Nong Chok, have lower crime rates (18-75 cases). These statistics highlight the need to improve urban safety and reduce crime rates in Bangkok. Additionally, the tourist criminal cases found in eight districts of Bangkok in 2022 indicate that Pathum Wan (1004), Watthana (1039), Sa Thon (1007), and Khlong Toei (1016) are the highest risk zones, which indicates these areas attract a considerable influx of people and may serve as hotspots for crime.

The processing of three machine learning models (Extra Tree, Random Forest, and LightGBM) revealed the ET model outperforms others. High-density Points of Interest (POI) areas, such as business, corporate, industrial, retail, catering, tourist attraction, and recreational centers, tend to attract more human traffic, which increases opportunities for criminal activity. This is due to the presence of potential targets in crowded environments and a lack of awareness among visitors, especially in the sensitive areas that are related to routine activity theory. The terms of influence factors investigated by feature importance are an important function in classification models to measure the performance of each factor. The results show the thirteen out of fourteen factors that influence models' accuracy with no difference from each other and found Business and Industrial (P8), Administration (P5), and Catering and Recreation (P2) are the most three factors that influence model accuracy. While NDVI (P13) has the least impact on the model. High population density correlates with increased opportunities for property crimes, in line with traditional criminological theories like routine activity theory, which suggest that increased human presence can both increase and decrease criminal activity, depending on the specific context and type of area.

The relation between green spaces and crime indicates that green spaces in urban areas that are well-designed and well-maintained would reduce property and violent crime rates, but poorly maintained green spaces could attract criminal activities. In areas with high levels of resource deprivation and physical disorder, parks may function as criminal marketplaces rather than providing social benefits. Facilities in green spaces and the socioeconomic context of surrounding neighborhoods have influenced the frequency and timing of criminal activities.

The effects of increased lighting on crime rates are more pronounced in urban areas with higher nighttime business activity and urbanization. Street lighting can reduce nighttime crimes and arrests in public spaces, not more than daytime crimes. Indicate that community pride and informal social control may be more effective than increased surveillance and deterrence.

The classification result shows the ratio of each risk level of crime across Bangkok. The analysis reveals that most areas within each land use category are classified as low-risk, accounting for over ninety percent of the total area. However, the presence of high and medium crime risk is the minimal proportion, but these data are very sensitive to investigations that are thorough. Urban and built-up areas are the highest proportion of the total of high and medium crime risk areas. This indicates a strong relationship between urbanization, population density, and increased crime risk. Areas with a large influx of tourists often result in crimes. As the eight districts were found to have crime incidents against tourists, the analysis focusses on high- and medium-risk classes, with the result revealing the crime risk level ratio in the area aligns with crime statistics against tourists in 2022 that raised concern for public and tourist safety.

The classification result also shows the 150 metro stations across Bangkok within a radius of 100 meters reveal 34 stations found the high crime risk, and 9 stations found the medium crime risk. Additionally, of the 25 metro stations located in the area with crime against tourists across eight districts, 11 stations found a high crime risk level, and one station found a medium crime risk level.

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Conclusion

Bangkok is the primary gateway to Thailand, functioning as an important connecting point for economic, cultural, and international linkages. As the capital and largest urban center, Bangkok is the top destination for students, workers, tourists, business professionals, and foreigners annually. As such, Bangkok attracts a considerable influx of people that may cause multidimensional problems to follow. Crime is one important big issue in Bangkok, especially crime against persons and property crime that can be partially attributed to the deterioration of social structures, and rapid urbanization resulting from migration led to overpopulation. Criminal activity in Bangkok impacts multiple dimensions of urban life, public safety, tourism, and economic development. Therefore, this study offers the classification of crime risk areas across Bangkok by integrating machine learning models, urban environmental data, and satellite imagery to effectively address crime, which is important for enhancing public safety and supporting sustainable urban development in Bangkok.

Machine learning models are important in the classification of crime risk areas through the analysis of multiple data sets, including crime risk point records, human activity, demographics, and socioeconomic factors, to predict high-medium risk areas with accuracy. These models can be utilized for predictive analytics for law enforcement agencies to predict possible crime hotspots and enhance safety measures. Furthermore, they support complex data processing for identifying correlations between crime incidents and the urban environment. This study uses three machine learning models (Extra Tree, Random Forest, and LightGBM) to analyze the factors that influence crime risk area by feature importance, which revealed thirteen out of fourteen factors influence model accuracy with no difference from each other, and Business and Industrial (P8), Administration (P5), and Catering and Recreation (P2) are the most factors that influence model accuracy. While NDVI (P13) has the least impact on the model. While analysis into specific groups of factors shows demographics is highest, followed by nighttime lights, POI, and greenspace.

Demographics represent population density related to areas with higher population density, especially in central Bangkok, tend to increase crime risks, while areas with lower population density show decreased vulnerability. Additionally, areas with higher light intensity are related to increased crime risk and areas with lower light intensity have reduced crime risks.

The processing of three machine learning models reveals an ET model that outperforms others with an accuracy of 97.53%. The classification result shows the ratio of each risk level of crime across 50 districts in Bangkok that aligns with the crime statistics 2022 in terms of crime against persons and property crime. This indicates that no district across Bangkok is truly safe. The analysis also reveals that most areas within each land use category are classified as low-risk, accounting for 94.89% of the total area. However, the proportion of high and medium crime risk still represents a small percentage, accounting for 2.59% and 2.52%, respectively, that are very sensitive to investigations that are thorough. The investigation shows Urban and built-up areas are the largest proportion of land use types in the total of high and medium crime risk areas, which emphasizes the correlation between urbanization, population density, and increased crime risk.

Crime against tourists in eight districts of Bangkok, including Bang Rak (1005), Samphanthawong (1003), Yannawa (1006), Bang Kho Laem (1008), Sathon (1007), Khlong Toei (1018), Pathum Wan (1004), and Watthana (1039), has an important influence on the tourism sector in the city. These districts, which attract a high influx of visitors because of famous attractions, shopping places, and nightlife, are especially vulnerable to crimes including theft, fraud, and other safety-related incidents. The occurrence of crime in these areas can influence tourists' perceptions of security and the image of Bangkok as a desirable travel destination. The areas with high tourist activity show increased crime risks, Pathum Wan, Watthana, and Khlong Toei are the highest proportion from the total area of high- and medium-risk areas in eight districts. The analysis of high- and medium-risk areas aligns with reporting incidents against tourists in eight districts in 2022. The presence of crime rates may lead to a reduction in tourist visits, impacting local businesses and the broader tourism industry. Ensuring safety in these districts, the result from this study is important information to enhance public confidence and support sustainable tourism growth

through crime prevention and public safety measures.

The Bangkok metro system, consisting of the BTS Skytrain, MRT, and Airport Rail Link, plays an important role in improving urban traffic congestion by providing efficient and accessible mass transit. It facilitates mobility across commercial, residential, and tourist zones, thereby enhancing economic productivity through improved business connectivity. The metro system is important for urban development, spatial planning, and the development of environmental sustainability in one of Southeast Asia's most densely populated megacities, while continuing expansions extend its reach. The classification results show that, of the 150 metro stations in Bangkok within a station's 100-meter radius, 34 are in high crime risk areas and 9 in medium crime risk areas. Furthermore, out of 25 metro stations located in districts with recorded crimes against tourists, 11 stations are classified as having a high crime risk level, while 1 station falls under the medium crime risk. These findings emphasize the relationship between metro station locations and crime risks, particularly in tourist density areas.

The findings emphasize that crime risk in Bangkok is influenced by multiple factors that cover human activity, socio-economics, and infrastructure and have an impact on safety, tourism, urbanization, social issues, transportation, and economic activity. Ensuring public safety is important for urban development, sustaining tourism, and supporting Thailand's GDP. Data-driven crime assessments are helpful policymakers in enhancing public safety, promoting economic growth, and ensuring sustainable urban development.

5.2 Recommendations and Future work

Future research should extend the crime risk analysis to include areas surrounding Bangkok, focusing on nearby neighborhoods and districts that might show continuous criminal crime risks. This would allow for a more comprehensive understanding of crime patterns and their broader urban context.

Future studies should classify, analyze, and investigate types of crime in these high-risk areas. Each type of crime has unique spatial and socio-economic influences, and a more focused investigation could help enhance prevention strategies for specific crime types.

Conducting studies to examine the long-term impacts of urbanization, infrastructure improvements, and crime prevention strategies on crime rates would provide insights into the dynamic relationships between these factors. Such studies would enhance our understanding of the temporal change of these factors, providing a more comprehensive view of their effects on urban crime patterns and the effectiveness of measures.

Finally, investigating the role of smart city technologies, such as advanced surveillance systems and AI-driven predictive policing, can enhance real-time monitoring and proactive response in high-risk crime areas, particularly around metro stations. These technologies offer the potential to improve situational awareness, optimize resource allocation, and facilitate data-driven decision-making, leading to more efficient crime prevention and public safety efforts.

REFERENCES

1. Nichaphat, S.T., R.; Warittha, P.; Suchanan C.; Yotsanan, S.; Pattakarn, G. *Revitalising Thailand's Tourism Sector*. 2021 [cited 2023 15/10/2023]; Available from: www.bot.or.th/content/dam/bot/documents/th/research-and-publications/research/discussion-paper-and-policy-paper/250624_WhitepaperVISA.pdf.
2. Toker, A.E., O., *Safety and security research in tourism: A bibliometric mapping*. European Journal of Tourism Research, 2023. **34**: p. 1-25.
3. Statista. *International Tourist Arrivals by World Region 2022*. 2023 [cited 2023 11/10/2023]; Available from: <https://www.statista.com/statistics/186743/international-tourist-arrivals-worldwide-by-region-since-2010/>.
4. UNWTO. *Tourism on Track for Full Recovery as New Data Shows Strong Start to 2023*. 2023 [cited 2023 10/11/2023]; Available from: <https://www.unwto.org/news/tourism-on-track-for-full-recovery-as-new-data-shows-strong-start-to-2023>.
5. Education, M.o. *National Strategy (2018-2037)*. 2023 [cited 2023 30/11/2023]; Available from: https://www.bic.moe.go.th/images/stories/pdf/National_Strategy_Summary.pdf.
6. Forum, W.E. *Travel & Tourism Development Index 2024*. 2024 [cited 2024 01/06/2024]; Available from: <https://www.weforum.org/publications/travel-tourism-development-index-2024/>.
7. Cities, S. *Safe Cities Index 2021*. 2021 [cited 2024 03/01/2024]; Available from: <https://safecities.economist.com/>.
8. INRIX. *Scorecard City 2022 - INRIX*. 2022 [cited 2024 03/01/2024]; Available from: <https://inrix.com/scorecard-city-2022/?city=Bangkok&index=32>.
9. Pineda-Jaramillo, J.F., M.; Pira, M.L.; Giuffrida, N.; Inturri, G.; Viti, F.; Ignaccolo, M., *A sentiment analysis approach to investigate tourist satisfaction towards transport systems: the case of Mount Etna*. Transportation Research Procedia, 2023. **69**: p. 400-407.
10. Fazio, M.L.P., M.; Inturri, G.; Ignaccolo, M., *Bus Rapid Transit vs. Metro. Monitoring on-board comfort of competing transit services via sensors*. 2020. 292-297.
11. Edelman, D., *Management of the Urban Environment in Three Southeast Asian Coastal Metros: Bangkok, Manila and Jakarta*. Advances in Applied Sociology, 2023. **13**: p. 765-839.
12. Iamtrakul, P.P., A.; Klaylee, J., *Analysis of Urban Sprawl and Growth Pattern Using Geospatial Technologies in Megacity, Bangkok, Thailand*. 2022. p. 109-123.
13. Haider, M.I., P., *Article Review: Theoretical Concepts of Crime and Practices in Urban Planning and Design Process for Safe Urban Life*. 2018. **12**: p. 7-24.
14. M., M., *12 Best Travel Insurance Companies Of August 2024*. Forbes Advisor, 2024.
15. Jomdet, T., *Public confidence in CCTV and Fear of Crime in Bangkok, Thailand*. International Journal of Criminal Justice Sciences (IJCJS), 2016. **11**(1): p. 17-29.

16. Ompad, D.G., S.; Vlahov, D., *Urbanicity, Urbanization, and the Urban Environment*. Macrosocial Determinants of Population Health, 2007.
17. Silva, P.L., L., *Urban Crime Occurrences in Association with Built Environment Characteristics: An African Case with Implications for Urban Design Sustainability*, 2020. **12**: p. 22.
18. Baja, E.B., E., *Criminal incidences in relation to built environment in Arba Minch City, Southern Ethiopian*. SN Social Sciences, 2024. **4**: p. 100.
19. Michael, T. *The Where, When and How of Crime: Principles of Environmental Criminology*. 2024 [cited 2024 24/11/2024]; Available from: <https://oercollective.caul.edu.au/criminology-criminal-justice/chapter/the-where-when-and-how-of-crime-principles-of-environmental-criminology/>.
20. Nian, G.S., J.; Huang, J., *Exploring the Effects of Urban Built Environment on Road Travel Speed Variability with a Spatial Panel Data Model*. ISPRS International Journal of Geo-Information, 2021. **10**(12): p. 829.
21. Lee, Y.S., D., *Identifying Relationship between Regional Centrality and POI Facilities: A Case Study of Seoul Metropolitan Area*. ISPRS International Journal of Geo-Information, 2024. **13**(1): p. 12.
22. Fan, D.Q., K.; Kang, C. *Understanding Urban Functionality from POI Space*. in *2018 26th International Conference on Geoinformatics*. 2018.
23. Danese, M.L., M.; Murgante, B. *Kernel Density Estimation Methods for a Geostatistical Approach in Seismic Risk Analysis: The Case Study of Potenza Hilltop Town (Southern Italy)*. in *Computational Science and Its Applications – ICCSA 2008*. 2008. Berlin, Heidelberg: Springer Berlin Heidelberg.
24. Du, Y.D., N., *A Systematic Review of Multi-Scale Spatio-Temporal Crime Prediction Methods*. ISPRS International Journal of Geo-Information, 2023. **12**: p. 209.
25. Sun, L.Z., G.; Zhao, D.; Ji, L.; Gu, H.; Sun, L.; Li, X., *Explore the Correlation between Environmental Factors and the Spatial Distribution of Property Crime*. ISPRS International Journal of Geo-Information, 2022. **11**: p. 428.
26. Amin, G.H., E.; Imtiaz, I.; Saqib, N.; Shahzad, M., *Ecotourism Potential Assessment for Gilgit-Baltistan, Pakistan using Integration of GIS, Remote Sensing, AHP and Crowd-sourced Data*. Geocarto International, 2021. **37**: p. 1-20.
27. Chitphong, S., *Safety conditions problem effecting international tourists visiting bangkok in participating in tourism activities*. 2019.
28. Patino, J.E.D., J.C.; Pardo-Pascual, J.E.; Ruiz, L.A., *Using remote sensing to assess the relationship between crime and the urban layout*. Applied Geography, 2014. **55**: p. 48-60.
29. Fondevila, G.V.-P., C.; Galindo Pérez, M.C.; Cafferata, F.G., *Crime deterrent effect of police stations*. Applied Geography, 2021. **134**: p. 102518.
30. Zeng, M.M., Y.; Wang, C., *The relationship between street environment and street crime: A case study of Pudong New Area, Shanghai, China*. Cities, 2021. **112**: p. 103143.
31. Zhang, X.L., L.; Xiao, L.; Ji, J., *Comparison of Machine Learning Algorithms for Predicting Crime Hotspots*. IEEE Access, 2020. **8**: p. 181302-181310.
32. Shepley, M.S., N.; Sadatsafavi, H.; Fournier, C.; Peditto, K., *The Impact of Green Space on Violent Crime in Urban Environments: An Evidence Synthesis*.

- International Journal of Environmental Research and Public Health, 2019. **16**: p. 5119.
33. Ogletree, S.S.L., L.R.; Powell, R.B.; White, D.L.; Brownlee, M.T. J., *Urban greenspace linked to lower crime risk across 301 major U.S. cities*. Cities, 2022. **131**: p. 103949.
 34. Venter, Z.S.S., C.; Faull, A.; Lancaster, L.; Breetzke, G.; Edelstein, I., *Is green space associated with reduced crime? A national-scale study from the Global South*. Science of The Total Environment, 2022. **825**: p. 154005.
 35. Fotios, S.R., C.; Farrall, S., *The Effect of Lighting on Crime Counts*. Energies, 2021. **14**: p. 4099.
 36. Chalfin, A.H., B.; Lerner, J.; Parker, L., *Reducing Crime Through Environmental Design: Evidence from a Randomized Experiment of Street Lighting in New York City*. Journal of Quantitative Criminology, 2021. **38**: p. 1-31.
 37. Mandalapu, V.E., L.; Vyas, P.; Roy, N., *Crime Prediction Using Machine Learning and Deep Learning: A Systematic Review and Future Directions*. 2023.
 38. Jenga, K.C., C.; Kar, G., *Machine learning in crime prediction*. Journal of Ambient Intelligence and Humanized Computing, 2023. **14**(3): p. 2887-2913.
 39. Zaidi, N.A.S.M., A.; Mostafa, S.A.; Razali, M.N. *A Classification Approach for Crime Prediction*. in *Applied Computing to Support Industry: Innovation and Technology*. 2020. Cham: Springer International Publishing.
 40. Abdulrahman, A.M., S.R.; Bander, A., *Crime Prediction Model using Three Classification Techniques: Random Forest, Logistic Regression, and LightGBM*. International Journal of Advanced Computer Science and Applications, 2024. **15**(1).
 41. Wheeler, A.P.S., W., *Mapping the Risk Terrain for Crime Using Machine Learning*. Journal of Quantitative Criminology, 2021. **37**(2): p. 445-480.
 42. Kim, S.L., S., *Nonlinear relationships and interaction effects of an urban environment on crime incidence: Application of urban big data and an interpretable machine learning method*. Sustainable Cities and Society, 2023. **91**: p. 104419.
 43. McCarty, D.h.w., K.; Lee, H., *Evaluation of Light Gradient Boosted Machine Learning Technique in Large Scale Land Use and Land Cover Classification*. Environments, 2020. **7**: p. 84.
 44. Geurts, P.E., D.; Wehenkel, L., *Extremely randomized trees*. Machine Learning, 2006. **63**(1): p. 3-42.
 45. Wang, J.H., J.; Shen, S.; Zhuang, J.; Ni, S., *Crime risk analysis through big data algorithm with urban metrics*. Physica A: Statistical Mechanics and its Applications, 2020. **545**: p. 123627.
 46. Kadar, C.P., I., *Mining large-scale human mobility data for long-term crime prediction*. EPJ Data Science, 2018. **7**(1): p. 26.
 47. Review, W.P. *Bangkok Population 2023*. 2023 [cited 2023 15/12/2023]; Available from: <https://worldpopulationreview.com/world-cities/bangkok-population>.
 48. BMA. *Department of Deputy BMA*. 2021 [cited 2023 15/12/2023]; Available from: http://office2.bangkok.go.th/ard/?page_id=4048.
 49. Vitchakorn, P., *The relationship between spatial patterns and crime against property in chanasongkram subdistrict*. Master's thesis, 2014.

50. Wassayos, N., *City safety policies lacking: MPB*, in *Bangkok Post*. 2022.
51. Chadchat, S. *Developing BKK Risk Map (BKK Risk Map)*. 2023; Available from: <https://www.chadchart.com/policy/621678c44e43cd8b4760bc3c/>.
52. Chen, Y.C., X.; Liu, Z.; Li, X., *Understanding the spatial organization of urban functions based on co-location patterns mining: A comparative analysis for 25 Chinese cities*. *Cities*, 2020. **97**: p. 102563.
53. Pineda-Jaramillo, J.V., F., *Analysis of tourist satisfaction towards urban metro systems: an hybrid Natural Language Processing and classification model approach*. 2023. 1-6.
54. Roy, G., *Travelers' online review on hotel performance – Analyzing facts with the Theory of Lodging and sentiment analysis*. *International Journal of Hospitality Management*, 2023. **111**: p. 103459.
55. Python. *A Practical Introduction to Web Scraping in Python*. 2023 [cited 2023 24/02/2023]; Available from: <https://realpython.com/python-web-scraping-practical-introduction/>.
56. Google. *Google Maps*. 2023 [cited 2023 09/10/2023]; Available from: https://en.wikipedia.org/wiki/Google_Maps.
57. Wikipedia. *Mass Rapid Transit Master Plan in Bangkok Metropolitan Region*. 2023; Available from: https://en.wikipedia.org/wiki/Mass_Rapid_Transit_Master_Plan_in_Bangkok_Metropolitan_Region.
58. He, L.P., A.; Liu, D., *Built environment and violent crime: An environmental audit approach using Google Street View*. *Computers, Environment and Urban Systems*, 2017. **66**: p. 83-95.
59. He, Q.L., J., *The roles of built environment and social disadvantage on the geography of property crime*. *Cities*, 2022. **121**: p. 103471.
60. Kamarajugedda, S. and E. Lo, *Modelling urban growth for bangkok and assessing linkages with road density and socio-economic indicators*. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2019. **XLII-4/W19**: p. 255-262.
61. (NESDC), O.o.t.N.E.a.S.D.C. *Gross Regional and Provincial Product, 2022 Edition*. 2022 [cited 2024 27/08/2024]; Available from: https://www.nesdc.go.th/nesdb_en/more_news.php?cid=156&filename=index.
62. Office, N.S. *Number of Population from Registration and Growth Rate by Region: 2012-2022*. 2022 20 Aug 2024 [cited 2024 26/08/2024]; Available from: https://www.nso.go.th/nsoweb/nso/statistics_and_indicators?impt_branch=300.
63. (NSO), N.S.O. *Latent Population Survey*. 2022 [cited 2024 27/08/2024]; 21]. Available from: https://www.nso.go.th/nsoweb/nso/survey_detail/yS.
64. Sims, K.R., A.; Bright, E.; Kaufman, J.; Pyle, J.; Epting, J.; Gonzales, J.; Adams, D.; Powell, E.; Urban, M.; Rose, A., *LandScan Global 2022*. 2023, Oak Ridge National Laboratory: Oak Ridge, TN.
65. ESA. *Spatial-Resolutions-Sentinel-2 MSI-User Guidez-Sentinel Online*. . 2023 [cited 2023 24/12/2023]; Available from: <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions/spatial>
66. Román, M.O.W., Z.; Sun, Q.; Kalb, V.; Miller, S.D.; Molthan, A.; Schultz, L.; Bell, J.; Stokes, E.C.; Pandey, B.; Seto, K.C.; Hall, D.; Oda, T.; Wolfe, R.E.; Lin, G.; Golpayegani, N.; Devadiga, S.; Davidson, C.; Sarkar, S.; Praderas, C.;

- Schmaltz, J.; Boller, R.; Stevens, J.; Ramos González, O.M.; Padilla, E.; Alonso, J.; Detrés, Y.; Armstrong, R.; Miranda, I.; Conte, Y.; Marrero, N.; MacManus, K.; Esch, T.; Masuoka, E.J., *NASA's Black Marble nighttime lights product suite*. Remote Sensing of Environment, 2018. **210**: p. 113-143.
67. Breiman, L., *Random Forests*. Machine Learning, 2001. **45**(1): p. 5-32.
 68. Tong, X.N., P.; Li, G.; Yuan, Q.; Liu, C.; Lu, H.; Li, G. *Urban Crime Trends Analysis and Occurrence Possibility Prediction based on Light Gradient Boosting Machine*. in *2021 IEEE 4th International Conference on Big Data and Artificial Intelligence (BDAI)*. 2021.
 69. Ke, G.M., Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T. *LightGBM: A Highly Efficient Gradient Boosting Decision Tree*. in *Neural Information Processing Systems*. 2017.
 70. Ghazwani, M.B., M.Y., *Computational intelligence modeling of hyoscine drug solubility and solvent density in supercritical processing: gradient boosting, extra trees, and random forest models*. Scientific Reports, 2023. **13**(1): p. 10046.
 71. Scikit-learn. *scikit-learn*. 2024 [cited 2024 04/01/2024]; Available from: https://scikit-learn.org/stable/getting_started.html.
 72. Vijayalakshmi, S.K., M.; Arun, M., *A study of various classification techniques used for very high-resolution remote sensing [VHRRS] images*. Materials Today: Proceedings, 2021. **37**: p. 2947-2951.
 73. Shi, W.M., Q.; Zhang, L.; Zhao, M.; Su, C.; Jancsó, T. *DSANet: A Deep Supervision-Based Simple Attention Network for Efficient Semantic Segmentation in Remote Sensing Imagery*. Remote Sensing, 2022. **14**, DOI: 10.3390/rs14215399.
 74. Mandrekar, J.N., *Receiver Operating Characteristic Curve in Diagnostic Test Assessment*. Journal of Thoracic Oncology, 2010. **5**(9): p. 1315-1316.
 75. Ting, K.M., *Confusion Matrix*, in *Encyclopedia of Machine Learning*, C. Sammut and G.I. Webb, Editors. 2010, Springer US: Boston, MA. p. 209-209.
 76. Suksonghong, K.G., K., *Dynamic prolonged effects of crime on tourism demand for Thailand National Parks*. Cogent Social Sciences, 2023. **9**(1): p. 2236360.
 77. Kaewumporn, K.W., S.; Panya, O., *Environmental Improvement and Urban Crime Prevention: A case Study of Selected Slum Communities in Bangkok, Thailand*. Journal of Population and Social Studies [JPSS], 2016. **21**(2): p. S-115-S-128.
 78. Prechathamwong, W.R., V., *Causal model of fear of crime among people in Bangkok*. Kasetsart Journal of Social Sciences, 2018.
 79. Sithiprom, T.S., A. *Comparative Feature Selection of Crime Data in Thailand*. in *Soft Computing in Data Science*. 2016. Singapore: Springer Singapore.
 80. Xu, W.H., P.; Duan, S.; Liu, C., *FIGAT: Accurately Classify Individual Crime Risks With Multi-Information Fusion*. IEEE Transactions on Services Computing, 2023. **16**(3): p. 1890-1903.
 81. Llaha, O. *Crime Analysis and Prediction using Machine Learning*. in *2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO)*. 2020.
 82. Yamashita, A., *Bangkok Metropolitan Area*, in *Urban Development in Asia and Africa: Geospatial Analysis of Metropolises*, Y. Murayama, et al., Editors. 2017, Springer Singapore: Singapore. p. 151-169.

83. Ketsil, T., *Environmental Factors and Hot Spot Areas of Juvenile Delinquency in Bangkok*. International Journal of Criminology and Sociology, 2021. **10**: p. 1088-1106.
84. Bowes, D.R., *A Two-Stage Model of the Simultaneous Relationship Between Retail Development and Crime*. Economic Development Quarterly, 2007. **21**(1): p. 79-90.
85. Fullerton, T.W., A., *Mexico Evidence on the Regional Retail Impacts of Violent Crime*. Journal of Economics and Public Finance, 2018. **4**: p. 244.
86. Rosenthal, S.S.R., A., *Violent crime, entrepreneurship, and cities*. Journal of Urban Economics, 2010. **67**(1): p. 135-149.
87. Foster, S.W., L.; Christian, H.; Knuiman, M.; Giles-Corti, B., *Planning safer suburbs: Do changes in the built environment influence residents' perceptions of crime risk?* Social Science & Medicine, 2013. **97**: p. 87-94.
88. Foster, S.K., M.; Wood, L.; Giles-Corti, B., *Suburban neighbourhood design: Associations with fear of crime versus perceived crime risk*. Journal of Environmental Psychology, 2013. **36**: p. 112-117.
89. Boivin, R.F., M., *Crimes by Visitors Versus Crimes by Residents: The Influence of Visitor Inflows*. Journal of Quantitative Criminology, 2018. **34**(2): p. 465-480.
90. McCormick, J.G.H., S.M., *Strategies in use to reduce incivilities, provide security and reduce crime in urban parks*. Security Journal, 2015. **28**(4): p. 374-391.
91. Ho, T.Z., J.; Brown, M.P., *Examining hotel crimes from police crime reports*. Crime Prevention and Community Safety, 2009. **11**(1): p. 21-33.
92. Ho, T.Z., J.; Brown, M.P.; Dooley, B., *Hotel crimes: An unexplored victimization in the hospitality industry*. Security Journal, 2017. **30**(4): p. 1097-1111.
93. Brunt, P.M., R.; Hambly, Z., *Tourist victimisation and the fear of crime on holiday*. Tourism Management, 2000. **21**(4): p. 417-424.
94. Buil-Gil, D.M., R.I., *Do tourists report crime to the police? An exploratory analysis in Barcelona*. Current Issues in Tourism, 2023. **26**(18): p. 2931-2947.
95. Mikow-Porto, V.S., T., *Why healthcare facilities are vulnerable to crime*. Journal of healthcare protection management : publication of the International Association for Hospital Security, 2013. **29**: p. 26-37.
96. Tippawan, R., *The Perception and Expectation Regarding Hospital Violence Management among Healthcare Workers in Emergency and Intensive Care Units*. Journal of Thailand Nursing and Midwifery Council, 2024. **39**(1): p. 33-46.
97. Deza, M.M., J.C.; Solomon, K., *Local access to mental healthcare and crime*. Journal of Urban Economics, 2022. **129**: p. 103410.
98. Sattayathamrongthian, M.J., W.; Makerd, P.; Wongthaworn, K., *Medical tourism in thailand: a case study of the bangkok metropolitan area*. The EUrASEANs: journal on global socio-economic dynamics, 2023: p. 289-298.
99. Ahmad, S., *Crime Prevention through Environmental Design in Healthcare Facility: A Case Study of Sehat Al Sharq at Al Khobar, KSA*. Qeios, 2023.
100. Ajala, O.O., B., *Influence of police stations' location on crime incidence in developing countries like Nigeria*. International Journal of Society Systems Science, 2018. **10**: p. 132.
101. Circo, G.M., E., *Estimating the impact of an integrated CCTV program on crime*. Journal of Experimental Criminology, 2021. **17**(1): p. 129-150.

102. Button, M.T., M., *Criminal Activity in the Financial Sector*, in *The Handbook of Security*, M. Gill, Editor. 2014, Palgrave Macmillan UK: London. p. 427-449.
103. Hannah, C.S., T.; & Subbaiyan, G., *The Impact of Design and Placement of ATM Deployments on Perceived Safety in India*. International Journal of Criminal Justice Sciences (IJCJS), 2016. **11**(1): p. 1.
104. Tin, D.I., F.; Ciottone, G.R., *Attacks on Educational Institutions*. Prehospital and Disaster Medicine, 2022. **37**(3): p. 333-337.
105. Solovjev, V., *Threats to the security of educational institutions (criminological study based on an anonymous online survey of students)*. Vestnik Tomskogo gosudarstvennogo universiteta. Pravo, 2022: p. 97-113.
106. Wood, L.S., C.; Kammer-Kerwick, M.; Follingstad, D.; Busch-Armendariz, N., *Climate Surveys: An Inventory of Understanding Sexual Assault and Other Crimes of Interpersonal Violence at Institutions of Higher Education*. Violence Against Women, 2017. **23**(10): p. 1249-1267.
107. Tillyer, M.S.W., R.J., *Busy Businesses and Busy Contexts: The Distribution and Sources of Crime at Commercial Properties*. Journal of Research in Crime and Delinquency, 2019. **56**(6): p. 816-850.
108. Twinam, T., *Danger zone: Land use and the geography of neighborhood crime*. Journal of Urban Economics, 2017. **100**: p. 104-119.
109. Smith, M.J.C., R.V., *Crime and Public Transport*. Crime and Justice, 2000. **27**: p. 169-233.
110. Poole, A.C.M., J.C.; Toohy, K.; Burraston, B., *Testing the Impact of Road Network Connectivity on Criminal Lethality*. Homicide Studies, 2018. **22**(3): p. 277-295.
111. Iwanski, N., et al. *A Computational Model for Predicting the Location of Crime Attractors on a Road Network*. in *2012 European Intelligence and Security Informatics Conference*. 2012.
112. Bernasco, W., *A sentimental journey to crime: effects of residential history on crime location choice*. Criminology, 2010. **48**(2): p. 389-416.
113. Bernasco, W.N., P., *How Do Residential Burglars Select Target Areas?: A New Approach to the Analysis of Criminal Location Choice*. The British Journal of Criminology, 2004. **45**(3): p. 296-315.
114. Shichor, D.D., D.; O'Brien, R., *Population density and criminal victimization. Some unexpected findings in central cities*. Criminology, 2006. **17**: p. 184-193.
115. Hipp, J.R.R., A., *Micro- and Macro-Environment Population and the Consequences for Crime Rates*. Social Forces, 2013. **92**(2): p. 563-595.
116. Boivin, R., *Routine activity, population(s) and crime: Spatial heterogeneity and conflicting Propositions about the neighborhood crime-population link*. Applied Geography, 2018. **95**: p. 79-87.
117. Sypion, N., *Exploring the Impact of Green Areas on Crime Rates in Urban Environments*. EUROPEAN RESEARCH STUDIES JOURNAL, 2023. **XXVI**: p. 456-461.
118. Demotto, N.D., C.P., *A GIS Analysis of the Relationship between Criminal Offenses and Parks in Kansas City, Kansas*. Cartography and Geographic Information Science, 2006. **33**(2): p. 141-157.
119. Kimpton, A.C., J.; Wickes, R., *Greenspace and Crime: An Analysis of Greenspace Types, Neighboring Composition, and the Temporal Dimensions of*

- Crime*. Journal of Research in Crime and Delinquency, 2017. **54**(3): p. 303-337.
120. Peng, C.S., W.; Zhang, X., *Crime under the Light? Examining the Effects of Nighttime Lighting on Crime in China*. Land, 2022. **11**(12): p. 2305.
 121. Welsh, B.F., D., *Effects of Improved Street Lighting on Crime*. Campbell Systematic Reviews, 2008. **4**: p. 1-51.
 122. Yiampisan, M.S., M., *Using the Kernel Density Estimation Surface for Criminal Pattern: A Case Study in Phranakhon District, Bangkok*. Journal of Architectural/Planning Research and Studies (JARS), 2010. **7**: p. 87-102.
 123. Cai, J.L., J.; Wang, S.; Yang, S., *Feature selection in machine learning: A new perspective*. Neurocomputing, 2018. **300**: p. 70-79.
 124. Kaneko, H., *Interpretation of Machine Learning Models for Data Sets with Many Features Using Feature Importance*. ACS Omega, 2023. **8**(25): p. 23218-23225.
 125. Oran, V. *Bangkok Transit Map*. 2024 [cited 2024 03/04/2024]; Available from: <https://www.bangkoktransitmap.com/>.

BIOGRAPHY

NAME Kriangkrai Manochai

DATE OF BIRTH 01 January 1986

PLACE OF BIRTH Chiangrai

PRESENT ADDRESS 222/270 Sunmoon Lake Bangpood Pakkred Nonthaburi
11120

EDUCATION Master degree

