Doctoral Dissertation

Grid-Level Snow Disaster Risk Assessment and Resilience Modeling Based on Interpretable Machine Learning and Multi-Source Spatial Data

Zhenyu, YANG

Supervisor: Hideomi, GOKON

Graduate School of Advanced Science and Technology
Japan Advanced Institute of Science and Technology
(School of Knowledge Science)

September, 2025

Dissertation Abstract

Approximately 60% of Japan's land area and 20% of its population are affected by heavy snowfall each year. Snow disasters often trigger traffic accidents and congestion, resulting in significant casualties and economic losses. These events pose severe challenges to urban transportation systems, residents' travel safety, and regional emergency response capacity. Traditional disaster response models struggle to accurately and promptly identify high-risk areas, limiting the effective allocation of disaster prevention resources. Against this backdrop, developing snow disaster risk identification and resilience assessment models based on multi-source spatial data and explainable machine learning is of great significance for enhancing urban disaster preparedness and promoting sustainable regional development.

This study takes the 2018 heavy snowstorm in Fukui Prefecture as a case study and proposes a grid-scale risk modeling framework integrating multi-source spatial data, including mobile GPS data, road data, land use data, digital elevation models (DEM), urban area data, snow depth data, traffic congestion data, nighttime light data (DNB), and the Normalized Difference Vegetation Index (NDVI). The dissertation consists of four core components: (1) Based on KDTree nearest-neighbor matching and singular value decomposition (SVD), the study performs dimensionality reduction on the spatiotemporal OD matrix to analyze the evolution of intercity travel patterns during the snowstorm and reveal disruptions in transportation demand structures; (2) A grid-level analysis of major roads in Fukui Prefecture is conducted to identify critical congestion points along National Route 8 and the Hokuriku Expressway. A Random Forest-based traffic congestion identification model is developed to assess snowstorm-induced traffic bottlenecks(3) A "triangular resilience framework" is established to evaluate robustness, vulnerability, and adaptability across land use types at a 500-meter grid level using elastic triangle theory, quantifying the disaster resilience of various urban functional zones; (4) Multiple machine learning models, including MLP, Decision Tree, Random Forest, SVM, LightGBM and XGBoost, are applied to classify high-risk snowstorm areas. SHAP values are used to interpret the models, revealing key influencing factors and nonlinear threshold effects, thereby generating interpretable spatial risk maps to support disaster early warning and emergency response.

The major findings are as follows:

(1) Using KDTree and SVD, the study identifies spatiotemporal patterns of transportation demand in Fukui during the snowstorm. The snow event can be divided into three phases: stable (Jan 27–Feb 2), disaster (Feb 3–Feb 11), and recovery (Feb 12–Feb 16). Intercity travel demand declined by 67.86% during the snowstorm. Demand patterns included regular demand (M_1) from cities such as Fukui, Sabae, Awara, and Sakai, and special demand (M_2), where northbound trips nearly ceased, shifting from a longitudinal pattern along National Route 8 and the Hokuriku Expressway to a horizontal pattern toward Ono

and Katsuyama. Travel outflows from Sakai dropped significantly, while Tsuruga's inflows increased, indicating the snowstorm's primary impact on northern Fukui, particularly Fukui City and Sakai.

(2) GPS data effectively captures population dynamics along road networks. In the case of the 2018 Fukui snow disaster, all locations where traffic volume surged sharply during the snowstorm were situated in urban areas, with predominant land use types being fields and forests. This phenomenon may be attributed to factors such as inadequate transportation infrastructure, the urban heat island effect, and steep terrain conditions. By integrating remote sensing data and GPS data with machine learning models, congested road sections can be effectively identified. Feature importance analysis highlighted the most influential variables in predicting congestion as follows: Snow Depth > Nighttime Light Difference > Elevation > Slope Angle > Urban Area > NDVI > Population Change > Forest > Field > Low-rise Buildings (Sparse) > Low-rise Buildings (Dense).

(3) The resilience of different land use types is evaluated at the 500-meter grid scale using the triangular resilience method. The results show that factory zones, agricultural areas, building land, and mid-/high-rise buildings demonstrate strong resilience, especially in their ability to absorb and recover from shocks. Prioritizing resilient building forms in urban planning can significantly improve disaster readiness. In contrast, sparse low-rise housing, parks, facilities, and dense low-rise neighborhoods exhibit weaker recovery and require targeted strengthening measures.

(4) By combining XGBoost and SHAP, an interpretable spatial risk assessment model is developed. The model quantifies high-risk areas and identifies key drivers with nonlinear relationships and threshold effects. Elevation, slope, DNB, NDVI, road density, and road width show significant nonlinear influences on snow risk. Snowstorms primarily affect primary and secondary roads, while rural roads are relatively less impacted. Real-time GPS monitoring and optimized road layouts can help improve transportation resilience during snow disasters.

The main innovations of this study are as follows:

(1) Integration of multi-source spatial data (GPS, land use, DEM, road networks, nighttime light, NDVI), providing a solid data foundation for high-precision modeling and interpretability analysis; (2) Combination of SVD and trajectory analysis to reveal dynamic disruptions in transportation demand during snowstorms; (3) Proposal of a "triangular resilience index" system (robustness, vulnerability, adaptability) to quantify resilience at a 500-meter grid level; (4) Application of XGBoost and SHAP to identify nonlinear relationships and threshold effects of key factors, enhancing interpretability and practical usability; (5) The proposed modeling framework demonstrates strong adaptability and transferability, applicable to other natural disasters such as floods, typhoons, and tsunamis. It supports the

investigation of critical impact mechanisms and resilience under different disaster scenarios.

This research provides scientific evidence for forecasting transportation risks during snow disasters in the Hokuriku region of Japan, optimizing urban functional zone layouts, and guiding disaster preparedness. It contributes to enhancing the overall resilience of urban systems in the context of climate change.

Keywords: Snow disaster; multi-source data; spatiotemporal characteristics; resilience; machine learning; SHAP

Acknowledgment

As time flies, it has already been three years since I joined Gokon Sensei's lab in the autumn of 2022. During this period, I had the privilege of meeting wonderful friends, acquiring new knowledge and methodologies, and learning to view the world from a broader and more rational perspective. Now, as I stand at the threshold of graduation, the confusion and uncertainties of the past have gradually faded, and this dissertation marks the closing of an important chapter in my life. The guidance of my teachers, the companionship of my lab mates through moments of both joy and hardship, and the dreams and aspirations nurtured during my doctoral journey have become invaluable treasures that I will carry with me forever.

I am deeply grateful to my parents for their unconditional love, sacrifices, and unwavering support. In every moment of setback and discouragement, it was their quiet strength and heartfelt encouragement that gave me the courage to stand up again and continue moving forward. Thank you for being my greatest source of strength.

I would like to express my heartfelt gratitude to my supervisor, Gokon Sensei, for his patient guidance, insightful academic instruction, and constant encouragement over the past three years. Your rigorous approach to research and your kindness have greatly enriched both my academic journey and personal growth. I am sincerely thankful for the valuable lessons you have taught me.

I am also thankful to myself for persevering through these challenging years. During my doctoral studies, I learned the importance of self-discipline and the joy of exploring the unknown. Every failure became a stepping stone for growth, and I have never lost the light in my eyes. This journey has shaped me into a more rational yet warm-hearted person. I hope to carry this passion, curiosity, and resilience with me into the future.

I would also like to express my heartfelt gratitude to Mr. Qing Yu from the Shenzhen Institute of Peking University. Thank you for generously sharing your instructional videos on GPS data processing. Your selfless contribution has greatly benefited my research and deepened my understanding of spatiotemporal data analysis.

I would also like to extend my sincere thanks to Taniguchi san from our lab and my friend Jianwen Sun. Thank you both for taking me fishing and introducing me to the joy of lure fishing. Jianwen, your impressive 78 cm sea bass has become my target to beat this year—it's been a true inspiration and a fun challenge to pursue.

Finally, I sincerely thank all the professors and experts who took the time to review my dissertation despite their busy schedules. Your thoughtful feedback and encouragement are deeply appreciated.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

Snow disasters refer to events caused by excessive snowfall, prolonged snow accumulation, or the persistence of snow layers [10]. These disasters primarily occur in regions such as the northeastern United States, southwestern Canada, Siberia, northeastern China, the Altai region, Europe, and Japan [11], as illustrated in Fig. 1.1. Compared to other natural disasters such as earthquakes, floods, landslides, hurricanes, and tsunamis, snow disasters have received relatively less research attention. However, their economic losses and human casualties are equally significant. Table 1.1 summarizes major snow disaster events in recent years that have resulted in documented fatalities and economic damage . Nevertheless, many snow disasters remain underreported due to incomplete statistical data. For instance, the 2009–2010 East Asia winter storm, the 2010 European snowstorm, the 2021 Storm Filomena in Spain, and the 2023 Canada ice storm all caused severe socio-economic impacts in the affected regions. However, due to data limitations, the full extent of their consequences remains difficult to quantify. This suggests that the impact of snow disasters is far more extensive than what existing data can capture. Therefore, strengthening snow disaster risk assessment and formulating effective disaster response strategies are of paramount importance.

Table 1.1: Serious Snow Events in Recent Years [5, 6, 2, 7, 8, 9]

Year	Country	Casualties	Economic Loss	Indirect Impact
2008	China	129 deaths	$207.76 billion	House collapses
2018	Japan	12 deaths, 26 seriously injured	$4.8 million	Traffic jam
2018	United Kingdom	95 deaths	–	Flight delays
2021	America	At least 246 deaths	$195 billion	Power outage
2024	China	4 deaths, 32 injured	$3.57 billion	Traffic congestion

Consequently, reducing the losses caused by snow disasters and building resilient cities

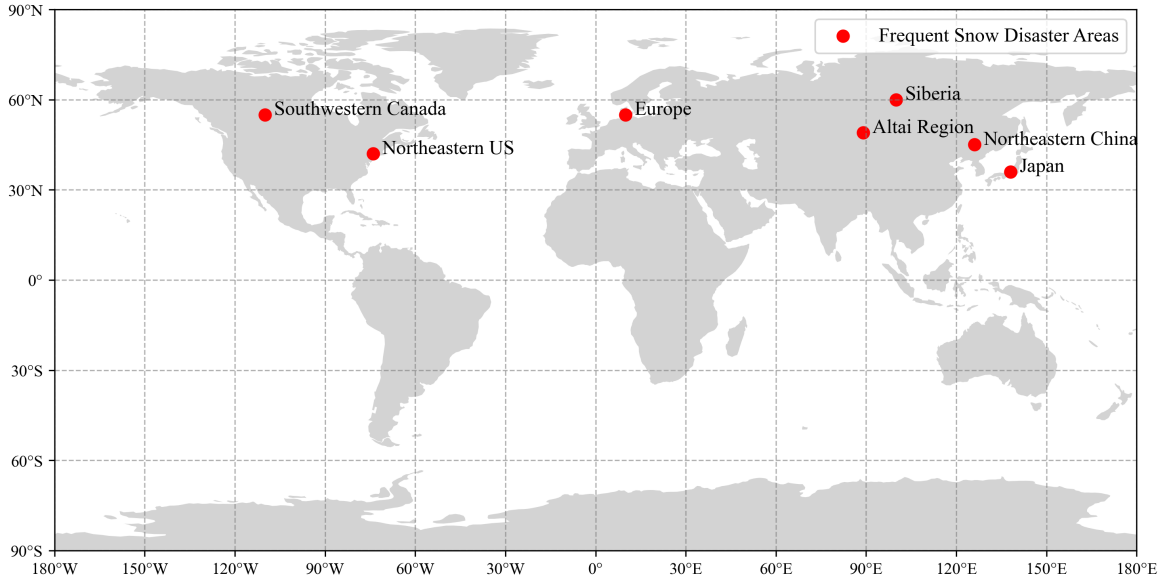


Figure 1.1: Frequent Snow Disaster Regions

has become a critical issue. The concept of resilient cities encompasses the capacity of urban areas to prepare for, absorb, recover from, and adapt to adverse events, thereby minimizing the social and economic impacts of disasters [12, 13]. This resilience framework involves multiple components, including risk assessment, sustainable urban planning, and community engagement [14].

In the context of snow disasters, resilient cities adopt specific strategies to address the unique challenges posed by heavy snowfall and icy conditions. For instance, urban planners can design buildings and infrastructure with enhanced snow load capacities and improve drainage systems to manage snowmelt effectively [15]. Additionally, transportation resilience is crucial, incorporating adaptive road maintenance plans, efficient snow removal systems, and real-time traffic management to prevent disruptions during severe snowstorms [16, 17]. Emergency preparedness and early warning systems further strengthen urban resilience by providing timely alerts and guiding residents and authorities in taking preventive measures [18].

Moreover, a data-driven approach to identifying and assessing high-risk areas for snow disasters in advance is essential. Such an approach enables cities to implement targeted interventions before disasters strike, mitigating the direct impacts of snow events and accelerating post-disaster recovery. Ultimately, this ensures a faster return to normalcy.

1.2 Research Gaps

1.2.1 Research Data Gaps

Previous studies on snow-related disasters have primarily relied on Electronic Toll Collection (ETC) data and satellite imagery [19, 20]. However, ETC data are limited to vehicles equipped with such systems, which introduces sampling bias and restricts comprehensive traffic monitoring. On the other hand, satellite-based observations often require ultra-high-resolution imagery and complex image processing workflows, and are further constrained by spatial resolution and temporal update frequency. In contrast, mobile phone-based GPS data provide broader spatial coverage, higher temporal resolution, and greater flexibility. In recent years, GPS data have emerged as a prominent source for analyzing mobility patterns and transportation demand [21, 22, 23].

Therefore, integrating GPS data with other geospatial datasets such as Digital Elevation Models (DEM), Normalized Difference Vegetation Index (NDVI), night-time light imagery, and land use information offers a promising approach for constructing comprehensive spatiotemporal databases. This integrated data framework is well-suited for evaluating the impact mechanisms of snow disasters on urban systems and supporting the assessment of urban resilience under extreme weather conditions.

1.2.2 Research Scope Gaps

Although a growing body of research has explored the concept of urban resilience, most existing studies remain focused on the city-wide or regional scale. In contrast, fine-grained analyses at the micro-grid level are relatively limited (see Table 1.2). In Japan, approximately 60% of the national land area and 20% of the population are affected by snow-related disasters each year [24]. However, most disaster resilience maps issued by national and local governments prioritize hazards such as earthquakes, typhoons, and tsunamis, with limited attention given to snow disasters.

Therefore, assessing spatial resilience characteristics under snowstorm conditions at the grid level is critical for identifying disparities in exposure and response capacity among different areas. Such fine-scale evaluations can provide practical insights for optimizing resource allocation and strengthening disaster preparedness at the local level.

1.2.3 Research Methodological Gaps

Current approaches to snow disaster risk assessment still face several notable limitations. First, many studies still rely on traditional risk assessment methods, such as the Analytic Hierarchy Process (AHP) and linear statistical models, which typically assume linear relation-

Table 1.2: Studies on Disaster Resilience

Reference	Data	Disaster	Research Method	Resolution
Yabe et al. [25]	GPS data	Earthquake	Mobilkit	Municipio and localidades levels
Hara et al. [26]	GPS mobile data	COVID-19	MSS	Prefecture level
Yabe et al. [3]	GPS data	Hurricane	BSTS	Regions level
Li et al. [27]	Snow depth, GDP, elevation and slope	Snow	AHP	City-level
Yang et al. [28]	GPS taxi data	Snow	The Bayesian models	National level
Jamal et al. [29]	Facebook data	Hurricane	Resilience curve	County level
This research	GPS mobile data, land use, NDVI, DEM, etc.	Snow	SVD, Resilience curve, XGBoost, SHAP, etc.	Grid level

Abbreviations: Mobilkit: Mobile Toolkit for Disaster Response; MSS: Mobile Spatial Statistics; BSTS: Bayesian Structural Time Series; AHP: Analytic Hierarchy Process.

ships among variables and fail to effectively capture the complex and nonlinear interactions among snow-related factors [30, 31]. Second, due to limitations in data acquisition, remote sensing–based monitoring methods suffer from insufficient spatial and temporal resolution, making it difficult to accurately reflect the impacts of snow disasters at micro-scales [32, 33]. Third, current research lacks model interpretability and practical usability, and effective tools that can provide clear policy recommendations are still missing [34].

In summary, current studies still lack a high-performance risk assessment tool that combines nonlinear modeling capabilities with interpretability. Therefore, it is urgent to introduce flexible, transparent, and spatially expressive machine learning methods to improve the accuracy, scientific value, and decision-making relevance of snow disaster risk assessment.

1.3 Research Objectives

This study aims to develop an interpretable spatial machine learning framework for identifying high-risk snow disaster areas at the micro-grid scale, thereby enhancing the effectiveness of pre-disaster risk assessment and post-disaster emergency response. In response to current limitations in data diversity, spatial resolution, and model interpretability, the

following three research objectives are proposed:

- **To construct a multi-source snow disaster analysis database**: This involves integrating various geospatial datasets, including mobile phone GPS data, Digital Elevation Models (DEM), Normalized Difference Vegetation Index (NDVI), nighttime light imagery, and land use data, to establish a high spatiotemporal resolution database for snow disaster risk assessment. Based on this database, the study further aims to identify the spatiotemporal patterns of intercity transportation demand during snowstorm events.
- **To evaluate resilience differences among land use types at the micro scale**: Taking Fukui Prefecture as the case study area, this research will use 500-meter grid cells and apply the Elastic Triangle Method to assess resilience from three dimensions: robustness, vulnerability, and recoverability. It seeks to quantitatively evaluate the spatial resilience characteristics of different land use types (e.g., building land, facility land, agricultural land) under snowstorm conditions, and to reveal their differential responses to snow disasters.
- **To develop an interpretable snow disaster risk identification model**: This objective involves combining machine learning techniques with SHAP (SHapley Additive exPlanations) values to build a spatially explicit, non-linear, and interpretable model. The model will quantify the relative impact of terrain, road infrastructure, and urban characteristics on snow disaster risks, and produce high-resolution risk prediction maps and influence factor distribution maps, thereby supporting targeted resource allocation and evidence-based policy-making.

1.4 Research Contributions

This study integrates multi-source spatial data and interpretable machine learning techniques to construct a micro-scale framework for snow disaster risk identification and resilience modeling. The main research contributions are as follows:

- **Proposed a multi-source data integration framework for snow disaster analysis**: This study integrates mobile phone GPS data, road network data, land use data, digital elevation model (DEM), nighttime light data, and NDVI to construct a high spatiotemporal resolution database for snow disaster risk analysis. This database provides a solid foundation for subsequent dynamic modeling and interpretable analysis.
- **Revealed the spatiotemporal evolution of intercity travel demand during snowstorms**: Using KDTree nearest-neighbor matching and Singular Value Decomposition

(SVD), this study identifies the dynamic changes in intercity transportation demand during snowstorm periods, offering a new data-driven perspective for understanding population mobility under disaster conditions.

- **Developed a grid-based resilience assessment framework using the "resilience triangle" indicators**: From the dimensions of robustness, vulnerability, and recoverability, this study quantitatively evaluates the snow disaster resilience of different land use types at a 500-meter grid scale. It reveals spatial disparities in disaster resilience across urban functional zones and provides reference for urban zoning and disaster mitigation planning.

- **Developed an interpretable machine learning model for snow disaster risk identification**: By combining machine learning models with SHAP (SHapley Additive exPlanations), this study quantifies the nonlinear influence and threshold effects of key factors on snow disaster risk. It also generates high-resolution risk prediction maps and factor distribution maps, enhancing the model's scientific validity and policy relevance.

- **Proposed a transferable and generalizable modeling framework**: The risk identification and resilience assessment framework developed in this study is adaptable to other natural disasters such as floods, typhoons, and tsunamis. It supports fine-scale disaster mechanism identification and resilience analysis under diverse disaster scenarios.

1.5 Thesis Structure

This dissertation is composed of eight chapters, with the content of each chapter outlined as follows. As illustrated in Fig. 1.2, this thesis focuses on four core research chapters, each addressing a key aspect of snow disaster analysis.

- **Chapter 1: Introduction**

  Introduces the research background and motivation, identifies existing problems and limitations in current studies, and clearly states the research objectives and contributions, laying the foundation for the subsequent chapters.

- **Chapter 2: Literature Review**

  Reviews the research progress in spatial data analysis under disasters, urban resilience, snow disaster studies, and interpretable machine learning. This chapter also discusses existing limitations and defines the academic gaps this research aims to fill.

- **Chapter 3: Study Area, Data Sources, and Preprocessing**

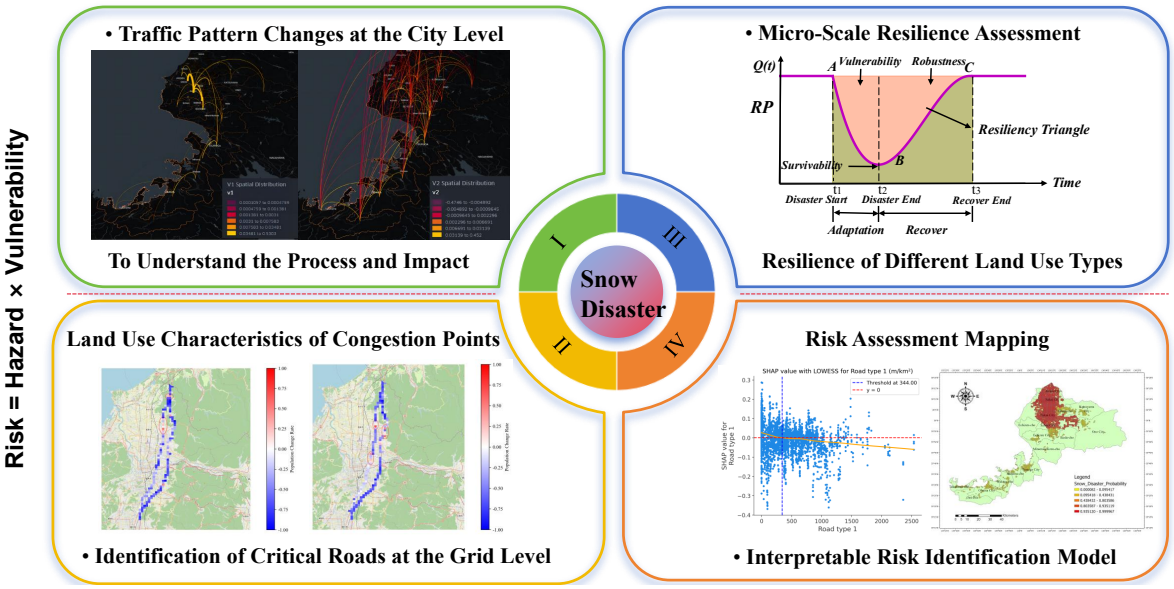


Figure 1.2: Four Core Research Chapters of This Thesis

Describes the study area (Fukui Prefecture) and the multi-source spatial datasets used, including mobile phone GPS data, road networks, land use data, Digital Elevation Model (DEM), nighttime lights, and NDVI. It also details the preprocessing methods applied to each data type.

- **Chapter 4: Spatiotemporal Analysis of Intercity Travel Demand During Snowstorms**

  Uses KDTree-based nearest neighbor matching and Singular Value Decomposition (SVD) to identify the evolution of intercity travel patterns during the 2018 Fukui snowstorm. The chapter analyzes disruptions in transportation demand and changes in spatial flow structures.

- **Chapter 5: Detecting High-Risk Traffic Congestion Areas**

  This chapter conducts grid-level analysis to investigate traffic congestion during snow disasters, focusing on critical corridors such as National Route 8 and the Hokuriku Expressway. The Random Forest model is employed to identify high-risk traffic congestion grid cells, providing insights for disaster management and emergency response.

- **Chapter 6: Micro-Scale Resilience Assessment of Land Use Types**

  Proposes a "resilience triangle" framework to evaluate urban resilience from the dimensions of robustness, vulnerability, and recoverability. At a 500-meter grid scale, this chapter assesses the resilience of different land use categories in snowstorm conditions.

- **Chapter 7: Risk Identification Modeling and Interpretability Analysis**

Builds a classification model using machine learning and SHAP theory to identify high-risk snow disaster areas. The model outputs are explained using SHAP values to uncover the key influencing factors and their nonlinear threshold effects, resulting in interpretable risk maps.

- **Chapter 8: Conclusion and Future Work**

  Summarizes the major findings of this study, discusses methodological and data-related limitations, and proposes future research directions, including the potential applicability of the proposed framework to other types of natural disasters.

This dissertation is composed of four core chapters (Chapters 4–7), each addressing a specific dimension of snowstorm-related transportation impact, but collectively contributing to an integrated analytical framework.

Chapter 4 adopts a macro-scale perspective to analyze how snow disasters affect regional mobility patterns, particularly intercity transportation. The findings from this chapter also help define the dependent variable (Y) used in subsequent spatial modeling tasks. Chapter 5 shifts the focus to road-level analysis by identifying traffic congestion hotspots through population-based road usage patterns. This complements the regional analysis in Chapter 4 and highlights that congestion frequently occurs in specific land use categories, providing the empirical motivation for Chapter 6. Chapter 6 examines the resilience of different land use types in absorbing and recovering from snowstorm-induced disruptions. This chapter deepens the analysis by evaluating spatial heterogeneity in transportation recovery across urban forms. Finally, Chapter 7 integrates multiple variables—including environmental, infrastructural, and topographical features—into a micro-scale spatial machine learning model. By leveraging explainable AI (e.g., SHAP values), it identifies nonlinear risk patterns and enhances the interpretability of snow disaster predictions.

Together, these four chapters construct a cohesive research framework, progressing from regional dynamics to fine-scale vulnerability assessment. This integrated approach contributes to a more comprehensive understanding of snowstorm risk and offers practical insights for disaster preparedness and resilient transportation planning.

Chapter 2

Literature Review

2.1 Overview of Snow Disaster Research

Snow disasters refer to events caused by excessive snowfall, prolonged snow accumulation, or the persistence of snow layers [10]. Snow disasters cause severe disruptions to aviation, transportation, agriculture, and livestock, leading to widespread flight delays, traffic congestion, power outages, damage to crops and livestock. However, studies specifically focused on snow disasters remain relatively scarce, as shown in Fig. 2.1.

Compared to sudden and destructive events such as floods or earthquakes, snow disasters affect urban systems in a more gradual, widespread, and systemically coupled manner. Snowstorm damage typically accumulates over time—through prolonged snowfall, road icing, and reduced visibility—and although it lacks the instantaneous shock of an earthquake, it can cause persistent and cumulative disruptions to transportation, logistics, and daily mobility. Urban transportation systems are particularly sensitive to snow events; even moderate snowfall can lead to widespread traffic paralysis if roads freeze or snow removal is delayed, thereby affecting essential functions such as commuting, emergency services, and supply chains. In contrast, floods tend to be geographically concentrated in low-lying areas, while earthquakes often lead to immediate structural damage.

Furthermore, snow disasters exhibit marked spatiotemporal heterogeneity in urban environments. Factors such as elevation, topography, road network density, and land use types contribute to varying levels of impact across different areas. This dual complexity—both geographical and functional—makes modeling urban snow disaster impacts more challenging than many other hazards. Consequently, a dynamic, high-resolution analytical approach is necessary. This study leverages mobile GPS data and machine learning methods to capture the evolving nature of snow disaster risk at fine spatial and temporal scales. These characteristics underscore the theoretical and practical significance of focusing on snow disasters and justify the need for high-frequency mobility data and explainable spatial modeling methods,

which form the foundation of this study.

Currently, research on snow disasters can be broadly divided into three main categories: snow disaster risk studies, snow disaster impact studies, and snow depth detection and monitoring, as shown in Fig. 2.2.

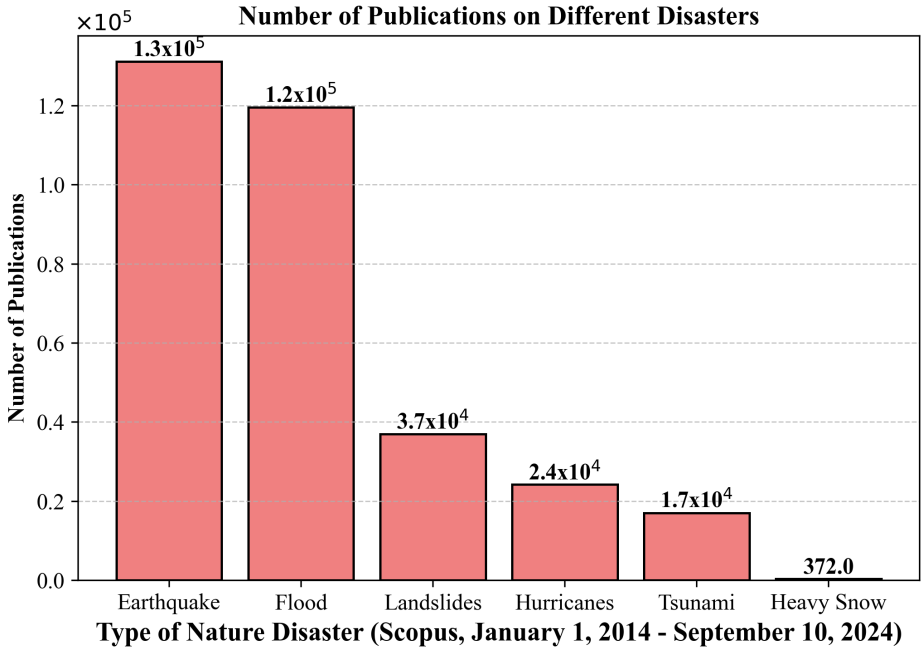


Figure 2.1: Publications of Disasters

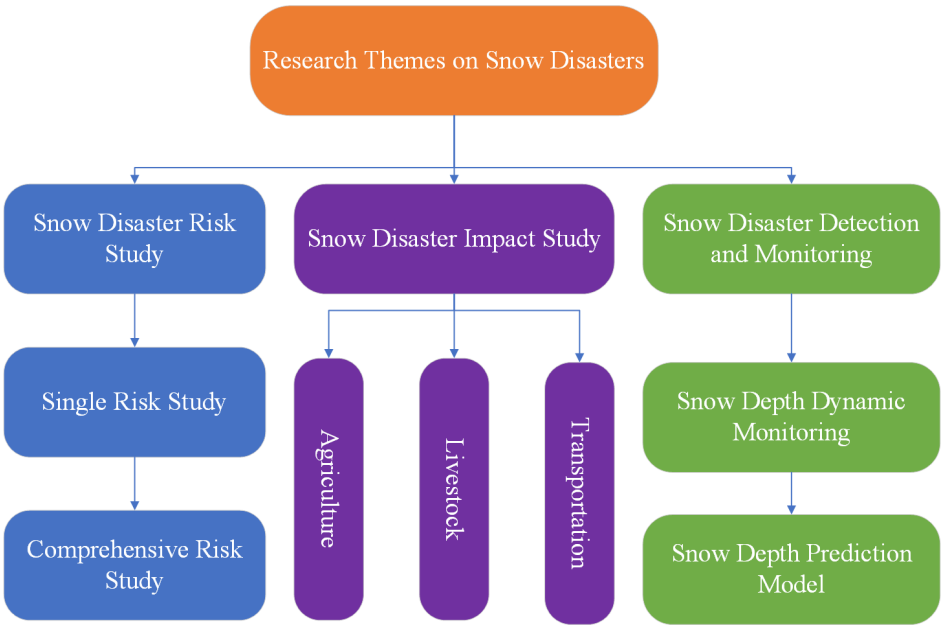


Figure 2.2: Research Themes on Snow Disasters

2.1.1 Snow Disaster Risk Studies

Snow disaster risk studies include assessments of single snow disaster risks and comprehensive regional risk assessments. Single risk assessments focus on specific events, while regional risk assessments consider multiple snow disaster factors to conduct systematic evaluations at a regional level [27]. For instance, Cappabianca et al. combined snowfall record analysis, avalanche dynamics simulations, and vulnerability relationships to assess avalanche risk [35]. Germain et al. reconstructed historical avalanche events using tree-ring data and analyzed avalanche risks in the Gaspé Peninsula [36]. Liu et al. developed a risk index model based on blizzard occurrence frequency to assess future blizzard risks in Northeast China [37]. Zhang et al. used long-term snowfall data from Heilongjiang Province to create a snow disaster index model for regional risk assessment [38]. Gao et al. developed a snow disaster risk model for Qinghai Province by integrating snow depth, terrain features, and socioeconomic factors [39]. Other studies have broadened the methodologies and data sources used for snow disaster risk assessment. For instance, Choubin et al. explored snow avalanche hazard prediction using machine learning models such as Support Vector Machine and Multiple Discriminant Analysis, underscoring the potential of advanced statistical techniques in hazard assessment [40]. Ahmad et al. highlighted the integration of remote sensing networks with real-time data to enhance risk detection and management, particularly in mountainous regions susceptible to heavy snowfall [41]. Sun et al. conducted a comprehensive assessment of snow disaster vulnerability, utilizing long-term remote sensing data and meteorological information to model impacts on the Qinghai-Tibet Plateau [42]. Additionally, Eckert and Giacona proposed a holistic paradigm for long-term avalanche risk management that addresses the dynamic nature of snow-related hazards [34]. Lee et al. applied a socio-economic approach to classify snow disaster risk levels in South Korea, considering economic impacts and resource vulnerabilities in their assessments [31]. Gascoin et al. offered recommendations on remote sensing for monitoring snow in mountainous areas, emphasizing the importance of spatial data in capturing snow distributions and enhancing predictive models [32]. Additionally, Papucci et al. [43] combined remote sensing and meteorological data to establish a forest snow disaster risk prediction model in Sweden, identifying key influencing factors such as tree height, snow depth, dominant tree species, temperature, wind speed, wind direction, soil depth, and clear-cutting distance.

Despite these advancements in snow disaster risk assessment, GPS data applications remain largely unexplored. In recent years, GPS data has gained increasing attention in disaster monitoring and risk assessment due to its high spatiotemporal resolution and real-time characteristics [25, 44, 45, 46, 47]. Therefore, integrating GPS data into snow disaster risk analysis is emerging as a promising trend to enhance predictive accuracy and regional adaptability.

2.1.2 Snow Disaster Impact Studies

Snow disaster impact studies evaluate the consequences of snow disasters on society and the environment, particularly in agriculture, livestock, and road transportation. For example, Gao et al. proposed an agricultural risk assessment model that detailed the impact of snowy weather on agricultural regions in China [48]. Wang et al. developed an integrated risk index (IRI) to assess overall snow disaster risks by combining historical data, snowfall frequency, environmental conditions, and livestock burden [49]. Leone et al. conducted a detailed analysis of avalanche risks in the Alps, focusing on their impact on local transportation [50]. Liu et al. combined road network, meteorological, and socioeconomic data in Guoluo Prefecture, Qinghai, using AHP and clustering analysis to quantify road snow disaster risks [51]. Call et al. reported that a 5.1 cm increase in snowfall significantly raised traffic accidents, with passenger vehicles being more affected than commercial vehicles [52]. Other studies have also examined the wide-ranging impacts of snow disasters across various sectors. For example, Zhang et al. analyzed the economic effects of urban transportation system disruptions due to snow disasters, revealing significant productivity losses due to commute delays and restricted logistics [53]. Similarly, Chen et al. utilized remote sensing data to study the impact of snow cover on agricultural yields in Northeast China, showing that prolonged snowfall directly affects crop growth and harvest timelines [54]. Wang and Zhang assessed the vulnerability of power infrastructure in snow-prone areas, indicating that snow accumulation and ice accretion on power lines can lead to prolonged outages, impacting energy distribution and public services [55]. Finally, Miller et al. highlighted the health impacts of snow disasters, documenting increases in hospital admissions due to traffic accidents and hypothermia during extreme snowfall events in northern climates [56].

Although these studies have provided valuable insights into the multifaceted impacts of snow disasters, they predominantly analyze large spatial scales (e.g., provincial or city levels), making it difficult to capture localized effects. Therefore, integrating high-resolution data, such as GPS data, into localized snow disaster impact assessments remains an important future research direction [57].

2.1.3 Snow Disaster Detection and Monitoring

The detection and monitoring of snow disasters are typically based on changes in snow depth as a core indicator. Snow depth detection and monitoring primarily involve snow depth monitoring and prediction models based on remote sensing technologies. Tanniru and Ramsankaran reviewed the application of passive microwave remote sensing technology in global and regional snow depth monitoring, exploring the strengths and limitations of various models [58]. Deems et al. analyzed the use of LiDAR in snow depth measurement, examin-

ing the measurement errors and technical parameters of airborne and ground-based LiDAR systems [59]. Lievens et al. discussed modern data assimilation methods, emphasizing how passive and active microwave remote sensing technologies estimate snow cover in mountainous regions [60]. Additional studies have enhanced snow depth detection and monitoring methodologies by leveraging diverse remote sensing technologies. For instance, Smith et al. investigated the accuracy of satellite-based Synthetic Aperture Radar (SAR) for snow depth estimation, noting its efficacy in forested and mountainous terrains [61]. Additionally, Brown and McCabe demonstrated the integration of multispectral and hyperspectral imagery to monitor snow cover and depth across large geographic scales, with improved spatial and temporal resolution [62]. Liu et al. employed unmanned aerial vehicle (UAV) technology to monitor snow depth variations in real time, providing high-resolution data in hard-to-reach areas [63]. Moreover, Zhang et al. evaluated the application of MODIS (Moderate Resolution Imaging Spectroradiometer) for seasonal snow monitoring, showcasing its effectiveness in capturing snow depth changes across different land covers [64]. Finally, Dawes et al. utilized a machine learning-based model combined with satellite data to predict snow depth and snow water equivalent, yielding accurate predictions in complex terrains [65]. Yu et al. [66] proposed a large-scale forest snow depth estimation method based on optimized feature selection and a deep neural network (DNN). Revuelto et al. [67] highlighted that under favorable lighting conditions, UAVs can provide high-precision snow distribution estimates over relatively small areas, serving as an alternative to complex and costly near-field remote sensing techniques. Bianchi et al. [68] utilized Sentinel-1 SAR imagery and developed a deep learning-based avalanche detection method, significantly improving the capability of SAR imagery to identify avalanche events. Additionally, Dai et al. [69] proposed a novel inversion algorithm for estimating snow depth (SD) and snow water equivalent (SWE) using AMSR-E passive microwave remote sensing data, enhancing the accuracy of SD and SWE estimation.

In summary, traditional models, such as the Analytic Hierarchy Process (AHP) and risk assessment theories, are commonly used in snow disaster risk studies, often assuming linear relationships between variables. However, natural phenomena often exhibit nonlinear patterns, and oversimplification to linear assumptions can lead to biased estimations. Recently, machine learning methods have gained attention for their ability to capture complex relationships among variables [70, 71, 72]. However, the black-box nature of machine learning models limits their interpretability [73, 74, 75]. SHAP theory addresses this by providing quantitative evaluations of the impacts of nonlinear relationships [76, 77]. Moreover, traditional satellite-based snow detection faces technical limitations due to image resolution. In contrast, mobile GPS data can capture individual locations and movement trajectories, offering fine-grained analyses of human mobility and behavior patterns [78]. Combining mobile GPS data, machine learning, and SHAP theory to develop an interpretable spatial

machine learning model to explore the nonlinear associations between multidimensional environmental factors and snow disasters presents a promising research direction.

2.2 Urban Resilience Studies

In recent years, terms such as ”urban resilience,” ”climate change,” ”vulnerability analysis,” ”complex networks,” ”optimization decision-making,” and ”recovery” have become focal points of current research. For example, Meerow et al. conducted a systematic review of the definitions and applications of urban resilience, proposing a comprehensive resilience framework [13]. Chelleri et al. explored the relationship between urban resilience and sustainable development, highlighting resilience as a critical tool for addressing complex urban challenges [79]. Hofmann evaluated the Rockefeller Foundation’s ”100 Resilient Cities” (100RC) initiative launched in 2013 and analyzed the resilience policies formulated to support disaster risk reduction under this program [80]. Glaeser reviewed extensive existing literature on the long-term impacts of natural disasters and discussed urban resilience in the context of the post-COVID-19 era [81]. Wang et al. utilized entropy-weighted TOPSIS to assess urban resilience and digital economy levels across 252 Chinese cities from 2011 to 2020, empirically analyzing the impact and mechanisms of digital economy development on urban resilience [82].

Overall, research on urban resilience began around 2006, evolving through three distinct phases.

2.2.1 Phase 1: Methodology Development (Before 2014)

The first phase primarily focused on developing methodologies to evaluate transportation networks. For example, Bruneau et al. [83] proposed a multidimensional framework encompassing technical, organizational, social, and economic aspects, which significantly influenced foundational approaches during this period. Cutter et al. [84] introduced the Social Vulnerability Index, emphasizing the importance of social factors in resilience assessments. Zobel and Khansa [85] described multi-event disaster resilience, employing probabilistic methods to simulate transportation system performance under consecutive disasters. Similarly, Frazier et al. (2014) utilized geospatial analysis to assess post-flood recovery, providing insights into the spatial resilience of transportation networks. Sun et al. [86] reviewed metrics and methods for measuring transportation infrastructure resilience, establishing a foundational framework for subsequent research.

2.2.2 Phase 2: Resilience Characteristics Analysis (2014–2017)

From 2014 to 2017, the second phase shifted focus towards understanding resilience characteristics and investigating the resilience of various transportation systems. Wang and Taylor [87] highlighted mobility patterns during disasters, offering early insights into how urban systems respond to extreme conditions. Zhou et al. [88] applied network modeling to study transportation resilience under uncertain conditions, advancing analytical tools in this period. Twumasi-Boakye and Sobanjo [89] proposed a systematic framework to evaluate the robustness of critical infrastructure. Chacon-Hurtado et al. [90] integrated economic factors into resilience frameworks, highlighting the interaction between financial systems and transportation network resilience.

2.2.3 Phase 3: Disaster Recovery and Data-Driven Research (Since 2017)

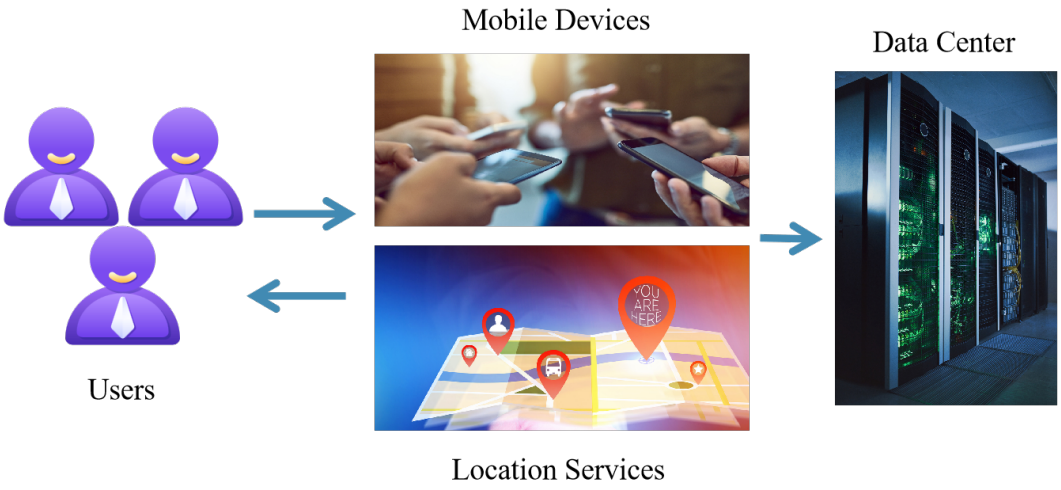
Since 2017, the third phase has emerged, driven by the increasing frequency of various disasters and extreme weather events [91]. This phase has seen a transition towards research on disaster mitigation, recovery, and system evolution. This shift has fostered the development of advanced resilience assessment methods to address the growing need for enhanced urban resilience [92]. Notably, since 2019, pandemic control and emergency management have become key concerns, with resilience assessment and construction of urban transportation networks emerging as critical research areas [93, 94]. These studies have further emphasized the importance of strengthening disaster resilience and exploring the relationship between transportation system resilience and urban resilience. An increasing number of studies have applied mobile data to disaster and resilience research, offering novel perspectives and methodologies for analyzing population mobility and its impacts during extreme weather events.

2.3 Applications of Mobile Phone GPS Data in Disaster and Transportation Studies

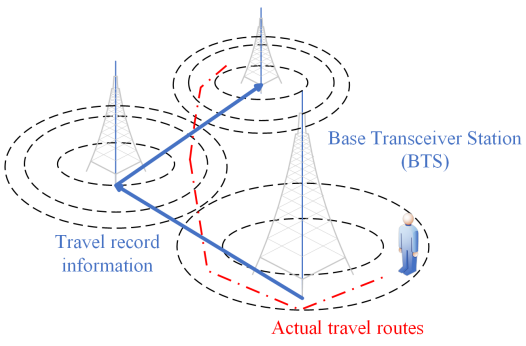
Mobile phone GPS data are primarily collected through Location-Based Services (LBS), which integrate Internet and Geographic Information Systems (GIS). Service providers offer applications such as location search, navigation, and real-time traffic monitoring based on mobile device positioning data. This process enables large-scale data collection, supporting mobility pattern analysis and transportation studies, as illustrated in Fig. 2.3.

After collection, the location data are transmitted to centralized data centers through

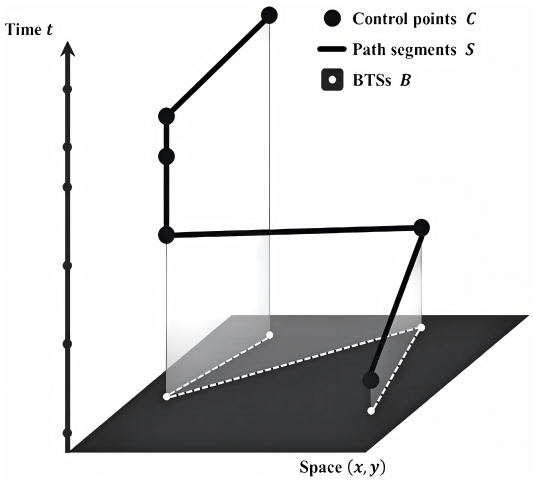
mobile networks, where they are processed and stored for further analysis. Additionally, data from Base Transceiver Stations (BTS) provide coarse-grained location information, which can complement GPS data to enhance spatiotemporal analysis accuracy. Using these data sources, spatiotemporal trajectory modeling is performed to reconstruct individual movement paths, facilitating disaster impact assessments and evacuation planning.



(a) Process of Mobile GPS Data Collection and Transmission



(b) Base Transceiver Station (BTS) Data Collection



(c) Spatiotemporal Trajectory

Figure 2.3: Process of Mobile GPS Data Collection, Transmission, and Spatiotemporal Modeling

In recent years, with the widespread availability of mobile phone GPS data, significant progress has been made in disaster management and transportation system optimization. Research efforts have primarily focused on disaster impact assessment, urban resilience analysis, human mobility and evacuation behavior modeling, and travel mode identification.

2.3.1 Applications in Disaster Management

In the field of disaster management, mobile location data has been extensively used to support disaster response and recovery efforts. Yabe et al. [95] systematically reviewed a decade of advancements in mobile location data applications for natural disasters and pandemics, emphasizing its critical role in enhancing disaster response capabilities. Hong et al. [96] analyzed resilience disparities among communities affected by natural disasters using large-scale mobile datasets, highlighting the importance of high-resolution data in disaster management. Hara and Yamaguchi [26] found that travel volumes in metropolitan areas of Japan significantly decreased during the COVID-19 state of emergency, demonstrating the substantial impact of policy interventions on human mobility.

In addition, research on disaster management tools continues to expand. Sukhwani et al. [97] investigated 33 disaster-related mobile applications in India and identified significant gaps in existing disaster management functionalities. Yeh et al. [98] proposed the Anomaly Detection of Population Distribution (ADPD) method based on mobile network data to detect anomalies during the 2018 Hualien earthquake in Taiwan. Santiago et al. [99] utilized big data from Google Popular Times to conduct spatiotemporal analyses, revealing that essential activities were less impacted than leisure activities during disasters. Saha et al. [100] designed a low-cost, GPS-enabled autonomous quadcopter, demonstrating its potential for public safety and emergency rescue operations.

To address navigation issues in disaster scenarios, Shahi et al. [101] developed a high-resolution map generation algorithm using drone image stitching techniques, suitable for situations where GPS signals are unavailable. Raei et al. [102] analyzed evacuation behavior and traffic dynamics during the 2021 wildfire in Lytton, Canada, using GPS data. Although data coverage was limited, the study provided valuable insights for evacuation planning and management in sparsely populated areas.

2.3.2 Applications in Transportation Systems

Extreme weather events often induce spatiotemporal variations in intercity transportation and micromobility patterns [103], leading to significant spatial heterogeneity in traffic behavior [104]. To mitigate transportation pressures caused by these variations, governments commonly implement measures such as road closures, snow clearance, and evacuation alerts. However, inappropriate evacuation policies may exacerbate traffic incidents and severe congestion. For instance, in January 2014, following a snowstorm warning in Atlanta, Georgia, the United States, a large number of residents simultaneously attempted to return home, resulting in massive vehicular gridlock [105]. Therefore, accurately understanding traffic demand and trends during snowstorms and assisting governments in issuing timely and

effective evacuation alerts are crucial for alleviating such issues.

In this context, Yang [106] proposed a framework for assessing the impacts of rainstorms and floods on transportation systems using GPS data. The framework employs changes in traffic flow and congestion indexes to identify disrupted and flood-damaged road sections, providing a scientific basis for improving transport system resilience. Wan et al. [107] applied a combination of Multi-Criteria Decision Making (MCDM) and GIS methods, utilizing long-term meteorological and traffic data to quantitatively assess the resilience of urban transportation in Changchun, China, under extreme climatic conditions. Their study revealed spatial disparities in transport resilience and proposed targeted enhancement strategies for vulnerable areas.

Regarding post-disaster human mobility patterns, Varol et al. [108] analyzed intra-city and inter-city movement behaviors following the 2020 Aegean Sea earthquake using Meta's Data for Good platform. They observed significant short-term changes in mobility behaviors, providing critical insights for disaster response and recovery planning.

Meanwhile, mobile GPS data has been extensively applied in travel mode identification and transportation optimization. Fan et al. [109] introduced the Multi-Scale Spatio-Temporal Attribute Fusion (MSAF) model for travel mode recognition based solely on GPS trajectories. Their model effectively captured spatio-temporal dependencies and achieved high classification accuracy, particularly addressing challenges related to sample imbalance in real-world datasets. Hosseini et al. [110] developed a machine learning framework using large-scale GPS trajectory data and random forest models to accurately detect various passenger trip phases, including access, egress, and waiting times at public transport stops, offering valuable support for urban transport planning and service improvements.

In summary, the widespread application of mobile phone GPS data has significantly advanced the fields of disaster management, urban resilience analysis, evacuation behavior modeling, and travel mode identification. These studies have not only enriched theoretical frameworks but also provided robust data support and technological tools for practical decision-making. With the continuous development of data integration and artificial intelligence technologies, GPS-based disaster and transportation research will play an increasingly vital role in enhancing emergency response capabilities and promoting sustainable urban development.

Chapter 3

Study Area, Data Sources, and Preprocessing

3.1 Research area

Fukui Prefecture is located in central Japan on Honshu Island, near the Sea of Japan. Fukui Prefecture, with a population of 762,679, accounts for approximately 6/1000 of Japan's total population and covers about 1/100 of Japan's land area. It includes both urbanized and rural areas. Its traffic demand and flow are shaped by a warm, humid climate and heavy snowfall in winter due to the Sea of Japan cold air mass, which significantly impacts traffic during snowstorms. From February 3rd to 9th, 2018, the Fukui snowstorm brought severe snowfall to Fukui Prefecture and surrounding areas. This event caused traffic disruptions, school closures, and affected residents. As a result of the snow disaster, around 1,500 vehicles became temporarily stranded on National Route 8 [1], as shown in Fig. 3.1. The historical monthly precipitation and snow depth, as well as daily snow accumulation during the 2018 snow disaster in Fukui City, are illustrated in Fig. 3.2.

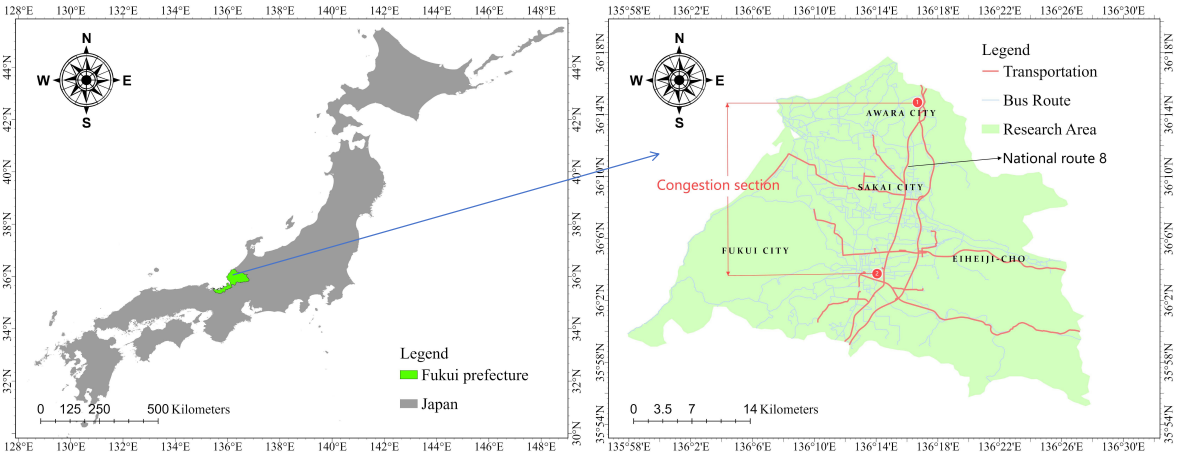
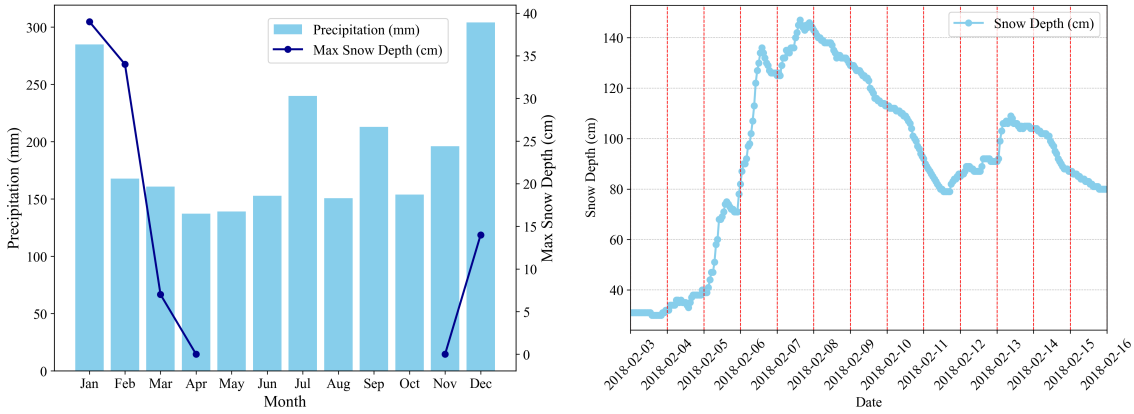


Figure 3.1: Research area and congestion section [1]



(a) Monthly Average Precipitation and Max Snow Depth (1991–2020)

(b) Daily Snow Depth During the February 2018 Snowstorm

Figure 3.2: Monthly and Daily Snow Depth Patterns in Fukui City

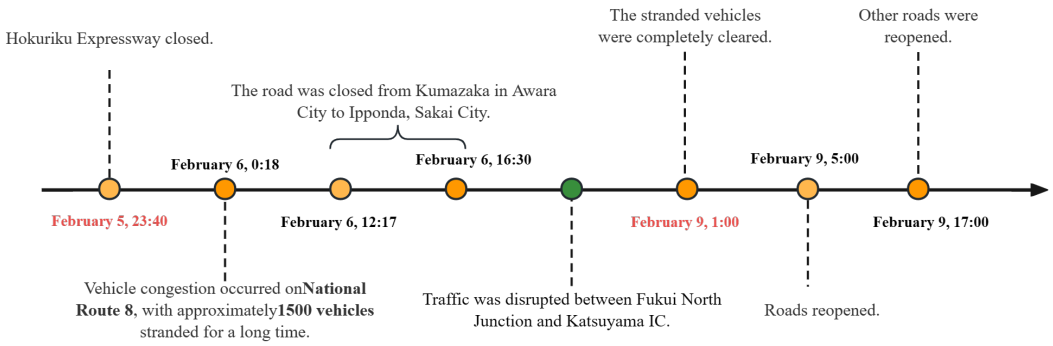


Figure 3.3: Process of a Snow Disaster Event [1, 2]

3.2 Research data

This study primarily utilizes mobile GPS data collected during the 2018 snowstorm in Fukui Prefecture, as well as road data, traffic congestion data, digital elevation data (DEM), and urban data, as detailed in Table 3.1.

3.2.1 Agoop GPS Data

This dataset was purchased from Agoop, a Japanese information service company. We obtained the mobile GPS data collected by Agoop during the snow disaster period in Fukui Prefecture in 2018, which is provided in the form of point data, as shown in Fig. 3.4. Agoop's mobile GPS data offer valuable information for studying population mobility, effectively capturing the spatiotemporal variations in population movement throughout the day. This

Table 3.1: Data Sources, Spatial, and Temporal Resolutions

Data	Spatial Resolution	Temporal Resolution	Source
Docomo GPS data	500 m	4 AM, 7 AM, 12 PM	Docomo Company
Agoop GPS data	Point-level	Minute-level	Agoop Company
Road density data	1 km	–	MLIT
Urban area data	–	–	MLIT
DEM data	500 m	–	GIS
Landuse data	500 m	–	MLIT
Night light data	500 m	Daily	LAADS
NDVI data	500 m	16 days	LAADS
Snow depth data	City level	Hourly	JMA
Traffic congestion data	City level	–	News and Reports [2, 4]

Abbreviations: MLIT: Ministry of Land, Infrastructure, Transport and Tourism; GSI: Geospatial Information Authority of Japan; LAADS: Level 1 and Atmosphere Archive & Distribution System; JMA: Japan Meteorological Agency; NDVI: Normalized Difference Vegetation Index.

comprehensive temporal coverage is critical for analyzing transportation demand and mobility patterns under snow disaster conditions. The dataset covers a period of 21 days, from January 27 to February 16, 2018. The data structure is presented in Table 3.2.

Table 3.2: Sample of GPS Records from Agoop GPS Data

Datetime	Daily ID	Longitude	Latitude	City Code
2018/1/27 0:00	d806d346bb	35.525728	134.117821	31201
2018/1/27 0:00	1503232ab1	36.317338	136.386391	17206
2018/1/27 0:00	e4e09937fac	36.085485	136.278997	18201
2018/1/27 0:00	06e4abad75	35.647763	139.746213	13103
2018/1/27 0:00	b53c758854	35.475786	135.794962	18204

Abbreviations: Daily ID: Identifier for the recorded individual; City Code: Current city code of the recorded individual.

Most of the sampling intervals in the Agoop GPS data are concentrated below 600 seconds, with a distinct peak around 300 seconds. This indicates that the majority of location points are recorded at intervals of approximately five minutes or less, ensuring sufficient temporal resolution to accurately capture short-term mobility behaviors. However, there is also a portion of the data with larger sampling intervals, which necessitates careful

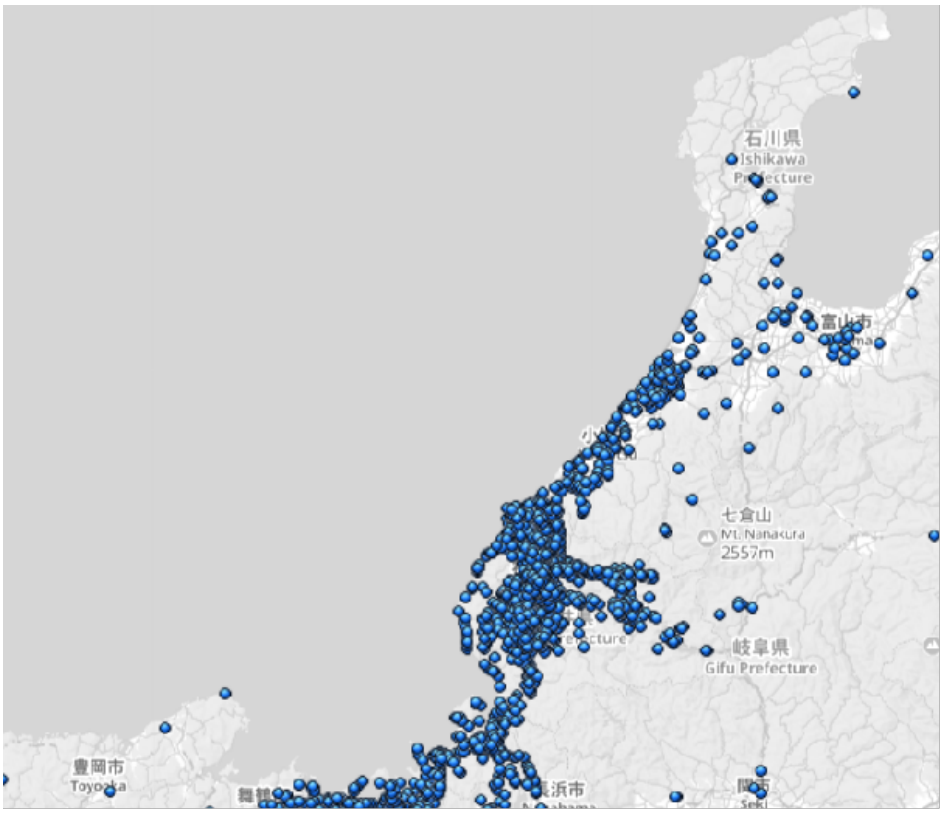


Figure 3.4: Spatial Distribution of Agoop GPS Data in Fukui Prefecture

preprocessing to maintain data quality and analysis accuracy.

3.2.2 Docomo GPS Data

The GPS data were provided by Docomo, a Japanese information service company, covering 21 days from January 27 to February 16, 2018, and recorded at three time points each day: 4:00 AM, 7:00 AM, and 12:00 PM. Docomo's GPS data are presented in a 500-meter grid cell format, as shown in Table 3.3. These data provide valuable insights into population distribution changes, which can be used to analyze population micromobility during snowstorms, as snowstorms may lead to localized disruptions and minor population redistributions.

Table 3.3: Sample of GPS Records from Docomo GPS Data

Date	Day_of_week	Time	Area	Residence	Age	Population
20180203	Saturday	1200	533630543	18201	50	90
20180204	Sunday	1200	543601634	18201	40	18
20180208	Thursday	1200	543621583	18201	50	65

It is important to note that due to its aggregated format and limited temporal resolution, the Docomo dataset cannot be used to analyze individual movement trajectories or to retrieve

precise origin-destination information. In contrast, the Agoop GPS data, purchased from Agoop Corporation, are provided as point-based data with a temporal resolution as high as one minute. This enables detailed behavioral analysis at the individual level and supports the reconstruction of precise movement trajectories. However, the spatial accuracy of the Agoop data may be affected by limitations in mobile base station coverage and app activity patterns. Particularly in sparsely populated or rural areas, data completeness may be inferior to that of the Docomo dataset, which is presented in Table 3.4.

Table 3.4: Comparison of Docomo and Agoop Mobile GPS Data

Aspect	**Docomo GPS Data**	**Agoop GPS Data**
Data Format	Aggregated Grid (500 m)	Point-based Data
Resolution	Hour (3 Time Points)	Up to Minute-level
Origin-Destination Info	Not Available	Available
Data Completeness	Higher in Rural Areas	Lower in Rural Areas
Suitable Analysis	Macro-level Population Distribution	Micro-level Behavioral and Trajectory Analysis

Note: Docomo data is more suitable for analyzing large-scale spatial distributions, while Agoop data enables fine-grained spatiotemporal mobility studies.

3.2.3 Road density data

Road infrastructure plays a critical role in assessing snow disaster risk, as road network configuration and distribution significantly influence traffic capacity and congestion during extreme weather. In this study, road density data were obtained from the Ministry of Land, Infrastructure, Transport, and Tourism (MLIT) and aggregated into a 1 km grid cell format, with units expressed as meters per square kilometer (m/km^2). Each grid cell represents the total length of roads, categorized by width: 5.5–13.0 m (Road type 1), 2.5–5.5 m (Road type 2), and under 2.5 m (Road type 3).To match other 500-meter grid cell data, the original 1-kilometer grid cell road density data was further processed into 500-meter grid cells. This classification and transformation enable a detailed analysis of the specific impacts of different road widths under snow disaster conditions. This categorization and transformation enable a detailed analysis of the specific impacts of different road widths under snow disaster conditions.

3.2.4 Urban area data

Urban areas are typically characterized by high population density and well-developed transportation infrastructure. Thus, incorporating urban area data can improve the accuracy of snow disaster risk assessments. These urban areas, defined as regions requiring comprehensive development, improvement, and conservation as unified cities, are represented in data published by the Ministry of Land, Infrastructure, Transport and Tourism (MLIT) in polygon format. We used Geographic Information System (GIS) software to convert these original polygons into a 500-meter grid cell format for analysis.

3.2.5 Traffic Congestion Data

Traffic congestion data enable us to identify specific road segments and cities affected by snowstorm-induced congestion, providing a basis for identifying potential risk areas during snowstorm events. The traffic congestion data were obtained from publicly available online news reports [1]. This dataset provides textual information indicating approximate congestion locations during snowstorm events, accompanied by a map highlighting these congested sections. The data specifically focuses on National Route 8 and the Hokuriku Expressway in the northern cities of Fukui Prefecture, including Fukui City, Sakai City, and Awara City, as shown in Fig. 3.1.

3.2.6 DEM Data

In mountainous or steep areas, snow tends to accumulate more easily and melts at a slower rate, significantly impacting road accessibility and congestion levels. To capture these effects, we utilized DEM data provided in raster format by the Geospatial Information Authority of Japan (GSI). This dataset includes variables such as average elevation and slope angle within 500-meter grid cells, which help identify high-risk areas during snowstorm events.

3.2.7 Night light Data

During snow disasters, transportation activity often decreases in affected areas, leading to reduced vehicle movements and diminished night-time lighting, especially from vehicle headlights. Therefore, analyzing variations in night light intensity can serve as an effective indicator for identifying snow-affected regions. This approach provides valuable support for spatially locating impacted areas and understanding the extent of disaster-induced disruptions. The night light data was obtained from the LAADS (Level 1 and Atmosphere Archive & Distribution System) platform. The night light data had already been preprocessed for cloud

removal.

3.2.8 NDVI Data

The NDVI (Normalized Difference Vegetation Index) is a widely used vegetation index that quantifies vegetation activity by comparing the difference between near-infrared (NIR) and red (RED) reflectance values [111]. It is calculated as follows:

$$NDVI = \frac{NIR - RED}{NIR + RED} \tag{3.1}$$

where *NIR* and *RED* represent the near-infrared and red light reflectance, respectively. High NDVI values typically indicate dense, healthy vegetation, while lower values correspond to sparse or stressed vegetation cover.

In the context of snow disasters, heavy snowfall can directly cover vegetation surfaces, significantly reducing their reflectance in the near-infrared band and leading to a marked decline in NDVI values. During snowstorm events, snow coverage can obscure vegetation surfaces, leading to a reduction in NDVI values. This phenomenon provides an indirect means to identify snow-affected areas through remote sensing observations.

The NDVI dataset used in this study had already been preprocessed for cloud removal, ensuring higher accuracy in detecting surface changes. With a spatial resolution of 500 meters and a temporal resolution of 16 days, this dataset supports the analysis of vegetation dynamics and snow disaster impacts over both spatial and temporal dimensions.

3.2.9 Landuse Data

The land use data were obtained from the Ministry of Land, Infrastructure, Transport and Tourism (MLIT) of Japan. These data are provided in a polygon format and have been preprocessed to ensure high accuracy and consistency, eliminating the need for additional cleaning procedures.

Land use data play a crucial role in understanding the spatial distribution of urban functional zones, including residential, commercial, industrial, agricultural, and green spaces. Accurate land use classification supports the analysis of how different regions respond to snow disasters, particularly in evaluating the vulnerability and resilience of various land use types. For instance, commercial and industrial zones may exhibit higher sensitivity to transportation disruptions, while residential areas are critical for assessing evacuation and shelter demands.

In this study, land use data were integrated with other spatial datasets to analyze the impacts of snow disasters on urban areas at a fine spatial resolution, providing important context for resilience assessments and policy recommendations.

3.2.10 Snow Depth Data

Snow depth is a critical parameter for assessing snow disaster severity, as it directly influences transportation disruptions, infrastructure load, and regional accessibility. Snow depth data were collected from the Japan Meteorological Agency (JMA). Fukui Prefecture consists of 17 cities, but only 6 of them have city-level snow depth monitoring stations, as shown in Fig. 3.5. The temporal resolution of the data is hourly.

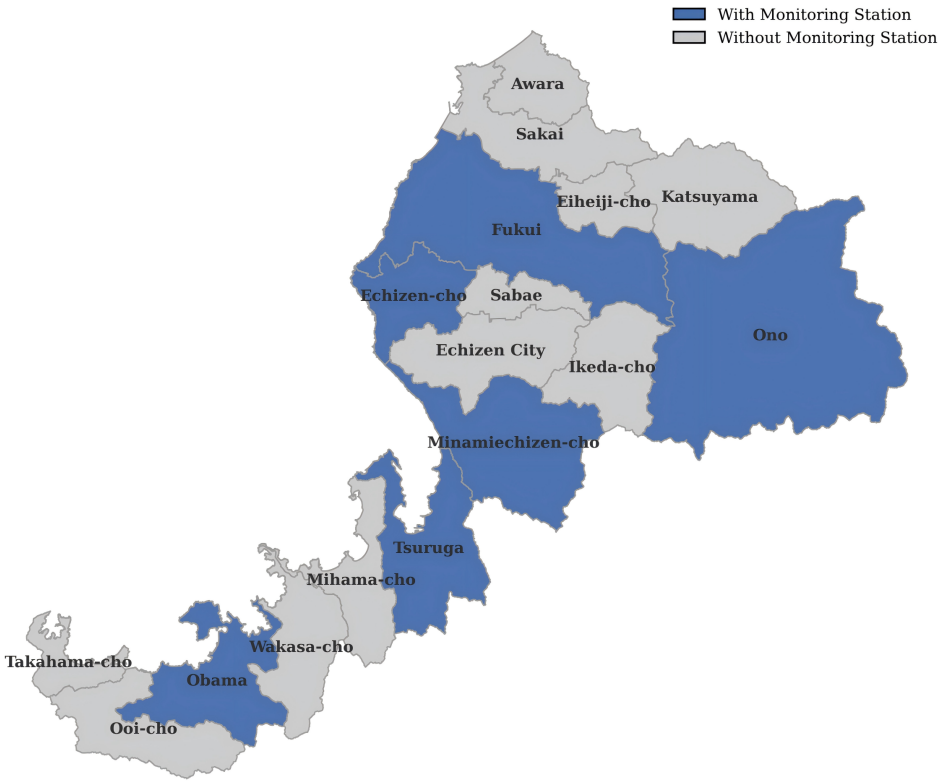


Figure 3.5: Snow Depth Monitoring Coverage in Fukui Prefecture

3.3 Data preprocessing

3.3.1 Agoop GPS data

Mobile data contains a significant amount of noisy data. In addition to standard data cleaning methods, we focused on cleaning redundant data, ping-pong data, and drift data.

- **Redundant Data:** This occurs when the same user has multiple records at the same timestamp in the mobile data. This may be caused by a short sampling interval or insufficient precision in the timestamp field. For consecutive records ($n \geq 3$) with the same location, only the first and the n-th records are retained. The intermediate records are considered redundant and are discarded.

- **Ping-Pong Data:** These are location data where the mobile device frequently switches between nearby base stations within a short period, leading to oscillating position records, as illustrated in Fig. 3.6. This often occurs due to unstable signals, dense base stations, or environments such as high-speed trains or underground metros.

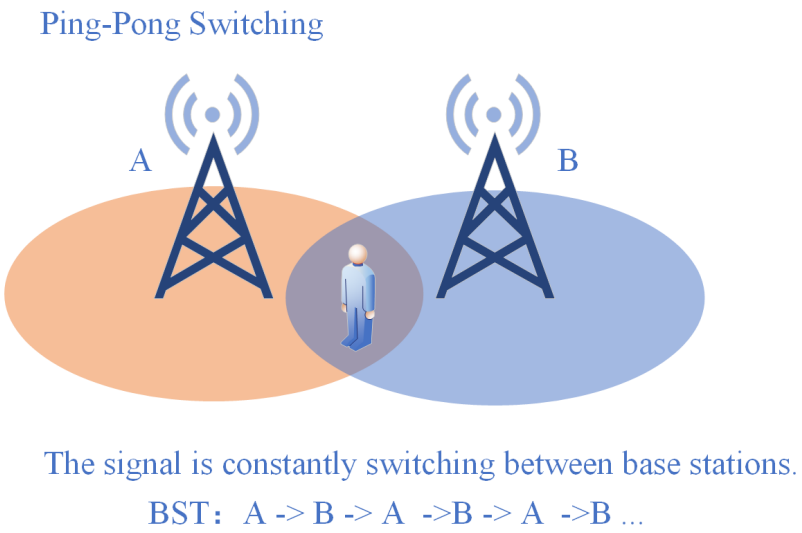


Figure 3.6: Illustration of Ping-Pong data

- **Drift Data:** Due to environmental interference, insufficient satellite coverage, or signal obstructions, discrepancies between the collected data and the actual situation may occur, leading to large deviations in recorded locations. These appear as sudden jumps in the trajectory. We first defined the geographic boundaries of the study area as (135.449339, 35.343678, 136.832362, 36.451457) and removed any data outside this region. Then, within the study area, we applied the following filters:

  - **Speed Threshold:** If the speed between the current trajectory point and the previous/next points is excessively high, the data is flagged as drift.
  - **Distance Threshold:** If the distance between the current trajectory point and the previous/next points is too large, the data is flagged as drift.
  - **Angle Threshold:** If the angle formed by the previous, current, and next trajectory points is too small, the data is flagged as drift.

  After completing all data cleaning procedures, the ArcGIS Pro ”Fishnet” tool can be used to generate grid cells for spatial analysis, as illustrated in Fig. 3.7. This grid-based method standardizes the spatial units for subsequent analysis and is equally applicable to both Agoop and Docomo datasets.

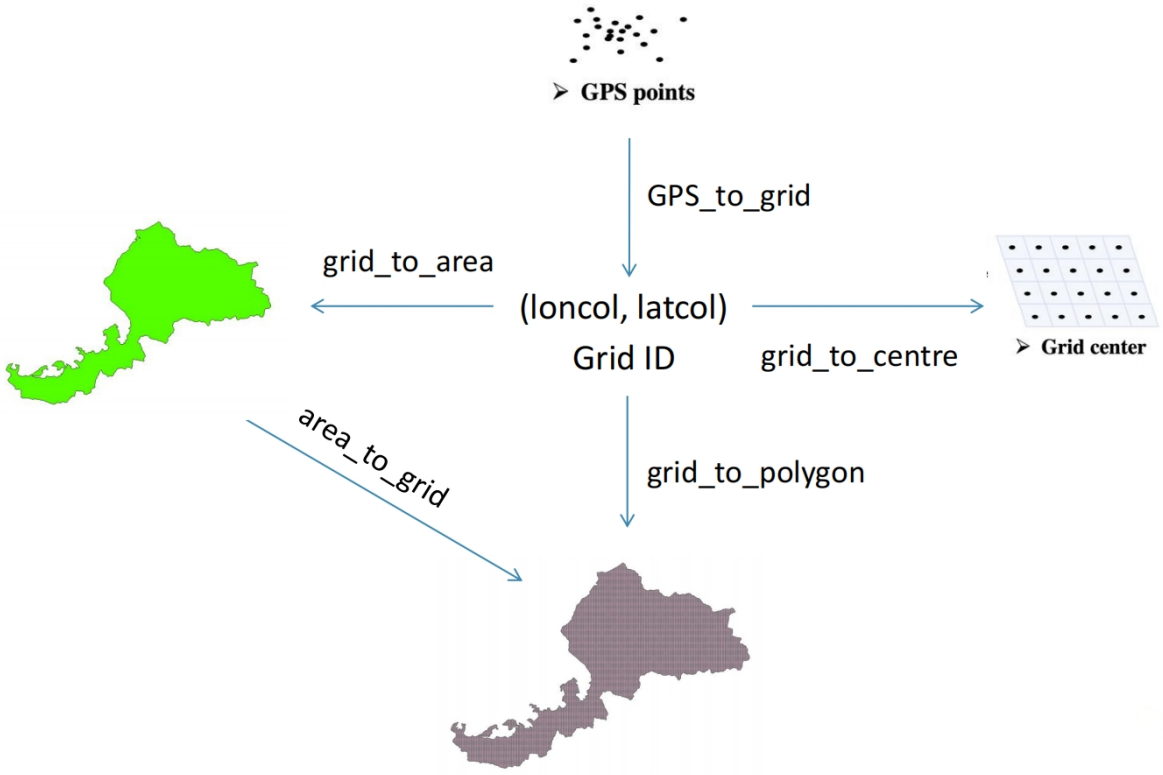


Figure 3.7: Processing Workflow of Agoop GPS Data Using ArcGIS Pro

3.3.2 Docomo mobile data

This dataset, purchased from Docomo, provides preprocessed population distribution data. We selected grid cells with at least two population records per day during both the pre-snowstorm and snowstorm periods. For missing time points, the data were interpolated using the average population between the same time points on the previous and following days.

The daily population fluctuation was calculated as the difference between the maximum and minimum population values each day. The average fluctuation from January 27 to February 2, 2018, was defined as the "normal value," and from February 3 to February 9, 2018, as the "abnormal value." The impact of the snowstorm on population fluctuation for each grid cell was then quantified as the difference between these two values, as detailed in the following equations:

$$P_{\text{fluctuation}} = P_{\text{daily\_max}} - P_{\text{daily\_min}} \tag{3.2}$$

$$D_{\text{fluctuation}} = D_{\text{daily\_max}} - D_{\text{daily\_min}} \tag{3.3}$$

$$\text{Difference} = D_{\text{fluctuation}} - P_{\text{fluctuation}} \tag{3.4}$$

- $P_{\text{daily\_max}}$: Maximum population before the snow disaster (January 27 to February 2, 2018).
- $P_{\text{daily\_min}}$: Minimum population before the snow disaster (January 27 to February 2, 2018).
- $P_{\text{fluctuation}}$: Population fluctuation before the snow disaster.
- $D_{\text{daily\_max}}$: Maximum population during the snow disaster (February 3 to February 9, 2018).
- $D_{\text{daily\_min}}$: Minimum population during the snow disaster (February 3 to February 9, 2018).
- $D_{\text{fluctuation}}$: Population fluctuation during the snow disaster.
- Difference: Change in population fluctuation attributed to the snow disaster.

3.3.3 Urban area data

The original data was in polygon format. We used ArcGIS Pro to spatially link and extract urban areas into grid cells, marking urban grid cells with a label of 1.

3.3.4 Traffic congestion data

This data was collected from online sources. Traffic congestion data is crucial for identifying areas affected by snow disasters and serves as the target variable in this study. In this context, snow disasters are defined as events caused by excessive snowfall, prolonged snow accumulation, or persistent snow coverage, leading to significant disruptions such as traffic congestion. Using a supervised learning model, we manually labeled grid cells experiencing traffic congestion during snowstorms as 1 (indicating a snow disaster) and those without congestion as 0 (indicating no snow disaster).

3.3.5 DEM data

The data can be directly downloaded from the GSI and is already cleaned and preprocessed.

3.3.6 Road density data

The original road dataset, initially available at a 1000-meter resolution, displayed significant missing values across most road types. To ensure data quality, only road

types with a missing data rate below 25% were included: Road type 1 (5.5 m ≤ width < 13.0 m), Road type 2 (3.0 m ≤ width < 5.5 m), Road type 3 (width < 3.0 m). To harmonize this with the 500-meter resolution of other datasets, we applied the following processing steps: First, grid cells with no recorded road data were removed. Next, the 1000-meter road data was spatially joined with the 500-meter grid cell of Fukui, associating each 1000-meter grid cell with its corresponding four 500-meter grid cells. Finally, assuming a uniform distribution, the road density from each 1000-meter grid cell was divided by four and allocated evenly among the corresponding 500-meter grid cells. This step ensured consistency with other variables and allowed for more meaningful spatial comparisons.

3.3.7 Night Light Data

To standardize the spatial resolution of the night light data, we used QGIS to generate a 500-meter grid mesh using the fishnet tool. The night light satellite imagery was then rasterized into 500-meter grid cells, and the average night light intensity was calculated for each grid. The processed night light dataset was subsequently spatially aligned with the GPS-based population grid to ensure consistent spatial referencing. This preprocessing enabled a unified spatial scale for further comparative analyses.

3.3.8 NDVI Data

The NDVI data underwent a similar preprocessing procedure. A 500-meter grid mesh was created using the fishnet tool in QGIS, and the NDVI satellite images were rasterized accordingly. The average NDVI value was computed for each grid cell. Finally, the processed NDVI dataset was spatially aligned with the GPS-based population grid to ensure consistency across datasets. This standardization facilitates integrated analysis and direct comparison between variables.

3.3.9 Land use data

First, grids without land use data records were removed. Next, ArcGIS Pro was used to spatially join the land use data with the processed GPS population data, resulting in population change rates for different land use types. This step ensures spatial consistency between the land use data and GPS population data, enhancing the scientific validity and significance of spatial comparisons.

3.3.10 Snowdepth data

Only six cities in Fukui Prefecture have official urban-level snow depth monitoring stations. For cities without monitoring stations, their snow depth values are assigned based on the closest monitored city in terms of geographical distance. To determine the nearest monitored city, we compute the Euclidean distance between the centroid of each unmonitored city and all monitored cities. The snow depth value of the nearest city is then assigned to the unmonitored city. Mathematically, this can be expressed as:

$$S(x) = S(x^*) \tag{3.5}$$

where:

- $S(x)$ is the estimated snow depth for an unmonitored city x.
- x^* is the centroid of the nearest monitored city, defined as:

$$x^* = \arg\min_{x_i \in M} d(x, x_i) \tag{3.6}$$

where:

- M is the set of cities with snow depth monitoring stations.
- $d(x, x_i)$ represents the Euclidean distance between the centroids of city x (unmonitored) and city x_i (monitored).

Based on the above method, the recorded snow depth data from the six cities are reasonably assigned to nearby cities without monitoring stations, as shown in Table 3.5.

Table 3.5: Snow Depth Monitoring Stations and Their Coverage in Fukui Prefecture

Monitoring Station	**Covered Cities and Towns**
Fukui Meteorological Station	Fukui City, Awara City, Sakai City, Eiheiji-cho
Echizen Meteorological Station	Echizen City, Sabae City, Echizen-cho
Minamiechizen-cho Meteorological Station	Minamiechizen-cho
Tsuruga Meteorological Station	Tsuruga City, Mihama-cho
Obama Meteorological Station	Obama City, Takahama-cho, Ooi-cho, Wakasa-cho
Ōno Meteorological Station	Ōno City, Katsuyama City, Ikeda-cho

Chapter 4

Intercity transportation demand in Fukui prefecture under 2018 heavy snowfall

Understanding intercity transportation demand and trends during snowstorms is crucial for mitigating traffic accidents and congestion. Heavy snow events can significantly disrupt regular travel patterns, leading to sudden shifts in demand and increased vulnerability of transportation networks. This chapter focuses on analyzing the characteristics of intercity transportation demand during the 2018 Fukui Heavy Snow Event using Agoop mobile GPS data. To capture the spatiotemporal variations of transportation demand, the K-dimensional Tree (KDTree) algorithm was employed for nearest neighbor matching, providing insights into the overall demand trends under severe snowstorm conditions. Additionally, Singular Value Decomposition (SVD) was utilized to decompose and reduce the dimensions of the spatiotemporal OD matrix, facilitating the identification of dominant intercity transportation flow patterns and their structural changes during the event.

4.1 Research Methodology

4.1.1 Research Framework

This study employs a combination of KDTree and SVD to analyze the spatiotemporal characteristics of intercity transportation demand during the 2018 heavy snowfall in Fukui Prefecture. The KDTree algorithm is utilized to efficiently identify road segments and areas that were severely affected by the snow disaster based on trajectory proximity analysis. Meanwhile, SVD is applied to decompose the spatiotemporal OD matrix, enabling the extraction of dominant transportation flow patterns and reducing data dimensionality for better interpretation of demand variations. The overall research framework is illustrated in Fig. 4.1.

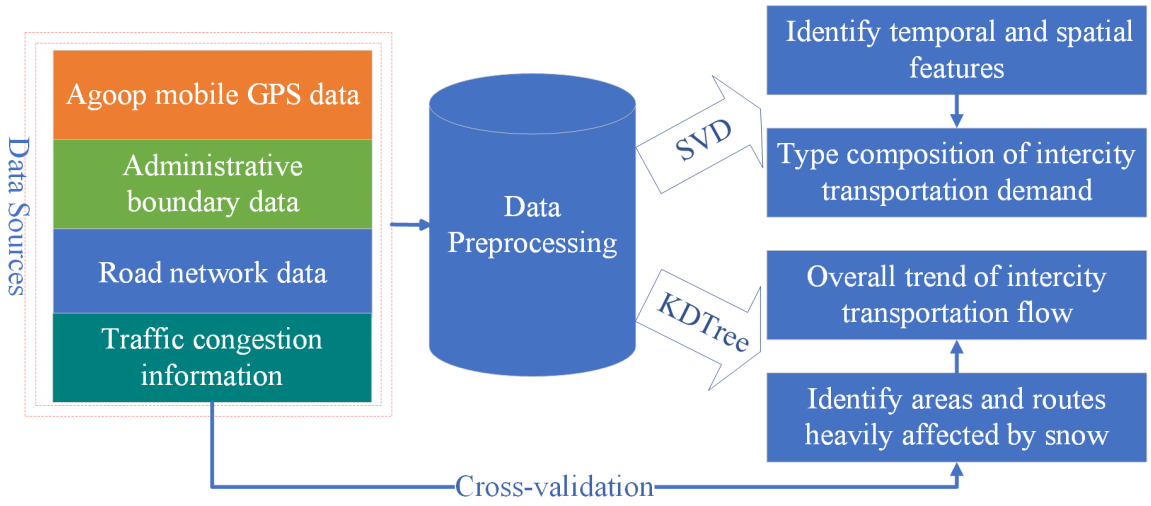


Figure 4.1: Research workflow

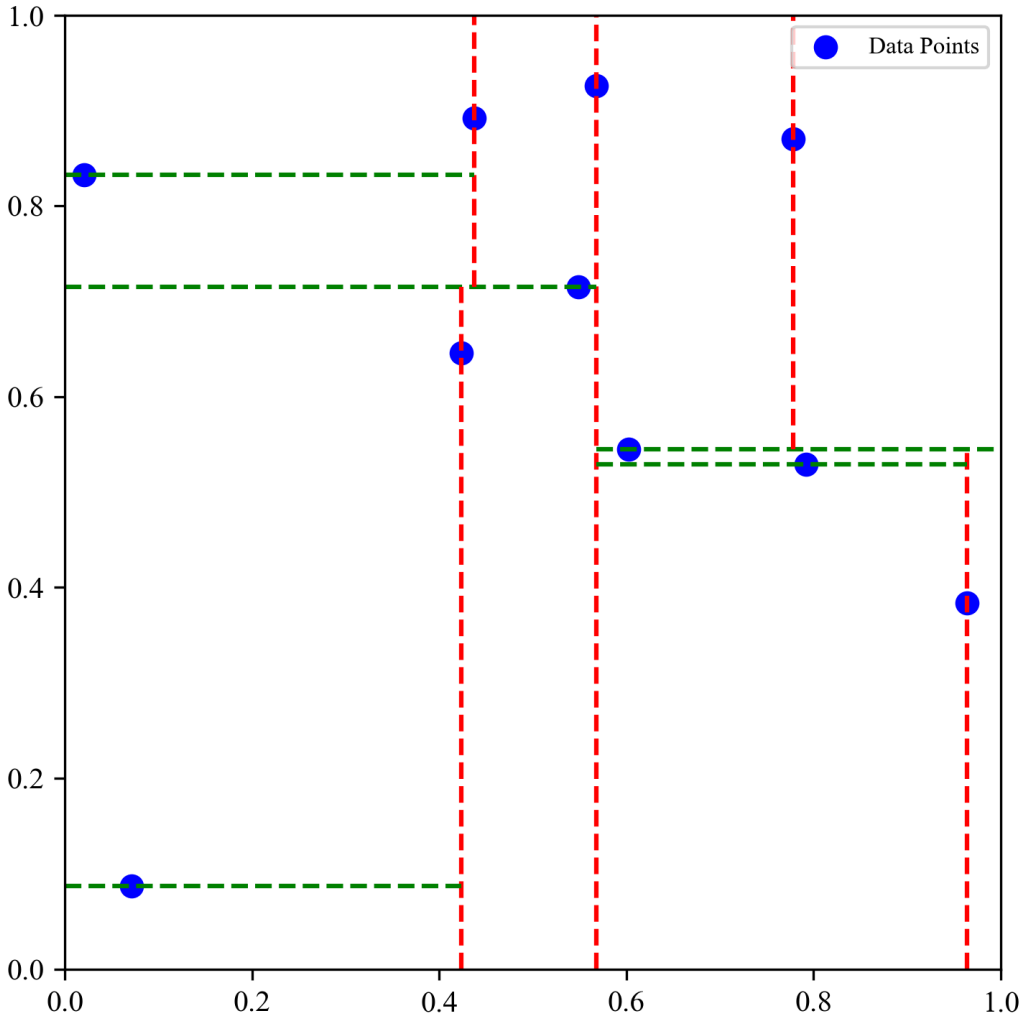


Figure 4.2: Illustration of KD-Tree Partitioning

4.1.2 K-dimensional Tree (KDTree) for Trajectory Matching

KDTree (K-Dimensional Tree) is a data structure designed for efficiently locating the nearest neighbors in high-dimensional spaces. It recursively partitions the space into multiple regions by constructing a binary tree to organize the data points. The partitioning process of KD-Tree is illustrated in Fig. 4.2.

Assume there are N data points in a k-dimensional space, denoted as $D = 0, 1, 2, \ldots, k$. At each node, the tree selects a dimension D for splitting. In the first layer (root node), the data is sorted along the first dimension, and the median point becomes the root. Points smaller than the median are assigned to the left subtree, and larger points to the right subtree. In the next layer, the second dimension is used for splitting, and the process repeats recursively.

This recursive construction continues until a leaf node is reached or a stopping condition is satisfied. At the i-th layer, the splitting dimension is determined by $d = i \bmod k$, and the splitting boundary is the median value m_d along this dimension. The corresponding formula is as follows [112]:

$$\text{Left subtree} = x \mid x_d < m_d, \quad \text{Right subtree} = x \mid x_d \geq m_d \tag{4.1}$$

In practice, the Agoop data is converted into a GeoDataFrame of point data, and the OSM road network data is transformed into a GeoDataFrame of line data. By utilizing KDTree, a spatial index is constructed based on the OSM road nodes, allowing the point dataset to be efficiently matched to the nearest road segments in the line dataset. The output is a GeoDataFrame containing all original columns from the point dataset, combined with the attributes of the matched road segments. Additionally, a new column is introduced to record the distance from each point to its nearest road segment. Each time an Agoop GPS point is processed, KDTree rapidly identifies the nearest OSM road node, effectively ”matching” the GPS point to the closest road segment.

4.1.3 Singular Value Decomposition (SVD)

Singular Value Decomposition (SVD) is a fundamental matrix factorization technique widely used for feature extraction and dimensionality reduction. It has been successfully applied in various fields such as recommendation systems, natural language processing, and computer vision.

Let M be an $m \times n$ matrix. The SVD of M is defined as [113]:

$$M = USV^T = \sum_{h=1}^{r} \delta_h u_h v_h^T \tag{4.2}$$

where:

- U is an $m \times m$ orthogonal matrix whose columns are the left singular vectors;
- S is an $m \times n$ diagonal matrix with non-negative singular values δ_h on the diagonal;
- V is an $n \times n$ orthogonal matrix whose columns are the right singular vectors;
- r is the rank of matrix M.

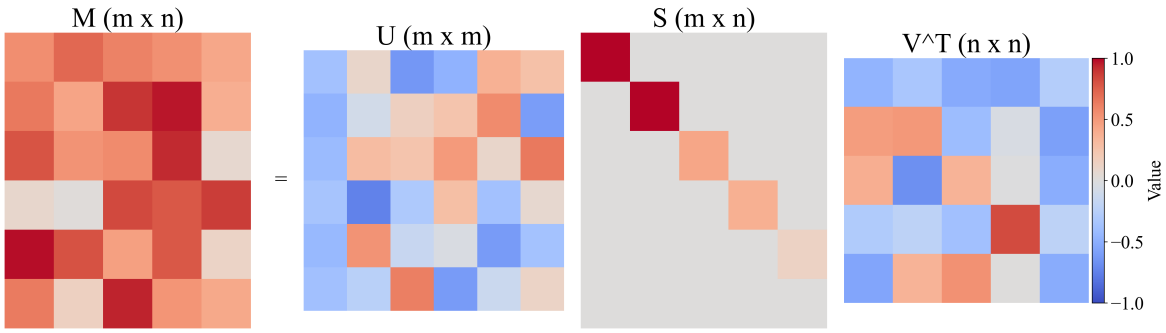


Figure 4.3: Visual representation of Singular Value Decomposition (SVD).

In this study, the origin-destination (OD) matrix derived from GPS trajectory data is decomposed using SVD to identify the dominant movement patterns during the snow disaster period. By analyzing the singular values and corresponding singular vectors, the primary spatiotemporal characteristics of intercity transportation flows are extracted, enabling a deeper understanding of how the snow disaster affected travel demand structures. The SVD-based analytical framework applied in this research is illustrated in Fig. 4.3.

4.1.4 Application of SVD in Spatio-Temporal OD Matrix for Intercity Transportation

Singular Value Decomposition (SVD) can be effectively applied to decompose spatiotemporal matrices, breaking them down into simpler, additive spatiotemporal variation patterns. In the context of intercity transportation, performing SVD on the origin-destination (OD) matrix enables the extraction of dominant travel patterns and the reduction of data complexity. Specifically, the OD matrix $\mathbf{M}$ can be represented as a collection of r distinct intercity transportation flow patterns.

In this decomposition:

- u_h represents the *temporal distribution* of the h-th type of intercity transportation flow,
- ${v_h}^T$ represents the *spatial distribution* of the h-th transportation flow pattern,
- δ_h indicates the *importance* (i.e., contribution) of the h-th type of flow pattern, quantified by the corresponding singular value.

To reduce dimensionality, the high-dimensional spatiotemporal OD matrix is projected into a lower-dimensional space by retaining only the top k components with the largest singular values (where $1 \leq k \leq r$). The remaining components, associated with smaller singular values, are discarded as they contribute less to the variance in the data. This process results in a reduced spatiotemporal OD matrix $\hat{\mathbf{M}}$, which preserves the most significant intercity transportation patterns while minimizing computational complexity.

4.2 Results and Discussions

4.2.1 Road Network Matching and Traffic Flow Visualization

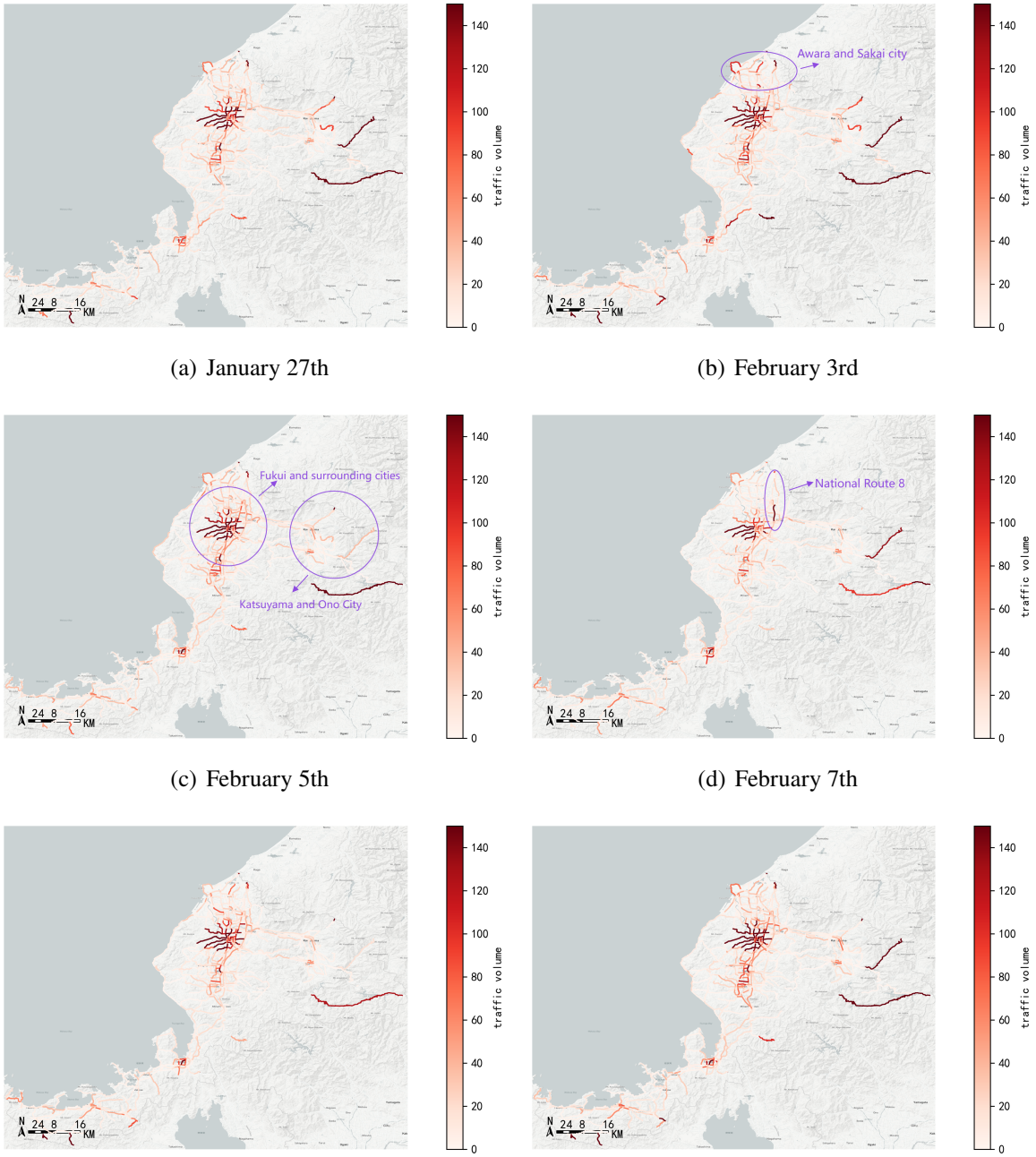
To observe the traffic flow trends during snowstorms, it is necessary to visualize the traffic distribution across various routes on the entire road network. This process aids in subsequent intercity travel statistics and streamlines the application of SVD. We downloaded the major and minor road data of Fukui Prefecture from OpenStreetMap (OSM) and combined it with the cleaned move point data from Agoop mobile data. Using the KDTree algorithm, we performed nearest neighbor matching. The matched results are shown in Figure 4.4.

The trajectory visualization analysis reveals that areas with high traffic volume are mainly concentrated in the northern region of Fukui Prefecture, radiating from Fukui City to surrounding cities such as Awara City, Sakai City, and Sabae City. Starting from February 3rd, the traffic pressure in Awara City and Sakai City significantly increased. From February 5th to February 7th, the traffic volume in Awara City, Sakai City, Fukui City, Katsuyama City, and Ono City noticeably decreased. By February 7th, the traffic pressure on certain sections of National Route 8 increased, which is consistent with the findings reported in the online network information [2]. Meanwhile, the traffic volume in Katsuyama City and Ono City almost disappeared. After February 13th, traffic patterns gradually began to return to their stable pre-snow disaster state.

4.2.2 Spatio-Temporal Characteristics of Intercity Transportation

For the period from January 27 to February 16, 2018, the spatio-temporal origin-destination (OD) matrix for intercity transportation in Fukui Prefecture was analyzed and dimensionally reduced using SVD. Suppose the spatio-temporal OD matrix of intercity transportation on a specified date s is denoted as $\mathbf{D}_s$, which contains n OD pairs. $\mathbf{D}_s$ is then transformed into a $1 \times n$ row vector $\mathbf{d}_s$, where n represents the number of OD pairs. The intercity transportation flow over m days can be represented as an $m \times n$ matrix $\mathbf{M}$.

To construct the OD matrix, it is first necessary to identify the movement and stay data from the Agoop dataset. This is typically accomplished using a time threshold method.



(a) January 27th

(b) February 3rd

(c) February 5th

(d) February 7th

(e) February 13th

(f) February 16th

Figure 4.4: Temporal variation of intercity traffic volume during the snow disaster period in Fukui Prefecture

Specifically, after dividing Fukui Prefecture into 500-meter grid cells, if a user remains in a particular grid for more than 30 minutes, it is considered a stay. The time period between two consecutive stays is identified as a trip.

After identifying all trips, the data is aggregated by intercity travel, where each unique OD pair forms the rows and the corresponding dates form the columns, thereby constructing the final OD matrix. The resulting OD matrix consists of 215 rows and 21 columns.

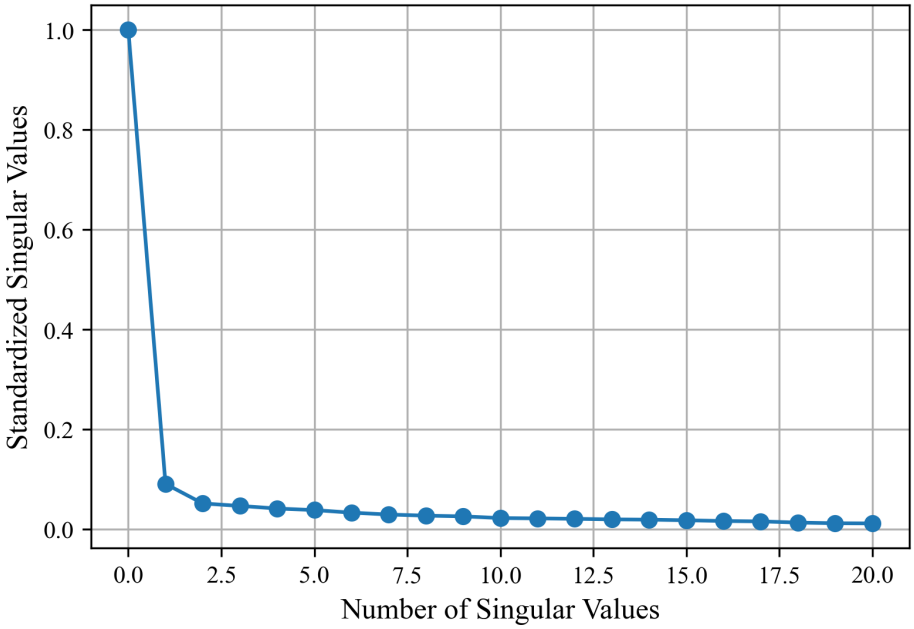


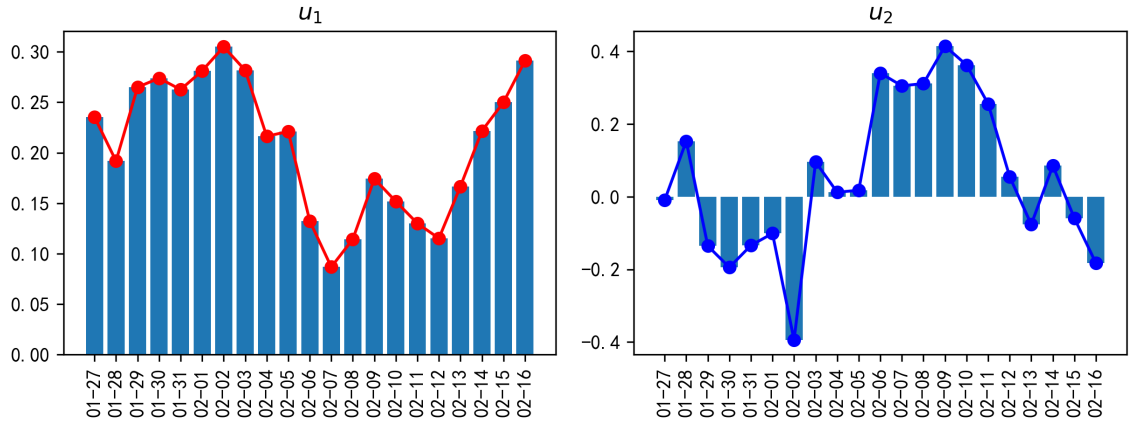
Figure 4.5: Standardized Singular Value Distribution

Singular values indicate the importance of each demand pattern within the spatio-temporal OD matrix. A larger singular value suggests that more of the original information in the matrix is retained. As illustrated in Fig. 4.5, the singular value distribution shows that the first two singular values are significantly larger than the others, with minimal variation observed after the third singular value.

By selecting the top k singular values, a lower-dimensional approximation of the original matrix can be obtained, effectively capturing the dominant intercity transportation patterns during the snow disaster period. By selecting the first two singular values, we can represent 88% of the original information contained in the spatiotemporal OD matrix through the corresponding demand patterns. These two singular values can be associated with two typical demand patterns and passenger flow types, denoted as M_1 and M_2. Additionally, M_1 and M_2 can be further decomposed into u_1, v_1 and u_2, v_2, where u_i represents temporal variations and v_i represents spatial variations.

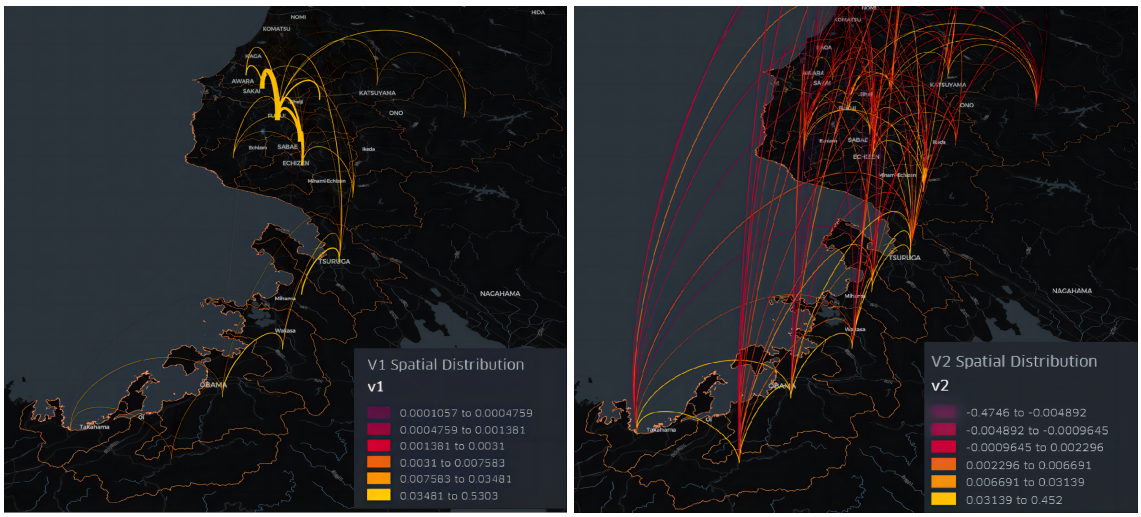
In Fig. 4.6(a) and Fig. 4.6(b), u_1 and u_2 represent the temporal unit vectors of transportation demand across different categories. The sign of the temporal unit vectors correspond to the direction of fluctuations in intercity transportation demand over time, while their absolute

values indicate the magnitude of these fluctuations. Similarly, in Fig. 4.6(c) and Fig. 4.6(d), v_1 and v_2 represent the spatial unit vectors of transportation demand for different categories, where positive values indicate an increase in traffic flow and negative values indicate a decrease.



(a) Temporal Distribution under M_1 Model

(b) Temporal Distribution under M_2 Model



(c) Spatiotemporal Distribution under M_1 Model

(d) Spatiotemporal Distribution under M_2 Model

Figure 4.6: Temporal and Spatial Distributions under the M_1 and M_2 Models.

Transportation Flow Type I (M_1) represents the daily intercity transportation demand within Fukui Prefecture and can be divided into three distinct stages: the stable stage from January 27 to February 2, the snowstorm stage from February 3 to February 11, and the recovery stage from February 12 to February 16. Starting on February 3, Transportation Flow Type I, represented by u_1, showed a marked decline, with its value dropping from 0.28 to a low of 0.09. This indicates that, due to the combined effects of the snowstorm and traffic control measures in some areas, the daily intercity transportation demand decreased by 67.86% within a short period. A brief rebound occurred from February 7 to February

9, coinciding with the implementation of traffic control and government policies, thereby reflecting the effectiveness of these measures. Spatially, Passenger Flow Type I primarily originates from four major urban areas: Fukui City, Sabae City, Awara City, and Sakai City.

Compared to u_1, u_2 exhibits greater volatility, suggesting it captures a different temporal variation pattern from the main trend, specifically representing the dynamic transportation demand during the snowstorm period, i.e., Transportation Flow Type II (M_2). Beginning on February 1, the value of u_2 initially dropped sharply from -0.10 to -0.39, before rapidly rising to a peak of 0.41. This indicates that the snowstorm caused significant dynamic changes, such as substantial disruptions in transportation during the initial stage followed by recovery or emergency transportation flow.

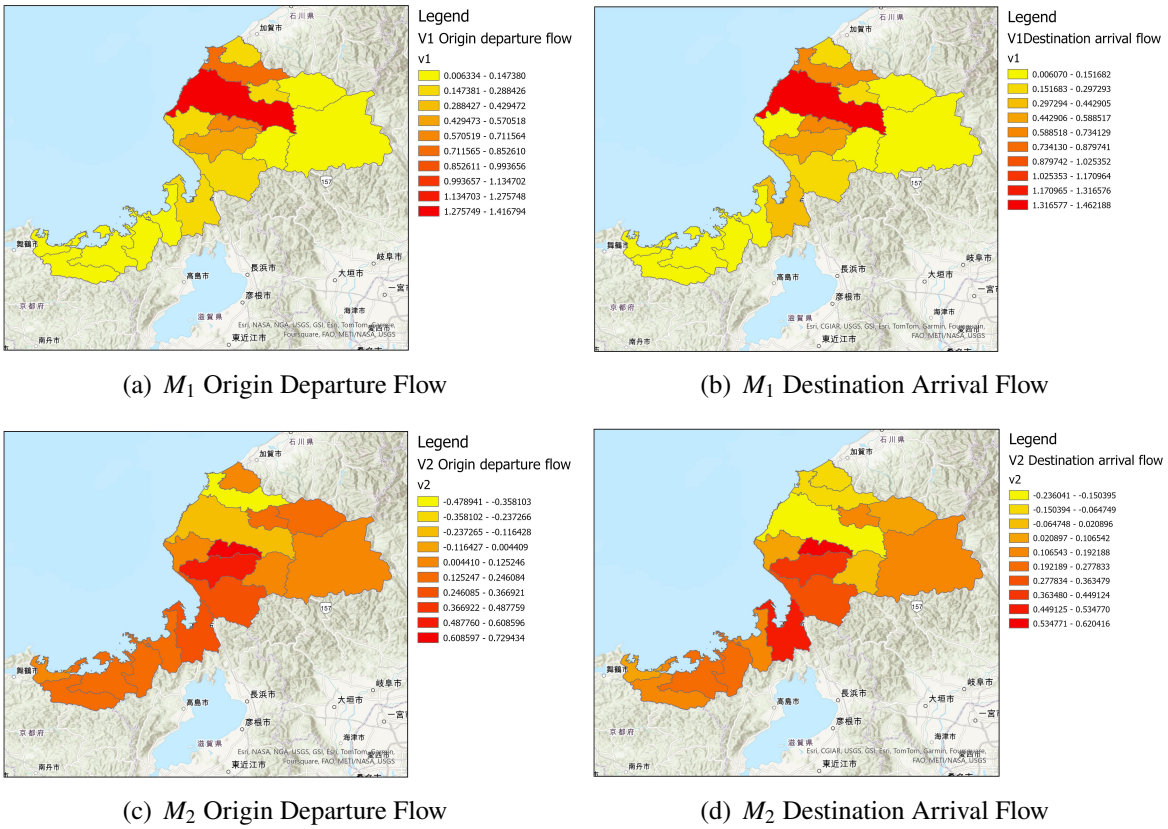
Spatially, most OD pairs showed minimal impact (values close to zero), but several exhibited significant positive and negative fluctuations. These fluctuations may indicate relatively small but crucial spatial changes in the time pattern corresponding to u_2. Positive peaks represent abnormal increases in transportation flow during specific periods, whereas negative peaks indicate decreases in transportation flow. Traffic volume from cities in the lower part of Fukui Prefecture to the upper part generally decreased, likely due to the concentration of the snowstorm in the upper part of Fukui. In contrast, in the upper part of Fukui, the distribution pattern shifted from a vertical distribution in the original M_1 mode to a horizontal distribution, with increased traffic flow from Fukui and its surrounding cities towards Ono city and Katsuyama city.

In Fig. 4.7, the departure and arrival volumes of intercity transportation demand for different cities are aggregated separately. The intensity of the colors reflects the magnitude of fluctuations in the corresponding directions.

By integrating the temporal and spatial unit vectors, it is possible to observe the fluctuations in various types of intercity transportation demand. If significant fluctuations are observed in the temporal unit vectors during a specific period, and corresponding significant fluctuations are noted in the departure and arrival flows of a particular city or between specific OD pairs, this indicates a substantial increase in that type of demand within the respective temporal and spatial scope. Conversely, if the temporal and spatial unit vectors exhibit opposite fluctuations, it suggests a significant decrease in that type of demand within the respective temporal and spatial range.

In the M_1 model, as shown in Fig. 4.7(a) and Fig. 4.7(b), the origin departure flows and destination arrival flows are mainly concentrated in Fukui City and Sakai City. However, in the M_2 model, the origin departure flows are primarily in Sabae City and Echizen City, with a significant decrease in departure flows from Sakai City, as shown in Fig. 4.7(c) and Fig. 4.7(d). Additionally, the destination arrival flows in Fukui City have greatly decreased, while the traffic volume in Tsuruga City has significantly increased. This indicates that the

snow disaster primarily affected the northern part of Fukui Prefecture, particularly Fukui City and Sakai City.



(a) M_1 Origin Departure Flow

(b) M_1 Destination Arrival Flow

(c) M_2 Origin Departure Flow

(d) M_2 Destination Arrival Flow

Figure 4.7: Origin Departure and Destination Arrival Flows under the M_1 and M_2 Models.

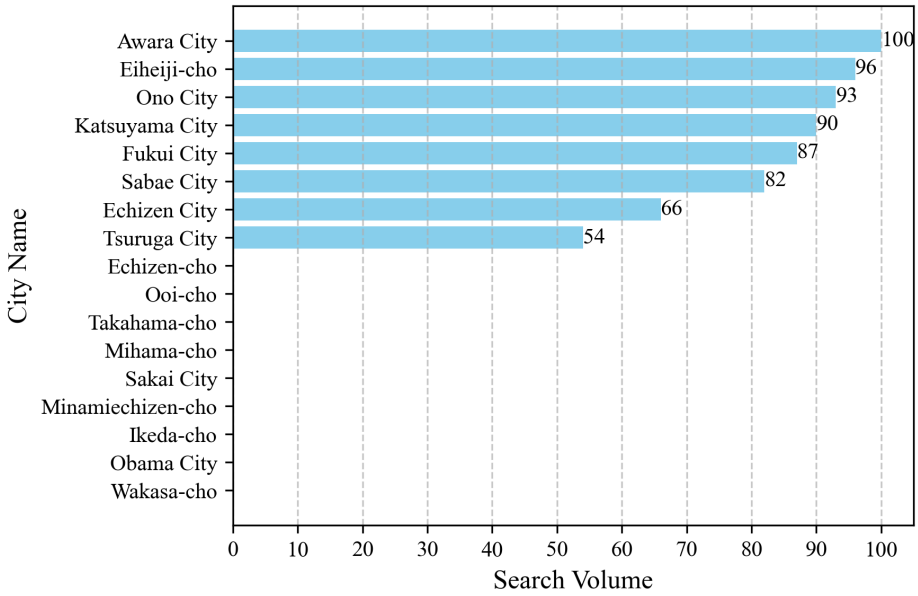


Figure 4.8: Ranking of Cities by Snowstorm-Related Searches Over the Past 10 Years

In addition, using Google Trends data to identify the cities with the highest search

volumes for snowstorm-related keywords in Fukui Prefecture over the past 10 years, as shown in Fig. 4.8, reveals that the cities with the highest search volumes largely align with those identified in this study as being affected by snowstorms. The scores range from 0 to 100: a score of 100 indicates the highest search volume, 50 represents half that, and 0 indicates insufficient data. This consistency in search patterns suggests that the spatiotemporal characteristics of snowstorms in Fukui Prefecture have remained similar over the past decade. Consequently, the traffic patterns observed during this particular snowstorm may reoccur in future events.

4.3 Discussion and Limitations

Spatial Resolution Limitation: The current analysis was conducted primarily at the city level. Although it effectively captured intercity transportation demand variations, it lacked finer-scale analysis, especially for critical transportation corridors such as National Route 8, which experienced severe congestion during the snowstorm. Future studies should consider a grid-level or road segment-level analysis to provide more precise spatial insights.

Case Selection and Representativeness: Although this study focused solely on the 2018 snow disaster in Fukui Prefecture, the selection of this case was both intentional and justified. The 2018 event was one of the most severe snowstorms in recent years, causing large-scale traffic paralysis—particularly on National Route 8—and drawing significant public and governmental attention. It provides a rich dataset for examining the impact of extreme snowfall on urban mobility and infrastructure.

In addition, an examination of Google Trends data over the past decade shows that the cities most affected in 2018—such as Fukui, Awara, and Sabae—frequently appear in search queries related to other major snowstorm events. This suggests that the spatial pattern of impact observed in 2018 shares partial consistency with broader long-term trends, thereby supporting its representativeness.

However, I acknowledge that relying on a single-event case study limits the temporal generalizability of the findings. Future research will aim to incorporate data from multiple snowstorm events, such as those in 2021 or 2024, to evaluate whether the observed spatial risk patterns are consistent across different temporal and regional contexts.

Interpretive Depth of Analytical Results: The SVD-based visualization in this chapter revealed concentrated traffic patterns during the snow disaster period. While the initial description highlighted severely affected areas such as Fukui, Awara, Sakai, and Sabae, a deeper interpretation underscores the broader significance of this finding: (1) the spatial clusters provide empirical support for defining the dependent variable Y in Chapters 5 and 7; (2) the clustering patterns directly align with the core research objective of identifying

vulnerable urban transportation corridors under snowstorm conditions; and (3) the results offer practical implications by identifying critical zones—particularly in northern Fukui Prefecture—that should be prioritized in future emergency preparedness, snow removal operations, and resource allocation strategies.

4.4 Conclusion

This study utilized the KDTree algorithm and SVD to decompose and analyze the traffic patterns and demand in Fukui Prefecture during the 2018 snow disaster. The findings are as follows:

- **Identification of Event Phases**: The study identified three phases of the event: the stable phase (January 27 to February 2), the snow disaster phase (February 3 to February 11), and the recovery phase (February 12 to February 16). During the snow disaster, intercity transportation demand dropped by 67.86% compared to the stable phase.
- **Spatial and Temporal Characteristics of Transportation under Different Modes**: Intercity transportation demand during the snow disaster included daily demand (M_1) and special demand (M_2). In the M_1 model, traffic primarily originated from Fukui City, Sabae City, Awara City, and Sakai City, with Fukui City and Sakai City being key points of departure and arrival. In the M_2 model, traffic from southern to northern Fukui Prefecture nearly ceased, shifting from a longitudinal pattern along National Route 8 and the Hokuriku Expressway to a horizontal distribution towards Ono city and Katsuyama city. Departure flows from Sakai City decreased significantly, while Tsuruga City saw a significant increase in traffic.
- **Impact Scope**: Snow disaster mainly affected northern Fukui Prefecture, especially Fukui City and Sakai City. The traffic pressure is mainly concentrated in Fukui City, Sabae City, Awara City, Echizen City, Sakai City, Katsuyama City, and Tsuruga City. Once a snow disaster occurs, these areas are more prone to traffic congestion or accidents. The government should strengthen the allocation of evacuation resources in these cities.
- **Validation with Google Trends**: An analysis using Google Trends data for snowstorm-related keyword searches over the past 10 years in Fukui Prefecture revealed that the highest-ranked cities largely correspond to those identified in this study as being affected by snowstorms. This consistency suggests that the spatiotemporal characteristics of snowstorms in Fukui Prefecture have remained similar over the past decade, and the traffic patterns observed during this snowstorm may reoccur in future events. This

indicates that the spatiotemporal characteristics of snowstorms in Fukui Prefecture have shown similar patterns over the past decade. Consequently, the traffic patterns observed during this particular snowstorm are likely to reoccur in future snowstorms. However, due to the lack of precise snowfall data for Fukui Prefecture over the past decade, and given that each snowstorm exhibits distinct spatiotemporal characteristics, the impact on the road network and traffic flow may vary. Therefore, the present study can only provide a general identification of areas and roads significantly affected by snowstorms. Accurately predicting the road segments most impacted in future snowstorms remains a key direction for further research.

Chapter 5

Detecting High-Risk Traffic Congestion Areas

In Chapter 4, the spatiotemporal variations of intercity transportation under snowstorm conditions were analyzed at the city level. However, critical congestion corridors, such as National Route 8 and the Hokuriku Expressway—frequently highlighted in government reports—were not examined at a finer spatial resolution. Therefore, this chapter focuses on grid-level analysis to identify high-risk traffic congestion areas during snow disasters.

Early identification of areas susceptible to traffic congestion caused by snowstorms is essential for formulating effective emergency response strategies and optimizing resource allocation. This study focuses on the 2018 heavy snowstorm in Fukui Prefecture, Japan, and integrates mobile GPS data with multi-source remote sensing datasets, including the Digital Elevation Model (DEM), land use data, nighttime light imagery, the Normalized Difference Vegetation Index (NDVI), urban area information, and other relevant spatial indicators. A spatial machine learning model based on the Random Forest algorithm is developed to identify potential congestion segments at a high spatial resolution.

5.1 Methodology

5.1.1 Methodology Framework

The Docomo data were spatially visualized and analyzed using ArcGIS Pro. Road-related features were extracted at the grid level for model training. A Random Forest algorithm was employed, using data from 10 cities in Fukui Prefecture, with 70% of the data randomly selected for training and the remaining 30% for testing. The trained model was then applied to the complete dataset covering all 17 cities in Fukui Prefecture for further evaluation.

The overall research workflow consists of four key steps, as illustrated in Fig. 5.1: (1) Data collection and feature computation; (2) Visualization of road traffic variations; (3) Model development and cross-validation; (4) Identification of congestion segments.

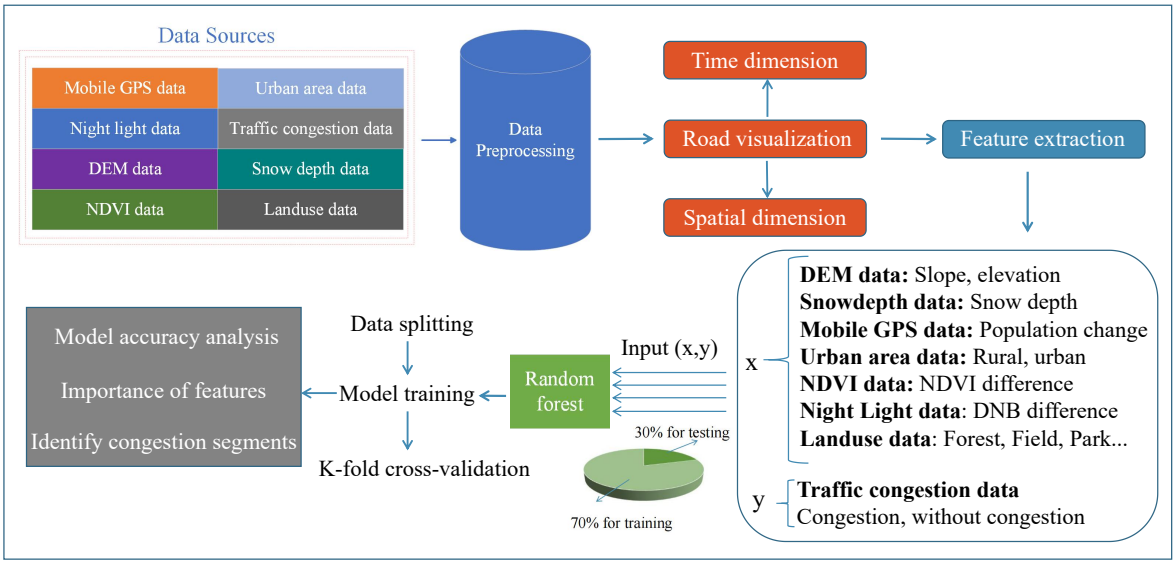


Figure 5.1: Methodology framework

5.1.2 Data preprocessing and feature extraction

The data preprocessing procedures have been comprehensively described in Chapter 3. This study primarily utilizes multiple datasets, including mobile GPS data collected during the 2018 snowstorm in Fukui Prefecture, as well as urban area data, traffic congestion data, digital elevation model (DEM) data, snow depth data, night light data, NDVI data, and land use data. The details of these data sources are listed in Table 5.1.

Table 5.1: Model Feature Types and Temporal-Spatial Resolutions

Data	**Spatial Resolution**	**Temporal Resolution**	**Variable Type**
Docomo Mobile GPS Data	500 m	4 AM, 7 AM, 12 PM	Feature
Road Network Data	–	–	Feature
Urban Area Data	–	–	Feature
DEM Data	500 m	–	Feature
Land Use Data	500 m	–	Feature
Night Light Data	500 m	Daily	Feature
NDVI Data	500 m	16 Days	Feature
Snow Depth Data	City Level	Hourly	Feature
Traffic Congestion Data	City Level	–	Target

5.1.3 Random Forest for Traffic Congestion Identification

After data preprocessing, a Random Forest (RF) model was employed to identify congested road segments. Random Forest was chosen for its strong predictive performance and robustness, which result from aggregating the outputs of multiple decision trees. Additionally, it offers good model interpretability, making it suitable for identifying key congestion-related factors.

Random Forest is an ensemble classification algorithm that improves accuracy by constructing multiple independent decision trees during training. Each tree votes for a class label, and the final prediction is made based on majority voting. This ensemble approach reduces overfitting and enhances generalization compared to single-tree models.

In Random Forest, the final classification $\hat{y}$ is determined by a voting mechanism across all trees. If there are T trees in the forest, each tree $h_t(x)$ provides a prediction for the input x. The final prediction is given by the mode of all tree outputs [114]:

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), \ldots, h_T(x)\} \tag{5.1}$$

where $h_t(x)$ is the prediction from the t-th tree, and mode represents the majority voting operation, selecting the class supported by the most trees. This majority voting mechanism ensures robustness against individual tree variance.

Instead of direct majority voting, we propose an alternative approach based on probability averaging to identify traffic congestion segments. Given T decision trees, each tree $P_t(Y = 1|X)$ predicts the probability of congestion given feature vector X. The overall probability is computed as:

$$P(Y = 1|X) = \frac{1}{T} \sum_{t=1}^{T} P_t(Y = 1|X) \tag{5.2}$$

where $P_t(Y = 1|X)$ represents the probability predicted by tree t.

To classify whether a road segment is congested, we apply a threshold-based decision rule:

$$\hat{Y}_i = \begin{cases} 1, & P(Y = 1|X_i) \geq 0.5 \quad \text{(Congested)} \\ 0, & P(Y = 1|X_i) < 0.5 \quad \text{(Non-congested)} \end{cases} \tag{5.3}$$

where X_i denotes the feature vector for a given road segment, and Y_i is the corresponding target variable indicating whether congestion occurs.

By aggregating probability-based outputs from multiple decision trees, Random Forest effectively captures complex relationships between input features and congestion status, reducing overfitting and improving prediction robustness.

5.2 Result and discussion

5.2.1 Spatiotemporal Visualization of Road Traffic Flow

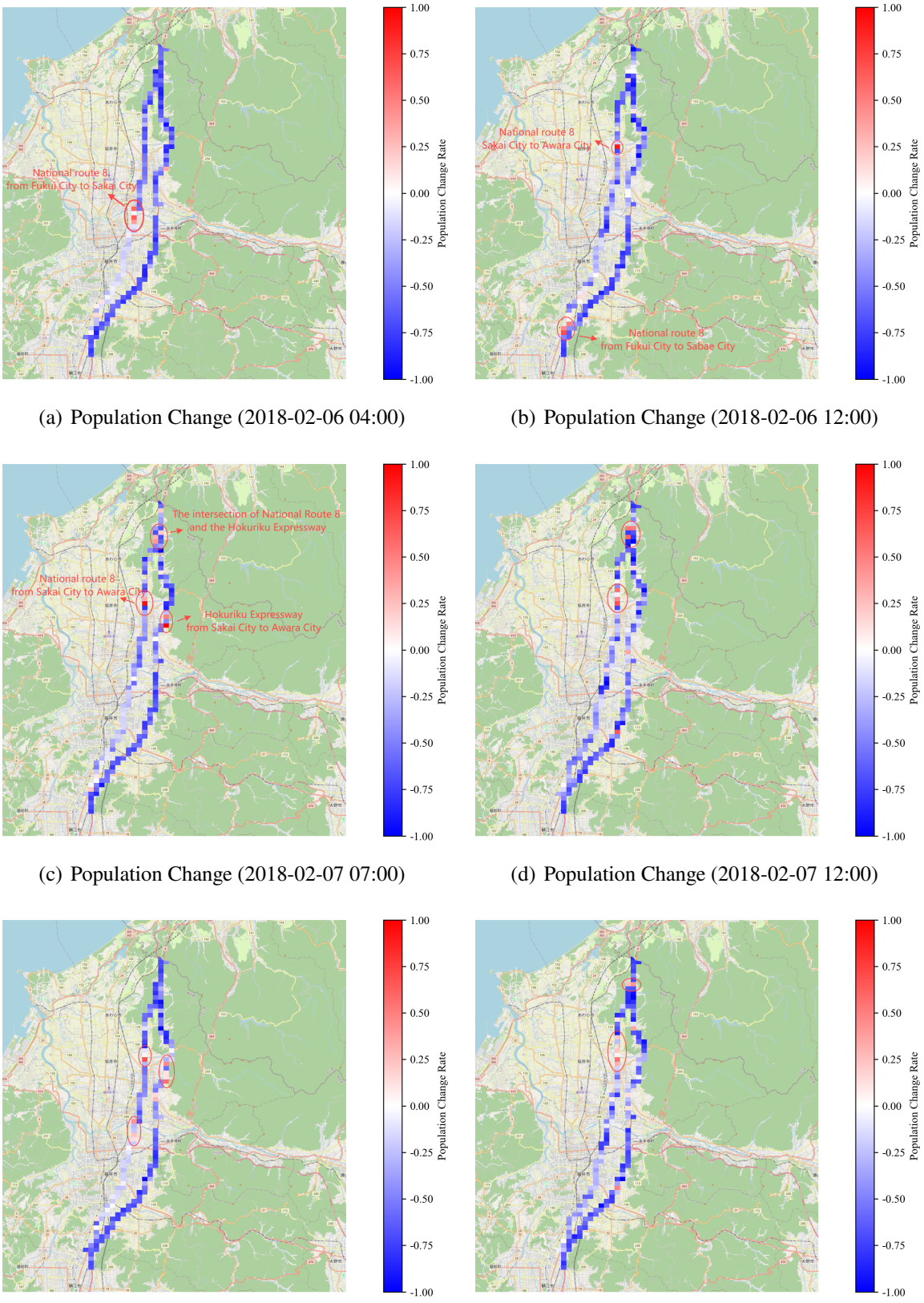
To analyze the variations in road traffic flow during snow disasters, we conducted a visualization analysis of the three cities reported to have experienced traffic congestion in the government report [2]. This study focuses on major roads, specifically National Route 8 and the Hokuriku Expressway, and applies a 500-meter grid to process and visualize the data. Similar to the previous GPS data processing approach, we calculated the average population values for each grid during the week before the snow disaster and during the disaster week. The population difference was then computed to identify regions where significant population changes occurred during the snow disaster.

To optimize the visualization, we selected time points with significant population changes, as shown in Fig. 5.2. In the visualization: Blue grids indicate population decreases, with deeper shades representing larger declines relative to normal conditions. Red grids indicate population increases, with deeper shades representing a more significant rise compared to normal conditions.

Based on the study background, a significant increase in red grids may indicate traffic congestion points. For example, at 4:00 AM on the 6th, traffic volume significantly increased on roads from Fukui City to Sakai City and Fukui City to Sabae City. After 7:00 AM, similar increases were observed on roads from Sakai City to Awara City. From 7:00 AM on the 7th, the traffic pressure visibly shifted from Fukui City to Sakai City and Awara City, with notable population surges in three key road segments: National Route 8 from Sakai City to Awara City, Hokuriku Expressway from Sakai City to Awara City, and the intersection of National Route 8 and the Hokuriku Expressway.

These traffic increases persisted until the 8th. The observed traffic congestion points align well with the Fig. 3.3, confirming that GPS-based population distribution effectively identifies congestion points on roads. If real-time data were fully available, it would theoretically enable quicker detection of congestion formation, thereby enhancing traffic monitoring and early warning capabilities.

However, it should be noted that not all increases in population necessarily indicate traffic congestion. In some suburban or boundary areas, red grids may reflect temporary vehicle or pedestrian accumulation due to detours or local conditions. Therefore, red grids in this study are interpreted as potential indicators of congestion.



(a) Population Change (2018-02-06 04:00)

(b) Population Change (2018-02-06 12:00)

(c) Population Change (2018-02-07 07:00)

(d) Population Change (2018-02-07 12:00)

(e) Population Change (2018-02-08 04:00)

(f) Population Change (2018-02-08 12:00)

Figure 5.2: Population Change Along National Route 8 and Hokuriku Expressway Over Time

5.2.2 Land Characteristics of Congestion Points

To investigate whether the traffic congestion points during the snow disaster share common spatial characteristics, we extracted the 500-meter grids with a significant increase in GPS-based population data from the road visualization map and analyzed their land-use characteristics.

The results indicate that all grids experiencing a sharp increase in traffic volume are located in urban areas, which aligns with the general understanding that urban areas tend to have higher traffic volumes and are more prone to congestion than rural areas. However, an intriguing finding is that the primary land-use types in these congestion-prone grids are field (40%) and forest (46.7%), as shown in Fig.5.3. This result deviates from conventional expectations, prompting us to explore possible explanations based on relevant literature. Three potential reasons are proposed:

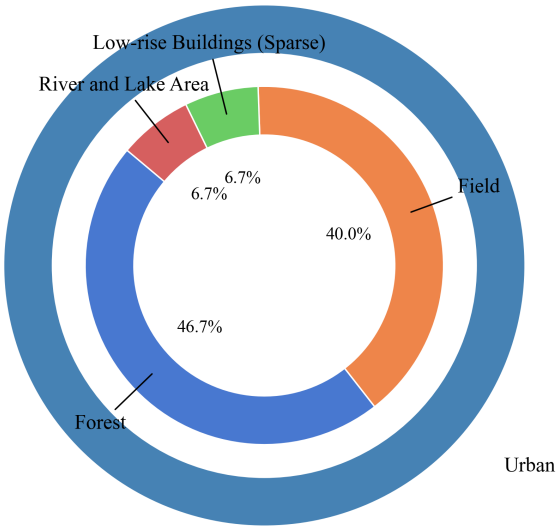


Figure 5.3: Land Type Distribution of Congestion Grids

Weak Transport Infrastructure: In urban core areas, such as commercial and residential districts, governments typically allocate more resources for road maintenance and snow removal. In contrast, urban peripheral areas, such as farmland and forests, tend to have weaker road infrastructure and receive fewer maintenance resources, leading to slower snow clearance and increased susceptibility to road closures and congestion during snow disasters [115].

Urban Heat Island Effect: In densely built urban cores, high vehicular activity and human movement enhance snow compression and melting, shortening the time required for roads to become passable. In contrast, forested and agricultural areas may experience reduced sunlight exposure due to tree cover, lowering ground temperatures and slowing snowmelt [116, 117]. Additionally, lower temperatures in these areas can lead to ice layer formation, creating hazardous driving conditions and increasing traffic congestion risks.

Steep Terrain: Many urban fringe areas are characterized by steep slopes, which increase the risk of snow-induced landslides and traffic accidents, further obstructing road accessibility [118, 119]. Moreover, roads with greater slopes tend to accumulate snow more easily, reducing road capacity and exacerbating congestion.

Building upon the above findings, we plan to integrate DEM data, Nighttime Light data, and NDVI data into a comprehensive model to further quantify the impact of environmental factors on road congestion during snow disasters. This study not only enhances our understanding of how snow disasters affect road traffic but also provides data-driven insights for future traffic planning and disaster response strategies.

Model results of of traffic congestion prediction

In this study, we first constructed a small-scale model using a Random Forest (RF) classifier for 10 cities in northern Fukui. We then validated this model by applying it to the full dataset covering all 17 cities in Fukui Prefecture. For both cases, the dataset was split into 70% training and 30% testing sets, and a 5-fold cross-validation was performed to ensure robustness. Table 5.2 presents the accuracy, precision, recall, and F1-score for both tests. Additionally, we analyzed the confusion matrix and the ROC curve of the RF model to further evaluate its classification performance, as shown in Fig. 5.4 and Fig.5.5.

Table 5.2: Comparison of Model Performance in Different Regions

Metric	**Northern Fukui (10 cities)**	**Fukui Prefecture (17 cities)**
Accuracy	0.8629	0.9459
Precision	0.7593	0.7500
Recall	0.7885	0.8667
F1-score	0.7736	0.8041

As shown in Table 5.2, the Random Forest model achieved higher overall performance when applied to the full Fukui Prefecture (17 cities), with an accuracy of 94.59% and an F1-score of 0.8041. In contrast, the model trained only on the northern part of Fukui (10 cities) yielded a slightly lower accuracy of 86.29% and an F1-score of 0.7736, indicating that expanding the training dataset improved classification performance.

Fig.5.4 shows the confusion matrix comparison, where the model trained on all 17 cities produced a significantly higher number of true negatives (TN) and slightly fewer false negatives (FN), reflecting better detection of non-congested grid cells. Fig.5.5 illustrates the ROC curves of both models, where the full-region model attained a higher AUC (0.98) compared to the northern-region model (0.93), further confirming its stronger predictive capability.

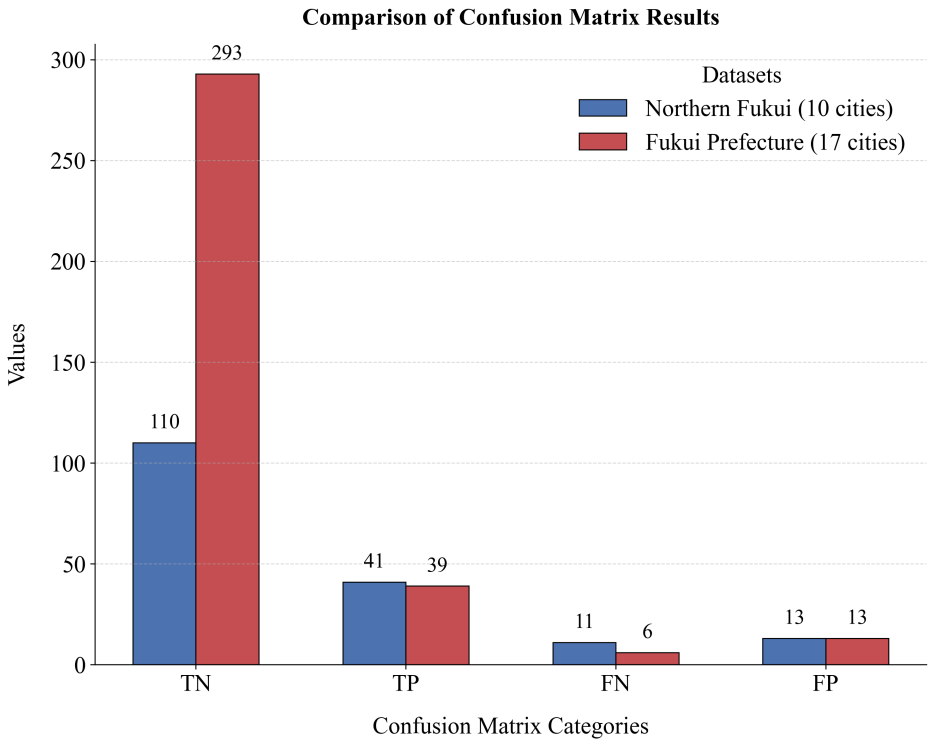


Figure 5.4: Confusion Matrix Results of Random Forest

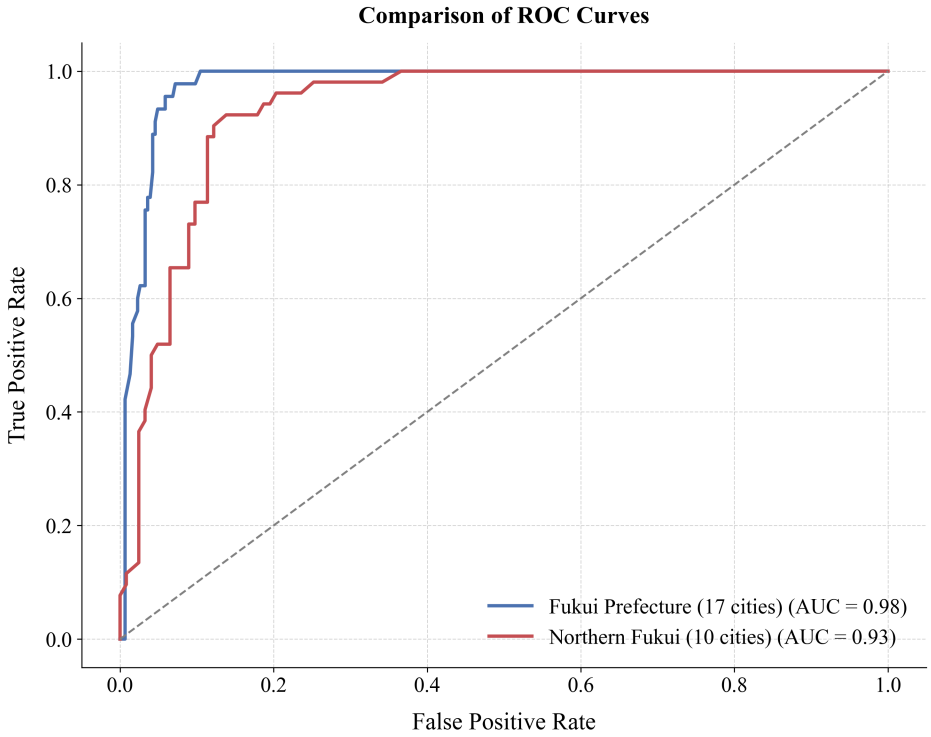


Figure 5.5: ROC Curve Results of Random Forest

5.2.3 Importance of features

Feature importance analysis quantifies the relative contribution of each variable to the model's output, with values being unitless. As shown in Figure 5.6, snow depth difference

exhibits the highest importance, reinforcing its direct relationship with snow disaster-induced traffic congestion. This aligns with previous studies [57], which suggest that snow accumulation significantly affects road accessibility and safety.

The second most influential factor is night light difference. During snow disasters, traffic volume decreases sharply due to hazardous conditions, leading to a reduction in night-time illumination. This phenomenon has also been reported in the literature [27], highlighting the necessity of integrating remote sensing data in snow disaster analysis.

Furthermore, the strong influence of elevation and slope angle on model output is consistent with our previous hypothesis that higher elevation and steeper slopes increase the probability of traffic congestion during snow disasters. This finding also supports our prior analysis of individual congestion points. In contrast, the lower importance of NDVI may be attributed to its coarse temporal resolution (16-day intervals), limiting its effectiveness in capturing rapid snow-induced changes.

By contrast, urban, population change, and land use types (except fields and forests) exhibit less importance of characteristics, indicating that their impact on snow-related traffic congestion is relatively limited in the current model. The feature importance analysis ranked key contributors as follows: Snow Depth > Nighttime Light Difference > Elevation > Slope Angle > Urban Area > NDVI > Population Change > Forest > Field > Low-rise Buildings (Sparse) > Low-rise Buildings (Dense).

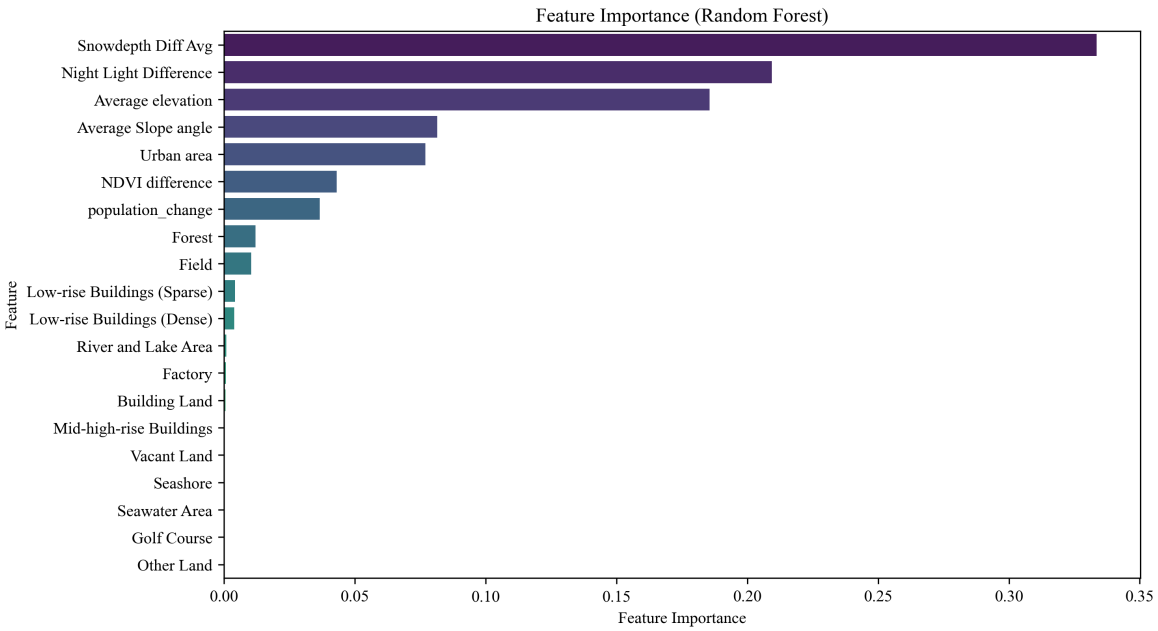


Figure 5.6: Importance of Features

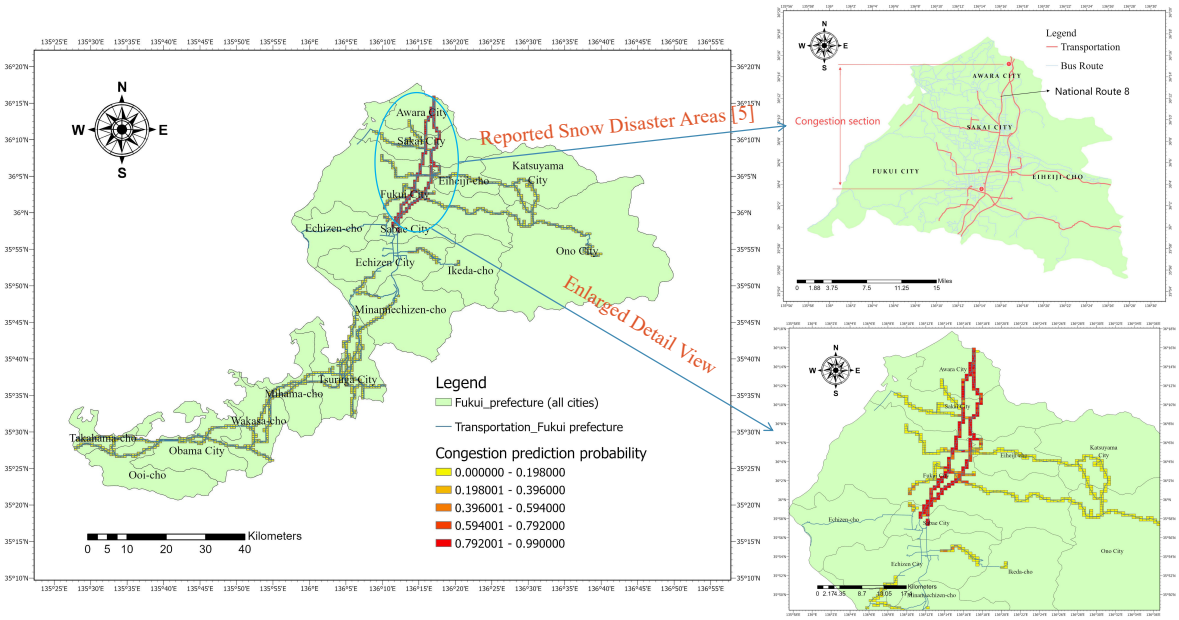


Figure 5.7: Predicted and Actual Congestion Segments [2]

5.2.4 Risk Map of Traffic Congestion from Model Prediction

Based on the developed spatial Random Forest model, we predicted the probability of snow disaster-induced traffic congestion for each 500-meter grid cell across Fukui Prefecture. The results were visualized as a risk map, as shown in Fig. 7.7, where redder grid colors indicate higher probabilities of snow disasters.

To validate the model's effectiveness, the predicted high-risk road segments were compared with government-reported congestion zones during the 2018 snowstorm event. The risk map highlights Fukui City, Sakai City, and Awara City as the most vulnerable areas, which is consistent with the affected regions reported in [2].

In summary, Fig. 5.7 lays a data-driven foundation for developing effective snow disaster response strategies in Fukui Prefecture. Compared to remote sensing data, which often suffer from limited spatial resolution and low temporal frequency, high-quality GPS data offer finer granularity and real-time potential. By integrating the complementary strengths of both data sources, there is promising potential to develop dynamic prediction models for snow disaster management in the future.

5.3 Limitations and discussions

Severe GPS Data Gaps: The substantial absence of GPS data in certain cities poses a significant challenge for accurate congestion prediction. As observed in the final prediction map, cities such as Sabae, Echizen City, and Ikeda-cho exhibit severe data gaps, limiting the model's ability to capture congestion patterns effectively.

Furthermore, the low temporal resolution of Docomo GPS data restricts real-time traffic analysis. With only three time points available per day, constructing a fully dynamic congestion monitoring model remains a challenge.

Potential for Dynamic Real-Time Detection: Although GPS data showed limited contribution in feature importance analysis, it effectively captures population dynamics on roads. If a complete, high-frequency GPS dataset were available, it would enable the development of a real-time congestion detection model, leveraging the integrated data framework proposed in this study.

5.4 Conclusion

This study utilized GPS data to reconstruct the real traffic conditions during the 2018 Fukui snow disaster. By integrating GPS data with multi-source remote sensing data, a spatial analysis model was developed using a Random Forest classifier. This model identifies road segments affected by snowstorm-induced congestion and generates a 500-meter grid-based congestion probability map. The key findings of this study are as follows:

GPS data effectively captures population dynamics on roads. With high-quality, high-frequency GPS datasets, real-time congestion detection would be feasible, offering valuable insights for emergency traffic management.

In the case of the 2018 Fukui snow disaster, all points where traffic volume surged sharply during the snowstorm were located in urban areas, with the predominant land types being field and forest. This phenomenon may be attributed to factors such as weak transport infrastructure, the urban heat island effect, and steep terrain.

By integrating remote sensing data and GPS data with machine learning models, congested road sections can be effectively identified. Feature importance analysis highlighted the most influential variables in predicting congestion as follows: Snow Depth > Nighttime Light Difference > Elevation > Slope Angle > Urban Area > NDVI > Population Change > Forest > Field > Low-rise Buildings (Sparse) > Low-rise Buildings (Dense).

This study demonstrates that integrating multi-source remote sensing data with GPS-based mobility data and machine learning techniques provides an effective framework for identifying traffic congestion patterns during extreme snowfall events. Future research should focus on improving the real-time availability of data and developing a dynamic real-time road monitoring model to enhance the accuracy and applicability of congestion prediction systems.

Chapter 6

Assessing Regional Resilience of Different Land Use Types

In the previous chapter, a grid-based analysis was conducted to identify traffic congestion points along National Route 8 and the Hokuriku Expressway under snowstorm conditions. A traffic congestion identification model was also developed. Notably, the results revealed that these congestion points appeared to be closely associated with specific land use types. To further assess resilience at the grid level, this chapter focuses on evaluating the snow disaster resilience of different land use types, providing a finer spatial understanding of regional vulnerability and recovery potential.

Understanding the resilience of areas with different land-use types can enhance a city's ability to respond to and recover from disasters. Based on Docomo mobile GPS data and the 2018 Fukui Prefecture snow disaster, this study explores the resilience of areas with different land-use types from the grid level. First, the resilience triangle method was employed to assess the robustness, vulnerability, and survivability of different land-use types at the 500-meter grid level. Second, Pearson correlation analysis and causal inference determined that snow depth is a causal factor leading to changes in grid resilience, thereby identifying the optimal time points for each grid to distinguish between disaster end and recovery start. Finally, a resilience map of Fukui Prefecture was created, integrating road data to identify cities with poor resilience and road sections prone to congestion during snow disasters.

6.1 RESEARCH METHODOLOGY

The primary methods employed in this study include the resilience triangle method and Pearson correlation analysis with causal inference. The resilience triangle method is used to explore resilience metrics for different land-use types, including robustness, vulnerability, and survivability. Pearson correlation analysis and causal inference are utilized to identify

key weather factors affecting resilience and to determine the optimal time points for each grid to transition from disaster end to recovery start. The research workflow is illustrated in Fig. 6.1.

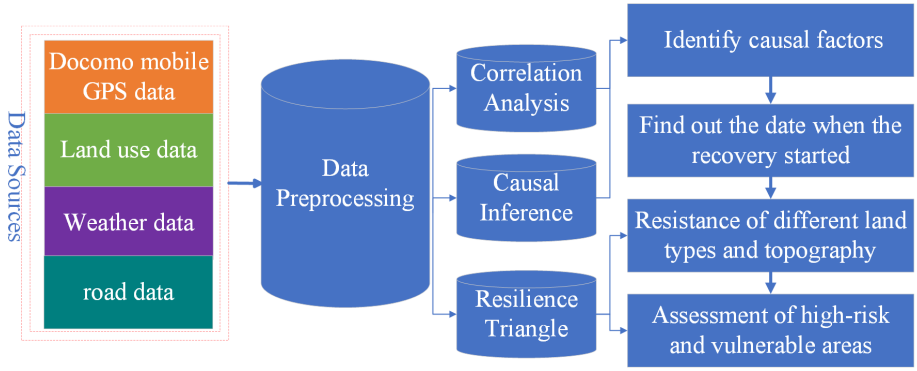


Figure 6.1: Methodology framework

6.1.1 Models for assessing regional resilience of different land use types

Resilience Triangle Method

The "Resilience Triangle" method evaluates the disaster resilience of a system by analyzing changes in system performance throughout the entire disaster process, emphasizing the process of performance change. This method quantifies resilience through mathematical integration to determine the resilience of the study area [120, 121], as shown in Fig. 6.2 . In this study, the method is applied to assess the resilience of different land use types during snow disasters by comparing the integral area of the resilience triangle parameters at the grid level.

- **Vulnerability**: The grid's resistance to disaster occurrence. The drop from point A to point B in the figure represents vulnerability.
- **Survivability**: The remaining capacity. Point B in the figure indicates the lowest performance level of the system during the disaster.
- **Robustness**: The grid's recovery rate. The speed and magnitude of the rise from point B to point C in the figure represent robustness.

In this study, the calculation and explanation of each parameter are as follows:

$$Q(t) = \frac{Q_{special}}{Q_{normal}} \tag{6.1}$$

$$RP = \int_{t1}^{t3} Q(t)dt \tag{6.2}$$

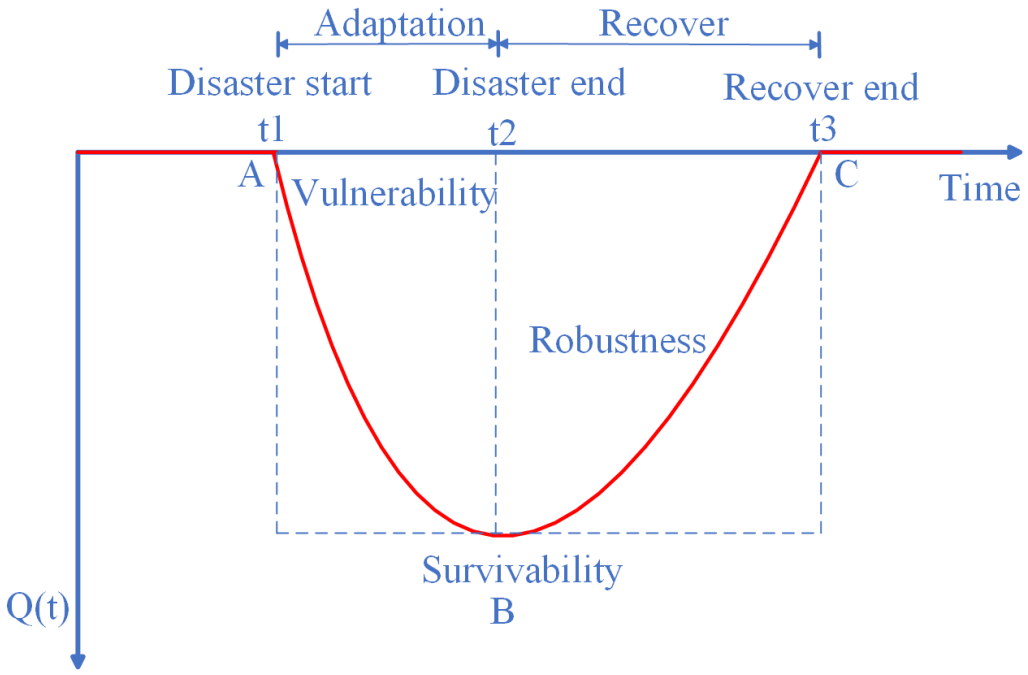


Figure 6.2: Description of the Resilience Triangle

- **Q(t)**: Population fluctuation curve over time.
- $\mathbf{Q_{normal}}$: Population fluctuation value during normal periods.
- $\mathbf{Q_{special}}$: Population fluctuation value after the start of the disaster.
- **Resilience Performance (RP)**: Resilience loss from t_1 to t_3.

Pearson Correlation Analysis

Pearson correlation analysis is a statistical method used to measure the linear relationship between two variables. By calculating the Pearson correlation coefficient (commonly denoted as r), the strength and direction of the relationship between the variables can be determined. The value of the Pearson correlation coefficient ranges from -1 to 1 [122], where:

- $r = 1$ indicates a perfect positive correlation, meaning both variables move in the same direction.
- $r = -1$ indicates a perfect negative correlation, meaning the variables move in opposite directions.
- $r = 0$ indicates no linear correlation, meaning there is no apparent linear relationship between the variables.

Causal Inference

The existence of a correlation between two variables does not necessarily imply a causal relationship between them [123]. Causal inference is commonly conducted using the Difference-in-Differences (DiD) method [124], which is a statistical approach for estimating the treatment effect between a "treatment group" and a "control group." However, DiD has several limitations. First, it relies on the parallel trends assumption, which requires that, in the absence of treatment, the difference between the treatment and control groups remains consistent over time. Second, when the outcome of interest evolves dynamically over time (e.g., disaster recovery patterns), the parallel trends assumption may not hold, rendering DiD unsuitable [3, 125]. Additionally, in this study, it is challenging to identify GPS data that simultaneously satisfy both time difference and group difference requirements. Therefore, the Bayesian Structural Time Series (BSTS) model was chosen, as its key advantage lies in its ability to infer causal effects without requiring a clearly defined control group, such as using GPS data from cities unaffected by snow disasters.

BSTS is a type of state-space model that probabilistically estimates directly observable time series data and unobservable latent variables, such as the trend of variable changes [126], as shown in Fig. 6.3. Observable variables and unobservable states (random variables) are correspondingly represented through the observation and state equations. The state equation indicates that the state evolves over time with error fluctuations. The state variable represents the hidden state within the system, which may influence the system's future state. The observation equation shows that the observed value is obtained by adding observational error (white noise) to the state vector at the same time point. The state equation and observation equation are calculated as follows:

$$x_t = x_{t-1} + \nu_t, \quad \nu_t \sim \mathcal{N}(0, \sigma_\nu^2) \tag{6.3}$$

$$y_t = x_t + \omega_t, \quad \omega_t \sim \mathcal{N}(0, \sigma_\omega^2) \tag{6.4}$$

- x_t: The state at time t.
- ν_t: The system noise at time t, following a normal distribution with a mean of zero and variance σ_ν^2.
- y_t: The observed value at time t.
- x_t: The state used in the observation equation.
- ω_t: The observation noise at time t, also following a normal distribution with a mean of zero and variance σ_ω^2.

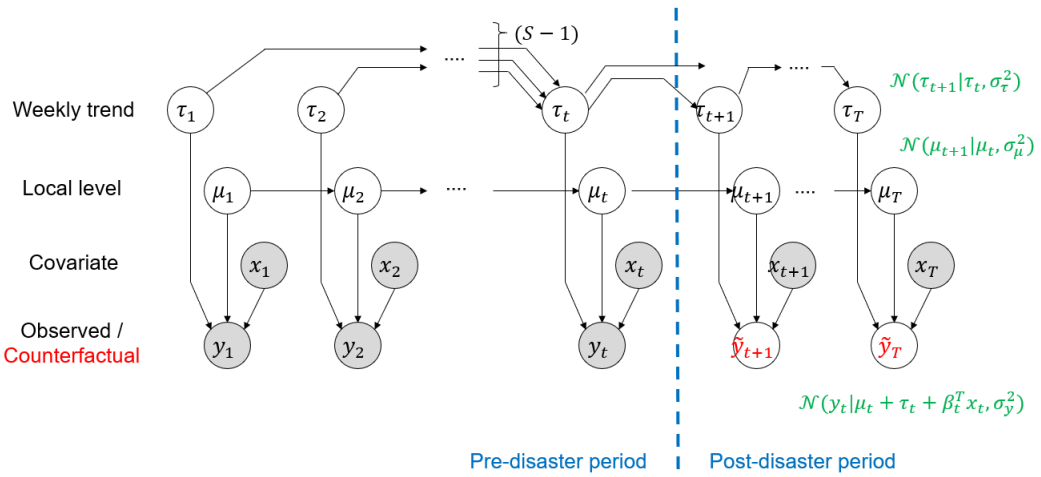


Figure 6.3: Graphical representation of the Bayesian structural time series model [3]

6.2 Results

6.2.1 Determine the start date of the grid recovery phase

After dividing Fukui Prefecture into grids of 500 meters each, there are a total of 16,815 grids. However, in the Docomo data, only 3,737 grids have complete data records at three time points each day. To reduce the impact of missing records when studying the resilience of different land types, we focus only on these 3,737 grids with complete records. Using the resilience triangle method, we can calculate the resilience indicators for each of these grids.

Fig. 6.4 shows the resilience triangle for a randomly selected grid ID, where the Q(t) curve illustrates changes in population movement during the snow disaster. From the figure, it can be seen that the Q(t) curve sharply declines starting from February 3, falling far below the pre-disaster levels, indicating a significant impact of the snow disaster on population movement within the grid. Subsequently, around February 5, there is a brief recovery in population movement, with the Q(t) value rising from -2.2 to -0.5. However, by February 12, the Q(t) curve drops to its lowest value again. This fluctuation is due to the intermittent nature of snowfall in the actual data, where population movement briefly recovers when snowfall stops, only to decline again as snowfall continues. Unlike the theoretical resilience triangle, the actual Q(t) curve exhibits multiple troughs due to the prolonged nature of the snow disaster. Therefore, identifying the optimal point among these troughs to distinguish the end of the disaster and the beginning of recovery becomes a challenge. The selection of this point directly affects the integral areas of Vulnerability and Robustness, significantly impacting the accuracy of subsequent conclusions.

Therefore, this study approaches the problem from the perspective of weather data, aiming to identify the causal factors of population changes and find the optimal segmentation point through the time series of weather changes. While we cannot obtain weather data at the

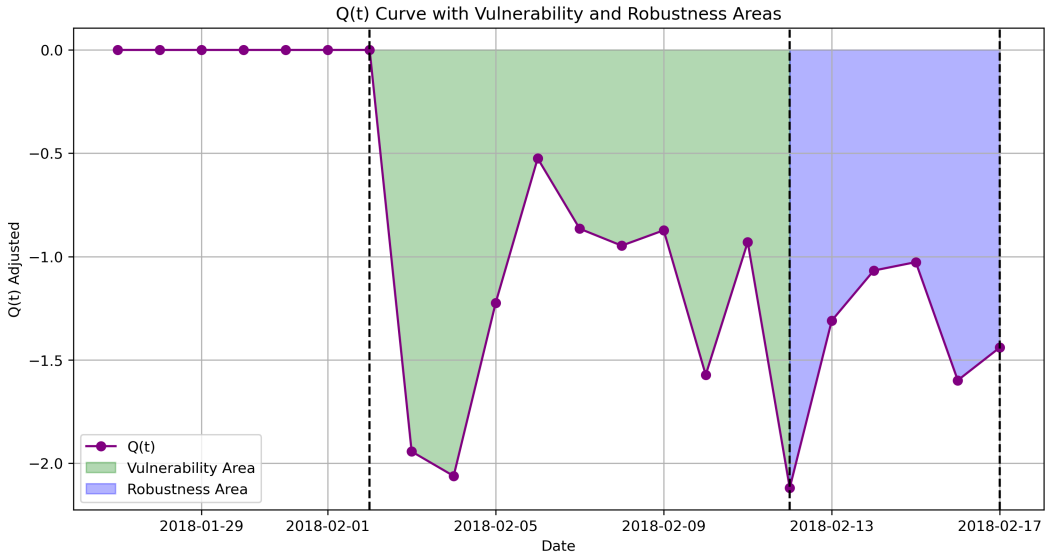


Figure 6.4: Resilience Triangles for Random Grid ID

500-meter level for each city, we have acquired 24-hour weather data at the city level for Fukui City. Through Pearson correlation analysis, we identified the most relevant meteorological factors correlated with population change rates. Based on this method, we can analyze the time series of weather changes to determine the optimal segmentation point, thereby improving the accuracy of the study's conclusions.

In light of potential multicollinearity among the meteorological variables, the Variable Inflation Factor (VIF) was calculated. Multicollinearity refers to a situation where several predictor variables in a model are highly correlated with each other. This can affect the accuracy and reliability of the model estimates. Due to the inconsistency in the units of different meteorological variables, a dimensionless normalization was performed. Moreover, to account for the substantial differences in total population among regions, the rate of population change was normalized. For certain meteorological variables with lengthy names, abbreviations were employed, with detailed correspondences provided in Fig. 6.5.

The three meteorological variables — snowfall, snow depth, and precipitation — are statistically significant and exhibit no multicollinearity, as shown in Table 6.1. Pearson's correlation coefficient indicates a strong positive association between the rate of population change and these weather variables.

Table 6.1: Correlation Analysis Results

Name	**Snow Fall**	**Snow Depth**	**Precipitation**
P-value	0.006092	0.000001	0.029048
Coefficient	0.51	0.77	0.41
VIF	2.658319	4.048196	1.623244

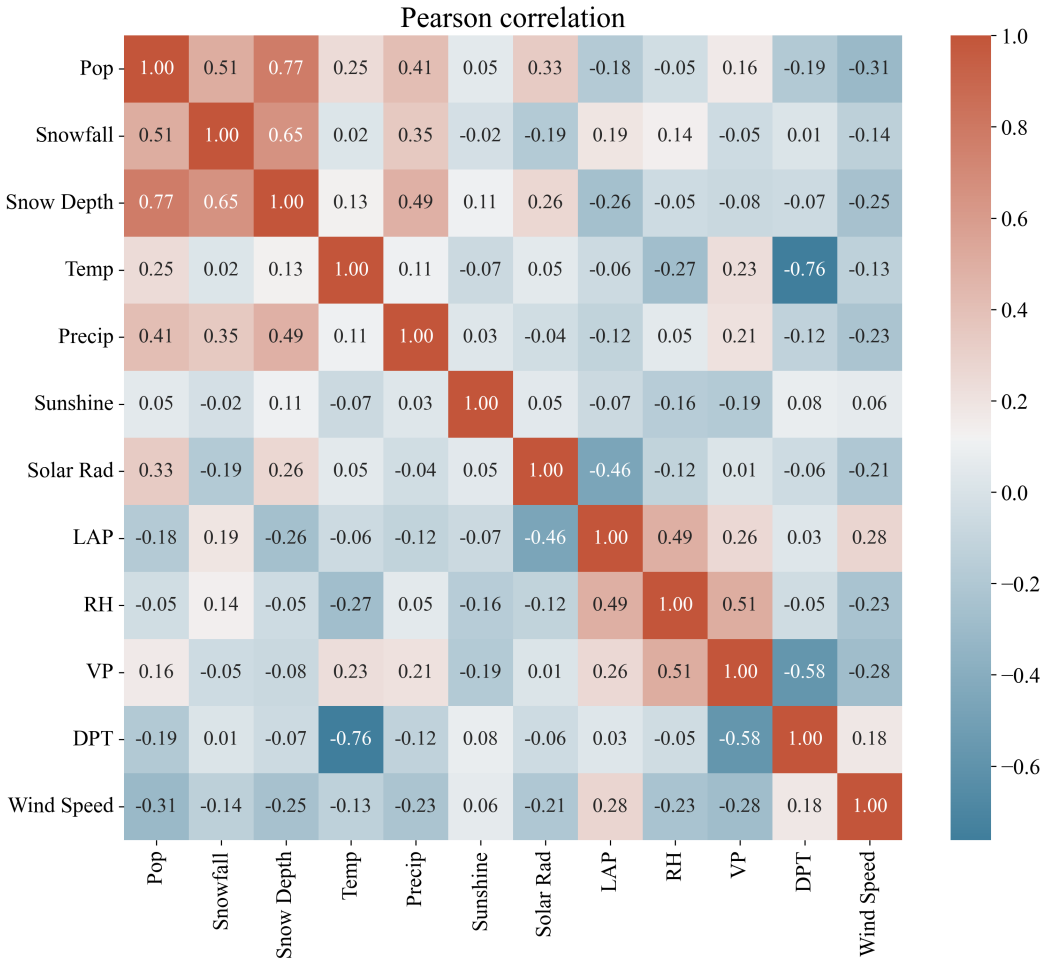
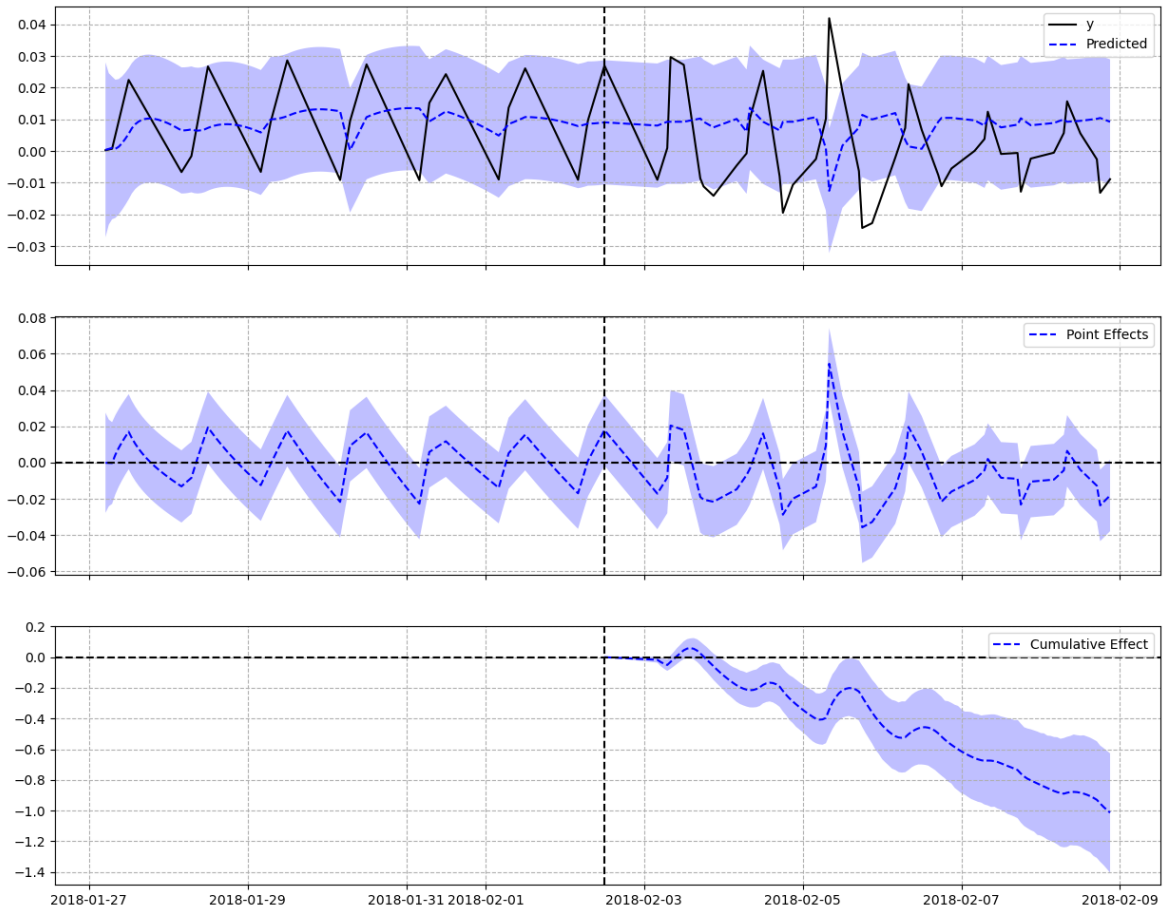


Figure 6.5: Pearson correlation between population change rate and weather variables. Abbreviations: Temp: temperature; Precip: precipitation; Sunshine: sunshine duration; Solar Rad: solar radiation; LAP: local atmospheric pressure; SLAP: sea-level atmospheric pressure; RH: relative humidity; VapPress: vapor pressure; DPT: dew point temperature.

However, the existence of a correlation between two variables does not necessarily imply a causal relationship between them. Additionally, given that precipitation levels are predominantly zero for most of the time, this dataset offers limited utility for subsequent analyses. Therefore, this study primarily focuses on investigating whether snow depth is the key factor influencing changes in population distribution, which is why causal inference is necessary [126].

Compared to traditional Difference-in-Differences (DID) methods, the Bayesian Structural Time Series (BSTS) model has the advantage of not requiring a control group, thus mitigating the challenges associated with collecting population data from cities with similar structural characteristics in the absence of snow disasters. The core of the model involves introducing intervention parameters, specifically the rate of snow depth change. We consider population change rates from January 27, 2018, to February 2, 2018, as the ”pre-intervention” period and those starting from February 3, 2018, as the ”post-intervention” period. By an-

alyzing pre-disaster actual population data in comparison to population predictions in the absence of snow disasters, this study aims to elucidate the causal impact of snow depth on population change rates.



Note: The first 1 observations were removed due to approximate diffuse initialization.

Figure 6.6: Causal Inference Results

In analyzing the impact of a snow disaster on population distribution, our study reveals notable differences between actual and predicted values. Considering the 95% confidence intervals (CIs) for both average and cumulative effects provides a measure of uncertainty in effect estimation across different dimensions. The 95% CI for the average effect reflects the uncertainty in our estimate of the intervention effect for each observation or time point. The 95% CI for the cumulative effect offers a measure of uncertainty for the accumulated amount of the intervention effect over the entire study period or after all observations are considered, as shown in Fig. 6.6. The actual average value during the disaster was 0.0, significantly lower than the predicted 0.01. Similarly, the cumulative actual value was 0.11, compared to a predicted 1.12, as shown in Table 6.2. These findings clearly indicate that both average and cumulative values were substantially reduced due to the snow disaster. The absolute effect of the disaster is evident, with an average decrease of -0.01 and a cumulative decrease of -1.01,

Table 6.2: Causal Inference Results

	Average	**Cumulative**
Actual	0.0	0.11
Prediction (s.d.)	0.01 (0.0)	1.12 (0.2)
95% CI	[0.01, 0.01]	[0.73, 1.51]
Absolute effect (s.d.)	-0.01 (0.0)	-1.01 (0.2)
95%CI	[-0.01, -0.0]	[-1.41, -0.63]
Relative effect (s.d.)	-90.44% (17.74%)	-90.44% (17.74%)
95% CI	[-125.48%, -55.95%]	[-125.48%, -55.95%]
Posterior tail-area probability p: 0.0		
Posterior prob. of a causal effect: 100.0%		

highlighting a negative impact on population distribution. In relative terms, there was an approximately 90.44% decrease in population distribution compared to what was expected, illustrating the severe impact of the disaster. The posterior tail-area probability is 0.0, and the probability of a causal effect is 100.0%, strongly suggesting that the observed changes were not due to random fluctuations but were indeed caused by the snow disaster. Overall, our analysis conclusively shows that the snow disaster had a significant and negative effect on population distribution.

Through Pearson correlation analysis and causal inference, we determined that snow depth is a causal factor affecting population change rates. We found that the snow depth increased sharply between February 3 and February 8. Therefore, for Fig. 6.4, we can determine the optimal points to differentiate between the end of the disaster and the beginning of recovery in two scenarios:

1. If the low point of the population change rate curve occurs after February 8, the first minimum point after February 8 should be selected as the optimal cut-off point to mark the end of the snow disaster and the beginning of recovery.

2. If the curve reaches its lowest point before February 8, that point should be chosen as the optimal cut-off point.

Similarly, this approach can be applied to the other 3,736 grids.

6.2.2 Resilience of Different Land Use Types

After calculating the resilience of all grids, we obtained box plots of the vulnerability and robustness for different land use types, as shown in Fig. 6.7 and Fig. 6.8. Additionally, we calculated the resilience parameters for each land use type, as presented in Table 6.3.

Combining these results, we found significant differences in the resilience of various land use types, which may reflect characteristics related to the ecological environment, usage intensity, and other factors. By analyzing these resilience indicators, we can gain a deeper understanding of the performance of different land use types in disaster response, providing a scientific basis for urban planning and disaster management.

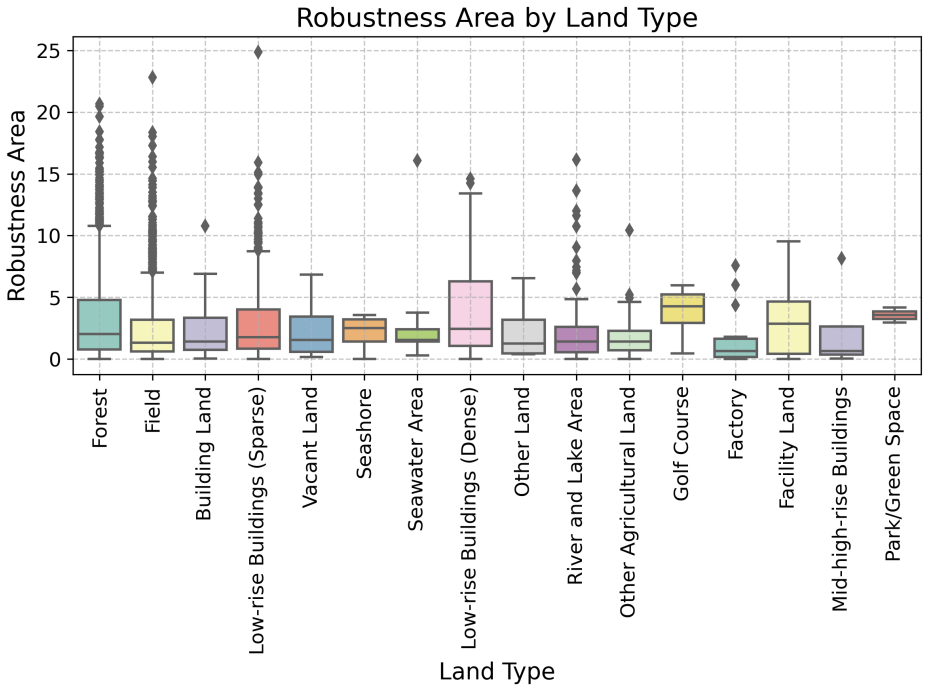


Figure 6.7: Robustness Distribution of Different Land Use Types

Forest: The robustness area of the forest (3.283256) and the vulnerability area (4.424078) are both large, indicating that the forest has weak resistance to disturbances and poor recovery ability. This may be due to the vast area and complex ecosystem of forests, which require a long time to recover after disturbances. The comprehensive RP (7.707334) further verifies this point. Fields: The robustness area (2.452680) and vulnerability area (3.766094) of fields are relatively moderate, showing good resistance to disturbances and recovery ability. The comprehensive RP (6.218774) of fields indicates that they have a certain resilience in the face of disasters and can quickly return to normal. Construction Land: The robustness area (2.298819) and vulnerability area (3.640278) of construction land are moderate, showing good resistance to disturbances and recovery ability. The comprehensive RP (5.939097) indicates that construction land has strong recovery ability in the face of disturbances and can quickly return to normal. Dense Low-rise Buildings and Sparse Low-rise Buildings: The robustness area (3.979128) and vulnerability area (4.018357) of dense low-rise buildings are large, indicating weak resistance and recovery ability. The robustness area (2.877072) and vulnerability area (4.105715) of sparse low-rise buildings are also large, showing similar weak resilience. The comprehensive RP of dense low-rise buildings and sparse low-rise buildings

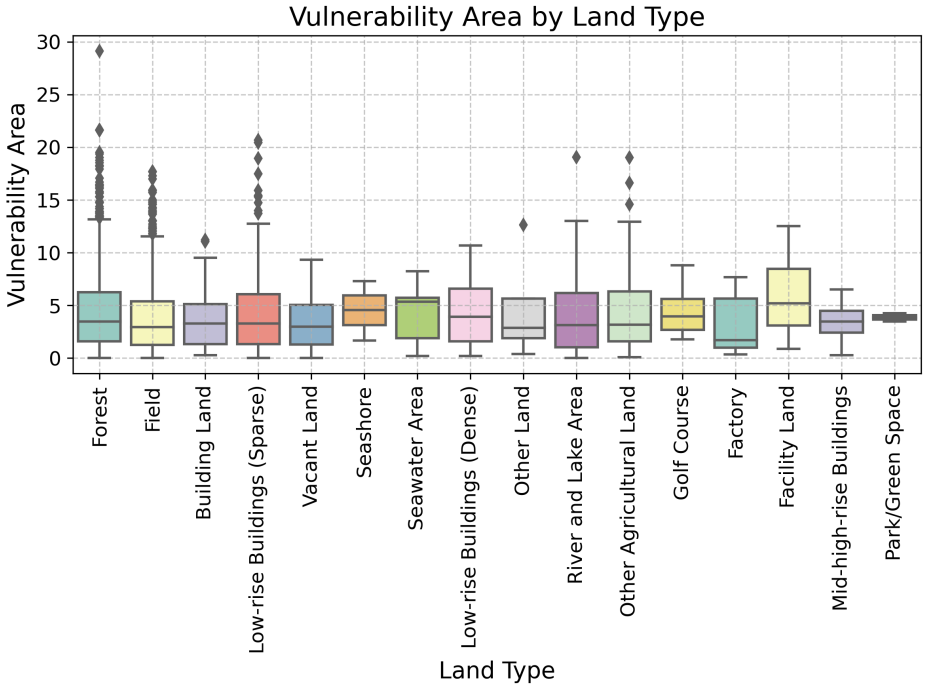


Figure 6.8: Vulnerability Distribution of Different Land Use Types

are (7.997485) and (6.982787), respectively, further indicating their vulnerability in the face of disasters. Mid-high-rise Buildings: The robustness area (2.362302) and vulnerability area (3.424012) of mid-high-rise buildings are small, with an RP of only 5.78631, indicating strong recovery ability in the face of disturbances and the ability to quickly return to normal. This may suggest that this building type has certain advantages in the face of disasters, possibly related to building structure and community mutual aid mechanisms. Parks/Green Spaces: The robustness area (3.550616) and vulnerability area (3.852316) of parks/green spaces are not significantly different. The comprehensive RP (7.402932) indicates average resistance and recovery ability. Golf courses show the weakest overall resilience in snow disasters, possibly because people prioritize material conditions in the face of disasters. Facility Land: The robustness area (2.930247) of facility land is small, but the vulnerability area (5.777300) is large, indicating poor resistance to disturbances but good recovery ability. Coast and River/Lake Areas: The robustness area (2.138968) and vulnerability area (4.511653) of the coast, and the robustness area (2.279636) and vulnerability area (4.023320) of river/lake areas show weak resistance and strong recovery ability.

Overall, factory, agricultural land, construction land, and mid-high-rise buildings show good comprehensive resilience, with strong resistance to disturbances and recovery ability. In contrast, low-rise buildings, parks, golf courses, and facility land have poor resilience and long recovery times. Forests, fields, sparse low-rise buildings, and river/lake areas have many outliers and need further analysis.

Factory, agricultural land, construction land, and mid-high-rise buildings demonstrated

Table 6.3: Resilience Metrics of Different Land Use Types

Land Type	Robustness	Vulnerability	RP
Vacant Land	2.144	3.277	5.422
Factory	1.683	2.997	4.681
Mid-high-rise Buildings	2.362	3.424	5.786
Building Land	2.299	3.640	5.939
Field	2.453	3.766	6.219
Other Agricultural Land	1.778	4.431	6.208
River and Lake Area	2.280	4.023	6.303
Low-rise Buildings (Sparse)	2.877	4.106	6.983
Forest	3.283	4.424	7.707
Low-rise Buildings (Dense)	3.979	4.018	7.997
Other Land	2.368	4.680	7.048
Seashore	2.139	4.512	6.651
Facility Land	2.930	5.777	8.708
Golf Course	3.833	4.477	8.311
Seawater Area	3.384	4.207	7.591
Park/Green Space	3.551	3.852	7.403

stronger resilience during disasters. This is primarily because these areas often serve as essential activity sites, such as workplaces or residential areas, which people prioritize maintaining during disasters. In contrast, recreational land uses, such as parks, golf courses, and facility grounds, exhibited weaker resilience due to their non-essential nature, leading to longer recovery times. This observation aligns with the findings of Santiago et al. [99], which highlighted that essential activities are less impacted by disasters compared to leisure activities. Additionally, mid-to-high-rise buildings showed greater resilience than low-rise buildings, likely because they are predominantly located in central urban areas with better-developed transportation infrastructure, further enhancing their recovery capacity [127].

6.2.3 Geospatial analysis of Risk

Based on the above research, we have drawn a comprehensive resilience map of Fukui Prefecture. In this study, the integral area of RP was calculated; hence, the larger the RP, the poorer the resilience, indicating higher risk. We also compared these results with the sections of roads where traffic congestion occurred, as shown in Fig. 6.9. The results show that the grids around the congested sections of National Route 8 are mainly in the third resilience zone [6.292-7.097] and the fourth resilience zone [5.487-6.291], with a small portion in the poorest

resilience zone [7.903-8.708]. Although it cannot be conclusively stated that poor resilience always leads to traffic congestion, areas with poor resilience are more likely to experience traffic congestion. Therefore, we can roughly estimate which cities and road sections have weaker resistance to snow disasters and are more prone to traffic congestion.

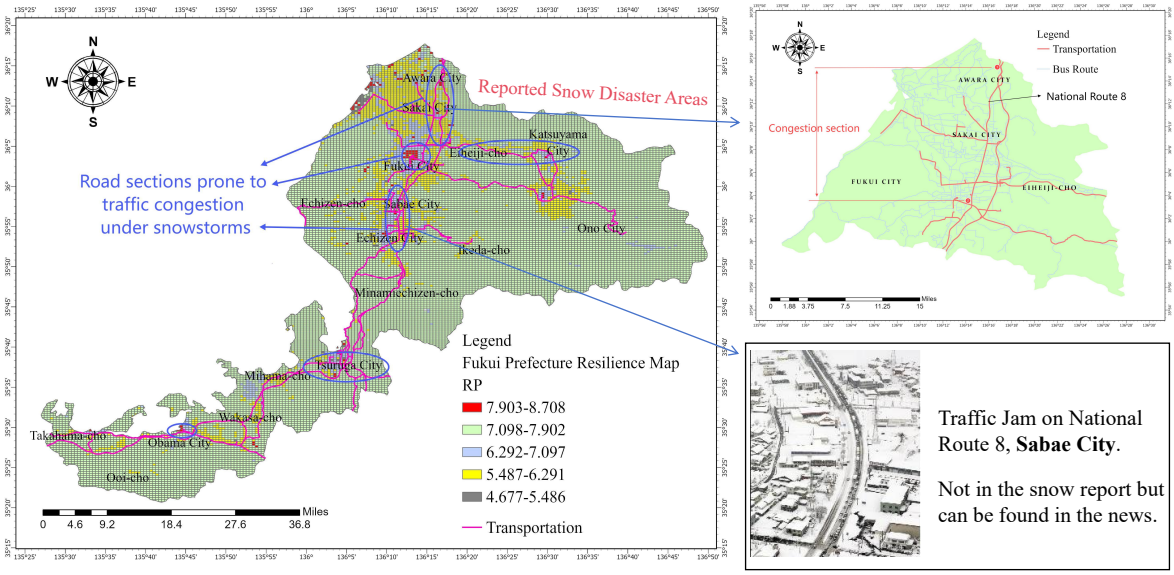


Figure 6.9: High-risk Traffic Jams in Fukui Prefecture [4, 1]

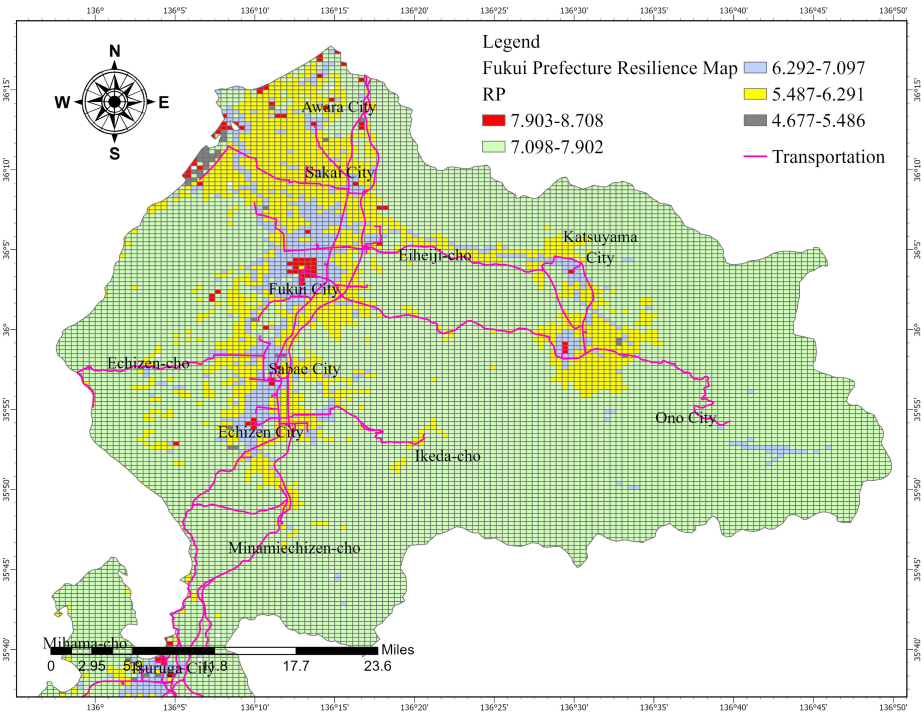


Figure 6.10: Resilience Map of Northern Fukui Prefecture

Specifically, the areas with poorer traffic resilience in Fukui Prefecture are mainly concentrated in the northern part, especially the vertical sections in Fukui City, Awara City,

and Sakai City, as well as the vertical sections from Sabae City to Echizen City and the horizontal sections from Eiheiji Town to Katsuyama City. In the southern part of Fukui Prefecture, the areas with poorer resilience are mainly the local sections in Obama City and Tsuruga City. This is consistent with the findings of Yang et al. [24], who identified that the 2018 Fukui snowstorm primarily affected the northern region of Fukui Prefecture. To enhance the readability of the resilience map, we have provided a magnified view of the northern part of Fukui Prefecture, as illustrated in Fig. 6.10.

6.3 Limitations and Discussions

In this study, Fukui Prefecture was divided into 500-meter grid cells, resulting in a total of 16,815 grids. However, after integrating GPS and land use data, only 3,737 grids contained complete data records. This data incompleteness may introduce certain biases in the resilience assessment. Due to budgetary constraints, it is currently challenging to obtain GPS data with more time-point records to address this issue.

The analysis in this study focused on the 2018 Fukui snowstorm as a case study. However, since 2018, no significant snowstorms have occurred in Fukui Prefecture, and financial limitations have hindered the acquisition of data for additional years. Future studies could integrate machine learning and SHAP theory to explore the underlying mechanisms of snowstorms. Additionally, the predictive risk maps generated from these approaches could be compared with actual road congestion data to validate the model's effectiveness. These findings can also be cross-referenced with the resilience maps generated in this study, further enhancing the scientific rigor and robustness of the conclusions.

In addition to data and case limitations, another practical factor that may influence the observed spatial patterns of resilience is the ease or difficulty of snow removal, which differs across land use types. Urban areas typically face challenges such as limited open space, high building density, and narrow roads, making timely snow clearance more difficult and recovery slower. In contrast, rural or industrial areas often have wider roads and more available space for snow disposal, enabling quicker mobility restoration. These operational differences may partly explain why regions with similar physical conditions (e.g., elevation, slope) display divergent resilience outcomes in the SHAP-based model. They may also account for outlier grids observed in certain land use categories. Acknowledging this logistical dimension helps enhance the interpretation of the model results and points to the importance of incorporating real-world maintenance capacity in future resilience assessments.

Moreover, the proposed model demonstrates high adaptability and is not limited to snowstorm scenarios. It can also be extended to other disaster types, such as floods, hurricanes, and tsunamis. This adaptability is attributed to the model's core methodology, which relies

on GPS data processing and resilience index computation. Researchers with access to GPS and land use data related to other disasters can apply this study's methodology to construct resilience maps. Furthermore, comparing the generated resilience maps with regions most severely affected by actual disasters can help validate the model's applicability and reliability.

6.4 Conclusion

This study utilized the resilience triangle method, Pearson correlation analysis, and causal inference to evaluate the regional resilience of different land use types during snow disasters, and explored the cities and road sections prone to traffic congestion under such conditions. The key findings are as follows:

Among various land use types, factories, agricultural land, construction land, and mid-high-rise buildings exhibit better resilience, recovering more quickly to pre-disaster population patterns. In contrast, low-rise buildings, parks, and facility land show poorer resilience, with longer recovery times. Notably, mid-high-rise buildings have smaller robustness and vulnerability areas, indicating strong recovery ability in the face of disturbances. This suggests that this building type may have certain advantages during disasters, potentially related to building structure and community mutual aid mechanisms. In urban planning, prioritizing such resilient building types could be beneficial.

Among weather factors, snow depth, snowfall, and precipitation show a positive correlation with changes in population distribution. Specifically, snow depth is not only highly correlated with population distribution changes but also serves as a causal factor influencing grid resilience.

Areas with poor traffic resilience in Fukui Prefecture are mainly concentrated in the northern part, particularly the vertical sections in Fukui City, Awara City, and Sakai City, as well as the vertical sections from Sabae City to Echizen City and the horizontal sections from Eiheiji-cho to Katsuyama City. In the southern part of Fukui Prefecture, areas with poor resilience are mainly localized sections in Obama City and Tsuruga City. During snow disasters, the government should focus on these cities and road sections, allocating more manpower and resources to help local residents combat the adverse effects of snow disasters.

This study analyzed and identified snow disaster resilience across different land types at the grid level, providing valuable insights into population dynamics and resilience factors during snow disasters. Furthermore, spatial risk analysis in Fukui Prefecture, Japan, was conducted to quantify and pinpoint high-risk and vulnerable areas, offering scientific support for local governments to formulate more targeted snow disaster mitigation policies. However, due to inconsistencies in data availability, only 3,737 out of the total 16,815 500-meter grids in Fukui Prefecture were included in the analysis. Addressing data-deficient areas through

improved data collection and integration to assess their resilience remains a crucial direction for future research.

Chapter 7

Machine Learning-Based Identification and Assessment of Snow Disaster Risks

In Chapter 4, the spatiotemporal variations of intercity transportation under snowstorm conditions were analyzed at the city level. Chapter 5 focused on grid-level analysis, identifying critical congestion points along National Route 8 and the Hokuriku Expressway, and developed a traffic congestion identification model based on the Random Forest algorithm. Chapter 6 further assessed the recovery capabilities of different land use types at the grid level using traditional resilience curve methods, providing in-depth insights into spatial resilience characteristics under snow disaster conditions.

However, the previous analyses have certain limitations. First, although machine learning models demonstrate strong predictive capabilities, they often lack interpretability due to their black-box nature. Second, traditional statistical methods are limited in capturing the complex nonlinear relationships between environmental factors and snow disaster risks.

To address these challenges, this chapter integrates multiple machine learning approaches with integrates multi-source data to develop an interpretable model for quantifying high-risk areas and examining key nonlinear relationships and threshold effects influencing snowstorm impact occurrence, offering actionable insights for mitigation strategies.

7.1 Methodology

7.1.1 Methodology framework

This study integrates multi-source data, including mobile GPS data, DEM data, road data, urban data, NDVI data, DNB data and traffic congestion data. We employed six machine learning models—Decision Tree, Random Forest, Support Vector Machine (SVM), Light Gradient Boosting Machine (LightGBM), Multilayer Perceptron (MLP), and Extreme

Gradient Boosting (XGBoost)—to capture complex nonlinear relationships among influencing factors and applied SHAP (SHapley Additive exPlanations) theory to interpret variable contributions. The overall workflow of the research is divided into three steps, as shown in Fig. 7.1: (1) Data collection and variable computation; (2) Model building and cross-validation; (3) Non-linear association modeling and result interpretation using SHAP values. (4) Geospatial analysis for snow disaster risk.

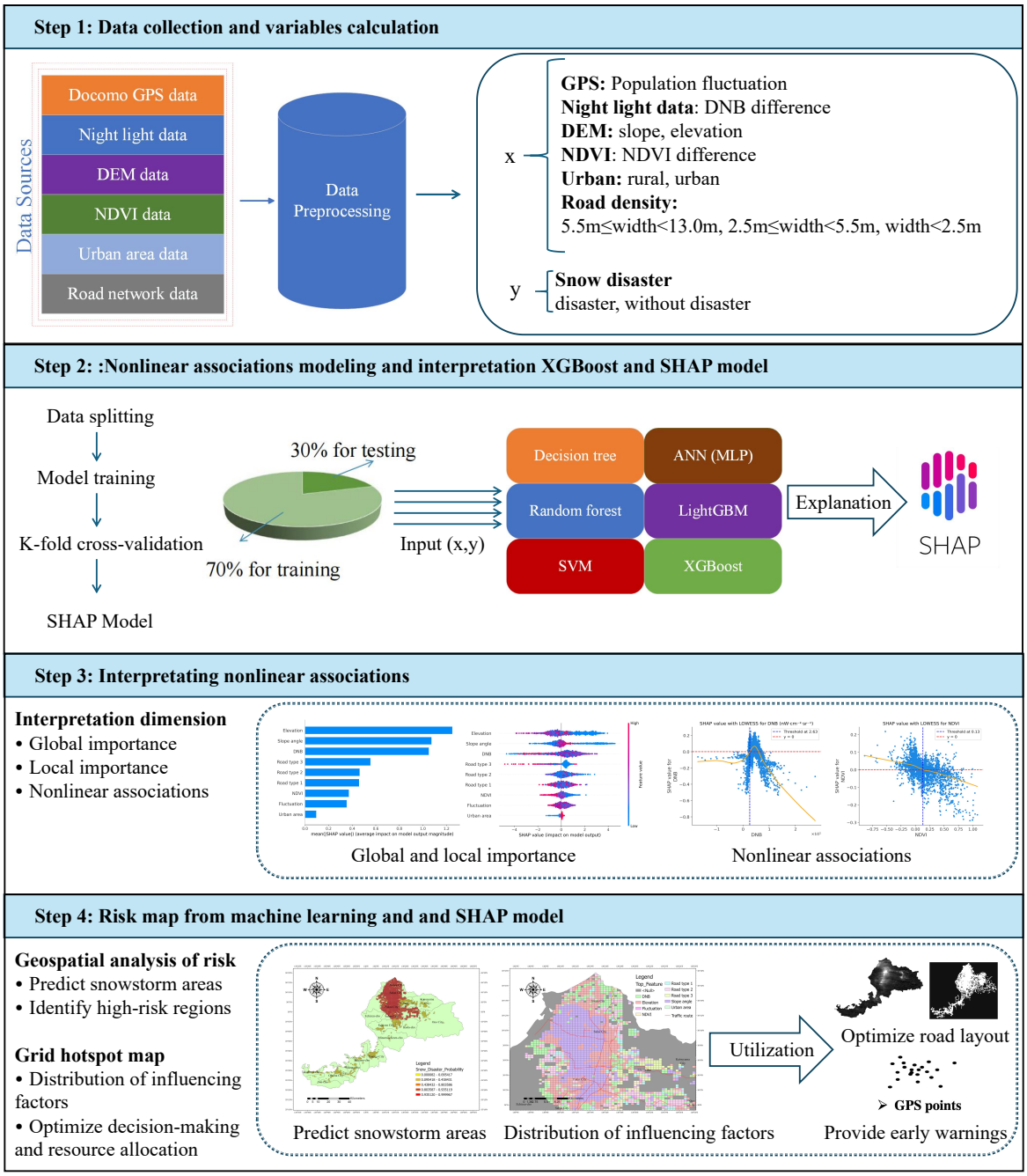


Figure 7.1: Methodology framework

The target variable y was defined based on the 2018 government report [1] and the study by Yang et al. [24], in which cities experiencing significant traffic congestion during

snowstorms were labeled as high-risk areas. The dataset, comprising samples from all 17 cities in Fukui Prefecture, was randomly split into 70% for training and 30% for testing. Although spatial partitioning was not explicitly applied, the random sampling ensured that data from all cities were represented in both training and testing sets, thereby reducing potential spatial bias.

To analyze snow disaster risk and its driving factors, we applied the optimal machine learning model selected from the six candidate models. Using SHAP (SHapley Additive exPlanations) theory, we calculated the contribution of each feature within every 500-meter grid cell. Based on the SHAP values, we determined the most influential feature for each grid and assigned it a distinct color for visualization. This process allowed us to generate a feature impact distribution map, highlighting the dominant factors driving snow disaster risks across the study area.

For the risk prediction map, the selected machine learning model was employed to estimate the probability of snow disaster occurrence for each grid cell. These probabilities were then visualized to create a probabilistic risk map of snow disasters. Both maps were generated using geospatial visualization techniques to ensure intuitive interpretation, providing a comprehensive understanding of spatial variability in snow disaster risks and their key contributing factors.

7.1.2 Data preprocessing and feature extraction

For details on data preprocessing, please refer to Chapter 3.

In this chapter, the modeling was conducted at a 500-meter grid level, which posed several constraints on variable compatibility. Notably, certain variables used in earlier chapters, such as snowfall data and detailed land use types, were excluded from the final model for the following reasons: First, snowfall data were only available at the city level, whereas other core predictors such as population fluctuation and road density were constructed at the grid level. Incorporating spatially coarse snowfall information into a high-resolution model could have introduced aggregation errors and undermined the model's predictive accuracy. Therefore, snowfall was excluded from this stage of the analysis. Second, the feature importance analysis presented in Chapter 5 revealed that most land use categories exhibited low predictive contributions, with the exception of "field" and "forest" types. To streamline the model and enhance interpretability, land use variables were also omitted in this chapter.

This decision-making process aims to maintain spatial consistency across variables and ensure model robustness. A detailed summary of variable selection criteria has been provided to help readers understand the rationale behind variable inclusion or exclusion across different stages of the research.

The study area in Fukui Prefecture was initially divided into a total of 16,815 grid cells,

each 500 meters in size. However, after integrating and cleaning all data and standardizing them to a 500-meter grid cell size, the number of available grid cells was reduced to 3480. Finally, we combined all the datasets and applied z-score normalization to the variables Elevation, Slope angle, Road type 1, Road type 2, Road type 3, NDVI, DNB and Fluctuation. The z-score normalization method was applied to standardize data, transforming each variable to have a mean of zero and a standard deviation of one. Consequently, values below the mean are represented as negative, while values above the mean are positive. The input data for the machine learning models is shown in Table 7.1:

Table 7.1: Sample of Data Input for Machine Learning Models

Variable Name	**Original Value**	**Description**	**Variable Type**
Road type 1	0.0	Density of roads with width 5.5 m $\leq$ width $<$ 13.0 m	Feature
Road type 2	247.5	Density of roads with width 3.0 m $\leq$ width $<$ 5.5 m	Feature
Road type 3	0.0	Density of roads with width $<$ 3.0 m	Feature
Elevation	96.4	Average elevation within the grid cell	Feature
Slope angle	8.0	Average slope angle within the grid cell	Feature
Fluctuation	-0.1076	Population fluctuation before and during snowstorm	Feature
Urban area	1	Indicator: 1 for urban, 0 for rural	Feature
NDVI	0.5748	Average Normalized Difference Vegetation Index	Feature
DNB	5.23	Average night light intensity (DNB)	Feature
Snow	0	Manually labeled snow disaster occurrence (1 for disaster, 0 for no disaster)	Target

7.1.3 Models for analyzing associations in snow disaster risk

Following data processing, we employed six machine learning models to analyze snow disaster risk: Decision Tree, Random Forest, Support Vector Machine (SVM), Light Gradient Boosting Machine (LightGBM), Multilayer Perceptron (MLP), and Extreme Gradient Boosting (XGBoost). The Decision Tree served as an interpretable baseline to reveal fundamental feature interactions; Random Forest enhanced prediction robustness by aggregating multiple decision trees; SVM identified optimal hyperplanes to separate classes, especially in high-dimensional spaces; LightGBM improved learning efficiency and accuracy using histogram-based algorithms and leaf-wise tree growth; MLP, as a neural network, captured

complex nonlinear patterns; and XGBoost optimized predictive accuracy through gradient boosting and regularization. These models provided complementary perspectives on snow disaster risk factors, ensuring both robustness and interpretability in the analysis, and supporting the selection of the best-performing model.

Multilayer Perceptron (MLP)

MLP is a feedforward neural network with an input layer, one or more hidden layers, and an output layer. By using hidden layers and activation functions, MLP captures non-linear relationships and increases network expressiveness. Common activation functions include Sigmoid, Tanh, and ReLU, while Softmax is often used for multi-class classification. Key hyperparameters include the number of hidden layers and their sizes.

The MLP utilizes activation functions to introduce non-linearity into the model. One common activation function for hidden layers is ReLU, defined as [128]:

$$f(x) = \max(0, x) \tag{7.1}$$

For multi-class classification, MLP often uses the Softmax function at the output layer to convert logits into probabilities:

$$P(y = j|x) = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \tag{7.2}$$

where z_j is the output score for class j, and K is the total number of classes. This transformation ensures that the output can be interpreted as a probability distribution over all possible classes.

Decision Tree

A decision tree is a model that learns from sample distribution patterns to construct a tree-like structure, representing the mapping between sample features and their classifications. Each node in a decision tree (except the leaf nodes) evaluates a specific feature, essentially posing a ”yes or no” question. Each branch represents an answer to this question, leading to subsequent nodes until a leaf node is reached. The final leaf node indicates the predicted category, assigning the sample to its most likely class after progressing through multiple feature-based decisions.

In constructing a decision tree, information gain is commonly used to determine the optimal attribute for splitting nodes. The information gain, $IG(D, a)$, is calculated as [129]:

$$IG(D, a) = H(D) - \sum_{v \in V} \frac{|D_v|}{|D|} H(D_v) \tag{7.3}$$

where $H(D)$ represents the entropy of the dataset D, and D_v is the subset of data for each value v of the attribute a. This formula calculates the reduction in entropy when splitting on attribute a, guiding the model in selecting the feature that best divides the data.

Random Forest

This method has been described in detail in Chapter 5 and is therefore not elaborated here.

Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised learning algorithm widely used for classification tasks. It aims to find an optimal hyperplane that maximizes the margin between two classes. SVM performs well in high-dimensional spaces and is effective when the number of features exceeds the number of samples [130].

For binary classification, SVM constructs a decision boundary defined by the hyperplane [131]:

$$w \cdot x + b = 0 \tag{7.4}$$

where w is the weight vector, x is the feature vector, and b is the bias. The decision rule is:

$$\hat{y} = \text{sign}(w \cdot x + b) \tag{7.5}$$

SVM can also incorporate non-linear decision boundaries using kernel functions, such as the Radial Basis Function (RBF), to map input data into higher-dimensional spaces where linear separation is feasible.

LightGBM (Light Gradient Boosting Machine)

LightGBM is an efficient gradient boosting framework that enhances training speed and prediction performance by optimizing decision tree growth. Compared to traditional GBDT methods, LightGBM demonstrates significant advantages in handling large-scale, high-dimensional, and sparse datasets [132]. The core algorithm of LightGBM is based on histogram techniques (which construct feature splits using discretized bins) and a leaf-wise growth strategy. Unlike XGBoost, which grows trees level-wise, LightGBM adopts a leaf-wise strategy that expands the leaf with the maximum loss reduction at each step to improve accuracy [133].

Its basic prediction function is defined as:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in \mathcal{F} \tag{7.6}$$

where f_k denotes the k-th decision tree, and $\mathcal{F}$ is the set of all weak learners (trees). LightGBM introduces two key techniques to improve efficiency and memory usage: GOSS (Gradient-based One-Side Sampling) and EFB (Exclusive Feature Bundling).

In this study, we enabled the `class_weight='balanced'` parameter to improve LightGBM's ability to identify snow disaster samples under class imbalance.

XGBoost

XGBoost is a boosting algorithm based on Gradient Boosting Decision Trees (GBDT). Unlike the parallel ensemble method of Random Forest, XGBoost uses a sequential approach, where each new tree is trained to minimize the residuals of previous trees. The final prediction is a weighted sum of outputs from multiple decision trees, combining weak learners to improve accuracy.

The prediction result of XGBoost, $\hat{y}_i$, is calculated as the weighted sum of outputs from all trees. Formally, it is defined as [134]:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in \mathcal{F} \tag{7.7}$$

where K is the number of trees, f_k represents the k-th tree, and $\mathcal{F}$ denotes the set of all possible trees. Each tree f_k contributes to reducing the residual errors from previous trees, making XGBoost a powerful tool for iterative error minimization and accuracy improvement.

7.1.4 SHAP for machine learning model explanation

In this study, SHAP (Shapley Additive exPlanations) is employed to interpret the contributions of different features to the model's predictions, helping to understand the factors influencing snow disaster risk. By providing insight into feature importance, SHAP enables a clearer understanding of how environmental and infrastructure variables impact the model's outputs, which aids in explaining the underlying patterns captured by the machine learning models used

SHAP is a game-theoretic approach for interpreting machine learning models [135]. It uses Shapley values to measure each feature's contribution to the model's prediction by assigning a portion of the model's output to each feature. The Shapley value, a weighted average of marginal contributions, reflects the impact of each feature on the current prediction.

The Shapley value for a feature i is calculated as follows [136]:

$$\varnothing_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n - |S| - 1)!}{n!} [f(S \cup \{i\}) - f(S)] \tag{7.8}$$

where N is the set of all n features, S is any subset excluding feature i, and $f(S \cup \{i\})$ and $f(S)$ denote model outputs with and without feature i, respectively. This calculation averages the marginal contributions of feature i across all subsets S, weighted by $\frac{|S|!(n-|S|-1)!}{n!}$ to account for subset size.

An interpretable model can be created using an additive feature attribution method, approximating the model by a linear function of input features:

$$g(z') = \varnothing_0 + \sum_{i=1}^{M} \varnothing_i z'_i \tag{7.9}$$

Here, $z' \in \{0, 1\}^M$ is a binary vector indicating if a feature is present ($z'_i = 1$) or absent ($z'_i = 0$). M is the number of features, $\varnothing_0$ is the base value (the expected output with no features), and $\varnothing_i$ is the Shapley value of feature i.

7.1.5 Risk map from machine learning prediction

To analyze snow disaster risk and its driving factors, we applied the optimal machine learning model selected from the six candidate models. Using SHAP (SHapley Additive exPlanations) theory, we calculated the contribution of each feature within every 500-meter grid cell. Based on the SHAP values, we determined the most influential feature for each grid cell and assigned it a distinct color for visualization. This process allowed us to generate a feature impact distribution map, highlighting the dominant factors driving snow disaster risks across the study area.

For the risk of snowstorm impact prediction map, the selected machine learning model was employed to estimate the probability of snow disaster occurrence for each grid cell. These probabilities were then visualized to create a probabilistic risk map of snow disasters. Both maps were generated using geospatial visualization techniques to ensure intuitive interpretation, providing a comprehensive understanding of spatial variability in snow disaster risks and their key contributing factors.

Based on the core prediction formula of XGBoost and SHAP theory [137, 138], the mathematical formulations for snow disaster risk prediction and feature impact analysis suitable for this study are as follows:

The snow disaster risk prediction probability can be mathematically expressed as:

$$P(\text{snow} = 1|\mathbf{X}) = \sum_{k=1}^{K} T_k(\mathbf{X}) \tag{7.10}$$

where $T_k(\mathbf{X})$ representing the k-th decision tree in the XGBoost ensemble. Each decision tree contributes incrementally to the overall prediction, and their collective output forms the basis for classifying snow disaster risks.

In order to map the spatial distribution of snow disaster influencing factors in Fukui Prefecture, it is necessary to identify the most important feature within each grid cell. The top feature can be determined using feature importance metrics based on SHAP values:

$$F_i = \arg\max_j |\phi_{ij}| \tag{7.11}$$

where $|\phi_{ij}|$ is the absolute SHAP value of feature j for instance i.

For comparison purposes, normalized SHAP probabilities can be utilized to provide a standardized measure of feature importance across grid cells:

$$P_{ij} = \frac{|\phi_{ij}|}{\sum_{k=1}^{M} |\phi_{ik}|} \tag{7.12}$$

The dominant feature is identified as:

$$F_i = \arg\max_j P_{ij} \tag{7.13}$$

where F_i represents the feature with the highest importance or normalized probability for a given instance i. The normalized SHAP probability, denoted as P_{ij}, is defined as the proportion of the total SHAP value contribution attributable to feature j for instance i. The total number of features in the dataset is represented by M. The absolute SHAP value, $|\phi_{ij}|$, indicates the magnitude of the contribution made by feature j to the model's prediction for instance i. These equations make it possible to calculate the feature with the highest probability within each grid cell based on SHAP.

These formulations form the basis for generating the probabilistic risk map and feature impact distribution map.

7.2 Result and discussion

7.2.1 Model results of snow disaster prediction

This study developed six models: Decision Tree, Random Forest, SVM, LightGBM, MLP, and XGBoost. Each model splits the dataset into 30% test and 70% training sets, and 5-fold cross-validation was performed. The Accuracy, Precision, Recall, and F1-score of the six methods are shown in Table 7.2, indicating that the XGBoost model has higher accuracy than the other models. We further analyzed the confusion matrix and ROC curve of the XGBoost model to verify its classification performance, as shown in Fig. 7.2.

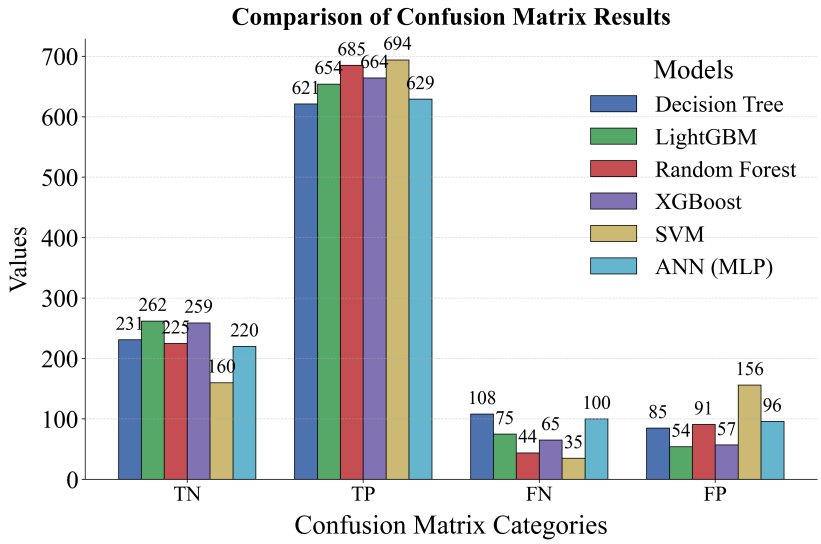
Table 7.2: Comparison of Model Performance

Model	Accuracy	Precision	Recall	F1-score
XGBoost	0.8833	0.9209	0.9108	0.9159
LightGBM	0.8766	0.9237	0.8971	0.9102
Random Forest	0.8708	0.8827	0.9396	0.9103
ANN (MLP)	0.8067	0.8624	0.8601	0.8613
Decision Tree	0.8153	0.8796	0.8519	0.8655
SVM	0.7569	0.8800	0.7545	0.8124

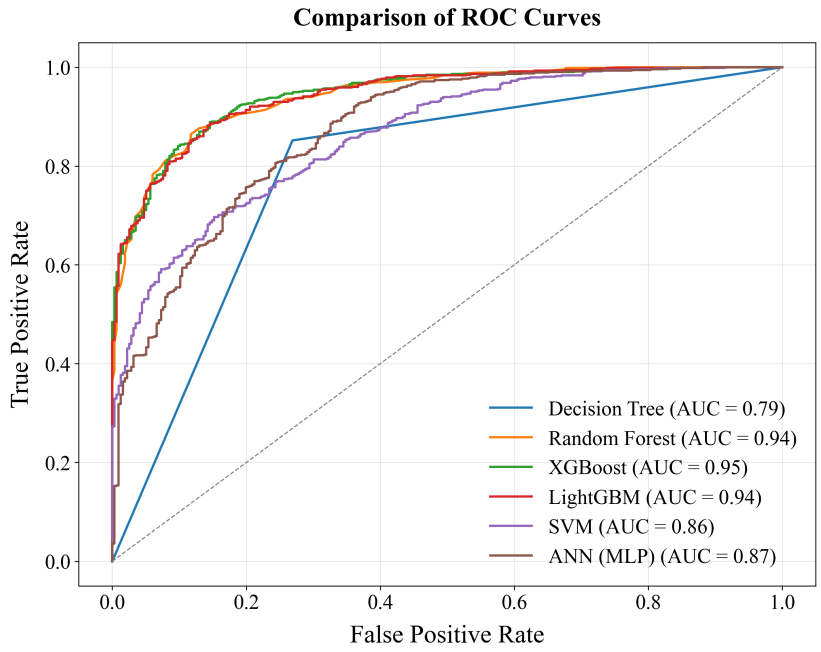
7.2.2 Global and local importance

The global variable importance (RI), measured by SHAP values, represents the relative contribution of each variable to the model output and is unitless. As shown in Fig. 7.3(a), terrain-related variables—particularly elevation and slope angle—exhibit the highest importance, underscoring their critical role in snow disaster prediction due to their influence on snow accumulation and road safety. The DNB ranks as the third most important variable, indicating that areas with intense human activity or developed urban infrastructure are more likely to be affected during snowstorms. Road-related variables (Road type 1, Road type 2, and Road type 3) also show substantial importance, highlighting the vulnerability of different road types under snow conditions. The NDVI ranks in the middle range, which may be attributed to its 16-day revisit cycle, limiting its ability to reflect real-time surface conditions during snow events. In contrast, the global importance of fluctuation and urban area is relatively low, suggesting their influence on snow disaster risk is more indirect or context-dependent within the current model.

The local interpretation plot (Fig. 7.3(b)) shows the SHAP values of each variable and their impact direction on the prediction. Each point represents a sample, with red



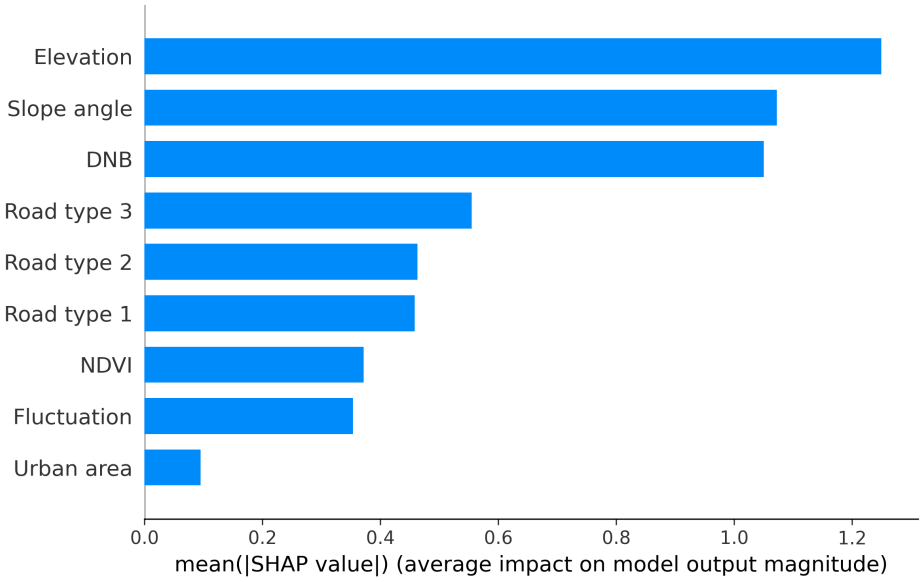
(a) Confusion Matrix



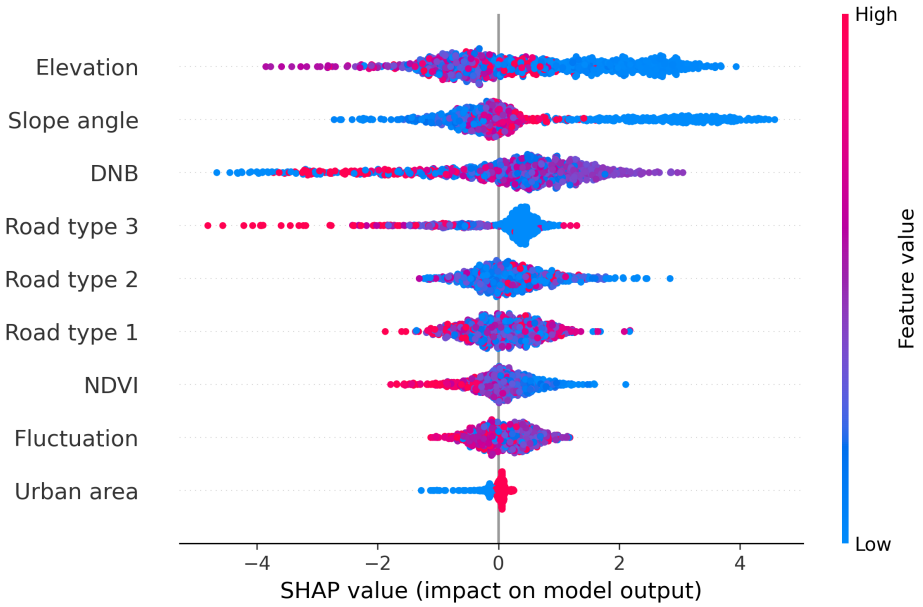
(b) ROC Curves

Figure 7.2: Results of Confusion Matrix and ROC Curve of Six Classification Methods

points indicating high feature values and blue points indicating low feature values. The x-axis represents the SHAP value: a positive value means the feature increases the model's prediction, while a negative value decreases it. For instance, for Elevation, red points are mainly concentrated on the negative side of the SHAP value. This suggests that in areas with higher elevation (i.e., red points), the SHAP values are usually negative, possibly because higher elevation regions tend to have lower traffic flow, and thus, less impact from snow disasters. On the other hand, in lower elevation areas (blue points), SHAP values are distributed across both the positive and negative axes, but with more concentration on the



(a) Global Variable Importance



(b) Local Variable Importance

Figure 7.3: Variable Importance of Model Variables

positive side. This indicates that in these lower elevation areas, the model tends to predict higher values, likely due to higher population density and greater traffic flow, leading to more severe snow disaster impacts.

For Slope angle, the negative axis portion might be similar to Elevation, reflecting the lower traffic volume in mountainous areas. However, there is a significant concentration of red points on the positive axis, revealing that some specific areas with larger slopes or roads experience intensified traffic problems during snow disasters, leading to a significant influence on the model output. For Road Type 3, lower road density primarily contributes positively to snow disaster prediction, while higher road density exhibits a negative contribution, reflecting its limited overall role in traffic during snow disasters. This may be attributed to its narrower width and lower utilization rates. Road Type 2 demonstrates a predominantly positive contribution to snow disaster prediction, as shown by the majority of points distributed on the positive SHAP value axis. High feature values (red points) are densely concentrated in this region, indicating that areas with longer road extensions of this type are more susceptible to snow disaster impacts. However, the presence of red points on the negative axis suggests that certain high-feature-value regions are associated with non-snow-disaster conditions, possibly due to favorable factors such as flat terrain or efficient snow-clearing capabilities that mitigate snow disaster risks. Low feature values (blue points) exhibit weaker but consistent positive contributions, emphasizing that even shorter extensions of this road type can influence snow disaster occurrence. Road Type 1 (Density of roads with width $5.5\,\mathrm{m} \leq \text{width} < 13.0\,\mathrm{m}$) shows a mixed distribution of blue (low feature value) and red (high feature value) points without a clear positive or negative impact, suggesting that the influence of Road Type 1 may depend on interactions with other features such as terrain, elevation, or the density of other road types.

High feature values of DNB (represented by red points) are primarily distributed on the positive SHAP axis, indicating that areas with stronger nighttime light intensity—typically associated with urbanization or heavy traffic—are more susceptible to snowstorm impacts. In contrast, high feature values of NDVI are predominantly located on the negative axis, suggesting that regions with high vegetation reflectance experience less snowstorm-related disruption. This finding supports the hypothesis that NDVI can serve as an effective indicator for snow disaster risk monitoring. However, the limited temporal resolution of NDVI data (e.g., 16-day intervals) may have restricted its global importance in the model, positioning it only at a moderate rank in the overall feature importance.

Interestingly, while variables like Fluctuation have low global importance, they still show significant local influence in certain regions. The red points for urban areas are mostly concentrated on the positive axis, indicating that snow disasters are particularly evident in urban areas, which also supports the conclusion that snow disasters have a stronger impact in

lower elevation areas.

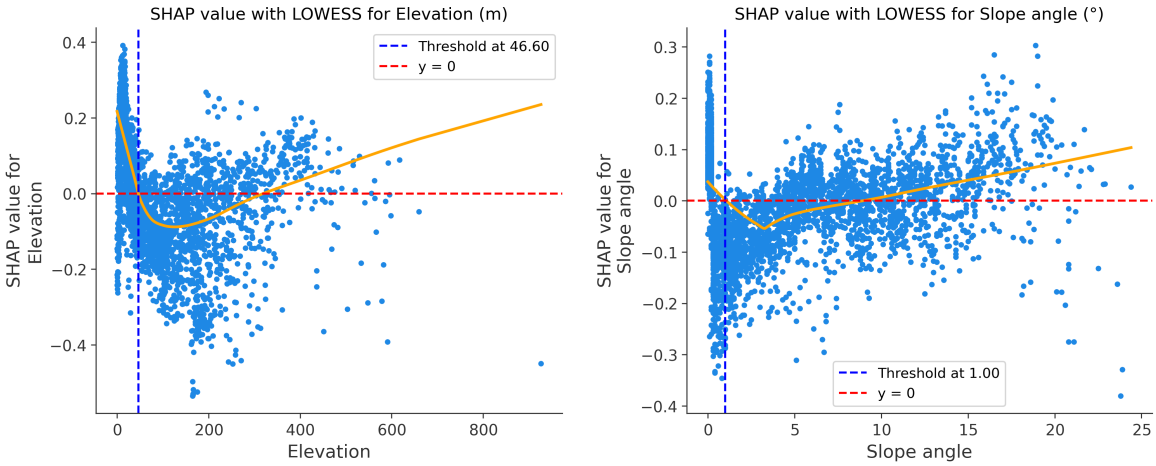
In summary, the model's predictions rely heavily on terrain features and road infrastructure, as well as human activity intensity. Elevation and slope angle are the most influential predictors, highlighting the importance of topographic conditions in shaping snow disaster risk. Among the road-related variables, Road Type 2 exhibits the highest SHAP value, suggesting that medium-width roads are more vulnerable to snow accumulation and traffic disruption. Road Type 1 and Road Type 3 also contribute notably to the model: the former may reflect the sensitivity of wide roads to increased traffic demand, while the latter, despite being narrower and less frequently used, still influences predictions in certain localized areas. In addition, DNB—a proxy for nighttime light intensity and human activity—emerges as a key factor, with high feature values consistently associated with elevated snowstorm risk. This indicates that urbanized or highly trafficked regions are more susceptible to snowstorm impacts. On the other hand, NDVI shows a moderate negative contribution, suggesting that areas with higher vegetation cover are generally less affected by snow-related disruptions. This aligns with the hypothesis that vegetation-rich regions may exhibit better snow absorption or fewer traffic-related vulnerabilities.

7.2.3 Interpreting Nonlinear Associations to Identify Key Snow Disaster Risk Factors

Utilizing Local Dependence Plots (LDPs), we can visualize the SHAP values for each variable in the LDPs [139]. Based on the LDPs, we explored the nonlinear patterns and threshold effects of various parameters.

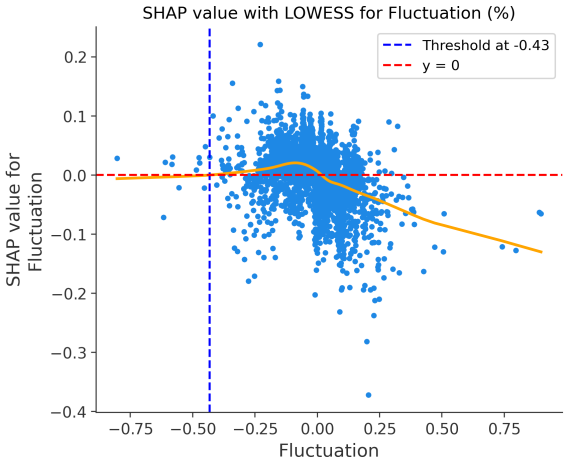
Elevation: As shown in Fig. 7.4(a), when the elevation is below approximately 46.6 meters (as indicated by the vertical threshold line), the SHAP values are predominantly positive. This suggests that low-elevation areas are more likely to experience severe snowstorm impacts. Such regions often correspond to densely populated urban areas with higher traffic volumes and building density, which can exacerbate the consequences of snow-related disruptions [140].

In the range of roughly 50 to 200 meters, the SHAP value tends to decrease and remains slightly negative. This implies a weakening influence of elevation on snowstorm impact in this interval. These mid-elevation zones are typically suburban or peri-urban areas with moderate development and infrastructure, thus experiencing less pronounced snowstorm effects. Above 400 meters, the SHAP values increase sharply and turn positive again, indicating that higher elevations are also associated with greater snowstorm risks. This may be due to lower ambient temperatures and higher snowfall accumulation at higher altitudes, combined with fragile infrastructure such as narrow mountain roads, which make transportation more vulnerable to

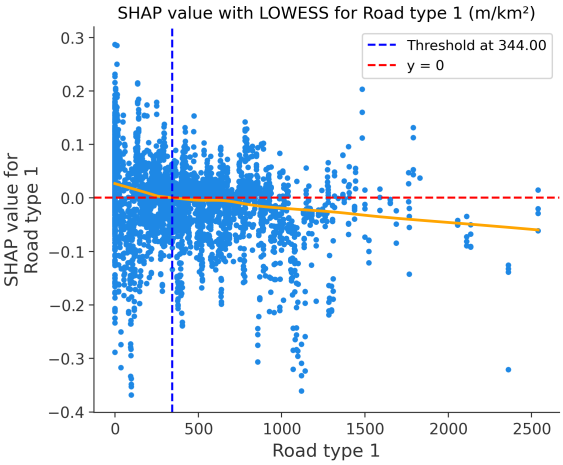


(a) Elevation SHAP Plot

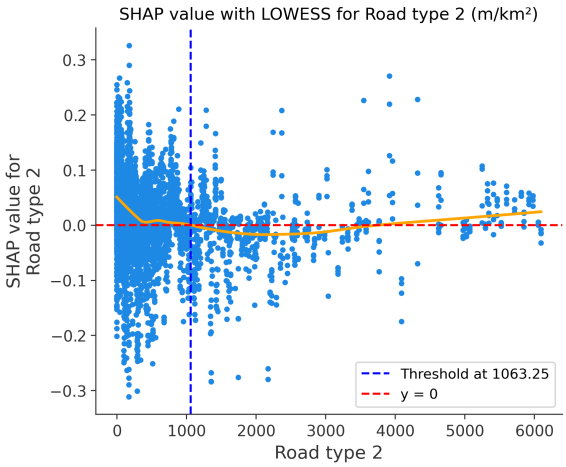
(b) Slope angle SHAP Plot



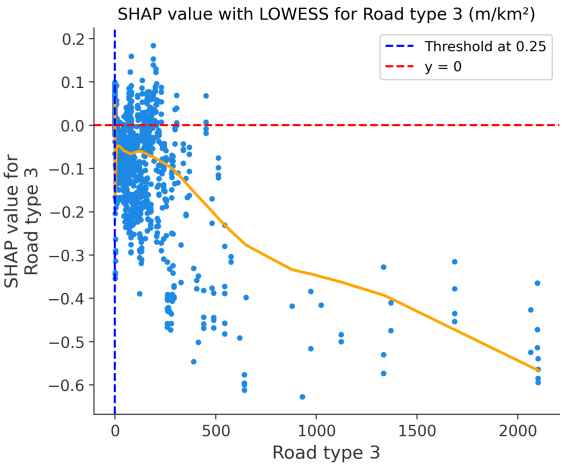
(c) Fluctuation SHAP Plot



(d) Road type 1 SHAP Plot

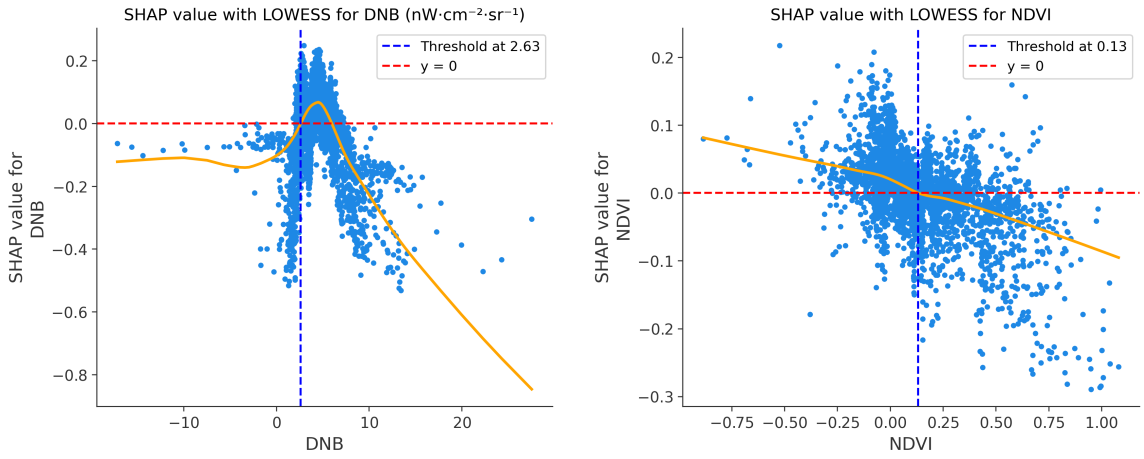


(e) Road type 2 SHAP Plot



(f) Road type 3 SHAP Plot

Figure 7.4: SHAP Value Plots for Different Variables



(a) Night Light (DNB) (b) NDVI

Figure 7.5: SHAP Value Plots for Remote Sensing Variables

snow disruption.

Slope Angle: As shown in Fig. 7.4(b), when the slope angle is below 1.0°, the SHAP value is close to zero or slightly positive, indicating minimal influence on snow disaster risk. These flat areas are generally safe and manageable during snow events. In the range between 1.0° and 9.9°, the SHAP value is consistently negative, suggesting that moderate slopes are associated with a reduced risk of snowstorm impacts. These areas may benefit from more efficient drainage or snow-clearing conditions, reducing the likelihood of disruptions.

When the slope angle exceeds 9.9°, the SHAP value increases sharply into the positive range, highlighting that steeper slopes significantly amplify snowstorm risk. This is likely due to the accumulation and sliding of snow on steep gradients, making roads more hazardous. Such terrain also presents challenges for snow removal, and increases the probability of vehicle skidding or accidents during icy conditions [141].

Population Fluctuation: In Fig. 7.4(c), the relationship between population fluctuation and snowstorm impact occurrence exhibits significant non-linearity. When the fluctuation value ranges from −0.25 to 0.05, the SHAP value gradually increases, peaking around −0.12. This indicates that lower levels of population fluctuation have a relatively positive contribution to snowstorm impact prediction, which aligns with common sense. Snowstorms often lead to reduced population mobility, as people tend to avoid unnecessary travel during severe weather. Although the SHAP threshold is located at −0.43, the values between −0.25 and 0 still exhibit negative contributions, implying that even moderate population reduction may indicate safer or less-affected areas. However, as the fluctuation value continues to rise beyond 0.25, the SHAP value decreases sharply. This suggests that areas with higher population fluctuation may have stronger resilience to snowstorms or are less affected by them.

Road Density: As shown in Fig. 7.4(d) to Fig. 7.4(f), the SHAP value analysis reveals

distinct patterns for different road types.

For **Road Type 1**, when the road density is less than 344.00 m/km^2, the SHAP values mostly fluctuate around zero, occasionally turning positive. This indicates that this variable has a minor impact on snowstorms, though it may occasionally increase the likelihood of snowstorm impact occurrence. However, when the road density exceeds 344.00 m/km^2, the SHAP values gradually decrease and become negative. This could suggest that as the density of wider roads increases, their higher traffic capacity reduces the likelihood of traffic congestion, thereby mitigating the impact of snowstorms on the transportation system.

For **Road Type 2**, when the road density is less than 1063.25 m/km^2, the SHAP values are mostly positive, indicating that these areas are more likely to be affected by snowstorms. This might be due to specific geographical or traffic characteristics of regions with lower road density, making them more susceptible to snowstorm impacts. However, when the density exceeds 1063.25 m/km^2, the SHAP values gradually turn negative. Similar to Road Type 1, this might indicate that as the density of wider roads increases, their higher traffic capacity reduces the likelihood of traffic congestion during snowstorms.

For **Road Type 3**, when the road density exceeds 0.25 m/km^2, the SHAP values are almost always negative, suggesting that regions dominated by Road Type 3 have a lower likelihood of snowstorm-related disruptions. This could be attributed to the fact that Road Type 3 represents the narrowest road types, often pedestrian paths or secondary roads, which have minimal impact on vehicular traffic. In practice, these areas are likely to have lower traffic volumes, making significant traffic problems during snowstorms unlikely.

NDVI: As shown in Fig. 7.5(b), NDVI exhibits a generally negative contribution to snowstorm impact prediction, especially when the NDVI value exceeds the threshold of 0.13. In this high-value region, SHAP values are clearly concentrated below the zero line, indicating that areas with high vegetation reflectance are less likely to be affected by snow disasters. This may be because regions with higher NDVI are often less urbanized and have lower traffic volumes, which reduces the likelihood of severe traffic disruptions during snowstorms. This observation supports the hypothesis that NDVI can serve as an effective indicator for snow disaster monitoring. However, due to the limited temporal resolution of NDVI (e.g., 16-day intervals), its global importance in the model remains moderate.

DNB: As illustrated in Fig. 7.5(a), DNB demonstrates a clear threshold behavior around 2.63 nW·cm^{-2}·sr^{-1}. When DNB values exceed this threshold, SHAP values rapidly increase and remain positive, indicating that areas with higher nighttime light intensity—typically representing urban regions or roads with substantial traffic—are more vulnerable to snowstorm impacts. In contrast, lower DNB values correspond to negative SHAP contributions, implying reduced risk in less developed or sparsely populated areas. This highlights the strong link between human activity intensity and snowstorm vulnerability, reinforcing DNB's importance

as a predictor of snowstorm-affected regions.

7.2.4 Interaction Analysis of Variables

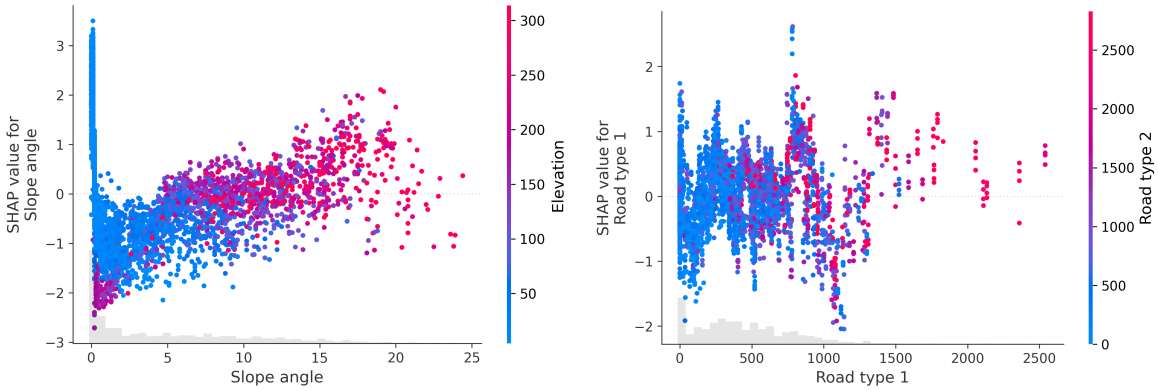
In addition, we selected several highly interactive factors to create interaction plots, as shown in Fig.7.6.

Interaction between Slope Angle and Elevation: As shown in Fig. 7.6(a), in low-elevation areas (represented by blue points), when the slope angle is small (approximately 0–5 degrees), SHAP values are predominantly negative and remain within a slightly negative range. This indicates that in relatively flat, low-elevation regions, slope angle has a minimal or even slightly negative correlation with the impact of snowstorms on traffic. However, once the slope angle exceeds approximately 5 degrees, SHAP values gradually rise towards positive values, suggesting that even in low-elevation areas, increased slope angles gradually amplify the negative effects of snowstorms on traffic conditions. In contrast, in high-elevation areas (represented by red points), the interaction between slope angle and elevation is more pronounced. As the slope angle reaches moderate levels (approximately 5–10 degrees), SHAP values shift noticeably from negative to positive. Moreover, as slope angle continues to increase, the positive effect on SHAP values steadily intensifies. This trend indicates that in high-elevation areas, even moderate slope angles can significantly exacerbate the negative impact of snowstorms on traffic.

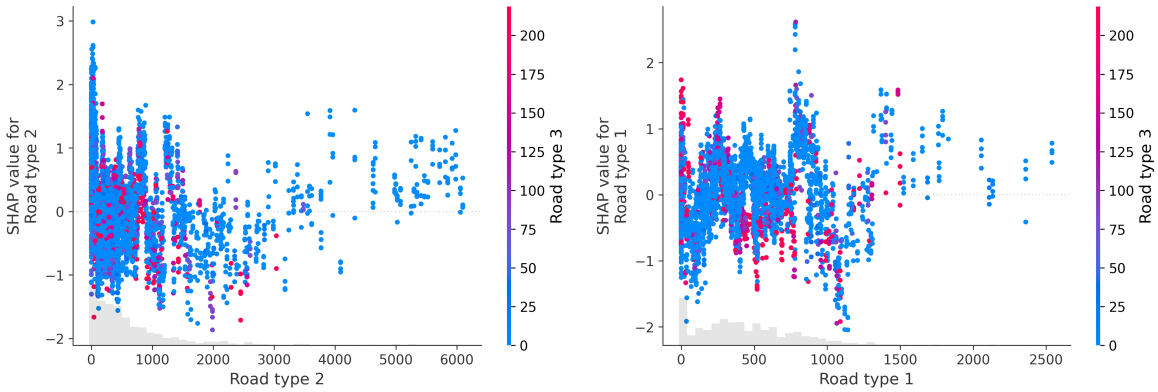
Interaction between Road Type 2 and Road Type 1:

As shown in Fig. 7.6(b), the interaction plot indicates a clear interplay between Road Type 2 and Road Type 1 densities regarding their impact on snowstorm-induced traffic congestion. When the density of Road Type 1 is relatively low (less than 1000 m/km^2), SHAP values predominantly exhibit negative values, implying a lower likelihood of severe congestion under snowstorm conditions. Interestingly, the color gradient reveals that these regions generally feature a lower density of Road Type 2 (represented primarily by blue dots), suggesting that low densities of both road types are associated with relatively minor impacts from snowstorms. However, as the density of Road Type 1 increases beyond approximately 1000 m/km^2, the SHAP values gradually shift towards positive, indicating an elevated risk of congestion during snowstorms. Notably, this shift coincides with a simultaneous increase in Road Type 2 density, represented by the transition towards red dots. This indicates a nuanced relationship: although higher Road Type 1 densities alone may increase congestion risk, the concurrent increase in Road Type 2 density initially mitigates this risk, highlighting the crucial role of Road Type 2 in enhancing traffic flow capacity and reducing vulnerability.

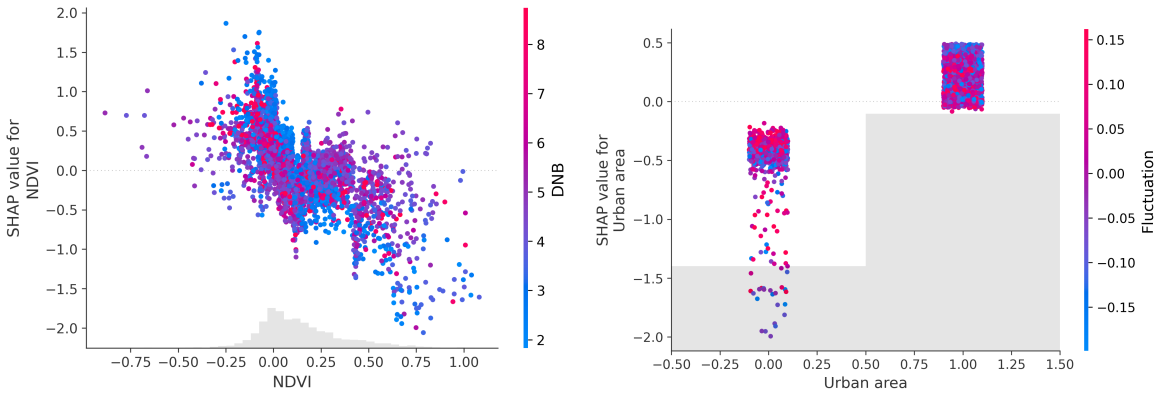
In high-density scenarios (especially when Road Type 1 exceeds 1500 m/km^2), SHAP values again become more positive, suggesting a complex dynamic where the protective capacity of wide roads (Road Type 2) might reach a threshold. This complexity underlines



(a) Interaction between Elevation and Slope Angle (b) Interaction between Road type 2 and Road type 1



(c) Interaction between Road type 2 and Road type 3 (d) Interaction between Road type 1 and Road type 3



(e) Interaction between NDVI and DNB (f) Interaction between Fluctuation and Urban

Figure 7.6: Interaction Plots for Different Variables

the importance of balanced infrastructure development; optimal densities of both road types are necessary to achieve effective traffic resilience during snowstorm events.

Interaction between Road Type 2 and Road Type 3: As illustrated in Fig. 7.6(c), when the density of Road Type 2 is relatively low (less than approximately 1000 m/km^2), the SHAP values are predominantly positive, suggesting a higher vulnerability to traffic disruptions caused by snowstorms. Notably, these areas are characterized by lower densities of Road Type 3, as indicated by the color gradient (primarily blue and purple). This observation implies that insufficient redundancy in the road network may exacerbate the negative impacts of snowstorms. As the density of Road Type 2 increases into the range of 1000 $\sim$ 3000 m/km^2, SHAP values progressively shift towards neutral or negative values, indicating a decreased likelihood of snowstorm-induced traffic impacts. This trend coincides with an increase in the density of Road Type 3 (represented by a transition to purple and red), highlighting that balanced densities of both road types effectively enhance regional resilience against snowstorms.

However, when the density of Road Type 2 further exceeds approximately 3500 m/km^2, Road Type 3 density becomes negligible, and SHAP values slightly shift back toward positive. This suggests that regions predominantly reliant on a single road type (Road Type 2), without adequate road-type diversity, may still be susceptible to the adverse effects of snowstorms.

Interaction between Road Type 1 and Road Type 3: As indicated by Fig. 7.6(d), SHAP values are predominantly positive when the density of Road Type 1 is low, suggesting that regions with sparse Road Type 1 infrastructure are more susceptible to snowstorm impacts. Within the medium-density range (500 $\sim$ 1500 m/km^2) of Road Type 1, there is a noticeable decrease in Road Type 3 density (as shown by the transition in colors), accompanied by a gradual decline of SHAP values towards negative. This implies that increased density of wide roads (Road Type 1) effectively mitigates the adverse effects of snowstorms. Furthermore, when the density of Road Type 1 exceeds approximately 1500 m/km^2, Road Type 3 density becomes nearly negligible, and SHAP values stabilize at negative values. This trend further emphasizes that areas characterized by high densities of wide roads (Road Type 1), even without significant support from Road Type 3, can still robustly mitigate snowstorm-induced traffic disruptions.

Interaction between NDVI and DNB: As shown in Fig. 7.6(e), when NDVI values are relatively low (approximately between 0.2 and 0.2), SHAP values are mostly positive, indicating that areas with poor vegetation cover are more likely to experience snowstorm-related traffic impacts. In these regions, DNB values are relatively high (represented by red and purple colors), suggesting strong nighttime lighting, which typically corresponds to urban or densely populated areas. This implies that urban regions with sparse vegetation are more vulnerable to traffic disruptions caused by snowstorms. As NDVI increases (beyond

approximately 0.4), SHAP values gradually shift to negative, and DNB values decrease (color transitions to blue), indicating that areas with good vegetation cover and weaker nighttime lighting are less prone to snowstorm-related traffic risks. This trend may be attributed to the fact that well-vegetated areas, such as rural or mountainous regions, tend to have lower population density and traffic volume, making them less severely affected by snowstorms.

Interaction between Urban Area and Population Fluctuation: As shown in Fig. 7.6(f), SHAP values in urban areas (Urban = 1) are generally close to zero or slightly positive, indicating that urban regions are more likely to experience snowstorm-related traffic disruptions. Although the population fluctuation values (Fluctuation) in these areas vary widely (ranging from blue to red), the overall risk of snowstorm impact remains relatively high. This may be attributed to the high population density and complex transportation systems in urban settings, where even with better infrastructure, extreme weather events can easily lead to congestion and service disruption. In contrast, SHAP values in non-urban areas (Urban = 0) are predominantly negative, suggesting a lower predicted impact of snowstorms on traffic. It is worth noting that some non-urban regions also exhibit high population fluctuation (as indicated by red points), but overall, regardless of fluctuation intensity, the snowstorm risk remains lower than in urban regions. This highlights urbanization itself as a key factor influencing transportation vulnerability during snowstorms.

Combining the Local Dependence Plots (LDPs) and interaction plots, it can be observed that in a 500-meter grid cell, the likelihood of snowstorm impact increases under the following conditions: elevation is below 46.6 m, slope angle exceeds 9.9°, the density of Road Type 1 is less than 344 m/km^2, the density of Road Type 2 is less than 1063.25 m/km^2, and Road Type 3 is nearly absent. Additionally, areas with low NDVI values (less than 0.2) and strong nighttime lighting (high DNB values) are more susceptible, particularly when located in urban settings. This vulnerability persists regardless of population fluctuation levels, suggesting that urbanization and road structure deficiencies are dominant factors contributing to traffic disruption during snowstorms.

7.2.5 Geospatial analysis of Risk

By selecting the optimal model from the six proposed methods and integrating SHAP theory, we predicted the snow disaster probability for each 500-meter grid cell, creating a probabilistic risk map for snow disasters in Fukui Prefecture, as shown in Fig. 7.7. Additionally, we visualized the most influential feature within each grid cell, generating a feature impact distribution map, as shown in Fig. 7.8. These findings provide valuable insights for local governments to optimize environmental variables on a small scale, thereby enhancing resilience against snow disasters and mitigating potential losses in future events.

In Fig. 7.7, we predicted the probability of snow disaster impact occurrence for each

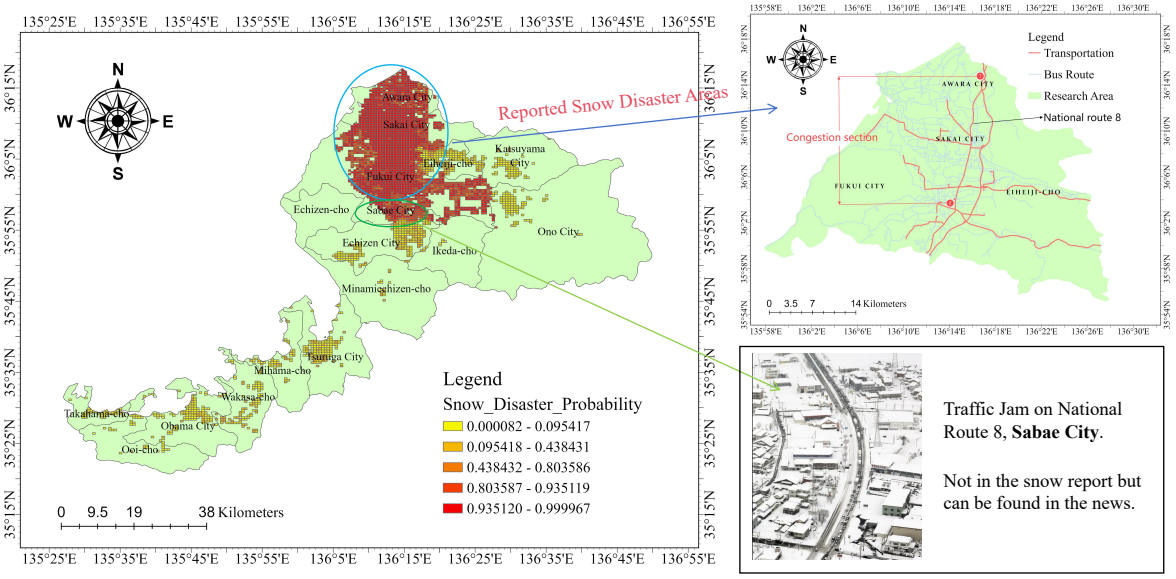


Figure 7.7: Predicted and Actual Snow Disaster Areas [1, 4]

500-meter grid cell. The redder the grid cell color, the higher the probability of a snow disaster. It is evident that Fukui City, Sakai City, and Awara City are high-risk areas for snow disasters, consistent with the affected regions reported in [1]. However, our model also predicted a high probability of snow disaster occurrence in Sabae City, which was not mentioned in the reports. Interestingly, we found online images showing traffic congestion in Sabae City caused by snow disasters [4], further validating the effectiveness and practical value of our model.

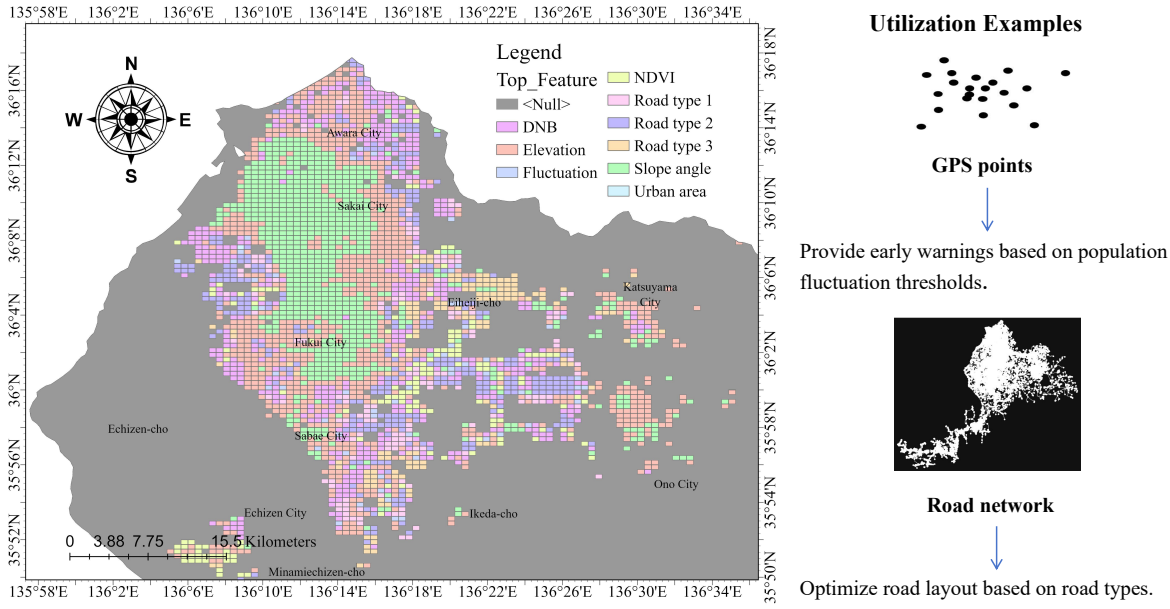


Figure 7.8: Spatial Distribution of Snow Disaster Influencing Factors in Fukui Prefecture

In Fig. 7.8, we identified the most influential factor for each grid cell, providing a refined reference for local governments to allocate resources and formulate snow disaster mitigation

strategies on a smaller scale. Given that the high-risk zones predicted in Fig. 7.7 are primarily concentrated in the northern part of Fukui Prefecture, this study focuses on analyzing the spatial distribution of environmental influencing factors in the northern region. This focus not only facilitates the precise identification of critical areas but also offers readers a clearer understanding of the distribution characteristics.

Specifically, factors such as elevation, slope, and urban areas are relatively challenging to modify. However, road width and road density can be optimized through future urban planning. For example, Road Type 3, which has narrower widths, shows a lower probability of snow disaster occurrence. Therefore, governments can focus more attention on Road Types 1 and 2. Additionally, since there is an interaction effect between Road Types 1 and 2, appropriately adjusting their combined layout may further enhance the region's resilience to snow disasters.

Furthermore, based on the previously calculated thresholds for various variables, it is also possible to implement early warning systems. For instance, with real-time GPS data, if the population fluctuation threshold in a specific area is detected to fall within the range of $[-0.25, 0]$, a warning could be issued advising people to avoid entering the area simultaneously, as it is highly likely to experience traffic congestion caused by snow disasters.

In summary, Fig. 7.7 and Fig. 7.8 provide valuable insights for the local government of Fukui, offering critical guidance for the efficient allocation of resources and manpower while also laying a data-driven foundation for formulating effective snow disaster response strategies.

7.3 Limitations and Discussions

To evaluate the performance of the selected models, we compared their predictive metrics in identifying snowstorm occurrences across the study area. Among the models tested, XGBoost achieved the highest accuracy, precision, and recall. This result is consistent with previous studies [142, 143], which have demonstrated that tree-boosting methods are particularly effective in handling imbalanced, nonlinear, and spatially heterogeneous data in disaster risk assessments. The superior performance of XGBoost can be attributed to its ability to capture complex feature interactions and provide hierarchical feature importance, which is essential when modeling the interplay among environmental, infrastructural, and topographical factors. In contrast, while MLP can model nonlinear relationships, it lacks interpretability and sometimes experiences unstable convergence with limited training samples. Random Forest also performed robustly but was slightly outperformed by XGBoost, potentially due to the lack of sequential boosting and adaptive weighting.

These findings demonstrate the effectiveness of the selected models within the con-

text of our case study. However, several limitations must be acknowledged regarding the generalizability and scope of the current analysis.

This study focuses on the 2018 Fukui snowstorm event, aiming to explore and identify key factors and their nonlinear effects associated with snowstorm impacts. However, to develop a more generalizable and robust predictive model, it is necessary to incorporate multiple snowstorm events across different regions and time periods. Although Fukui is one of the most snowstorm-prone areas in Japan, the collection of data from other snowstorm events in the region has been constrained by current budgetary limitations and challenges in data accessibility. Furthermore, due to the relatively limited number of systematic studies on snow disasters, it remains difficult to extract high-quality, comparable data from the existing literature.

In addition, while our study highlights the novelty of applying explainable machine learning methods to snowstorm impact assessment, we recognize that similar techniques have already been utilized in the context of other natural hazards, such as tsunamis and earthquakes [144, 145]. Thus, although our application is relatively new in the domain of snow-related disasters, the methodological foundation builds upon and extends previous work in the broader field of disaster science. Within this context, the primary contribution of our study lies in the integration of multi-source data and the development of an interpretable modeling framework tailored to the snowstorm scenario, enabling the identification of critical influencing factors and nonlinear effects.

Additionally, while interpreting the model outputs, special attention should be given to variables like elevation. The SHAP analysis indicated that grid cells with elevations below 46.6 meters tended to show higher predicted risk. This pattern should not be simplistically interpreted as a linear relationship such as "low elevation = high population density = high risk." While low-lying urban areas do often have denser infrastructure and road networks, which increases their vulnerability, some low-elevation regions may actually consist of agricultural fields or paddy lands with sparse population and minimal traffic activity. Such spatial heterogeneity within the same elevation range can lead to misinterpretation if land use context is not considered. Acknowledging this limitation is important to prevent overgeneralization and to guide future research toward integrating elevation, land use, and mobility factors more holistically in risk interpretation.

This study investigated variable interactions primarily among road types (Type 1, Type 2, and Type 3), as road structure plays a central role in determining traffic flow and disruption during snowstorm events. While these interactions yielded insightful results, the current analysis did not explore more complex and practically meaningful combinations, such as "road type × slope" or "land use × snowfall depth." These combinations may reveal nonlinear risk patterns and offer more precise guidance for disaster response planning. For example,

secondary roads in steep terrain may be more prone to traffic blockages under snowy conditions, and industrial or residential land use may be associated with differing snow removal capacities. Integrating these interaction effects into the modeling framework could improve the interpretability and decision-support capabilities of snow disaster risk models. Due to time and computational constraints, these combinations were not analyzed in the current study but are planned as a key extension in future research.

Moreover, the proposed modeling framework is highly adaptable and not limited to snowstorm scenarios. The methodology can be extended to other types of natural disasters, such as floods, hurricanes, and tsunamis. Its core strength lies in its ability to accommodate structurally similar input data, allowing for the investigation of key impact mechanisms under diverse disaster contexts.

7.4 Conclusion

This study integrates multiple machine learning algorithms and SHAP theory to investigate the influencing factors and nonlinear relationships related to the 2018 Fukui snowstorm. It produces a 500-meter grid cell-scale snowstorm impact prediction map and spatial distribution maps of influencing factors. The main findings are as follows:

This study developed an interpretable machine learning framework to investigate the nonlinear relationships between environmental factors and snowstorm impact occurrence. Among the six models evaluated, XGBoost, MLP, Random Forest, Decision Tree, SVM and LightGBM, XGBoost emerged as the best-performing algorithm, achieving a prediction accuracy of 0.8833. The feature importance analysis ranked key contributors as follows: Elevation > Slope > DNB > Road type 3 > Road type 2 > Road type 1 > NDVI > Fluctuation > Urbanization.

Combining the Local Dependence Plots (LDPs) and interaction plots, it was observed that in a 500-meter grid cell, snowstorm impact occurrence is more likely when the elevation is below 46.6 m, the slope angle exceeds 9.9°, the density of Road type 1 is less than 344 m/km^2, the density of Road type 2 is less than 1063.25 m/km^2, and Road type 3 is nearly absent, the population fluctuation falls within $[-0.25, 0]$, and the area is classified as urban. Furthermore, two remote sensing-based indicators—nighttime light intensity (DNB) and vegetation index (NDVI)—also contribute to snowstorm impact occurrence. Specifically, DNB values above 2.63 nW $\cdot$ cm^{-2} $\cdot$ sr^{-1} are associated with increased risk, reflecting the vulnerability of highly urbanized or trafficked areas. In contrast, NDVI values above 0.13 are associated with reduced snowstorm impact, suggesting that areas with higher vegetation coverage may exhibit enhanced resilience.

In addition, among the three road types, areas with high densities of Road type 3 are

less likely to experience snowstorms due to their narrow widths, low traffic volumes, and predominant use as pedestrian pathways. In contrast, snowstorms are more prevalent in areas dominated by Road type 2, which serves as the primary traffic infrastructure. Regions with interactions between Road types 1 and 2 also tend to experience traffic congestion caused by snowstorms; however, the likelihood decreases as the densities of these two road types increase. Optimizing the spatial configuration of Road types 1 and 2 can therefore effectively mitigate the impacts of snowstorms.

The 500-meter grid cell risk prediction results indicate that snowstorms are primarily concentrated in Fukui City, Sakai City, and Awara City [1]. Additionally, Sabae City, which was not included in post-disaster reports but was reported in media coverage [4], was also identified as a high-risk area, further validating the model's applicability. The spatial distribution maps of influencing factors provide actionable insights for local governments to implement early warning systems and road management strategies. Although terrain-related features such as elevation and slope are challenging to modify, snowstorm resilience can be enhanced through the monitoring of GPS data for preemptive warnings or by optimizing the density and spatial configuration of Road types 1 and 2, thereby enhancing regional traffic resilience. These findings underscore the potential of this model to guide practical snowstorm management and mitigation efforts.

This study identified and analyzed the key driving factors behind snowstorm risk and their nonlinear relationships, deepening the understanding of the spatial variability and underlying mechanisms of snowstorm risks. Additionally, spatial risk analysis was conducted in Fukui Prefecture, Japan, to quantify and locate high-risk and vulnerable areas, providing scientific support for local governments to formulate more precise snowstorm mitigation policies. However, due to varying levels of data availability, only 3480 grid cells out of a total of 16,815 500-meter grid cells in Fukui Prefecture were ultimately used for analysis. Evaluating and predicting the areas with missing data will be an important direction for future research.

Chapter 8

Conclusion and Future Work

8.1 Conclusion

This dissertation focuses on snow disaster risk identification and assessment, addressing the challenges from four key perspectives: intercity transportation demand analysis, resilience evaluation of different land-use types, and interpretable machine learning-based spatial risk modeling. The main conclusions are as follows:

- **Analysis of Intercity Transportation Demand under Snowstorms**
  Based on Agoop mobile GPS data and Singular Value Decomposition (SVD) methods, this study revealed the intercity travel demand changes and localized pattern shifts during the 2018 Fukui snowstorm, laying the foundation for the target variable construction in subsequent chapters. Three distinct phases were identified: the stable phase (January 27 to February 2), the snow disaster phase (February 3 to February 11), and the recovery phase (February 12 to February 16). During the snow disaster phase, intercity transportation demand decreased by 67.86% compared to the stable phase. Travel demand was classified into daily demand (M_1) and special demand (M_2). M_1 traffic primarily originated from Fukui City, Sabae City, Awara City, and Sakai City, with Fukui and Sakai serving as key departure and arrival hubs. In the M_2 model, travel from southern to northern Fukui nearly ceased, shifting from a longitudinal distribution along National Route 8 and the Hokuriku Expressway to a horizontal pattern toward Ono and Katsuyama. Departure flows from Sakai City declined significantly, while Tsuruga City experienced a substantial increase, indicating that the snow disaster mainly affected northern Fukui Prefecture, particularly Fukui City and Sakai City.

- **Detecting High-Risk Traffic Congestion Areas**
  A Random Forest model integrating multi-source spatial data was constructed to identify critical factors influencing traffic congestion under snowstorm conditions at the

grid level. The results show that: Severe traffic congestion during the 2018 Fukui snowstorm primarily occurred in Sabae, Fukui, and Awara. Initial congestion points were frequently located on intercity roads near administrative boundaries, especially in areas classified as "field" or "forest" land-use types. The model achieved a prediction accuracy of 94.59% when trained on data from 10 cities and tested on 7 other cities within Fukui Prefecture. Feature importance analysis revealed that the top influential factors are: Snow Depth > Nighttime Light Difference > Elevation > Slope Angle > Urban Area > NDVI > Population Change > Forest > Field > Low-rise Buildings (Sparse) > Low-rise Buildings (Dense).

- **Micro-Scale Resilience Assessment of Land-Use Types**
  A micro-scale resilience assessment framework was developed based on the Resilience Triangle Theory, utilizing 500-meter grid data to quantify the resilience of different land-use types under snowstorm conditions. The results indicate that land-use types such as factories, agricultural land, building land, and mid- to high-rise buildings exhibit better resilience—particularly mid- to high-rise buildings, which demonstrate strong resistance and recovery capabilities when subjected to disturbances. In contrast, areas dominated by sparse low-rise buildings, parks, facility lands, and dense low-rise buildings exhibit weaker resilience and require longer recovery times. Spatially, areas with poor transportation resilience were mainly concentrated in northern Fukui, particularly along the vertical corridors through Fukui City, Awara City, and Sakai City, and from Sabae City to Echizen City, as well as the horizontal corridor from Eiheiji-cho to Katsuyama City.

- **Interpretable Machine Learning Model for Snow Disaster Risk Identification**
  An interpretable snow disaster risk identification model was developed by integrating the XGBoost classifier with SHAP theory. This model effectively identified critical factors influencing snow disaster risks, such as elevation, slope, road density, and remote sensing indicators, revealing significant nonlinear threshold effects and spatial heterogeneity. The XGBoost model outperformed MLP, Random Forest, Decision Tree, SVM, and LightGBM models, achieving a prediction accuracy of 0.8833.

  Elevation, slope, NDVI, DNB, population fluctuation, road density, and road width exhibited strong nonlinear impacts and threshold effects on snowstorm risk occurrence. Urban areas with elevation below 46.6 m, slopes exceeding 9.9°, a density of major roads (Road Type 1) less than 344 m/km^2, a density of minor roads (Road Type 2) less than 1063.25 m/km^2, and nearly absent rural roads (Road Type 3), coupled with population fluctuations in the range of $[-0.25, 0]$, are particularly vulnerable to snow disasters. Furthermore, remote sensing indicators show that areas with nighttime light

(DNB) values above 2.63 $nW \cdot cm^{-2} \cdot sr^{-1}$—typically urban or high-traffic regions—face increased snowstorm risks. In contrast, areas with NDVI values above 0.13, indicating higher vegetation coverage, are associated with lower snowstorm impact.

Snow disaster resilience for mitigating traffic congestion can be improved by incorporating early warnings based on GPS data and optimizing the spatial configuration of major and minor roads.

In summary, this research enriches the academic field of snow disaster risk management by integrating multi-source spatial data, advanced machine learning models, and explainable AI techniques. The results provide valuable theoretical foundations and decision-making references for governments to formulate more scientific and effective disaster prevention and mitigation strategies.

8.2 Future Work

Although this study has made significant progress in snow disaster risk analysis, several limitations remain, pointing to directions for future research:

- **Improvement of Data Accuracy**
  Currently, snow depth monitoring stations are available in only six cities, and the data are aggregated at the city level. This coarse spatial resolution introduces considerable limitations when integrating with other high-resolution spatial datasets. Future research should focus on acquiring or simulating high-precision snow depth data to enhance spatial analysis accuracy and improve the reliability of risk assessments.

- **Dynamic Risk Monitoring and Early Warning Systems**
  Due to the limited spatial and temporal resolution of snow depth data, the current study developed a static spatial risk assessment model. In future work, real-time meteorological data should be incorporated to establish a dynamic snow disaster early warning framework, enabling more timely and effective risk prediction and management. This shift from static to real-time dynamic risk prediction would provide critical decision support for practical disaster prevention and emergency response.

- **Model Generalization and Applicability**
  This study focused primarily on Fukui Prefecture, and the generalization capability of the proposed models has not been fully validated in other regions. Future studies should conduct cross-regional empirical validations to assess the robustness and applicability of the models in areas with different geographical and socio-economic characteristics.

Moreover, the current research framework could be extended to other natural disasters such as typhoons, floods, and earthquakes, contributing to the development of a comprehensive multi-hazard risk assessment system and enhancing regional disaster prevention and mitigation capabilities.

Publications

Journal Papers

1. Yang, Z., & Gokon, H. (2024). A Study of Intercity Transportation Demand in Fukui Prefecture Under 2018 Heavy Snowfall: Based on Mobile GPS Data. *Intelligence, Informatics and Infrastructure*, 5(2), 57-65.

2. Yang, Z., Gokon, H., & Liu, Z. (2025). Assessing Regional Resilience of Different Land Use Types During Snowstorms Using Mobile Data. *Progress in Disaster Science*, 100412.

3. Yang, Z., Gokon, H., et al. (2025). Machine Learning-Based Identification and Assessment of Snow Disaster Risks Using Multi-Source Data: Insights from Fukui Prefecture, Japan. *Progress in Disaster Science*, 100426.

4. Yang, Z., & Gokon, H. (Accepted). Detecting High-Risk Traffic Congestion Areas During Snow Disasters: A Random Forest-Based Spatial Model. *Intelligence, Informatics and Infrastructure*.

Conference Presentations

1. Yang, Z. Y. Exploring the Relationship Between Snow Depth and Population Mobility Using Mobile GPS and Weather Data. *International Conference: Big Data for Disaster Response and Management in Asia and the Pacific*, Tohoku University, Japan. (Oral Presentation)

2. Yang, Z. A Study of Intercity Transportation Demand in Fukui Prefecture Under 2018 Heavy Snowfall: Based on Mobile GPS Data. *Intelligence, Informatics and Infrastructure Symposium*. (Online Oral Presentation)

3. Yang, Z. Y. Assessing Regional Resilience of Different Land Use Types During Snowstorms Using Mobile Data. *16th International Conference on Applied Energy (ICAE2024)*, Japan. (Oral Presentation)

4. Yang, Z. Y. *An Interpretable Machine Learning Model for Snowstorm Risk Identification and Assessment*. IEEE IGARSS 2025, Brisbane, Australia, August 3–8, 2025. (Accepted)

Bibliography

[1] Overview of the 2018 heavy snowfall in fukui prefecture (in japanese). https://www.pref.fukui.lg.jp/doc/kikitaisaku/disaster_archive_fukui_d/fil/11ooyuki_H30_2.pdf.

[2] Overview of the 2018 heavy snowfall in fukui prefecture, 2018. Available at https://www.kkr.mlit.go.jp/bousai/saigairaiburari/disasters_detail.php?id=44 (in Japanese).

[3] Takahiro Yabe, Yunchang Zhang, and Satish Ukkusuri. Quantifying the economic impact of extreme shocks on businesses using human mobility data: a bayesian causal inference approach. *arXiv preprint arXiv:2004.11121*, 2020.

[4] Overview of the 2018 heavy snowfall in fukui prefecture (in japanese). https://www.yomiuri.co.jp/s/ims/fukuisnow2018/. Accessed: 2024-12-03.

[5] B. Zhou, L. Gu, Y. Ding, L. Shao, Z. Wu, X. Yang, C. Li, Z. Li, X. Wang, Y. Cao, and B. Zeng. The great 2008 chinese ice storm: Its socioeconomic–ecological impact and sustainability lessons learned. *Bulletin of the American Meteorological Society*, 92(1):47–60, 2011.

[6] J. He, W. Duan, Y. Zhou, and Y. Su. Impact of media information on social response in disasters: A case study of the freezing-rain and snowstorm disasters in southern china in 2008. *International Journal of Disaster Risk Science*, 15(1):73–87, 2024.

[7] K. Greening and A. Hodgson. Atmospheric analysis of the cold late february and early march 2018 over the uk. *Weather*, 74(3):79–85, 2019.

[8] N. M. Flores, H. McBrien, V. Do, M. V. Kiang, J. Schlegelmilch, and J. A. Casey. The 2021 texas power crisis: Distribution, duration, and disparities. *Journal of Exposure Science Environmental Epidemiology*, 33(1):21–31, 2023.

[9] TI Hua, SO Jianyang, GA Jingjing, ZH Luqiang, LI Kunyu, and YA Minyue. Comparative analysis of two rainfall/snowfall and freezing weather events impact on the highway in early 2024. *Torrential Rain and Disasters*, 43(4):490–498, 2024.

[10] Vesela Radović and Isabel Iglesias. Extreme weather events: Definition, classification, and guidelines towards vulnerability reduction and adaptation management. In *Climate Action*, pages 464–476. Springer, 2019.

[11] W. Shijin, Y. Yuande, and C. Yanjun. Global snow-and ice-related disaster risk: A review. *Natural Hazards Review*, 23(4):03122002, 2022.

[12] W. K. Davies. Resilient cities: Coping with natural hazards. In *Theme Cities: Solutions for Urban Problems*, pages 311–357. Springer International Publishing, Cham, 2015.

[13] S. Meerow, J. P. Newell, and M. Stults. Defining urban resilience: A review. *Landscape and Urban Planning*, 147:38–49, 2016.

[14] C. Albert, S. Rufat, and C. Kuhlicke. Five principles for climate-resilient cities. *Nature*, 596(7873):486–486, 2021.

[15] M. Karamouz, A. Hosseinpour, and S. Nazif. Improvement of urban drainage system performance under climate change impact: Case study. *Journal of Hydrologic Engineering*, 16(5):395–412, 2011.

[16] Y. B. Choi, R. W. Kim, and I. B. Lee. Numerical analysis of snow distribution on greenhouse roofs using cfd–dem coupling method. *Biosystems Engineering*, 237:196–213, 2024.

[17] B. Chen and X. Fan. Msgc-yolo: An improved lightweight traffic sign detection model under snow conditions. *Mathematics*, 12(10):1539, 2024.

[18] G. Peng, M. Xu, and H. Tan. Phase transition in a new heterogeneous macro continuum model of traffic flow under rain and snow weather environment. *Physica A: Statistical Mechanics and its Applications*, 637:129585, 2024.

[19] N-E El Faouzi, Romain Billot, and Salim Bouzebda. Motorway travel time prediction based on toll data and weather effect integration. *IET intelligent transport systems*, 4(4):338–345, 2010.

[20] Evan A Thaler, Ryan L Crumley, and Katrina E Bennett. Estimating snow cover from high-resolution satellite imagery by thresholding blue wavelengths. *Remote Sensing of Environment*, 285:113403, 2023.

[21] Claudia Bergroth, Olle Järv, Henrikki Tenkanen, Matti Manninen, and Tuuli Toivonen. A 24-hour population distribution dataset based on mobile phone data from helsinki metropolitan area, finland. *Scientific data*, 9(1):39, 2022.

[22] Mohamed Batran, Mariano Gregorio Mejia, Hiroshi Kanasugi, Yoshihide Sekimoto, and Ryosuke Shibasaki. Inferencing human spatiotemporal mobility in greater maputo via mobile phone big data mining. *ISPRS International Journal of Geo-Information*, 7(7):259, 2018.

[23] Ying Lv, Danyue Zhi, Huijun Sun, and Geqi Qi. Mobility pattern recognition based prediction for the subway station related bike-sharing trips. *Transportation research part C: emerging technologies*, 133:103404, 2021.

[24] Zhenyu Yang and Hideomi Gokon. A study of intercity transportation demand in fukui prefecture under 2018 heavy snowfall: Based on mobile gps data. *Intelligence, Informatics and Infrastructure*, 5(2):57–65, 2024.

[25] Takahiro Yabe, Nicholas KW Jones, Nancy Lozano-Gracia, Maham Faisal Khan, Satish V Ukkusuri, Samuel Fraiberger, and Aleister Montfort. Location data reveals disproportionate disaster impact amongst the poor: A case study of the 2017 puebla earthquake using mobilkit. *arXiv preprint arXiv:2107.13590*, 2021.

[26] Y. Hara and H. Yamaguchi. Japanese travel behavior trends and change under covid-19 state-of-emergency declaration: Nationwide observation by mobile phone location data. *Transportation Research Interdisciplinary Perspectives*, 9:100288, 2021.

[27] Hao Li, Wenshuang Xi, Lijuan Zhang, and Shuying Zang. Snow-disaster risk zoning and assessment in heilongjiang province. *Sustainability*, 13(24):14010, 2021.

[28] Zhenzhen Yang, Feng Liu, Ziyou Gao, Huijun Sun, Jiandong Zhao, Davy Janssens, and Geert Wets. Estimating the influence of disruption on highway networks using gps data. *Expert Systems with Applications*, 187:115994, 2022.

[29] Tasnuba Binte Jamal and Samiul Hasan. Understanding the loss in community resilience due to hurricanes using facebook data. *International journal of disaster risk reduction*, 97:104036, 2023.

[30] W. Sun, P. Bocchini, and B. D. Davison. Resilience metrics and measurement methods for transportation infrastructure: The state of the art. *Sustainable and Resilient Infrastructure*, 5(3):168–199, 2020.

[31] H. Lee, D. Kim, and G. Chung. Classification of risk levels for snow damage estimation considering socioeconomic factors in south korea. *Applied Water Science*, 14(11):241, 2024.

[32] S. Gascoin, K. Luojus, T. Nagler, H. Lievens, M. Masiokas, T. Jonas, and P. De Rosnay. Remote sensing of mountain snow from space: status and recommendations. *Frontiers in Earth Science*, 12:1381323, 2024.

[33] J. Wang, H. Li, X. Hao, X. Huang, J. Hou, T. Che, and Z. Wang. Remote sensing for snow hydrology in china: challenges and perspectives. *Journal of Applied Remote Sensing*, 8(1):084687–084687, 2014.

[34] N. Eckert and F. Giacona. Towards a holistic paradigm for long-term snow avalanche risk assessment and mitigation. *Ambio*, 52(4):711–732, 2023.

[35] F. Cappabianca, M. Barbolini, and L. Natale. Snow avalanche risk assessment and mapping: A new method based on a combination of statistical analysis, avalanche dynamics simulation and empirically-based vulnerability relations integrated in a gis platform. *Cold Regions Science and Technology*, 54(3):193–205, 2008.

[36] D. Germain. Snow avalanche hazard assessment and risk management in northern quebec, eastern canada. *Natural Hazards*, 80:1303–1321, 2016.

[37] D. Liu, L. Zhang, S. Jiang, S. Shi, and Y. Li. Hazard prediction and risk regionalization of snowstorms in northeast china. *Physics and Chemistry of the Earth, Parts A/B/C*, 116:102832, 2020.

[38] L. Zhang, Y. Wang, and Q. Sun. Seasonal snow monitoring using modis data across diverse land covers. *Environmental Monitoring and Assessment*, 192(12):789, 2020.

[39] J. Gao, X. Huang, X. Ma, Q. Feng, T. Liang, and H. Xie. Snow disaster early warning in pastoral areas of qinghai province, china. *Remote Sensing*, 9(5):475, 2017.

[40] B. Choubin, M. Borji, A. Mosavi, F. Sajedi-Hosseini, V. P. Singh, and S. Shamshirband. Snow avalanche hazard prediction using machine learning methods. *Environmental Earth Sciences*, 78(18):1–14, 2019.

[41] R. Ahmad. Smart remote sensing network for disaster management: an overview. *Telecommunication Systems*, 87:213–237, 2024.

[42] X. Sun, L. Miao, X. Feng, and X. Zhan. Snow disaster risk assessment based on long-term remote sensing data: A case study of the qinghai–tibet plateau region in xizang. *Remote Sensing*, 16(10):1661, 2024.

[43] E. Papucci, V. Manabe, S. Suvanto, S. Adler, and I. Bohlin. Mapping the risk for wind and snow damage using nfi field plots and auxiliary remote sensing and weather data. In *ForestSAT 2024*, Rotorua, New Zealand, 2024.

[44] E. Ubaldi, T. Yabe, N. K. Jones, M. F. Khan, S. V. Ukkusuri, R. Di Clemente, and E. Strano. Mobilkit: A python toolkit for urban resilience and disaster risk management analytics using high frequency human mobility data. *arXiv preprint*, arXiv:2107.14297, July 2021.

[45] D. Cicek and B. Kantarci. Use of mobile crowdsensing in disaster management: A systematic review, challenges, and open issues. *Sensors*, 23(3):1699, January 2023.

[46] A. Wu, X. Yan, E. Kuligowski, R. Lovreglio, D. Nilsson, T. J. Cova, Y. Xu, and X. Zhao. Wildfire evacuation decision modeling using gps data. *International Journal of Disaster Risk Reduction*, 83:103373, December 2022.

[47] R. Verma, S. Mittal, Z. Lei, X. Chen, and S. V. Ukkusuri. Comparison of home detection algorithms using smartphone gps data. *EPJ Data Science*, 13(1):6, December 2024.

[48] J. Gao. Analysis and assessment of the risk of snow and freezing disaster in china. *International Journal of Disaster Risk Reduction*, 19:334–340, 2016.

[49] S. Wang, S. Chen, and Y. Wei. Risk prevention and control strategies for the severely affected areas of snow disaster in the three rivers source region (trsr), china. *Sciences in Cold and Arid Regions*, 11(3):248–252, 2019.

[50] F. Leone, A. Colas, Y. Garcin, N. Eckert, V. Jomelli, and M. Gherardi. The snow avalanches risk on alpine roads network. assessment of impacts and mapping of accessibility loss. *Journal of Alpine Research — Revue de géographie alpine*, 102-4, 2014.

[51] Z. Liu, G. L. Hou, and F. Li. Preliminary study of risk assessment of the failure of road traffic caused by snow disaster in guoluo prefecture. In *Medicine and Biopharmaceutical: Proceedings of the 2015 International Conference*, pages 1420–1431. 2016.

[52] D. A. Call and G. A. Flynt. The impact of snowfall on crashes, traffic volume, and revenue on the new york state thruway. *Weather, Climate, and Society*, 14(1):131–141, 2022.

[53] Y. Zhang, H. Li, and J. Wang. Economic impacts of snow-related disruptions in urban transportation systems. *Journal of Urban Economics*, 128:103–115, 2022.

[54] L. Chen, X. Yang, and Q. Liu. Effects of snow cover on agricultural yields in northeast china: A remote sensing perspective. *Agricultural and Forest Meteorology*, 306:108462, 2021.

[55] S. Wang and M. Zhang. Vulnerability of power infrastructure to heavy snowfall and ice accretion in northern regions. *Energy Policy*, 165:112942, 2023.

[56] A. Miller, R. Johnson, and P. Thompson. Health impacts of extreme snowfall events: Hospital admissions due to accidents and hypothermia. *Environmental Health Perspectives*, 128(4):47012, 2020.

[57] Z. Yang, H. Gokon, and Z. Liu. Assessing regional resilience of different land use types during snowstorms using mobile data. *Progress in Disaster Science*, page 100412, February 2025.

[58] S. Tanniru and R. A. A. J. Ramsankaran. Passive microwave remote sensing of snow depth: Techniques, challenges and future directions. *Remote Sensing*, 15(4):1052, 2023.

[59] J. S. Deems, T. H. Painter, and D. C. Finnegan. Lidar measurement of snow depth: a review. *Journal of Glaciology*, 59(215):467–479, 2013.

[60] C. Largeron, M. Dumont, S. Morin, A. Boone, M. Lafaysse, S. Metref, and S. A. Margulis. Toward snow cover estimation in mountainous areas using modern data assimilation methods: A review. *Frontiers in Earth Science*, 8:325, 2020.

[61] T. Smith, J. Anderson, and R. White. Evaluation of synthetic aperture radar (sar) for snow depth estimation in forested and mountainous regions. *Remote Sensing of Environment*, 270:112870, 2022.

[62] R. Brown and M. F. McCabe. Multispectral and hyperspectral imagery for snow cover and depth monitoring: Advances in spatial and temporal resolution. *International Journal of Applied Earth Observation and Geoinformation*, 98:102304, 2021.

[63] Y. Liu, X. Chen, and H. Zhang. Real-time snow depth monitoring using unmanned aerial vehicle technology in remote regions. *Journal of Mountain Science*, 20(5):1153–1165, 2023.

[64] F. Zhang, L. Zhang, Y. Huang, N. Wang, and Y. Li. Hazard assessment and regionalization of snowstorms in heilongjiang province from 1961 to 2015. *Physics and Chemistry of the Earth, Parts A/B/C*, 115:102833, 2020.

[65] N. Dawes, A. Rowlands, and S. Huang. Predicting snow depth and snow water equivalent using machine learning and satellite data in complex terrains. *Hydrology and Earth System Sciences*, 27(3):1451–1468, 2023.

[66] G. Yu, L. Gu, X. Li, and X. Fan. Forest snow depth estimation based on optimized features and dnn network using c-band sar data. *IEEE Geoscience and Remote Sensing Letters*, March 2024.

[67] J. Revuelto, E. Alonso-Gonzalez, I. Vidaller-Gayan, E. Lacroix, E. Izagirre, G. Rodríguez-López, and J. I. López-Moreno. Intercomparison of uav platforms for mapping snow depth distribution in complex alpine terrain. *Cold Regions Science and Technology*, 190:103344, 2021.

[68] F. M. Bianchi, J. Grahn, M. Eckerstorfer, E. Malnes, and H. Vickers. Snow avalanche segmentation in sar images with fully convolutional neural networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:75–82, November 2020.

[69] L. Dai, T. Che, J. Wang, and P. Zhang. Snow depth and snow water equivalent estimation from amsr-e data based on a priori snow characteristics in xinjiang, china. *Remote Sensing of Environment*, 127:14–29, December 2012.

[70] A. Karimpour, A. Hosseinzadeh, and R. Kluger. A data-driven approach to estimating dockless electric scooter service areas. *Journal of Transport Geography*, 109:103579, 2023.

[71] M. Saarela and S. Jauhiainen. Comparison of feature importance measures as explanations for classification models. *SN Applied Sciences*, 3(2):272, 2021.

[72] G. Grekousis, Z. Feng, I. Marakakis, Y. Lu, and R. Wang. Ranking the importance of demographic, socioeconomic, and underlying health factors on us covid-19 deaths: A geographical random forest approach. *Health & Place*, 74:102744, 2022.

[73] N. Savage. Breaking into the black box of artificial intelligence, 2022.

[74] V. Hassija, V. Chamola, A. Mahapatra, A. Singal, D. Goel, K. Huang, and A. Hussain. Interpreting black-box models: a review on explainable artificial intelligence. *Cognitive Computation*, 16(1):45–74, 2024.

[75] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi. A survey of methods for explaining black box models. *ACM Computing Surveys (CSUR)*, 51(5):1–42, 2018.

[76] A. Salih, Z. Raisi-Estabragh, I. B. Galazzo, P. Radeva, S. E. Petersen, G. Menegaz, and K. Lekadir. Commentary on explainable artificial intelligence methods: Shap and lime. *arXiv preprint arXiv:2305.02012*, 2023.

[77] W. E. Marcílio and D. M. Eler. From explanations to feature selection: assessing shap values as feature selection mechanism. In *2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 340–347. IEEE, 2020.

[78] F. Zangenehnejad and Y. Gao. Gnss smartphones positioning: Advances, challenges, opportunities, and future perspectives. *Satellite Navigation*, 2:1–23, 2021.

[79] P. J. G. Ribeiro and L. A. P. J. Gonçalves. Urban resilience: A conceptual framework. *Sustainable Cities and Society*, 50:101625, 2019.

[80] S. Z. Hofmann. 100 resilient cities program and the role of the sendai framework and disaster risk reduction for resilient cities. *Progress in Disaster Science*, 11:100189, 2021.

[81] E. L. Glaeser. Urban resilience. *Urban Studies*, 59(1):3–35, 2022.

[82] H. Wang, G. Peng, and H. Du. Digital economy development boosts urban resilience—evidence from china. *Scientific Reports*, 14(1):2925, 2024.

[83] M. Bruneau, S. E. Chang, R. T. Eguchi, G. C. Lee, T. D. O'Rourke, A. M. Reinhorn, M. Shinozuka, K. Tierney, W. A. Wallace, and D. von Winterfeldt. Framework for resilience in the face of disasters. *Journal of Engineering Mechanics*, 129(8):845–852, 2003.

[84] S. L. Cutter, L. Barnes, M. Berry, C. Burton, E. Evans, E. Tate, and J. Webb. A place-based model for understanding community resilience. *Global Environmental Change*, 18(4):598–606, 2008.

[85] C. W. Zobel and L. Khansa. Characterizing multi-event disaster resilience. *Computers & Operations Research*, 42:83–94, 2014.

[86] R. Sun, G. Gao, Z. Gong, and J. Wu. A review of risk analysis methods for natural disasters. *Natural Hazards*, 100(2):571–593, 2020.

[87] Q. Wang and J. E. Taylor. Patterns and limitations of urban human mobility resilience under the influence of multiple types of natural disaster. *PLOS ONE*, 11(1):e0147299, 2016.

[88] X. Zhou, H. Chen, and Y. Wang. Transportation network resilience metrics under uncertainty. *Transport Policy*, 56:1–10, 2017.

[89] R. Twumasi-Boakye and J. Sobanjo. Civil infrastructure resilience: state-of-the-art on transportation network systems. *Transportmetrica A: Transport Science*, 15(2):455–484, 2019.

[90] D. Chacon-Hurtado, L. L. Losada-Rojas, D. Yu, K. Gkritza, and J. D. Fricker. A proposed framework for the incorporation of economic resilience into transportation decision making. *Journal of Management in Engineering*, 36(6):04020084, 2020.

[91] K. K. Sangha, J. Russell-Smith, J. Evans, and A. Edwards. Methodological approaches and challenges to assess the environmental losses from natural disasters. *International Journal of Disaster Risk Reduction*, 49:101619, 2020.

[92] Research on the resilience of urban traffic systems: Concepts, characteristics, and issues, 2017. Available at https://www.planning.org.cn/news/view?id=15175 (in Chinese).

[93] Jie Ren. Research on road network characteristics and resilience measurement of five new towns in shanghai. *City Planning Review*, 46(9):82–92, 2022.

[94] Ye Li, Xinghua Liu, and Qing He. Urban transportation system resilience during the covid-19 pandemic. *Urban Transport of China*, 18(3):80–87, 2020.

[95] T. Yabe, N. K. Jones, P. S. C. Rao, M. C. Gonzalez, and S. V. Ukkusuri. Mobile phone location data for disasters: A review from natural hazards and epidemics. *Computers, Environment and Urban Systems*, 94:101777, 2022.

[96] B. Hong, B. J. Bonczak, A. Gupta, and C. E. Kontokosta. Measuring inequality in community resilience to natural disasters using large-scale mobility data. *Nature Communications*, 12(1):1870, 2021.

[97] V. Sukhwani and R. Shaw. Operationalizing crowdsourcing through mobile applications for disaster management in india. *Progress in Disaster Science*, 5:100052, 2020.

[98] E. H. Yeh, P. Lin, and M. W. Huang. Adpd: Anomaly detection for population distribution in geo-space using mobile networks data. *IEEE Internet of Things Journal*, 9:22774–22784, 2022.

[99] E. Santiago-Iglesias, J. Carpio-Pinedo, W. Sun, and J. C. García-Palomares. Frozen city: Analysing the disruption and resilience of urban activities during a heavy snowfall event using google popular times. *Urban Climate*, 51:101644, 2023.

[100] HimadriNath Saha, Srijita Basu, Supratim Auddy, Ratul Dey, Arnab Nandy, Debjit Pal, Nirjhar Roy, Subhadeep Jasu, Ankita Saha, SoummyoPriyo Chattopadhyay, et al. A low cost fully autonomous gps (global positioning system) based quad copter for disaster management. In *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, pages 654–660. IEEE, 2018.

[101] Gaurav Shahi, Vikas Sagar, Kumod Kumar Gupta, and Pradeep Kumar Arya. Algorithmic framework for aerial image stitching to create high-resolution maps for navigation in a gps-denied environment. In *International Conference on Advances in Information Communication Technology & Computing*, pages 397–434. Springer, 2024.

[102] Bahareh Raei, Max Kinateder, Noureddine Bénichou, Islam Gomaa, and Xin Wang. Are the data good enough? spatial and temporal modeling of evacuee behavior using gps data in a small rural community. *International Journal of Disaster Risk Reduction*, 116:105054, 2025.

[103] V. S. Brum-Bastos, J. A. Long, and U. Demšar. Weather effects on human mobility: a study using multi-channel sequence analysis. *Computers, Environment and Urban Systems*, 71:131–152, 2018.

[104] Y. Yuan, X. Yang, J. Zhang, D. Song, and X. Yue. Adaptive behavior of intercity travelers within urban agglomeration in response to adverse weather: Accounting for multilayer unobserved heterogeneity. *Transport Policy*, 153:141–158, 2024.

[105] J. Hurley. Snow jam 2014: Planning atlanta, ryan gravel, and a car-dependent region, 2014. Online Resource.

[106] Zhenzhen Yang. Assessing the impacts of rainstorm and flood disaster for improving the resilience of transportation system. *Journal of Advanced Transportation*, 2024(1):6687438, 2024.

[107] Ziyang Wan, Qiuling Lang, Yichen Zhang, Jiquan Zhang, Yanan Chen, Gexu Liu, and Huanan Liu. Improving the resilience of urban transportation to natural disasters: the case of changchun, china. *Scientific Reports*, 15(1):1116, 2025.

[108] Cigdem Varol, Gizem Hayrullahoglu, Emrah Soylemez, Necibe Aydan Sat, Elif Varol, and Nazlı Tunar Ozcan. The movement pattern changes of population following a disaster: Example of the aegean sea earthquake of october 2020. *International Journal of Disaster Risk Reduction*, 112:104743, 2024.

[109] Kunkun Fan, Daichao Li, Xinlei Jin, and Sheng Wu. A multi-scale attributes fusion model for travel mode identification using gps trajectories. *Geo-spatial Information Science*, pages 1–15, 2024.

[110] Seyed Hassan Hosseini, Siavash Pourkhosro, Guido Gentile, and Lory Michelle Bresciani Miristice. Gps-based trip phase and waiting time detection to and from public transport stops via machine learning models. *Transportation Research Procedia*, 78:530–537, 2024.

[111] Sha Huang, Lina Tang, Joseph P Hupy, Yang Wang, and Guofan Shao. A commentary review on the use of normalized difference vegetation index (ndvi) in the era of popular remote sensing. *Journal of Forestry Research*, 32(1):1–6, 2021.

[112] Byungjoon Chang, Woong Seo, and Insung Ihm. On the efficient implementation of a real-time kd-tree construction algorithm. In *GPU Computing and Applications*, pages 207–219. Springer, 2014.

[113] Qing Yu and Jian Yuan. Transbigdata: A python package for transportation spatio-temporal big data processing, analysis and visualization. *Journal of Open Source Software*, 7(71):4021, 2022.

[114] M. Haggag, E. Rezk, and W. El-Dakhakhni. Machine learning prediction of climate-induced disaster injuries. *Natural Hazards*, 116(3):3645–3667, 2023.

[115] N. Pressman and X. Zepic. *Planning in Cold Climates: A Critical Overview of Canadian Settlement Patterns and Policies*. Wiley, 1982.

[116] C. Yang, F. Yan, and S. Zhang. Comparison of land surface and air temperatures for quantifying summer and winter urban heat island in a snow climate city. *Journal of Environmental Management*, 265:110563, July 2020.

[117] S. B. Malevich and K. Klink. Relationships between snow and the wintertime minneapolis urban heat island. *Journal of Applied Meteorology and Climatology*, 50(9):1884–1894, September 2011.

[118] S. Cai, S. S. Ishak, and J. Hu. Assessment of vehicle performance in harsh environments using lsu driving simulator and numerical simulations. Technical report, Southwest Region University Transportation Center (US), December 2015.

[119] S. Stanchi, M. Freppaz, E. Ceaglio, M. Maggioni, K. Meusburger, C. Alewell, and E. Zanini. Soil erosion in an avalanche release site (valle d'aosta: Italy): towards a winter factor for rusle in the alps. *Natural Hazards and Earth System Sciences*, 14(7):1761–1771, July 2014.

[120] Junqing Tang, Song Han, Jing Wang, Baojie He, and Jinhan Peng. A comparative analysis of performance-based resilience metrics via a quantitative-qualitative combined approach: Are we measuring the same thing? *International Journal of Disaster Risk Science*, 14(5):736–750, 2023.

[121] M Bevilacqua, FE Ciarapica, and G Marcucci. A modular analysis for the supply chain resilience triangle. *IFAC-PapersOnLine*, 51(11):1528–1535, 2018.

[122] Ton J Cleophas, Aeilko H Zwinderman, Ton J Cleophas, and Aeilko H Zwinderman. Bayesian pearson correlation analysis. *Modern Bayesian statistics in clinical research*, pages 111–118, 2018.

[123] J. Runge, A. Gerhardus, G. Varando, V. Eyring, and G. Camps-Valls. Causal inference for time series. *Nature Reviews Earth & Environment*, 4(7):487–505, 2023.

[124] Roger A Morbey, Andre Charlett, Daniel Todkill, and Alex J Elliot. Validation of a difference-in-differences investigation tool (did it) for quantifying local outbreaks. *medRxiv*, pages 2024–09, 2024.

[125] Kay H Brodersen, Fabian Gallusser, Jim Koehler, Nicolas Remy, and Steven L Scott. Inferring causal impact using bayesian structural time-series models. 2015.

[126] W. J. Requia and H. F. A. de Melo. Effectiveness of public policies related to traffic emissions in improving air quality in brazil: A causal inference study using bayesian structural time-series models. *Atmospheric Environment*, 319:120291, 2024.

[127] Jonas Wassmer, Bruno Merz, and Norbert Marwan. Resilience of transportation infrastructure networks to road failures. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 34(1), 2024.

[128] G. Dudek. Multilayer perceptron for short-term load forecasting: from global to local approach. *Neural Computing and Applications*, 32(8):3695–3707, 2020.

[129] Z. Mohammadi-Pirouz, K. Hajian-Tilaki, M. Sadeghi Haddat-Zavareh, A. Amoozadeh, and S. Bahrami. Development of decision tree classification algorithms in predicting mortality of covid-19 patients. *International Journal of Emergency Medicine*, 17(1):1–18, 2024.

[130] Zhihui Lai, Guangfei Liang, Jie Zhou, Heng Kong, and Yuwu Lu. A joint learning framework for optimal feature extraction and multi-class svm. *Information Sciences*, 671:120656, 2024.

[131] Sancho Salcedo-Sanz, José Luis Rojo-Álvarez, Manel Martínez-Ramón, and Gustavo Camps-Valls. Support vector machines in engineering: an overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 4(3):234–267, 2014.

[132] John Hancock and Taghi M Khoshgoftaar. Leveraging lightgbm for categorical big data. In *2021 IEEE Seventh International Conference on Big Data Computing Service and Applications (BigDataService)*, pages 149–154. IEEE, 2021.

[133] Mengran Zhu, Ye Zhang, Yulu Gong, Kaijuan Xing, Xu Yan, and Jintong Song. Ensemble methodology: Innovations in credit default prediction using lightgbm, xgboost, and localensemble. In *2024 IEEE 4th International Conference on Electronic Technology, Communication and Information (ICETCI)*, pages 421–426. IEEE, 2024.

[134] K. Abraham, M. Abdelwahab, and M. Abo-Zahhad. Classification and detection of natural disasters using machine learning and deep learning techniques: A review. *Earth Science Informatics*, 17(2):869–891, 2024.

[135] M. T. Ribeiro, S. Singh, and C. Guestrin. "why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144, 2016.

[136] L. Xiao, S. Lo, J. Liu, J. Zhou, and Q. Li. Nonlinear and synergistic effects of tod on urban vibrancy: Applying local explanations for gradient boosting decision tree. *Sustainable Cities and Society*, 72:103063, 2021.

[137] A. S. Albahri, Y. L. Khaleel, M. A. Habeeb, R. D. Ismael, Q. A. Hameed, M. Deveci, and L. Alzubaidi. A systematic review of trustworthy artificial intelligence applications in natural disasters. *Computers and Electrical Engineering*, 118:109409, 2024.

[138] M. Lichter, A. T. Vafeidis, R. J. Nicholls, and G. Kaiser. Exploring data-related uncertainties in analyses of land area and population in the "low-elevation coastal zone" (lecz). *Journal of Coastal Research*, 27(4):757–768, 2011.

[139] H. U. Liwei, H. E. Yu, H. O. U. Zhi, Z. H. A. N. G. Ruijie, C. H. E. N. Chen, and L. I. U. Bing. Multi-dimensional coupling study on traffic accident risk of highway in mountainous areas. *China Safety Science Journal*, 34(5):17, 2024.

[140] Michal Lichter, Athanasios T Vafeidis, Robert J Nicholls, and Gunilla Kaiser. Exploring data-related uncertainties in analyses of land area and population in the "low-elevation coastal zone"(lecz). *Journal of Coastal Research*, 27(4):757–768, 2011.

[141] HU Liwei, HE Yu, HOU Zhi, ZHANG Ruijie, CHEN Chen, and LIU Bing. Multi-dimensional coupling study on traffic accident risk of highway in mountainous areas. *China Safety Science Journal*, 34(5):17, 2024.

[142] Feng Wang, Xinrang Wang, and Sai Li. Explainable machine learning for predictive modeling of blowing snow detection and meteorological feature assessment using xgboost-shap. *PloS one*, 20(3):e0318835, 2025.

[143] Muhammad Sakib Khan Inan and Istiakur Rahman. Explainable ai integrated feature selection for landslide susceptibility mapping using treeshap. *SN Computer Science*, 4(5):482, 2023.

[144] Mario Di Bacco, James H Williams, Daisuke Sugawara, and Anna Rita Scorzini. Towards multi-variable tsunami damage modeling for coastal roads: Insights from the application of explainable machine learning to the 2011 great east japan event. *Sustainable Cities and Society*, 115:105856, 2024.

[145] Xiaojian Zhang, Xilei Zhao, Dare Baldwin, Sara McBride, Josephine Bellizzi, Elizabeth S Cochran, Nicholas Luco, Matthew Wood, and Thomas J Cova. Modeling protective action decision-making in earthquakes by using explainable machine learning and video data. *Scientific reports*, 14(1):5480, 2024.