| | |
|---|---|
| Title | Landslide Risk Assessment and Bridge Disaster Management Using Machine Learning and Deep Learning Models |
| Author(s) | 任，福東 |
| Degree Grantor | 北海道大学 |
| Degree Name | 博士(工学) |
| Dissertation Number | 甲第16343号 |
| Issue Date | 2025-03-25 |
| DOI | https://doi.org/10.14943/doctoral.k16343 |
| Doc URL | https://hdl.handle.net/2115/95010 |
| Type | doctoral thesis |
| File Information | Ren_Fudong.pdf |

# Landslide Risk Assessment and Bridge Disaster Management Using Machine Learning and Deep Learning Models

By

**Fudong REN**

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Engineering

Supervision and Guidance: **Prof. Koichi ISOBE**

Examination committee:

**Prof. KOICHI ISOBE**

**Prof. YOICHI WATABE**

**Prof. SATOSHI NISHIMURA**

**Prof. TOMOHITO YAMADA**

Thesis No.

Division of Field Engineering for Environment
Graduate School of Engineering, Hokkaido University

March 2025

**Contents**

1.Introduction........................................................................................................1
1.1. Research Background...........................................................................1
1.2. Previous Research ...................................................................................3
1.2.1. Landslide Risk Assessment Based on Machine Learning...........................3
1.2.2. Landslide Risk Assessment Based on CNN and sampling method ............4
1.2.3. Bridge Risk Assessment with DGP and DCA ............................................5
1.3. Research Aim ...........................................................................................5
2.Regional Dependence Evaluation of Landslide Risk Using Machine Learning.........7
2.1. Purpose ......................................................................................................7
2.2. Study Area and Data...................................................................................7
2.2.1. Study Area...............................................................................................7
2.2.2. Historical Landslide Records (HLS)..........................................................8
2.2.3. Data-use ...................................................................................................9
2.4. Methodology ............................................................................................13
2.4.1. Workflow Overview..................................................................................13
2.4.2. Machine Learning Models ........................................................................14
2.4.3. Evaluation Method....................................................................................17
2.4.4. Information Gain Ratio (IGR) Method .......................................................19
2.5. Results ......................................................................................................19
2.5.1. Importance of Features ............................................................................19
2.5.2. Result of Evaluation..................................................................................20
2.6. Discussion and conclusion .......................................................................22
3. Landslide Risk Assessment Using CNN with Different Window Size and Sampling Method...........................................................................................................23
3.1. Research Purpose .....................................................................................23
3.2. Study Area and Data..................................................................................24
3.3. Method......................................................................................................24
3.3.1. CNN ........................................................................................................25
3.3.2. CNN Windows Sizes.................................................................................27
3.3.3. Sampling Techniques ...............................................................................27
3.3.4. Evaluation Method....................................................................................28
3.4. Results and Discussion..............................................................................29
3.4.1. Window Size ............................................................................................29
3.4.2. Sampling Method......................................................................................32
3.4.3. Influence of Window Size on False Negatives and False Positives............37
3.4.4. Visualization.............................................................................................39
3.4.5. Versatility ................................................................................................48

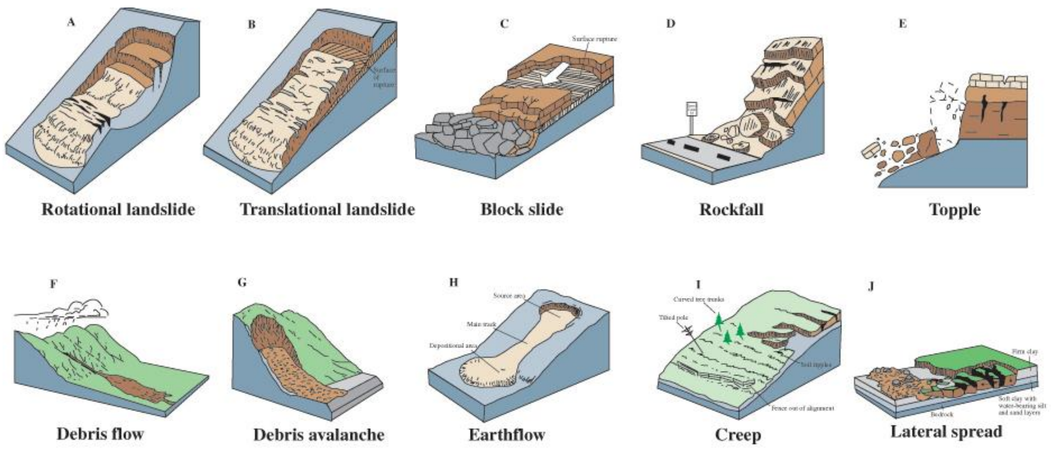
3.5. Conclusion........................................................................................50
3.6. Future works........................................................................................51
4. Bridge Risk Management Based on Machine Learning with Driftwood Generation Potential by Deep Learning ........................................................................54
4.1. Research Purpose ........................................................................................54
4.2. Study Area and Data........................................................................................54
4.2.1. Study Area........................................................................................54
4.2.2. Data Used in This Study ........................................................................55
4.3. Method........................................................................................59
4.3.1. Landslide Risk Assessment........................................................................59
4.3.2. Driftwood Generation Potential and Bridge Risk........................................59
4.4 Result and Discussion ........................................................................................63
4.4.1. Feature Importance ........................................................................................63
4.4.2. Landslide Risk Assessment........................................................................64
4.4.3. Driftwood Generation Potential (DGP) and Driftwood Capture Amount (DCA) ........................................................................................66
4.4.4. Bridge Disaster Risk Assessment........................................................68
4.5. Conclusion........................................................................................72
4.6. Future works........................................................................................73
5. Summary and Conclusion........................................................................................75
Acknowledgements........................................................................................78
References........................................................................................79

# 1.Introduction

## 1.1. Research Background

Landslides are natural phenomena characterized by the downward movement of soil or rock masses on a slope, often posing a significant threat to human lives, infrastructure, and the environment. These events are influenced by various internal factors such as topography, lithology, and geological structure, as well as external factors including earthquakes, precipitation, and human activities like slope excavation, water storage, and drainage. The occurrence of landslides has led to substantial losses, impacting industrial and agricultural production, and causing harm to individuals and property. Globally, thousands of casualties have been reported, and numerous homes and structures destroyed by landslides.

To illustrate, in 2018, the landslide in Shum Shui Po, Hong Kong, caused widespread property damage due to prolonged rainfall, demonstrating the interplay between precipitation and slope stability. Similarly, the 2014 Oso landslide in Washington State, USA, resulted in 43 fatalities, emphasizing the devastating human and economic impacts of landslides. Understanding the mechanisms behind these events is crucial for prevention and mitigation. Figure 1 depicts the major types of landslide movements and provides examples of landslide events.




(a) The major types of landslide movements

(b) Examples of landslide occurrence in CHINA

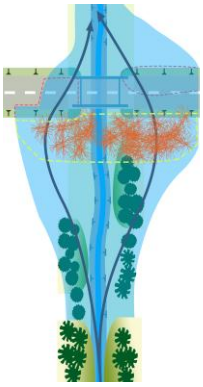
**Figure 1.** The major types of landslide movements and examples of landslide occurrence.

Creating landslide susceptibility maps has become a priority for ensuring the sustainability of human-environmental, cultural, economic, and social systems. Such assessments integrate geological, hydrological, and anthropogenic factors, enabling the identification of high-risk zones and the formulation of strategies to mitigate landslide risks.

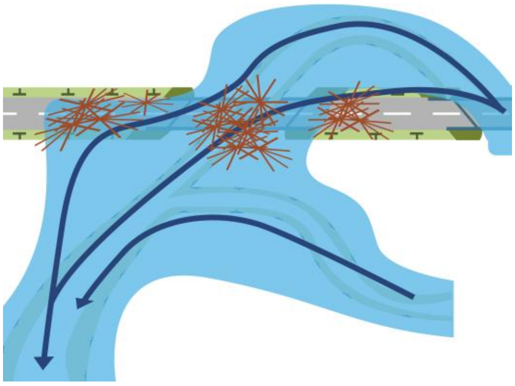
Rainfall is one of the primary triggers of landslides. In recent years, rainfall patterns have become increasingly concentrated and intense, a trend projected to continue under changing climatic conditions. Heavy rainfall saturates slopes, weakening the soil's cohesive strength and dramatically increasing landslide risks. Moreover, this intensified rainfall contributes to the rise of river flows, transporting sediment and driftwood downstream. These materials, especially driftwood, frequently become trapped at bridges, obstructing river channels and exacerbating upstream flooding and erosion.

Establishing a rational and efficient method for predicting the volume of driftwood that may accumulate at river bridges has become a pressing challenge. Research into the mechanisms of driftwood generation, its movement downstream, and its deposition in waterways is advancing. In Japan, for instance, the aftermath of Typhoon Hagibis in 2019 underscored the importance of effective driftwood management, as massive accumulations of driftwood caused significant infrastructure damage. Figure 2 illustrates the causes of driftwood and forms of bridge damage due to driftwood.

In response, the Erosion and Sediment Control Department of the Ministry of Land, Infrastructure, Transport, and Tourism issued a directive in August 2023, emphasizing the urgent need for comprehensive driftwood management strategies[1]. These efforts aim to mitigate the cascading effects of extreme weather events, safeguarding communities and critical infrastructure.



(a) River channel blockage due to driftwood and embankment erosion behind the abutment



(b) Subsidence of bridge piers due to riverbed scouring and channel blockage from driftwood

**Figure 2.** Bridge damage caused by driftwood, as reported in survey findings from the Civil Engineering Bureau, Road Division, Hokkaido Prefectural Government.

(a) Kamimemuro Bridge

(b) Simoshintoku River JR Bridge

**Figure 3.** Examples of Bridge damage occurrence caused by driftwood.

## 1.2. Previous Research

### 1.2.1. Landslide Risk Assessment Based on Machine Learning

Chen's doctoral dissertation explores the application of artificial intelligence (AI) in evaluating earthquake-induced landslide susceptibility, integrating geological and seismic datasets to enhance model performance. By leveraging domain-specific factors such as seismic intensity, the study demonstrates significant improvements in predictive accuracy over traditional methods[2]. Similarly, Lee et al. employed Support Vector Machines (SVMs) to map landslide susceptibility in Gangwon Province, Korea. Their research, using diverse datasets including topographical, hydrological, and geological variables, highlights the capability of SVMs to manage nonlinear relationships and high-dimensional data[3]. Lin et al. combined SVMs with the Newmark displacement model to assess earthquake-triggered landslides, successfully merging machine learning's predictive strengths with physical insights into slope stability under seismic loading[4].

Mondal and Mandal utilized logistic regression for landslide susceptibility mapping in the Balason River basin of the Darjeeling Himalaya. This research analyzed factors such as slope, land use, and precipitation, illustrating that simpler statistical models can still provide robust results[5]. Kadavi et al. compared logistic regression and decision tree models for mapping susceptibility in Gangwon-do, South Korea, noting the trade-offs between interpretability and the ability to handle nonlinear relationships[6]. Addressing data imbalance in susceptibility datasets, Sameen et al. proposed a systematic sample subdividing strategy, achieving enhanced predictive accuracy and generalizability[7]. Adnan et al. advanced the field by combining multiple machine learning algorithms and optimizing ensemble techniques, improving spatial agreement in susceptibility predictions[8]. Kawabata and Bandibas applied Artificial Neural Networks (ANNs) for mapping, emphasizing the critical role of high-quality input data in ensuring model reliability[9].

These studies collectively reveal a transition from traditional statistical approaches to advanced AI techniques, such as SVMs and ANNs, while also showcasing hybrid methods that combine machine learning with physical models. Despite their predictive superiority, AI models demand extensive data preparation and parameter tuning, whereas statistical models remain valuable for their simplicity and interpretability.

### 1.2.2. Landslide Risk Assessment Based on CNN and sampling method

Sameen et al. addressed the challenges of data imbalance in landslide susceptibility datasets, demonstrating the benefits of a systematic sample subdividing strategy in improving model performance[7]. Wei et al. introduced a feature enhancement framework for landslide detection using remote sensing data, significantly improving detection accuracy through advanced preprocessing techniques[10]. Wei et al. integrated spatial response features with machine learning classifiers, such as random forests and support vector machines, achieving notable improvements in susceptibility mapping[11]. He et al. developed a unified network incorporating spatial neighborhood information and superimposed landslide factors, providing a comprehensive view that enhanced mapping reliability[12].

Jiang et al. applied convolutional neural networks (CNNs) to high-resolution optical images for landslide detection and segmentation along the Sichuan-Tibet corridor, achieving superior performance compared to traditional methods[13]. Jiang et al. further demonstrated the advantages of CNNs in capturing complex patterns and spatial relationships, outperforming conventional machine learning techniques[14]. Hussain et al. evaluated deep learning and machine learning models, emphasizing CNNs' strength in handling high-dimensional datasets and extracting spatial features[15]. Similarly, Ghorbanzadeh et al. compared various methods and confirmed CNNs’ potential for large-scale, high-resolution susceptibility mapping[16]. Liu et al. conducted a comparative study, highlighting CNNs' superior performance in capturing spatial correlations and high-level features[17].

These studies underscore the rapid advancements in CNN-based methods for landslide detection and susceptibility mapping. CNNs demonstrate consistently higher predictive accuracy and efficiency, albeit with challenges related to data quality and interpretability.

#### 1.2.3. Bridge Risk Assessment with DGP and DCA

Minami et al. investigated a woody debris disaster in the Yosasa River, Tochigi Prefecture, analyzing its mechanisms and emphasizing preventive measures for river management[18]. Yamada et al. provided a quantitative analysis of damage caused by Typhoon Etau, offering valuable data for riparian forest management[19]. Yano et al. assessed debris disaster risk at bridges along the Kagetsu River, introducing a novel approach to quantifying debris risks at infrastructure sites[20]. Dozono further developed a comprehensive risk assessment framework for driftwood disasters at bridges, emphasizing the importance of understanding driftwood generation potential[21].

Bangnira et al. identified bridges prone to instream wood accumulation in the UK, advocating for integrated hydrological and structural analyses to address debris risks[22]. Yano et al. analyzed the causes of driftwood generation during the Northern Kyushu heavy rain event, providing actionable recommendations for preparedness[23]. Kakihara et al. constructed a topography-based model to estimate driftwood risk, offering insights into spatial risk distribution[24]. Tanaka and Nagashima employed terrain analysis to assess woody debris risks, highlighting the significance of slope stability in mitigating debris disasters[25].

These studies highlight the necessity of integrating hydrological, geomorphological, and infrastructural analyses for managing driftwood and woody debris risks. Future research should aim to refine predictive models and foster cross-disciplinary collaboration to enhance disaster mitigation efforts.

### 1.3. Research Aim

This study aims to improve the predictive accuracy of landslide risk assessments and extend the scope of research to driftwood management. Leveraging advanced machine learning techniques and GIS-based analyses, the research integrates multiple methodological components to construct a comprehensive framework addressing both landslide and driftwood hazards.

The foundation of this study lies in developing detailed spatial databases that incorporate geological, topographical, and climatic factors along with historical landslide records. These databases enable the application of traditional machine learning models, such as logistic regression, decision trees, and random forests, to generate initial landslide susceptibility maps (LSMs). These maps establish a baseline for understanding landslide-prone areas and set the stage for further advancements.

Building on this groundwork, the study progresses to implementing convolutional neural networks (CNNs) to overcome the limitations of traditional models. CNNs offer superior performance by effectively analyzing spatial features and addressing the challenges of varying landslide scales. By employing different window sizes, the study identifies optimal configurations, with larger window sizes like 19×19 pixels proving particularly effective in capturing critical contextual information. Furthermore, the use of downsampling and oversampling techniques ensures robust training and balanced datasets, enhancing the overall reliability of CNN models.

The research then extends the application of CNN-generated LSMs to evaluate driftwood generation potential. By identifying high-risk landslide zones with steep slopes exceeding 30 degrees, the study estimates driftwood volumes and assesses their impact on riverine and bridge infrastructures. This phase integrates environmental data with structural characteristics, enabling precise predictions of risks posed by driftwood accumulation during extreme weather events.

The relationship between the study's components is critical. The transition from traditional machine learning models in the initial phase to advanced CNN models highlights the evolution in predictive capabilities. Insights from CNN-based landslide risk assessments inform the subsequent analysis of driftwood risks, demonstrating the adaptability of the methodologies to interconnected environmental challenges. This integration ensures a holistic approach to disaster risk management.

By connecting these parts, the study not only refines landslide risk prediction but also offers actionable solutions for mitigating associated hazards like driftwood. The findings contribute to safer infrastructure planning and enhanced disaster resilience, particularly in regions vulnerable to natural hazards.

# 2.Regional Dependence Evaluation of Landslide Risk Using Machine Learning

## 2.1. Purpose

This study aims to advance the evaluation of landslide hazards by integrating GIS technology with machine learning techniques, while also examining how factors influencing landslides vary across different regions. The research primarily focuses on historical landslides (HLS). To achieve this, the study pursues two specific goals.

The first goal is the construction of a comprehensive spatial database for geological disasters within the study regions. This database integrates a broad range of factors such as topographic features, geological structures, land use patterns, and climatic conditions, providing a robust foundation for landslide risk analysis. By systematically organizing these factors, the database aims to enhance the understanding of landslide-prone areas and their underlying mechanisms.

The second goal involves applying advanced Python-based machine learning models to generate accurate landslide susceptibility maps (LSMs). These maps identify areas at higher risk of future landslides, offering actionable insights for risk mitigation and disaster management. By leveraging the predictive power of machine learning, the study ensures adaptability to varying regional conditions and improved accuracy in assessing landslide susceptibility.

## 2.2. Study Area and Data

### 2.2.1. Study Area

The primary objective of this study is to utilize Convolutional Neural Networks (CNN) to assess the risk of landslide disasters in Japan. Three regions were selected based on specific criteria: Niigata Prefecture (NIG area), Iwate and Miyagi Prefectures (IWT-MYG area), and Hokkaido (HKD area). These areas, as illustrated in figure 4, were chosen due to their historical susceptibility to landslides, availability of comprehensive data sets, and their geographic diversity, which includes varied topographical and climatic conditions. Additionally, these regions have experienced significant earthquakes in the past, accompanied by numerous earthquake-induced landslides, making them pertinent study areas for future research on earthquake-triggered landslides. This selection is intended to ensure that the findings are robust and applicable across different landscapes, thus enhancing the study’s relevance and applicability to landslide risk assessment.

2.2.1.1. Niigata Prefecture (NIG area)

The first study area shown figure 4 (a) is Tokamachi, Kawaguchi, Horinouchi, Ojiya, Yamakoshi and Nagaoka in Niigata Prefecture, located in the central Japan. The latitude and longitude range from 138° 40.019' E to 138° 58.08' E and 37° 0.6048' N to 37° 35.2644' N. The total area of the area is 1795.36 $km^2$. The elevation range of the study area is from 14 to 1021.5 m, while the slope range is from 0 to 78.01 degrees.

2.2.1.2. Iwate Prefecture, Miyagi Prefecture (IWT-MYG area)

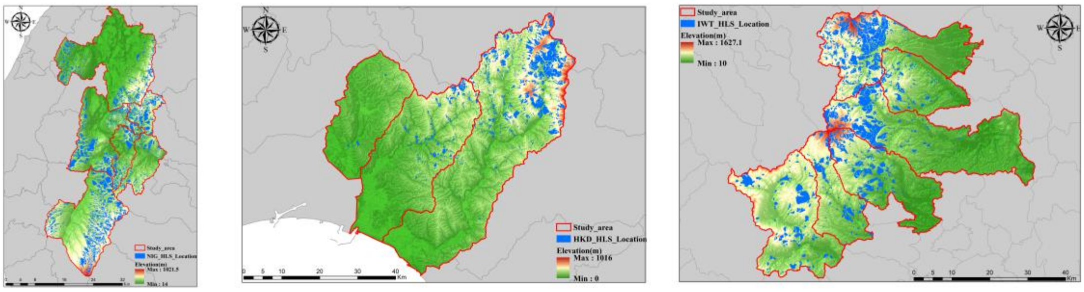
The second study area shown in figure 4 (b) is Ichinoseki, Oshu and the surrounding area of Iwate Prefecture, and Kurihara and the surrounding area of Miyagi Prefecture located in northeastern Japan. The latitude and longitude range from 140° 40.019' E to 141° 15.888' E and 38° 41.1228' N to 39° 10.869' N. The total area is 2447.05 $km^2$. The elevation range of the study area is from 10 to 1627.1 m, while the slope range is from 0 to 71.32 degrees. In the following, "IWT-MYG" is abbreviated as "IWT."

2.2.1.3. Hokkaido (HKD area)

The third study area shown in figure 4 (c) is Mukawa, Abira, and Atsuma in Hokkaido in the northern part of Japan, with the longitude and latitude range from 41° 44' 11.6 "E to 142° 20' 2.4 "E and 42° 31' 42.6 "N to 42° 59' 13.7 "N, respectively. The total area is 1353 $km^2$. The study area's elevation ranges from 0 to 1016 m, while the slope ranges from 0 to 76.05 degrees.

2.2.2. Historical Landslide Records (HLS)

The landslide data utilized in this research were sourced from the National Research Institute for Earth Science and Disaster Resilience, JAPAN. The landslide topographic distribution map [26] was created through the three-dimensional interpretation of 1:40,000 monochrome aerial photographs captured in the 1970s. This map delineates cliffs and mobile objects, showcasing their spatial distribution on a 1:50,000 topographic map. Through this landslide topographic distribution map, we gain insights into the historical location, magnitude, and alterations of landslides. Refer to figure 4 for visualization.




(a) NIG (b) HKD (c) IWT-MYG

**Figure 4.** Study area and distributions of historical landslides; (a) NIG study area, (b) HKD study area and (c) IWT-MYG study area.

### 2.2.3. Data-use

Landslides can arise from a multitude of factors, each interconnected in complex ways. Through the examination of landslide causation, certain factors have emerged as particularly relevant for research purposes.

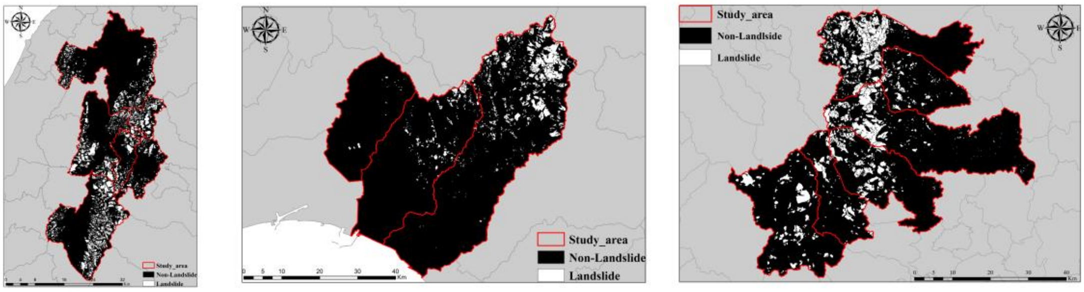
Precipitation's role in landslide occurrence has been extensively studied worldwide over many years, and this study similarly acknowledges its significance as a key influencing factor. Due to the absence of specific occurrence times for historical landslide data in the study area, the assessment of past landslides initially utilized the average annual precipitation over the past 30 years as a reference.

Considering insights from prior research and a diverse range of data sources, this study identifies 15 factors associated with historical landslides: elevation, slope, aspect, curvature, lithology [27] , soil type [28], land use[29], precipitation [30, 31], distance from rivers[32], faults [33], roads [29], vegetation [34], catchment area[35], and openness [36]. These factors are depicted in figure 5. Table 1 illustrates the range of the dataset. Furthermore, the Digital Elevation Model (DEM) used in this study has a resolution of 10 meters, and all landslide records and associated feature resolutions prepared are consistent with the resolution of the DEM.

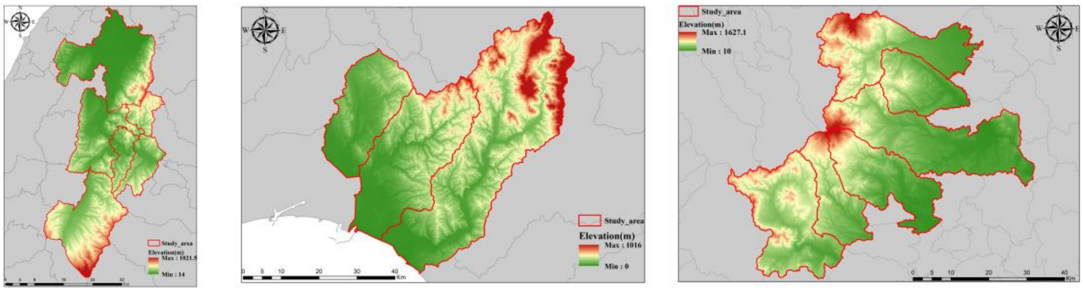
**Table 1.** Features and data range in the study area

| Code | Feature | Resource | NIG | HKD | IWT |
|---|---|---|---|---|---|
| A0 | HLS | NIED | 0,1 | 0,1 | 0,1 |
| A1 | DEM(m) | GIAJ | 14 to 1021.5 | 0 to 1016 | 10 to 1627.1 |
| A2 | Slope (°) | DEM | 0 to 78.01 | 0 to 76.05 | 0 to 71.32 |
| A3 | Aspect | DEM | | -1 to 360 | |
| A4 | Curvature | DEM | -84.3 to 115.5 | -63.0 to 57.0 | -80.5 to 76.6 |
| A5 | Rain(mm) | JMA | 2185.37 to 2795.37 | 1060.82 to 1215.38 | 1222.54 to 2130.98 |
| A6 | River(m) | MLIT | 0 to 2612.16 | 0 to 3560.9 | 0 to 4909.76 |
| A7 | Road(m) | MLIT | 0 to 6402.81 | 0 to 16048.8 | 0 to 23740.1 |
| A8 | Fault(m) | GSJ | 0 to 22333.4 | 0 to 16798.8 | 0 to 14769.6 |
| A9 | Landuse | MLIT | | 11 types | |
| A10 | Soil | MLIT | | 37 types | |
| A11 | Lithology | GSJ | | 27 types | |
| A12 | Vegetation | BiodiC-J | | 33 types | |
| A13 | Catchment Area | DEM | 1 to 6359 | 1 to 13026 | 1 to 13460 |
| A14 | Openness-Down | DEM | 0 to 88.06 | 4.04 to 87.87 | 0 to 87.6 |
| A15 | Openness-Up | DEM | 0 to 77.96 | 4.04 to 76.09 | 0 to 74.78 |

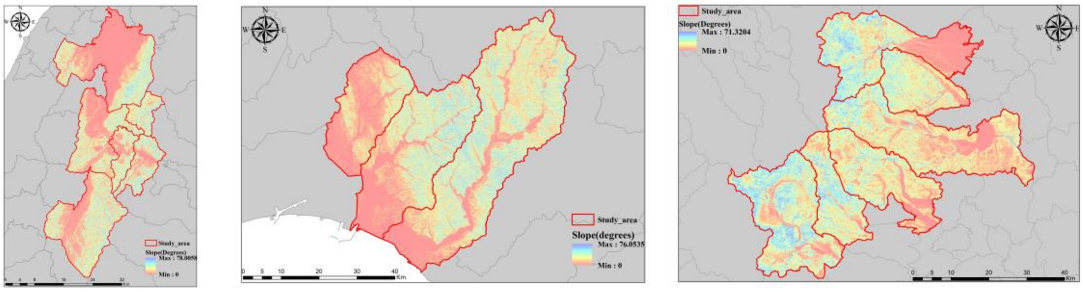
[1] NIED: National Research Institute for Earth Science and Disaster Resilience.
[2] GIAJ: Geospatial Information Authority of Japan.
[3] DEM: Digital Elevation Map with 10 m x 10 m resolution published by Geospatial Information Authority of Japan.
[4] JMA: Japan Meteorological Agency.
[5] MLIT: Ministry of Land, Infrastructure, Transport and Tourism.
[6] GSJ: Geological Survey of Japan, National Institute of Advanced Industrial Science and Technology.
[7] BiodiC-J: Biodiversity Center of Japan.



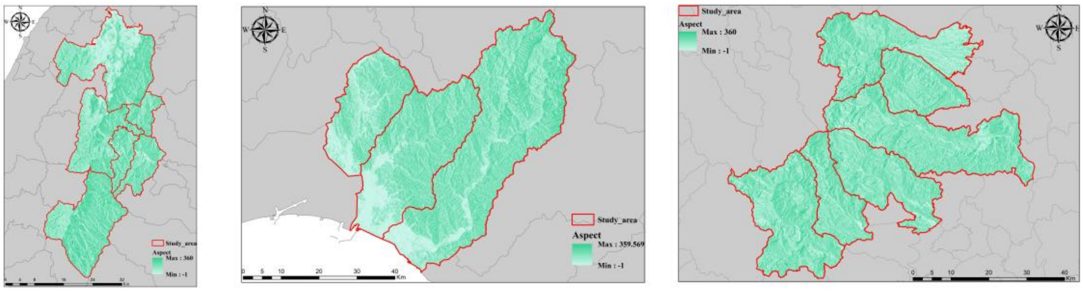

(**a**) HLS



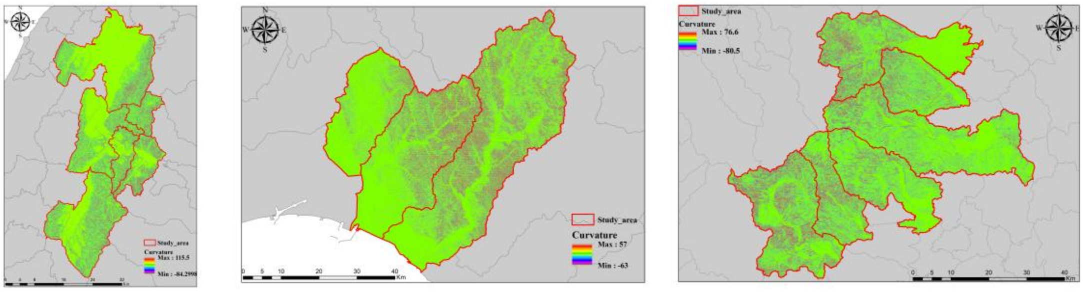

(**b**) Elevation (m)



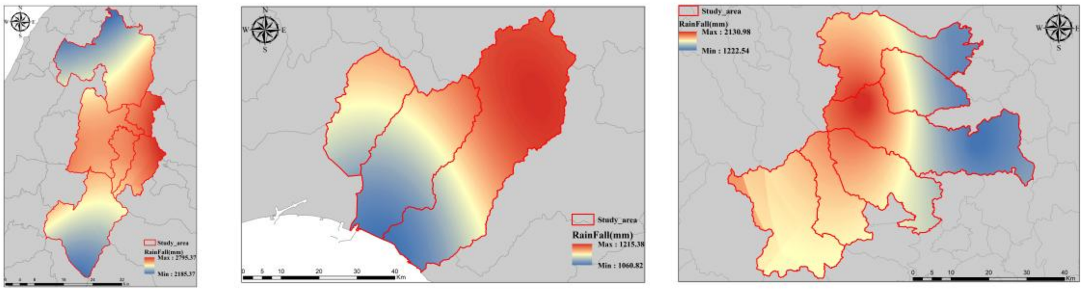
(**c**) Slope (degrees)



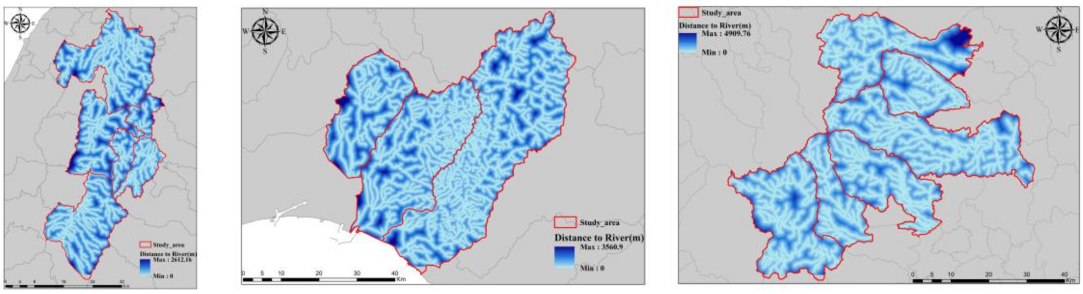
(**d**) Aspect



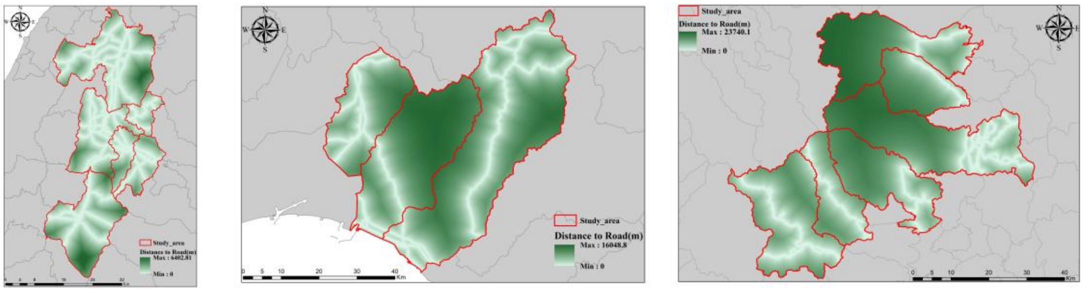
(**e**) Curvature



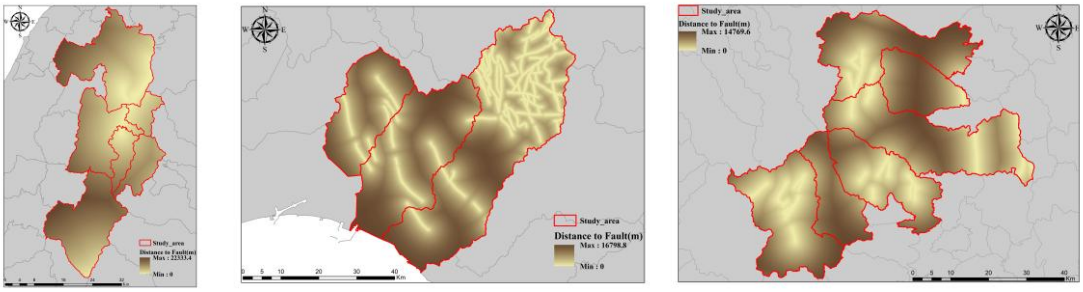
(**f**) Average rainfall (mm)



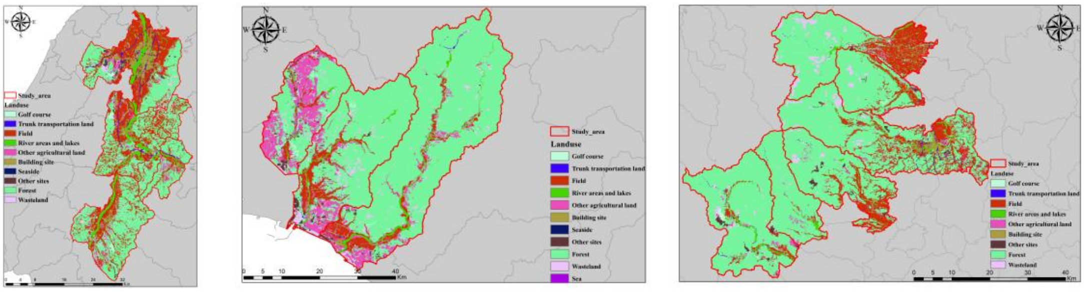
(**g**) Distance to rivers (m)



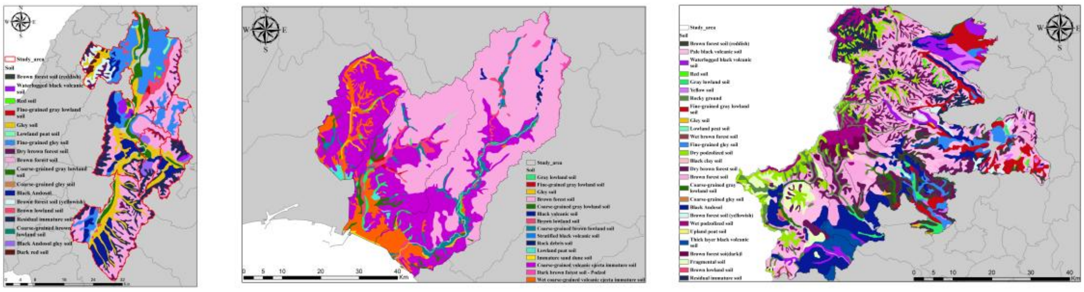
(**h**) Distance to roads (m)



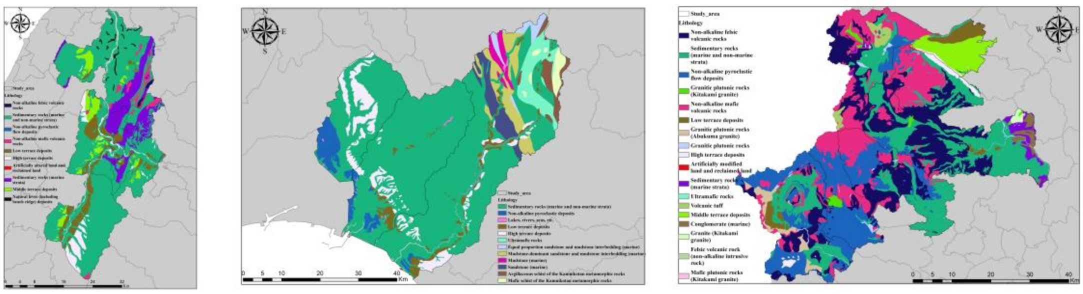
(**i**) Distance to faults (m)



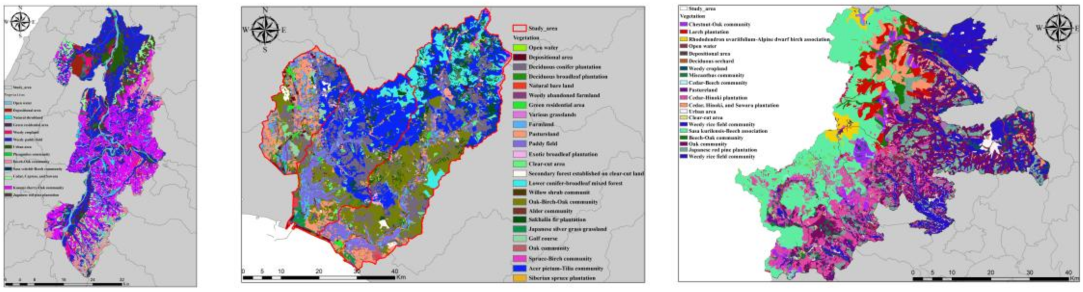
(j) Landuse



(k) Soil



(l) Lithology



(m) Vegetation

**Figure 5.** Maps of landslide features (NIG, HKD and IWT-MYG).

## 2.4. Methodology

### 2.4.1. Workflow Overview

The methodological framework combined GIS data processing with machine learning modeling to create landslide susceptibility maps and evaluate factor importance. Key steps show as figure 6.

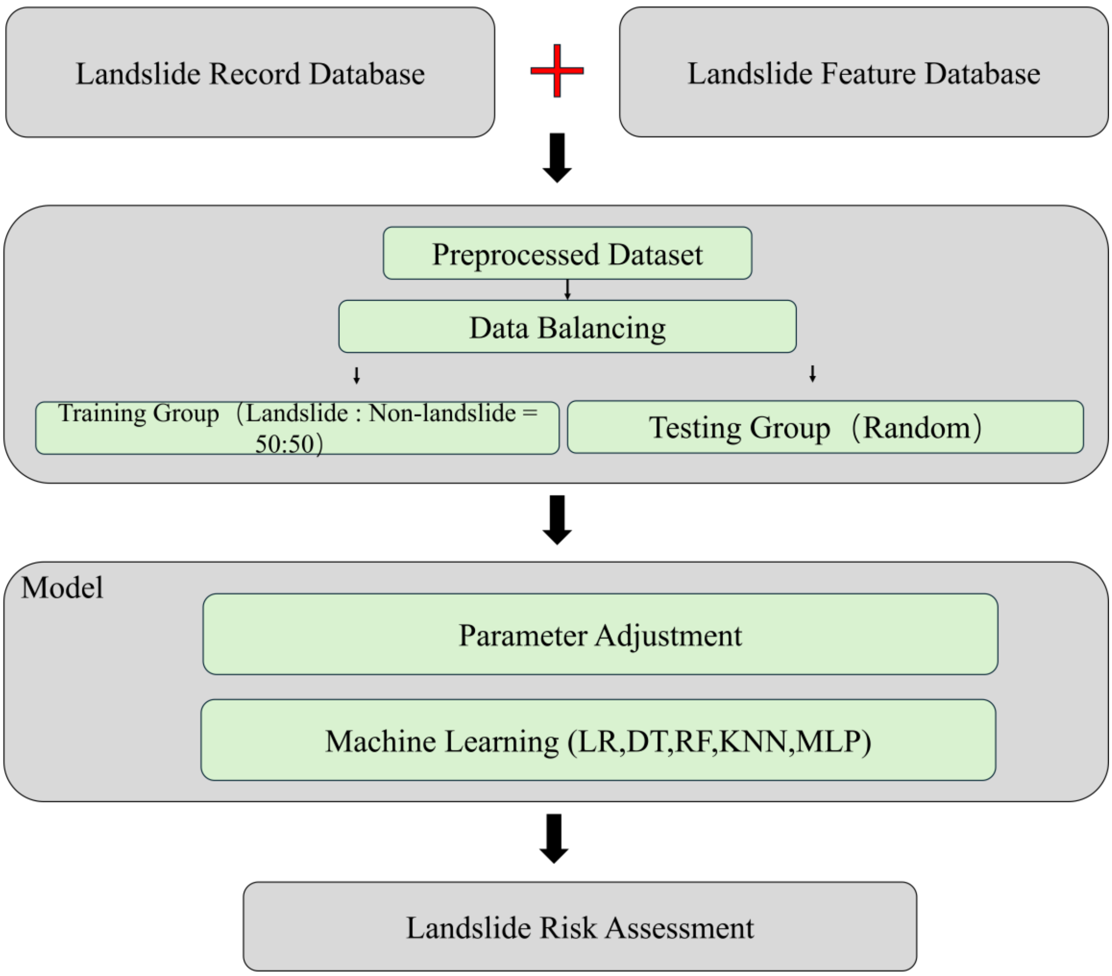



**Figure 6.** Flowchart of this study.

#### 2.4.2. Machine Learning Models

##### 2.4.2.1. Logistic Regression

Logistic regression is an extremely easy to understand model that is equivalent to $y = f(x)$, showing the relationship between the independent variable x and the dependent variable y. The Logistic regression (LR) prediction equation is shown below:

$$P(Y = 1|x) = \frac{1}{1 + e^{-(w^T x + b)}} \tag{1}$$

In other words, the logarithmic probability of output $Y = 1$ is a model represented by a linear function of input x, which is a logistic regression model. When the value of $(w^T x + b)$ is closer to positive infinity, the probability value of $P(Y = 1|x)$is closer to 1. Therefore, the idea of logistic regression is to first fit the decision boundary (not limited to linear, but also polynomial), and then establish the relationship between the boundary and the probability of classification, to obtain the probability in the case of two classifications.

One of the main advantages of logistic regression is its interpretability, allowing us to clearly understand the contribution of each independent variable to the outcome. The logistic regression model also has a low computational cost, making it ideal for preliminary analysis on large-scale datasets and especially for classification problems that require quick and easily interpretable results.

##### 2.4.2.2. Decision Tree (DT) and Random Forest (RF)

The decision tree model is a tree structure consisting of nodes that represent features and corresponding decision rules. It involves classifying or regressing instances by recursively dividing data into sub-regions based on features, continuing until a condition is met and forming leaf nodes. Decision trees are particularly effective in modeling non-linear relationships by creating complex decision boundaries through recursive partitioning, allowing them to adapt to non-linear data distributions without making strong assumptions about the underlying relationships.

Random forest is an ensemble learning algorithm that integrates multiple decision trees. Each decision tree acts as a classifier, and for a given input, multiple trees produce classification results. The random forest aggregates these votes and assigns the most common category as the final output. By combining multiple decision trees, random forests enhance the stability and accuracy of the model, making them well-suited for high-dimensional data and complex classification tasks. These models are selected for their flexibility, high performance, and ability to effectively capture non-linear relationships.

2.4.2.3. KNN

KNN (K-Nearest Neighbor) is one of the simplest machine learning algorithms, which can be used for classification and regression, and is a supervised learning algorithm. It is based on the idea that if most of the K most similar (i.e., most neighboring) samples in the feature space belong to a particular class, then that sample also belongs to that class. In other words, this method only determines the category of the sample to be classified according to the category of the nearest one or several samples in the classification decision. The mathematical description of KNN is as follows：

$$X = (x_1, x_2, x_3, \ldots, x_n), Y = (y_1, y_2, y_3, \ldots, y_n)$$

$$\left(\sum_{i=1}^{n} |x_i - y_i|^p\right)^{\frac{1}{p}} \tag{2}$$

The *n* means feature numbers, *p* normally equals 1 (Manhattan distance), 2 (Euclidean distance) and ∞(Chebyshev distance).

A major advantage of KNN is that it does not require any assumptions and directly classified based on the category of the nearest samples. This makes it highly adaptable in scenarios involving small sample sizes and multi-class problems. The reason for choosing KNN is its straightforward approach, which is well-suited for preliminary data exploration and handling unstructured data that is difficult to model.

2.4.2.4. MLP

Artificial neural networks are divided into multiple layers and single layers. Each layer contains several neurons. Each neuron is connected by a direct arc with variable weights. The network is trained through repeated learning and training of known information, and through gradual adjustments. The method of changing the weight of neuron connection can achieve the purpose of processing information and simulating the relationship between input and output.

It does not need to know the exact relationship between input and output and does not need many parameters. It only needs to know the non-constant factors that cause output changes, that is, non-constant parameters. Therefore, compared with traditional data processing methods, neural network technology has obvious advantages in processing fuzzy data, random data, and nonlinear data. It is especially suitable for systems with large scales, complex structures and unclear information. An ANN with multiple hidden layers is called a multilayer perceptron (MLP). The basic mathematical expression of neuron is:

$$y = f\left(\sum_{i=1}^{n} w_n x_n + b\right) \tag{3}$$

Suppose $W = [w_1, w_2, w_3, \dots, w_n]$ is the weight vector, and $X = [x_1, x_2, x_3, \dots, x_n]$ is the input vector.

$$y = f(W^T X + b) \tag{4}$$

The network can learn and train on known information repeatedly, making it highly effective in processing dynamic and complex data relationships. The reason for choosing MLP is its powerful modeling capability, which allows it to perform exceptionally well in complex nonlinear problems.

2.4.2.5. Machine Learning Strategy

When training with datasets, it's common to split the data into two parts: a training set and a test set. The training set is used to train model parameters, while the test set evaluates the model's performance on unseen data. This separation helps ensure the model's effectiveness.

However, models often perform well on training data but less so on test data. To address this, cross-validation is used. This method involves dividing the initial dataset into K sub-samples, using one for validation and the rest for training. This process is repeated K-times, rotating the validation subset each time. The results are averaged to provide a robust estimate of model performance. The K-fold cross-validation, particularly 5-fold and 10-fold, is widely used due to its efficiency in utilizing data for both training and validation as shown in figure 7.

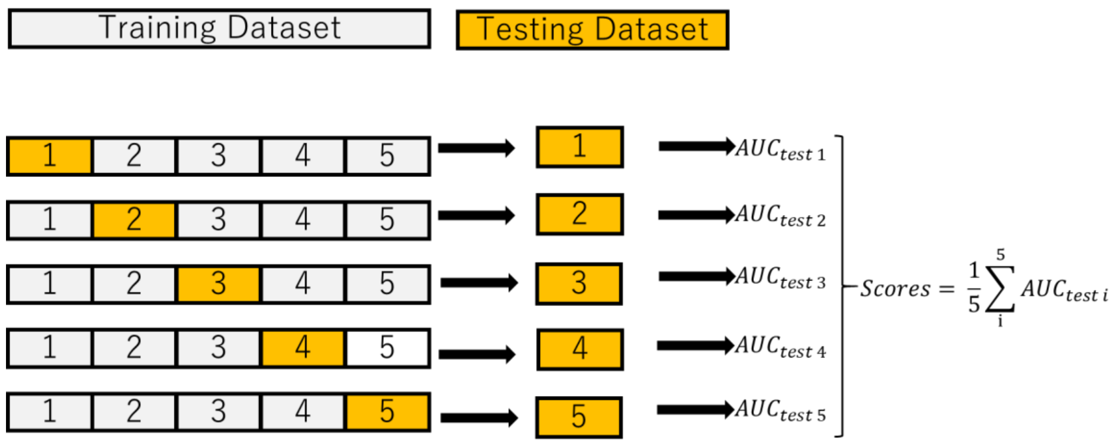



**Figure 7.** Diagram of Cross-validation.

2.4.3. Evaluation Method

**AUC and ROC**

A confusion matrix is a table that is often used to evaluate the performance of a classification algorithm. It provides a comprehensive way to assess the performance of a model by breaking down the predicted and actual classes into four categories. The confusion matrix is typically arranged as shown in Table 2.

**Table 2.** The Confusion matrix.

| | | Predicted category | |
|---|---|---|---|
| | | **Positive (1)** | **Negative (0)** |
| **True category** | **Positive (1)** | True positive (TP) | False negative (FN) |
| | **Negative (0)** | False positive (FP) | True negative (TN) |

AUC stands for Area Under the ROC Curve, and it is a metric used to evaluate the performance of a binary classification model. The ROC (Receiver Operating Characteristic) curve is a graphical representation that illustrates the trade-off between true positive rate and false positive rate at various classification thresholds. A higher AUC indicates better discriminative ability of the model. It measures the model's ability to distinguish between positive and negative instances.

The ROC curve plots False Positive Rate on the abscissa and True Positive Rate on the ordinate. The calculation methods for these two indicators are as follows:

$$True\ Positive\ Rate = \frac{TP}{TP + FN} \tag{5}$$

$$False\ Positive\ Rate = \frac{FP}{FP + TN} \tag{6}$$

**Accuracy (ACC)**

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{7}$$

Precision and Recall correspond to looking at TP from two different dimensions that the prediction effect of positive examples. Precision is subjectively, the effect of positive examples being correctly divided; the denominator of Precision is (TP+FP), which represents the number of all predicted positive examples. Precision represents the proportion of all positive examples that are actually positive. Precision represents the subjective prediction effect of positive examples. Recall is objectively; the positive example is correctly classified. The denominator of Recall is (TP+FN), which represents the number of all actual positive instances. Recall represents the proportion of predictions that are correct among all actual positive examples. Recall indicates whether the prediction effect of the positive example is good or not.

In binary classification tasks, precision (P-score) and recall (R-score) are commonly used to evaluate model effectiveness. Precision measures the accuracy of positive predictions, while recall assesses the ability of the model to identify all relevant instances. However, when these metrics yield divergent results, it complicates the comparison of different models' performance. For instance, consider two models, A and B: Model A exhibits a higher recall, whereas Model B demonstrates greater precision. To resolve this dilemma and provide a balanced evaluation, the F-score is employed. This metric integrates precision and recall into a single score, offering a comprehensive measure of a model’s accuracy and completeness. Specifically, the F1 score is used in our analysis, assigning equal importance to both precision and recall, thereby facilitating nuanced comparisons between models where traditional metrics might conflict.

**Precision(P-score)**

$$\text{Psocre} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{8}$$

**Recall(R-score)**

$$\text{Rscore} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{9}$$

**F1-score(F-score)**

$$\mathrm{F1score} = \frac{2 \times \mathrm{Psocre} \times \mathrm{Rscore}}{\mathrm{Psocre} + \mathrm{TRscore}} \tag{10}$$

2.4.4. Information Gain Ratio (IGR) Method

In this research, the information gain ratio (IGR) method was employed using the training dataset. A higher IGR value for a landslide causative factor indicates its greater importance to the models. Factors with an IGR of zero signify that they contribute nothing to the models and should be excluded [37, 38].

## 2.5. Results

2.5.1. Importance of Features

As depicted in figure 8, the IGR values of all thirteen factors exceeded zero, suggesting their contribution to landslide modeling in our study. Consequently, all selected landslide causative factors were utilized for modeling. However, there were slight variations in the IGR among the three regions. The IGR of the NIG and IWT regions exhibited similarity, while the IGR of the HKD region displayed distinct differences.

In all study areas, the features with high IGR value focus on topography and rain data, Geological data and distance to road, river and faults with low IGR value. Compared with Geological data, according to the landslide distribution map and feature distribution map, landslides have a larger distribution range of topography and rain data, but the distribution range is smaller in geological data, At the same time, the impact of topography and rain data on landslides is more intuitive. For example, in mountainous areas with high altitudes, large slopes, and dense rainwater, landslides are more likely to occur than in other areas. In the study area, landslides are concentrated in one or few types of soil, land-use and lithology. In the HKD area, all IGR values are smaller than IWT area and NIG area. The main reason is that the landslides in distribution HKD are concentrated in the upper right part of this area, and the relative number of recorded landslides is far less than that in the other two study areas.

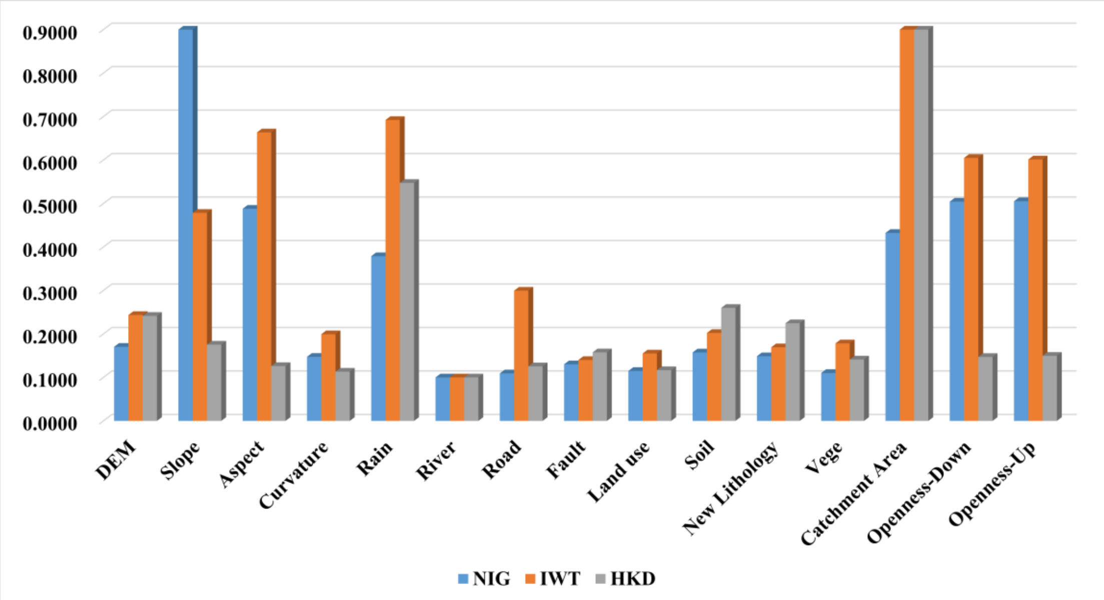



**Figure 8.** Information gain ratio of the landslide feature.

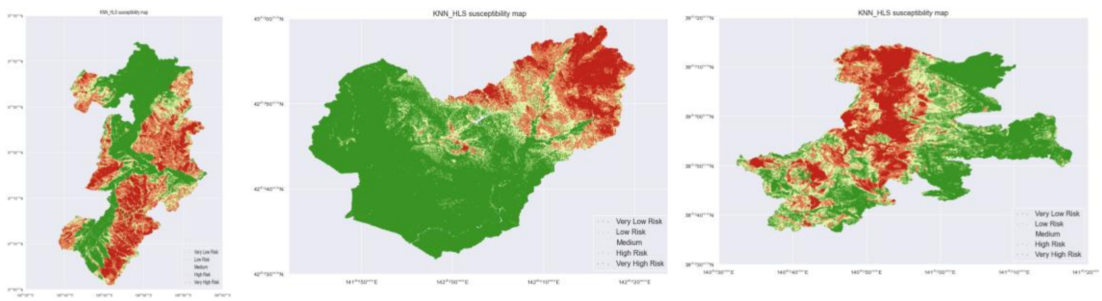
### 2.5.2. Result of Evaluation

Table 3 shows the results by some kinds of machine learning models. According to the results of AUC alone among the three regions, the KNN model shows the best performance for the NIG region (AUC = 0.931), and DT model indicates the best performance for the IWT-MYG region (AUC = 0.961), and the ANN results are the best for the HKD area (AUC = 0.969). For all models, the results obtained for the IWT-MYG area, and the HKD area are better than for the NIG area. However, there is possibility that AUC alone cannot evaluate the performance of the model comprehensively, so Accuracy (ACC), Precision (P-score) and Recall (R-score) are also introduced to further evaluate the model performance. The P-score is not high for all three regions because the model does not perform well in predicting negative samples, resulting in an increase in FP. Accuracy is defined as the proportion of correctly predicted sample number to the total sample numbers. Therefore, ACC results includes the following two tendencies; 1) in case that the P-score is low, and the ACC is low, it means that the model on actual negative examples shows poor performance, and the decrease in the number of correctly predicted samples (TP+TN), resulting in low ACC results, 2) in case that the P-score is low, but the ACC is relatively good, it means that the imbalance of positive and negative examples in the data set. Although the prediction for the actual negative examples is not good, the actual negative examples number is large enough in the data used in this series of training, so the number of correct predicted samples (TP+TN) will not be greatly affected, resulting in relatively good ACC results. The R-score is holistically better in the results in all regions, except the R-score of the LR model in three regions, which shows that the LR model has poor detection performance for actual positive examples, resulting in a decrease in TP.

**Table 3.** Results obtained by machine learning.

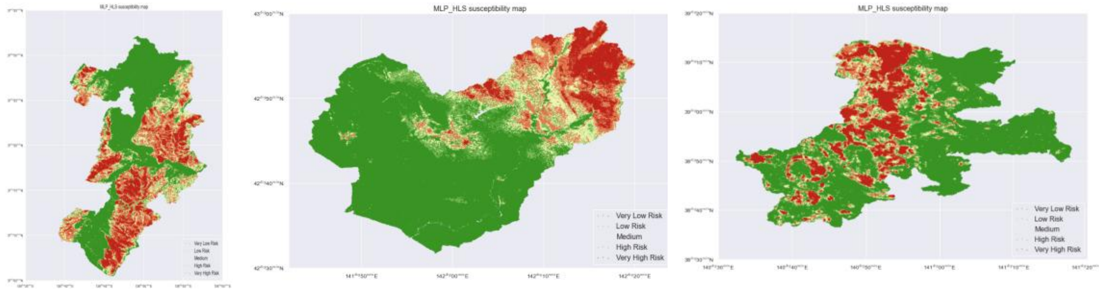
| Area | Model | ACC | P-score | R-score | F-score | AUC |
|---|---|---|---|---|---|---|
| NIG | LR | 0.704 | 0.278 | 0.617 | 0.383 | 0.744 |
| | DT | 0.873 | 0.547 | 0.866 | 0.670 | 0.873 |
| | RF | 0.654 | 0.281 | 0.864 | 0.422 | 0.654 |
| | KNN | 0.832 | 0.466 | 0.884 | 0.610 | 0.931 |
| | ANN | 0.827 | 0.458 | 0.854 | 0.596 | 0.919 |
| | MLP | 0.811 | 0.433 | 0.866 | 0.577 | 0.914 |
| IWT-MYG | LR | 0.802 | 0.331 | 0.495 | 0.396 | 0.769 |
| | DT | 0.908 | 0.600 | 0.893 | 0.718 | 0.961 |
| | RF | 0.756 | 0.323 | 0.783 | 0.457 | 0.844 |
| | KNN | 0.848 | 0.458 | 0.863 | 0.598 | 0.927 |
| | ANN | 0.880 | 0.526 | 0.828 | 0.644 | 0.938 |
| | MLP | 0.876 | 0.516 | 0.857 | 0.644 | 0.942 |
| HKD | LR | 0.873 | 0.202 | 0.635 | 0.307 | 0.873 |
| | DT | 0.952 | 0.480 | 0.918 | 0.630 | 0.952 |
| | RF | 0.726 | 0.413 | 0.805 | 0.551 | 0.839 |
| | KNN | 0.912 | 0.327 | 0.924 | 0.483 | 0.967 |
| | ANN | 0.913 | 0.328 | 0.914 | 0.482 | 0.969 |
| | MLP | 0.898 | 0.314 | 0.928 | 0.465 | 0.958 |

In the landslide susceptibility map (LSM) as shown in figures 9 and 10, the sensitivity is divided into 5 equal parts, while showing each sensitivity represents a different risk level and expressing the results on the coordinates. Judging from the evaluation indicators and the results of LSM, for all three regions, the performance of the KNN model and the MLP model are better than that of the other models, meaning that it is suitable to apply to practical use.




(a)NIG (b)HKD (c)IWT

**Figure 9.** LSM based on machine learning models (KNN).




(a)NIG (b)HKD (c)IWT

**Figure 10.** LSM based on machine learning models (MLP).

## 2.6. Discussion and conclusion

In this study, based on previous research on machine learning and landslides, the datasets generated by ArcGIS were analyzed using several different machine learning models to produce a landslide sensitivity map.

In conventional machine learning, datasets are typically divided according to a fixed proportion, with distinct training and test groups set for analysis. However, in this research, to enable the model to better grasp the relationship between features and labels, we used completely different datasets as the training and test groups. The training group consisted of a dataset with an equal proportion of landslide and non-landslide instances (50% each), while the test set was a completely random dataset. Analysis of model scores demonstrated the feasibility of this approach. For the study areas, the test group results outperformed those of the training group. However, because we used the grid search method for parameter optimization—analyzing one parameter at a time within a defined range to determine the optimal values, the process incurred a relatively high time cost but proved to be the most effective.

Regarding the landslide sensitivity map, certain areas of the HLS sensitivity map derived from the LR and RF models were vague in classifying landslide hazard levels, rendering such results impractical for real-world applications. This indicates that the LR and RF models struggled to clearly distinguish between landslide and non-landslide areas. The KNN and MLP (ANN) models produced results that closely resembled the landslide distribution maps in ArcGIS, though achieving this accuracy required substantial computational effort.

In the next chapter, we will build on our machine learning research and utilize the databases established in this study. Specifically, we will employ CNN models in combination with various window sizes and sampling methods to evaluate landslide risk, highlighting the advantages of CNN over traditional machine learning methods.

# 3. Landslide Risk Assessment Using CNN with Different Window Size and Sampling Method

## 3.1. Research Purpose

Building on previous research, traditional machine learning models were applied to assess landslide susceptibility across Niigata Prefecture (NIG), Iwate and Miyagi Prefectures (IWT-MYG), and Hokkaido (HKD). Utilizing ArcGIS, models such as logistic regression, decision trees, K-nearest neighbors (KNN), artificial neural networks (ANN), and random forests (RF) were employed to create landslide sensitivity maps based on regional characteristics and a comprehensive set of fifteen internal and external factors, including topography, soil, lithology, precipitation, and human activities. These models were trained on historical and earthquake-induced landslide data.

Results revealed that while machine learning models can effectively predict landslides, their performance heavily depends on regional conditions and high-quality data. Logistic regression and random forests were less effective, whereas KNN and ANN demonstrated better performance in distinguishing landslide-prone areas. Notably, feature importance varied across regions, with slope and lithology being consistently influential yet differing in impact depending on location. These findings emphasized the need for tailored approaches to account for regional differences in landslide risk prediction.

Building on these insights, this study aims to develop a novel approach to landslide risk assessment by leveraging a convolutional neural network (CNN) that utilizes varying input window sizes to enhance spatial feature analysis. By dividing the study areas into smaller image segments, the method increases sensitivity to subtle landscape details. To address the challenge of data imbalance in landslide prediction, we integrate downsampling, which reduces the prevalence of non-landslide data, and oversampling, through a random window-moving method, to ensure balanced data representation. Preliminary results indicate that larger window sizes, particularly 19×19 pixels, significantly enhance predictive performance, and the random window-moving method further improves model accuracy. These innovations aim to develop more robust and adaptable tools for landslide risk assessment tailored to the unique characteristics of diverse regions.

### 3.2. Study Area and Data

In this chapter, the database used is the same as in the previous chapter; however, the manner in which the data is utilized differs. In traditional machine learning models, the data was employed in the form of point data, where each point represented specific spatial characteristics. In contrast, the CNN model utilizes the data in the form of TIFF images, enabling the analysis of spatial patterns and features in a more comprehensive and continuous manner. This shift in data representation allows the CNN model to capture nuanced spatial details and relationships that may not be fully utilized in point-based approaches.

### 3.3. Method

The process of this study is outlined in figure 11 and can be divided into four main stages. First, the database is established, and all feature data is integrated and standardized using ArcGIS. Next, the data undergoes preprocessing where the established Landslide Record Database and Landslide Feature Database are segmented according to predefined window sizes. Then, Balancing the data, a trained model is developed, and its performance is assessed. Lastly, the preprocessed data is fed back into the trained model for testing and prediction in the fourth stage.

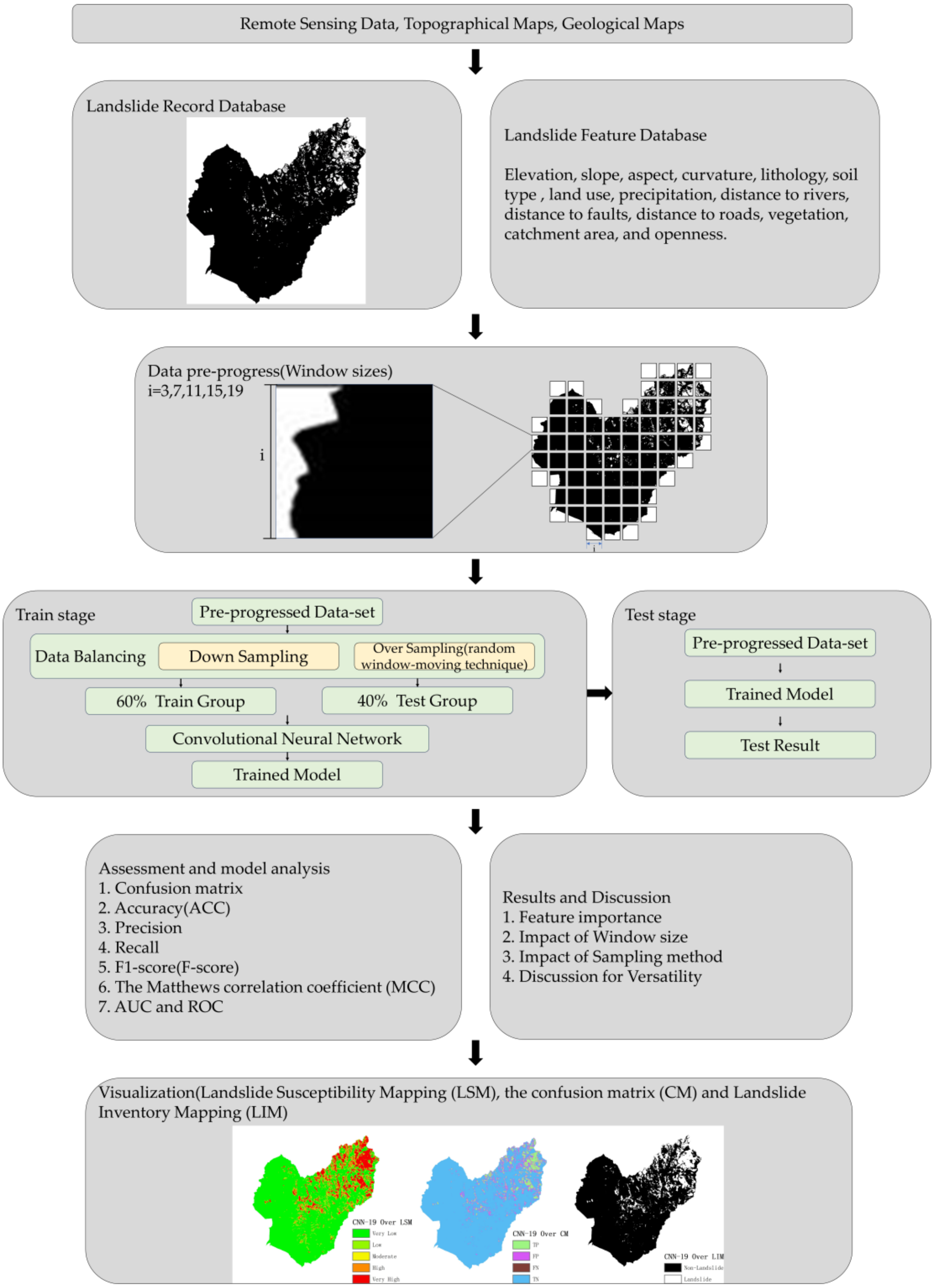



**Figure 11**. Flowchart of this study.

### 3.3.1. CNN

CNN, short for Convolutional Neural Network, is a type of artificial neural network known for its ability to effectively extract feature representations from images. This capability allows CNNs to recognize visual patterns in images without the need for complex, human-designed rules [39].

The structure of a CNN can be broken down into three main layers: the convolutional layer, the pooling layer, and the fully connected layer. In the convolutional layer, filters are applied to the input image to detect various features. Given a picture size of 500×500 and passing through 50 filter convolution layers, the resulting dimension would be 500×500×50. To manage this large amount of data, the output is downscaled in the pooling layer, effectively reducing the dimensionality while preserving important features. The fully connected layer is responsible for classification, where the features extracted from the previous layers are used to classify objects. The key operations performed in any CNN can be succinctly described by the following equation[40]:

$$O^l = P\left(\sigma\left(O^{l-1} \times W^l + b^l\right)\right) \tag{11}$$

Where the $O^{l-1}$ is the output feature data from the previous layer of the lth layer, the $W^l$ and the $b^l$ indicate the weights of the layer, respectively, that convolve the $O^{l-1}$ by the linear convolution, and the $\sigma(\cdot)$ denotes the non-linearity function outside the convolutional layer. These steps are often followed by a pooling operation which is represented by P in Equation (1)[40].

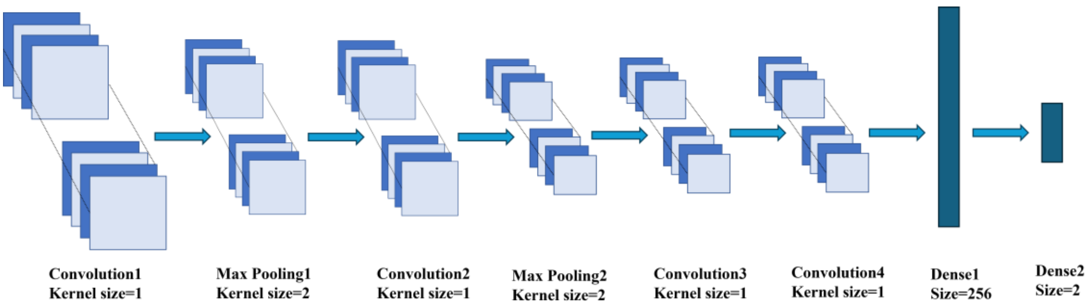



**Figure 12.** Diagram of CNN model structure.

In this project, we split the dataset into 60% for training and 40% for testing (train-Prop = 0.6). The model is trained using the Adam optimizer with an initial learning rate of 0.001. The architecture starts with several convolutional layers, beginning with a 512-channel 1x1 filter layer with 'valid' padding and ReLU activation, followed by a 0.25 Dropout and a 2x2 Max Pooling layer to reduce dimensionality and enhance spatial in-variance. Additional convolutional layers include 256, 128, 64, and 32 channel 1x1 filters, each followed by a 0.25 Dropout to mitigate overfitting. The network concludes with a Flatten layer, a 256-node Dense layer with ReLU activation, another 0.25 Dropout, and a final 2-node Dense layer with soft max activation suitable for binary classification tasks.

### 3.3.2. CNN Windows Sizes

In this research, multiple input window sizes were employed for landslide detection, including 3×3 pixels, 7×7 pixels, 11×11 pixels, 15×15 pixels, 19×19 pixels, 23×23 pixels and 27×27 pixels determined through cross-validation. The utilization of various input window sizes was necessitated by the intricate nature of the features' shapes and sizes within the study area. Notably, the presence of both large and small landslides, often exhibiting diverse shapes, posed a challenge for detection[16]. Some landslides appeared elongated and potentially thin, resembling unsealed roads rather than typical landslide formations. Moreover, the study area exhibited varying aspects, slopes, and flow directions, with individual landslides encompassing different aspects. Furthermore, most landslides displayed a blend of topographic features, further complicating their recognition [16]. To address these complexities, different sizes of input windows were employed, following a similar approach used for detecting complex-shaped objects in urban areas[40]. For clarity, the notation "CNN-i" will be adopted to represent different window sizes, 3×3 pixels, 7×7 pixels, 11×11 pixels, 15×15 pixels, 19×19 pixels, 23×23 pixels and 27×27 pixels, correspond"CNN-3, CNN-7, CNN-11, CNN-15, CNN-19, CNN-23 and CNN-27", respectively.

Additionally, the DEM data used in this study had a resolution of 10m×10m, where one pixel could cover up to 270 m×270 m. When employing DEM data with different resolutions, such as 30m×30m, one pixel can cover up to 810m×810m. Therefore, when using different DEM datasets, it is essential to appropriately scale the window sizes to ensure that the spatial representation remains consistent and effective for detecting landslides across various resolutions. This scaling is vital to maintain the accuracy and reliability of the model when applied to different geographic extents and resolutions.

### 3.3.3. Sampling Techniques

Deep learning methodologies, particularly Convolutional Neural Networks (CNN), require a substantial number of sample patches for effective training[41], typically in the form of a labeled training dataset [42]. Maintaining a balanced distribution of samples is equally crucial. To mitigate the significant imbalance between landslide and non-landslide data, this study employs two distinct sampling strategies: downsampling and oversampling. Downsampling is implemented by randomly removing non-landslide data to equalize the presence of both categories in the dataset, thereby preventing model bias towards the more prevalent class. Conversely, oversampling is utilized to enhance the representativeness of the landslide data by artificially expanding the training dataset. This is achieved through techniques such as translating, rotating, and randomly mirroring images or input windows, as suggested by previous research [38]. These methods ensure a more balanced dataset, which is crucial for training robust machine learning models that can accurately predict diverse landslide occurrences. Directly comparing the impact of the random window-moving technique with downsampling can provide clear insights into their effects on dataset balance and model performance, especially in cases of class imbalance. The random window-moving technique increases minority class samples by creating variations of existing samples, enriching the dataset. In contrast, downsampling reduces the size of the majority class, leading to faster processing but potentially losing valuable information. Highlighting these differences can shed light on the strengths and weaknesses of each method.

In this study, the random window-moving technique was employed specifically due to landslide-related considerations [43], as illustrated in the schematic diagram depicted in figure 13. The implementation details of the random window-moving technique are as follows: The window size was set to half of the input window size. The stride was set to half of the moving window size to ensure substantial overlap and continuous coverage across the image. Boundary handling was managed through padding, maintaining consistency in feature dimensions across the entire image. Window starting positions were selected entirely at random, enhancing the diversity of the sampled image segments and potentially increasing the robustness of the model to variations in input data. Additionally, care was taken to ensure that the data is centered within the window, which helps in better capturing the local features within the window and improves the model's recognition ability.

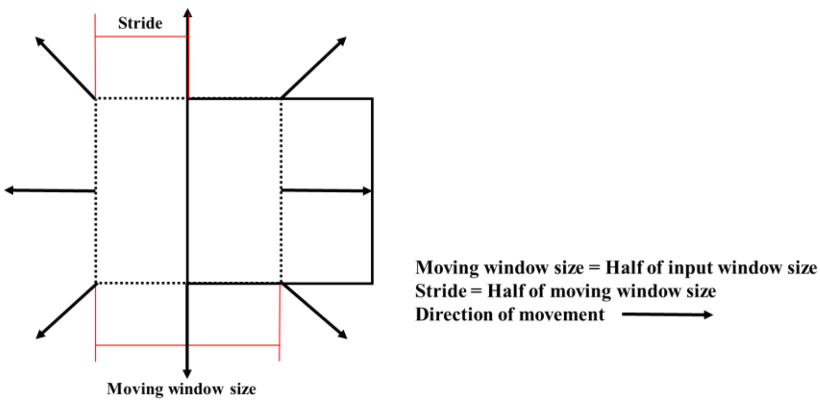



**Figure 13**. Schematic diagram of shifting method.

### 3.3.4. Evaluation Method

In ML, the Matthews correlation coefficient (MCC) has been used as a measure of binary classifications, even if the two classes are of very different sizes [44]. The MCC is defined as follows:

$$\mathrm{MCC} = \frac{\mathrm{TP} \times \mathrm{TN} - \mathrm{FP} \times \mathrm{FN}}{\sqrt{(\mathrm{TP} + \mathrm{FP})(\mathrm{TP} + \mathrm{FN})(\mathrm{TN} + \mathrm{FP})(\mathrm{TN} + \mathrm{FN})}} \tag{12}$$

### 3.4. Results and Discussion

3.4.1. Window Size

3.4.1.1. Window Size and Landslide Scale

Firstly, our study extended the analysis of CNN models' performance across different landslide areas scale, randomly selected 24 landslides of different scales with a unified $9\mathrm{km}^2$ study area in the three study areas of HKD, NIG, and IWT to study the input window size and landslide scale. The changes in landslide scale and AUC of different CNN-i models are shown in figure 14.

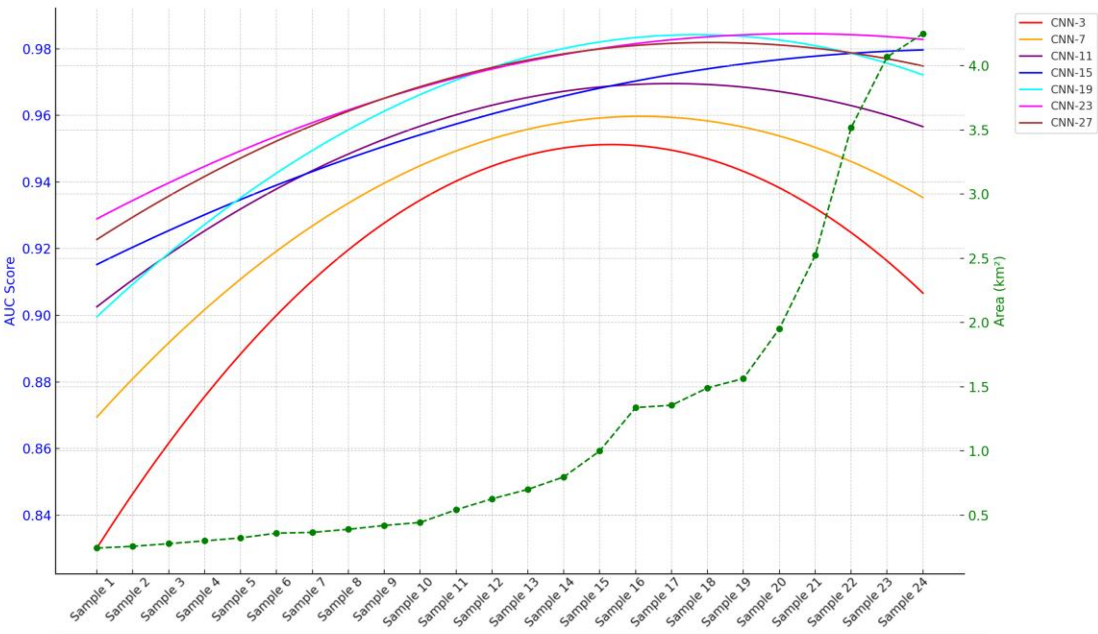



**Figure 14.** Variation of landslide areas across landslide scale samples and AUC of CNN-i models across landslide scale.

The analysis of CNN models, ranging from CNN-3 to CNN-27, illustrates their effectiveness in predicting landslide scales through a comparison of average AUC values for all samples and Pearson correlation coefficients (as shown in figure 15). The study emphasizes the crucial role of input window size in determining model accuracy and its correlation with actual landslide dimensions. Models like CNN-19 and CNN-23 stand out, achieving the highest AUC scores, which suggests their strong ability to capture critical features with well-sized input windows. In contrast, the smallest (CNN-3) and largest (CNN-27) window sizes exhibited lower correlation coefficients, likely due to underfitting and overfitting, respectively. The optimal range of window sizes, identified between 15x15 and 23x23 pixels, strikes a balance by providing sufficient contextual information without introducing excessive noise. This range enhances classification accuracy and maintains strong correlations with real-world landslide data, optimizing both predictive power and model relevance.

Building on these findings, the decision was made to employ CNN models ranging from CNN-15 to CNN-23 to analyze the entire study area, while also examining the results of CNN-3 to CNN-11 for comparison. Models within the CNN-15 to CNN-23 range, with their medium-sized input windows, are particularly effective at capturing essential landscape features related to landslide scales, thereby ensuring high performance. This effectiveness is reflected in their high average AUC values, indicating strong discriminative capabilities in classifying landslide-prone and non-landslide areas. Among these, CNN-15 stands out, exhibiting the highest Pearson correlation coefficients, which validate a strong linear relationship between the model's predictions and actual landslide sizes. This correlation is vital because it confirms the models' capacity to generate results that accurately reflect real-world geological conditions. The adaptability of these models across different terrains further enhances their suitability for broad-scale geographical analyses, making them valuable tools for reliable and precise landslide prediction. This precision is crucial for comprehensive landslide risk assessment and management. By strategically selecting CNN-15 to CNN-23 models, the approach optimizes performance across key metrics, setting the stage for future improvements. This could include hybrid or ensemble methods, which have the potential to further increase both accuracy and applicability in various environmental contexts.

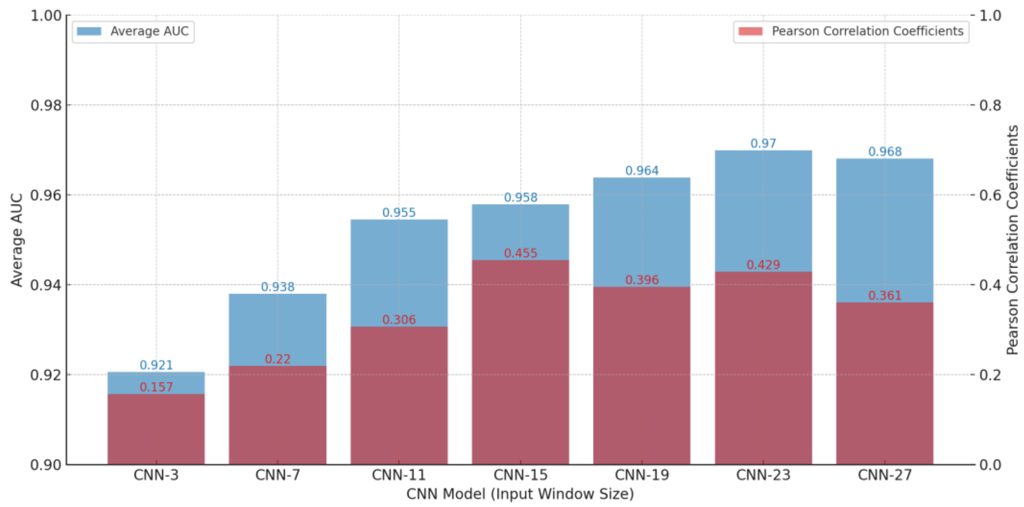



**Figure 15.** Average AUC and Pearson correlation coefficients of CNN Models for Landslide scale.

3.4.1.2. Window Size and Whole Study Area

Based on the results from the trained model for each study area (HKD, NIG, and IWT) using downsampling and oversampling methods while varying window sizes. The findings across these study areas confirm that regardless of the sampling method, increasing the window size leads to improvements in evaluation indices such as AUC, P-score, and R-score. For instance, in the HKD region, the trained model's accuracy improved from 0.907 with a 3×3 pixels window size to 0.943 with a 19×19 pixels window size. Similarly, the AUC increased from 0.964 to 0.982(in Table 4-a). When applied to the test data, the model also showed improvements, with accuracy increasing from 0.887 to 0.913 and the P-score from 0.216 to 0.271 as the window size increased (in Table 5-a). Comparison with the results in Table 3 for ACC and AUC from machine learning demonstrates that the CNN model outperforms traditional machine learning approaches across all window sizes, indicating a more robust performance.

Larger window sizes allow the model to capture a wider range of environmental variables simultaneously, enabling it to understand more complex interactions. For example, larger windows can incorporate diverse topographical gradients, soil types, and hydrological features that smaller windows may overlook, leading to a more detailed comprehension of landslide triggers. By offering a broader spatial context, CNN models can distinguish more effectively between landslide-prone and stable areas. This is particularly advantageous in complex terrains where local conditions may not fully represent regional dynamics, and larger windows provide a macroscopic view that helps in identifying patterns indicating instability.

Among the various CNN configurations tested, CNN-19 emerges as the best model for predicting landslide scales, outperforming models like CNN-15 and CNN-23. CNN-19's 19x19 input window strikes an ideal balance—large enough to include essential terrain features for accurate landslide modeling, yet small enough to avoid incorporating irrelevant data that could lower performance. This optimal window size allows CNN-19 to capture a comprehensive landscape view while maintaining generalization across diverse terrains without the underfitting or overfitting issues associated with smaller (CNN-15) or larger (CNN-23) windows.

In addition to high AUC values, CNN-19 maintains strong Pearson correlation coefficients, ensuring that its predictions not only effectively classify landslide-prone areas but also align closely with actual landslide dimensions, reflecting a reliable geological understanding. While CNN-15 may be underfit in complex terrains due to its smaller window, and CNN-23 may be overfit by incorporating excessive noise, CNN-19 strikes the right balance, making it the most effective in capturing the dynamics of landslides.

However, larger window sizes come with challenges, such as increased computational demand and higher memory requirements, which can limit real-time application. Additionally, the resolution of the input data is critical for selecting the optimal window size. Future research could explore multi-scale analysis, training models with different window sizes to improve robustness, or integrating CNNs with other machine learning methods to mitigate the limitations of single-scale models, potentially resulting in more accurate landslide prediction systems.

Furthermore, the inherent imbalance in the test data structure may lead to a disparity between the test and trained model results, with test results often underperforming compared to the trained models.

### 3.4.2. Sampling Method

According to Tables 4-7 and Figures 16-18, the evaluation indices such as AUC, P-score, and R-score obtained from models trained using the oversampling method tend to be higher than those using the downsampling method. This improvement is evident in both the training and test results.

The superiority of oversampling is particularly pronounced in the test model, where it outperforms the downsampling technique. Firstly, oversampling increases the number of samples in the entire dataset, which is crucial for the CNN model. Secondly, the oversampled data in this study consists of landslide data samples, making them distinct from non-landslide data. Achieving data alignment and sample balance is essential in deep learning, and oversampling effectively addresses these concerns.

As a result, oversampling increases the representation of landslide instances in the training data, significantly enhancing the model's ability to detect subtle cues associated with landslide occurrences. This method reduces the model’s bias towards the more frequent non-landslide class. However, it may also introduce variance by potentially overfitting to the landslide class, particularly if the oversampling technique simply replicates existing samples without introducing variability.

On the other hand, while downsampling helps balance the classes by reducing the size of the dominant non-landslide class, it risks losing valuable information critical for identifying landslides. This approach can inadvertently simplify the model's view of the landscape, potentially leading to underfitting and poorer generalization on new, unseen data.

In summary, both oversampling and downsampling have certain drawbacks. Further research should investigate adaptive sampling techniques that dynamically adjust the sampling method based on real-time model performance and feedback. This approach could optimize the balance between bias and variance dynamically, improving model accuracy and generalizability. Additionally, exploring advanced oversampling techniques that introduce more realistic variability into the training data could help develop more robust and accurate landslide prediction models.

**Table 4-a.** Trained CNN-i model results based on HKD area with downsampling method.

| **Area** | **Model** | **Confusion matrix** | | | | **ACC** | **P-score** | **R-score** | **F-score** | **AUC** |
|---|---|---|---|---|---|---|---|---|---|---|
| | | TP | FP | TN | FN | | | | | |
| HKD (Down) | CNN-3 | 45835 | 5759 | 43881 | 3437 | 0.907 | 0.888 | 0.930 | 0.909 | 0.964 |
| | CNN-7 | 47525 | 5837 | 43744 | 1806 | 0.923 | 0.891 | 0.963 | 0.926 | 0.971 |
| | CNN-11 | 47170 | 4520 | 44981 | 2241 | 0.932 | 0.913 | 0.955 | 0.933 | 0.977 |
| | CNN-15 | 48138 | 4552 | 44786 | 1436 | 0.939 | 0.914 | 0.971 | 0.941 | 0.981 |
| | CNN-19 | 48015 | 4457 | 45211 | 1229 | 0.943 | 0.915 | 0.975 | 0.944 | 0.982 |
| | CNN-23 | 46608 | 3179 | 46189 | 2936 | 0.938 | 0.936 | 0.941 | 0.938 | 0.982 |

**Table 4-b.** Trained CNN-i model results based on NIG area with downsampling method.

| **Area** | **Model** | **Confusion matrix** | | | | **ACC** | **P-score** | **R-score** | **F-score** | **AUC** |
|---|---|---|---|---|---|---|---|---|---|---|
| | | TP | FP | TN | FN | | | | | |
| NIG (Down) | CNN-3 | 46111 | 10928 | 36115 | 979 | 0.874 | 0.808 | 0.979 | 0.886 | 0.933 |
| | CNN-7 | 46364 | 10302 | 36799 | 668 | 0.883 | 0.818 | 0.986 | 0.894 | 0.944 |
| | CNN-11 | 46643 | 10986 | 36233 | 271 | 0.880 | 0.809 | 0.994 | 0.892 | 0.945 |
| | CNN-15 | 46611 | 10036 | 37114 | 372 | 0.889 | 0.823 | 0.992 | 0.900 | 0.949 |
| | CNN-19 | 46759 | 10286 | 36878 | 210 | 0.888 | 0.820 | 0.996 | 0.899 | 0.950 |
| | CNN-23 | 46402 | 9645 | 37483 | 603 | 0.891 | 0.828 | 0.987 | 0.901 | 0.953 |

**Table 4-c.** Trained CNN-i model results based on IWT area with downsampling method.

| Area | Model | Confusion matrix TP | FP | TN | FN | ACC | P-score | R-score | F-score | AUC |
|---|---|---|---|---|---|---|---|---|---|---|
| IWT (Down) | CNN-3 | 106053 | 15428 | 98251 | 7548 | 0.899 | 0.873 | 0.934 | 0.902 | 0.959 |
| | CNN-7 | 109538 | 16030 | 97962 | 3750 | 0.913 | 0.872 | 0.967 | 0.917 | 0.966 |
| | CNN-11 | 109430 | 14585 | 99164 | 4101 | 0.918 | 0.882 | 0.964 | 0.921 | 0.970 |
| | CNN-15 | 109643 | 13044 | 100619 | 3974 | 0.925 | 0.894 | 0.965 | 0.928 | 0.973 |
| | CNN-19 | 109426 | 13511 | 99928 | 4415 | 0.921 | 0.890 | 0.961 | 0.924 | 0.969 |
| | CNN-23 | 109232 | 13207 | 100595 | 4246 | 0.923 | 0.892 | 0.963 | 0.926 | 0.973 |

**Table 5-a.** Test group results using trained CNN-i model based on HKD area with downsampling method.

| Area | Model | Confusion matrix TP | FP | TN | FN | ACC | P-score | R-score | F-score | MCC |
|---|---|---|---|---|---|---|---|---|---|---|
| HKD (Down) | CNN-3 | 115233 | 418198 | 3222142 | 8407 | 0.887 | 0.216 | 0.932 | 0.351 | 0.418 |
| | CNN-7 | 119219 | 425302 | 3215038 | 4421 | 0.886 | 0.219 | 0.964 | 0.357 | 0.429 |
| | CNN-11 | 118046 | 336527 | 3303813 | 5594 | 0.909 | 0.260 | 0.955 | 0.408 | 0.472 |
| | CNN-15 | 120168 | 336599 | 3303741 | 3472 | 0.910 | 0.263 | 0.972 | 0.414 | 0.480 |
| | CNN-19 | 120754 | 324446 | 3315894 | 2886 | 0.913 | 0.271 | 0.977 | 0.425 | 0.490 |
| | CNN-23 | 116595 | 231030 | 3409310 | 7045 | 0.937 | 0.335 | 0.943 | 0.495 | 0.541 |

**Table 5-b.** Test group results using trained CNN-i model based on NIG area with downsampling method.

| Area | Model | Confusion matrix TP | FP | TN | FN | ACC | P-score | R-score | F-score | MCC |
|---|---|---|---|---|---|---|---|---|---|---|
| NIG (Down) | CNN-3 | 115298 | 344422 | 1120262 | 2368 | 0.781 | 0.251 | 0.980 | 0.399 | 0.430 |
| | CNN-7 | 116072 | 319706 | 1144978 | 1594 | 0.797 | 0.266 | 0.986 | 0.419 | 0.451 |
| | CNN-11 | 117009 | 341023 | 1123661 | 657 | 0.784 | 0.255 | 0.994 | 0.406 | 0.441 |
| | CNN-15 | 116790 | 311512 | 1153172 | 876 | 0.803 | 0.273 | 0.993 | 0.428 | 0.461 |
| | CNN-19 | 117145 | 319558 | 1145126 | 521 | 0.798 | 0.268 | 0.996 | 0.423 | 0.456 |
| | CNN-23 | 116266 | 298753 | 1165931 | 1400 | 0.810 | 0.280 | 0.988 | 0.437 | 0.468 |

**Table 5-c.** Test group results using trained CNN-i model based on IWT area with downsampling method.

| Area | Model | Confusion matrix TP | FP | TN | FN | ACC | P-score | R-score | F-score | MCC |
|---|---|---|---|---|---|---|---|---|---|---|
| IWT (Down) | CNN-3 | 265495 | 599087 | 3789656 | 18604 | 0.868 | 0.307 | 0.935 | 0.462 | 0.491 |
| | CNN-7 | 274943 | 622882 | 3765861 | 9156 | 0.865 | 0.306 | 0.968 | 0.465 | 0.501 |
| | CNN-11 | 274090 | 567355 | 3821388 | 10009 | 0.876 | 0.326 | 0.965 | 0.487 | 0.520 |
| | CNN-15 | 274207 | 503575 | 3885168 | 9892 | 0.890 | 0.353 | 0.965 | 0.516 | 0.546 |
| | CNN-19 | 273211 | 520643 | 3868100 | 10888 | 0.886 | 0.344 | 0.962 | 0.507 | 0.536 |
| | CNN-23 | 273526 | 506337 | 3882406 | 10573 | 0.889 | 0.351 | 0.963 | 0.514 | 0.543 |

**Table 6-a.** Trained CNN-i model results based on HKD area with oversampling method.

| Area | Model | Confusion matrix | | | | ACC | P-score | R-score | F-score | AUC |
|---|---|---|---|---|---|---|---|---|---|---|
| | | TP | FP | TN | FN | | | | | |
| HKD (Over) | CNN-3 | 1426211 | 124347 | 1331308 | 30406 | 0.947 | 0.920 | 0.979 | 0.949 | 0.984 |
| | CNN-7 | 1445542 | 110236 | 1345726 | 10768 | 0.958 | 0.929 | 0.993 | 0.960 | 0.990 |
| | CNN-11 | 1449994 | 85648 | 1370168 | 6462 | 0.968 | 0.944 | 0.996 | 0.969 | 0.993 |
| | CNN-15 | 1452864 | 74311 | 1381416 | 3681 | 0.973 | 0.951 | 0.997 | 0.974 | 0.995 |
| | CNN-19 | 1454869 | 59765 | 1396738 | 900 | 0.979 | 0.961 | 0.999 | 0.980 | 0.997 |
| | CNN-23 | 1386938 | 118275 | 1338441 | 68618 | 0.936 | 0.921 | 0.953 | 0.937 | 0.979 |

**Table 6-b.** Trained CNN-i model results based on NIG area with oversampling method.

| Area | Model | Confusion matrix | | | | ACC | P-score | R-score | F-score | AUC |
|---|---|---|---|---|---|---|---|---|---|---|
| | | TP | FP | TN | FN | | | | | |
| NIG (Over) | CNN-3 | 581436 | 121195 | 464813 | 4304 | 0.893 | 0.828 | 0.993 | 0.903 | 0.952 |
| | CNN-7 | 581505 | 98173 | 487608 | 4462 | 0.912 | 0.856 | 0.992 | 0.919 | 0.966 |
| | CNN-11 | 580933 | 78093 | 507844 | 4878 | 0.929 | 0.882 | 0.992 | 0.933 | 0.974 |
| | CNN-15 | 581178 | 64862 | 521115 | 4593 | 0.941 | 0.900 | 0.992 | 0.944 | 0.981 |
| | CNN-19 | 584992 | 68492 | 516619 | 1645 | 0.940 | 0.895 | 0.997 | 0.943 | 0.983 |
| | CNN-23 | 577578 | 108612 | 477954 | 7604 | 0.901 | 0.842 | 0.987 | 0.909 | 0.955 |

**Table 6-c.** Trained CNN-i model results based on IWT area with oversampling method.

| Area | Model | Confusion matrix | | | | ACC | P-score | R-score | F-score | AUC |
|---|---|---|---|---|---|---|---|---|---|---|
| | | TP | FP | TN | FN | | | | | |
| IWT (Over) | CNN-3 | 1713553 | 180576 | 1573831 | 43035 | 0.936 | 0.905 | 0.976 | 0.939 | 0.979 |
| | CNN-7 | 1721697 | 163870 | 1591188 | 34240 | 0.944 | 0.913 | 0.981 | 0.946 | 0.983 |
| | CNN-11 | 1737835 | 149781 | 1606272 | 17107 | 0.952 | 0.921 | 0.990 | 0.954 | 0.987 |
| | CNN-15 | 1718580 | 117579 | 1637565 | 37271 | 0.956 | 0.936 | 0.979 | 0.957 | 0.986 |
| | CNN-19 | 1717931 | 81962 | 1673660 | 37442 | 0.966 | 0.954 | 0.979 | 0.966 | 0.991 |
| | CNN-23 | 1750909 | 273065 | 1482606 | 4415 | 0.921 | 0.865 | 0.997 | 0.927 | 0.972 |

**Table 7-a.** Test group results using trained CNN-i model based on HKD area with oversampling method.

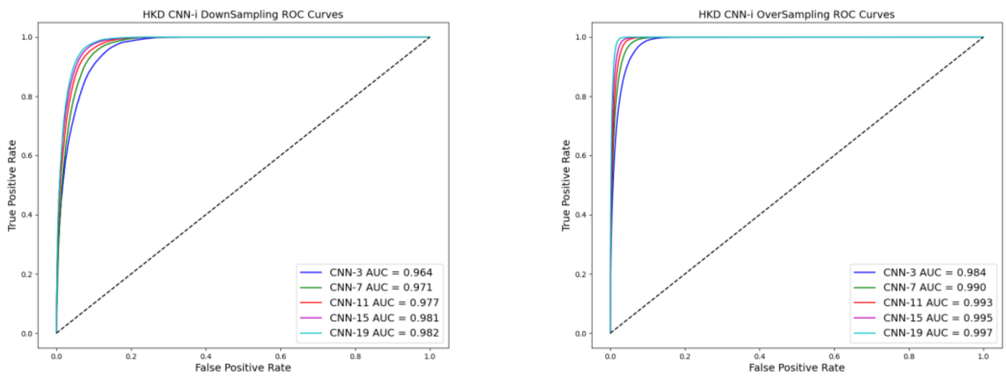
| Area | Model | Confusion matrix | | | | ACC | P-score | R-score | F-score | MCC |
|---|---|---|---|---|---|---|---|---|---|---|
| | | TP | FP | TN | FN | | | | | |
| HKD (Over) | CNN-3 | 121085 | 311546 | 3328794 | 2555 | 0.917 | 0.280 | 0.979 | 0.435 | 0.499 |
| | CNN-7 | 122725 | 274251 | 3366089 | 915 | 0.927 | 0.309 | 0.993 | 0.471 | 0.532 |
| | CNN-11 | 123090 | 212257 | 3428083 | 550 | 0.943 | 0.367 | 0.996 | 0.536 | 0.586 |
| | CNN-15 | 123334 | 184167 | 3456173 | 306 | 0.951 | 0.401 | 0.998 | 0.572 | 0.616 |
| | CNN-19 | 123564 | 147155 | 3493185 | 76 | 0.961 | 0.456 | 0.999 | 0.627 | 0.662 |
| | CNN-23 | 117825 | 294533 | 3345807 | 5815 | 0.920 | 0.286 | 0.953 | 0.440 | 0.498 |

**Table 7-b.** Test group results using trained CNN-i model based on NIG area with oversampling method.

| Area | Model | Confusion matrix | | | | ACC | P-score | R-score | F-score | MCC |
|---|---|---|---|---|---|---|---|---|---|---|
| | | TP | FP | TN | FN | | | | | |
| NIG (Over) | CNN-3 | 116797 | 302690 | 1161994 | 869 | 0.808 | 0.278 | 0.993 | 0.435 | 0.467 |
| | CNN-7 | 116759 | 245016 | 1219668 | 907 | 0.845 | 0.323 | 0.992 | 0.487 | 0.515 |
| | CNN-11 | 116676 | 194057 | 1270627 | 990 | 0.877 | 0.375 | 0.992 | 0.545 | 0.567 |
| | CNN-15 | 116748 | 159136 | 1305548 | 918 | 0.899 | 0.423 | 0.992 | 0.593 | 0.611 |
| | CNN-19 | 117332 | 169400 | 1295284 | 334 | 0.893 | 0.409 | 0.997 | 0.580 | 0.600 |
| | CNN-23 | 116154 | 271395 | 1193289 | 1512 | 0.828 | 0.300 | 0.987 | 0.460 | 0.489 |

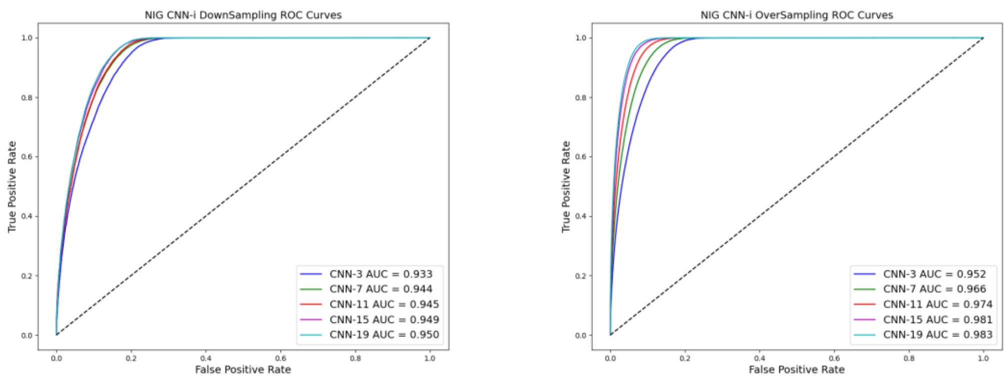
**Table 7-c.** Test group results using trained CNN-i model based on IWT area with oversampling method.

| Area | Model | Confusion matrix | | | | ACC | P-score | R-score | F-score | MCC |
|---|---|---|---|---|---|---|---|---|---|---|
| | | TP | FP | TN | FN | | | | | |
| IWT (Over) | CNN-3 | 277195 | 450606 | 3938137 | 6904 | 0.902 | 0.381 | 0.976 | 0.548 | 0.575 |
| | CNN-7 | 278574 | 408079 | 3980664 | 5525 | 0.911 | 0.406 | 0.981 | 0.574 | 0.599 |
| | CNN-11 | 281337 | 373809 | 4014934 | 2762 | 0.919 | 0.429 | 0.990 | 0.599 | 0.623 |
| | CNN-15 | 278007 | 294255 | 4094488 | 6092 | 0.936 | 0.486 | 0.979 | 0.649 | 0.664 |
| | CNN-19 | 278052 | 204274 | 4184469 | 6047 | 0.955 | 0.576 | 0.979 | 0.726 | 0.732 |
| | CNN-23 | 283393 | 683656 | 3705087 | 706 | 0.854 | 0.293 | 0.998 | 0.453 | 0.496 |



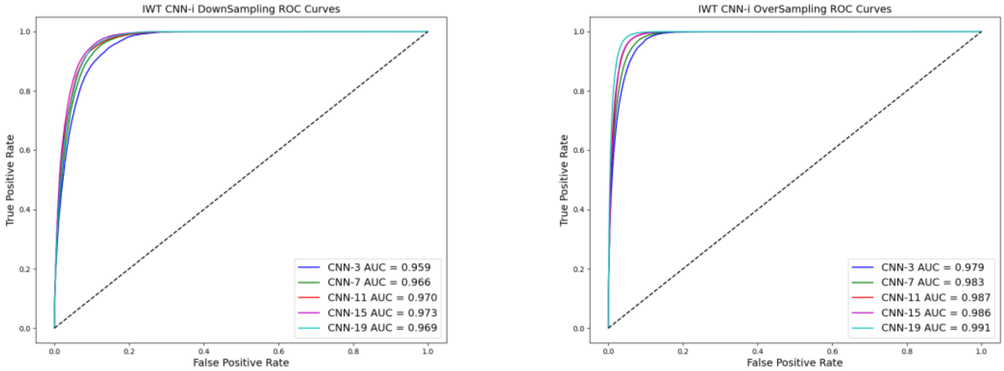

(a) (b)

**Figure 16.** ROC curve in downsampling (a) and oversampling (b) in HKD based on CNN-i.




(a) (b)

**Figure 17.** ROC curve in downsampling (a) and oversampling (b) in NIG based on CNN-i.




(a) (b)

**Figure 18.** ROC curve in downsampling (a) and oversampling (b) in IWT based on CNN-i.

### 3.4.3. Influence of Window Size on False Negatives and False Positives

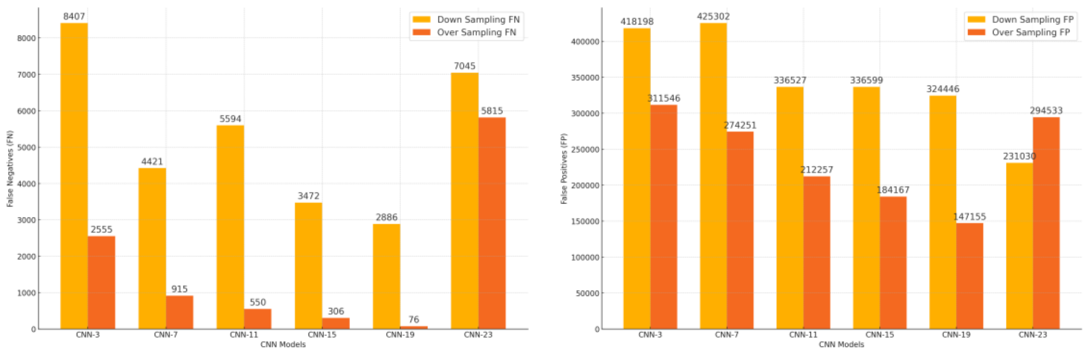
This study evaluates the impact of different sampling methods and window sizes on CNN performance across three regions: HKD, NIG, and IWT (shown in Figures 19-21). The results are presented in terms of false negatives (FN) and false positives (FP), comparing downsampling and oversampling approaches for each model.

The analysis reveals that window size plays a critical role in minimizing FN errors. In the HKD region, an optimum window size is identified where FN values are significantly reduced, as highlighted in the figure. When the window size is too small, the models tend to misclassify many positive cases, leading to high FN rates. Conversely, excessively large window sizes also result in increased FN errors, suggesting a trade-off between spatial resolution and prediction accuracy. A similar pattern is observed in the NIG and IWT regions, where models with intermediate window sizes generally exhibit better performance in reducing FN.

In applications such as landslide susceptibility mapping and other disaster-related tasks, FN errors are particularly critical. A high FN rate means that actual hazardous areas are misclassified as non-hazardous, leading to severe consequences, such as inadequate risk assessment and failure to issue early warnings. This misclassification can increase human and economic losses, particularly in regions prone to natural disasters. Therefore, minimizing FN should be prioritized when optimizing CNN models for hazard detection. The results indicate that over-sampling generally reduces FN errors more effectively than down-sampling, but its impact varies across different window sizes and study regions.

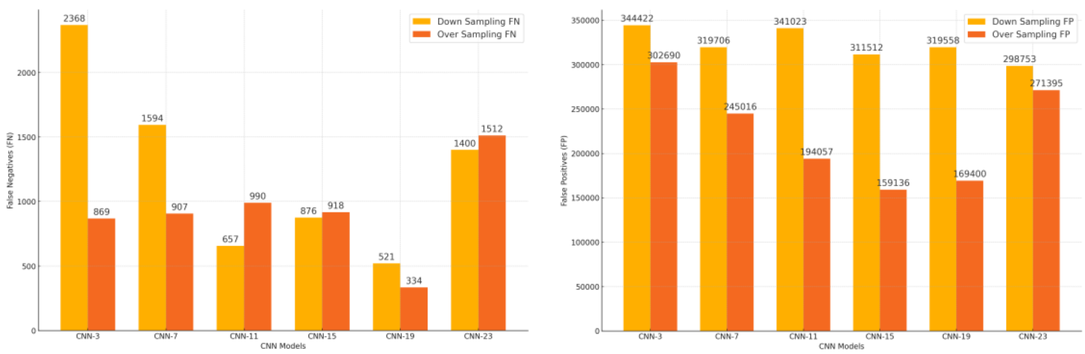
While reducing FN is crucial for hazard detection, the trade-off with FP must also be considered. FP errors indicate that non-hazardous areas are incorrectly identified as hazardous, potentially leading to unnecessary interventions and economic inefficiencies. The results show that over-sampling tends to reduce FN at the cost of increased FP, whereas down-sampling leads to higher FN but lower FP. The balance between these errors is highly dependent on the application: in high-risk environments, prioritizing FN reduction may be more appropriate, even at the expense of a higher FP rate.

From the experimental results, it is evident that both window size and sampling strategy significantly affect the model's predictive ability. An intermediate window size generally provides the best trade-off between reducing FN and FP. Furthermore, while over-sampling improves sensitivity by lowering FN, it may introduce higher FP rates, necessitating careful selection of the sampling approach based on application priorities. Given the high cost of FN in hazard detection, model configurations should emphasize reducing FN errors while maintaining an acceptable FP level. These findings provide insights into optimizing CNN-based disaster prediction models and suggest that region-specific tuning of hyperparameters is essential for achieving optimal results.



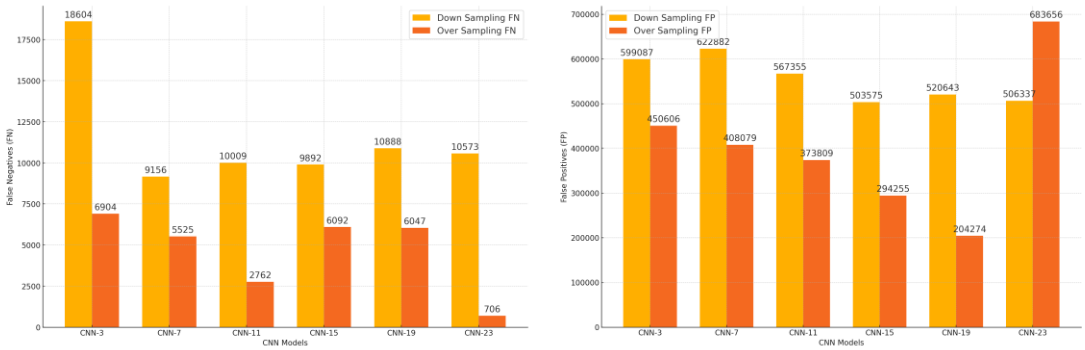

(a) False Negatives (b) False Positives

**Figure 19.** Details of False Negatives (a) and False Positives (b) in HKD based on CNN-i.




(a) False Negatives (b) False Positives

**Figure 20.** Details of False Negatives (a) and False Positives (b) in NIG based on CNN-i.



(a) False Negatives (b) False Positives

**Figure 21.** Details of False Negatives (a) and False Positives (b) in IWT based on CNN-i.

### 3.4.4. Visualization

Figures 22 to 39 illustrate the Landslide Susceptibility Mapping (LSM), Landslide Inventory Mapping (LIM), and the confusion matrix (CM) based on CNN models ranging from CNN-15 to CNN-23. The LSM categorizes landslide risk into five levels: very low, low, moderate, high, and very high, enabling precise risk assessment across various geographic regions. LIM, on the other hand, uses the outcomes from the trained CNN models to test and predict the original dataset. The key difference between LSM and LIM is that LSM assesses potential hazard levels, while LIM focuses on the detection of actual landslide events. The confusion matrix (CM) visualization further provides a detailed analysis of the models' performance, giving insights into accuracy, recall, precision, and misclassification rates within each category. This helps to identify both the strengths and areas where the model needs improvement.

Across different study areas, CNN-19 stands out, outperforming both CNN-15 and CNN-23 in terms of accuracy and predictive reliability, particularly when paired with sophisticated sampling techniques.

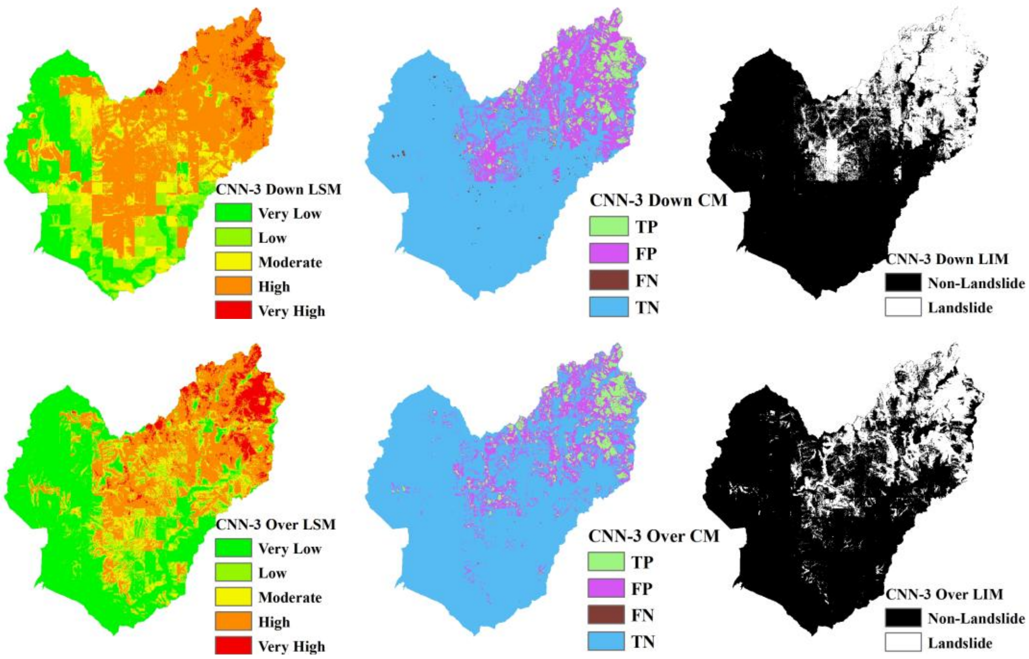



**Figure 22.** LSM, LIM and CM of HKD in CNN-3 (Upper: DownSampling, lower: OverSampling).

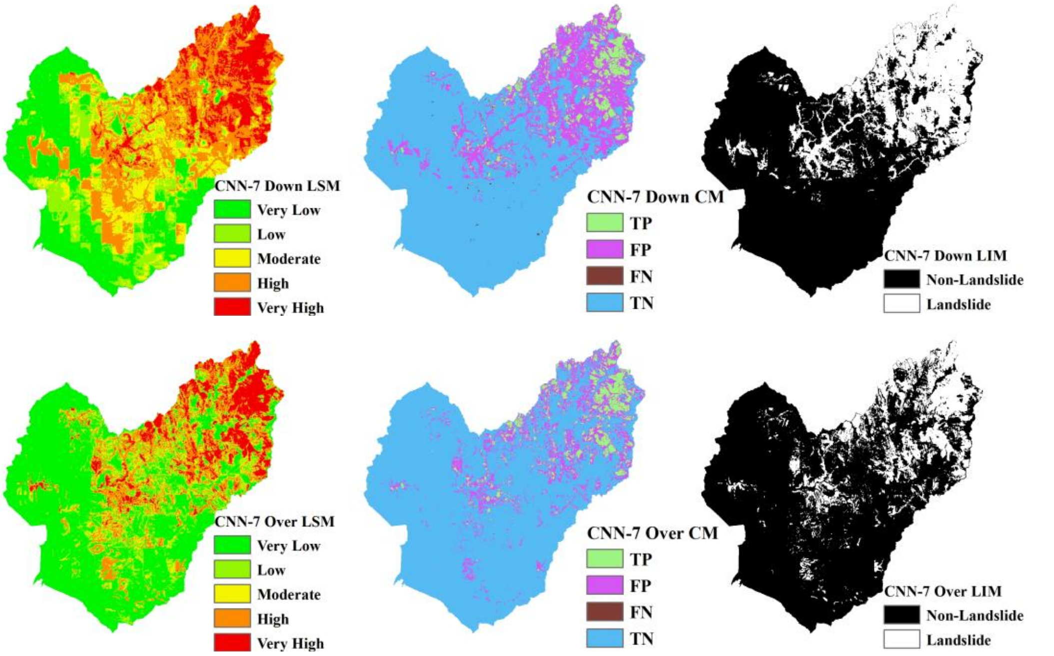



**Figure 23.** LSM, LIM and CM of HKD in CNN-7 (Upper: DownSampling, lower: OverSampling).

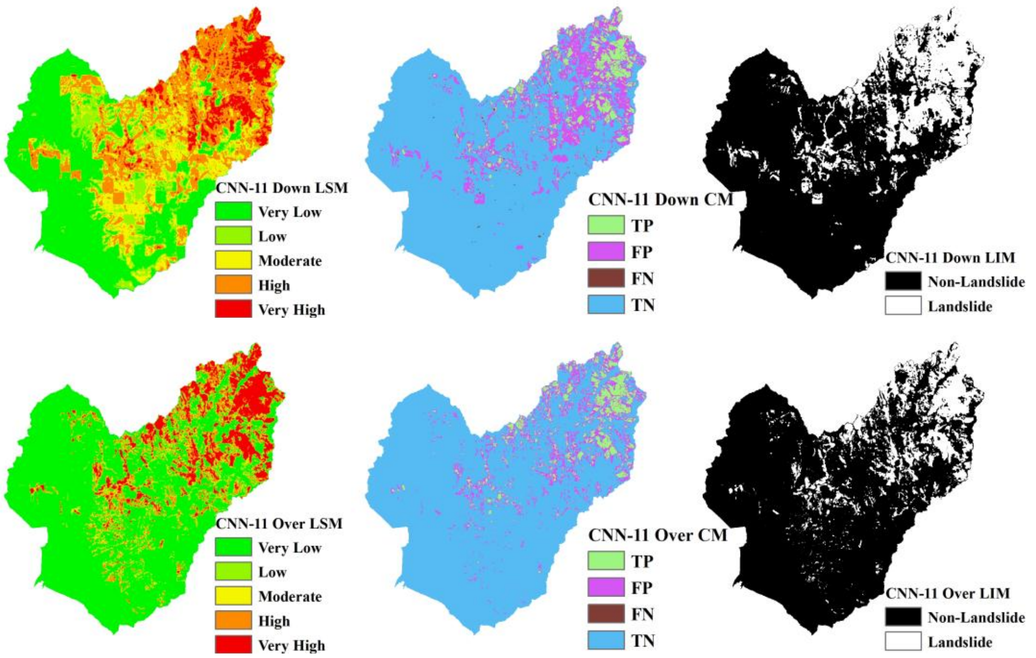



**Figure 24.** LSM, LIM and CM of HKD in CNN-11 (Upper: DownSampling, lower: OverSampling).

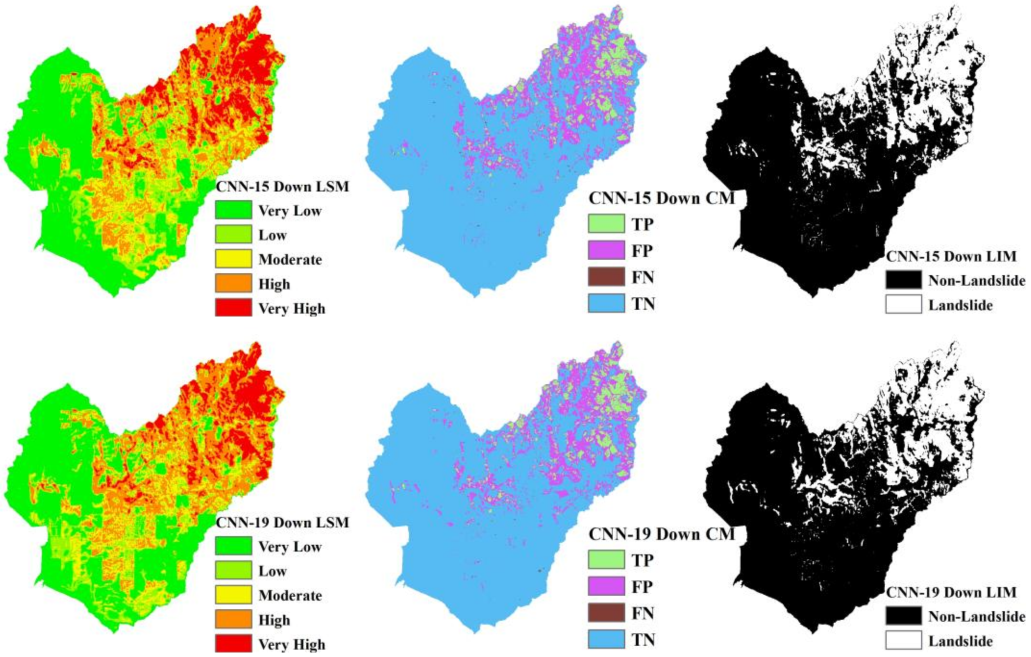



**Figure 25.** LSM, LIM and CM of HKD in CNN-15 (Upper: DownSampling, lower: OverSampling).

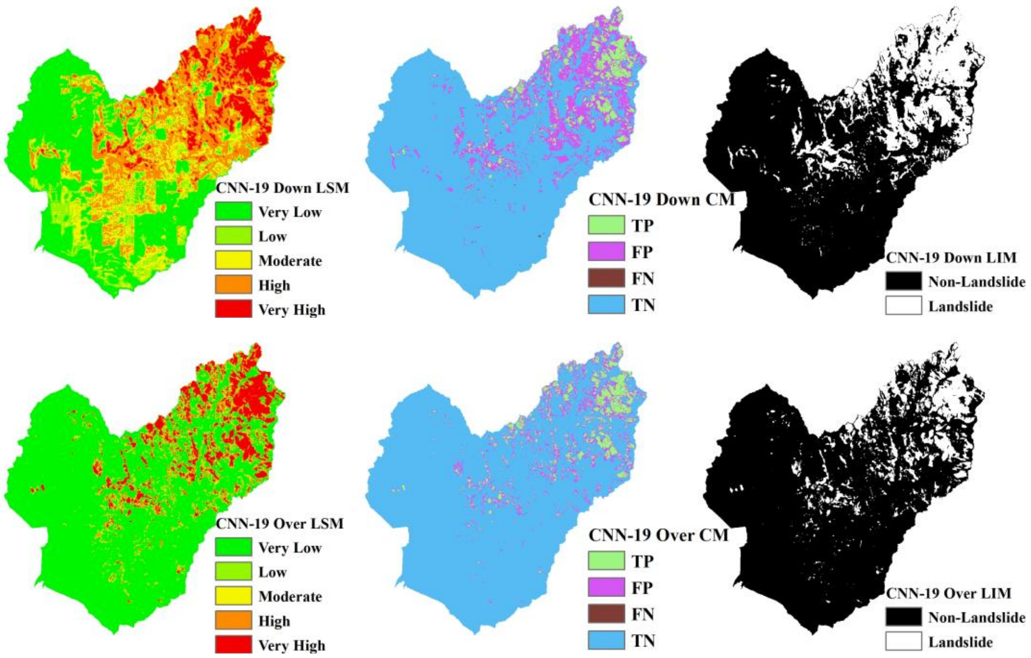



**Figure 26.** LSM, LIM and CM of HKD in CNN-19 (Upper: DownSampling, lower: OverSampling).

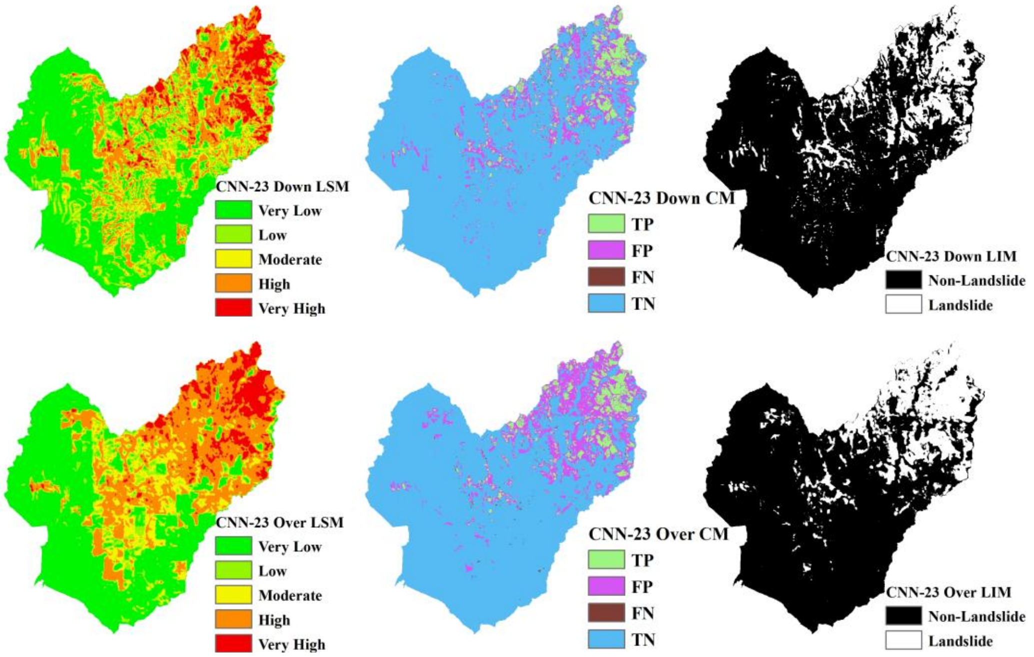



**Figure 27.** LSM, LIM and CM of HKD in CNN-23 (Upper: DownSampling, lower: OverSampling).

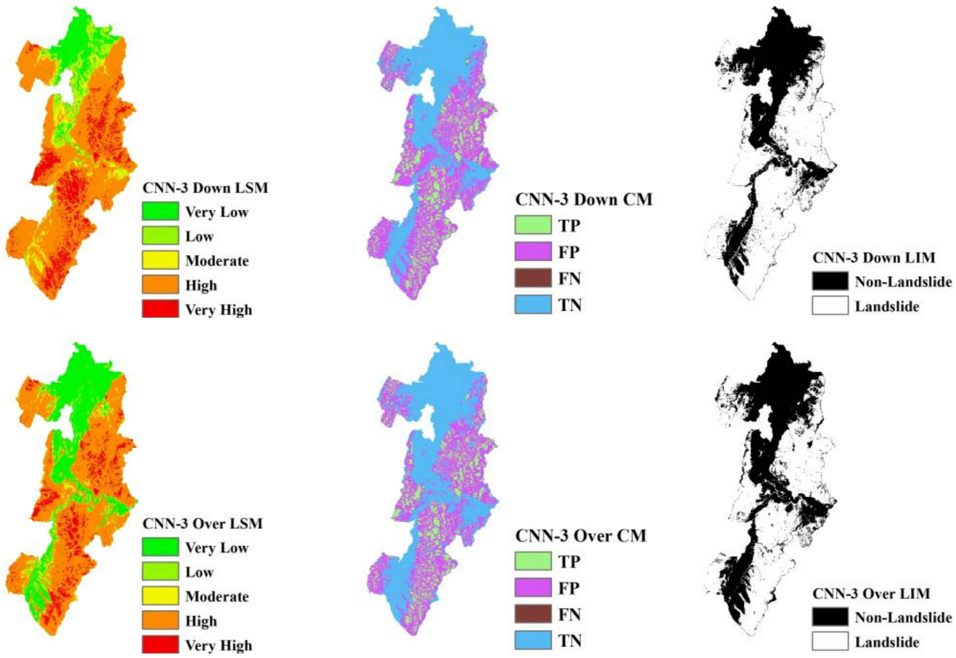



**Figure 28.** LSM, LIM and CM of NIG in CNN-3 (Upper: DownSampling, lower: OverSampling).

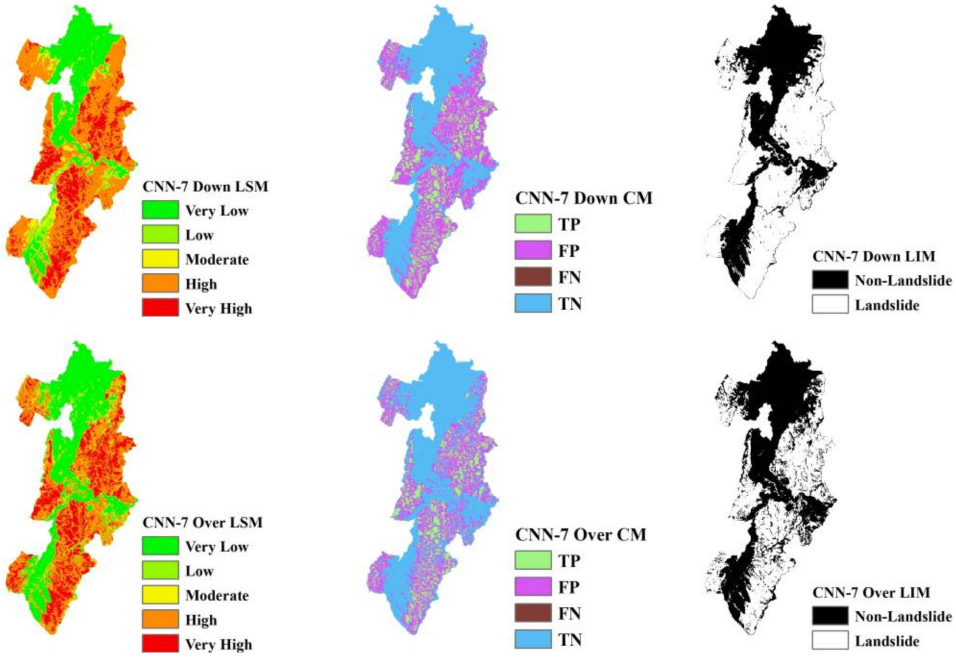



**Figure 29.** LSM, LIM and CM of NIG in CNN-7 (Upper: DownSampling, lower: OverSampling).

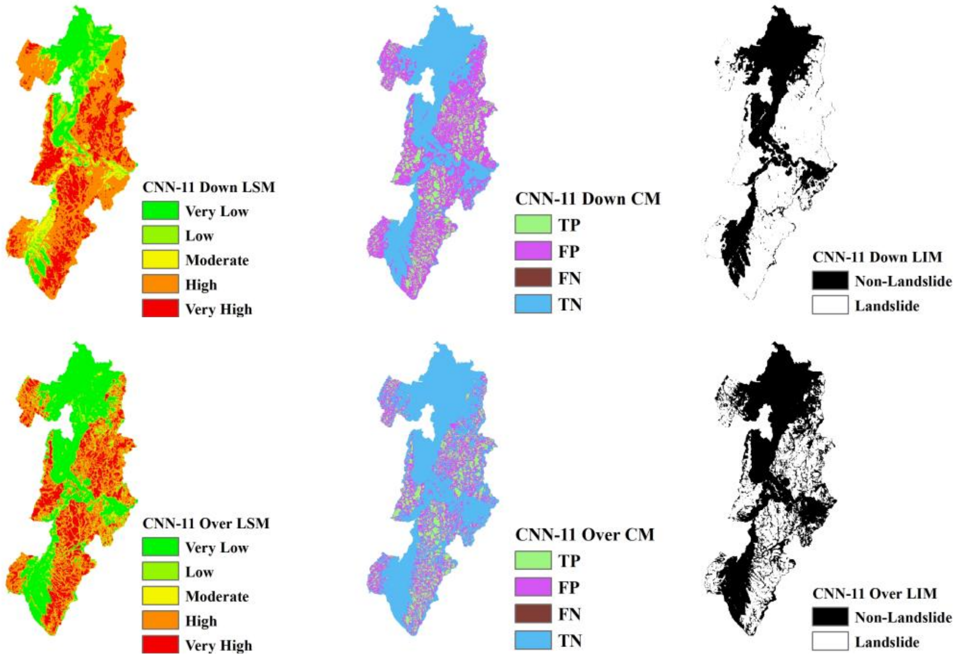



**Figure 30.** LSM, LIM and CM of NIG in CNN-11 (Upper: DownSampling, lower: OverSampling).

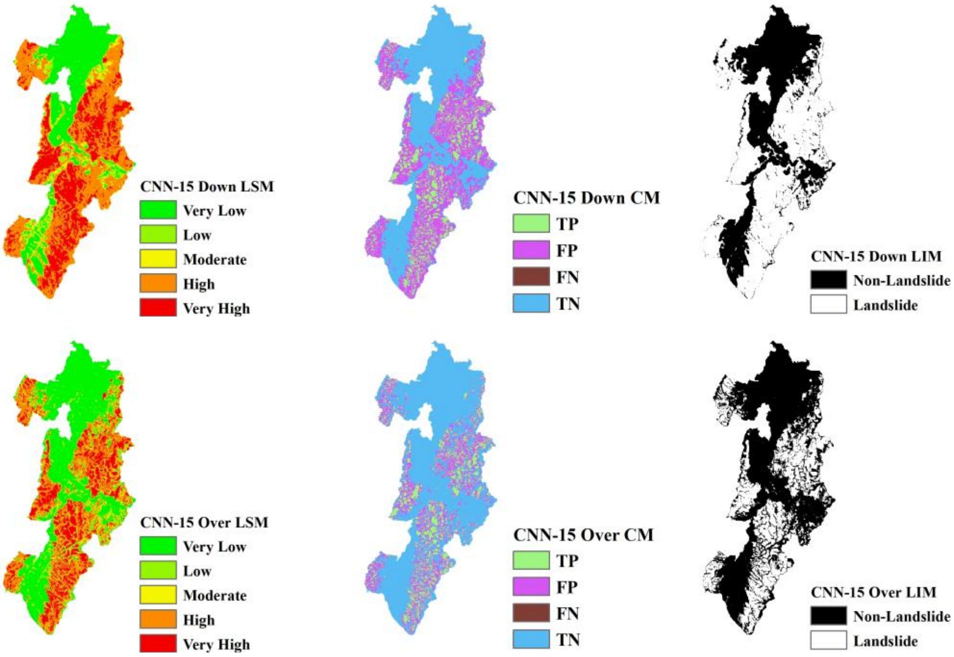



**Figure 31.** LSM, LIM and CM of NIG in CNN-15 (Upper: DownSampling, lower: OverSampling).

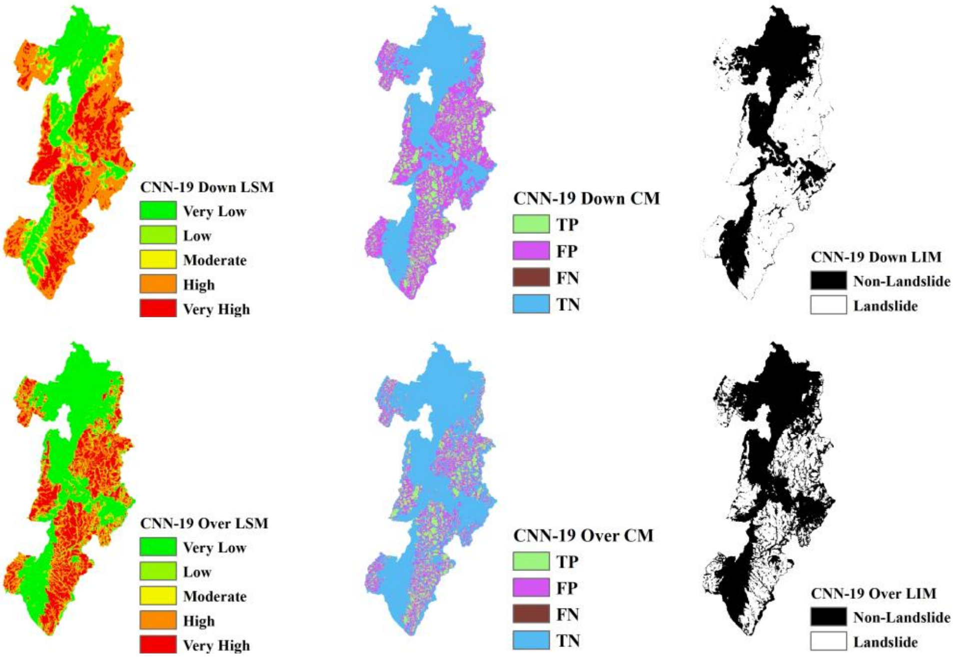



**Figure 32.** LSM, LIM and CM of NIG in CNN-19 (Upper: DownSampling, lower: OverSampling).

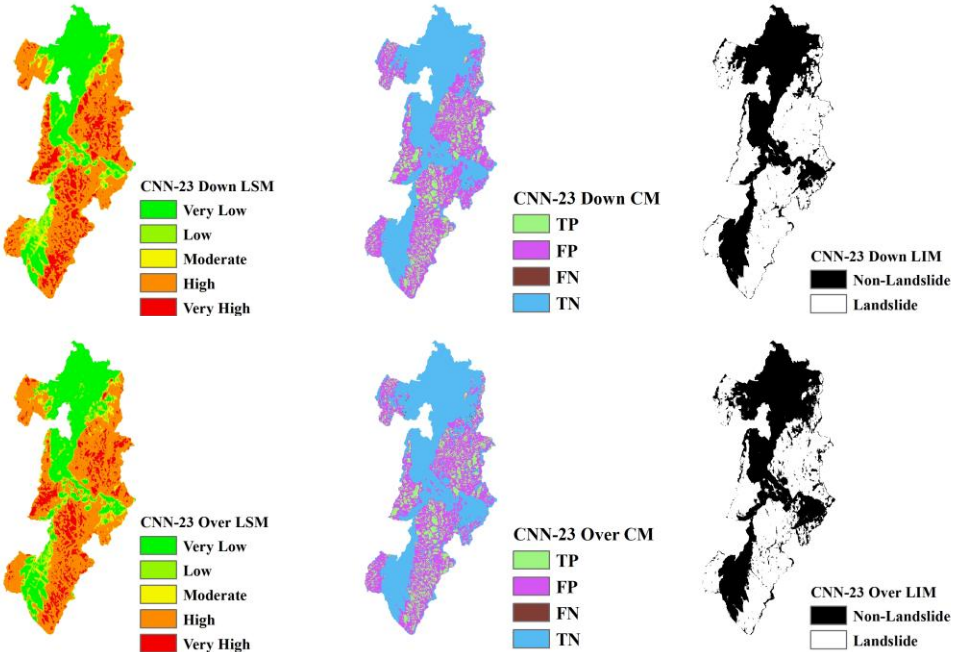



**Figure 33.** LSM, LIM and CM of NIG in CNN-23 (Upper: DownSampling, lower: OverSampling).

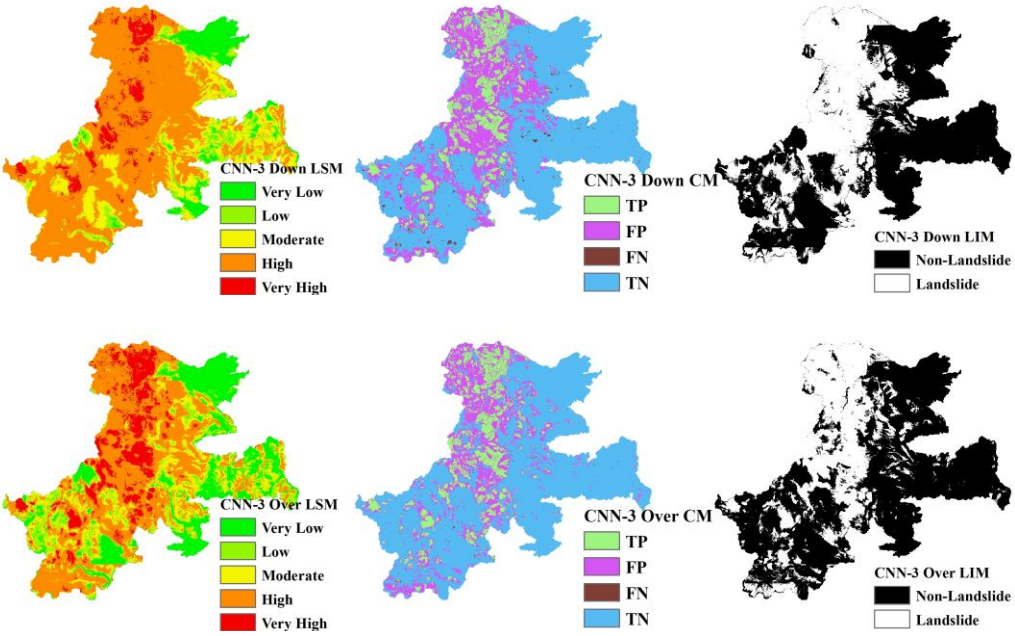



**Figure 34.** LSM, LIM and CM of IWT in CNN-3 (Upper: DownSampling, lower: OverSampling).

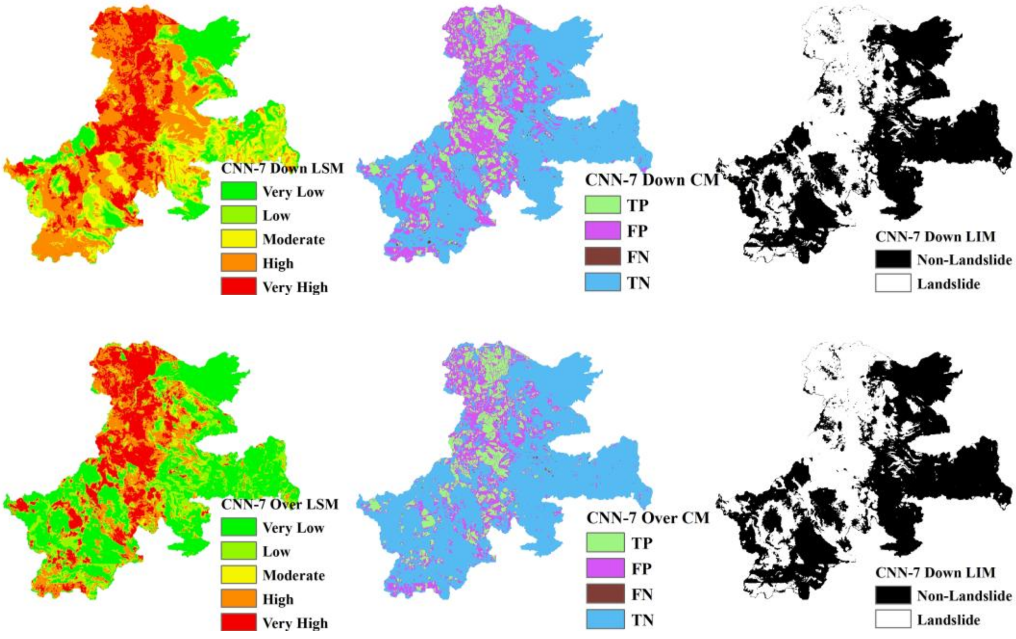



**Figure 35.** LSM, LIM and CM of IWT in CNN-7 (Upper: DownSampling, lower: OverSampling).

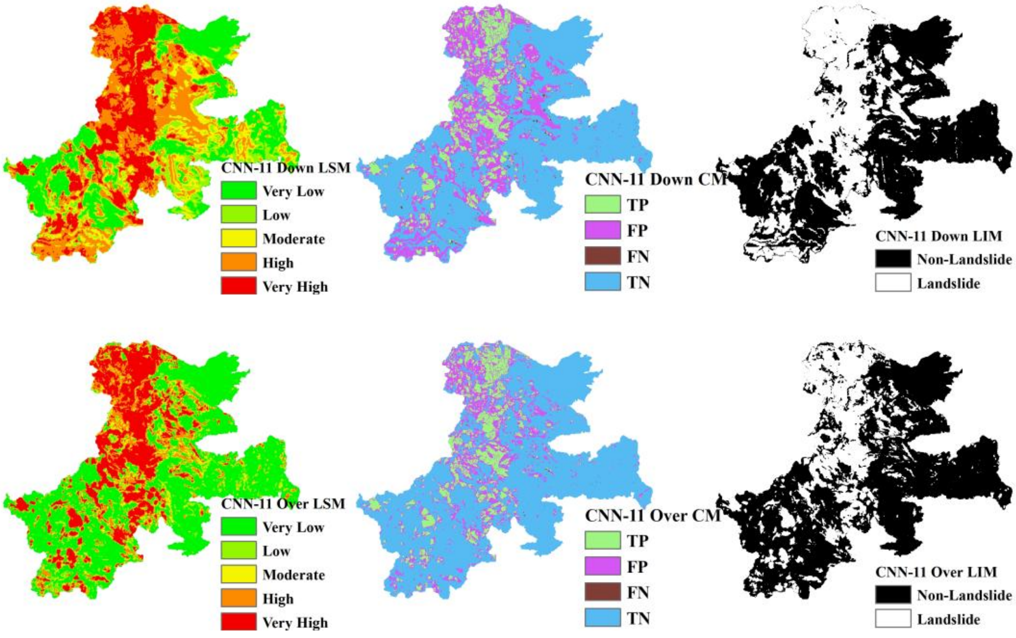



**Figure 36.** LSM, LIM and CM of IWT in CNN-11 (Upper: DownSampling, lower: OverSampling).

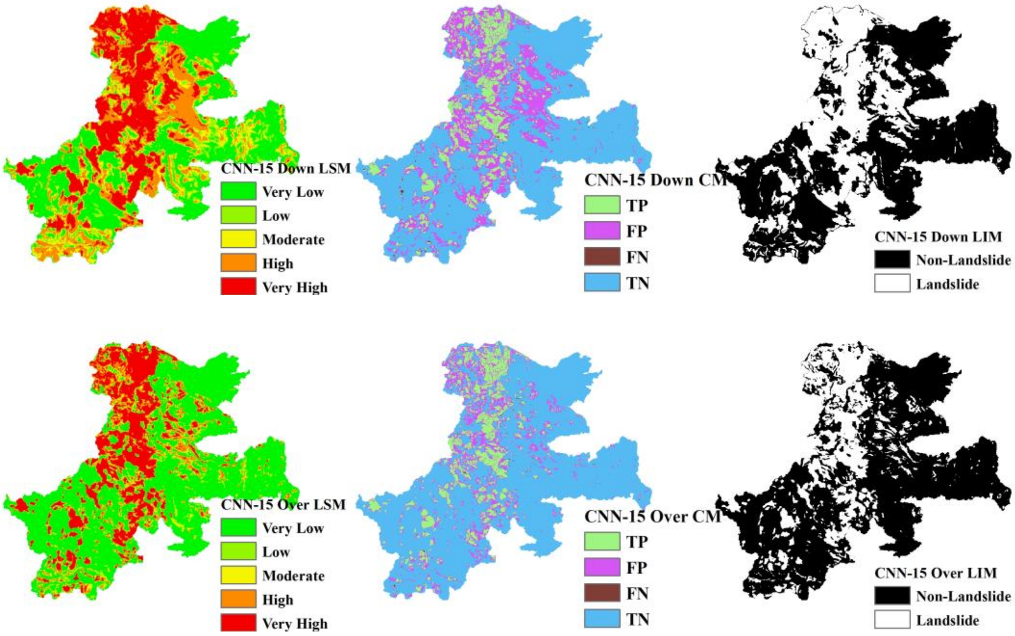



**Figure 37.** LSM, LIM and CM of IWT in CNN-15 (Upper: DownSampling, lower: OverSampling).

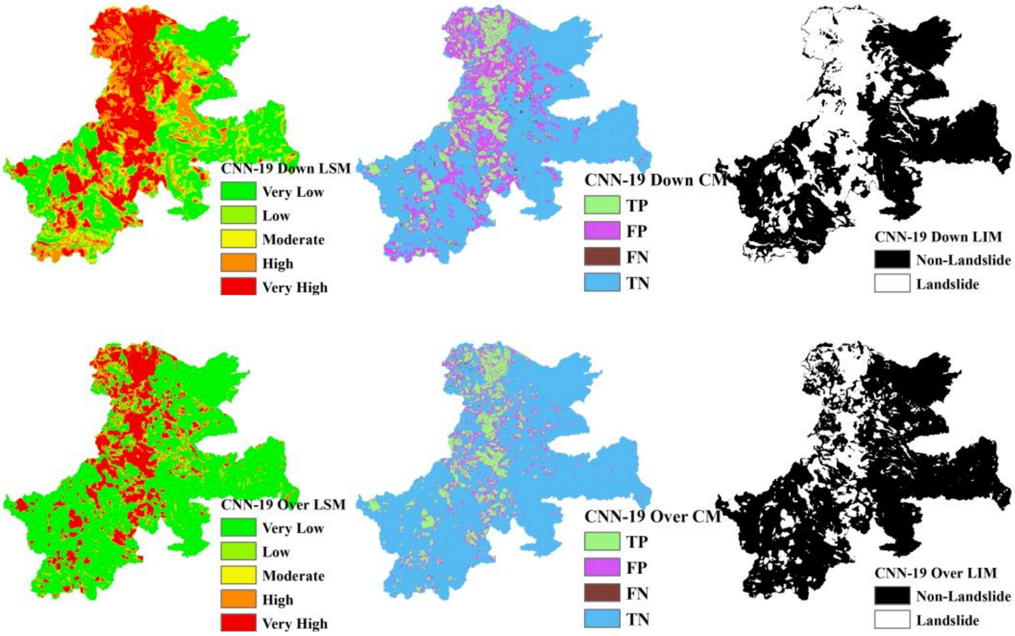



**Figure 38.** LSM, LIM and CM of IWT in CNN-19 (Upper: DownSampling, lower: OverSampling).

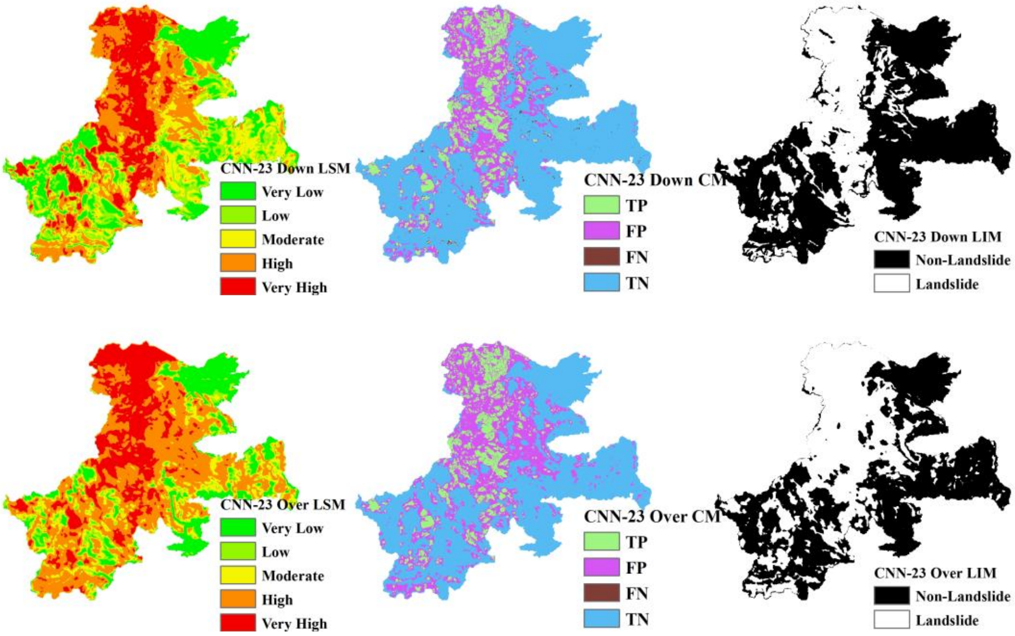



**Figure 39.** LSM, LIM and CM of IWT in CNN-23 (Upper: DownSampling, lower: OverSampling).

3.4.5. Versatility

The accuracy and confusion matrix of the results shown in Table 8 was obtained by applying the CNN models trained on data from one study area to test data from a different study area. According to these results, When HKD is the training area, better results are obtained compared to when NIG and IWT are used as training regions. Similarly, when evaluating based on R-score and AUC, better results are achieved when HKD is the training area. Additionally, LSM corroborates these findings, as depicted in Figures 40 to 42. Although the accuracy is insufficient to perfectly guarantee the versatility of the proposed CNN model, all obtained results are deemed acceptable to some extent due to the significant amount of non-landslide data.

As the possible reason, regarding the results of the Information Gain Ratio (IGR), notable differences are observed between the HKD area and the NIG and IWT regions. In the HKD area, no single feature exhibits significant influence, whereas in the NIG and IWT regions, certain features demonstrate a significantly higher influence compared to others, leading to pronounced regional characteristics. Consequently, the model is influenced by these regional characteristics during the training process. However, such strong regional characteristics are absent in the HKD area. Therefore, when the HKD area is used as the training area, the model's performance when applied to other areas is generally better compared to when the NIG and IWT areas are used as training areas. The variation in IGR among different features across regions suggests that certain features are more influential in some areas than in others. This disparity implies that a model's ability to generalize across regions can be enhanced by identifying and focusing on features that offer the most predictive power universally, rather than those that are region-specific.

Based on the versatility tests conducted in this research, it must be acknowledged that there are significant challenges in achieving versatility. However, the results also point to specific directions for future efforts. To improve the universality of models, two key aspects must be prioritized. Firstly, when selecting features, it is crucial to avoid those with strong regional characteristics, as these can limit the model's applicability across different contexts. Secondly, during the model training and testing processes, it is essential to select appropriate sampling methods and model structures. This ensures the development of robust models that are universally applicable. Despite these focused efforts, it is important to recognize that achieving high-precision generalization with the current models remains a challenging task. Therefore, continued research and refinement of models are necessary to enhance their predictive accuracy and generalizability. Additionally, at the current stage, the best approach still involves assessing landslide risks based on specific regional historical landslide records and characteristics. This method allows for a more targeted and effective risk assessment, leveraging localized knowledge and data to predict and mitigate potential landslide events effectively.

**Table 8.** Different test area results based on different trained area CNN-19 model with oversampling.

| **Train** | **Test** | **Model** | **Confusion matrix** | | | | **ACC** | **P-score** | **R-score** | **F-score** | **AUC** | **MCC** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | TP | FP | TN | FN | | | | | | |
| HKD (Over) | IWT | CNN-19 | 126527 | 524167 | 3864576 | 157572 | 0.854 | 0.194 | 0.445 | 0.271 | 0.785 | 0.225 |
| | NIG | | 46008 | 166957 | 1297727 | 71658 | 0.849 | 0.216 | 0.391 | 0.278 | 0.672 | 0.213 |
| NIG (Over) | IWT | | 71519 | 582257 | 3806486 | 212580 | 0.830 | 0.109 | 0.252 | 0.153 | 0.700 | 0.082 |
| | HKD | | 62254 | 624565 | 3015775 | 61386 | 0.818 | 0.091 | 0.504 | 0.154 | 0.785 | 0.153 |
| IWT (Over) | HKD | | 36024 | 239514 | 3400826 | 87616 | 0.913 | 0.131 | 0.291 | 0.180 | 0.618 | 0.154 |
| | NIG | | 4803 | 12500 | 1452184 | 112863 | 0.921 | 0.278 | 0.041 | 0.071 | 0.385 | 0.081 |

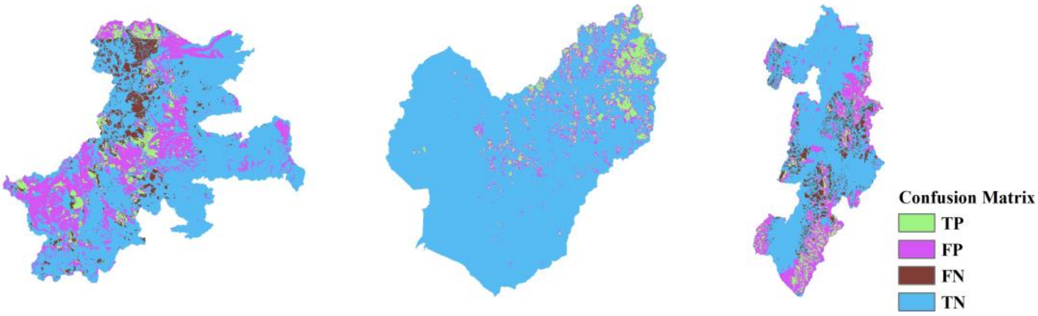



**Figure 40.** CM based on the CNN-19 with the oversampling method in HKD. (Left to Right: IWT, HKD and NIG).

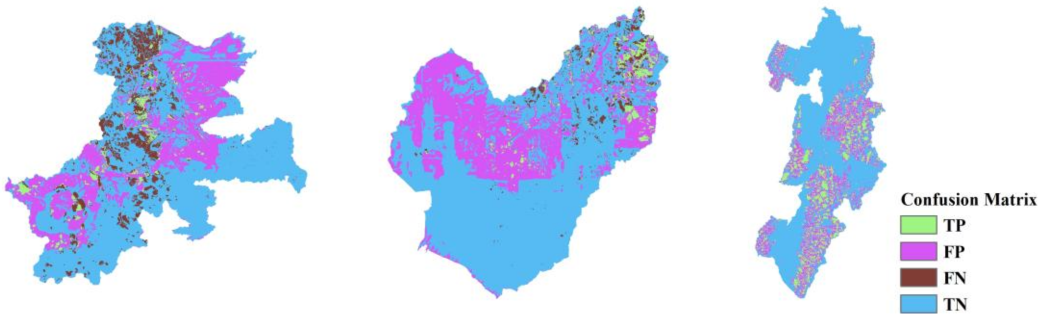



**Figure 41.** CM based on the CNN-19 with the oversampling method in NIG. (Left to Right: IWT, HKD and NIG).

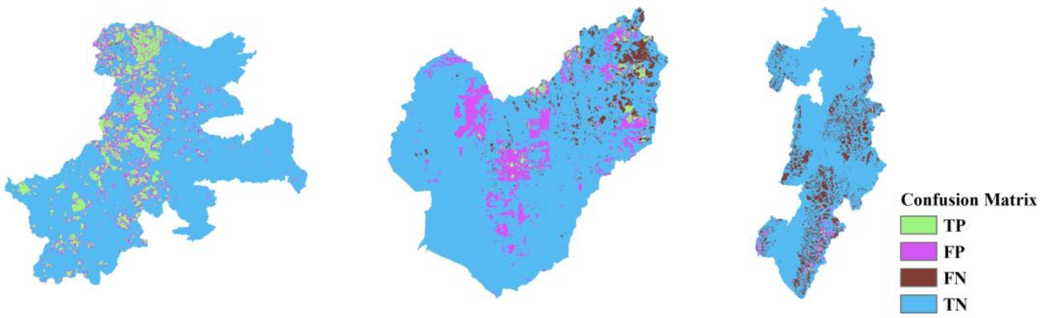



**Figure 42.** CM based on the CNN-19 with the oversampling method in IWT. (Left to Right: IWT, HKD and NIG).

## 3.5. Conclusion

This study extensively explores the effectiveness of convolutional neural network (CNN) models across three distinct regions, employing various evaluation metrics to assess predictive accuracy, model adaptability, and the impact of regional characteristics and sampling techniques on model performance.

The analysis reveals that increasing the window size significantly improves the model's predictive capabilities by enabling the capture of more comprehensive environmental features, crucial for accurately predicting landslide occurrences. The generalizability of CNN-19 across diverse geographical settings further underscores its suitability for comprehensive analysis across extensive study areas. It adapts well to different environmental conditions, ensuring consistency in performance and reliability in predictions. The strategic selection of CNN-19 for landslide scale analysis leverages its capabilities to optimize performance on critical metrics, making it the most appropriate model for such applications. Its ability to balance feature capture with accurate and relevant output makes it a pivotal tool for landslide risk assessment and management, potentially guiding future enhancements and the exploration of hybrid or ensemble models to elevate accuracy and applicability in varied environmental settings. Furthermore, the study explores the implications of varying window sizes, sampling methods, and regional feature sensitivity on model adaptability. Information Gain Ratio (IGR) results indicate that incorporating regional characteristics into model training enhances both adaptability and generalizability, enabling more robust and reliable predictions across diverse landscapes.

However, a significant disparity between the performance of trained models and test models is noted, primarily due to data imbalances within the test datasets. This difference is more pronounced in scenarios where there is a considerable gap between the number of landslide and non-landslide data points in the original dataset. In addressing these imbalances, our comparative analysis of downsampling and oversampling techniques demonstrates that oversampling notably enhances model performance in the test sets. This method proves particularly beneficial in imbalanced scenarios, improving model accuracy and overall effectiveness by providing a more balanced representation of both landslide and non-landslide conditions.

### 3.6. Future works

**Landslide Type Differentiation**

Developing CNN models that classify different types of landslides, such as debris flows, rockfalls, and shallow or deep-seated landslides, will enhance risk assessment accuracy. Integrating geotechnical and geophysical data can help distinguish between rainfall-induced and earthquake-induced landslides. Additionally, time-series analysis can track the progression of slow-moving landslides versus sudden failures, improving early warning capabilities. Incorporating unsupervised learning techniques, such as clustering algorithms, can further automate landslide detection and classification in large-scale datasets.

**Exploring Advanced Models**

Applying Generative Adversarial Networks (GANs) to generate synthetic landslide-prone areas can improve training data diversity and model robustness. Utilizing Recurrent Neural Networks (RNNs) and their variants (e.g., LSTMs, GRUs) will allow models to capture temporal patterns in landslide occurrences, particularly in relation to seasonal variations in rainfall and snowmelt. Hybrid deep learning architectures that integrate CNNs with transformers or attention mechanisms can enhance feature extraction from high-resolution satellite images.

**Real-Time Disaster Monitoring with CNNs**

Developing a real-time prediction system by integrating CNNs with GIS and remote sensing data can support continuous landslide monitoring. Implementing early warning systems based on CNN models can provide probabilistic risk assessments using real-time meteorological data. A cloud-based deployment framework would enable dynamic updates and on-the-fly assessments of potential landslide-prone areas.

**Incorporating Precipitation Data Differentiation**

Distinguishing between different precipitation sources, such as rainfall and snowmelt, rather than treating all precipitation as a single influencing factor, will improve prediction accuracy. Incorporating seasonal hydrological variations can help account for the delayed impact of snowmelt in mountainous regions, which significantly contributes to landslide occurrences. Multi-source satellite imagery (e.g., MODIS, Sentinel-1/2, Landsat) can be utilized to track snow cover changes, ensuring a more comprehensive understanding of how snowmelt affects soil saturation and slope stability.

**Expanding the Study Area and Data Sources**

Extending the research to different geographic regions with varying climatic conditions will assess the transferability of the trained CNN model across multiple landscapes. Incorporating high-resolution digital elevation models (DEMs) can improve topographical feature extraction and refine landslide susceptibility mapping. Leveraging multi-temporal and multi-spectral remote sensing datasets will provide a more detailed characterization of terrain and land-use changes over time.

By incorporating landslide type differentiation, advanced models, and real-time monitoring techniques, future research can provide more refined risk assessments and mitigation strategies, allowing authorities to prioritize response efforts based on the specific characteristics and severity of landslide-prone areas.

# 4. Bridge Risk Management Based on Machine Learning with Driftwood Generation Potential by Deep Learning

## 4.1. Research Purpose

Building on the demonstrated feasibility of traditional machine learning and CNN models in landslide risk prediction, this chapter utilizes the models and results from previous sections to advance the study of driftwood generation and its associated risks.

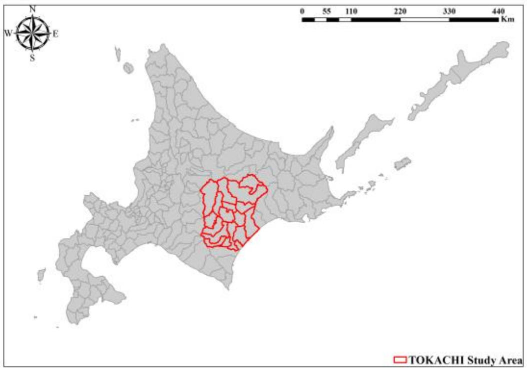
The innovation of this study lies in its refined approach to identifying sources of driftwood generation compared to previous research. The primary goal is to develop a more accurate and robust method for predicting and managing the risks associated with driftwood in river environments, thereby improving safety in bridge infrastructure. While earlier studies primarily focused on areas with slopes exceeding 30° as sources, this study advances the methodology by integrating landslide risk maps generated from CNN and traditional machine learning predictions. This integration enables more precise delineation of the impact of slope collapses on driftwood generation potential.

Furthermore, the calculated driftwood generation potential and driftwood capture amount are used as key features in assessing bridge risk. These features, combined with existing environmental and bridge characteristics, enhance the predictive framework for bridge risk. Additionally, this approach facilitates a comparison between CNN-based and traditional machine learning techniques in bridge disaster risk assessment, providing valuable insights for infrastructure safety planning.

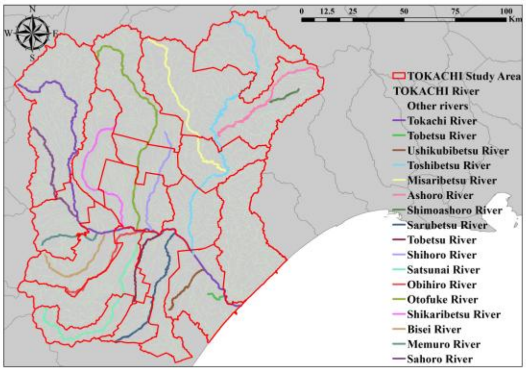
## 4.2. Study Area and Data

### 4.2.1. Study Area

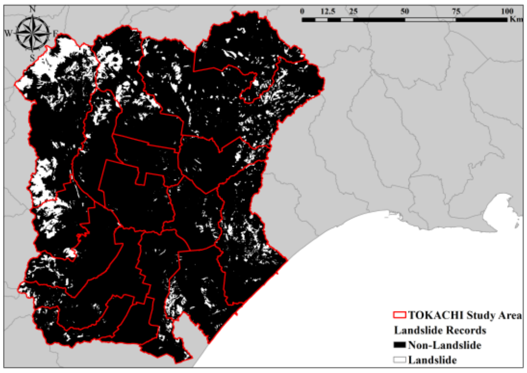
The focus of this study is the Tokachi River basin, as shown in Figure 43(a)(b), located in the central and eastern parts of Hokkaido, near Obihiro City in Japan. In 2016, Typhoon No. 10 inflicted severe damage on the area, leading to the discharge of substantial quantities of driftwood. Over 20,000 $\boldsymbol{m^3}$ of driftwood accumulated near the mouth of the Tokachi River, causing significant disruptions. Additionally, several bridges in the area sustained significant damage, exhibiting various types of failure patterns [45].



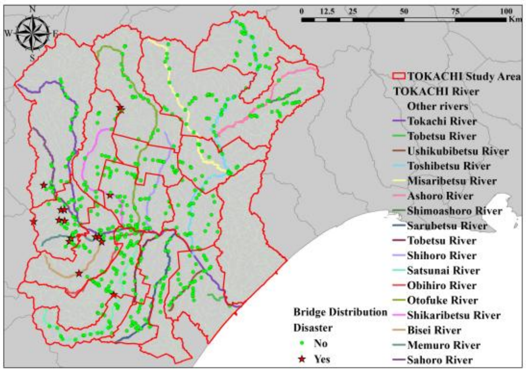

(a) Location of study area




(b) Distribution of major rivers




(c) Distributions of historical landslides




(d) Distribution of bridges

**Figure 43.** (a) Location of study area, (b) Distribution of major rivers, (c) Distributions of historical landslides, (d) Distribution of bridges in the study area.

### 4.2.2. Data Used in This Study

#### 4.2.2.1. Date used for landslide risk assessment

In this study, we divided the entire target basin into 10-meter mesh grids to evaluate slope disaster risk and estimate driftwood potential. For landslide risk prediction, we utilized the landslide topography distribution map database provided by the Disaster Prevention Science Institute [26]. Data for landslide factors were generated using national numerical information from the Geospatial Information Authority of Japan and the Ministry of Land, Infrastructure, Transport, and Tourism [46]. Terrain information, including elevation, slope angle, slope direction, and average annual rainfall, was derived from 10-meter DEM data provided by the Geospatial Information Authority of Japan's base map information download service, using ArcGIS tools. Qualitative factors such as terrain, soil, and surface conditions were also extracted every 10 meters from the same service [47]. Additionally, a vegetation classification map was created using vegetation survey data from the Ministry of the Environment [48]. These are detailed in table 9 and figure 44.

**Table 9.** Features and data range in the study area for landslide risk.

| Feature | Resource | TOKACHI |
|---|---|---|
| HLS | NIED | 0, 1 |
| DEM (m) | GIAJ | 0 to 2136 |
| Slope (°) | DEM | 0 to 82.8999 |
| Aspect | DEM | -1 to 359.61 |
| Curvature (dimensionless) | DEM | -116.591 to 42.9292 |
| Average rainfall (mm) | JMA | 761.015 to 1543.07 |
| Distance to River (m) | MLIT | 0 to 3517.74 |
| Distance to Road (m) | MLIT | 0 to 55769.9 |
| Distance to Fault (m) | GSJ | 0 to 48467.8 |
| Landuse | MLIT | 12 types |
| Soil | MLIT | 46 types |
| Lithology | GSJ | 66 types |
| Vegetation | BiodiC-J | 79 types |
| Catchment Area | DEM | 1 to 12762 |
| Openness (°) | DEM | 0 to 90 |

[1] NIED: National Research Institute for Earth Science and Disaster Resilience.

[2] GIAJ: Geospatial Information Authority of Japan.

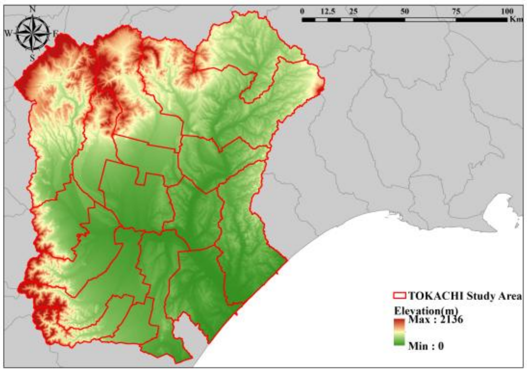
[3] DEM: Digital Elevation Map with 10 m x 10 m resolution published by Geospatial Information Authority of Japan.

[4] JMA: Japan Meteorological Agency.

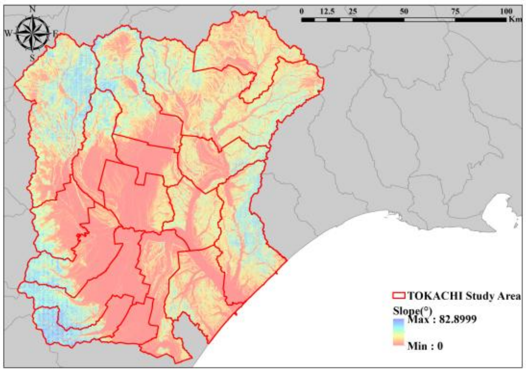
[5] MLIT: Ministry of Land, Infrastructure, Transport and Tourism.

[6] GSJ: Geological Survey of Japan, National Institute of Advanced Industrial Science and Technology.

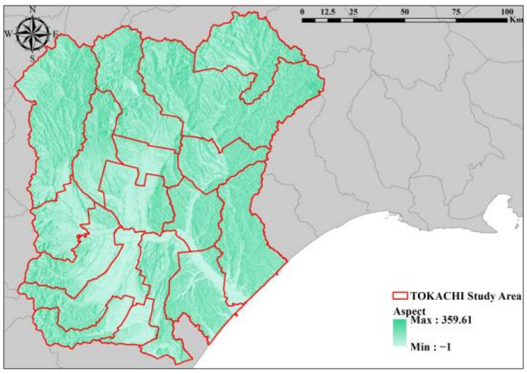
[7] BiodiC-J： Biodiversity Center of Japan.



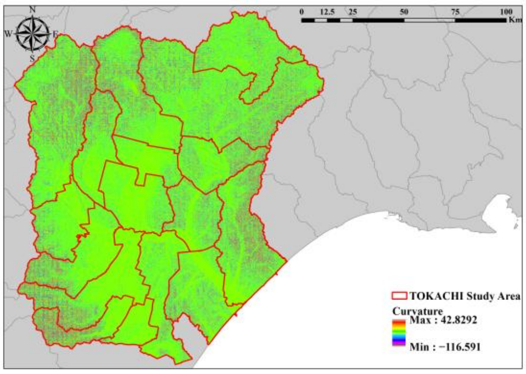

(a) Elevation (m)



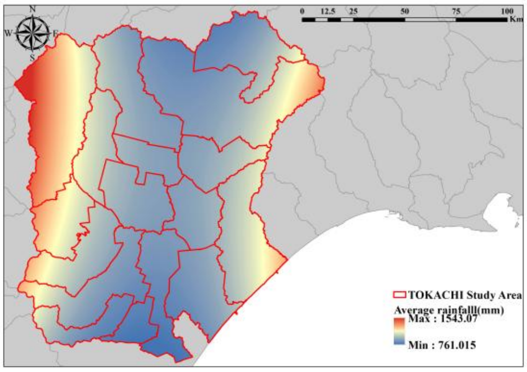

(b) Slope (degrees)



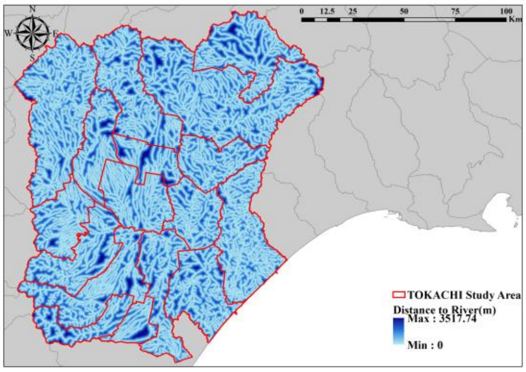

(c) Aspect



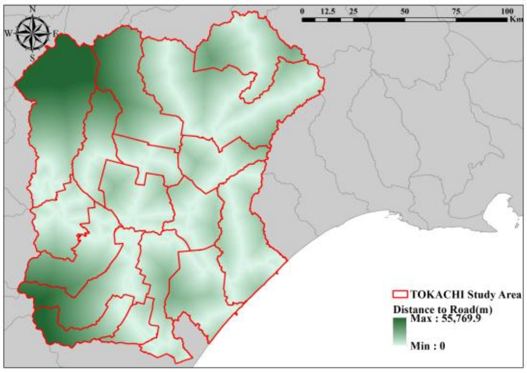

(d) Curvature



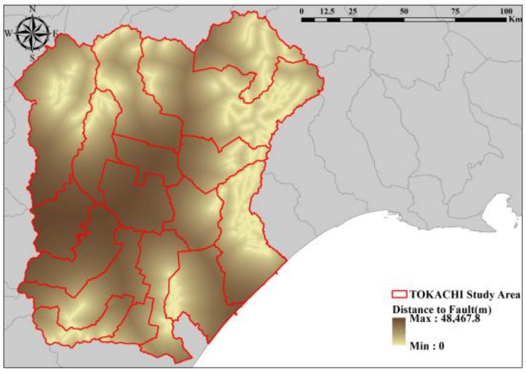

(e) Average rainfall (mm)



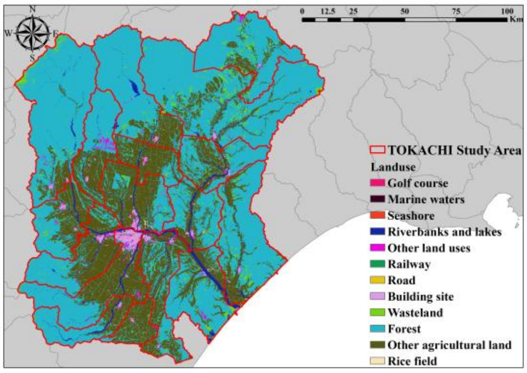

(**f**) Distance to rivers (m)



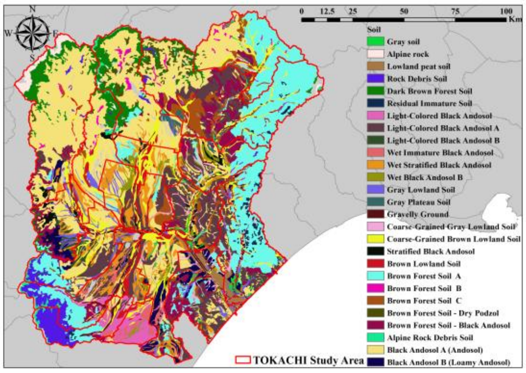

(g) Distance to roads (m)



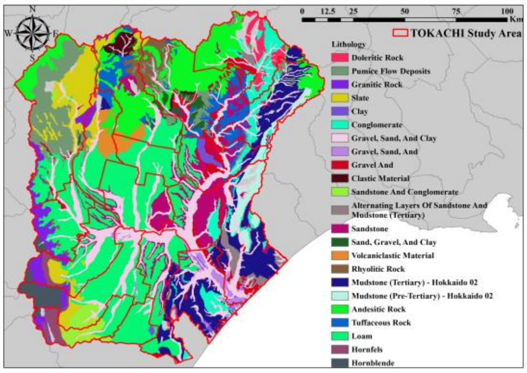

(h) Distance to faults (m)



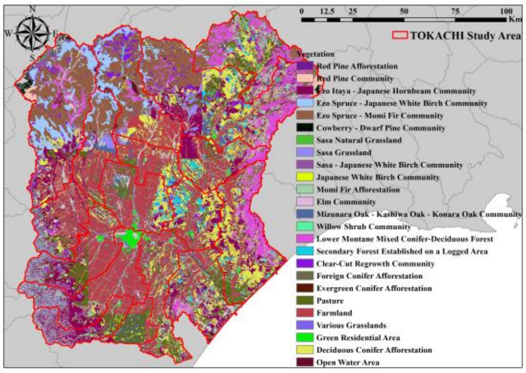

(i) Landuse




(j) Soil




(k) Lithology




(l) Vegetation

**Figure 44.** Maps of landslide features in Tokachi study area.

4.2.2.2. Data used to assess the risk of river crossing bridges during heavy rain

The dataset for disaster risk prediction focuses on road bridges along Tokachi prefectural roads and national highways. It excludes structures such as overpasses, rail bridges and the bridges for which data were not available from the general bridge diagrams [49]. In total, the dataset encompasses 515 bridges, 16 of which sustained damage. The history of bridge disasters was compiled by reviewing local government websites and disaster reports from the Japanese Geotechnical Society[50-52], specifically for the heavy rainfall event in August 2016. Bridge registers were provided by the Ministry of Land, Infrastructure, Transport and Tourism Hokkaido Development Bureau, and general bridge diagrams were supplied by the Hokkaido Department of Construction, as shown in Table 10. Figure 43(d) illustrates the distribution of bridges in the study area.

**Table 10.** Features and data range in the study area used for bridge risk assessment.

| Feature | Resource | TOKACHI |
|---|---|---|
| Bridge Disaster History | MLIT-HRDB<br>HPG-CD | 0, 1 |
| River Width (m) | GIAJ | 0.3 to 99.920436 |
| Curvature (dimensionless) | GIAJ | -116.591 to 42.9292 |
| Catchment Area | DEM | 1 to 12762 |
| Riverbed Slope (%) | DEM | 0 to 4.22 |
| Surface Geology | MLIT | 6 types |
| Intersection Angle (°) | GIAJ | 33.3273 to 145.3979 |
| Span Number | | 1 to 23 |
| Completion Year | MLIT-HRDB | 6 types |
| Bridge length (m) | HPG-CD | 3.0 to 984.2 |
| Impediment ratio of river flow | | 0 to 0.0932 |
| Driftwood Capture Amount ($m^3$) | | 0 to 319.406113 |

[1] MLIT-HRDB: Ministry of Land, Infrastructure, Transport and Tourism, Hokkaido Regional Development Bureau, Construction Department, Road Construction Division.
[2] HPG-CD: Hokkaido Prefectural Government, Construction Department, Civil Engineering Bureau Road Division.
[3] GIAJ: Geospatial Information Authority of Japan.
[4] DEM: Digital Elevation Map with 10 m x 10 m resolution published by Geospatial Information Authority of Japan.
[5] MLIT: Ministry of Land, Infrastructure, Transport and Tourism.

### 4.3. Method

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation, as well as the experimental conclusions that can be drawn.

4.3.1. Landslide Risk Assessment

The detailed workflow of landslide risk can be broadly categorized into four stages. Initially, the database is established, and all feature data is integrated and unified using ArcGIS. Subsequently, the data is preprocessed, and the significance of features is assessed. Following that, a trained model is constructed, and its performance is evaluated. Finally, the preprocessed data is reintroduced into the trained model for testing and prediction in the fourth step.

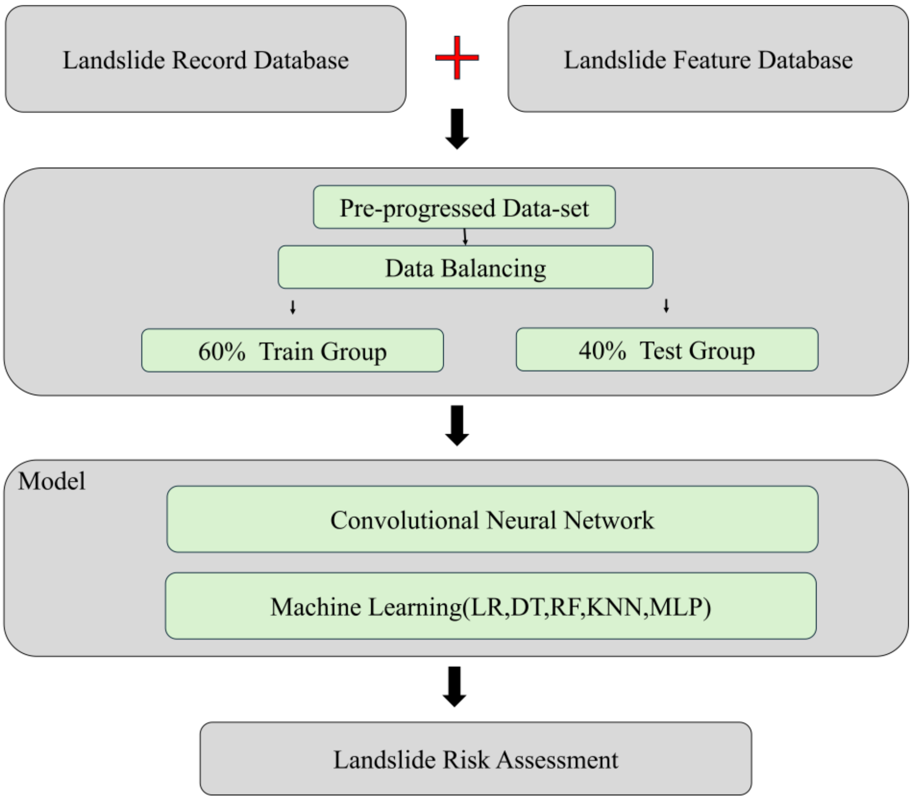



**Figure 45.** Flowchart of landslide risk assessment.

4.3.2. Driftwood Generation Potential and Bridge Risk

The origins of driftwood include the outflow of standing trees, fallen trees, logged timber, and materials used for construction. In our study, we identify landslides as the primary cause of driftwood generation, particularly in mountainous regions where landslides involve coniferous or broadleaf tree species. We assume that tree collapses due to landslide failures are the main natural source of driftwood. This study focuses on the sliding of standing trees associated with slope failures to understand the mechanism of tree movement and its environmental impact. By integrating these natural sources with bridge feature data, we use a machine learning model to predict and quantify the risk to bridges, effectively assessing bridge risk within our framework based on the potential for driftwood generation and capture, shown in figure 46.

As previously discussed, the estimation of driftwood generation potential in this study is based on the proposed methodology [20, 21]. The potential volume of driftwood is calculated by assuming that all areas identified as high-risk for landslides actually experience collapses. This approach allows for the estimation of the maximum potential driftwood volume that could be generated by landslides.

$$V_i = \sum_k \beta_k A_{ki} \tag{13}$$

Here, $V_i$ $(km^3)$ represents the driftwood generation potential at any given bridge, $\beta_k (m^3/km^2)$ denotes the volume of driftwood runoff per unit area of forest for each specific tree type $k$, and $A_{ki}$ $(km^2)$ represents the forest area within the driftwood supply source at location $i$ in the watershed. The value of $\beta_k$ varies based on forest vegetation types [53], with coniferous trees assigned a value of 1000, broadleaf trees 100, and barren lands 0, as illustrated in figure 47(a).

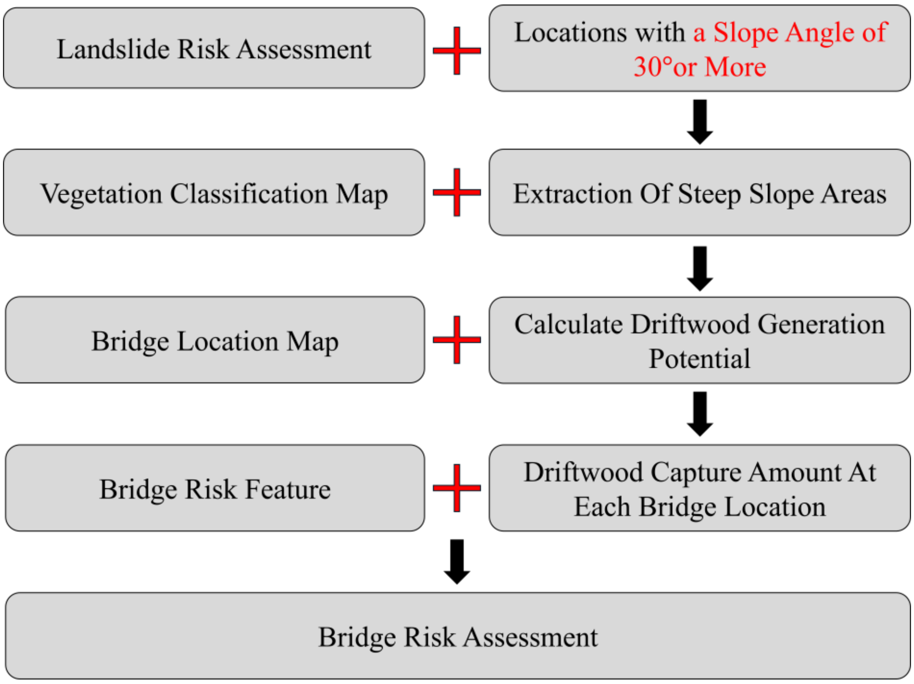
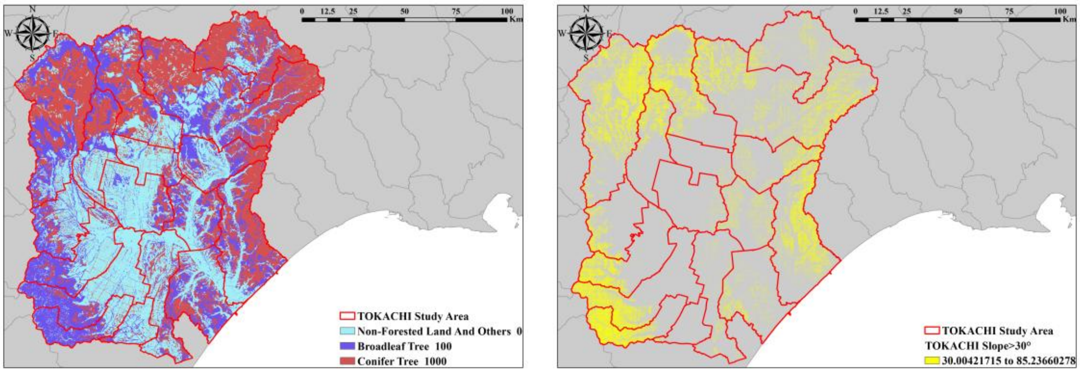



**Figure 46.** Flowchart for assessing the risk of river-crossing bridges during heavy rain.




(a) Vegetation reclassification

(b) Landslide susceptibility area with slopes exceeding 30°

**Figure 47.** Distribution map of (a) vegetation reclassification and (b) landslide susceptibility area with slopes exceeding 30 degrees.

The driftwood capture rate estimates the amount of driftwood intercepted by a specific bridge. Driftwood originating upstream is carried downstream and often accumulates at river crossings, including bridges and dams. The likelihood of this driftwood being captured is influenced by factors such as the river's width and the bridge's span. The formula that describes this relationship is provided below.

$$\alpha_i = \frac{L_w}{W_i} \times \frac{1}{k} \tag{14}$$

Here, $L_w$ represents the assumed length of driftwood determined from field surveys in forests within the watershed, However, in this study, assumed $L_w = (\mathrm{DSM} - \mathrm{DEM})/2$, which will be explained in detail later. $W_i(m)$ represents the minimum channel width at high water levels at bridge $i$, and $k$ is a coefficient representing the difficulty of driftwood accumulation at the bridge. Due to insufficient data on bridges and driftwood, it is impossible to determine the constant $k$, but regardless of the numerical value, it is possible to represent the amount of driftwood captured relatively [21]. In this study, we set $k$ = 1 so that the driftwood capture rate does not exceed 1, calculate the relative amount of driftwood captured by each bridge, and compare the driftwood capture trends between different bridges.

To determine the amount of driftwood each bridge captures based on the calculated driftwood generation potential, as shown in figure 48, the following recurrence formula is used:

$$\tilde{V}_i = \left(V_i - \tilde{V}_{i-1} - \tilde{V}_{i-2} \cdots \tilde{V}_1\right)\alpha_i = \alpha(V_i - \sum_{k=1}^{i-1} \tilde{V}_i) \qquad (15)$$

Here, $\tilde{V}_i$ represents the amount of driftwood captured at bridge $i$, and $\alpha_i$ represents the driftwood capture rate for bridge $i$

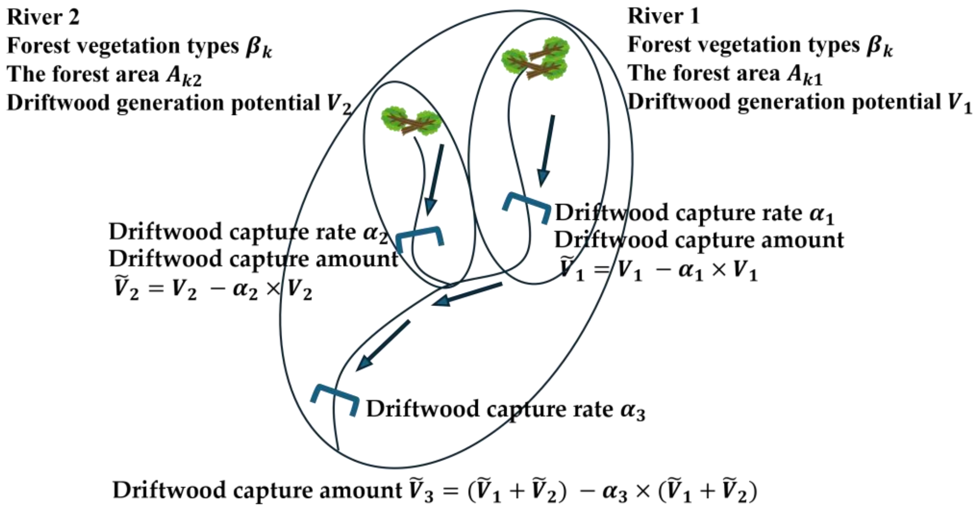



**Figure 48.** Method for calculating the Driftwood capture amount.

In this research, areas with slopes exceeding 30 degrees, as indicated on the landslide susceptibility map (Figure 47(b)), are identified as potential driftwood sources. By concentrating on high-risk landslide zones with steep slopes, the study targets primary sources of driftwood, which are largely derived from trees that fall due to landslides. The maximum distance $L_{max}$ that driftwood can travel to reach river channels is determined using an empirical formula that estimates soil displacement from slope collapses, taking into account the slope's vertical profile [26]. Trees located where the distance from the river to the slope exceeds the estimated maximum distance are excluded in the driftwood production estimate, while trees within this range are included as shown in figure 49 [20, 54].

$$\frac{H}{L_{max}} = 0.73 \tan\theta - 0.21 \tag{16}$$

Here, $H', L', \theta$ ($\tan\theta = H'/L'$) represents the height, length and angle of inclination of the slope. $H$ and $L_r$ represent the elevation and distance of the slope from the nearest river channel, respectively.

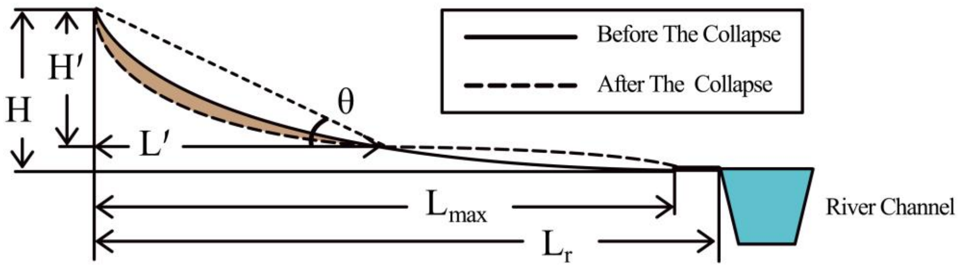



**Figure 49.** Method for calculating the travel distance of sediment from landslide collapses [20, 54].

### 4.4 Result and Discussion

4.4.1. Feature Importance

In this study, we employed the Information Gain Ratio (IGR) method to analyze the training dataset. A higher IGR value indicates a greater importance of a causative factor for landslides or bridges within the models. Factors with an IGR of zero are deemed non-contributive and are recommended for exclusion from the models [37, 38]. As illustrated in figures 50 and 51, the IGR values for all 15 factors associated with landslides were above zero, confirming their relevance in modeling within this study. And 12 factors related to bridges, only when span=1, span number and impediment ratio of river flow are 0, so they will be removed in the subsequent study. Notably, as shown in figure 44, the driftwood capture amount plays a crucial role in predicting the hazard levels of bridges, providing a robust basis for this research. Consequently, all identified causative factors for landslides risk and bridge damage risk were included in the modeling process.

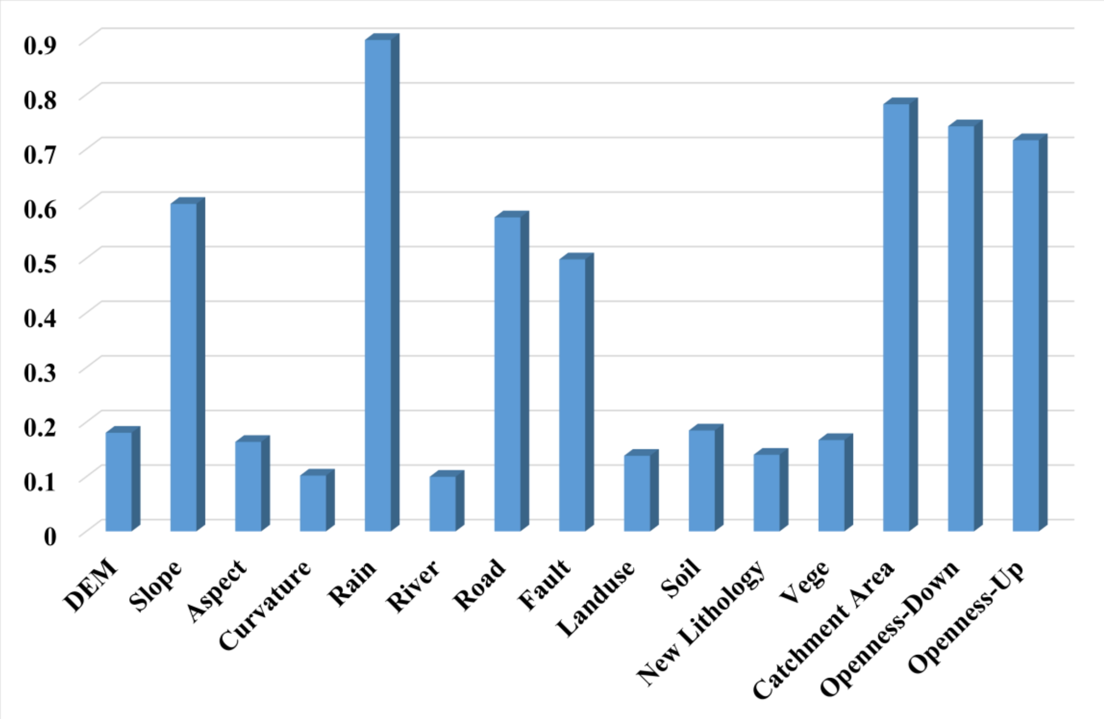



**Figure 50.** Information gain ratio of the features used for the landslide risk assessment.

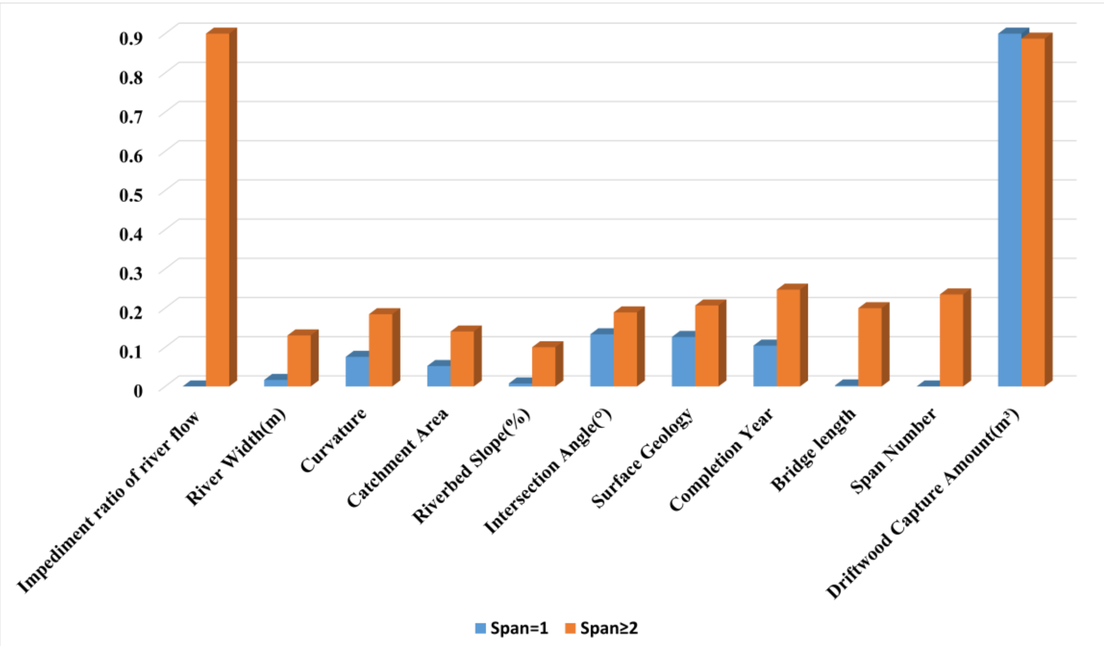



**Figure 51.** Information gain ratio of the features used for the bridge disaster risk assessment.

### 4.4.2. Landslide Risk Assessment

Table 11 and figure 52 depict a comparative analysis of different machine learning models applied for landslide susceptibility mapping (LSM) in the Tokachi region. Each model's performance in classifying the risk levels of landslide occurrence is assessed using Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) metrics, and their predictive power visualized through landslide susceptibility maps.

Decision Trees (DT) exhibit an AUC of 0.832, demonstrating good performance. K-Nearest Neighbors (KNN) reaches an impressive AUC of 0.946, indicating excellent prediction accuracy. Logistic Regression (LR) shows a lower AUC of 0.681, suggesting less effectiveness in this application. Random Forests (RF) have a high AUC of 0.889, reflecting strong predictive capabilities. Multi-Layer Perceptron (MLP) and Convolutional Neural Networks (CNN) lead with AUCs of 0.982 and 0.993 respectively, suggesting outstanding performance in predicting landslide susceptibility.

The susceptibility maps generated by CNN and MLP models are detailed, showing varying risk levels across the region. In CNN-based map, it highlights a broader area under 'very high' and 'high' risk categories compared to the MLP model. CNN seems to identify more potential high-risk zones, which could indicate a higher sensitivity to factors contributing to landslides. In MLP-based map, Displays a more conservative distribution of 'high' and 'very high' risk areas. This might suggest that the MLP model prioritizes minimizing false positives or is less sensitive to the variables that influence landslide occurrence.

The CNN model's superior performance in ROC and AUC assessments and its detailed visualizations on the susceptibility map suggest that CNN is particularly effective for applications involving spatial data. CNN excels at identifying patterns and spatial relationships, which are essential for analyzing landslide susceptibility. They are especially skilled at capturing contextual information from surrounding areas, providing significant advantages over other machine learning models that rely on pointwise evaluations for tasks like image and pattern recognition. The CNN model effectively manages complex, non-linear interactions among variables that contribute to landslides. Meanwhile, the MLP model also performs well, with a high AUC of 0.982, but it distributes risk differently. This feature could be advantageous in scenarios where minimizing overprediction of risk areas is crucial, thereby avoiding unnecessary costs from excessive mitigation measures. This nuanced understanding of each model's strengths enables more informed decision-making when selecting the appropriate tool for specific landslide risk assessment situations.

The difference in performance and risk mapping between CNN and other models like KNN and RF also highlights the importance of choosing the right algorithm based on the specific characteristics of the dataset and the objectives of the analysis. While KNN and RF show robustness and high accuracy, their simplistic or ensemble nature may not capture as detailed spatial patterns as CNN, CNN's and MLP's result shown in figure 53. Based on these results, by selecting CNN's result, the driftwood generation potential is calculated.

**Table 11.** Landslide Risk Prediction AUC for different models.

| **Model** | **DT** | **KNN** | **LR** | **RF** | **MLP** | **CNN** |
|---|---|---|---|---|---|---|
| AUC | 0.832 | 0.946 | 0.681 | 0.889 | 0.982 | 0.993 |

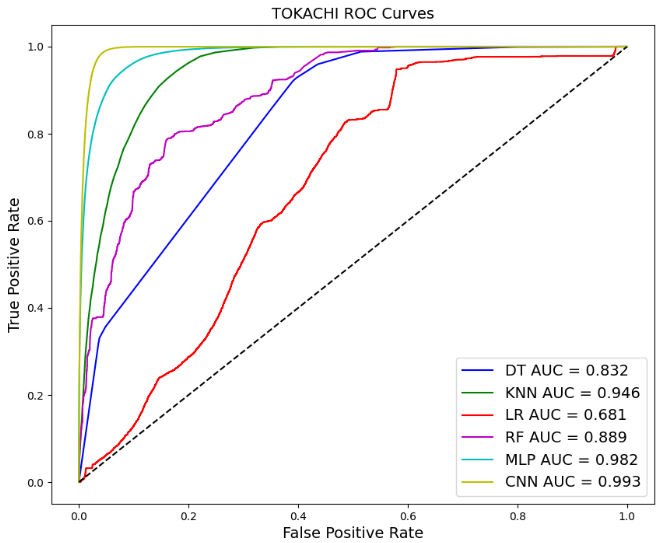
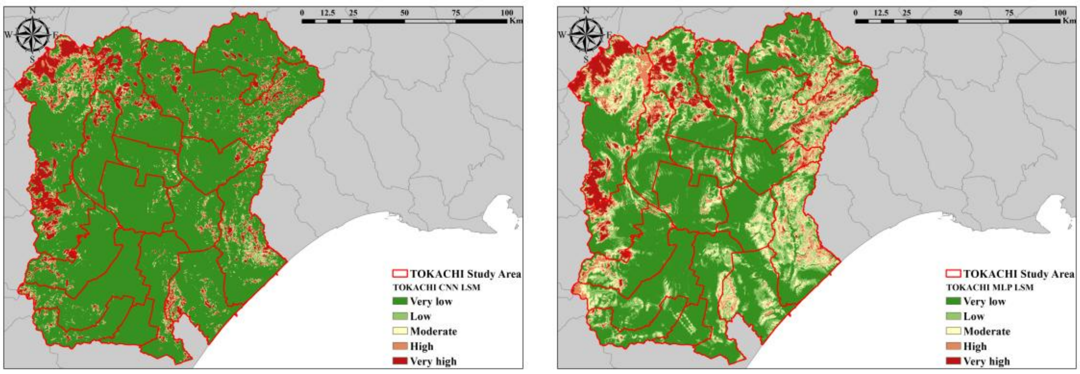



**Figure 52.** Landslide Risk Prediction AUC for different models.




(a) CNN (b) MLP

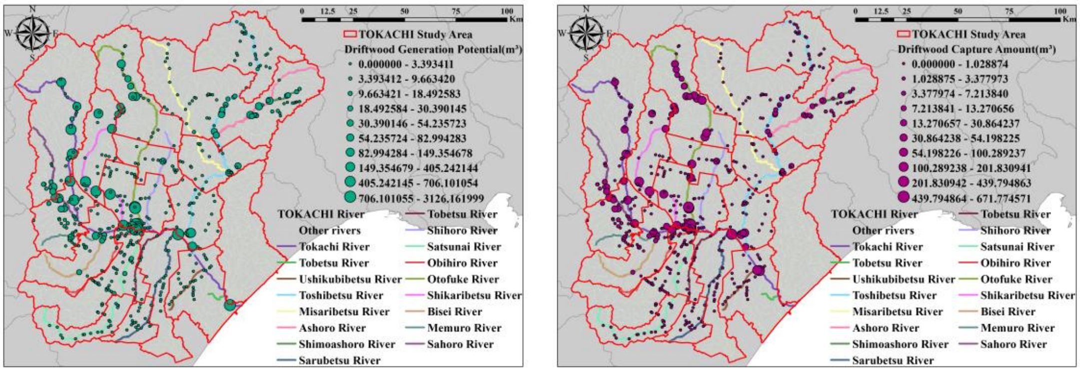
**Figure 53.** Landslide risk assessment using CNN and MLP.

4.4.3. Driftwood Generation Potential (DGP) and Driftwood Capture Amount (DCA)

In this study, we estimated the potential driftwood length at the landslide site by calculating the difference between the Digital Surface Model (DSM) and the Digital Elevation Model (DEM). While the DEM provides terrain data excluding artificial structures, the DSM includes the heights of all surface objects, such as vegetation. Therefore, tree heights can be determined by subtracting the DEM from the DSM. For this analysis, DSM data [55] was obtained from the Japan Aerospace Exploration Agency (JAXA). We calculated the driftwood length as half of the observed height difference during descent [20], resulting in an average estimate of 5.24 m in the landslide-prone area, obtained using convolutional neural networks (CNN).

Regions with very high susceptibility are concentrated in the northern and mountainous areas, defined by steep slopes and loose soil, which make them highly prone to landslides during heavy rainfall or seismic events (Figure 53). Figure 54(a) shows that approximately 17% of the areas predicted to be at high risk for landslides are identified as the area having driftwood source. In steep upstream regions, landslides dislodge trees and vegetation, significantly increasing driftwood generation, especially following intense rainfall events. In contrast, areas with low susceptibility experience minimal driftwood production due to stable terrain and a lower risk of landslides.

High driftwood capture zones are commonly found along major river channels and at confluence points, particularly in downstream areas (Figure 54(b)). These zones function as natural driftwood traps due to reduced flow velocities or shifts in channel width. The overlap between driftwood generation and capture areas indicates that driftwood from landslide-prone upstream regions is transported downstream, where it accumulates at confluences or areas with slower flow. This buildup can present significant risks to infrastructure and exacerbate flooding. Areas with high landslide susceptibility are also driftwood hotspots, as landslides uproot vegetation, introducing it into river systems as driftwood. Major river channels, confluences, and bends, where flow slows, align with high driftwood capture zones, facilitating accumulation. Effective management in the Tokachi area should focus on stabilizing slopes in landslide-prone areas and monitoring driftwood buildup in rivers. Preventive actions, such as reforestation, slope reinforcement, and installing driftwood barriers, can help mitigate the impacts of landslides and driftwood. Driftwood generated from landslides poses risks to both the environment and human infrastructure.




(a) Driftwood generation potential (DGP) (b) Driftwood capture amount (DCA)

**Figure 54.** Calculation results of driftwood generation potential (DGP) and driftwood capture amount (DCA) based on CNN.

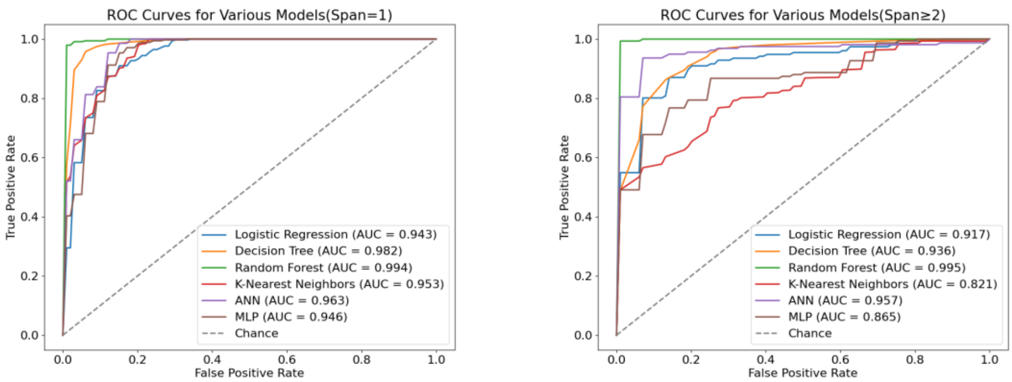
4.4.4. Bridge Disaster Risk Assessment

In the field of predictive modeling for bridge disaster prediction, a thorough analysis encompassing both single span (Span=1) as shown in figure 55(a) and Table 12-a and multi-span bridges (Span≥2) as shown in figure 55(b) and Table 12-b reveals distinct outcomes and feature influences (Figure 51) for different bridge structures. Decision Tree and Random Forest models emerge as the top performers in both categories. For Span=1 bridges, these models achieved high accuracy and strong AUC scores, demonstrating their effectiveness in simpler structural contexts. The Decision Tree offered a well-balanced trade-off between precision and recall. However, the Random Forest model's near-perfect scores across both span categories, particularly for multi-span bridges, suggest a potential overfitting issue due to its complexity and the smaller dataset size for Span≥2, raising concerns about its generalizability.

Additionally, other models such as Logistic Regression, K-Nearest Neighbors (KNN), Artificial Neural Network (ANN), and Multi-Layer Perceptron (MLP) exhibited varying results. Logistic Regression struggled with precision in both datasets, achieving moderate accuracy for Span=1 but a significant drop for Span≥2, despite a high recall, indicating challenges with class discrimination. KNN consistently encountered issues with precision, while ANN and MLP performed reasonably well for Span=1 but saw declines in Span≥2. These results highlight the need for model adjustments or improvements when working with more complex datasets, where simpler models or those with built-in regularization may provide more consistent performance.

The significance of various predictive features, as shown in figure 51, further highlights the varying impacts of environmental and structural variables on bridge safety. For Span=1 bridges, "Driftwood Capture Amount ($m^3$)" stands out as a crucial factor, far surpassing other features, emphasizing the importance of driftwood deposition in influencing the stability of single-span bridges. In contrast, for Span≥2 models, both the " Impediment ratio of river flow " and "Driftwood Capture Amount ($m^3$)" play a significant role, indicating that debris accumulation is a major risk for larger, more complex bridge structures. This variation in feature importance suggests that while factors such as "Curvature," "Intersection Angle," and "Surface Geology" maintain moderate relevance across both bridge types, the specific dynamics of bridge span and structure require customized modeling approaches.

This thorough analysis highlights the importance of considering bridge type, as well as relevant structural and environmental features, when developing predictive models for bridge disasters. Tailoring model strategies to the specific conditions of each bridge type and prioritizing the most influential features can help reduce risks such as overfitting in more complex models, ultimately enhancing the accuracy and reliability of predictions related to bridge safety.




(a) Span=1 (b) Span≥2

**Figure 55.** Comparison of AUC for Bridge Disaster Risk Prediction in training groups using different models based on various bridge structures.

**Table 12-a.** Bridge disaster risk assessment results for a Span=1 bridge, obtained from each model.

| | **Model** | **Confusion matrix** | | | | **ACC** | **P-score** | **R-score** | **F-score** | **AUC** |
|---|---|---|---|---|---|---|---|---|---|---|
| | | **TP** | **FP** | **TN** | **FN** | | | | | |
| Span=1 | Logistic Regression | 6 | 41 | 303 | 1 | 0.880 | 0.128 | 0.857 | 0.222 | 0.940 |
| | Decision Tree | 5 | 2 | 342 | 2 | 0.989 | 0.714 | 0.714 | 0.714 | 0.990 |
| | Random Forest | 5 | 11 | 333 | 2 | 0.963 | 0.313 | 0.714 | 0.435 | 0.986 |
| | K-Nearest Neighbors | 7 | 47 | 297 | 0 | 0.866 | 0.130 | 1.000 | 0.230 | 0.952 |
| | ANN | 6 | 20 | 324 | 1 | 0.940 | 0.231 | 0.857 | 0.364 | 0.968 |
| | MLP | 5 | 24 | 320 | 2 | 0.926 | 0.172 | 0.714 | 0.278 | 0.934 |

**Table 12-b.** Bridge disaster risk assessment results for a Span≥2 bridge, obtained from each model.

| | Model | Confusion matrix TP | FP | TN | FN | ACC | P-score | R-score | F-score | AUC |
|---|---|---|---|---|---|---|---|---|---|---|
| Span≥2 | Logistic Regression | 8 | 32 | 123 | 1 | 0.799 | 0.200 | 0.889 | 0.327 | 0.891 |
| | Decision Tree | 5 | 6 | 149 | 4 | 0.939 | 0.455 | 0.556 | 0.500 | 0.849 |
| | Random Forest | 7 | 1 | 154 | 2 | 0.982 | 0.875 | 0.778 | 0.824 | 0.994 |
| | K-Nearest Neighbors | 8 | 66 | 89 | 1 | 0.591 | 0.108 | 0.889 | 0.193 | 0.746 |
| | ANN | 8 | 6 | 149 | 1 | 0.957 | 0.571 | 0.889 | 0.696 | 0.984 |
| | MLP | 8 | 32 | 123 | 1 | 0.799 | 0.200 | 0.889 | 0.327 | 0.885 |

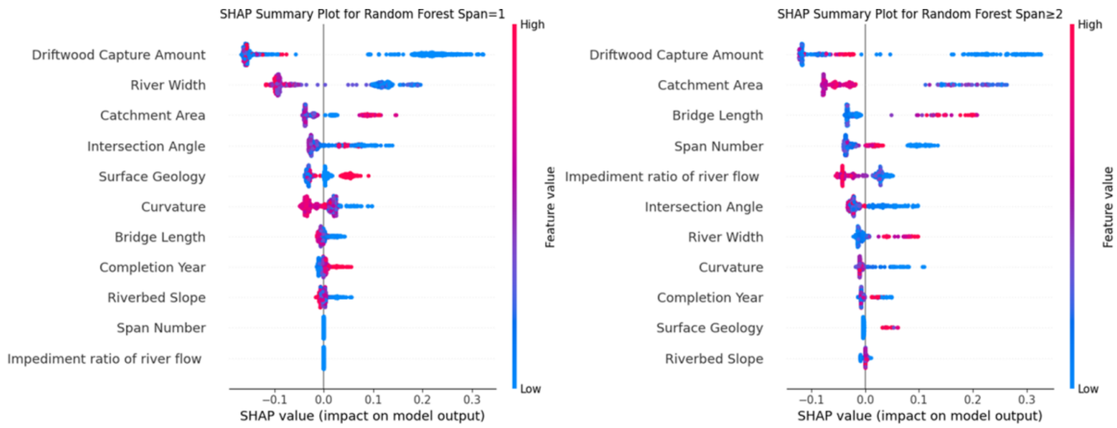



**Figure 56.** SHAP Value results for bridge disaster risk based on Randon Forest.

The analysis of the driftwood capture amount and bridge disaster data reveals a significant spatial pattern. Areas with higher driftwood capture amounts, particularly in downstream sections of major rivers such as the Tokachi River, exhibit a higher concentration of disaster-affected bridges. This observation aligns with the hypothesis that higher driftwood capture amounts increase the risk of bridge disasters. However, the SHAP analysis shows an inverse relationship, with higher driftwood capture amounts being associated with non-disaster predictions. This contradiction likely stems from the imbalanced dataset, where non-disaster bridges dominate, causing the model to misinterpret the true relationship.

The SHAP analysis (shown in Figure 56) indicates that for Span = 1 bridges, driftwood capture amount is the most significant feature influencing disaster predictions, reflecting the vulnerability of smaller-span bridges to direct driftwood impacts. For Span ≥ 2 bridges, the risk factors are more complex, with both driftwood capture amount and impediment ratio of river flow playing significant roles. The interaction between these features suggests that larger-span bridges are influenced not only by direct driftwood impacts but also by water flow obstruction, which can amplify the risk.

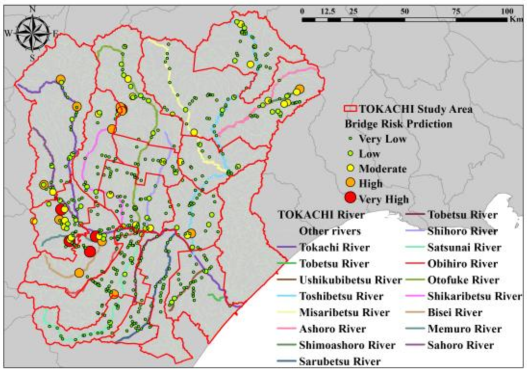
Spatially, high driftwood capture amounts are concentrated in downstream river sections, which overlap with regions of higher disaster occurrence. However, the disaster risk is not solely determined by driftwood capture amount. Other factors, such as river width, flow obstruction, and bridge design, interacting with driftwood capture amounts to influence risk. For instance, upstream regions with lower driftwood capture amounts may experience high disaster risks due to concentrated water flow, while wide downstream sections may have mitigating effects but remain vulnerable due to large debris volumes.

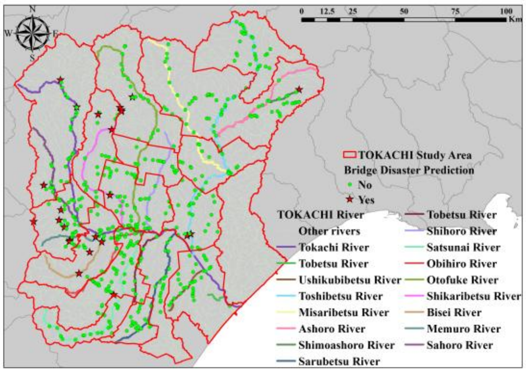
In conclusion, the analysis reveals that while higher driftwood capture amounts are strongly associated with bridge disaster risk, the model has not fully captured this relationship, likely due to data imbalance and feature interactions. The interplay between driftwood capture amount and other factors, such as flow obstruction and bridge span, highlights the complexity of disaster mechanisms. Refinements to the model and adjustments to the dataset are necessary to more accurately reflect these relationships and improve disaster risk predictions.

Figure 57 illustrates bridge risk and disaster predictions, categorizing bridges into five risk levels, ranging from very low to very high. The highest-risk bridges are predominantly located in the southern and southwestern regions of the study area. A clear correlation is observed between the driftwood capture amounts shown in figure 54(b) and the bridge risk predictions in figure 57. Specifically, areas with significant driftwood accumulation align with regions where bridges are assessed as having a higher risk of disaster. This connection is likely due to the impact of driftwood on bridge structures during flood events. Large quantities of driftwood can obstruct water flow, increase hydraulic pressure on bridge piers, and cause direct damage to bridge components, all of which heighten the risk of failure or severe damage, particularly in areas prone to heavy rainfall and flooding.

The overlap between driftwood-heavy zones and high-risk bridges underscores the need for targeted interventions to mitigate risk and prevent disasters. Strategies such as reinforcing bridge structures, ensuring unobstructed river flow, and establishing proactive inspection programs are essential to reducing the chances of bridge damage and failure. Additionally, installing driftwood capture devices upstream of bridges could alleviate pressure on vulnerable structures by decreasing the volume of driftwood reaching critical areas.

In conclusion, the visual data in these maps highlight the relationship between natural driftwood dynamics and bridge vulnerability. The findings suggest that managing driftwood accumulation and addressing structural risks can significantly enhance regional resilience against flood-related disasters. By understanding the links between driftwood capture, bridge risk levels, and disaster predictions, local authorities can develop targeted, prioritized plans to improve bridge safety and mitigate potential hazards in the Tokachi area.




(a) Bridge disaster prediction results (b) Distribution map of high disaster risk bridges

**Figure 57.** Prediction results for bridge disaster risk.

### 4.5. Conclusion

In this study, we have demonstrated the critical application of advanced machine learning techniques, particularly Convolutional Neural Networks (CNN), in assessing and predicting the risks associated with bridge infrastructure in the Tokachi area of Hokkaido, Japan. Given the increasing prevalence of extreme weather events and their resultant impacts such as landslides and driftwood accumulation, our research highlights the imperative need for sophisticated risk assessment models that can accurately forecast infrastructure vulnerabilities and guide preemptive measures.

Our research applied a comprehensive approach that integrated environmental and structural data with convolutional neural networks (CNN) and conventional machine learning techniques to predict and evaluate disaster risks for bridges. The study differentiated between single-span and multi-span bridges, finding that while "Driftwood Capture Amount ($m^3$)" was a key predictor for single-span bridge risk, both " Impediment ratio of river flow " and "Driftwood Capture Amount ($m^3$)" were critical for multi-span bridges. This distinction is essential for accurate risk assessment and supports more informed decision-making. Additionally, the study quantified the impact of factors like slope inclination, vegetation types, and river dynamics on landslides and driftwood accumulation, enabling detailed risk profiles for bridges that guide targeted intervention and maintenance strategies.

The findings from our study underscore the superior performance of CNN in spatial data analysis and pattern recognition, which are crucial for detecting areas with high susceptibility to landslides and driftwood blockages. These capabilities allow for a more nuanced understanding of risk distributions, facilitating targeted interventions that are both cost-effective and timely. Moreover, the application of these models has proven to be instrumental in identifying potential hotspots for disaster, thereby enhancing the strategic planning of infrastructure resilience measures.

Furthermore, the integration of machine learning into risk assessment processes not only improves the accuracy of predictions but also provides a dynamic tool for infrastructure management under changing climatic conditions. This adaptability is particularly relevant in the context of global climate change, which is expected to intensify the frequency and severity of extreme weather events.

In conclusion, our research contributes to the ongoing efforts in disaster risk reduction by providing a methodological framework that technology to enhance the bridge against environmental threats. The insights gained from this study advocate for the broader adoption of machine learning techniques in civil engineering practices, particularly for regions prone to natural disasters. Future studies could expand upon this work by incorporating real-time data analytics and exploring the potential of hybrid models that combine the strengths of multiple machine learning approaches to further refining risk assessment and management strategies.

### 4.6. Future works

#### Future Considerations for Driftwood Sources

Driftwood accumulation poses a significant risk to bridge infrastructure, particularly during extreme weather events. In this study, the maximum driftwood volume is modeled based on landslides as the primary source. However, under extreme weather conditions, driftwood can also originate from riparian forests, where strong winds, heavy rainfall, or flooding dislodge trees and vegetation, contributing to driftwood transport. Future research should focus on differentiating between landslide-induced and riparian forest-derived driftwood, ensuring a more precise risk assessment.

To enhance risk prediction, integrating multi-source remote sensing data (e.g., LiDAR, satellite imagery) can improve driftwood source identification and tracking. Hydrodynamic modeling should also be considered to simulate driftwood transport mechanisms under various weather conditions, helping to understand how driftwood accumulates at bridge piers or in river channels. Additionally, machine learning models can be refined to predict driftwood distribution patterns by incorporating variables such as river discharge, vegetation cover, and terrain conditions.

**Enhancing Bridge Structural Data for Risk Assessment**

A comprehensive assessment of bridge vulnerability requires detailed structural data. Currently, risk evaluation primarily relies on environmental factors such as slope, vegetation, and hydrology. However, future research should focus on integrating high-resolution structural data, including bridge material properties, pier height, span length, deck design, and foundation type, to improve the accuracy of risk predictions.

Advanced techniques such as structural health monitoring (SHM) using IoT-based sensors and UAV inspections can provide real-time insights into bridge stability under extreme weather conditions. Additionally, finite element modeling (FEM) can be applied to simulate the impact of landslide debris and driftwood collisions on bridge structures. Deep learning models can also be trained using a combination of structural, environmental, and hydrodynamic data to improve risk assessment accuracy for different bridge designs.

By distinguishing driftwood origins, improving transport modeling, and integrating detailed bridge structural data, future studies can develop more robust risk assessment frameworks for infrastructure protection, ensuring effective mitigation strategies and improved bridge resilience against extreme weather events.

# 5. Summary and Conclusion

The study provides a robust exploration of the interplay between landslide risks, driftwood generation, and bridge vulnerabilities under extreme weather events, with a focus on integrating advanced machine learning techniques, particularly Convolutional Neural Networks (CNN). This research builds upon conventional machine learning methods, aiming to refine predictive accuracy and resilience strategies for infrastructure and environmental challenges.

Landslides are a critical natural hazard influenced by geological, topographical, and climatic factors. The integration of Geographic Information Systems (GIS) with machine learning models has significantly enhanced landslide susceptibility mapping. The research systematically employs various machine learning models, including Logistic Regression (LR), Decision Trees (DT), Random Forests (RF), K-Nearest Neighbors (KNN), Artificial Neural Networks (ANN), and CNN. Each model is evaluated on its capacity to predict landslide-prone areas across diverse regions in Japan, namely Niigata Prefecture (NIG), Iwate and Miyagi Prefectures (IWT-MYG), and Hokkaido (HKD).

Traditional models like LR and RF demonstrated limitations in capturing the complex, nonlinear relationships inherent in landslide data, as reflected in their lower Area Under the Curve (AUC) scores. In contrast, ANN and CNN models exhibited superior performance, with CNN achieving the highest AUC values across all regions. CNN's capability to process spatial data, combined with its convolutional architecture, allows for the extraction of nuanced patterns and contextual information from satellite-derived images. The study highlights that varying input window sizes, particularly 15×15 to 23×23 pixels, significantly enhance model performance, balancing feature representation and noise reduction. This methodological refinement provides a blueprint for adapting CNN-based models to analyze other geohazards.

Driftwood, a secondary hazard of landslides, poses significant risks to riverine and bridge infrastructures. The study innovatively integrates landslide susceptibility maps generated by CNN models with additional environmental data to estimate driftwood generation potential (DGP). Areas with steep slopes exceeding 30 degrees and dense vegetation were identified as primary driftwood sources. Using Digital Surface Models (DSM) and Digital Elevation Models (DEM), the study calculates the volume of driftwood produced during landslide events, with results indicating a strong correlation between high-risk landslide zones and regions of substantial driftwood output.

This integration provides a comprehensive understanding of how natural processes like landslides cascade into other hazards. By focusing on upstream areas with steep inclines and vulnerable vegetation, the study establishes a targeted approach for managing driftwood risks, which are exacerbated during heavy rainfall or seismic activity.

Building upon the landslide and driftwood analyses, the research shifts its focus to evaluating bridge vulnerabilities in the Tokachi River basin of Hokkaido, Japan. Here, the study assesses the risks posed by driftwood accumulation on single-span (Span=1) and multi-span (Span≥2) bridges. By combining CNN-based driftwood predictions with structural and hydrological data, the research identifies critical factors influencing bridge safety.

For single-span bridges, the "Driftwood Capture Amount ($m^3$)" emerges as the most significant predictor of disaster risk. Meanwhile, multi-span bridges are primarily influenced by a combination of " Impediment ratio of river flow " and "Driftwood Capture Amount ($m^3$)," reflecting the complexity of these structures in managing debris flow. The study’s results highlight spatial patterns where high driftwood accumulation correlates with increased bridge risk, providing actionable insights for regional disaster management.

This study reveals critical insights into the dynamics of natural hazards and their interactions with infrastructure vulnerabilities. The superior performance of CNN models in predicting both landslide and driftwood risks underscores their potential as primary tools for spatial data analysis. These models excel in capturing contextual information from diverse environmental features, thereby offering higher predictive accuracy compared to traditional approaches. Moreover, optimal input window sizes for CNNs, identified between 15×15 and 23×23 pixels, ensure that the model effectively balances detailed feature capture and computational efficiency. The findings also highlight regional differences in the importance of risk factors such as slope and vegetation, emphasizing the need for tailored strategies to address diverse geographical and climatic conditions. Lastly, driftwood accumulation is identified as a significant risk factor for bridge safety, necessitating proactive measures to mitigate associated hazards.

Future studies should explore hybrid modeling approaches that combine CNNs with ensemble methods like Random Forests or Gradient Boosting Machines to further enhance predictive robustness. Additionally, the development of adaptive sampling strategies that dynamically adjust based on real-time model performance would improve class balance and accuracy. Extending the current methodological framework to other natural hazards, such as floods or wildfires could also broaden its applicability, enhancing disaster resilience. Finally, incorporating real-time hydrological data and field surveys would refine predictions of driftwood dynamics, offering more precise risk assessments and mitigation strategies for critical infrastructure.

This study underscores the transformative potential of integrating advanced machine learning techniques with GIS and environmental data for comprehensive hazard risk assessments. By leveraging CNN models, the research achieves unprecedented accuracy in predicting landslide-prone areas, estimating driftwood generation, and assessing bridge vulnerabilities. The findings provide critical insights for disaster mitigation, infrastructure planning, and environmental management, emphasizing the importance of interdisciplinary approaches to address the challenges posed by natural hazards. This framework serves as a foundation for future innovations aimed at safeguarding communities and ecosystems in the face of escalating climate-induced risks.

## Acknowledgements

First and foremost, I would like to express my deepest gratitude to my advisor, Associate Professor Koichi Isobe, for his invaluable guidance, encouragement, and unwavering support throughout my doctoral studies. His expertise and dedication have been instrumental in shaping this research and my academic growth.

This study was partially funded by the Japan Science and Technology Agency (JST) SPRING Grant Number JPMJSP2119, and the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Numbers 23K04019 and 24K00985. I am also sincerely grateful to Sato Kogyo Co., Ltd. for their financial support.

I would like to acknowledge the Hokkaido Regional Development Bureau of the Ministry of Land, Infrastructure, Transport and Tourism, Hokkaido Prefecture, and the Ministry of Land, Infrastructure, Transport and Tourism's database for providing bridge inspection reports, general maps, bridge databases for national and prefectural highway bridges, and various terrain information, which were critical to the success of this research. I also wish to thank the Japan Meteorological Agency database for supplying precipitation data, which significantly contributed to the analysis and findings of this study.

This work also benefited greatly from the contributions of several graduates, including Wu Zhenyu, Ota Ryohei, Watanabe Sakurako, Taniguchi Naoki, Orito Kazuki, Kusumoto Yumi, Ohira Natsuo, Tateyama Tomoya, and Li Diyuan. I am profoundly grateful for their insights and collaboration.

Finally, I am deeply thankful to my family and friends for their unwavering encouragement and support, which have been a source of strength throughout this journey. This achievement would not have been possible without their love and understanding.

Thank you all for making this endeavor meaningful and rewarding.

## References


1. Ministry of Land, Infrastructure, Transport and Tourism, Basic Considerations for the Driftwood Management Plan During Sediment and Flood Disasters (Trial Version) [Draft]
2. Chen, X., Application of Artificial Intelligence to Assessment of Earthquake-induced Landslide Susceptibility, in Chinese Doctoral Dissertations Full-text Database. 2007.
3. Lee, S., S.-M. Hong, and H.-S. Jung, A support vector machine for landslide susceptibility mapping in Gangwon Province, Korea. Sustainability, 2017. **9**(1): p. 48.
4. Lin, Q., Y. Liu, L. Liu, and Y. Wang, Earthquake-triggered Landslide Susceptibility Assessment Based on Support Vector Machine Combined with Newmark Displacement Model. Journal of Geo-Information Science, 2017. **12**.
5. Mondal, S. and S. Mandal, RS & GIS-based landslide susceptibility mapping of the Balason River basin, Darjeeling Himalaya, using logistic regression (LR) model. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, 2018. **12**(1): p. 29-44.
6. Kadavi, P.R., C.-W. Lee, and S. Lee, Landslide-susceptibility mapping in Gangwon-do, South Korea, using logistic regression and decision tree models. Environmental Earth Sciences, 2019. **78**: p. 1-17.
7. Sameen, M.I., B. Pradhan, D.T. Bui, and A.M. Alamri, Systematic sample subdividing strategy for training landslide susceptibility models. Catena, 2020. **187**: p. 104358.
8. Adnan, M.S.G., M.S. Rahman, N. Ahmed, B. Ahmed, M.F. Rabbi, and R.M. Rahman, Improving spatial agreement in machine learning-based landslide susceptibility mapping. Remote Sensing, 2020. **12**(20): p. 3347.
9. Kawabata, D. and J. Bandibas, Landslide susceptibility mapping using geological data, a DEM from ASTER images and an Artificial Neural Network (ANN). Geomorphology, 2009. **113**(1-2): p. 97-109.
10. Wei, R., C. Ye, T. Sui, H. Zhang, Y. Ge, and Y. Li, A feature enhancement framework for landslide detection. International Journal of Applied Earth Observation and Geoinformation, 2023. **124**: p. 103521.
11. Wei, R., C. Ye, T. Sui, Y. Ge, Y. Li, and J. Li, Combining spatial response features and machine learning classifiers for landslide susceptibility mapping. International Journal of Applied Earth Observation and Geoinformation, 2022. **107**: p. 102681.
12. He, Y., Z.a. Zhao, W. Yang, H. Yan, W. Wang, S. Yao, L. Zhang, and T. Liu, A unified network of information considering superimposed landslide factors

sequence and pixel spatial neighbourhood for landslide susceptibility mapping. International Journal of Applied Earth Observation and Geoinformation, 2021. **104**: p. 102508.

13. Jiang, W., J. Xi, Z. Li, M. Zang, B. Chen, C. Zhang, Z. Liu, S. Gao, and W. Zhu, Deep Learning for Landslide Detection and Segmentation in High-Resolution Optical Images along the Sichuan-Tibet Transportation Corridor. Remote Sensing, 2022. **14**(21): p. 5490.
14. Jiang, Z., M. Wang, and K. Liu, Comparisons of Convolutional Neural Network and Other Machine Learning Methods in Landslide Susceptibility Assessment: A Case Study in Pingwu. Remote Sensing, 2023. **15**(3): p. 798.
15. Hussain, M.A., Z. Chen, Y. Zheng, Y. Zhou, and H. Daud, Deep Learning and Machine Learning Models for Landslide Susceptibility Mapping with Remote Sensing Data. Remote Sensing, 2023. **15**(19): p. 4703.
16. Ghorbanzadeh, O., T. Blaschke, K. Gholamnia, S.R. Meena, D. Tiede, and J. Aryal, Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. Remote Sensing, 2019. **11**(2): p. 196.
17. Liu, R., X. Yang, C. Xu, L. Wei, and X. Zeng, Comparative Study of Convolutional Neural Network and Conventional Machine Learning Methods for Landslide Susceptibility Mapping. Remote Sensing, 2022. **14**(2): p. 321.
18. Minami, N., Y. Doi, N. Osanai, S. Takesaki, Y. Nakayama, K. Nonaka, and S. Mashiko, Report of woody debris disaster at Yosasa River on August 1998 in Tochigi prefecture, Japan. Sabo Gakkaishi, 2000. **53**(4): p. 44-51.
19. Yamada, K., Y. Nagasaka, H. Sato, T. Tsushima, and T. Abe, Quantitative study of the area of damage and the amount of lost or trapped timber in a riparian forest in the Appetsu River basin caused by Typhoon 0310 (Etau). Sabo Gakkaishi, 2006. **59**(1): p. 13-20.
20. Yano, S., S. Tsuchihashi, S. Dozono, K. Kazama, and T. Kita, Assessment of woody debris disaster risk at all bridges along the Kagetsu River according to its generation potential concept. Journal of Japan Society of Civil Engineers, Ser. B1 (Hydraulic Engineering), 2016. **72**(4): p. I_289-I_294.
21. Dozono, S., Development of a Risk Assessment Method for Driftwood Disaster at Bridges Based on the Potential for Driftwood Generation in River Basins. 2017, Kyushu University.
22. Bangnira, F., B. Marti-Cardona, B. Imam, and V. Ruiz-Villanueva, Identifying bridges prone to instream wood accumulation: insights from bridges across the UK. Natural Hazards, 2024. **120**(1): p. 25-40.

23. Yano, S., R. Okubo, A. Tsusue, D. Takemura, K. Tomita, K. Kasama, and Y. Nihei, Analysis on cause of driftwoods generation by 2017 northern Kyushu heavy rain. Journal of Japan Society of Civil Engineers, Ser. B1 (Hydraulic Engineering), 2018. **74**(5): p. I_1063-I_1068.
24. Kakihara, N., Y. Minowa, and K. Nagashima, Model construction for estimating risk of driftwood by topographical factors. The Japanese Forest Society Congress, 2022. **133**: p. 525.
25. Tanaka, Y. and K. Nagashima, Woody debris risk assessment based on terrain analysis of slope failure. The Japanese Forest Society Congress, 2021. **132**: p. 548.
26. National Research Institute for Earth Science and Disaster Resilience,Japan, Landslide topography GIS data
27. Liu, R., N. Sun, and G. Tang, Analysis of Geological Environment and Causes of Landslides in Guangdong. Tropical Geography, 2010. **1**: p. 13-17.
28. Lee, S. and K. Min, Statistical analysis of landslide susceptibility at Yongin, Korea. Environmental geology, 2001. **40**(9): p. 1095-1113.
29. Glade, T., Landslide occurrence as a response to land use change: a review of evidence from New Zealand. Catena, 2003. **51**(3-4): p. 297-314.
30. Ray, R.L. and J.M. Jacobs, Relationships among remotely sensed soil moisture, precipitation and landslide events. Natural Hazards, 2007. **43**: p. 211-222.
31. Sun, L., B. Ma, L. Pei, X. Zhang, and J.L. Zhou, The relationship of human activities and rainfall-induced landslide and debris flow hazards in Central China. Natural Hazards, 2021. **107**: p. 147-169.
32. Larsen, I.J. and D.R. Montgomery, Landslide erosion coupled to tectonics and river incision. Nature Geoscience, 2012. **5**(7): p. 468-473.
33. Tibaldi, A., L. Ferrari, and G. Pasquarè, Landslides triggered by earthquakes and their relations with faults and mountain slope geometry: an example from Ecuador. Geomorphology, 1995. **11**(3): p. 215-226.
34. Gonzalez-Ollauri, A. and S.B. Mickovski, Hydrological effect of vegetation against rainfall-induced landslides. Journal of Hydrology, 2017. **549**: p. 374-387.
35. Catani, F., N. Casagli, L. Ermini, G. Righini, and G. Menduni, Landslide hazard and risk mapping at catchment scale in the Arno River basin. Landslides, 2005. **2**: p. 329-342.
36. Yokoyama, R., M. Sirasawa, and Y. Kikuchi, Representation of topographical features by opennesses. Journal of the Japan society of photogrammetry and remote sensing, 1999. **38**(4): p. 26-34.

37. Shirzadi, A., K. Solaimani, M.H. Roshan, A. Kavian, K. Chapi, H. Shahabi, S. Keesstra, B.B. Ahmad, and D.T. Bui, Uncertainties of prediction accuracy in shallow landslide modeling: Sample size and raster resolution. Catena, 2019. **178**: p. 172-188.
38. Wang, Y., Z. Fang, and H. Hong, Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County, China. Science of the total environment, 2019. **666**: p. 975-993.
39. Ding, A., Q. Zhang, X. Zhou, and B. Dai. Automatic recognition of landslide based on CNN and texture change detection. in 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC). 2016. IEEE.
40. Zhang, C., I. Sargent, X. Pan, H. Li, A. Gardiner, J. Hare, and P.M. Atkinson, An object-based convolutional neural network (OCNN) for urban land use classification. Remote sensing of environment, 2018. **216**: p. 57-70.
41. Dixit, M., R. Kwitt, M. Niethammer, and N. Vasconcelos. Aga: Attribute-guided augmentation. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
42. Liu, B., X. Wang, M. Dixit, R. Kwitt, and N. Vasconcelos. Feature Space Transfer for Data Augmentation. in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018.
43. Ghorbanzadeh, O., S.R. Meena, H.S.S. Abadi, S.T. Piralilou, L. Zhiyong, and T. Blaschke, Landslide mapping using two main deep-learning convolution neural network streams combined by the Dempster–Shafer model. IEEE Journal of selected topics in applied earth observations and remote sensing, 2020. **14**: p. 452-463.
44. Matthews, B.W., Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochimica et Biophysica Acta (BBA)-Protein Structure, 1975. **405**(2): p. 442-451.
45. Ministry of Land, Infrastructure, Transport and Tourism, Tokachi River basin sediment disaster countermeasure examination: Considering the characteristics of the Tokachi River basin
46. Ministry of Land, Infrastructure, Transport and Tourism, Download site for digital land information. *https://nlftp.mlit.go.jp/ksj/.*
47. Geospatial Information Authority of Japan, 10m Digital Elevation Model. *https://fgd.gsi.go.jp/download/mapGis.php?tab=dem.*
48. Ministry of the Environment,Japan, Natural Environment Survey Web-GIS. *http://gis.biodic.go.jp/webgis/index.html.*
49. Hokkaido Regional Development Bureau,Japan, Hokkaido Bridge and Road Large Structure Data Storage and Management System.

https://www.hkd.mlit.go.jp/ky/category/u23dsn00000001h3.html.
50. Cabinet Office, Japan,, Damage caused by Typhoon No. 10 of 2016. *https://www.bousai.go.jp/updates/h28typhoon10/pdf/h28typhoon10_22.pdf.*
51. Japanese Geotechnical Society, Report on the investigation into geological disasters caused by the heavy rains in Hokkaido in August 2016
52. Hokkaido Shimizu Town: Hokkaido, Japan, Report on the response to the heavy rain disaster caused by Typhoon No. 10 in 2016. *https://www.town.shimizu.hokkaido.jp/disaster/files/taihuu10gouooamesaigaihoukokusyo.pdf.*
53. Forestry Agency,Japan, Guidelines for preventing debris flows and driftwood.
54. Moriwaki, H., A Prediction of the Runout Distance of a Debris. Landslides, 1987. **24**(2): p. 10-16_1.
55. Japan Aerospace Exploration Agency, ALOS Global Digital Surface Model (DSM) "ALOS World 3D - 30m (AW3D30)". *https://www.eorc.jaxa.jp/ALOS/jp/dataset/aw3d30/aw3d30_j.htm.*