Investigation of Chiang Rai irregular migration and its environmental factors using Machine Learning

JANJIRA KHUENMAMUEANG

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

การศึกษาการย้ายถิ่นฐานแบบผิดปกติของจังหวัดเชียงรายและปัจจัยด้านสิ่งแวดล้อมโดยใช้ Machine Learning

จันทร์จิรา คืนมาเมือง

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาวิชาภูมิสารสนเทศศาสตร์
คณะภูมิสารสนเทศศาสตร์ มหาวิทยาลัยบูรพา
2568

Investigation of Chiang Rai irregular migration and its environmental factors using Machine Learning

JANJIRA KHUENMAMUEANG

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

The Thesis of Janjira Khuenmamueang 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	
....	.... Principal examiner
(Professor Dr. Hong Shu)	(Professor Dr. Timo Balz)
	.... Member
Co-advisor	(Professor Dr. Milan Konecny)
....	.... Member
(Assistant Professor Dr. Phattraporn Soytong)	(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 Pongkittiwiboon)
....

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)
....

65910032: MAJOR: GEOINFORMATICS; M.Sc. (GEOINFORMATICS)
KEYWORDS: irregular migration, remote sensing data, machine learning

JANJIRA KHUENMAMUEANG : INVESTIGATION OF CHIANG RAI IRREGULAR MIGRATION AND ITS ENVIRONMENTAL FACTORS USING MACHINE LEARNING. ADVISORY COMMITTEE: HONG SHU, Ph.D. PHATTRAPORN SOYTONG, Ph.D. 2025.

Irregular migration constitutes a multifaceted challenge in Southeast Asia. Thailand serves as a significant transit and destination for migrant workers from neighboring countries, especially Myanmar, Laos, and Cambodia. Chiang Rai Province is located in the uppermost region of Thailand, which is joint between Myanmar and Laos. Its geographical and socioeconomic conditions attract a lot of migrant workers, both legal and illegal. This study investigates the environmental factors influencing irregular migration along Thailand-Myanmar and Laos borders in Chiang Rai province. The study utilized geospatial data, including Remote Sensing (RS) and Geographic Information System (GIS) data, combined with arrest point data of irregular migration along the border from 2019 to 2023; three machine learning algorithms consisting of XGBoost, Random Forest, and LightGBM were applied to analyze and predict irregular migration patterns.

The primary objectives are threefold: (1) to analyze the variables associated with irregular migration in Chiang Rai Province between 2019 and 2023 (2) to evaluate and compare the predictive performance of machine learning algorithms for accurately predicting irregular migration (3) to create a risk map visualizing the spatial distribution of irregular migration incidents within the study area.

The study focuses on four districts of Chiang Rai province which consist of Mae Sai, Mae Chan, Chiang Saen, and Mae Fa Luang. Their different geographic characteristics present challenges for irregular migration management. Mae Sai represents the highest irregular migration incidents. Meanwhile, Mae Fa Luang and Mae Chan reported fewer cases. The study shows that the topography of the study area influences migration patterns. Remote sensing data, including the Normalized Difference Vegetation Index (NDVI) and Bare Soil Index (BSI), were utilized to

assess land cover and vegetation density. Moreover, to understand human activity and the physical landscape, Nighttime light data and Digital Elevation Models (DEM) were applied in this study.

The results show that irregular migration activity was highest in April or during the summer season and lowest during the dry/winter months of October and December. Temporal analysis also marked a preference for nighttime crossings to avoid detection. The highest number of irregular migration cases was recorded in 2022, with 1,882 cases, while in 2019, 56 incidents were recorded, the lowest number out of all five years.

Myanmar nationals make up the majority of irregular migrants (1,991 cases), while Thai and Chinese nationals were recorded in smaller proportions. The study indicated that 55% of migrants were male, while 45% were female; this shows that migration pressures affect both genders almost equally.

XGBoost exhibited the highest predictive accuracy among the other machine learning algorithms, with an R-squared value of 0.91 and the lowest Root Mean Square Error (RMSE). Feature importance analysis across all models consistently identified road networks as the most significant predictor of irregular migration. The transportation infrastructure is facilitating movement across borders. Elevation and proximity to rivers are also key factors that light the physical environment's impact on irregular migration routes. In addition, The Random Forest model, which is less accurate than XGBoost, also highlighted Nighttime light as a crucial factor, indicating that irregular migrants prefer areas with the dance of human activity.

An irregular migration risk map was generated using the best machine learning algorithm. The map presents the high-risk area along the border, especially near Mae Sai. Mae Sai district is in a critical environment; its road networks and challenging terrain converge to create migration hotspots. The risk map from this study is very beneficial, especially for border authorities to have more focus and effective monitoring and intervention efforts.

This study illustrates the application of geospatial techniques and machine learning to analyze the irregular migration patterns in Chiang Rai Province along

Thai-Myanmar and Laos border. Environmental factors, geographic data, and irregular migration arrest data are important tools for investigating migration patterns in border areas. Developing more accurate predictive models will be able to support policymakers and border control authorities to understand essential insights. As irregular migration remains a critical issue in Thailand and Southeast Asia, applying predictive analytics and geospatial technologies will help in effective border management and immigration control strategies.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor, Prof. Hong SHU, for his invaluable guidance throughout this study journey. His expert advice and encouragement have been instrumental in completing my thesis. I am equally grateful to Asst. Prof. Phattharaphorn Soithong, my co-advisor, has unwavering support in all aspects of my research and university life both in Thailand and China.

I am sincerely thankful to the Chiang Rai Border Control Authority for providing essential arrest point data, which played a crucial role in training the algorithms used in this study. Special thanks go to my classmate, Japhichom Phongphat, for his assistance with technical methods and his generous suggestions that enriched my research.

I am fortunate to have had the encouragement and camaraderie of my fellow SCGI#5 classmates and my senior, Ms. Sunantha Ousaha, whose support from the beginning has been invaluable. I could not have come this far without their kindness and friendship.

I also 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 for their academic guidance. I also appreciate the Geo-Informatics and Space Technology Development Agency of Thailand (GISTDA) for the opportunity to pursue my studies further.

I would like to thank the commander of the Royal Thai Police for granting me this opportunity to enhance my experience and reach new milestones in my professional journey. To my family, I am deeply grateful for their understanding, unwavering support, and belief in me. Lastly, I want to express my gratitude for persevering along this journey. You did it!

Janjira Khuenmamueang

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1 INTRODUCTION

1.1 Background

Thailand is a crucial transit and destination country for migrant laborers in Southeast Asia. Thailand's geographical and cultural affinities and economic conditions draw laborers from adjacent nations (Khanawiwat, 2020). The International Organization for Migration (IOM) predicted approximately 4.9 million migrants in Thailand in 2018, according to the Thailand Migration Report 2019. The largest groups of Thailand's migrant population are the migrants from Myanmar, Laos, Cambodia, and Vietnam (Harkins, 2019). Furthermore, the investigations indicated that the population of migrants with irregular status exceeded 800,000 in the same year, a figure significantly high relative to the total number of migrants residing in the nation (Khanawiwat, 2020). Irregular migration denotes the movement of humans that transpires outside the legal parameters set by laws, regulations, or international accords governing entry into or departure from the country of origin, transit, or destination (MIGRATION). Illegal immigration is a complex and multifaceted issue with significant implications for sending and receiving countries.

The World Migration Report (2022) indicates that Myanmar-Thailand corridor is among the top 15 international migration routes in 2020 (McAuliffe, 2021). The international boundary between Thailand and Myanmar spans 2,401 kilometers, making it the longest among neighboring countries (Onanong Thippimol, 2011). The nature of Thailand-Myanmar and Laos border region has rendered it a significant transit hub for illegal immigration. The study conducted by the International Labor Organization (ILO) and IOM indicates that 91 percent of workers from Myanmar accessed Thailand via irregular channels (Harkins, 2019).

The primary group of irregular migrants originates from Myanmar, particularly in the northern region, which exhibits the most significant arrest rates and an increasing influx along the border (Aenihon, 2018). Chiang Rai province is situated in the northernmost region of Thailand. Its border adjoins Laos and Myanmar. The landscape of Chiang Rai consists of mountains, forests, and rivers that correspond to the distinctive qualities of each region's limits (Onanong Thippimol, 2011). Chiang Rai is a border commercial conduit with Myanmar, Laos, and southern China. This region also consists of communities of several ethnic groups. A significant number of foreign labor migrants reside and work both legally and illegally in this area.

Chiang Rai province's topography constitutes the primary obstacle to issues faced by illegal migrant workers. A long border characterized by mountains and forests, and narrow and long river-separated countries between Thailand and Myanmar facilitates migrants to escape and cross the border irregularly. Furthermore, economic expansion in Thailand has led to an increase in population movement in different areas and heightened border crossings, resulting in numerous issues such as social security, the economy, and public health in Chiang Rai and other areas (Jakkawat, 2017).

Irregular migration is a prominent issue that Thailand has confronted for an extended period, primarily because of numerous irregular migrant routes, particularly at Chiang Rai international border with Myanmar and Laos. This study will utilize machine learning algorithms to examine the environmental factors linked to irregular migration in Chiang Rai Province. The study concentrates on four principal districts along Thailand-Myanmar and Laos: Chiang Saen, Mae Sai, Mae Chan, and Mae Fa Luang. The study employs remote sensing data, geographical information system (GIS) data, and data on irregular movement arrest points. Machine learning algorithms which include XGBoost (XGB), Random Forest (RF), and LightGBM (LGBM), will be utilized to evaluate the optimal accuracy performance method. The variables and environmental factors related to irregular migration and risk maps will be presented for border management to enhance the understanding of significant variables and improve future border control in the study area.

1.2 Scientific Questions

- What variables are significantly associated with the incidence of irregular migration in Chiang Rai Province, Thailand during the period from 2019 to 2023?
- How do the predictive performances of XGBoost (XGB), Random Forest (RF), and LightGBM (LGBM) differ in modeling irregular migration patterns in Chiang Rai Province, and which algorithm provides the highest accuracy in prediction?
- How does machine learning algorithm generate and display a risk map showing the spatial distribution of irregular migration in Chiang Rai Province, Thailand between 2019 and 2023?

1.3 Objectives

- To analyze the variables associated with irregular migration in Chiang Rai Province, situated along the Thailand-Myanmar and Laos border, from 2019 to 2023.
- To assess and compare the performance of machine learning algorithms between XGBoost (XGB), Random Forest (RF), and LightGBM (LGBM) for accurate calculation of irregular migration on the Chiang Rai border, from 2019-2023.
- To create a risk map displaying irregular migration episodes in the study area.

1.4 Structure of Thesis

The study examines irregular migration in Chiang Rai Province, focusing on Thailand-Myanmar and Laos borders from 2019 to 2023. The study focuses on four key districts: Chiang Saen, Mae Sai, Mae Chan, and Mae Fa Luang. After reviewing issues, various factors, and related works about irregular migration, the data from remote sensing, GIS data, and irregular migration arrest point data were analyzed in the study. Machine learning algorithms, such as XGBoost, Random Forest, and LightGBM, are conducted to analyze the best performance algorithm for calculating irregular migration patterns. The performance of these algorithms will be evaluated and compared using evaluation metrics. The study's results involve analyzed data about the variables and environmental factors influencing irregular movement and risk mapping, highlighting hotspot areas of events within the study region. The study's results will be addressed at the conclusion. Structure of Thesis is shown in Figure 1.

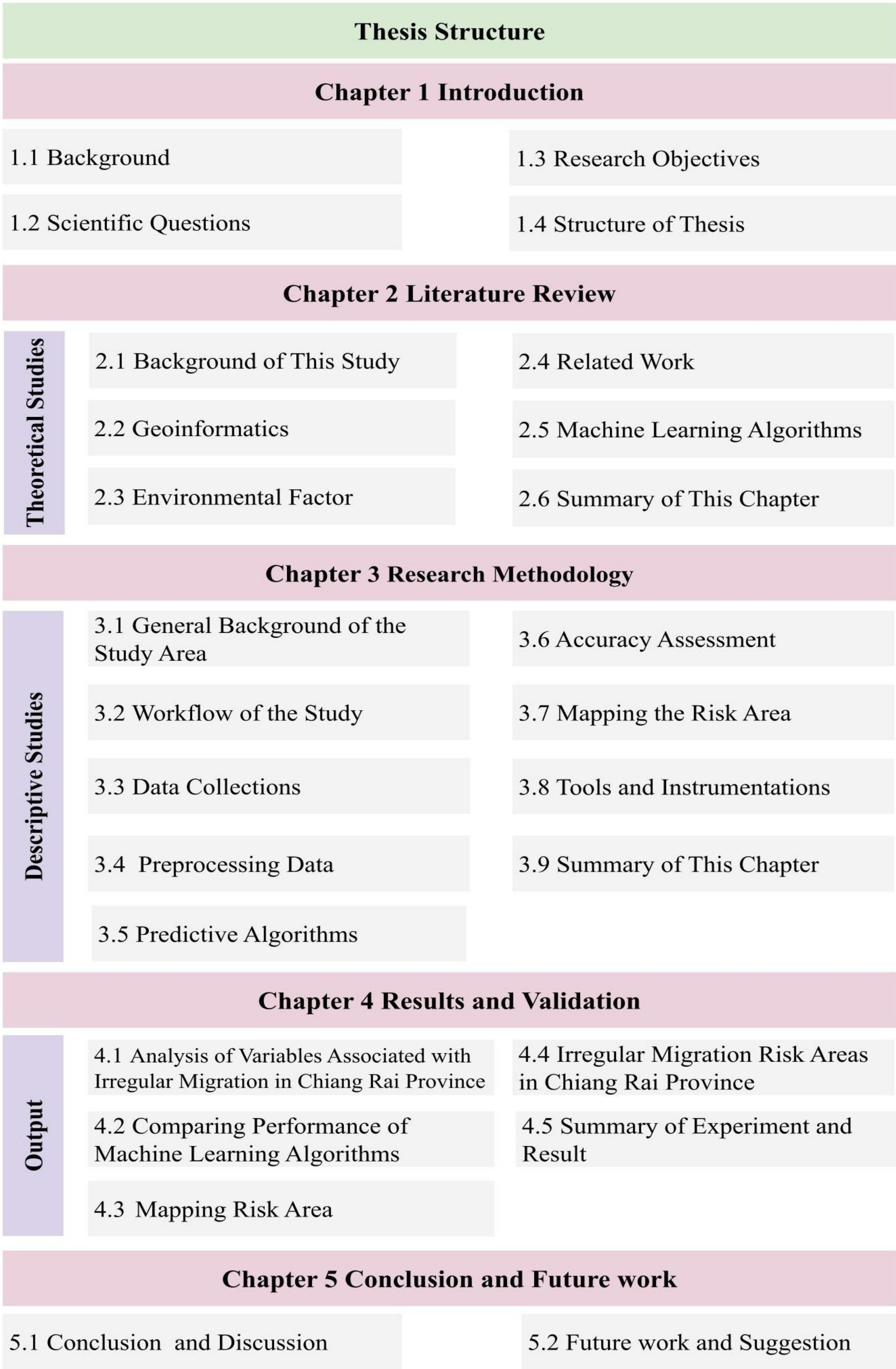

Figure 1 Structure of Thesis

CHAPTER 2 LITERATURE REVIEW

2.1 Background of This Study

2.1.1 Irregular Migration in Thailand

An irregular migrant is a person who lacks legal status in transit or a host nation due to unlawful arrival, violation of a condition of admission, or visa expiration. The concept includes, among others, people who entered a transit or host nation legitimately but stayed longer than permitted or later accepted illicit work. A migrant with irregular status is an individual who has entered a foreign country without adhering to the official immigration process, lacks a valid work visa, or has subsequently become irregular after entry (Migration, 2019). The term “illegal migrants” is commonly used. However, “irregular” is preferred over “illegal” since “illegal” conveys a criminal connotation and is perceived as denying migrants’ humanity (Organization, 2014).

A significant population of irregular migrant workers from neighboring nations is working in Thailand. Since Thailand offers higher wages and employment prospects in the service, agriculture, and fishery industries. The influx of foreign workers or irregular migrants is not solely a breach of immigration rules; it is also influenced by economic considerations, including high demand and low wages for irregular laborers. The diminishing natural borders in Thailand and the minimal expense associated with unlawful border crossings incentivize irregular migration (Basir, 2020).

In 2020, the vast majority of refugees from Asian countries resided in neighboring countries. As in the past, Geography is still one of the primary factors influencing migration, despite the expansion of globalization. Many individuals who migrate across international boundaries do so to neighboring nations, which may be easier to reach, more familiar, and more uncomplicated to return to. In the same year, Thailand was the sixteenth-ranked international migration destination, while Myanmar was the nineteenth-ranked international migration origin country. Furthermore, the record of international migration between nations revealed that

Myanmar-Thailand is in the top fifteen most significant corridors in the world. This migration pattern indicates that there were large amounts of Myanmar migrants residing in Thailand in 2020; around 1,848,270 people. Myanmar has a large refugee population and displacement of people living in other countries around the world. However, according to the current statistics, many more people are displaced by natural disasters than by conflict and violence; the largest new internal displacements in Asia were caused by natural disasters. A pandemic also affects the movement of migration.

In these few years, COVID-19 has had a significant impact on mobility and migration. To prevent the spreading of COVID-19, Asian nations implemented some initial international and domestic restrictions connected to the virus throughout 2020. By mid-June 2021, global travel limitations, such as screening arrivals, were still in effect in most Asian nations. International controls, such as bans on arrivals and total border closures, decreased far more than quarantine measures throughout time, with the latter decreasing significantly more. COVID-19 negatively affects migrants across the entire international migration cycle and increases their previous vulnerabilities. Thousands of migrants, including seasonal laborers, temporary residence holders, international students, migrants traveling for medical treatment and reintegration, mariners, and others, were stuck by the border closures. It compelled immobility by stressing particular forms of mobility or by relegating movement to informal channels such as irregular routes (McAuliffe, 2021).

In addition, numerous Myanmar migrants are drawn to Thailand due to elevated labor costs, economic challenges in Myanmar, ongoing conflicts among various factions, the extensive border facilitating escape, and lenient penalties under Thai law. Moreover, illicit immigration is less arduous than legal immigration (Trimek, 2020). One study revealed that illegal immigrants were mainly smuggled into natural routes at night, when no officers were patrolling, and then moved to the meeting point where smugglers were waiting to facilitate their entry into the city. Some migrants smuggle by taking public buses or hiding vegetables, fruits, or other items on trucks. In contrast, some can pass through various checkpoints by bribing officers or deceiving them by using false government officials' vehicles to pass quickly (Sapprasert, 2017). Many factors and conditions are related to irregular migration in Thailand. Addressing Thailand's challenges in border control and the need for more effective monitoring strategies is necessary.

2.2 Geoinformatics

Geoinformation includes Remote Sensing technology, Geographic Information Systems, and Global Navigation Satellite System. Details of these technologies are as follows:

2.2.1 Geographic Information System

A Geographic Information System (GIS) is a computer-based system for the collecting, storage, management, analysis, and representation of geo-referenced data to facilitate decision-making. GIS is a computer-based tool for mapping and analyzing objects and occurrences on Earth. GIS system consists of software, hardware, data, and personnel that facilitate the entry, manipulation, analysis, and presentation of information associated with a specific location on the Earth's surface. Its most significant applications are in resource management, utility management, telecommunications, urban and regional planning, vehicle routing, parcel delivery, and all scientific disciplines related to the Earth's surface (Ershad & Ali, 2020a).

2.2.2 Remote Sensing

Remote sensing (RS) is defined as the art, science, and technology employed to determine, quantify, and evaluate the characteristics of object features or targets located on, above, or below the Earth's surface, without direct contact between the sensors and the observed targets or events. This enables the collection of information regarding object features by the detection and recording of reflected or emitted energy, followed by processing, analysis, and implementation of that data (Awange & Kiema, 2019). Remote sensing is the measurement of physical quantities at long distances, generally via quantitative spectroscopic techniques. Implementing satellite technology to examine aspects of the Earth's atmosphere and surface is an expanding area characterized by multiple uses and the continual development of new or better approaches (Taylor, 1996).

In recent years, the application of remote sensing data for natural resource mapping and as input for environmental process modeling has become increasingly popular. Remotely sensing data abilities from diverse sensors across various platforms, characterized by a broad spectrum of spatiotemporal, radiometric, and spectral resolutions, have established remote sensing as a primary source of data for large-scale applications and studies (Melesse et al., 2007).

2.2.3 Global Navigation Satellite System

A Global Navigation Satellite System (GNSS) offers constant positioning on or near the Earth's surface through the use of radio signals generated from a network of orbiting satellites. The system includes approximately 20 or more satellites, ground stations for monitoring and control, and receivers utilized by users to passively gather signals for locating. GNSS satellites emit carrier signals within the 1.1–1.8 GHz spectrum, modified by pseudo-random noise (PRN) algorithms to reduce interference and facilitate range measurements. By calculating the distances to at least four satellites and utilizing satellite positional data, a receiver can determine its three-dimensional coordinates and temporal information (Axelrad, 2010).

Global Positioning Systems (GPS) is a satellite-based navigation system consisting of a network of 24 satellites launched in orbit by the U.S. Department of Defense. GPS was initially created for military purposes; however, in the 1980s, the government allowed it to be used for civilian purposes. GPS operates throughout all weather conditions, globally, and at all times. There are no subscription fees or setup charges to use GPS (Ershad & Ali, 2020b). GPS and GIS are widely used to collect and combine spatial data from diverse sources. Recent advancements in satellite photography and remote sensing offer scientists obtain spatial data at various resolutions (Gotway & Young, 2002).

2.3 Environmental Factor

Migration is related to various factors. Environmental factors can serve as both push and pull factors in migration (van der Geest et al., 2010). This study focuses on environmental factors that affect irregular migration in the study area.

2.3.1 Normalized Difference Vegetation Index

Normalized Difference Vegetation Index (NDVI) is calculated from the ratio of red and near-infrared in the electromagnetic spectrum. NDVI has become the primary way to describe crops. Area covered around the world Vegetation classification and dynamics and the life cycle of plants (Kogan, 1995). The NDVI measures the amount of green vegetation by considering that during photosynthesis, plants absorb visible light and strongly reflect near-infrared light, which is not used for photosynthesis. NDVI values relate to green biomass, green leaf area index (LAI) (Caruso et al., 2023), and the percentage of vegetation cover (Shahriar Pervez et al., 2014).

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

In this study, NDVI is utilized to assess environmental conditions, these fluctuations can provide insight into environmental stressors that may influence migration patterns.

2.3.2 The Bare Soil Index

The Bare Soil Index (BSI) is a spectral index employed to identify and quantify areas of bare soil, determined by a mix of visible and shortwave infrared bands (Salas & Kumaran, 2023). BSI is calculated from the reflection values of the blue, red, near-infrared (NIR), and shortwave infrared (SWIR) bands, indicating the spectral properties of bare soil surfaces. The BSI index is very helpful for evaluating soil exposure, land degradation, and desertification processes.

$$BSI = \frac{(SWIR + red) - (NIR + blue)}{(SWIR + red) + (NIR + blue)}$$

BSI produces a method for identifying areas with minimal to absent vegetation cover, distinguishing between bare soil and other land cover categories. It is a helpful method for environmental evaluations regarding land management, agriculture, and the analysis of spatial patterns of human land alteration (Nascimento et al., 2021). Analyzing BSI values enables the observation of temporal variations in bare soil cover related to both natural phenomena and human activity, which offers insights into land-use dynamics that may influence irregular moving patterns.

2.3.3 Nighttime Light

Satellites such as the Defense Meteorological Satellite Program Operational Line Scanner (DMSP/OLS) and the Visible Infrared Imaging Radiometer Suite (VIIRS) collect artificial light emanating from the Earth's surface during nighttime. Initially developed in the 1970s for the detection of cloud formations, DMSP/OLS further began the recording of visible and near-infrared emissions, including urban lights and fires. A digital archive developed by the National Oceanic and Atmospheric Administration (NOAA) in 1992 significantly enhanced the accessibility of nighttime light data, facilitating its wide utilization in urbanization monitoring, population density estimation, socioeconomic analysis, and environmental impact assessments, despite the initial limitations of early data (Huang et al., 2014).

Nowadays, Nighttime Light (NTL) data is highly valued for its ability to reflect human activity, making it an essential tool for studying urbanization, economic indicators such as GDP and population, and the environmental impacts of human development. NTL operates as reliable data for assessing economic activity and development in subnational regions of developing countries (Bruederle & Hodler, 2018). The benefit of it consists of its ability to connect remote sensing technology with socioeconomic research, offering an independent and objective examination of human activities. In densely populated urban areas, NTL data can be employed to analyze and organize human activities according to land use patterns (Ren et al., 2020).

2.3.4 Digital Elevation Model

A Digital Elevation Model (DEM) is a fundamental geographical dataset used in GIS to illustrate the Earth's surface. Digital Elevation Models (DEMs) are a digital array of ground elevation information that commonly employs a grid pattern to represent continuous terrain variation. DEMs represent a "bare" land surface, omitting non-ground elements such as flora, buildings, and other structures, concentrating exclusively on the authentic height of the terrain (Zhou, 2017).

The essential purpose of DEMs is to examine the Earth's physical attributes, especially as anthropogenic activities progressively modify natural terrains. According to Burrough (1986), DEMs offer a gridded matrix representation of relief crucial for analyzing and predicting topographic phenomena. Their progression typically encompasses four phases: (i) Data acquisition by technologies such as remote sensing, LiDAR, or photogrammetry; (ii) Data modeling utilizing techniques like image processing and photogrammetry; (iii) Data management encompassing data coding, organizing, and spatial database management; and (iv) Application development.

DEMs provide many applications, encompassing photogrammetry, remote sensing, hydrology, geomorphology, civil and geological engineering, urban planning, environmental management, and telecommunications. They also function in specific domains such as military route planning, aviation simulation, and video game production. DEMs can be depicted as contour maps, illustrating lines of uniform elevation, or as point height grids, where terrain heights are collected at regular or irregular intervals (Lakshmi & Yarrakula, 2019).

Various stakeholders generate and utilize DEMs in diverse fields such as geomorphometry, cryosphere science, soil science, precision agriculture, natural hazard management, and telecommunications. DEMs are a crucial quantitative instrument in remote sensing and spatial analysis which also help in supplying critical data for modeling, planning, and environmental evaluation (Guth et al., 2021). Technology development has significantly enhanced the accuracy and resolution of DEMs, primarily via high-resolution satellite-based models. As DEM technology advances, it yields progressively accurate topographic data that tackles practical difficulties in scientific, technical, and commercial domains.

2.3.5 LandScan Global Population Database

LandScan Global Population Database, introduced in 1998, provides statistics on the distribution of the global population. It created a high-resolution global population database with a spatial resolution of 30 by 30 seconds (Dobson et al., 2000). The LandScan online application and demographic datasets, developed by

East View Geospatial using data from Oak Ridge National Laboratory, were initially developed for U.S. military and intelligence communities and serve a variety of uses (Schuster, 2020).

LandScan is the integration of both land cover and land use, representing changes caused by humans and activities. This combination provides a more accurate comprehension of population dynamics. It enhances this methodology by providing a spatial resolution of approximately 90 meters, efficiently recording population patterns during both day and night. This high-resolution model of LandScan enhances locational precision and offers significant insights for applications including urban planning, disaster preparedness, public health, and sustainable development (Bhaduri et al., 2007).

This model is customized to account for each country's specific data conditions and geographical characteristics. It helps investigate border areas where complicated and various activities happen within 24 hours.

2.3.6 Land Use Land Cover

A Land Use and Land Cover (LULC) analysis is a vital insight into land dynamics, including land use change and urban development. The land cover indicates the physical characteristics on the Earth's surface, including water bodies, plants, soil, and infrastructure, which are essential to biophysical processes and global environmental changes. LULC allows us to manage the use of natural resources, environmental modeling, and identifying habitats for biodiversity, which is significant. Land cover is frequently represented using maps, enabling observation and examination of physical surface characteristics. Land use is related to how humans adjust or manage land cover for specific purposes such as agriculture, urban development, or extraction of resources. Land cover represents the Earth's natural and constructed environments, while land use reflects human impact upon those landscapes, generally necessitating socio-economic context or field validation to comprehend its purpose (Montalván-Burbano et al., 2021).

2.3.7 River

Natural landscapes and human activities were influenced by the river for a long time. Rivers are water bodies flowing downslope via natural open channels under gravity's force. It consists of streams with a significant volume and flow. These dynamic geomorphological systems are crucial in shaping the Earth's surface by transporting water, sediments, and other materials. Rivers not only flow toward the ocean but are also able to terminate inland because of various geographical, climatic, or anthropogenic factors. Rivers are essential components of drainage basins, they drive erosion and sediment deposition processes while supporting diverse and complex ecosystems essential to environmental balance (Latrubesse & Park, 2017).

2.3.8 Road

Road information is an important data source of a basic geographic database. Roads are a key element of human infrastructure, connecting between locations. A road is a lane, course, or route that has been cleared or enhanced to facilitate travel by foot or various forms of transportation. Roads typically comprise one or more roadways featuring several lanes, walkways, and adjacent areas. Public roads are classified by their function and usage. Different types of roads offer different transportation purposes, such as highways, motorways, principal and secondary roads, and local streets (Dubey & Gangwar, 2017).

Roads establish connections between communities. Roads established trade routes and patterns and encouraged the discovery and geographical development of the nation. Roads enable the movement of armies and able to unite people. Moreover, roads influence migration patterns, roads can be barriers or can serve as escape routes (Conover, 2020). Road networks are major influential factors in the development of any nation (Babu & Manoj, 2020).

2.3.9 Slope

Slopes define the sloped surfaces of the Earth's topography, characterized by vertical or sub-horizontal irregularities. The origins and structures of these slopes vary (Nikiforov et al., 2020). Slope in GIS is generally generated using Digital Elevation

Models (DEMs) by neighborhood-based methods that average neighboring elevation values. These methods are particularly appropriate for the analysis of coarse-resolution DEMs, commonly employed in broad landscape assessment (Ashraf et al., 2012).

2.4 Related Works

Several methodologies for comprehending the relationship between migration patterns and environmental, demographic, and geographical variables through traditional procedures and advanced analytical techniques. A study in Ghana investigates the correlation between migration and environmental dynamics by including census data and RS, NDVI. This study shows that rural population density and availability of natural resources play a more substantial influence on migratory patterns (van der Geest et al., 2010).

Conversely, studies examining nine Asian mega deltas investigate urban migration through the analysis of gridded population density data, nighttime light changes, and DEM. Urban expansion driven by migration is particularly apparent in low-lying coastal areas at elevations below 10 meters, characterized by dense populations that are susceptible to environmental risks. Night light statistics underscore significant urban expansion in these deltaic regions (Small et al., 2018).

A study employs machine learning techniques, specifically Random Forest and SPRUCE (Spatial Prediction using Random Forest to Uncover Connectivity across Environments), to examine the environmental and geographical determinants affecting people's migration patterns. The study employs high-density SNP array data and coalescent-based MAPS (Migration Analysis of Population Structures) to examine more than 20 spatial variables, including climate and topography. The study identified precipitation, the minimum temperature of the coldest month, and altitude as significant contributors (Pless et al., 2023).

Machine learning is similarly utilized in a study on migration in Bangladesh, employing Random Forest models to examine a dataset of 2,000 variables from 1,700 households. This method recognizes affluence and household structure as crucial elements affecting migration choices. Utilizing extensive datasets, machine learning methodologies reveal novel potential to identify patterns in intricate social and environmental interactions that conventional approaches may neglect (Best et al., 2022).

Irregular migration is also one of the unlawful acts in Thailand, which can be categorized as a crime. The examination of crime patterns and criminology has advanced considerably with the incorporation of spatial analysis and machine learning methodologies to investigate environmental and geographic determinants of criminal behavior. Recent research has applied GIS tools to map and analyze spatial variables, such as proximity to infrastructure, including roads, settlements, and checkpoints, as well as terrain characteristics. A discovery from these studies reveals that both terrain and accessibility play critical roles in shaping migration and trafficking routes (Karrachitawaragul, 2016).

Alongside GIS-based studies, research employing spatial and temporal analysis has investigated many crime categories. The spatial-temporal cube models and autocorrelation methods were utilized in this study. Finally, the research showed the specific geographic and temporal patterns that illustrate the variability of particular criminal activities over time and space (Cheng et al., 2022). A separate area of research has concentrated on data-driven methodologies by integrating statistical models with field surveys to evaluate the impact of urban planning features, such as street configurations and commercial density, on the likelihood of street crime. The negative binomial regression was analyzed. This study found that the urban design components impact the probability of street crimes (Zeng et al., 2021).

Remote sensing and high-resolution photography utilization can enhance urban structures' impact comprehension on crime. Researchers have found that diverse and chaotic urban layouts correlate with increased crime rates, corroborating theories such as the "broken windows" hypothesis and Crime Prevention Through Environmental Design (CPTED). The observations indicated the significance of urban planning and architecture in creating secure settings (Patino et al., 2014). Furthermore, the utilization of extensive urban data, encompassing points of interest (POIs) and smart city complaint data has revealed non-linear correlations between urban settings and criminal activity. This study underscores the impact of urban characteristics, including retail density and green space, by CPTED principles, emphasizing their capacity to either alleviate or intensify criminal behavior (Kim & Lee, 2023).

These studies collectively demonstrate a multi-faceted approach to crime analysis, employing spatial, environmental, and machine-learning techniques to produce insights that guide crime prevention and urban planning initiatives. Recent evidence from these research initiatives highlights the importance of spatial analysis and data-driven methodologies in modern criminology. This facilitates a more profound comprehension of the intricate relationships among the built environment, geographic variables, and criminal activity.

Recent technological breakthroughs, particularly in machine learning, have significantly improved crime prevention and border security procedures. This study examines the utilization of several machine learning methodologies to enhance crime forecasting and strengthen border security through the automation of surveillance, intruder identification, and monitoring of unlawful activities. Principal approaches encompass supervised, unsupervised, and reinforcement learning, each offering distinct benefits for augmenting surveillance efficacy and refining crime prevention tactics (Kaur, 2021).

Machine learning applications have been rigorously assessed to determine their efficacy in forecasting criminal behavior. Researchers have discovered major determinants of crime, including time, location, and socioeconomic characteristics, through a comprehensive analysis of crime-related data sources, such as criminal records and demographic and geographical information. This systematic literature review illustrates that machine learning algorithms, including decision trees, support vector machines, neural networks, and clustering methods, provide varied functionalities suited to the data attributes and the particular crime under examination (Jenga et al., 2023).

Furthermore, comparative studies have underscored the efficacy of supervised learning techniques, particularly the Random Forest algorithm, in forecasting crime patterns with high accuracy. The consistent success of supervised models across studies reveals their reliability and practicality, making them highly suitable for crime prediction tasks across various contexts (Alsubayhin et al., 2023). Additional studies utilizing logistic regression, neural networks, and ensemble models applied predictive analysis to identify crime patterns, particularly temporal and spatial factors, to enhance crime forecasting (Anneleen, 2017).

Advances in machine learning have progressively facilitated precise predictions of crime patterns, equipping law enforcement with the knowledge to foresee and alleviate criminal activity efficiently. This study investigates the utilization of diverse machine learning algorithms such as Support Vector Machine (SVM), Random Forest, XGBoost, linear regression, and K-Nearest Neighbors (K-NN) to forecast spatio-temporal crime patterns and evaluate model efficacy across various crime classifications and predictive scenarios. A study specifically examined SVM, Random Forest, and XGBoost in predicting 25 separate criminal categories, revealing that XGBoost attained the maximum accuracy, surpassing the other models in classifying and predicting these crimes (Almuhanna et al., 2021).

Building on this, another study focused on predicting theft using spatiotemporal data, where the efficacy of multiple algorithms (SVM, linear regression, XGBoost, Random Forest, and K-NN) was assessed. XGBoost again demonstrated superior predictive capability, consistently achieving the most accurate results among the models tested (Djon et al., 2023). Finally, a broader study examining general trends in criminal behavior also employed SVM, K-NN, Random Forest, and XGBoost. This study showed that XGBoost and Random Forest delivered excellent accuracy and efficiency, especially when analyzing crime trends over shorter durations (Castro, 2020). Together, these studies illustrate the utility of machine learning algorithms, particularly XGBoost and Random Forest, in delivering high-precision crime predictions across various contexts. With the rising availability of complex spatio-temporal data, machine learning has become an invaluable tool for predicting and understanding crime patterns, aiding law enforcement agencies in implementing proactive measures.

Moreover, by integrating public security, meteorological, points of interest (POI), human mobility, and public service data, studies introduced an innovative temporal and spatial correlation framework (TCP). This model showcased the influence of temporal and spatial patterns on crime prediction accuracy, highlighting the potential of advanced machine learning frameworks to enhance predictive modeling for public safety (Zhao & Tang, 2017).

As global migration rises, comprehending the intricate aspects influencing individuals' migration choices has become imperative. Recently, machine learning methodologies have shown considerable potential in precisely predicting migration trends by analyzing vast social, economic, and demographic datasets. The study employed a machine learning framework incorporating a customized loss function for neural networks and then compared these to conventional migration models. The findings indicated that their machine-learning methodology was superior in forecasting domestic and foreign migrations (Caleb Robinson, 2018). The study utilized Random Forest algorithms to examine a substantial survey. The study identified crucial factors influencing migration decisions and analyzed their impact using survival analysis. The variables such as financial resources and the structure of households are important determinants that can accurately predict migration patterns (Best et al., 2022). It delineates main migration determinants such as financial assets and familial composition and analyzes their influence on migration trends. Collectively, these methodologies seek to elucidate how machine learning might enhance understanding of migration patterns and augment forecast precision in local and international migration scenarios.

This investigation underscores the considerable potential of machine learning to enhance crime prediction and border protection. The accuracy is significantly contingent upon the choice of algorithms, characteristics of the dataset, and contextual factors. Continued study into the integration of varied data sources and the advancement of complex modeling approaches might enhance these applications, positioning machine learning as an essential tool for proactive and adaptive crime prevention efforts worldwide.

2.5 Machine Learning Algorithms

Three machine learning algorithms were employed to predict performance in this study: XGBoost (XGB), Random Forest (RF), and LightGBM (LGBM) algorithms. The selected baseline algorithms enable a comprehensive evaluation of the algorithms across various categories of complexity. The algorithms were also applied to evaluate the explicability of irregular migration in Chiang Rai province, Thailand.

2.5.1 Extreme Gradient Boosting

Extreme Gradient Boosting is the full name of the XGBoost algorithm (Karthikraja et al., 2022) and was developed by Tianqi Chen and Carlos Guestrin in 2016. It is an ensemble Machine Learning (ML) algorithm dependent on decision trees and employs a gradient-boosting method. This algorithm has been guiding many cutting-edge industry applications. XGBoost is a faster algorithm than others because of its parallel and distributed computing. It was designed with a strong emphasis on device optimization and machine learning concepts (Ali et al., 2023). Machine learning methods are used under the gradient boosting framework. In gradient boosting, each subsequent tree is built gradually to reduce the error of the previous tree (Zhang et al., 2020). XGBoost commences with a solitary leaf and incrementally incorporates additional branches into the tree until the optimal split is identified. With XGBoost, simultaneous training of multiple trees is impossible; nevertheless, concurrent creation of individual tree nodes is feasible. XGBoost implements a distributed weighted quantile sketch technique, for facilitating the identification of optimal split points and accommodates weighted datasets (Ali et al., 2023) (Chen & Guestrin, 2016).

2.5.2 Random Forest

Random Forests (RF) algorithm was developed by Leo Breiman. This algorithm is very flexible and capable of addressing categorical response variables in classification tasks and continuous response variables in regression problems (Cutler et al., 2011). Random Forest integrates the Bagging Ensemble Learning Theory with a random subspace approach. Bagging ensemble learning technology trains many decision trees, amalgamating them into RF. In data classification issues, the classification outcomes are dictated by the mode of the results produced by all decision trees (Zhang et al., 2022) Moreover, RF can utilize both categorical and continuous predictor variables, allowing them to be well-suited for a broad spectrum of data types and analytical applications. RF constructs a large ensemble of decision trees during the training phase. It predicts the most frequent class among the trees for classification tasks. The regression tasks compute the average prediction across the trees. It is effective when dealing with datasets that exhibit non-linear relationships between variables (Singh et al., 2022).

A major advantage of the RF algorithm is its high accuracy, especially when working with small or limited datasets. The potential of the RF algorithm is to handle complex data sources and uncover subtle interactions between factors, which is an excellent choice for challenging analytical scenarios However, its performance can degrade in high-dimensional datasets, where the increased complexity may lead to overfitting and reduced predictive power (Kotsiopoulos et al., 2021).

2.5.3 LightGBM

Gradient boosting is a powerful method with elevated predictive accuracy. The protracted computing time has limited its applications in predicting extensive compound libraries or creating in silico predictive models requiring regular retraining. LightGBM, a recent improvement of the gradient boosting algorithm, maintains its predictive solid capability while addressing scalability and extended computing duration by implementing a leaf-wise tree expansion strategy and introducing innovative techniques (Zhang et al., 2019). LightGBM is a new gradient-boosting framework based on decision tree algorithms introduced by Microsoft. This algorithm supports many algorithms, such as GBM, GBDT, Gradient Boosted Regression Tree (GBRT), and Multiple Additive Regression Tree (MART). The result of LightGBM is scalable, accurate, and efficient (Alkasassbeh et al., 2020).

2.6 Summary of This Chapter

This chapter is divided into four parts. First is the background of the study, explaining the irregular migration situation in Thailand and the factors influencing migration. Migration depends on different situations and conditions in countries of origin and destination. The second part mentions Geoinformatics, including Geographic Information Systems (GIS), Remote Sensing (RS), and Global Navigation Satellite Systems (GNSS). Geoinformatics is a crucial tool for collecting, managing, and analyzing spatial data, providing essential data to understand the drives of irregular migration.

The third part of this chapter is about environmental factors, with nine factors related to this study consisting of indices: Normalized Difference Vegetation Index (NDVI) and the Bare Soil Index (BSI) to evaluate vegetation health and land cover. Digital Elevation Model (DEM) to examine the Earth's physical attributes. Nighttime light data measures human activity through artificial light emissions and population data from LandScan Global, which analyzes the relationship between population density, human activity, and migration patterns. In addition, LULC, river, road, and slope were also mentioned in this study to provide a deeper understanding of each environmental factor related to migration flows.

The related studies applying machine learning and geoinformatics to examine migration and crime patterns were put at the end of this chapter. It introduces three machine learning algorithms utilized in the study, including XGBoost, Random Forest, and LightGBM, to predict irregular migration in Chiang Rai province, Thailand. The selected algorithms will be applied for predictive accuracy, and migration patterns based on environmental and geographic data will be interpreted, preparing for further analysis in the following chapters. Table 1 shows the summaries of some related works which include several factors and methods.

Table 1 Summaries of Some Related Work

Year	Factors	Method	Reference
2010	Population, Green Space	Mixed-methods approach	(van der Geest et al., 2010)
2018	Population, Green Space, Elevation	Traditional spatial analysis methods	(Small et al., 2018)
2023	Climate, Elevation, LULC	RF, SPRUCE	(Pless et al., 2023)
2022	Population, Economic Indicators, Geographic Variables, Environment-Related Factors	RF, Survival Analysis	(Best et al., 2022)
2016	Proximity to roads, settlements, checkpoints, Slope, LULC	GIS	(Karrachitawaragul, 2016)
2021	Road, LULC, Environmental Features	GIS, Negative binomial regression	(Zeng et al., 2021)
2014	Human Mobility, Vegetation Dynamics	RS, Census Data Analysis, Cross-Analysis	(Patino et al., 2014)
2023	POI, Road, LULC, Population	LightGBM, SHAP	(Kim & Lee, 2023)
2021	-	SVM, RF, XGBoost, Linear regression, K-NN	(Almuhanna et al., 2021)

CHAPTER 3 RESEARCH METHODOLOGY

This chapter includes a general background of the study area, workflow, data collection, preprocessing data, machine learning algorithms, and algorithm accuracy comparison. The methodology of this study contains XGBoost, Random Forest, and LightGBM algorithms using remote sensing data from Sentinel 2 and irregular migration arrest point data from an agency responsible for ensuring the security and control of Thailand borders for the calculation of indices. All steps will be detailed in the following.

3.1 General Background of the Study Area

Chiang Rai province is located in upper northern Thailand. Thailand's northernmost province is approximately 19 degrees north to 20 degrees 30 minutes north latitude and from 99 degrees 15 minutes east to 100 degrees 45 minutes east longitude. Its location is around 829 kilometers from Bangkok, the capital of Thailand. The area is approximately 11,678.369 square meters. The north is connected to The Republic of the Union of Myanmar (Myanmar) and the Lao People's Democratic Republic (Laos). The south is connected to Lampang and Phayao provinces. The east is connected to the Laos and Phayao Provinces. The west is connected to Myanmar and Chiang Mai Province. The border distance between Chiang Rai province and Myanmar is 153 kilometers, and the Chiang Rai border connects to Laos at 155 kilometers.

Chiang Rai experiences three distinct seasons. Summer generally runs from mid-February until mid-May, averaging 30.6 degrees Celsius, and the highest temperature reaches 35.4 degrees Celsius. The rainy season extends from mid-May to mid-October, with 145 rainy days and a total annual precipitation of 2,042.6 millimeters. Winter occurs between mid-October and mid-February. The average lowest temperature is 15.1 degrees Celsius. The Minimum temperature is 10.2 degrees Celsius (Rai, 2022). Chiang Rai is divided into 18 districts. The districts are divided into 124 sub-districts and 1,751 villages. The population of Chiang Rai is 1,298,977 people in 2023 (Office, 2024).

The Golden Triangle is where the borders of Thailand, Laos, and Myanmar converge. The Mekong River separates this province from Laos; Mae Sai River and Ruak River separate it from Myanmar. The main river flows through Chiang Rai's town Kok River. The eastern region of the province is predominantly a flat plain, whereas the western region is characterized by mountainous terrain. The province's average elevation is 580 meters, with a maximum elevation of 1,998 meters, with a maximum elevation of 1,998 meters (Akber & Shrestha, 2015).

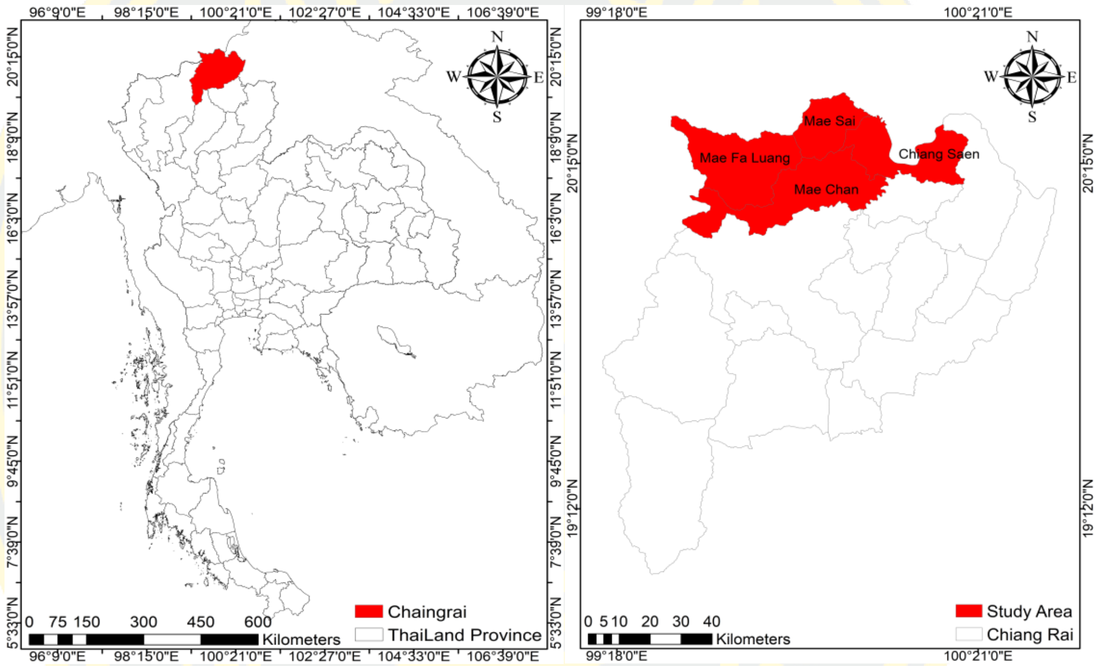

Figure 2 The Study Area

Chiang Rai Province is located in a strategically significant region, connected to neighboring countries, and has numerous natural entry and exit points. The population has ethnic diversity, including a hidden population from surrounding countries seeking employment, resulting in a multicultural society, which is prominent in tourism trade and investment. It is an important contribution to issues of illegal immigration by foreign workers, drug trafficking, and transnational crime (Office of Social Development and Human Security, 2021).

The government record shows that four districts of Chiang Rai along the Thai-Myanmar and Laos border, Mae Chan, Chiang Saen, Mae Sai, and Mae Fa Luang districts, have a large amount of irregular migration traveling through natural channels, so this study will focus on these study areas.

3.2 Workflow of the study

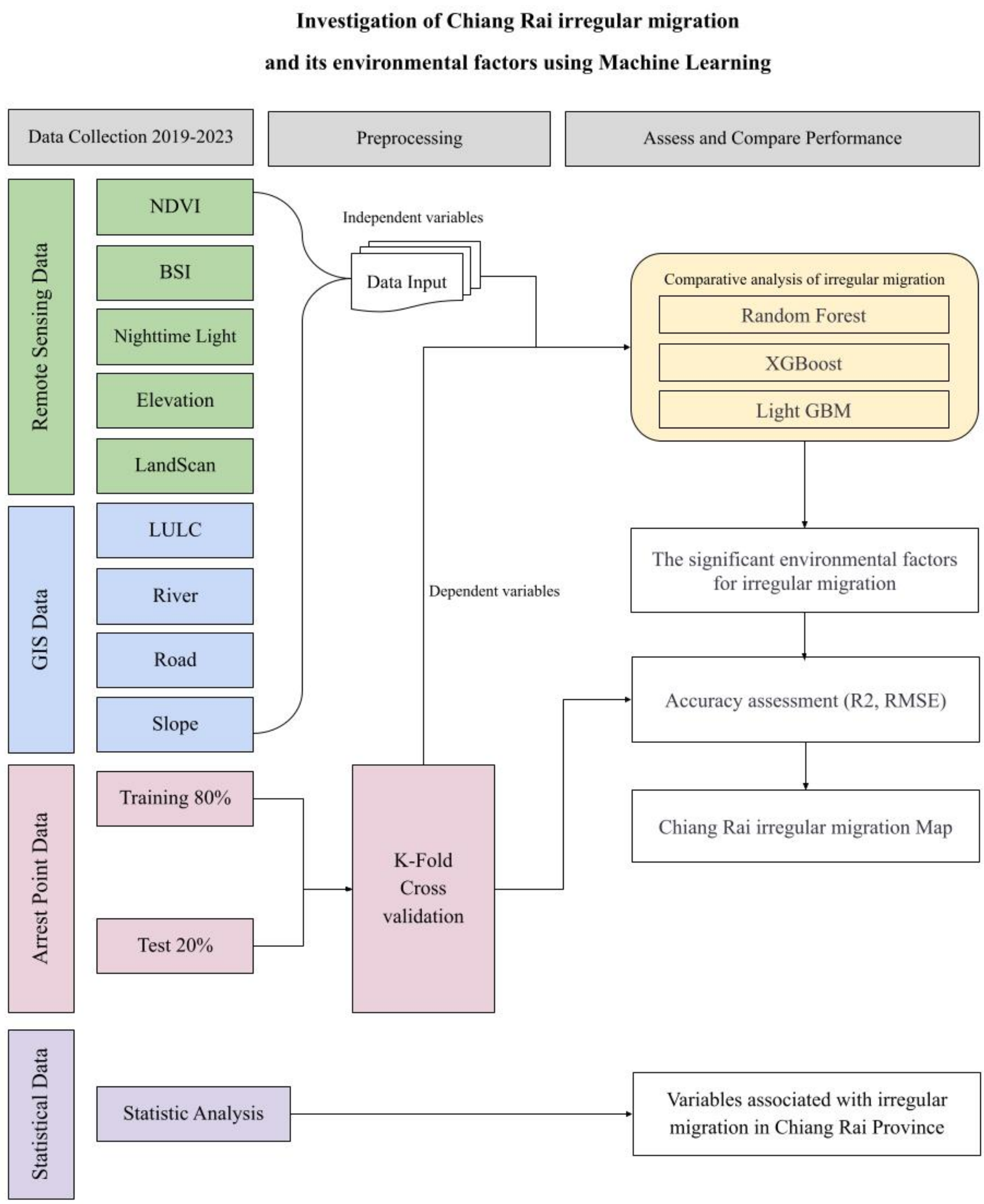

Figure 3 The Technical Framework of This Study

The general idea of this study is shown in Figure 3. This study examines the environmental and geographic determinants affecting irregular migration in Chiang Rai Province, Thailand, specifically along the borders with Myanmar and Laos, from 2019 to 2023. The process employs a systematic approach, commencing with extensive data acquisition from many sources. Remote sensing datasets, encompassing NDVI, BSI, nighttime light data, elevation, and LandScan population data, are integrated with GIS datasets, including land use/land cover, river networks, road networks, and slope data. Furthermore, irregular movement data, indicated by arrest locations, is included to document migration occurrences within the study region.

In the preprocessing step, the datasets are structured to isolate independent variables based on environmental and geographic parameters. In contrast, the dependent variable, representing migratory occurrences, is sourced from the arrest point data. The dataset is divided into two parts, training (80%) and testing (20%) subsets to facilitate model validation.

Three machine learning algorithms, including Random Forest, XGBoost, and LightGBM, are utilized to study the data and evaluate the relative significance of components affecting irregular migration. A K-fold cross-validation process is employed to improve algorithm dependability, and algorithm performance is assessed using R-squared (R^2) and root mean square error (RMSE) metrics. The algorithm with the greatest prediction accuracy is employed to ascertain the region's principal environmental and spatial determinants of irregular migration. A risk map of irregular migration for Chiang Rai Province has been created, providing a geographical depiction of high-risk zones and movement hotspots. Statistical data from irregular migration records will be also analyzed to find variables associated with irregular migration in the study area.

3.3 Data Collections

The research study utilized Remote Sensing data, Geographic Information data (GIS), and irregular migration point data to investigate irregular migration and its environmental factors in Chiang Rai province from 2019-2023. The details are as follows:

3.3.1 Remote Sensing Data

The study employed high-resolution remote sensing data from the Sentinel-2 satellite to examine environmental factors influencing irregular migration patterns in Chiang Rai province. Sentinel-2, with a spatial resolution of 10 meters, provided detailed imagery essential for capturing fine-scale information on land cover and soil conditions across the study area. From 2019 to 2023, this five-year dataset allowed for a longitudinal analysis of environmental dynamics and their potential connections to migration trends.

The Harmonized Sentinel-2 MSI (MultiSpectral Instrument) product facilitated the extraction of critical spectral indices. Four bands that are susceptible to vegetation and soil characteristics were utilized: Band 8 (near-infrared), Band 4 (red), Band 2 (blue), and Band 11 (shortwave infrared). Two indices derived from this data, Normalized Difference Vegetation Index (NDVI) and the Bare Soil Index (BSI), were central to the study's environmental analysis. NDVI, calculated from the near-infrared and red bands, measures vegetation health and density, with higher values indicating denser, healthier vegetation and lower values indicating sparse or degraded areas. The BSI, calculated using visible and shortwave infrared bands, provides insights into bare soil exposure and land degradation, offering valuable information about land use changes that may impact migration patterns.

Additionally, the study incorporated a digital elevation model (DEM) from the COPERNICUS/DEM/GLO30 dataset to analyze the study area's topographic characteristics. With an approximate spatial resolution of 30 meters, the DEM was resampled using bilinear interpolation to meet the analysis's spatial requirements. This topographic data enabled a more precise assessment of potential migration routes and obstacles within the region's varied landscape.

Nighttime Light data captures artificial lighting levels at night from the NOAA/VIIRS/DNB/MONTHLY_V1/VCMSLCFG dataset. It offers insights into human activity and settlement patterns, in which brighter lighting areas are often associated with higher economic activity or population density. Nighttime light provides a valuable dataset for identifying potential transit zones or destinations for migrants. It highlights economic and social factors that impact migration trends in densely populated areas.

LandScan Global Population Data at 1 km resolution was utilized to assess population distribution across the area. LandScan provides solution population estimates for evaluating how population density interacts with environmental factors that may drive or restrict migration. This study overlapped the population data and other environmental variables to understand the interaction between human distribution and factors collectively shaping irregular migration dynamics.

In this study, Sentinel-2 imagery, DEM, nighttime light, and population data were used to spatially analyze environmental factors and irregular migration incidents in the study area. This approach provides land cover mapping, soil conditions, topography, human activity, and population distribution. These datasets and indices provide detailed information on the environmental factors of migration.

3.3.2 Geographic Information System data

The GIS datasets utilized in this study are essential for analyzing the spatial and environmental aspects related to irregular migration in Chiang Rai province, situated near the Thai-Myanmar and Laos borders. These databases offer significant insights into the terrain attributes, infrastructure, and natural elements that affect migration patterns. Primary data sources encompass Land Use/Land Cover (LULC) statistics obtained from the Land Development Department of Thailand, which provides comprehensive classifications of land utilization within the study area, including agricultural, residential, forested, and more categories. This dataset is crucial for pinpointing regions that may either promote or obstruct migration, as various land cover types influence accessibility, visibility, and appropriateness for cross-border transit. By comprehending the variations in land usage throughout the region, it becomes feasible to identify particular locations where migration is more probable.

Data on rivers, obtained from governmental open data sources, offers an extensive overview of river networks within the research field. Comprehending the location and attributes of river networks aids in recognizing potential routes and barriers for irregular migration. Likewise, road data, sourced from government open data repositories, delineates the current transportation infrastructure, encompassing primary highways, secondary roads, and access points. Roads are essential in establishing migration routes, as they provide accessible avenues for migrants to traverse the terrain.

Slope data offers insights into the topography and gradient of the terrain within the study area. Derived from governmental data systems, slope data is especially pertinent in areas characterized by difficult or mountainous terrain, such as those next to the Thai-Myanmar and Laos borders.

All datasets from 2019 to 2023 were processed and analyzed utilizing ArcMap, a GIS software often employed for geographic data management and analysis. ArcMap facilitated the integration and superimposition of several data layers, permitting comprehensive geographical analysis of migratory trends within the study area. The analysis identified possible migration hotspots, routes, and barriers by integrating land use, river, road, and slope data.

This GIS-based analysis offers a thorough spatial comprehension of the environmental and infrastructure elements affecting irregular migration in Chiang Rai Province. The amalgamation of various statistics facilitates a multi-faceted perspective of the study area, guiding focused border management tactics by pinpointing regions particularly susceptible to irregular movement. This methodology is essential for improving the efficacy of monitoring and intervention initiatives in areas where physical and infrastructural elements influence migration patterns.

3.3.3 Arrest Points Data

The study employed an extensive dataset of arrest locations linked to illegal migration events, diligently documented by government border control officials in Chiang Rai Province throughout five years (2019-2023). This information is crucial for examining the spatial, temporal, and demographic trends of irregular migration along the Thai-Myanmar and Laos borders. Arrest point data collection encompasses critical information regarding each occurrence, including the date, time, and exact location, along with demographic variables such as the gender and nationality of the apprehended individuals. These variables establish a solid foundation for examining the determinants of irregular migration and recognizing significant trends in migration dynamics within this region.

The dataset, recorded and managed in Microsoft Excel, is systematically organized to ensure effective data management and processing. To uphold credibility and consistency, the dataset was obtained from authenticated government sources, therefore guaranteeing the veracity of the analysis. Each variable provides distinct insights: temporal data (date and time) provides the analysis of seasonal and daily trends, indicating periods of heightened or diminished migration activity, whilst spatial data permits the identification of regional hotspots for irregular migration. Demographic factors, including sex and nationality, augment the dataset's analytical depth by revealing population-specific trends and identifying groups potentially predisposed to irregular movement.

This dataset of arrest points performed two functions in the study: it facilitated the descriptive analysis of irregular migratory patterns and provided the basis for training and testing machine learning algorithms. Machine learning algorithms depend on extensive, high-quality data to generate precise predictions and identify significant patterns. This study enhanced algorithm training by providing arrest point data, which offered a thorough perspective on historical migration episodes, allowing algorithms to learn from empirical data. Incorporating characteristics from the arrest dataset into these algorithms sought to enhance their predictive precision in identifying risk factors and high-risk areas for irregular migration.

The arrest point dataset is fundamental to this work, facilitating comprehensive descriptive analysis and the creation of predictive models for irregular migration. The integration facilitates a comprehensive analysis of spatial, temporal, and demographic variables affecting migration patterns, aiding the development of machine learning algorithms that can guide targeted border control operations and offer insights into movement trends in Chiang R ai Province. The data collection is shown in Table 2.

Table 2 Data Collection

Type Sources	Dataset	Index	Spatial Resolution	Period (Year)	Source
Remote Sensing	Sentinel-2	NDVI	10 meters	2019-2023	Sentinel-2 (Harmonized)
		BSI			
	GLO-30	Digital Elevation Model	Resampled		COPERNICUS/DEM/GLO30
	VIIRS	Nighttime Light			NOAA/VIIRS/DNB/MONTHLY_V1
	Land Scan Population	Population Data			LandScan Population Database
GIS	LULC	-	10 meters		Land Development Department
	River	-			Government Open Data
	Road	-			
	Slope	-			
Arrest point data	Arrest points of irregular migration in the study area	-	-		Border control agency

3.4 Preprocessing Data

Each index and data were processed separately in this stage, with the data resampled to a consistent spatial resolution. The indices are then merged into a composite image using the add Bands function in Google Earth Engine. This composite image is subsequently exported as a GeoTIFF file for further analysis. GIS data and arrest point data were integrated and analyzed by ArcMap to illustrate environmental conditions in the study area for a comprehensive spatial assessment.

3.4.1 Normalized Difference Vegetation Index

Normalized Difference Vegetation Index (NDVI) was calculated using Sentinel-2 data from the COPERNICUS/S2_SR_HARMONIZED dataset, specifically utilizing bands 4 (red) and 8 (near-infrared). NDVI is a well-established metric for assessing vegetation health and density. The normalization process was carried out in Google Earth Engine, adjusting the NDVI values to a standardized range between 0 and 1. This analysis covered the period from 2019 to 2023, providing a comprehensive temporal overview. NDVI was further employed to assess forest density and evaluate the environmental conditions that may influence the accessibility of areas by irregular migrants. The Normalized Difference Vegetation Index in the study area is displayed in Figure 4.

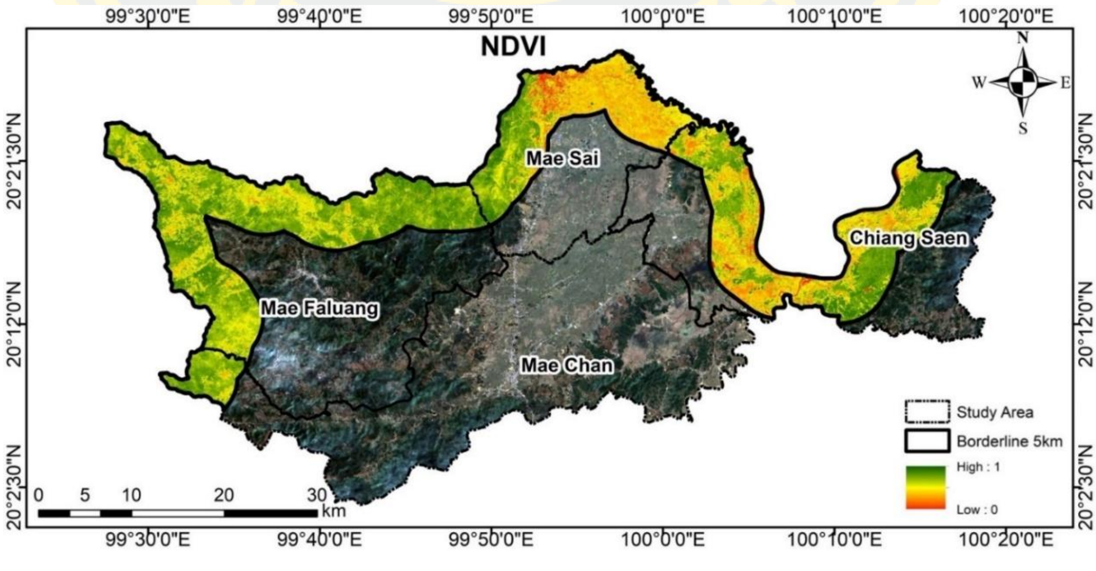

Figure 4 he Normalized Difference Vegetation Index in the Study Area

3.4.2 Bare Soil Index

The Bare Soil Index (BSI) was computed with Sentinel-2 data from the COPERNICUS/S2_SR_HARMONIZED dataset. The BSI formula consists of visible, red, blue, near-infrared (NIR), and shortwave infrared (SWIR) bands, employing Sentinel-2 bands 2 (blue), 4 (red), 8 (NIR), and 11 (SWIR). The data was normalized and resampled in computation in Google Earth Engine to ensure a consistent spatial resolution, adjusting the BSI values to a standardized range between 0 and 1. This analysis presents a temporal evaluation of bare soil conditions during the study period to evaluate environmental factors that influence irregular border crossings. The data of Bare Soil Index in the study area is displayed in Figure 5.

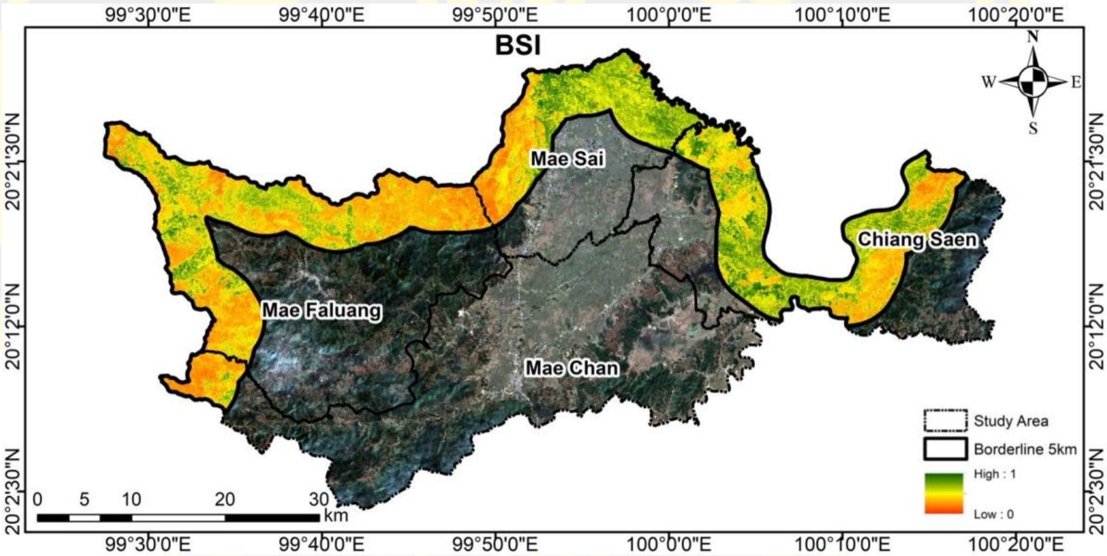

Figure 5 The Bare Soil Index in the Study Area

3.4.3 Nighttime Light Data

The nighttime light data originated from NOAA/VIIRS/DNB/MONTHLY_V1 dataset, providing global observations of nighttime illumination. This dataset utilizes the Day/Night Band (DNB) of the Visible Infrared Imaging Radiometer Suite (VIIRS) to record nighttime light signals, with a spatial resolution of approximately 500 meters. The data were processed in Google Earth Engine, where they were calibrated and normalized to ensure temporal consistency. The investigation covered the years 2019 to 2023, providing insights into the spatial and temporal distribution of human activity via light intensity. These data are essential for investigating how the presence or absence of nightlights may indicate patterns of irregular migration activity, particularly in border regions where such movement is more likely to occur. The data of Nighttime Light in the study area is displayed in Figure 6.

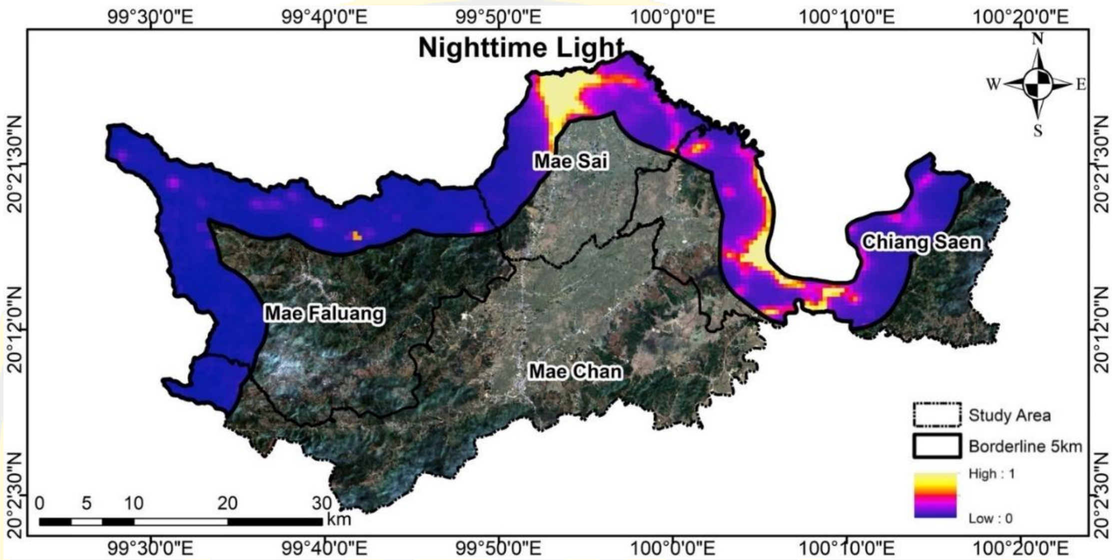

Figure 6 Nighttime Light Data in the Study Area

3.4.4 Digital Elevation Model

Digital Elevation Model (DEM) data was sourced from the COPERNICUS/DEM/GLO30 collection, which provides global elevation information with a spatial resolution of 30 meters. Google Earth Engine was utilized to resample the elevation data by bilinear interpolation, ensuring stability throughout the study area. Then, the normalization scale range of values to a standardized scale of 0 to 1. The DEM data from 2019-2023 were applied in the study which will provide more understanding of topography and its impact on environmental conditions. The DEM data of the study area is shown in Figure 7.

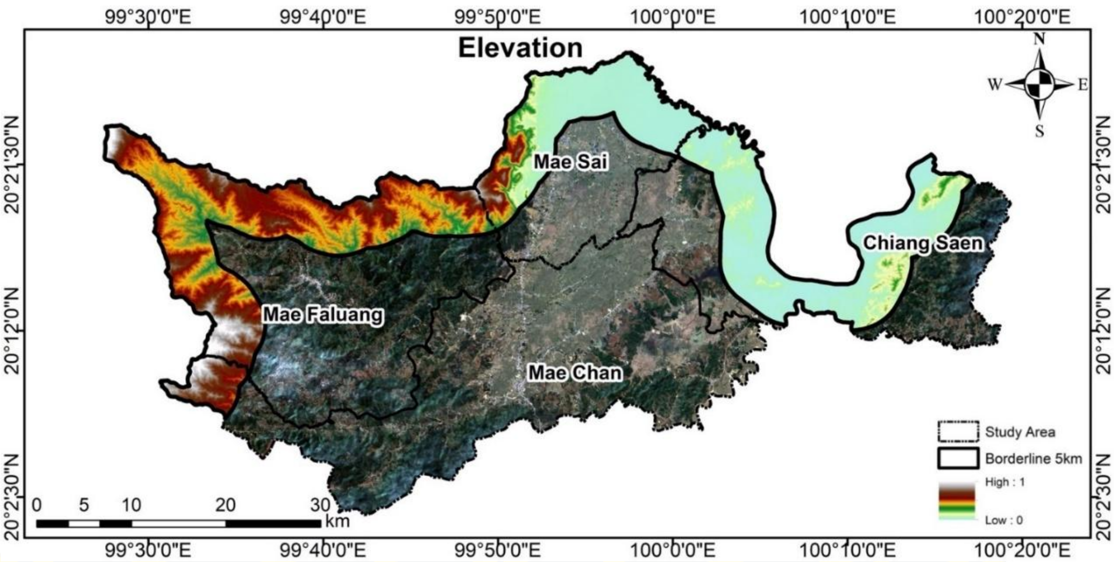

Figure 7 Digital Elevation Model in the Study Area

3.4.5 LandScan Global Population Data

The LandScan Global Population Data was applied to evaluate population distribution and density within the study area. This dataset, generated by the Oak Ridge National Laboratory, provides high-resolution population estimates at a spatial resolution of 1 km², indicating the average population. The LandScan data was integrated into research using Google Earth Engine, enabling an in-depth examination of human settlement patterns associated with environmental factors causing irregular migration. The dataset covers from 2019 to 2023 and was essential for identifying areas of elevated population density that may correlate with heightened migration activities. This investigation focused on clarifying the relationship between population distribution and environmental and geographical data, focusing on the relation between demographic parameters and irregular migration patterns along the border. Employing the bilinear resampling method. Subsequently, normalize the range of values to a standardized scale of 0 to 1. LandScan Population data in the study area is shown in Figure 9, Dark colors signify high population density in the study area.

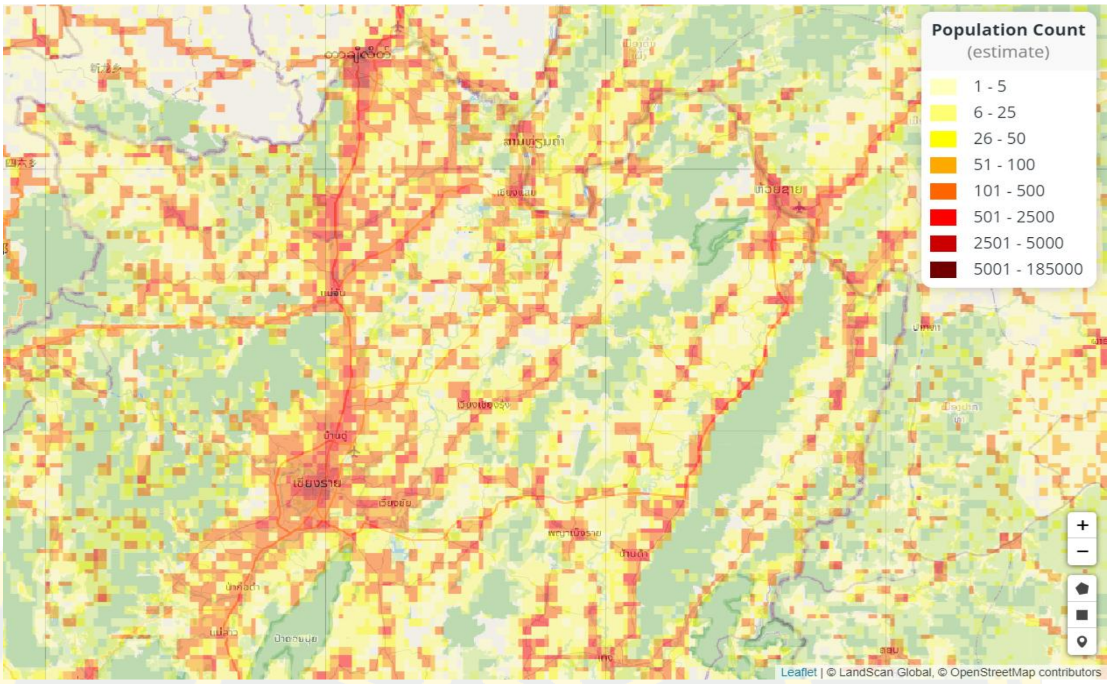

Figure 8 LandScan Data of Population Density in Chiang Rai, Thailand

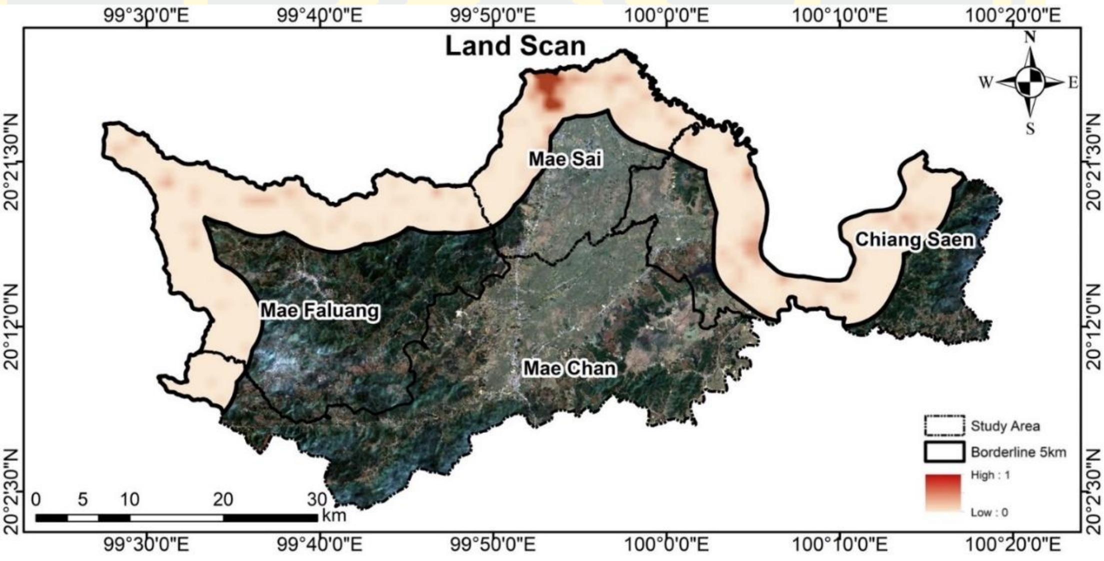

Figure 9 LandScan Population Data in the Study Area

3.4.6 Land Use Land Cover

The Land Use/Land Cover (LULC) data was sourced from the land development department's open data. This dataset classifies land cover into specific categories, such as forests, agriculture, wetlands, and urban regions, ensuring the examination of how different land uses could impact irregular migrating patterns. The LULC data was examined in Google Earth Engine, from 2019 to 2023. This study intended to discover how different land use configurations affect migration trends by evaluating the relationship between land cover types and migration in border areas. The integration of LULC data and further spatial data offered a comprehensive perspective on environmental factors influencing irregular movement dynamics.

LULC data are displayed in Figure 10, characterized by five colors representing five zones. The agricultural zone is shown in yellow, which covers most of the land in the study area. Green represents the forest area, red represents the urban area, the miscellaneous area is shown in gray, and the water area is shown in blue color.

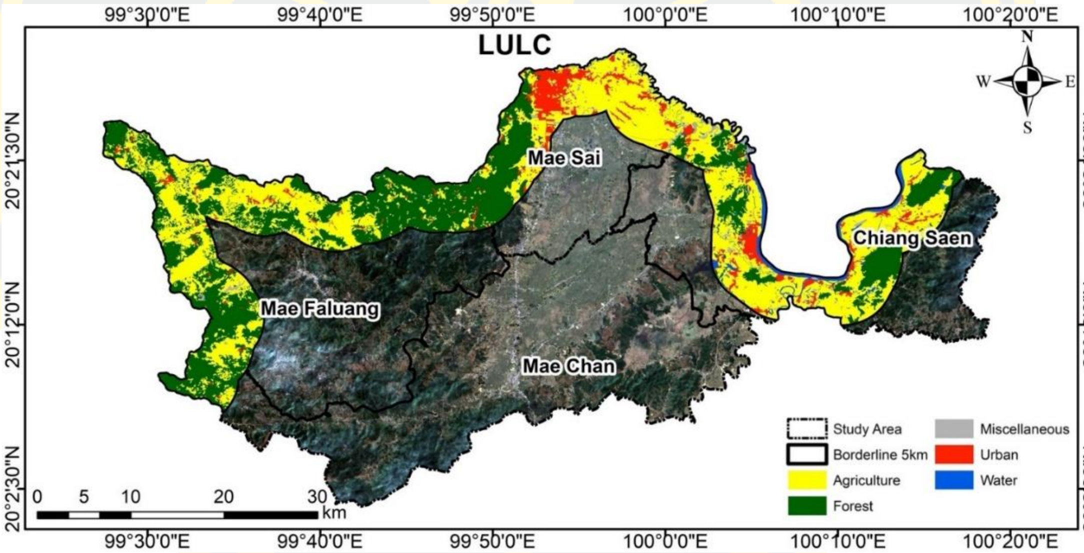

Figure 10 Land Use Land Cover in the Study Area

3.4.7 River Data

The river data were sourced from government open data to study the hydrological features of the study area. This dataset includes information on river locations, widths, and flow characteristics for understanding how waterways can serve as natural barriers or pathways for movement. The river data were integrated into the analysis using ArcMap for spatial assessment of the correlation between river adjacency and migration routes. The analysis covered the period from 2019 to 2023 to discover and understand how seasonal variations in river conditions impact migration behavior in the study area. The river network data in the study area are shown in Figure 11, the blue line represents waterways.

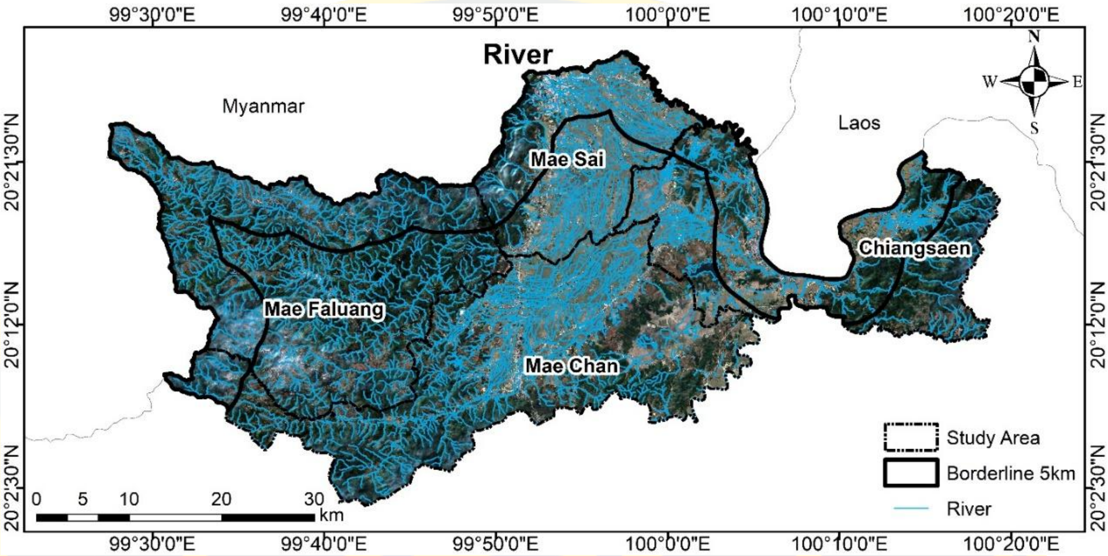

Figure 11 River Data in the Study Area

The map in Figure 12 shows river density across the study area, employing Kernel Density Estimation (KDE). KDE is a statistical technique for estimating the spatial density of points or events within a specified area or network. This methodology is extensively employed in spatial analysis to identify "hot spots" or clusters, emphasizing areas with a significant concentration of events or activities. Areas with a higher concentration of points demonstrate higher density, creating a gradient from high-density to low-density areas. KDE provides spatial distribution patterns that may not be immediately apparent from raw point data (Okabe et al., 2009).

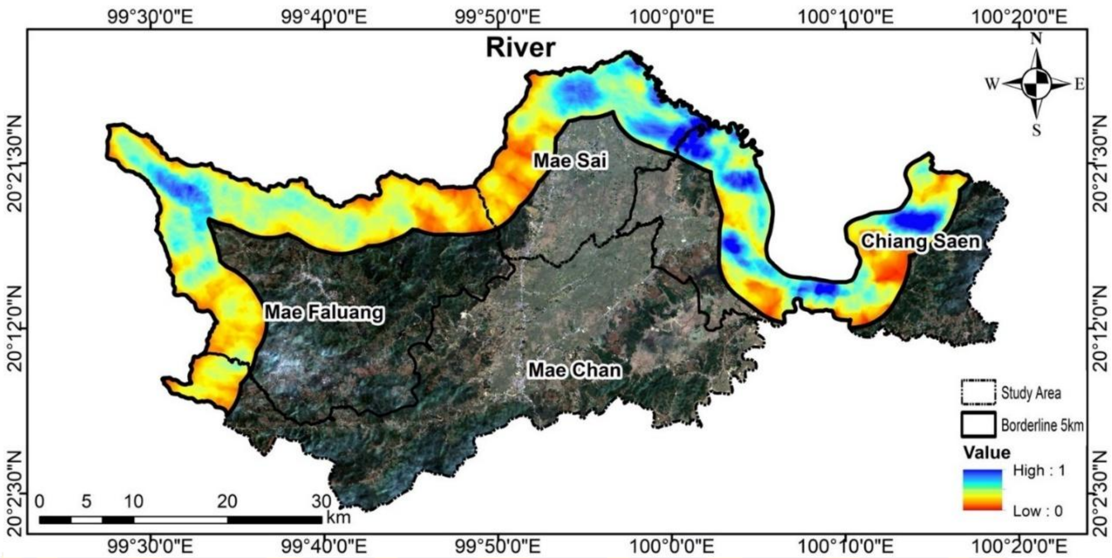

Figure 12 Kernel Density of River

The color gradient on the map illustrates river density, with cooler colors (blue) representing higher densities and warmer colors (yellow to orange) signifying lower density values. The high-density areas, indicated in blue, are concentrated around the northern border areas, especially along the northern Myanmar and Laos borders. These regions are marked by a dense river network.

3.4.8 Road Data

The road data were obtained from government open data sources to analyze the distribution and density of transportation infrastructure within the study area. This dataset includes comprehensive information on various road types, such as highways, local roads, and trails. Geographic Information System (GIS) techniques were used to analyze road density, and the total length of roads was calculated within specified locations to generate a spatial density map. This analysis, covering from 2019 to 2023, analyzes the impact of road density on irregular movement patterns. The road network data map of the study area is represented in Figure 13; orange lines represent the road network.

Figure 14 displays the road density in the study area. This density is estimated using Kernel Density Estimation (KDE). The KDE method provides a distinct visual depiction of regions with higher concentrations and accessibility of road networks.

The darker portions on the map signify areas of elevated road density, whereas the lighter regions indicate lower density. The map indicates that the districts of Mae Sai and Chiang Saen have elevated road density, as shown by darker colors. On the contrary, Mae Fa Luang and certain areas of Mae Chan show lower road density, as seen by lighter colors. These regions are mainly rural and mountainous, resulting in restricting road accessibility.

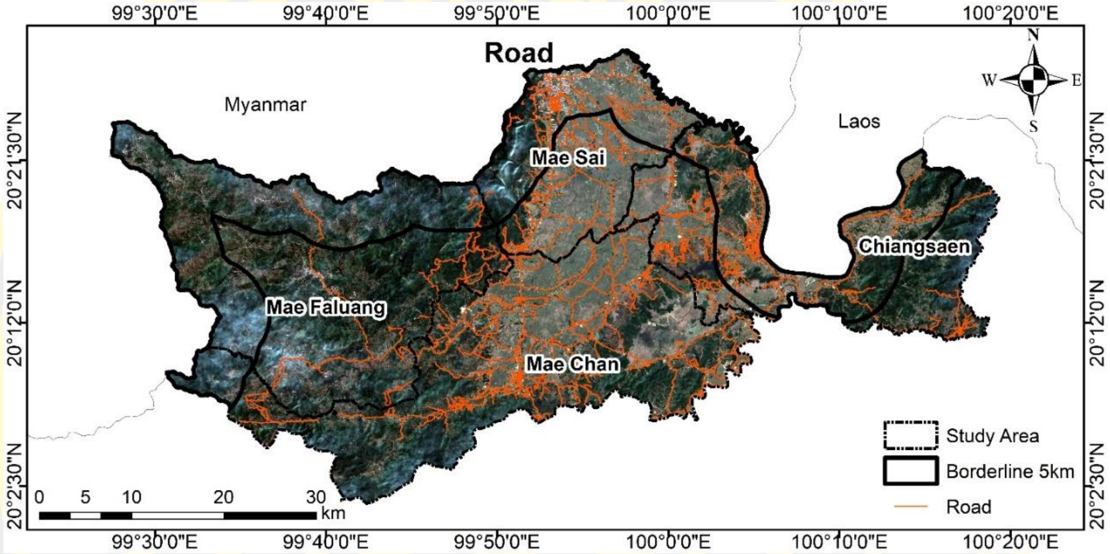

Figure 13 Road Network Data in the Study Area

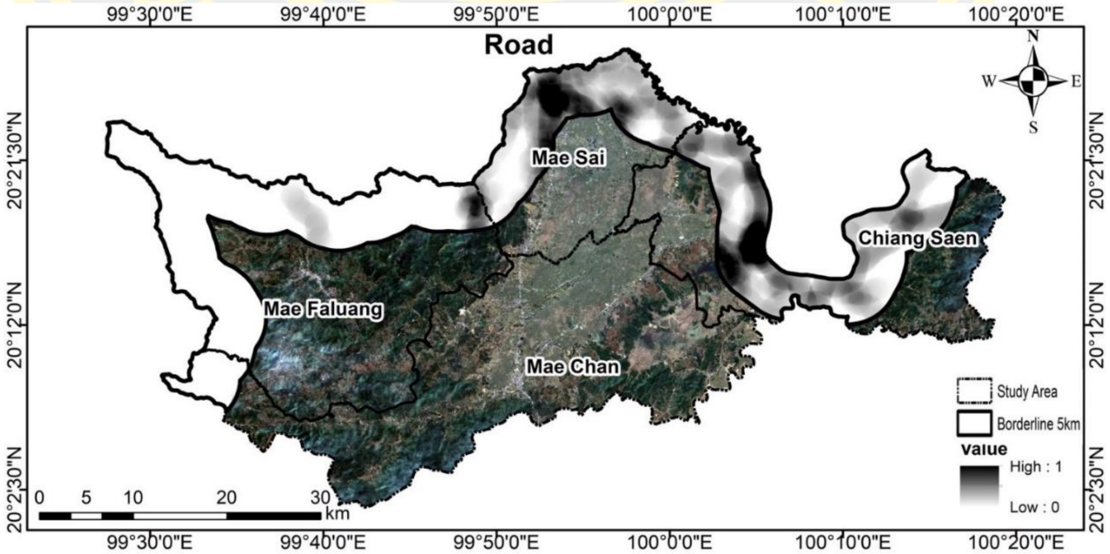

Figure 14 Kernel Density of Road

3.4.9 Slope Data

Slope data was gathered from official open data sources, it provided topographical information of the study area. This dataset was generated by Digital Elevation Models (DEMs) and includes measurements of terrain inclination, expressed in degrees or as a percentage. The slope analysis was conducted using Geographic Information System (GIS) techniques such as ArcMap to calculate slope gradients across the study area, from 2019-2023. This study will show how variations in terrain slope impact irregular migration patterns. The slope data is shown in Figure 15.

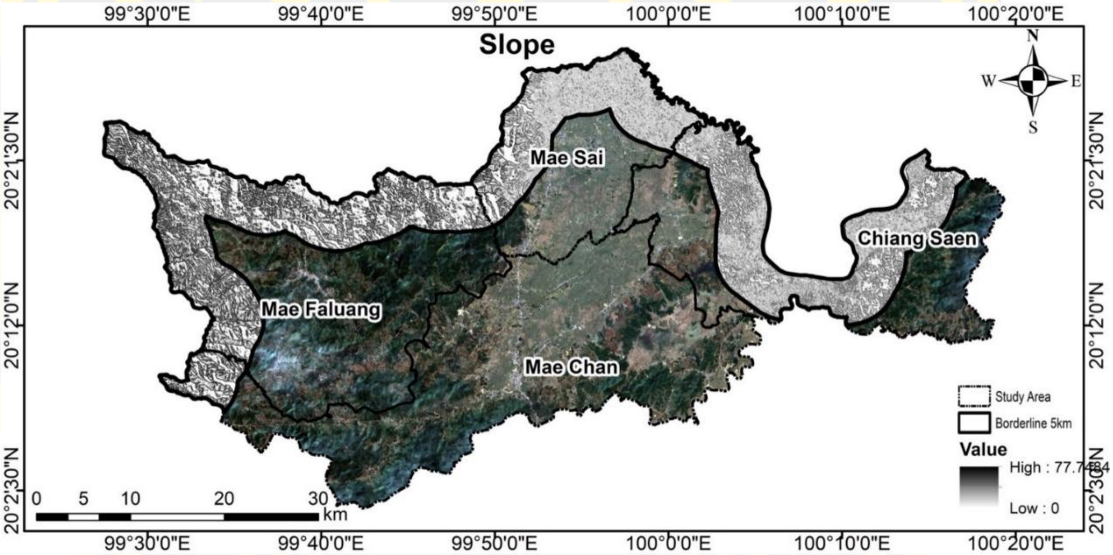

Figure 15 Slope Data in the Study Area

3.4.10 Arrest Point Data

The Arrest point data were collected from related authority records who were in charge of border control. The record detailed the locations of irregular immigration arrests along the Chiang Rai province, Thailand border. This data sets a thorough data cleaning data to and reliability before being converted into a CSV format for further analysis. Once in the appropriate format, the data were imported into ArcMap, where the arrest points were transformed into spatial features for analysis. A 5-kilometer buffer zone was then created along the borderline to delineate the study area, focusing on the investigation of migration patterns irregularly. This buffered dataset served as an input for sampling in machine learning algorithms, enabling the generation of random points in surrounding areas for comparative analysis. Figure 16 illustrates the arrest points, whereas Figure 17 depicts the survey of irregular migration arrest points.

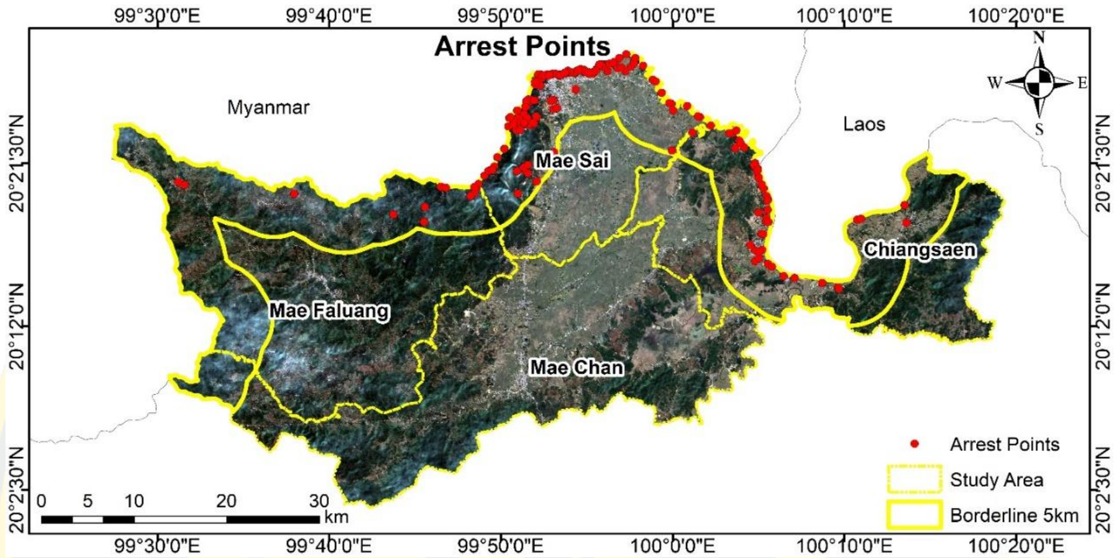

Figure 16 Arrest Point of Irregular Migration

Figure 17 The Survey of Irregular Migration Arrest Points in the Study Area

3.5 Predictive Algorithms

In this study, machine learning algorithms consist of Extreme Gradient Boosting (XGBoost), Random Forest (RF), and LightGBM, which are implemented in Python 3.11 via the integration of libraries such as GDAL, Numpy, Matplotlib, Scikit-learn, and the XGB Python Library within the Jupyter Notebook environment, which choose to predict and estimate to obtain effective results and calculate feature importance from each model using the machine learning permutation importance calculation technique to rank each variable relevance.

3.5.1 Extreme Gradient Boosting

XGBoost is a strong supervised learning technique that applies the gradient boosting framework to iteratively create decision trees, with each tree correcting the errors of its predecessor. This method generates a highly precise and robust model. XGBoost is used to examine environmental and geographical elements affecting irregular migration, improving its predictions by clarifying migration patterns.

The first step is to employ the Geospatial Data Abstraction Library (GDAL) to process feature data (X), including terrain, border proximity, and socio-economic conditions, with the target variable (y) denoting migration events. GDAL guarantees uniform array generation based on raster dataset attributes, with pixel values retained in a NumPy array for analysis.

The features (X) and labels (y) are combined into a cohesive 2-D array (DataX) with the NumPy toolkit. The train_test_split function from sci-kit-learn divides the dataset into 80% for training and 20% for testing, allowing validation of the model's performance on unseen data.

During the third step, the XGBoost algorithm is instantiated, and cross-validation measures are established for assessing performance through training. The algorithm is trained using the fit approach, methodically building decision trees to reduce prediction errors and clarify the connections between environmental variables and migration. The algorithm's performance is evaluated using the test data after training.

The trained XGBoost algorithm finally predicts irregular migration across the study area, and the results are exported into a new raster file.

3.5.2 Random Forest Regression

Random Forest is a method of ensemble learning that independently and continuously creates numerous decision trees, effectively mitigating the danger of overfitting. This makes it suitable for examining complex data, such as irregular patterns of migration affected by environmental and topographical variables.

The method begins with the ingestion of environmental feature data (X) and the goal variable (y) denoting migration occurrences, using the Geospatial Data Abstraction Library (GDAL). The raster datasets, including satellite or geographic data, are arranged and standardized for study. Pixel values from each band are obtained and stored in a NumPy array after that.

The feature (X) and target (y) arrays are combined into a singular 2-D array utilizing NumPy. This integrates predictor variables with migration incident labels. The dataset is divided into training (80%) and testing (20%) subsets via the train_test_split tool from sci-kit-learn, providing the validation of the model's predicted performance on unseen data.

After preparing the data, a Random Forest algorithm is instantiated and trained on the training dataset through the fit approach. The model establishes the relationship between environmental variables and migration incidents, with its efficacy assessed using the test data.

The trained Random Forest algorithm eventually predicts migration patterns over the research region and exports these predictions into a new raster file.

3.5.3 LightGBM Regression

LightGBM is a rapid and scalable gradient-boosting algorithm that creates trees in a leaf-wise manner, enabling faster convergence and increased accuracy, making it suitable for managing extensive data sets associated with environmental and migration factors.

The process starts with data ingestion using the Geospatial Data Abstraction Library (GDAL), which evaluates the feature data (X) indicative of environmental conditions and the target variable (y) relating to irregular migration incidents. The data is organized into arrays, which are dynamically modified based on the characteristics of the raster dataset, providing uniformity across all bands. The pixel data are then placed in a NumPy array for processing.

The NumPy module is then implemented to combine the feature and target arrays into a cohesive 2-D array referred to as DataX. The dataset is split into training (80%) and testing (20%) subsets via train_test_split from sci-kit-learn, allowing the evaluation of the algorithm's performance on unfamiliar data.

The LightGBM algorithm is instantiated with optimized hyperparameters and trained on the training dataset. After completion of training, the model's efficacy is assessed using the test data, and the results are recorded for comparison with alternative models.

The trained LightGBM algorithm eventually forecasts migratory patterns throughout the research area, producing class labels that are exported as a new raster file.

3.5.4 Cross-Validation

This study employed cross-validation to assess the accuracy of the predictive models. K-fold cross-validation, a technique that splits the dataset into k subgroups, was employed. The model is trained using k-1 folds and evaluated on the remaining fold. This procedure is executed k times, with each fold utilized as the test set once. Averaging the performance indicators from each fold yields a more dependable assessment of the model's generalization capability. This approach is particularly helpful for evaluating, selecting, and enhancing hyperparameters, offering a more accurate assessment of the model's efficacy.

XGBoost (XGB), Random Forest (RF), and LightGBM were trained and evaluated on an 80/20 division of the dataset. To ensure the stability and reliability of the models for predicting irregular migration patterns, a 5-fold cross-validation procedure was employed. The data was randomly split into five categories according to serial numbers or time. Four subsets were used for model training, while the remaining subsets were employed for validation. The process was executed five times, allowing each fold to serve as the test set precisely once.

After the completion of the 5-fold cross-validation, important performance metrics, which included the mean cross-validated R^2 (coefficient of determination), RMSE (Root mean square error), and MAE (mean absolute error), were calculated. These metrics offer an in-depth evaluation of each algorithm's predicted precision and its capacity to generalize to novel data. This cross-validation method additionally prevents overfitting, ensuring that the models perform effectively on both training and test datasets (Santos et al., 2018).

3.6 Accuracy Assessment

This study investigated the prediction performance of the XGBoost (XGB), Random Forest (RF), and LightGBM algorithms for accuracy. The algorithms were trained on 80% of the dataset, and the remaining 20% was assigned for validation. The assessment focused on comparing the actual irregular migration statistics with the projected numbers generated by each algorithm. For assessing algorithm accuracy, key regression metrics were utilized: R-squared (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE).

R^2 measures the level to which the algorithm accounts for the variance in the observed data. It indicates the proportion of variability in the dependent variable that is accounted for by the model. A higher R^2 value indicates that the algorithm accurately fits the data and clarifies the underlying patterns. The formula for this metric is as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (O_i - P_i)^2}{\sum_{i=1}^{n} (O_i - \bar{O}_i)^2}$$

RMSE measures the average magnitude of prediction errors, showing the point of how the model's predictions diverge from actual values. Reduced RMSE values indicate better precision in model predictions, as it considers larger error through squaring before a typical basis, hence making it sensitive to outliers. The formula for this metric is as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2}$$

MAE, which computes the average absolute deviation between anticipated and actual values, serves as an additional metric for assessing prediction accuracy. Similar to RMSE, reduced MAE values indicate superior model performance. Conversely, unlike RMSE, MAE treats all mistakes uniformly, rendering it less responsive to greater discrepancies (Gond et al., 2023). The formula for this metric is as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |P_i - O_i|$$

Where P_i is the predicted value, O_i is the observed value, and n is the number of observations, where $\bar{O}_i$ is the mean of the observed values.

This study considered models with elevated R^2 values and diminished RMSE and MAE values as superior in forecasting irregular migratory patterns. These metrics provide an extensive evaluation of the algorithm's predictive abilities, facilitating the identification of the most precise method for predicting migration in the study area.

3.7 Mapping the Risk Area of Irregular Migration in The Study Area

After training the algorithms for predicting irregular migration and identifying key influencing factors, their performance was assessed using R-squared (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE), tailored to the specific conditions of the study area. The trained models were then used to calculate migration probabilities for each pixel over the study period. To analyze the spatial distribution of irregular migration, zonal statistical techniques were applied using ArcMap.

This approach enabled the identification of subdistricts along Thai-Myanmar and Laos borders, facilitating the calculation of migration severity indices. Average migration probabilities were computed for each subdistrict, providing insights into the localized effects of irregular migration. These computations offered a detailed representation of migration trends and allowed for an accurate prediction of migration hotspots. The outcome was a comprehensive map depicting the spatial distribution of irregular migration along the border from 2019 to 2023.

3.8 Tools and Instrumentations

Various tools and instruments were applied in this study during data collection, preprocessing data, and analysis processes, as shown in Table 3.

Table 3 Tools and Instrumentation

Tools and Instrumentation	Usability
Microsoft Excel	Collecting irregular migration data
Google Earth Engine (GEE)	Extracting data by NDVI, BSI, NT, and DEM from Sentinel-2 L2A
ArcMap	Preprocessing
Python	Processing ML
Jupyter notebook	Processing ML

3.9 Summary of This Chapter

This chapter presents the materials and methods for investigating irregular migration in Chiang Rai Province through machine learning algorithms and integrated data analysis. The study utilizes a combination of Remote Sensing data, GIS datasets, and irregular migration arrest records to examine migration patterns and their environmental drivers from 2019 to 2023. Data were sourced from three primary categories: Sentinel-2 satellite imagery for environmental indices such as NDVI and BSI; GIS data for LULC, rivers, roads, and slopes; and arrest point data provided by border control authorities.

Chiang Rai Province is situated in northern Thailand. Chiang Rai contains complex terrain and ethnic diversity. The study focuses on four key districts including

Mae Chan, Chiang Saen, Mae Sai, and Mae Fa Luang, identified as key areas for irregular migration. Data preprocessing included resampling the datasets to maintain consistent spatial resolution, followed by integration for analysis. Additional data such as nighttime light observations, population density from the LandScan database, and DEM were incorporated to evaluate the impact of environmental and socio-economic factors on migration patterns.

Three machine learning algorithms, XGBoost, Random Forest, and LightGBM, were applied to predict irregular migration trends. 80% of the dataset was trained, and the remaining 20% was reserved for testing. Algorithm performance was evaluated using key metrics: R-squared (R^2), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). K-fold cross-validation was employed to ensure model reliability and generalization. This study aims to identify critical factors influencing irregular migration. The result of the study will be shown in the next chapter.

CHAPTER 4 RESULTS AND VALIDATION

This chapter presents the experiment and the result of applying machine learning algorithms to investigate irregular migration and its related factors in four districts of Chiang Rai province, next to Thai-Myanmar and Laos border. The risk map of irregular migration will also be shown to illustrate a clear picture of the study area. The results are the following.

4.1 Analysis of Variables Associated with Irregular Migration in Chiang Rai Province

4.1.1 Irregular Migration Incidents Categorized by Districts

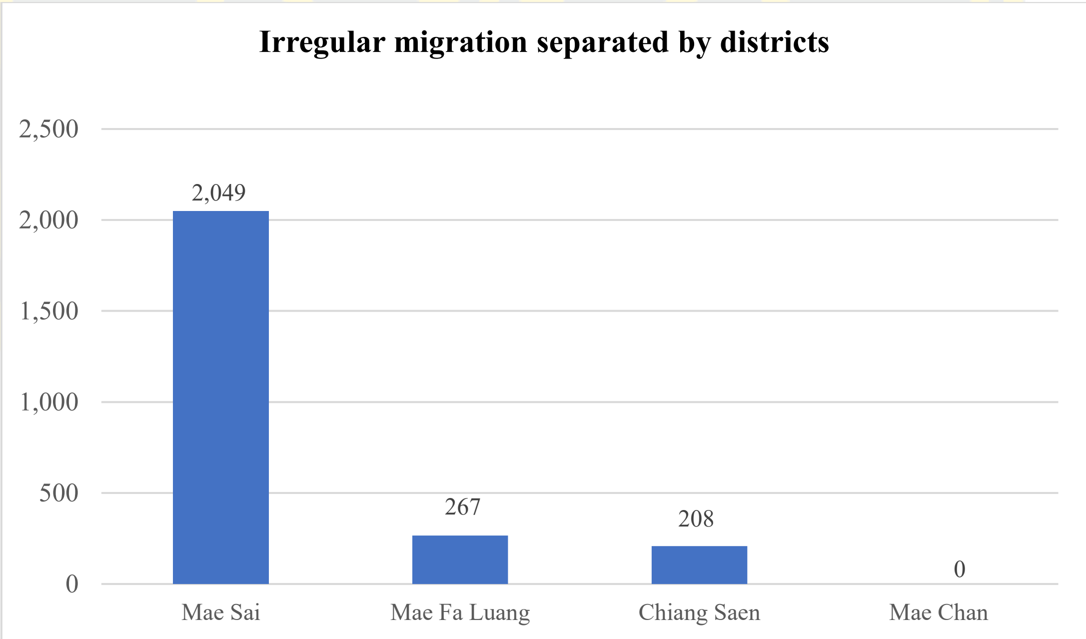

Figure 18 Irregular Migration Incidents Categorized by Districts

Figure 18 illustrates the distribution of irregular migration occurrences among four districts in Chiang Rai province. Mae Sai district documented the highest incidence of irregular migration, with 2,049 cases, thereby establishing it as the principal hotspot in the region. This notable concentration indicates that Mae Sai exhibits traits such as geographic position, border accessibility, or established migration pathways that enable cross-border mobility.

Mae Fa Luang and Chiang Saen districts documented significantly fewer occurrences, totaling 267 and 208, respectively. Despite being significant, these values are far lower than those recorded in Mae Sai, suggesting that irregular movement activity is less prevalent in these regions. Mae Chan district reported no illegal movement activity.

The study highlights Mae Sai as the primary hub for irregular movement in Chiang Rai province, whereas other districts exhibit significantly lower occurrence rates. This distribution pattern offers critical insights for targeted interventions and resource allocation, highlighting the necessity to focus on border control and monitoring initiatives in the high-incidence region of Mae Sai district.

4.1.2 Monthly Distribution of Irregular Migration on Chiang Rai Border

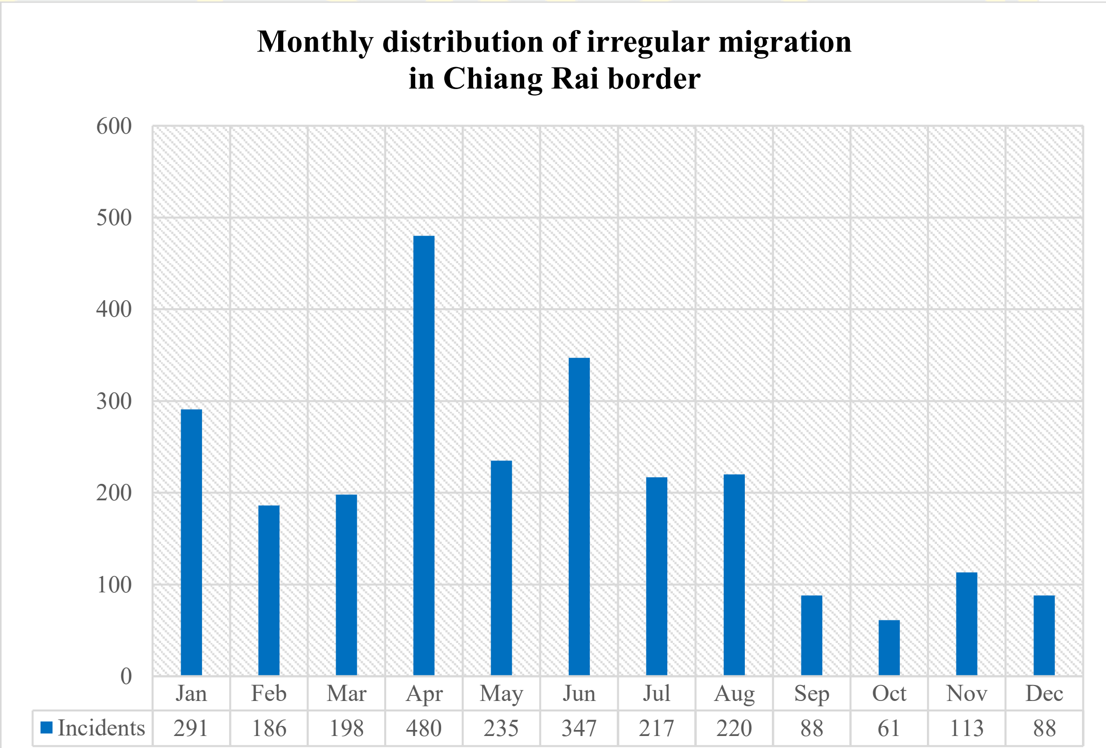

Figure 19 Monthly Distribution of Irregular Migration on the Chiang Rai border

Figure 19 depicts the monthly distribution of irregular movement occurrences along the Chiang Rai border. The statistics indicate clear seasonal variations in migration patterns, characterized by significant peaks and troughs during the year.

The peak occurrence of irregular movement transpires in April, with a documented total of 480 incidents. June follows this peak, accounting for 347 instances. Migration remains at a high level from January to June. The elevated frequency of migration during certain months indicates a possible correlation between migratory behavior and seasonal or environmental conditions that promote cross-border mobility at particular times of the year. The rise in events throughout these months also indicates environmental and socioeconomic factors or agricultural cycles that influence migration patterns and facilitate increased migration during this period.

In the second half of the year, migratory incidences decline significantly. October experienced the lowest number of incidents (61), followed by September (88 incidents). This phase of the dry season likely discourages migration due to challenging topography and reduced labor demand. Migration activity had a slight increase in November, with 113 events. The figure decreased again in December (88 incidents) and then started rising again at the beginning of the year. The low number of events around the end of the year suggests that irregular migration is comparatively less active at the end of the year. This diminished activity is associated with seasonal climatic circumstances or more effective border control tactics throughout these months.

The monthly distribution reveals a cyclical trend in irregular migration occurrences at Chiang Rai border, characterized by beginning and mid-year peaks and diminished activity in the final months. This periodic trend highlights the necessity for focused, timely border management efforts, especially during peak periods, to effectively handle the evolving nature of irregular migration.

4.1.3 Yearly Distribution of Irregular Migration on Chiang Rai Border

Figure 20 depicts the annual distribution of irregular movement occurrences at Chiang Rai border from 2019 to 2023. The data indicate significant fluctuations in migration trends across the five years, with a particularly pronounced peak in 2022.

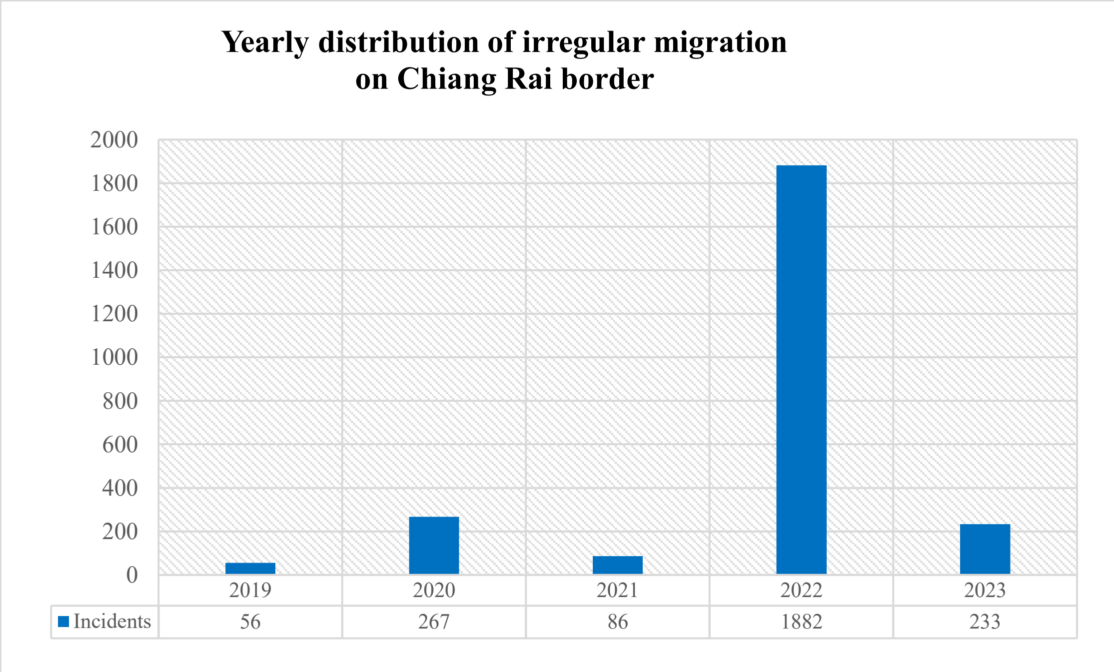

Figure 20 Yearly Distribution of Irregular Migration on Chiang Rai border

In 2022, irregular migration events escalated significantly, with 1,882 cases, the highest recorded figure in this period. The significant increase in 2022 indicates an extraordinary year for migrant activity, potentially affecting resource allocation and management techniques at the border. Conversely, the years 2019, 2020, 2021, and 2023 exhibit significantly reduced levels of irregular migration. In 2020 and 2021, irregular migration episodes experienced a little uptick, totaling 267 and 86 cases, respectively, however, remained comparatively low relative to the peak observed in 2022. As of 2023, the number of events had diminished to 233, reflecting a reduction from the 2022 apex, while still elevated compared to the figures observed before 2022.

This annual distribution indicates that 2022 was an anomaly characterized by exceptionally high irregular movement activity, whilst subsequent years had more consistent but diminished levels of events. These trends underscore the necessity for flexible border control strategies that account for yearly variations in migration patterns, especially in reaction to abrupt surges as shown in 2022.

4.1.4 Interval of Irregular Migration in The Study Area

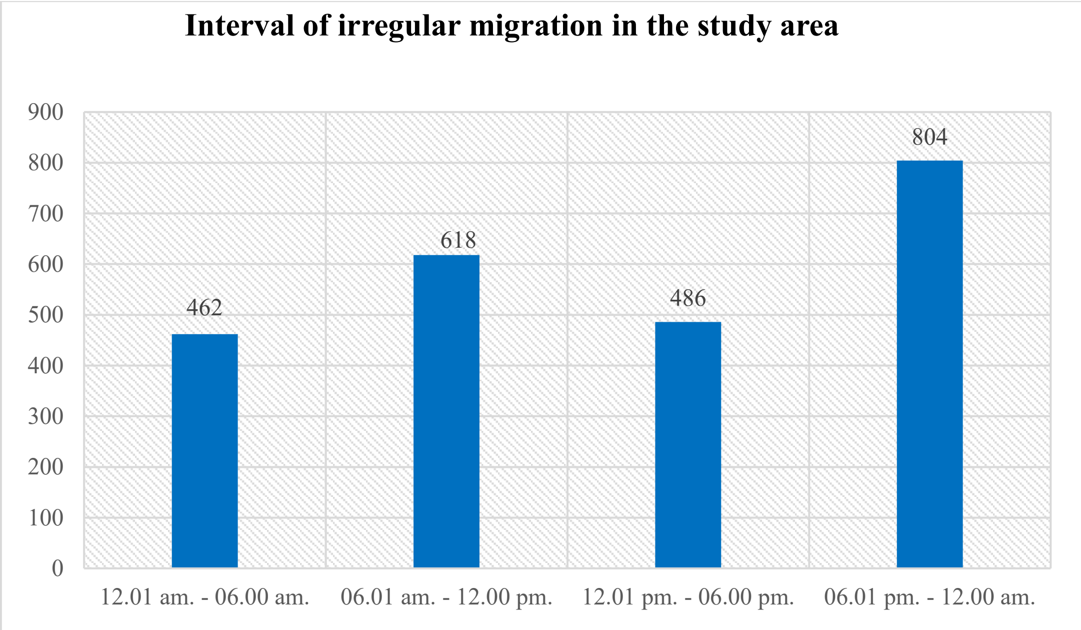

Figure 21 Interval of Irregular Migration in the Study Area

Figure 21 illustrates the distribution of irregular movement episodes categorized by time intervals within the study area, emphasizing the fluctuations in migration activity at various times of the day. The data reveals a distinct pattern, with the peak concentration of irregular migration happening in the late evening and nighttime.

The period from 6:01 p.m. to 12:00 a.m. documented the highest incidence, amounting to 804 occurrences. This surge indicates that migrants may favor nocturnal movement to decrease the chance of detection in obscurity. The second largest interval, from 6:01 a.m. to 12:00 p.m., documented 618 events, signifying a modest level of activity throughout the morning hours.

Conversely, the periods from 12:01 a.m. to 6:00 a.m. and 12:01 p.m. to 6:00 p.m. exhibited less migratory activity, recording 462 and 486 instances, respectively. These intervals may indicate suboptimal conditions for migration, potentially attributable to enhanced visibility during daytime or intensified border monitoring at specific times.

The data provides a clear temporal distribution of irregular migration incidents within the study area. However, it should be noted that authorities did not record the time of activity at the start of 2019, with only 2,370 of 2,524 cases including time data. Consequently, this gap may introduce some inaccuracy in the interval-based analysis of irregular migration activities in the region. The temporal distribution of irregular migration episodes indicates a predilection for nocturnal movement, especially in the late evening hours. This trend highlights the necessity for focused surveillance and enforcement strategies at peak hours, particularly in the evening and early night, to successfully manage irregular migrant flows in the studied area.

4.1.5 Irregular Migration Categorized by Nationality

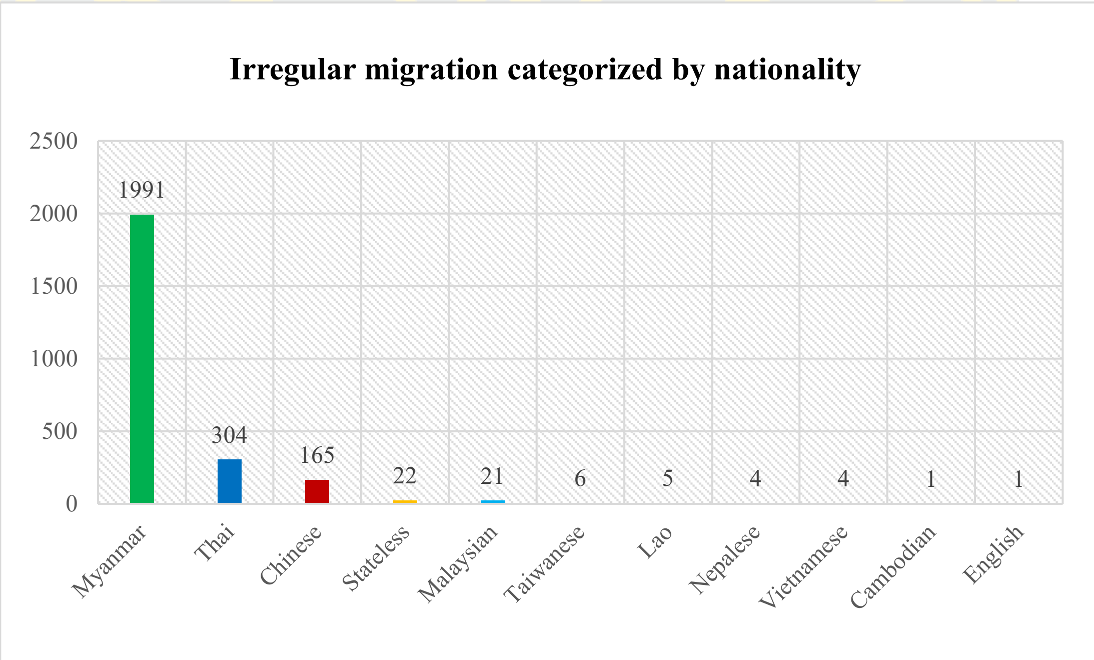

Figure 22 Irregular Migration Categorized by Nationality

Figure 22 depicts the distribution of irregular migration episodes classified by nationality within the study area. The findings indicate that the predominant number of irregular migrants originates from Myanmar, totaling 1,991 cases. This predominant presence indicates that Myanmar is the principal nation of origin for irregular migrants in this region, perhaps attributable to its geographical proximity and maybe socio-economic or political factors influencing migration.

Thai citizens constitute the second-largest group, with 304 instances, succeeded by Chinese nationals with 165 cases. The figures, albeit markedly lower than those from Myanmar, suggest a degree of irregular cross-border travel among local and adjacent communities, maybe influenced by distinct circumstances relative to those affecting migrants from Myanmar.

Other nations are represented in far lower quantities, with stateless individuals being 22 and Malaysians totaling 21. There are limited instances of irregular movement from Taiwanese (6), Lao (5), Nepalese (4), Vietnamese (4), Cambodian (1), and English nationals (1), indicating that such migration from these nationalities is negligible.

The data indicate that Myanmar is the primary source of irregular migration in the study area, with a notable decline in cases from other nations. This distribution highlights the necessity for migration control methods specifically targeting Myanmar nationals but also acknowledging the existence of irregular migrants from other nations, albeit in significantly lesser quantities.

4.1.6 Irregular Migration Categorized by Gender

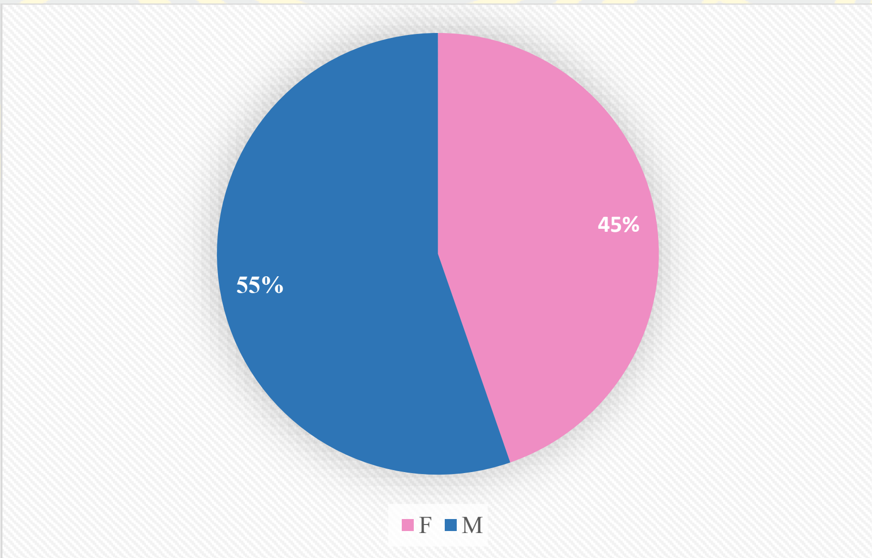

Figure 23 Irregular Migration Categorized by Gender

The irregular migrants in the study area were categorized by gender as shown in Figure 23. The record indicates that 1,396 irregular migrants were male (M), representing around 55% of the irregular migrant population. While 1,128 were female (F), comprising the remaining 45%. Even though the number of female migrants was less than males but still presents a high proportion which shows that Thailand is a destination country for all genders.

This analysis indicates a concentration of irregular movement in Mae Sai, with seasonal peaks occurring in April and a significant rise in 2022. The temporal distribution indicates that nighttime is the favored period for movement, with Myanmar nationals being the majority. The gender distribution indicates a balanced representation, with a little greater number of male migrants.

4.2 Comparing Performance of Machine Learning Algorithms

Three machine learning algorithms were utilized in this study including XGBoost (XGB), Random Forest (RF), and LightGBM (LGBM). After comparing all algorithms, the best performance algorithm will be applied to the study.

4.2.1 Performances of Machine Learning Algorithms

Table 4 Summary of Key Hyperparameters for ML in This Study

Characteristics	Descriptions
Model	XGBoost, Random Forest, LightGBM
Parameter	n_estimators = 200, K_fold = 5
Dependent variable	Arrest Point
Independent variables	NDVI, BSI, NTL, DEM, LandScan, LULC, River data, Road data, Slope data

The models were executed within a Jupyter Notebook environment. The dataset was divided into a training set of 80% and a test set of 20%. Model parameters were refined by 5-fold cross-validation (CV). The mean CV score was computed to assess model reliability and accuracy. The analysis utilized data from remote sensing, GIS, and environmental variables. Each model produced different outcomes upon testing. The performance of the three machine learning algorithms was evaluated, revealing varying levels of accuracy across measures and datasets. The summary parameter of Machine Learning is shown in Table 4.

Table 5 and Figure 24 compare the effectiveness of three machine learning algorithms including XGBoost, Random Forest, and LightGBM in predicting irregular migration incidents within the study. The assessment metrics comprise the coefficient of determination (R^2), Root Mean Square inaccuracy (RMSE), and Mean Absolute Error (MAE), providing a thorough evaluation of the predicted accuracy and inaccuracy of each algorithm.

XGBoost had superior, attaining an R^2 value of 0.91, signifying that it explains 91% of the variance in the data, establishing it as the most precise model of the three. Furthermore, it had the lowest error rates, with an RMSE of 0.15 and an MAE of 0.05, indicating its exceptional capacity to provide accurate predictions with a minimum divergence from actual values.

Random Forest algorithm exhibited a performance R^2 value of 0.85, which, although comparatively high, is inferior to that of XGBoost, suggesting somewhat reduced predictive accuracy. The error metrics, with an RMSE of 0.19 and an MAE of 0.12, exceed those of XGBoost, indicating a larger degree of prediction inaccuracy.

LightGBM demonstrated commendable performance, attaining an R^2 value of 0.87, so situating it between XGBoost and Random Forest regarding prediction accuracy. The RMSE and MAE values of 0.18 and 0.09, respectively, are marginally higher than those of XGBoost but lower than those of Random Forest, indicating that LightGBM is a robust alternative with competitive predictive efficacy.

In this study, XGBoost surpasses the other algorithms, attaining the highest R^2 value and the lowest RMSE and MAE, signifying the most precise and dependable forecasts. Consequently, XGBoost is recognized as the optimal algorithm for this investigation, rendering it the most appropriate selection for predicting irregular migration events.

Table 5 Performances of Machine Learning Algorithms

ML	Accuracy Assessment	Performance of ML
XGB	R2	0.912
	RMSE	0.147
	MAE	0.054
RF	R2	0.846
	RMSE	0.195
	MAE	0.117
LGBM	R2	0.873
	RMSE	0.177
	MAE	0.091

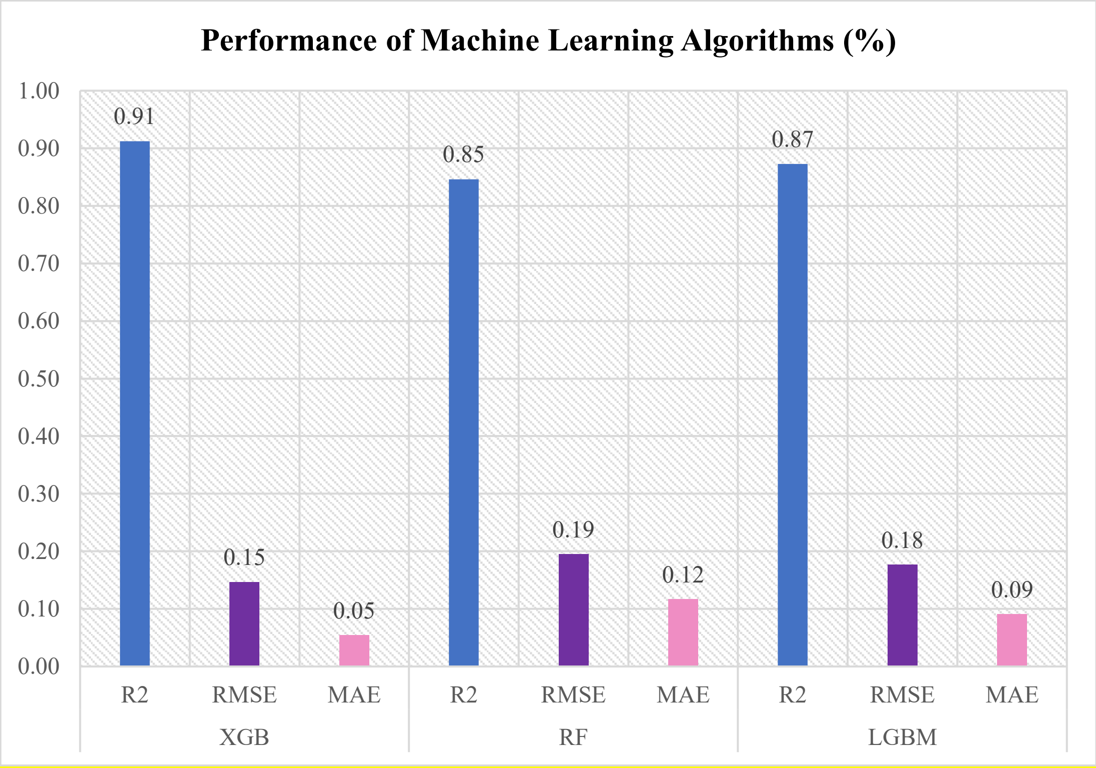

Figure 24 Performances of Machine Learning Algorithms

4.2.2 Significance of Variables

4.2.2.1 Significance Variables in XGBoost:

Figure 25 demonstrates the significance of several environmental variables in predicting irregular migration in Chiang Rai province, utilizing the XGBoost algorithm. The contribution of each factor is represented as a percentage, indicating its significance in the model's predicted accuracy. Roads were the most significant factor, accounting for 32.49% of the predictions of the algorithms, indicating that closeness to road networks is crucial in irregular migration patterns. This indicates that migrants rely on road infrastructure to enable cross-border transit. Elevation, with a significance of 14.02%, is the second most significant element.

Rivers, accounting for 13.83% of the model, significantly influence the area, acting as natural conduits or obstacles along the Chiang Rai border. Nighttime light data (NTL), representing 11.33% of algorithm significance, offers insights into human activity levels or population density, indicating that regions with more nighttime light attract or concentrate migratory flows. Population density data (LandScan), reflecting a contribution of 7.29%, further substantiates this, suggesting that migratory pathways are affected by the concentration of settlements in particular areas. Land Use Land Cover (LULC), with a significance of 6.64%, indicates the influence of land types on migratory patterns, as various terrain types may promote or obstruct mobility.

The slope, accounting for 5.46%, influences the ease of navigation across the terrain, with higher gradients potentially hindering movement, while gentler inclines may enhance it. Normalized Difference Vegetation Index (NDVI) indicates a vegetation density significance of 5.18%, implying that regions with extensive vegetation may offer concealment or hinder mobility. The Bare Soil Index (BSI), with a minimal significance of 3.76%, suggests that exposed soil regions may facilitate or impede movement based on particular topographical characteristics. This XGBoost algorithms analysis highlights the primary impact of road networks on irregular migration in Chiang Rai Province, with additional influences from topographical features like elevation and proximity to rivers.

Table 6 Significance Variables in XGBoost

ML	Variables	Significant Value	%
XGBoost	Road	0.324898	32.49
	Elevation	0.140192	14.02
	River	0.138279	13.83
	NTL	0.113314	11.33
	LandScan	0.072863	7.29
	LULC	0.06641	6.64
	Slope	0.054597	5.46
	NDVI	0.051818	5.18
	BSI	0.03763	3.76

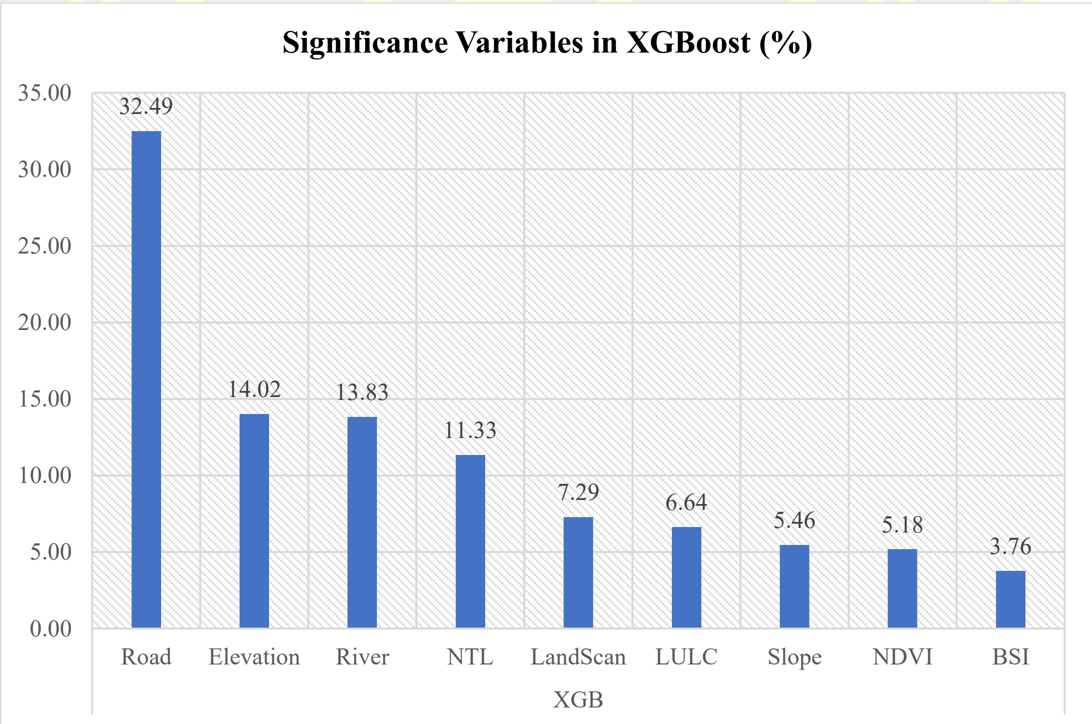

Figure 25 Significance Variables in XGBoost

4.2.2.2 Significance Variables in Random Forest:

Table 7 illustrates the significance of different environmental factors in forecasting irregular migration in Chiang Rai province, Thailand, utilizing Random Forest (RF) algorithm. The contribution of each variable is represented as a percentage in Figure 26, signifying its relative significance in the algorithm's predictive efficacy.

Road networks are recognized as the primary predictor, contributing 23.56%, highlighting the significance of road access in determining migration paths. This elevated percentage indicates that migrants deliberately use road infrastructure to enable cross-border transit. Elevation and nighttime light data (NTL) are tightly correlated, with significance values of 17.62% and 17.39%, respectively, signifying that particular altitudinal zones and regions with increased nocturnal lighting are essential determinants. These considerations suggest that migrants are attracted to specific elevations and areas with noticeable human activity, which could provide both access routes and potential concealment.

Rivers account for 13.33% of the model's predictive capability, possibly acting as natural corridors along the border, affecting migration patterns. Population density data (LandScan) is significant at 10.64%, suggesting regions with differing population densities encourage migratory movements based on localized population concentrations. Normalized Difference plant Index (NDVI), contributing 7.18%, indicates how plant density influences migration paths, as dense foliage can provide cover or impede movement.

The slope and Bare Soil Index (BSI) demonstrate diminished relevance, contributing 4.54% and 4.34%, respectively. The slope impacts the ease of traversing terrain, as higher inclines may impede mobility, whereas BSI, indicating bare soil regions, could change route selection depending on particular terrain features. Land Use Land Cover (LULC) has the least relevance at 1.40%, indicating that classifications of land types (such as urban or wooded regions) exert negligible impact on migration predictions inside the RF algorithm.

Random Forest algorithm identifies road networks, elevation, and densely populated areas from nighttime light as the primary variables influencing irregular migratory patterns in Chiang Rai province, with additional influences from rivers and population density.

Table 7 Significance Variables in Random Forest

ML	Variables	Significant Value	%
RF	Road	0.235618	23.56
	Elevation	0.176223	17.62
	NTL	0.173866	17.39
	River	0.13331	13.33
	LandScan	0.106426	10.64
	NDVI	0.071808	7.18
	Slope	0.045391	4.54
	BSI	0.043395	4.34
	LULC	0.013963	1.40

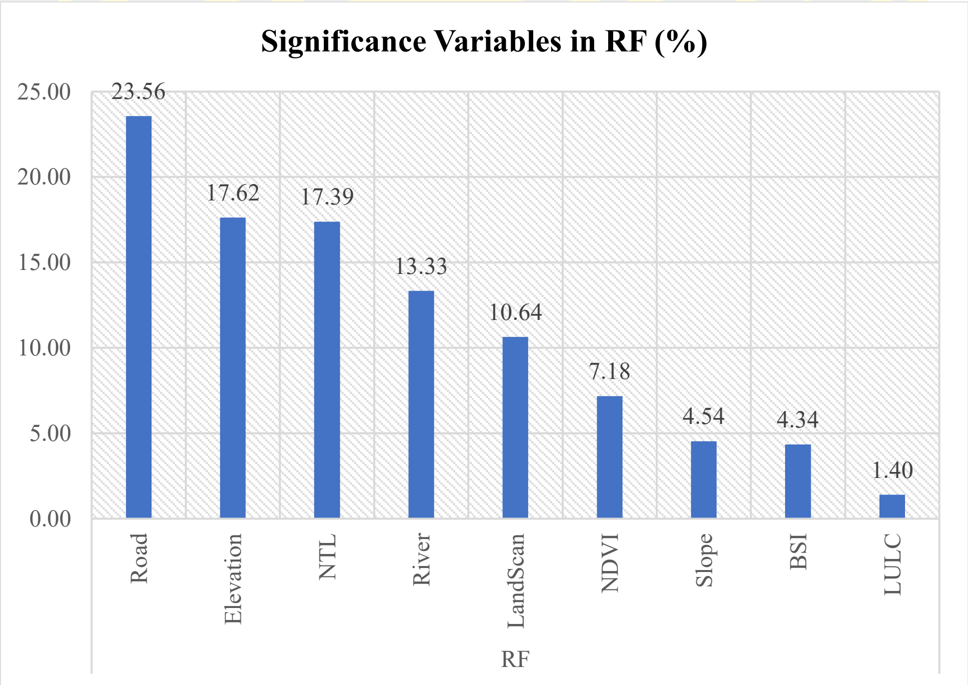

Figure 26 Significance Variables in Random Forest

4.2.2.3 Significance Variables in LightGBM:

Table 8 demonstrates the significance of different environmental factors in forecasting irregular migration in Chiang Rai province, utilizing the LightGBM algorithm. The contribution of each variable is represented as a percentage in Figure 27, signifying its relative significance in the algorithm's predictive efficacy.

Road networks serve as the primary predictor, accounting for 14.69% of the algorithm's accuracy. The significant contribution indicates that proximity to roads substantially influences migration patterns, due to migrants' need for road infrastructure for border crossing. Rivers rank as the second most significant component, contributing 13.54%, suggesting that these natural water features can act as both conduits and obstacles, affecting the direction and intensity of migration patterns.

The Normalized Difference Vegetation Index (NDVI) and slope are significant, accounting for 12.60% and 12.55%, respectively. The importance of NDVI indicates that regions with dense vegetation provide concealing chances or pose obstacles to movement, while slope indicates that topographical inclines affect the ease of traversal. The elevation, contributing 11.88%, suggests that specific altitudinal zones are purposefully chosen by migrants to enhance mobility or for purposes of concealment.

Bare Soil Index (BSI), nighttime light data (NTL), and population density (Landscan) each exhibit comparable contributions, with significant values of 11.85%, 11.75%, and 11.71%, respectively. BSI indicates regions with exposed soil, potentially impacting route selection due to terrain features, whereas NTL and LandScan offer insights into human activity and population density, which influence migration routes by identifying areas with greater human presence or settlement concentration. Ultimately, Land Use Land Cover (LULC) exhibits the least relevance at 2.42%, indicating that classifications of land types exert negligible influence on migration predictions inside the LightGBM algorithm.

LightGBM algorithm highlights the significance of road networks, rivers, and vegetation density in forecasting irregular movement in Chiang Rai province, with additional influences from slope, elevation, and human activity indicators.

Table 8 Significance Variables in LightGBM

ML	Variables	Significant Value	%
LightGBM	Road	0.14693	14.69
	River	0.1354	13.54
	NDVI	0.12598	12.60
	Slope	0.12552	12.55
	Elevation	0.11879	11.88
	BSI	0.11847	11.85
	NTL	0.11754	11.75
	LandScan	0.11709	11.71
	LULC	0.24243	24.24

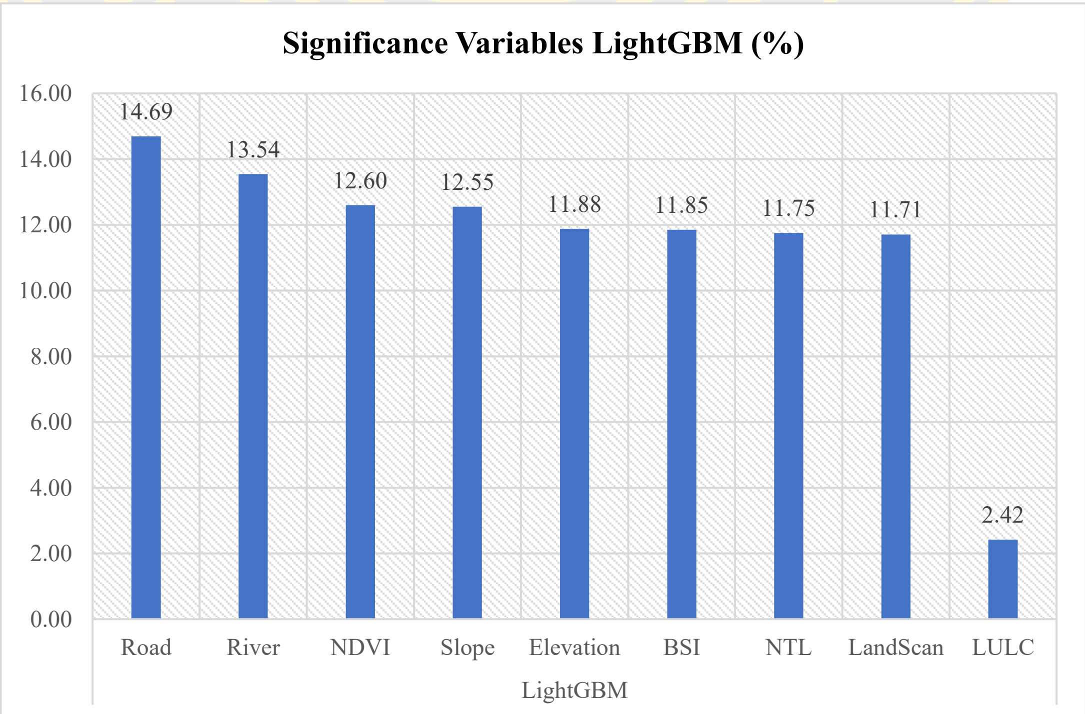

Figure 27 Significance Variables in LightGBM

4.3 Mapping the Risk Area for Irregular Migration

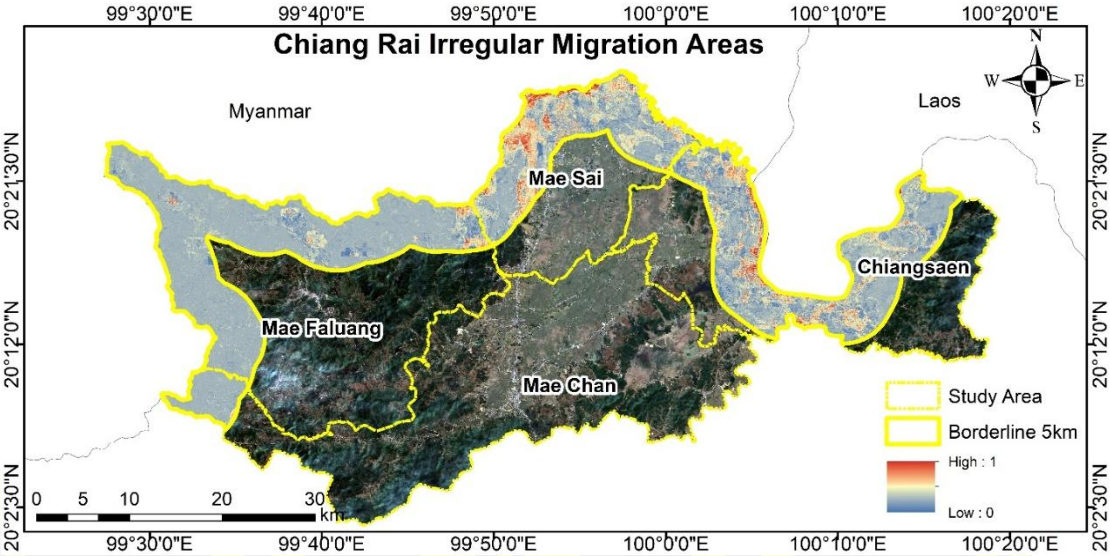

Figure 28 Mapping Risk Area for Irregular Migration

Figure 28 illustrates the spatial distribution of irregular migration risk areas in Chiang Rai Province, emphasizing the border districts next to Myanmar and Laos: Mae Sai, Mae Fa Luang, Mae Chan, and Chiang Saen, which is defined by a yellow dashed boundary. A solid yellow line delineates a 5-kilometer buffer along the borders with Myanmar and Laos. These regions are where migration enforcement and monitoring efforts are concentrated, due to their vulnerability to cross-border movements, both legal and illegal. A heatmap overlay illustrates the intensity of irregular migration activity. The map employs a color gradient to represent risk levels, from low-risk areas (blue) to high-risk areas (red).

High-risk areas, highlighted in red, are mainly in Mae Sai and particular parts of Chiang Saen, making these districts the most significant migratory hotspots in Chiang Rai Province. Mae Sai shows the most concentrated clustering of high-risk areas, especially along its northern and northeastern frontier with Myanmar. The district's geographical location, infrastructure, and connectivity to urban areas enhance its status as a migration hub. In Chiang Saen, high-risk areas cluster around the Mekong River, which serves as a natural boundary with Laos. The river's accessibility, along with moderate slopes and adjacent road networks, makes Chiang Saen an essential area for migration control.

The moderate-risk areas in Mae Fa Luang are shaped by its rugged mountainous landscape, which is joined by navigable paths. Despite the rough terrain restricting extensive mobility, smaller groups of migrants utilize these natural channels to avoid detection. In Mae Chan, moderate-risk areas are predominantly located along the roads connecting Mae Sai and Chiang Saen. These locations, however less accessible than border areas, function as possible spillover zones for migration activities originating from high-risk districts.

The majority of low-risk areas, shown in blue, are found in the interior regions of Mae Fa Luang and Mae Chan, far from direct border access or critical migration pathways. These regions lack the necessary infrastructure and environmental variables that encourage irregular migration lack the necessary infrastructure and environmental variables that encourage irregular migration.

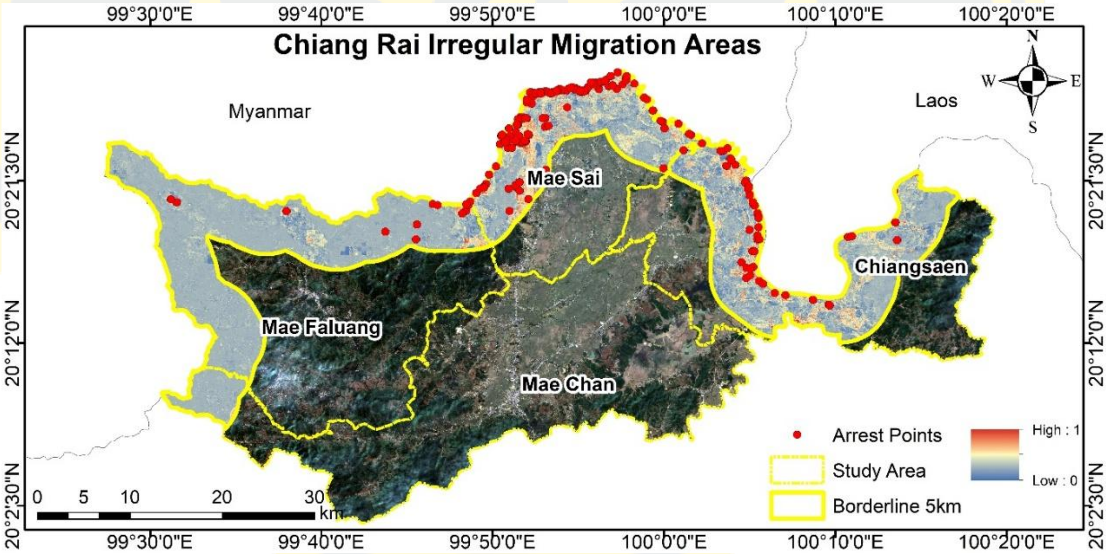

Figure 29 Mapping Risk Area with Arrest Point for Irregular Migration

Figure 29 shows the spatial distribution of irregular movement risks over Chiang Rai Province, integrating arrest point data (red dots) with a gradient map of migration risks. The map defines areas of low risk (blue) to high risk (red) and offers a comprehensive visual illustration of the correlation between migration hotspots and enforcement measures.

High-risk areas, indicated in red, are especially prevalent in the Mae Sai and Chiang Saen districts. These regions also demonstrate the highest concentration of arrest points, signifying increased migrating activity. In Mae Sai, arrest spots are in transit corridors and urban areas. In Chiang Saen, arrest locations are distributed near the river, emphasizing its significance as a crossing point.

Mae Fa Luang contains fewer high-risk zones and arrest points than Mae Sai and Chiang Saen. Arrest points are irregularly distributed and situated in areas of moderate risk, indicating isolated migration activity in this region. There were no arrest points for Mae Chan during the study period. The next part will clarify further specifics regarding areas at risk for irregular migration.

4.4 Irregular Migration Risk Areas in Chiang Rai Province

4.4.1 Mae Sai District

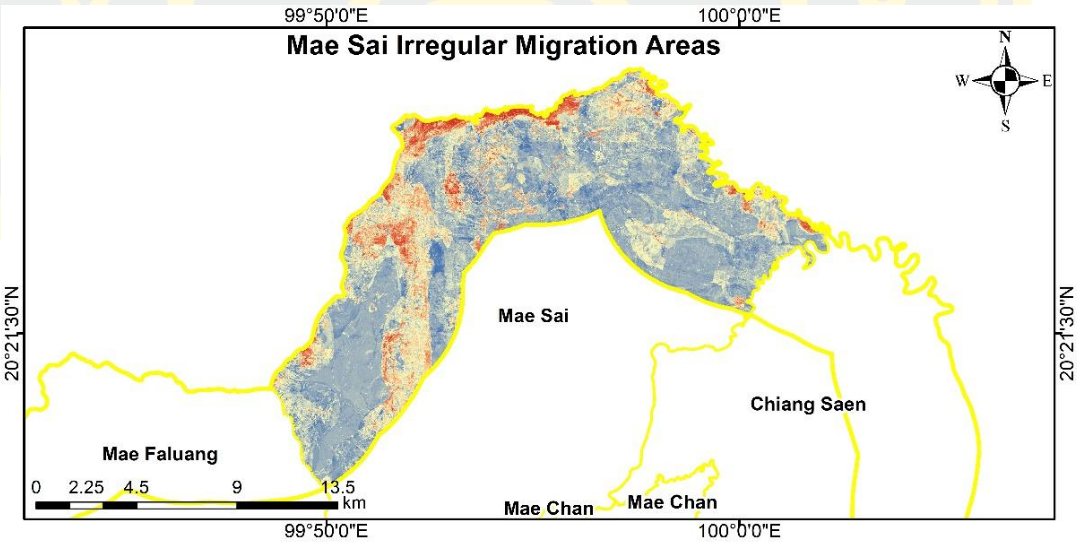

Figure 30 Mae Sai Irregular Migration Risk Area

The map in Figure 30 highlights particular areas of irregular migration risk within Mae Sai district. The color gradient utilized in the map, transitioning from blue to red, graphically depicts the differing levels of migration risk, with red denoting high-risk locations and blue signifying lower-risk zones. Since Mae Sai is a cluster area of irregular migration, the study will concentrate on each part of this region to illustrate and understand the environment that affects irregular migration.

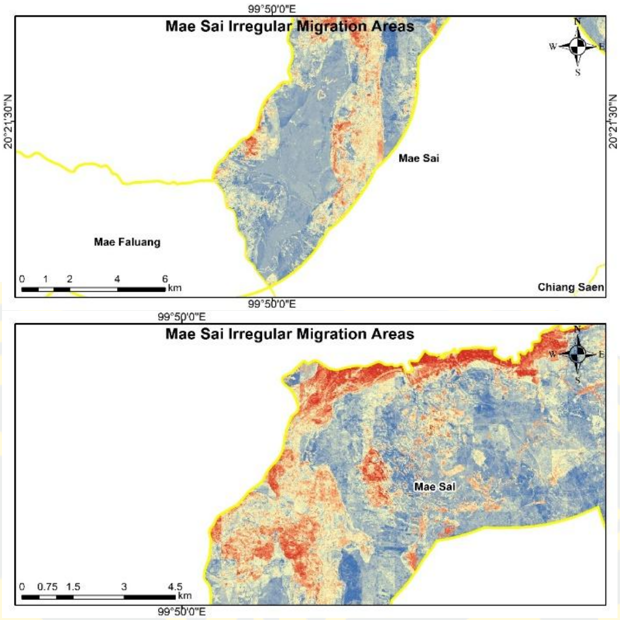

Figure 31 Mae Sai Irregular Migration Risk Area (1,2)

The area on the left side of Mae Sai is mainly covered with mountainous terrain, which creates natural corridors that facilitate irregular migration by offering concealed and accessible routes. The area's geographical closeness to the two nations is designated as a special control zone, where military personnel patrol and establish checkpoints, leading to many arrests. Most of the arrest locations in these areas were on the road. Figure 32 shows the sample pictures of this area.

Figure 32 Sample of the Risk Area: Mae Sai (1,2)

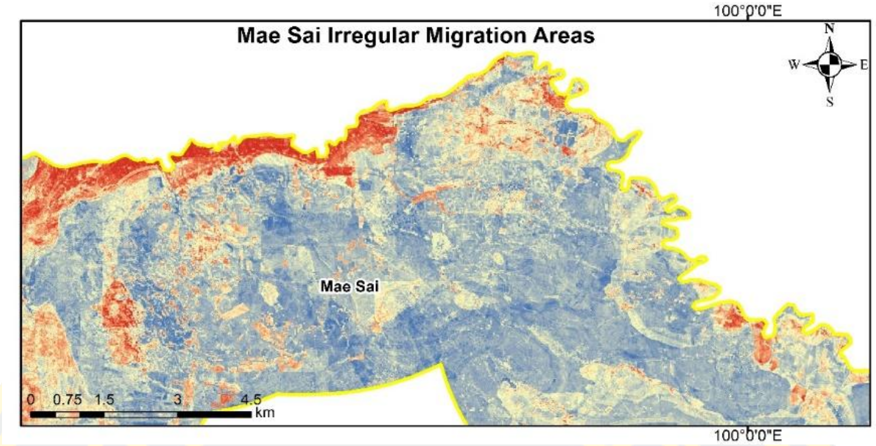

Figure 33 Mae Sai Irregular Migration Risk Area (3)

Figure 33 illustrates an irregular movement risk zone in Mae Sai, a significant urban locality in northern Thailand adjacent to the Myanmar border. Mae Sai is recognized for its importance as a tourist attraction and an economic zone. The site's location and attractions attract numerous visitors, resulting in heightened cross-border activity. The influx of individuals exacerbates the difficulties associated with border management and migration control.

Figure 34 Sample of the Risk Area: Mae Sai (3)

The Sai River divides Mae Sai from Myanmar, a natural obstacle to migration that complicates irregular travel. Nonetheless, During the dry season, this river becomes shallow, facilitating migrants' crossing the border by foot. Mae Sai continues to be a main point for cross-border operations. Its designation as a border town and the existence of formal immigration checks between the two nations. However, the amount of irregular migration was significantly the highest in this area. The sample pictures of this area are shown in Figure 34.

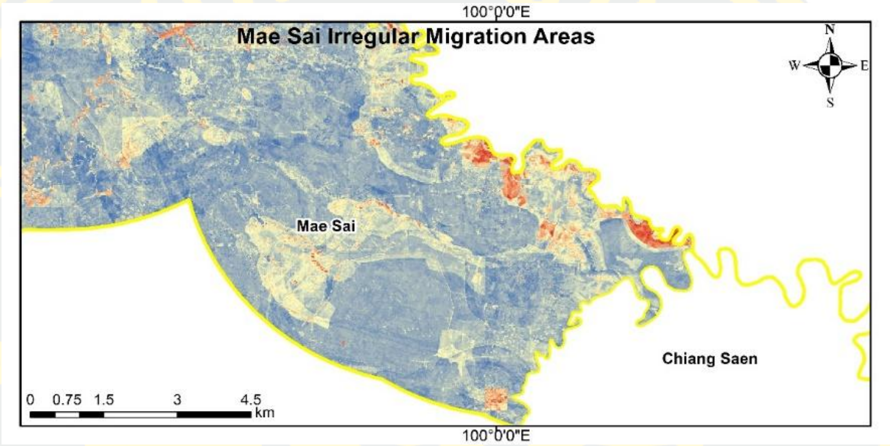

Figure 35 Mae Sai Irregular Migration Risk Area (4)

Figure 35 outlines irregular migration risk zones, mainly within an agricultural region of Mae Sai. There is a river which constitutes a natural demarcation between Thailand and Myanmar. The red regions on the map denote locations with elevated migration risk. Nonetheless, these red zones are restricted in this agricultural area, signifying a diminished frequency of irregular migration apprehensions in comparison to other regions of Mae Sai. This indicates that irregular migration is either less prevalent in this region or more challenging to identify due to the expansive agricultural terrain and less border patrol.

The agricultural fields and riverbank provide potential cover and facilitate easier border access. Nevertheless, the sparse distribution of red areas in this section of Mae Sai suggests that more migration activity happens in other regions, where a more significant number of risk points are located. Samples of risk areas are shown in Figure 36.

This spatial concentration indicates that Mae Sai needs targeted assistance, especially in its northern border areas. The insights obtained from this map may guide local authorities in strategic planning, facilitating optimal resource allocation, and implementing targeted initiatives to reduce irregular migration in high-risk border areas.

Figure 36 Figure Sample of the Risk Area: Mae Sai (4)

4.4.2 Chiang Saen District

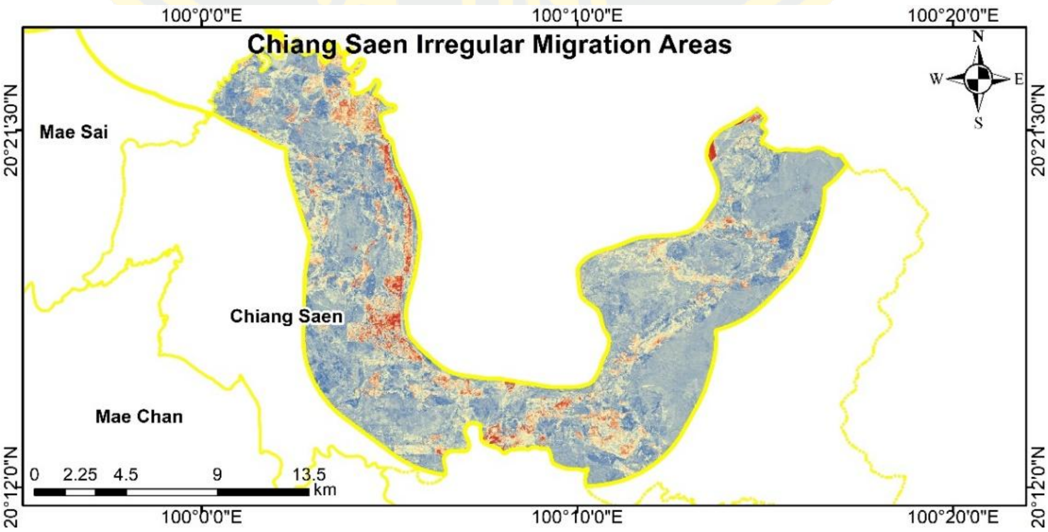

Figure 37 Chiang Saen Irregular Migration Risk Area

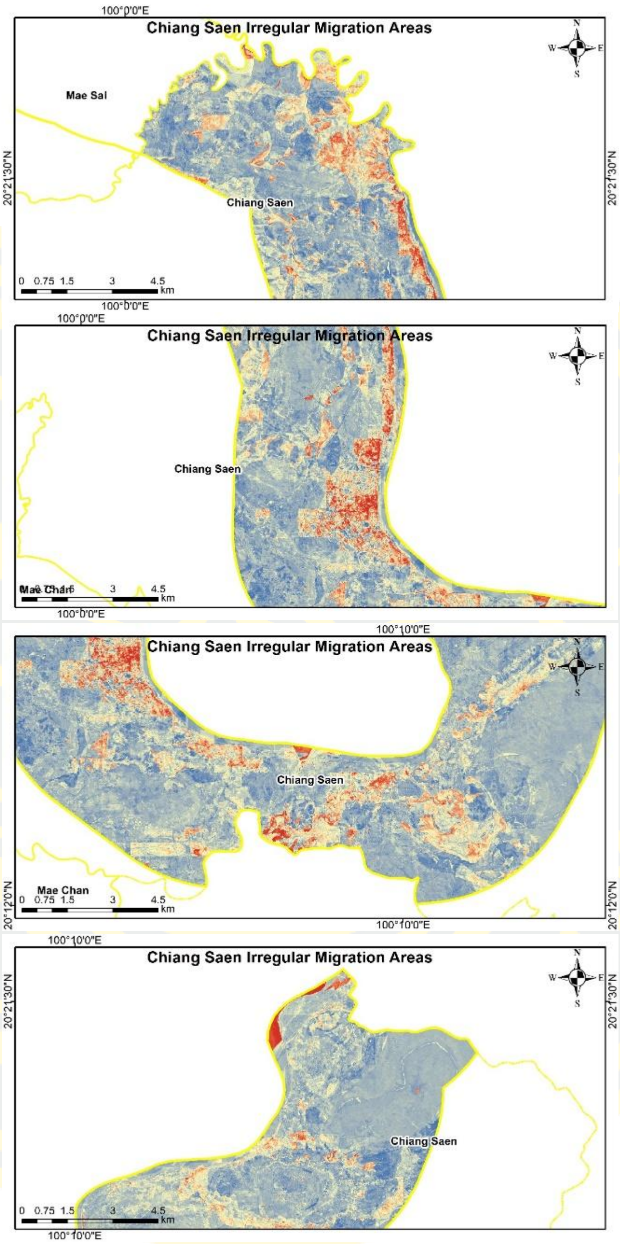

Figure 38 Chiang Saen Irregular Migration Risk Area, Highlighting in Each Zone

The map in Figures 37 and 38 illustrates areas of irregular migration risk in the Chiang Saen district, a strategically significant border region in northern Thailand. Chiang Saen borders Myanmar to the north and Laos, starting from the Golden Triangle, a notable convergence points between the three nations. Chiang Saen's location renders it an important area for overseeing and regulating cross-border travel, owing to its proximity to several international borders.

Areas with elevated risk for irregular migration are primarily located along the river and its surrounding transportation networks. The river is a natural conduit for transboundary migration, enabling access among Thailand, Myanmar, and Laos. The existence of risk zones along the roadways indicates that these routes are frequently utilized by migrant's post-border crossing, either for transit or as covert paths to evade discovery.

Moderate-risk zones in Chiang Saen are located close to roadways connecting the district to neighboring regions of Chiang Rai. These paths are frequently employed by migrants after crossing the border to avoid detection. Low-risk locations are located further interior, distanced from the river and primary transportation lines, leading to low migration activity in these regions. Samples of risk areas are shown in Figure 39.

Figure 39 Sample of the Risk Area: Chiang Saen

4.4.3 Mae Fa Luang District

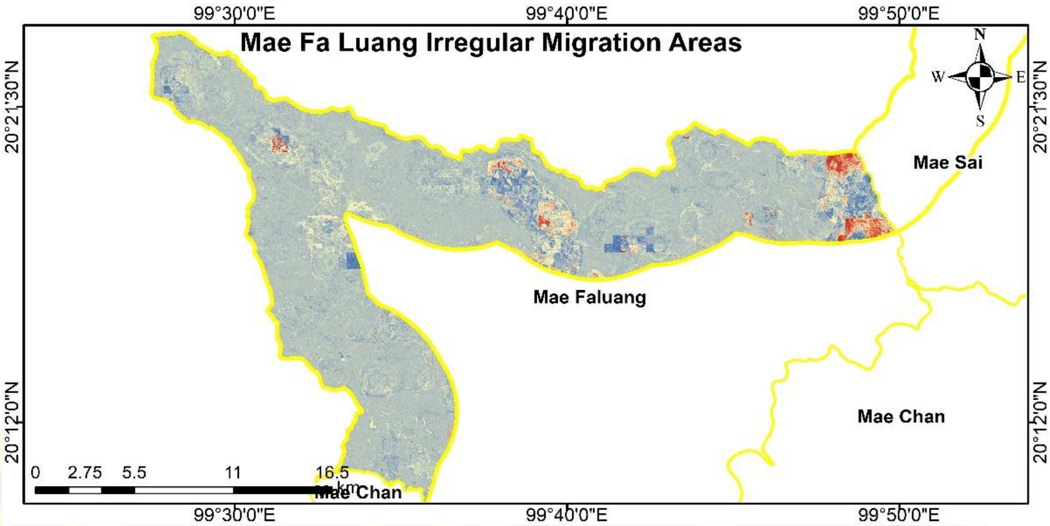

Figure 40 Mae Fa Luang Irregular Migration Risk Area

The map in Figure 40 delineates areas of irregular migration risk in Mae Fa Luang, a district adjacent to the Myanmar border. Most of Mae Fa Luang is covered by forest and agricultural areas. The district is situated in a mountainous region and possesses a greater elevation than other districts. This challenging landscape probably influences migration patterns and complicates monitoring efforts. The map depicts a few high-risk zones in Mae Fa Luang, suggesting low irregular movement activity in this region. The primary risk zone is located on the eastern side, close to Mae Sai. Mae Sai is recognized for elevated migration activity, which may intermittently extend into this region of Mae Fa Luang. Samples of risk areas are shown in Figure 41, which were obtained from Google Street View.

Mae Fa Luang possesses limited road infrastructure, which may mitigate migration risks. The lack of roads results in fewer accessible pathways, complicating the movement of migrants through the district. The interplay of dense forest, elevated terrain, and restricted road access likely deters irregular migration, redirecting it to more accessible neighboring regions, such as Mae Sai.

Figure 41 Sample of the Risk Area: Mae Fa Luang

4.4.4 Mae Chan District

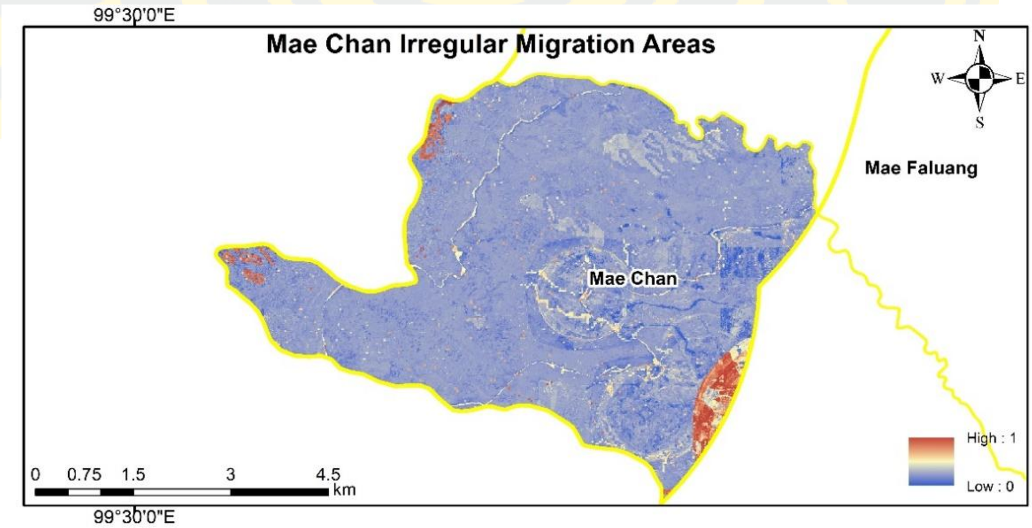

Figure 42 Mae Chan Irregular Migration Risk Area

Multiple factors are likely responsible for this minimal risk. Mae Chan is an inland district. Its center location renders it less accessible for migrants seeking direct entry or exit routes across the border. Migrants favor border regions with direct access to adjacent countries over inland locations.

4.5 Summary of Experiment and Result

This chapter delineates the results of statistical analysis variables associated with irregular migration in the study area during the study period. The study employs machine learning algorithms to analyze irregular movement patterns and their related characteristics across four districts in Chiang Rai province, situated along the Thai-Myanmar and Laos borders. This study's findings show the spatial, temporal, and demographic distribution of migration movement, the importance of environmental factors in forecasting migration trends, and the machine learning algorithms' performances.

According to the analysis of variables related to irregular migration, the major hotspot with the highest incidence of irregular movement cases (2,049) is Mae Sai district. Conversely, the districts of Mae Fa Luang and Chiang Saen reported markedly fewer incidents while Mae Chan reported no event which indicates that irregular movement is localized in particular places within the region.

Seasonal trends were identified, with significant migration events happening in April, followed by June and January. Although other years saw comparatively stable but diminished activity levels, the migration rate significantly increased in 2022, with 1,882 incidents compared with the other four years.

The study indicated a predilection for nocturnal migration during a daily cycle, especially between 6:01 p.m. and 12:00 a.m. Data on nationality shows that the majority of irregular migrants are from Myanmar, with lesser numbers from Thailand, China, and other nations. Myanmar is the primary source of irregular migration in the study area. Furthermore, the number of males comprised 55% and females 45% of the migrant population, which indicated a relatively even distribution. This finding shows that Chiang Rai is a destination for both sexes.

This study evaluated the predictive performance of three machine learning algorithms consisting of XGBoost, Random Forest, and LightGBM. The result reveals that XGBoost is the most dependable algorithm for forecasting migration events which exhibited exceptional accuracy, attaining an R^2 of 0.91 with an RMSE of 0.15 and an MAE of 0.05. Variable relevance analysis of the algorithms consistently revealed that roads, elevations, and rivers significantly impact migration patterns. Roads are the most pivotal predictor, indicating that proximity to transportation networks and geographic factors substantially influence migration patterns in the study area.

A risk map was created to illustrate high-risk zones for irregular migration around the borders of Mae Sai, Mae Fa Luang, Mae Chan, and Chiang Saen. High-risk areas are mainly located in Mae Sai and particular parts of Chiang Saen, making these districts the most significant migratory hotspots in Chiang Rai Province. Mae Sai shows the most concentrated clustering of high-risk areas, especially along its northern and northeastern frontier with Myanmar. The district's geographical location, infrastructure, and connectivity to urban areas enhance its status as a migration hub. The road networks and transportation routes in Mae Sai allow rapid transit, while urban centers draw migrants for economic prospects or temporary sanctuary. Furthermore, Mae Sai's relatively flat topography compared to adjacent districts, along with its closeness to legal checks, creates an ideal environment for irregular migration.

In Chiang Saen, high-risk areas cluster around the Mekong River, serving as a natural boundary with Laos. The accessibility, along with moderate slopes and adjacent road networks, makes Chiang Saen an essential area for migration control. The proximity to the Golden Triangle region, a notable center for cross-border activity, enhances the risk.

The moderate-risk areas in Mae Fa Luang are shaped by its rugged mountainous landscape, which is joined by navigable paths and pathways. Despite the rough terrain restricting extensive mobility, smaller groups of migrants utilize these natural channels to avoid detection. Forested regions within the district offer concealment, leading to moderate risk ratings in certain areas.

In Mae Chan, moderate-risk areas are predominantly located along the roads connecting Mae Sai and Chiang Saen. These locations, however less accessible than border areas, function as possible spillover zones for migration activities originating from high-risk districts.

Most low-risk areas are found in the interior regions of Mae Fa Luang and Mae Chan, far from direct border access or important migration pathways. The hilly topography of Mae Fa Luang significantly restricts travel, as steep slopes make large-scale migration impractical. Also, the inner areas of Mae Chan have restricted access to key border crossings and transit corridors, leading to low migrant activity.

CHAPTER 5 CONCLUSION AND FUTURE WORK

5.1 Conclusion and Discussion

Thailand's border is characterized by mountains, dense forests, and rivers. Complex geography plays a crucial role in facilitating irregular migration. Chiang Rai province is located in the uppermost part of Thailand, connecting to Myanmar and Laos, making it a risk area for irregular movement. The topography in the border area and socioeconomic challenges lead to obstacles for authorities in controlling irregular migration.

The primary initiative of this study is to investigate the significant factors affecting irregular migration in Chiang Rai province from 2019-2023. The study covers four district areas: Mae Sai, Mae Chan, Chiang Saen, and Mae Fa Luang. Several sources, such as Remote Sensing, GIS data, and irregular migration arrest point data from border authorities, were applied in the study. Three machine learning algorithms consisting of XGBoost, Random Forest, and LightGBM were employed to examine the most accurate algorithms for predicting irregular migration patterns. Finally, the study seeks to develop a comprehensive risk map that outlines irregular movement incidents within the study area, providing critical insights to improve border management methods.

This study comprehensively analyses the various variables influencing irregular migration in Chiang Rai province. The findings underscore the substantial influence of the region's geographical and social landscape on migratory patterns, designating Chiang Rai as a pivotal transit hub. Mae Sai area is notable for its strategic position at the border intersection with Myanmar and Laos, establishing it as a natural center for cross-border migration, in contrast to the more arduous terrains of Mae Fa Luang and Mae Chan districts. Socioeconomic and policy-related issues significantly influence migration trends. Myanmar is the primary source of irregular migration in the study area. The demand for male labor in Thailand's informal sectors draws migrants, especially males from Myanmar, motivated by economic instability and restricted employment opportunities.

The migration rate significantly increased in 2022 compared with the other four years. The COVID-19 pandemic significantly impacted travel trends, since lockdowns in 2020 resulted in a rise in irregular migration due to the closure of formal routes, compelling people to utilize informal channels. In 2021, rigorous border controls initially diminished migratory events; nevertheless, a rebound transpired in 2022 when limitations were relaxed. This study underscores the significant influence of restrictive border policies on migration patterns, indicating that irregular migration frequently increases when official avenues are inaccessible, especially in Southeast Asia, where labor mobility is crucial to the economy.

The study also discerned seasonal movement patterns, peaking in April during the summer season, with diminished rates in the dry months of October. This study suggests that irregular migration is comparatively less active at the end of the year. These findings align with prior research indicating that precipitation and minimum temperature influence the migration rate by around 40% (Pless et al., 2023). These variations are influenced by environmental and socioeconomic factors, including terrain accessibility and labor requirements associated with planting and harvesting seasons. This alignment with natural and agricultural cycles illustrates how migration patterns correspond with environmental conditions and seasonal labor availability, highlighting the direct influence of these factors on migratory trends.

Furthermore, the study revealed that most migratory episodes transpire throughout the night. These findings align with prior research illegal immigrants mostly move through natural routes at night when no officers are patrolling. Stricter enforcement during the morning might be the reason for fewer incidents in the study area. The observation that most incidents happen during nighttime hours suggests that border enforcement should be strengthened during these times to effectively prevent irregular migration (Sapprasert, 2017).

The utilization of machine learning algorithms provides significant predictive insights on irregular migration activity. XGBoost demonstrated the highest predictive accuracy among the investigated algorithms, attaining an R^2 value of 0.91 with an RMSE of 0.15 and an MAE of 0.05 and efficiently elucidating the intricate links between environmental factors and migration. Feature relevance analysis underscored the significance of roads, elevation, and rivers, affirming that geographical attributes are crucial in determining migrants' route selection and movement viability.

This study identified road networks, elevation, and rivers as the primary environmental factors affecting migration, with road networks being the most significant factor, accounting for 32.49% relevance in the XGBoost algorithm. This discovery highlights that roads facilitate not just rapid and accessible pathways but also assist migrants in avoiding detection. A prior study also reveals that illegal immigrants mostly use main roads for transport, as well as forest trails and indirect routes to avoid detection at state investigation posts (Karrachitawaragul, 2016). The elevation is the second significant, accounting for 14.02 %. The gentler inclines in regions such as Mae Sai and Chiang Saen facilitate transportation, in stark contrast to the mountainous obstacles present in other locales. River was the third significant factor, accounting for 13.83 %. River crossings, especially during low water conditions, provide supplementary access locations. Meanwhile, bare soil conditions have little direct influence on migrants' movement decisions, as they were the last significant factor for the irregular movement, accounting for only 3.76%.

The creation of a risk map identifying high-risk migration areas offers significant insights into border control initiatives. Risk maps define high-risk zones in Mae Sai and certain areas of Chiang Saen as the essential locations for migration controls. These districts are significantly impacted by their proximity to international borders, accessible topography, and transportation networks, making them crucial to enforcement initiatives. Moderate-risk zones, such as Mae Fa Luang and particular areas of Chiang Saen, serve as transitional regions, often connecting high-risk zones to adjacent places. Low-risk areas, primarily in Mae Chan and the interior of Mae Fa Luang, are marked by inadequate infrastructure, challenging topography, and lower connection, which restricts their migration activities.

The study result emphasized that irregular movement in Chiang Rai is shaped by a combination of spatial, temporal, demographic, and environmental factors, with infrastructure, geographical location, and border proximity playing crucial roles. Environmental factors influence migration patterns across countries over time, but the size of the effects depends on the economic and sociopolitical context and the environmental factors considered (Hoffmann et al., 2020). There are various factors influencing irregular migration, and the conditions in both origin and destination countries are important. The problem with the manpower and expertise of the border control agency also influences irregular migration, which also requires attention (Asst. Prof. Pol.Col. Dr. Prapon Sahaphanthana, 2020)

This study provides essential insights for improving border control procedures and comprehending the environmental factors influencing irregular migration. Policymakers and border authorities can utilize these studies to enhance management measures of irregular migration along Thailand's northern border.

5.2 Future Work and Suggestions

The results of this study can be used as a guideline for border control planning and management in Chiang Rai Province and other border areas in Thailand. Future studies should enlarge the study area to investigate additional borders over the country for more comprehension of irregular migration dynamics. Moreover, expanding the study period would provide insights into long-term migration trends and a deeper evaluation of border control measures.

Further geographic and socioeconomic factors should be applied to enhance the predictive ability of machine learning algorithms. Including these factors would offer a better understanding of irregular migration drivers and improve predictive models. Future studies should investigate the motivations and context behind migration decisions. Comprehending the economic, political, and social circumstances in origin and destination regions would increase awareness of irregular migration.

Future research should concentrate on selecting and optimizing the models. To enhance their accuracy, a comprehensive examination of algorithm performance across various settings and measures will be helpful. This includes adjusting hyperparameters, identifying essential features, and evaluating models on diverse datasets. It is useful to enhance models and broaden their application to more locations and circumstances. Researchers can apply these models to diverse migration situations, refine them with updated data, and adapt them to address challenges in various border regions.

The tools and results obtained through this work, including risk maps and predictive models, should be applied and integrated into border control practices. Applying the guidelines from this study and collaborating with local authorities and policymakers would enhance the effectiveness of border management and irregular migration control in the region.

REFERENCES

Aenihon, P. (2018). *Foreign Workers: Implementation and Administration in Thailand*. Mahachulalongkornrajavidyalaya University,Thailand. Retrieved 1/10/2023 from https://so02.tci-thaijo.org/index.php/jam/article/download/139681/115408

Akber, M., & Shrestha, R. (2015, 01/02). Land use change and its effect on biodiversity in Chiang Rai province of Thailand. *Journal of Land Use Science, 10*, 108-128. https://doi.org/10.1080/1747423X.2013.807315

Ali, Z., Abduljabbar, Z., Tahir, H., Sallow, A., & Almufti, S. (2023, 05/31). Exploring the Power of eXtreme Gradient Boosting Algorithm in Machine Learning: a Review. *12*, 320-334. https://doi.org/10.25007/ajnu.v12n2a1612

Alkasassbeh, M., Abbadi, M., & Al-Bustanji, A. (2020). LightGBM Algorithm for Malware Detection. In (pp. 391-403). https://doi.org/10.1007/978-3-030-52243-8_28

Almuhanna, A. A., Alrehili, M. M., Alsubhi, S. H., & Syed, L. (2021, 6-7 April 2021). Prediction of Crime in Neighbourhoods of New York City using Spatial Data Analysis. 2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA),

Alsubayhin, A., Ramzan, M., & Alzahrani, B. (2023). Crime Prediction Using Machine Learning: A Comparative Analysis. *Journal of Computer Science, 19*(9). https://doi.org/10.3844/jcssp.2023.1170.1179

Anneleen, R., Wim, Hardyns., Lieven, Pauwels. (2017). The use of predictive analysis in spatiotemporal crime forecasting: Building and testing a model in an urban context. https://doi.org/doi: 10.1016/J.APGEOG.2017.06.011

Ashraf, M. I., Zhao, Z., Bourque, C. P.-A., & Meng, F.-R. (2012). GIS-evaluation of two slope-calculation methods regarding their suitability in slope analysis using high-precision LiDAR digital elevation models. *Hydrological Processes, 26*(8), 1119-1133. https://doi.org/https://doi.org/10.1002/hyp.8195

Asst. Prof. Pol.Col. Dr. Prapon Sahaphanthana, e. a. (2020). *The threat from transnational crime in Thailand and ASEAN, a case for illegal immigrants to commit crimes in Thailand*. Thailand Science Research and Innovation (TSRI).

Awange, J., & Kiema, J. (2019). Fundamentals of Remote Sensing. In J. Awange & J. Kiema (Eds.), *Environmental Geoinformatics: Extreme Hydro-Climatic and*

Food Security Challenges: Exploiting the Big Data (pp. 115-123). Springer International Publishing. https://doi.org/10.1007/978-3-030-03017-9_7

Axelrad, P. (2010). Global Navigation Satellite Systems. https://doi.org/ https://doi.org/10.1002/9780470686652.eae294

Babu, S., & Manoj, B. S. (2020). Toward a Type-based Analysis of Road Networks. *ACM Trans. Spatial Algorithms Syst., 6*(4), Article 28. https://doi.org/10.1145/3397579

Basir, S. (2020, 01/31). Irregular Migrations in Southeast Asia: Challenges for Protection and Migration. *Indonesian Journal of International Law, 17*. https://doi.org/10.17304/ijil.vol17.2.782

Best, K., Gilligan, J., Baroud, H., Carrico, A., Donato, K., & Mallick, B. (2022, 2022/03/25). Applying machine learning to social datasets: a study of migration in southwestern Bangladesh using random forests. *Regional Environmental Change, 22*(2), 52. https://doi.org/10.1007/s10113-022-01915-1

Bhaduri, L., Bright, E., Coleman, P., & Urban, M. (2007, 06/01). LandScan USA: A High Resolution Geospatial and Temporal Modeling Approach for Population Distribution and Dynamics. *GeoJournal, 69*, 103-117. https://doi.org/10.1007/s10708-007-9105-9

Bruederle, A., & Hodler, R. (2018). Nighttime lights as a proxy for human development at the local level. *PLOS ONE, 13*(9), e0202231. https://doi.org/10.1371/journal.pone.0202231

Caleb Robinson, B. D. (2018). *A Machine Learning Approach to Modeling Human Migration* Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies, Menlo Park and San Jose, CA, USA.

Caruso, G., Palai, G., Tozzini, L., D'Onofrio, C., & Gucci, R. (2023, 2023/12/01/). The role of LAI and leaf chlorophyll on NDVI estimated by UAV in grapevine canopies. *Scientia Horticulturae, 322*, 112398. https://doi.org/https://doi.org/10.1016/j.scienta.2023.112398

Castro, Ú. R., M. and Brandão, W. (2020). Predicting Crime by Exploiting Supervised Learning on Heterogeneous Data. Proceedings of the 22nd International Conference on Enterprise Information Systems - Volume 1: ICEIS,

Chen, T., & Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System*

Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA. https://doi.org/10.1145/2939672.2939785

Cheng, W., Rao, Y., Tang, Y., Yang, J., Chen, Y., Peng, L., & Hao, J. (2022). Identifying the Spatio-Temporal Characteristics of Crime in Liangshan Prefecture, China. *International Journal of Environmental Research and Public Health, 19*(17), 10862. https://www.mdpi.com/1660-4601/19/17/10862

Conover, T. (2020). *The Routes of Man: How Roads Are Changing the World and the Way We Live Today*. https://doi.org/https://doi.org/10.5860/choice.47-6830.

Cutler, A., Cutler, D., & Stevens, J. (2011). Random Forests. In (Vol. 45, pp. 157-176). https://doi.org/10.1007/978-1-4419-9326-7_5

Djon, D., Jhawar, J., Drumm, K., & Tran, V. (2023). A Comparative Analysis of Multiple Methods for Predicting a Specific Type of Crime in the City of Chicago. *arXiv preprint arXiv:2304.13464*.

Dobson, J., Bright, E., Coleman, P., Durfee, R., & Worley, B. (2000, 07/01). LandScan: A Global Population Database for Estimating Populations at Risk. *Photogrammetric Engineering and Remote Sensing, 66*, 849-857.

Dubey, S., & Gangwar, P. K. (2017). Road: An overview. *International Journal of Advance Research and Innovative Ideas in Education, 3*, 3399-3401.

Ershad, M., & Ali, E. (2020a). *Geographic Information System (GIS): Definition, Development, Applications & Components*.

Ershad, M., & Ali, E. (2020b, 04/08). Global Positioning System (GPS): Definition, Principles, Errors, Applications & DGPS.

Gond, S., Gupta, N., Patel, J., & Dikshit, P. K. S. (2023, 2023/03/02). Spatiotemporal evaluation of drought characteristics based on standard drought indices at various timescales over Uttar Pradesh, India. *Environmental Monitoring and Assessment, 195*(3), 439. https://doi.org/10.1007/s10661-023-10988-2

Gotway, C. A., & Young, L. J. (2002, 2002/06/01). Combining Incompatible Spatial Data. *Journal of the American Statistical Association, 97*(458), 632-648. https://doi.org/10.1198/016214502760047140

Guth, P. L., Van Niekerk, A., Grohmann, C. H., Muller, J.-P., Hawker, L., Florinsky, I. V., Gesch, D., Reuter, H. I., Herrera-Cruz, V., Riazanoff, S., López-Vázquez, C., Carabajal, C. C., Albinet, C., & Strobl, P. (2021). Digital Elevation Models: Terminology and Definitions. *Remote Sensing, 13*(18), 3581. https://www.mdpi.com/2072-4292/13/18/3581

Harkins, B. (2019). *Thailand Migration Report 2019*. United Nations Thematic Working Group on Migration in Thailand. Retrieved 3/10/2023 from https://thailand.un.org/en/50831-thailand-migration-report-2019

Hoffmann, R., Dimitrova, A., Muttarak, R., Crespo Cuaresma, J., & Peisker, J. (2020, 2020/10/01). A meta-analysis of country-level studies on environmental change and migration. *Nature Climate Change, 10*(10), 904-912. https://doi.org/10.1038/s41558-020-0898-6

Huang, Q., Yang, X., Gao, B., Yang, Y., & Zhao, Y. (2014). Application of DMSP/OLS Nighttime Light Images: A Meta-Analysis and a Systematic Literature Review. *Remote Sensing, 6*(8), 6844-6866.

Jakkawat, C. (2017). *Foreign Workers Management and National Security in Chiangrai Province*. Army War College. Retrieved 2/10/2023 from http://awc.ac.th/storage/research/87.pdf

Jenga, K., Catal, C., & Kar, G. (2023, 2023/03/01). Machine learning in crime prediction. *Journal of Ambient Intelligence and Humanized Computing, 14*(3), 2887-2913. https://doi.org/10.1007/s12652-023-04530-y

Karrachitawaragul, C. (2016). *The Application of Geographic Information System for Analysis of Illegal Migration and Drug Trafficking Pattern: A Case Study of Tak Province*. Retrieved 3/01/2024 from https://ir.swu.ac.th/jspui/handle/123456789/28470

Karthikraja, C., Senthilkumar, J., Hariharan, R., Usha Devi, G., Suresh, Y., & Mohanraj, V. (2022). Retracted: An empirical intrusion detection system based on XGBoost and bidirectional long-short term model for 5G and other telecommunication technologies. *Computational Intelligence, 38*(4), 1216-1231. https://doi.org/https://doi.org/10.1111/coin.12497

Kaur, H. (2021). MACHINE LEARNING AND BORDER SECURITY. *Palarch's Journal of Archaeology Of Egypt/Egyptology*.

Khanawiwat, S. C. T. (2020). *The Report on the Route of Migration from Myanmar and*

Cambodia to Thailand. Plan International Thailand. Retrieved 5/10/2023 from https://plan-international.org/uploads/sites/82/2022/02/migration_route_report.eng_.fairfish.pdf

Kim, S., & Lee, S. (2023, 2023/04/01/). 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, 91*, 104419. https://doi.org/https://doi.org/10.1016/j.scs.2023.104419

Kogan, F. N. (1995, 1995/01/01/). Application of vegetation index and brightness temperature for drought detection. *Advances in Space Research, 15*(11), 91-100. https://doi.org/https://doi.org/10.1016/0273-1177(95)00079-T

Kotsiopoulos, T., Sarigiannidis, P., Ioannidis, D., & Tzovaras, D. (2021, 2021/05/01/). Machine Learning and Deep Learning in smart manufacturing: The Smart Grid paradigm. *Computer Science Review, 40*, 100341. https://doi.org/https://doi.org/10.1016/j.cosrev.2020.100341

Lakshmi, S., & Yarrakula, K. (2019, 01/14). Review and critical analysis on digital elevation models. *Geofizika, 35*, 129-157. https://doi.org/10.15233/gfz.2018.35.7

Latrubesse, E., & Park, E. (2017). Rivers and Streams. In. https://doi.org/10.1002/9781118786352.wbieg1027

McAuliffe, M. a. A. T. e. (2021). *World Migration Report 2022*. International Organization for Migration. Retrieved 2/10/2023 from https://publications.iom.int/books/world-migration-report-2022

Melesse, A. M., Weng, Q., Thenkabail, P. S., & Senay, G. B. (2007). Remote Sensing Sensors and Applications in Environmental Resources Mapping and Modelling. *Sensors, 7*(12), 3209-3241.

MIGRATION, I. O. F. *Key Migration Terms*. Retrieved 1/10/2023 from https://www.iom.int/key-migration-terms

Migration, I. O. f. (2019). *Glossary on Migration*. International Organization for Migration. Retrieved 5/01/2024 from https://publications.iom.int/books/international-migration-law-ndeg34-glossary-migration

Montalván-Burbano, N., Velastegui-Montoya, A., Gurumendi-Noriega, M., Morante-

Carballo, F., & Adami, M. (2021). Worldwide Research on Land Use and Land Cover in the Amazon Region. *Sustainability, 13*(11), 6039. https://www.mdpi.com/2071-1050/13/11/6039

Nascimento, C. M., de Sousa Mendes, W., Quiñonez Silvero, N. E., Poppiel, R. R., Sayão, V. M., Dotto, A. C., Valadares dos Santos, N., Accorsi Amorim, M. T., & Demattê, J. A. M. (2021, 2021/01/01/). Soil degradation index developed by multitemporal remote sensing images, climate variables, terrain and soil attributes. *Journal of Environmental Management, 277*, 111316. https://doi.org/https://doi.org/10.1016/j.jenvman.2020.111316

Nikiforov, A. R., Nikiforova, A. A., & Petlukova, E. A. (2020, 2020/04/01). Taluses and Downfall Slopes as Geomorphological Refugia of Flora Relict Endemics of the Mountain Crimea and their Modern Phytoindication. *IOP Conference Series: Earth and Environmental Science, 459*(4), 042016. https://doi.org/10.1088/1755-1315/459/4/042016

Office, C. R. P. S. (2024). *Chiang Rai Provincial Statistical Report 2024*. Statistical Forecasting Bureau,

National Statistical Office. Retrieved 10/03/2024 from https://chiangrai.nso.go.th/

Office of Social Development and Human Security, C. R. P. (2021). *Chiang Rai Provincial Social Situation Report 2021*. Retrieved 11/03/2024 from https://www.m-society.go.th/ewtadmin/ewt/mso_web/download/article/article_20211110150503.pdf

Okabe, A., Satoh, T., & Sugihara, K. (2009, 2009/01/01). A kernel density estimation method for networks, its computational method and a GIS‐based tool. *International Journal of Geographical Information Science, 23*(1), 7-32. https://doi.org/10.1080/13658810802475491

Onanong Thippimol, T. S., Dulyapak Preecharushh, Supalak Ganjanakhundee, Akkharaphong Khamkhun. (2011). *Boundaries of Siam/Thailand-Malaysia-Burma-Laos-Cambodia*. The Foundation for The Promotion of Social Science and Humanities Textbooks Project.

Organization, I. L. (2014). *Migration irregular status and deportation: An outline of international standards regional trends and good practices*. Retrieved 5/01/2024 from https://www.ilo.org/resource/migration-south-east-asia-policy-brief-migration-irregular-status-and

Patino, J. E., Duque, J. C., Pardo-Pascual, J. E., & Ruiz, L. A. (2014, 2014/12/01/). Using remote sensing to assess the relationship between crime and the urban layout. *Applied Geography, 55*, 48-60. https://doi.org/https://doi.org/10.1016/j.apgeog.2014.08.016

Pless, E., Eckburg, A. M., & Henn, B. M. (2023). Predicting Environmental and Ecological Drivers of Human Population Structure. *Molecular Biology and Evolution, 40*(5). https://doi.org/10.1093/molbev/msad094

Rai, P. o. o. C. (2022). *Chiang Rai Provincial Development Plan 2023-2027*. Retrieved 5/01/2024 from https://chiangrai.gdcatalog.go.th/dataset/planning

Ren, Z., Liu, Y., Chen, B., & Xu, B. (2020). Where Does Nighttime Light Come From? Insights from Source Detection and Error Attribution. *Remote Sensing, 12*(12), 1922. https://www.mdpi.com/2072-4292/12/12/1922

Salas, E. A. L., & Kumaran, S. S. (2023). Hyperspectral Bare Soil Index (HBSI): Mapping Soil Using an Ensemble of Spectral Indices in Machine Learning Environment. *Land, 12*(7), 1375. https://www.mdpi.com/2073-445X/12/7/1375

Santos, M. S., Soares, J. P., Abreu, P. H., Araujo, H., & Santos, J. (2018). Cross-Validation for Imbalanced Datasets: Avoiding Overoptimistic and Overfitting Approaches [Research Frontier]. *IEEE Computational Intelligence Magazine, 13*(4), 59-76. https://doi.org/10.1109/MCI.2018.2866730

Sapprasert, C. A. (2017). *Problems of Illegal Immigrants with Thailand's security (Chantaburi and Trat Provinces)* Army War College]. http://awc.ac.th/storage/research/24.pdf

Schuster, J. (2020, 07/20). LandScan. *The Charleston Advisor, 22*, 37-40. https://doi.org/10.5260/chara.22.1.37

Shahriar Pervez, M., Budde, M., & Rowland, J. (2014, 2014/06/01/). Mapping irrigated areas in Afghanistan over the past decade using MODIS NDVI. *Remote Sensing of Environment, 149*, 155-165. https://doi.org/https://doi.org/10.1016/j.rse.2014.04.008

Singh, S., Gole, A., Kanherkar, P., Abhishek, P., & Shelke, P. (2022). *Crime Prediction using Machine Learning*. https://doi.org/10.1109/ICETEMS56252.2022.10093303

Small, C., Sousa, D., Yetman, G., Elvidge, C., & MacManus, K. (2018, 2018/06/01/).

Decades of urban growth and development on the Asian megadeltas. *Global and Planetary Change, 165*, 62-89. https://doi.org/https://doi.org/10.1016/j.gloplacha.2018.03.005

Taylor, F. W. (1996, 1996/09/01). Remote sensing of the earth from space. *Contemporary Physics, 37*(5), 391-405. https://doi.org/10.1080/00107519608217544

Trimek, A. B. a. J. (2020). *Myanmar Smugglers in Thailand after the Opening of ASEAN Community (AC)* RSU National Research Conference 2020. Retrieved 6/01/2024 from https://rsucon.rsu.ac.th/proceeding/article/2353

van der Geest, K., Vrieling, A., & Dietz, T. (2010). Migration and environment in Ghana: a cross-district analysis of human mobility and vegetation dynamics. *Environment & Urbanization, 22*, 107 - 123.

Zeng, M., Mao, Y., & Wang, C. (2021, 2021/05/01/). The relationship between street environment and street crime: A case study of Pudong New Area, Shanghai, China. *Cities, 112*, 103143. https://doi.org/https://doi.org/10.1016/j.cities.2021.103143

Zhang, C., Jia, D., Wang, L., Wang, W., Liu, F., & Yang, A. (2022, 2022/10/01/). Comparative research on network intrusion detection methods based on machine learning. *Computers & Security, 121*, 102861. https://doi.org/https://doi.org/10.1016/j.cose.2022.102861

Zhang, J., Mucs, D., Norinder, U., & Svensson, F. (2019, 09/27). LightGBM: An Effective and Scalable Algorithm for Prediction of Chemical Toxicity-Application to the Tox21 and Mutagenicity Data Sets. *Journal of Chemical Information and Modeling, 2019*. https://doi.org/10.1021/acs.jcim.9b00633

Zhang, W., Zhang, R., Wu, C., Goh, A. T. C., Lacasse, S., Liu, Z., & Liu, H. (2020, 2020/07/01/). State-of-the-art review of soft computing applications in underground excavations. *Geoscience Frontiers, 11*(4), 1095-1106. https://doi.org/https://doi.org/10.1016/j.gsf.2019.12.003

Zhao, X., & Tang, J. (2017). *Modeling Temporal-Spatial Correlations for Crime Prediction* https://doi.org/10.1145/3132847.3133024

Zhou, Q. (2017). Digital Elevation Model and Digital Surface Model. In. https://doi.org/10.1002/9781118786352.wbieg0768

BIOGRAPHY

NAME	Pol.Sgt.Janjira Khuenmamueang
DATE OF BIRTH	10 March 1992
PLACE OF BIRTH	Chiang Rai
PRESENT ADDRESS	39/459 Phipornpong 4 village, Vibhavadi 84 Road, Sanambin, Donmueang. Bangkok. 10210.
POSITION HELD	Immigration officer
EDUCATION	Public Administration, Chiang Mai University