Climate Services 33 (2024) 100440

Contents lists available at ScienceDirect

Climate Services

journal homepage: www.elsevier.com/locate/cliser

Original research article

Projection and identification of vulnerable areas due to heavy snowfall using machine learning and K-means clustering with RCP scenarios

Moon-Soo Song [a], Jae-Joon Lee [b], Hong-Sic Yun [c], Sang-Guk Yum [d,*]

[a] Department of Safety & Disaster Prevention Engineering, KyungWoon University, Gumi-si, Gyeongsangbuk-do 39160, Republic of Korea
[b] Department of Fire Safety Engineering, Jeonju University, Jeonju-si, Jeollabuk-do Republic of Korea
[c] School of Civil, Architectural Engineering & Landscape Architecture, Sungkyunkwan University, Suwon 16419, Republic of Korea
[d] Department of Civil Engineering, College of Engineering, Gangneung-Wonju National University, Gangneung 25457, Republic of Korea

ARTICLE INFO

Keywords:
Snowfall projection
Machine learning
GIS
RCP
Image clustering

ABSTRACT

Heavy snowfall is a natural disaster that causes extensive damage in South Korea. Therefore, predicting heavy snowfall occurrence, identifying vulnerable areas, and establishing response plans to reduce risk are crucial. In this study, to project heavy snowfall, meteorological and geographic data from the past 30 years were collected, and four machine learning algorithms were trained and compared: multiple linear regression, support vector regression, random forest regressor (RFR), and extreme gradient boosting. We observed that the RFR model (R^2 = 0.64) demonstrated the most optimal performance in projecting snowfall compared to other models. Representative concentration pathway (RCP) scenario data was input into the RFR model to generate projection data up to 2100. Projection results of more than 48.2 cm based on heavy snowfall events in the past 20 years were observed 17 times in RCP2.6, 19 times in RCP4.5, 16 times in RCP6.0, and 17 times in RCP8.5. The annual GIS-based projected snowfall images for the RCP8.5 scenario were classified into five distinct groups using K-means clustering. These groups were then further divided based on the vulnerability of regions, including Gangwon-do, Jeollabuk-do, and northern Gyeonggi-do. Our study can aid decision-making on policies related to heavy snowfall disaster prevention standards, snow removal plans, budgeting, and the establishment of mid- to long-term climate change adaptation plans for government, public institutions and private organizations.

Practical implications

Considering the growing body of evidence pointing to climate change-induced shifts in climate conditions and the increased incidence of extreme weather events, our research narrows its focus to a critical facet of this challenge: heavy snowfall in South Korea. This meteorological phenomenon carries profound implications for urban and rural societies, with particular regard to the disruption of transportation networks and urban infrastructure. Effective responses to heavy snowfall necessitate the ability to project short-term, medium-term, and long-term snowfall patterns while simultaneously developing robust adaptation strategies.

Drawing upon data from Representative Concentration Pathway (RCP) scenarios, our study employs advanced machine learning techniques to project snowfall trends in South Korea. The results of our analysis reveal a troubling prognosis: both the frequency of heavy snowfall and snowfall depth are expected to increase across most regions of the country. This underscores the pressing need for proactive adaptation measures that can help mitigate urban risks and safeguard communities. Distinguishing our work from prior research endeavors that have primarily concentrated on quantitative aspects of future snowfall, this paper aims to translate

Abbreviations: ANN, Artificial neural network; ASOS, Automated synoptic observing system; BART, Bayesian Additive Regression Trees; BANN, Bayesian Artificial Neural Network; R^2, Coefficient of determination; DT, Decision tree; XGB, Extreme gradient boosting; GBM, Gradient boosting machine; IPCC, Intergovernmental Panel on Climate Change; KMA, Korea Meteorological Administration; MAE, Mean absolute error; MOIS, Ministry of the Interior and Safety; MLR, Multiple linear regression; RF, Random forest; RFR, Random forest regressor; RCP, Representative concentration pathway; RMSE, Root mean square error; SWE, Snow Water Equivalent; SVM, Support vector machine; SVR, Support vector regression; TOL, Tolerance; TPI, Topographic Position Index; TWI, Topographic Wetness Index; VIF, Variance inflation factor.

* Corresponding author.
E-mail addresses: songms0722@ikw.ac.kr (M.-S. Song), safety_lee@jj.ac.kr (J.-J. Lee), yoonhs@skku.edu (H.-S. Yun), skyeom0401@gwnu.ac.kr (S.-G. Yum).

https://doi.org/10.1016/j.cliser.2023.100440
Received 27 November 2022; Received in revised form 20 September 2023; Accepted 20 December 2023
Available online 25 December 2023

intricate climate projection data into readily understandable visual representations through image clustering. This novel approach promises to facilitate comprehension and decision-making among policymakers and stakeholders alike. To equip policy decision-makers with the insights necessary for formulating response plans that minimize the risks associated with heavy snowfall, it is essential to gain a comprehensive understanding of evolving snowfall patterns. This entails identifying years marked by heightened snowfall activity and pinpointing areas that are particularly vulnerable to these meteorological shifts.

We are acutely aware of the inherent uncertainties associated with climate projection models and RCP scenarios, which may introduce subjectivity into decision-making processes. Nevertheless, in the face of mounting extreme climate events, the potential utility of snowfall forecast maps as communication tools within climate change adaptation policies cannot be understated. Moreover, the incorporation of detailed spatial indicators, such as population density, building distribution, and regional adaptive capacity, enables a more robust assessment of vulnerability, utilizing the snowfall forecast data as a central component in climate exposure analysis in subsequent studies. In summary, this research endeavors to shed light on the repercussions of climate change-driven heavy snowfall in South Korea and the imperative need for pragmatic adaptation strategies. By making climate data more accessible and actionable, we aim to empower decision-makers to safeguard urban communities and infrastructure against the challenges posed by heavy snowfall events in our changing climate.

Data availability

Data will be made available on request.

Introduction

The sixth assessment report of the Intergovernmental Panel on Climate Change (IPCC) stated that the abnormal climate observed worldwide is due to the rapid climate change caused by global warming (IPCC, 2014; IPCC, 2023). Heavy snowfall, one of the abnormal climate phenomena, frequently occurs in the northern mid-latitudes (Krasting et al., 2013) and causes significant direct and indirect socio-economic damage. In February 2021, shipments of COVID-19 vaccines to New York, USA, were suspended due to the heaviest recorded snowfall in the past decade (France24, 2021). In January 2019, a snowstorm in Austria killed four people and isolated 12,000 tourists (United Press International, 2019). In March 2018, heavy snowfall and cold waves in Europe killed 16 people, and over 350 flights were canceled (Associated Press, 2018). In December 2020, a snowstorm in Japan left around 1,000 cars stranded and about 10,000 households without power (Deutsche Welle, 2020).

According to Article 3, No. 1 of the Framework Act on the Management of Disasters and Safety in South Korea, heavy snowfall is classified as a major natural disaster. The damages caused by heavy snowfall have been incurred nationwide in the safety fields of roads, logistics, transportation, and facilities. According to the "Disaster Annual Report 2019" published by the Ministry of Interior and Safety (MOIS), which annually establishes and publishes major statistics on the damage and recovery status of natural disasters, typhoons, heavy rainfall, and heavy snowfall damage have accounted for approximately 53.85 % ($1550 million), 35.21 % ($1014 million), and 6.47 % ($186 million), respectively, of the total damage caused by natural disasters over the past decade (2010–2019) (MOIS, 2020). Heavy snowfall has caused extensive damage in Korea; thus, studies on heavy snow projection and damage reduction are required.

To mitigate the impact of heavy snowfall, research on projecting snowfall has primarily been carried out within the realms of meteorology and climatology. More recently, such research has also expanded into the domain of disaster management. First, previous studies on snowfall prediction conducted in South Korea, our study area, were reviewed. Meteorological data, such as temperature, precipitation, relative humidity, wind speed, solar insolation, snowfall depth data, and geographical data, such as latitude, longitude, and elevation, were collected and utilized for developing a snowfall prediction model using machine learning techniques. In addition, the predictive snowfall using climate change RCP scenario data provided by the KMA was determined (Kim et al., 2013; Park et al., 2014; Kim et al., 2014; Park et al., 2016; Oh et al., 2020). Tabari et al. (2010) compared the predicted snow depth results, derived using MLR, ANN and neural network-genetic algorithm (NNGA), using latitude, longitude, elevation, snow cover, and snow density as the input variables. Comparing the R^2 and RMSE values of the model determined that the NNGA yielded optimum results with R^2 and RMSE values of 0.70 and 0.202 cm, respectively. Hamidi et al. (2018) predicted monthly snowfall in Iran using SVM, RF, and MLR methods. Their study was conducted using time-series forecasting, and monthly snowfall observation data were used as input variables. The performance of each model was evaluated using the RMSE and R^2 values, and the SVM model exhibited exceptional performance with an R^2 value of 0.95. This value was applied for snowfall prediction in the area. Zhang et al. (2019) performed snowfall predictions for mountainous regions. Eight factors, including average temperature, relative humidity, wind speed, latitude, longitude, elevation, slope, and slope direction, were used as input parameters for the MLR and RF models to project snowfall. The coefficient of determination of the RF model was 0.74, which was superior to that of the linear regression model. Vafakhah et al. (2022) used various machine learning algorithms, such as BANN, SVM, Cubists, and RF, to predict Snow Water Equivalent (SWE) in the Sohrevard watershed in Iran. Nine geo-environment variables (altitude, slope, eastness, profile curvature, plan curvature, solar radiation, TPI, TWI, and wind exposition index) were used as SWE influencers. Based on the results obtained from the error metrics, the RF algorithm showed optimum performance. Meng and Zhu (2023) have provided a description of a study that evaluated seven machine learning algorithms for modeling monthly and seasonal snowfall in the lower peninsula of Michigan based on selected environmental and climatic variables using 65 years of on-site snowfall observations. Monthly temperature variables (maximum, minimum and average), maximum and minimum vapor pressure, latitude, longitude, and elevation of each station, among other factors, were utilized as input variables. The Bayesian Additive Regression Trees (BART) model emerged as the best fit for estimating monthly average snowfall, outperforming the RF model, with an R^2 value of 0.88. Snowfall is a nonlinear process in which precipitation, temperature, relative humidity, and geographic variables are variously related. Studies have been conducted using statistical and machine learning techniques that can consider the nonlinear relationship of factors. This is because nonlinear activation functions (Sigmoid and Tanh) are used in machine learning algorithms to explain the nonlinear relationship between weather factors.

Table 1 summarizes the machine learning algorithms and input factors used in previous studies. Recent research has incorporated machine learning models such as RF, SVR, and MLR, among others, with RF demonstrating favorable performance.

Classifying, identifying, and predicting specific patterns and key characteristics of spatiotemporal climate and environmental data is significant for various purposes, such as identifying extreme weather patterns, studying the effects of climate change, and responding to disasters (Chattopadhyay et al., 2020). Recently, to identify disaster-vulnerable areas, studies on image grouping have been conducted by analyzing images, such as satellite and LiDAR images, using the K-means clustering algorithm. Chattopadhyay et al. (2020) studied how to cluster spatiotemporal RGB image data into n classes using unsupervised techniques, such as K-means. Ibrahim et al. (2021) collected flood image samples comprising land and river areas to develop an unmanned aerial

Table 1
Heavy snowfall projection: Overview of variables and methods.

Method	Input variables	Output variable	Methods	Performance
Tabari et al., 2013	Precipitation, elevation, latitude, longitude	Snow depth	MLR, ANN, NNGA	NNGA
Kim et al., 2013	Minimum temperature, maximum temperature, average temperature, precipitation	Snow depth	ANN, MLR	ANN > MLR
Park et al., 2014	Minimum temperature, maximum temperature, average temperature, precipitation	Snow depth	ANN	ANN
Kim et al., 2014	Minimum temperature, maximum temperature, average temperature, precipitation	Snow depth	ANN	–
Park et al., 2016	Minimum temperature, maximum temperature, average temperature, precipitation	Snow depth	MLR	–
Hamidi et al., 2018	Snow depth	Snow depth	RF, SVM, MARS	RF > SVM > MARS
Zhang et al., 2019	Average temperature, relative humidity, latitude, longitude, aspect, elevation, wind speed, slope	Snowfall	RF, MLR, RNN	RF > RNN > MLR
Oh et al., 2020	Temperature, snow depth, rate of humidity change, solar radiation rate	Snow Melting Depth	MLR	–
Feng et al., 2022	Average temperature, average atmospheric pressure, sunshine duration, wind speed, snow depth, longitude, latitude, elevation, slope angle	snow density	MLR RF, XGB, LGBM	RF > XGB > LGBM > MLR
Vafakhah et al., 2022	Altitude, slope, curvature, solar radiation, TPI, TWI, wind exposition	SWE	BANN, SVM, Cubist, RF	RF > Cubist > BANN > SVM
Meng & Zhu, 2023	Monthly temperature variables, maximum and minimum vapor pressure, Latitude, Longitude, and elevation of each station, etc.	Snowfall	BART, RF, SVM, etc.	BART > RF > SVM

vehicle (UAV)-based flood detection automation system. Flood images were expressed using RGB and HSI color models, and flood-vulnerable areas were classified using the K-means clustering image segmentation method. Shafapourtehrany et al.(2022) used the K-means clustering algorithm to group large GIS-based image data collections, performing seismic vulnerability mapping in Istanbul, Turkey. This can improve the innovation and model performance required for seismic vulnerability mapping.

In this study, we centered our research on South Korea as our primary study area. We deliberately selected input variables that had not been previously applied in domestic studies, namely longitude, latitude, and elevation. Our primary objective was to project snowfall depth using machine learning algorithms that have demonstrated efficacy in previous research. To achieve this, we focused on four leading approaches proven to be most effective in various applications: MLR, SVR, RFR, and XGB (Bedi et al., 2020). Furthermore, we extended our analysis by applying the projection model derived from our research to RCP climate change scenarios. This enabled us to forecast snowfall distribution in a geographic information system (GIS) context. Subsequently, we performed image clustering using the K-means clustering technique. The clustering results allowed us to identify areas that are particularly vulnerable to heavy snowfall. The insights gained from these clustered areas hold significant implications for future heavy snow disaster management efforts undertaken by government and public institutions. This study contributes to a more comprehensive understanding of the spatial distribution of heavy snowfall and facilitates proactive planning and risk mitigation in the face of changing climate conditions.

Materials and methods

Data description

Study area and field data

South Korea is located in East Asia and has an area of 100,410 km^2. Its geographical coordinates are 124°–132°E, 33°–43°N. Climatically, South Korea has a relatively long winter, with an average temperature of 3.8 °C, an average maximum temperature of 9.2 °C, an average minimum temperature of − 1.2 °C, an average precipitation of 8.9 mm, and an average snowfall of 5.7 mm during the winter season (October to April) for the last 10 years. Fig. 1 shows the study area and 102 ASOSs and 8 provinces in South Korea.

In a comprehensive examination of methodologies and factors employed in prior studies, we observed that nine specific variables had been consistently and widely utilized in the construction of snowfall projection models. These nine variables are categorized into two main groups: meteorological factors, which encompass minimum temperature, maximum temperature, average temperature, precipitation, relative humidity, and snowfall; and geographic factors, comprising latitude, longitude, and elevation data sourced from ASOSs, as shown in Table 2. Daily meteorological data over the past 30 years (1991–2020) during the winter (October to April) were collected from 102 ASOSs nationwide under the KMA.

The datasets with missing independent variables were eliminated because machine learning is difficult to perform when there are missing values in the dataset (Pratama et al., 2016). Among the 945,748 daily datasets collected, 42,701 were selected after excluding non-snowy days and datasets with missing values. In addition, multicollinearity analysis was performed. Multicollinearity results in an inaccurate analysis due to the strong correlations between the independent variables in the regression analysis. A general diagnostic index of multicollinearity states that multicollinearity occurs when the TOL is below 0.1 or the VIF is above 10 (Ainiyah et al., 2016). A high VIF indicates high collinearity (Mallick et al., 2021). Among the independent variables, we performed multicollinearity analysis on the meteorological factors (average temperature, minimum temperature, maximum temperature, daily precipitation, and average relative humidity) and snowfall. Table 3 shows the results of the multicollinearity analysis. The VIF of the average temperature was 21.738. After dimensionality reduction, multicollinearity analysis was repeated by excluding the average temperature from the independent variables. The variance expansion coefficient of the variables was ≤ 2, and it was verified that multicollinearity was absent.

RCP scenario data

To produce projected snowfall data for each RCP scenario, RCP scenario data was downloaded from the Korea Meteorological Administration's "Climate Information Portal" (https://www.climate.go.kr/home/). RCP scenario data provide grid size (135, 12.5, and 1 km), scenario type, climate data type, climate projection model, weather

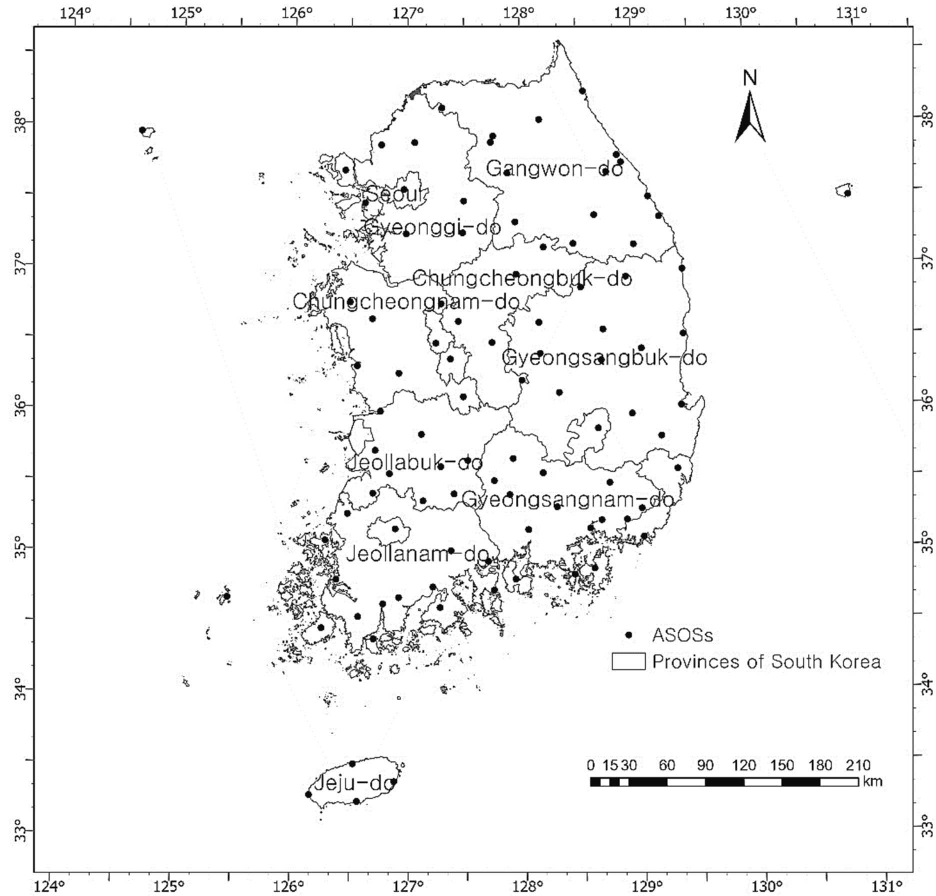

Fig. 1. Study area: 102 ASOSs and 8 provinces in South Korea.

Table 2
Meteorological and geographic factors.

	Input Variables	Output Variables
Meteorological factors (daily)	Minimum temperature (°C), maximum temperature (°C), average temperature (°C), precipitation (mm), relative humidity (%)	Snowfall (cm)
Geographic factors	Latitude (°), longitude (°), and elevation (m)	

variable, and file format, among others, and can be downloaded by designating each data element desired by the user (Lee et al., 2020). For GIS analysis, each ASCII file is reclassified by year, header information input (e.g., number of grids in x and y directions, grid size, and left/bottom grid positions), data matrix structure change, and data refinement work of resolution conversion and spatial disaggregation. A total of 467,200 files were reproduced and used for analysis, combining four types of RCP scenarios and 29,200 units of data for the next 80 years (2021 to 2100). As shown in Table 4, the climate projection model was selected as HadGEM3-RA, which is a regional climate model for South Korea. It is known that it properly simulates the statistical characteristics of Korean observation data and has excellent climate forecast reproducibility compared to other models (Reichler and Kim, 2008; S. Kim et al., 2019; Lee et al., 2020; Z. Kim et al., 2022). In addition, factors such as RCP scenario, grid size (1 km), and climate factors (maximum temperature, minimum temperature, precipitation, and relative humidity) were selected.

Approaches and processing procedures

Fig. 2 shows a schematic of the main structure of this study, including data collection and pre-processing, model development and comparison, application of RCP scenarios, and image clustering. The details of the algorithm and datasets are described below.

Application and comparison of machine learning algorithms for the snowfall projection model

The pre-processed datasets comprised the final seven inputs (minimum temperature, maximum temperature, daily precipitation, average

Table 3
Multicollinearity analysis.

1st	Input Variables	TOL	VIF	2nd	Input Variables	TOL	VIF
	Average temperature (°C)	0.046	21.738		Average temperature (°C)	-	-
	Minimum temperature (°C)	0.104	9.585		Minimum temperature (°C)	0.533	1.877
	Maximum temperature (°C)	0.149	6.689		Maximum temperature (°C)	0.561	1.783
	Precipitation (mm)	0.816	1.226		Precipitation (mm)	0.816	1.226
	Relative humidity (%)	0.849	1.178		Relative humidity (%)	0.849	1.178
Output variables: snowfall (cm)							

Table 4
RCP Scenario Data Information.

Category	Selected factors
RCP scenario type	RCP2.6, RCP4.5, RCP6.0, RCP8.5
Area	South Korea
Climate model	HadGEM3-RA
Climate factors	Minimum temperature, Maximum temperature, Precipitation, Relative humidity
Spatial resolution	1 km
Time resolution	Daily
Time range	2021–2100
File format	ASCII (including header information for GIS utilization)

relative humidity, latitude, longitude, and elevation) and one output variable. Furthermore, four machine learning algorithms (MLR, SVR, RFR, and XGB) were trained. The snowfall projection model was developed on a Jupyter Notebook (64-bit Windows 10) using Python 3.7. The optimal hyperparameters for each algorithm were selected and applied using a grid search technique during the learning process. Regression model optimization using machine learning refers to estimating a hyperparameter that minimizes a predefined loss function in the training data (Luo, 2016). This study applied the grid search and k-fold cross-validation methods to select the optimal hyperparameter. The range of each parameter was set, the accuracy of the model generated according to the combinations was measured, and the optimal parameter that provided maximum accuracy was selected (Claesen and De Moor, 2015). For the k-fold cross-validation method, the datasets were k-equalized into sets of the same size. The k-1 among the divided datasets was used as the training data, and the remaining dataset was used as the testing data. This method was used to verify the performance of the model. In this study, five-fold cross-validation was applied (Vabalas et al., 2019). The model performance was evaluated by comparing the snowfall estimated by the trained model with the actual snowfall value measured at the observation station. The optimal model was determined by comparing and verifying the accuracy of the models using MAE, RMSE, and R^2.

MLR

Linear regression is an extensively used regression analysis model, and it has been used by researchers since the invention of artificial intelligence (Chaloulakou et al., 2003). This method derives the results of independent and dependent variables using a one-dimensional linear predictive equation. When the cost function has a minimum value, the

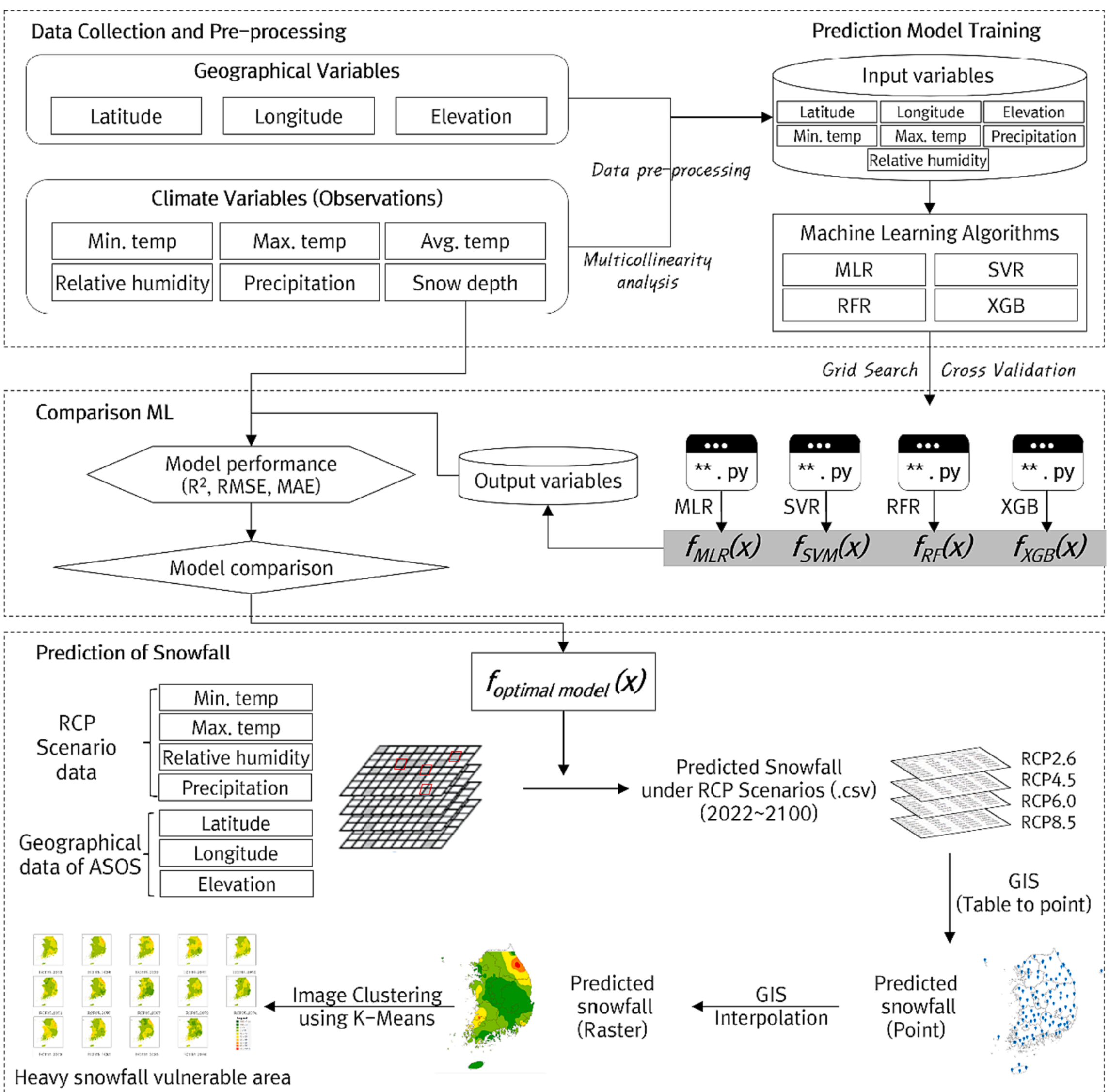

Fig. 2. Research workflow.

derived equation is the optimal predictive model. The least-squares or gradient descent method is mainly used to determine the minimum value of the loss function (Liu et al., 2021). Linear regression analysis refers to the estimation of a dependent variable using a statistical method considering the independent variables (X_1, X_2, …, X_k) expected to affect the dependent variable (Y) significantly. Multiple regression analysis was performed for the independent variables (factors affecting snowfall) in this study. Additionally, the variables were adjusted and analyzed after multicollinearity analysis was performed.

SVR

SVM (Cortes and Vapnik, 1995) is a supervised machine learning algorithm for classification problems. The input variable is built into a high-dimensional functional space using a linear or nonlinear kernel function depending on the relationship between the dependent and independent variables. A linear model was developed in the feature space to maintain a balance between error minimization and overfitting (Bansal et al., 2021). SVR (Vapnik et al., 1997) is an extension of SVM applicable to classification problems and projection fields, such as regression analysis (Bermolen and Rossi, 2009). SVR learns in a direction that maximizes the distance between the separation hyperplane and support vector within a threshold (Carrera and Kim, 2020).

RFR

The RF algorithm is a DT-based algorithm (Breiman, 2001). It is a model of an ensemble technique developed by combining multiple DTs with different structures and performances. Furthermore, it outputs classification or average predictions (regression analysis) from multiple DTs constructed during the training process. The RFR compensates for the bias introduced by a single DT owing to the randomness. Therefore, it does not easily overfit and provides high accuracy and a fast training speed (Babar et al., 2020). The RFR algorithm randomly selects data (bootstrapping) and learns individually. Bagging is an abbreviation for bootstrap and aggregation, a concept that collects models generated from each bootstrap sample. Aggregating refers to merging datasets formed by bootstrapping, and a random subspace is applied to train the datasets. Determining the split point of the DT based on the split function implies that learning is performed by randomly selecting a number of variables less than the variables of the input data. Contrary to the DT algorithm, where the error is transferred at each intermediate node in RF, the error generated in the intermediate node of each tree is not transmitted to the terminal node and converges to the limit value. This improves the performance of the predictive model by minimizing the correlation between individual trees (Ganguly et al., 2019).

XGB

XGB (Chen and Guestrin, 2016) is known for its powerful performance, as demonstrated by recent studies. In addition, it has been extensively used in various applications. XGB is an algorithm based on GBM, a boosting model comprising a series of basic regression trees using a sequential ensemble technique (Zhu et al., 2021). This method improves error by sequentially repeating the learning prediction for several weak learners and assigning weights when the predicted values differ from the input data. The residual error of the model derived from Tree 1 was checked, and a predictive model that reduced the residual error of Tree 1 was derived from Tree 2. Subsequently, the residuals in Tree 2 were checked, and a predictive model that reduces the residuals in Tree 2 was derived using Tree 3. This method derives a model from the final tree with small residuals as the final prediction model while repeating this process (Zhu et al., 2021). Furthermore, XGB exhibits exceptional performance in classification and regression problems. The weight of the hidden layer is unknown in the case of commonly used ANN-based algorithms. Therefore, the correlation between each variable and the prediction model remains unknown. However, XGB has the advantage of being able to analyze the feature importance of variables.

GIS processing for snowfall projection images

By inputting the meteorological factors and geographic factor values extracted from the location of the ASOSs of the RCP scenario data into the optimal model, predictive snowfall data for each RCP scenario is generated. Using ArcGIS Pro, they are spatialized into point data with attribute values (latitude and longitude) and created as a 5 m grid raster data image of nationwide distribution through inverse distance weighting interpolation (Mitas and Mitasova, 1999; Murphy et al., 2010; Jang et al., 2015; Bienvenido-Huertas et al., 2022). Based on the predicted snowfall image, K-means image clustering was performed to predict areas and periods vulnerable to heavy snow. Snowfall projection images of the RCP8.5 scenario (2021–100) were graded into 10 levels, including 3 cm for snow removal operation, 5 cm for heavy snowfall advisory, and 20 cm for heavy snowfall warning.

K-means clustering

K-means clustering is a representative non-hierarchical clustering analysis that minimizes the average Euclidean distance between patterns and the center of the cluster to which the pattern belongs and clusters large amounts of data based on similar properties (Marta-Almeida et al., 2016; Ruela et al., 2020). In this study, feature vectors were extracted from predicted snowfall RGB images (.png) using the VGG16 model and grouped according to image characteristics using K-means clustering (Abid et al., 2021). The k-value is an important parameter of the grouping process. The elbow method is used to select an appropriate k-value, as the oldest method for determining the true number of clusters in a dataset (Kodinariya and Makwana, 2013). In this study, the optimal number of clusters was calculated using the elbow method. During clustering, the sum of the distances between clusters drops sharply in one section, and the value at this point was used as the optimal number of clusters (Bienvenido-Huertas et al., 2021).

Results and discussion

Snowfall projection by RCP scenarios

The applicability of $f_{MLR}(x)$, $f_{SVR}(x)$, $f_{RFR}(x)$, and $f_{XGB}(x)$, which were the optimal models for each algorithm, was evaluated using hyperparameters. The optimum hyperparameter results of each machine learning algorithm were derived through grid search and five-fold cross-validation (Table 5).

The RFR model exhibited MAE, RMSE, and R^2 values of 1.65 cm, 3.35 cm, and 0.64, respectively, using performance evaluation criteria. Additionally, it exhibited a higher projection accuracy than the three models (MLR, SVR, and XGB) (Table 6). The XGB model exhibited a performance similar to that of the RFR model because it was close to the evaluation standard value obtained based on the RFR model. For snowfall projection, ensemble models, such as RFR and XGB, demonstrated better performance than single regression models, such as MLR and SVR.

The projected daily snowfall values obtained using the MLR, SVR, RFR, and XGB models and the observed snowfall values are shown in Fig. 3 using scatter plots of (a) MLR, (b) SVR, (c) RFR, and (d) XGB. We

Table 5
Results of hyperparameter tuning.

Models	Hyperparameters	Optimal hyperparameters	
SVR	Kernel	Linear, Polynomial, Sigmoid, RBF	RBF
	Cost	0.01, 0.1, 1, 10, 100	1
	γ	0.01, 0.1, 1, 10, 100	1
RFR	max_features	4, 8, 10, 12, 14, 16, 18, 20	4
	n_estimators	10–1000	100
	max_depth	4, 6, 8, 10, 12	10
XGB	max_features	4, 8, 10, 12, 14, 16, 18, 20	4
	n_estimators	10–1000	20
	max_depth	4, 6, 8, 10, 12	6

Table 6
Comparative statistics of projection models.

	Criteria	MAE(cm)	RMSE(cm)	R^2
Models				
MLR	2.32	4.22	0.45	
SVR	1.73	3.91	0.53	
RFR	1.65	3.35	0.64	
XGB	1.64	3.44	0.62	

observed that the snowfall simulation of the RFR and XGB models exhibited better performance than that of the other two models. The RFR and XGB models accurately evaluated the nonlinear relationship between the predictor and independent variables using a coefficient of determination. The MLR and SVR models partially interpreted the variance in snowfall. For the field observation data, there are a few datasets for high snowfall and several datasets for low snowfall. The imbalance in datasets was analyzed as the result of underestimating the MLR and SVR models (Park et al., 2021). Finally, a comparison between the statistical criteria of the four models demonstrated that the RFR was the optimum model for predicting snowfall. The predictive performance of the RFR model was exceptional because assuming a correlation between the dependent and independent variables in this model was unnecessary. In addition, the performance of the ensemble models (RFR and XGB) was better than that of the single regression models (MLR and SVM). They are less sensitive to datasets with inappropriate error distributions (Zhang et al., 2019).

Applying RCP scenario data to the RFR model (Fig. 4), the trend of projected snowfall variation by year (2021–2100) and ten-year units (the 2020 s–2090 s) was analyzed. For comparison with the current trend, we used the annual and monthly statistics for each meteorological station data for the last 21 years (2000–2020) from "Statistics Korea" (Statistics Korea, 2022). Among all ASOS points, the peak snowfall was 87.7 cm, observed in Daegwallyeong in 2001. The average value of the peak snowfall for each point was 48.2 cm, and the average snowfall was 10.3 cm. Heavy snowfall events projected to be above the average value of the peak snowfall (48.2 cm) and the national peak snowfall (87.7 cm) in the past 21 years are shown in all scenarios. Extreme snowfall (48.2 cm) was observed 17 times in RCP2.6, 19 times in RCP4.5, 16 times in RCP6.0, and 17 times in RCP8.5 in each scenario; the vulnerable areas with high snowfall projected were Daegwallyeong, Yeongwol, Taebaek, Gangneung, Jecheon, Chungju, and Jeongeup (not shown). This amount of snowfall exceeds past heavy snow events, indicating that heavy snow damage may occur in these vulnerable cities. Among the scenarios, RCP2.6 and RCP8.5 show the highest projected snowfall in the 2070 s and 2090 s, and RCP4.5 and RCP6.0 show the highest projected snowfall in the 2020 s and 2030 s. The projected snowfall is expected to increase compared to the current average (10.3 cm). However, in all scenarios except RCP2.6, the projected peak snowfall tends to decrease overall (see trendlines of Fig. 4) because the maximum and minimum temperature factors used in machine learning increased in the future. Increased precipitation may contribute to the increase in snowfall due to climate change, but the projected snowfall also tends to decrease gradually because the temperature increases and the number of occurrences below the freezing point decreases (Kim et al., 2014). Based on our analysis using RCP climate change scenarios, it is projected that the average annual snowfall will decrease in the future, but the tendency for heavy snowfall will increase. It is important to note that only one regional climate model was employed in this analysis, and the RCP scenario itself entails inherent uncertainties.

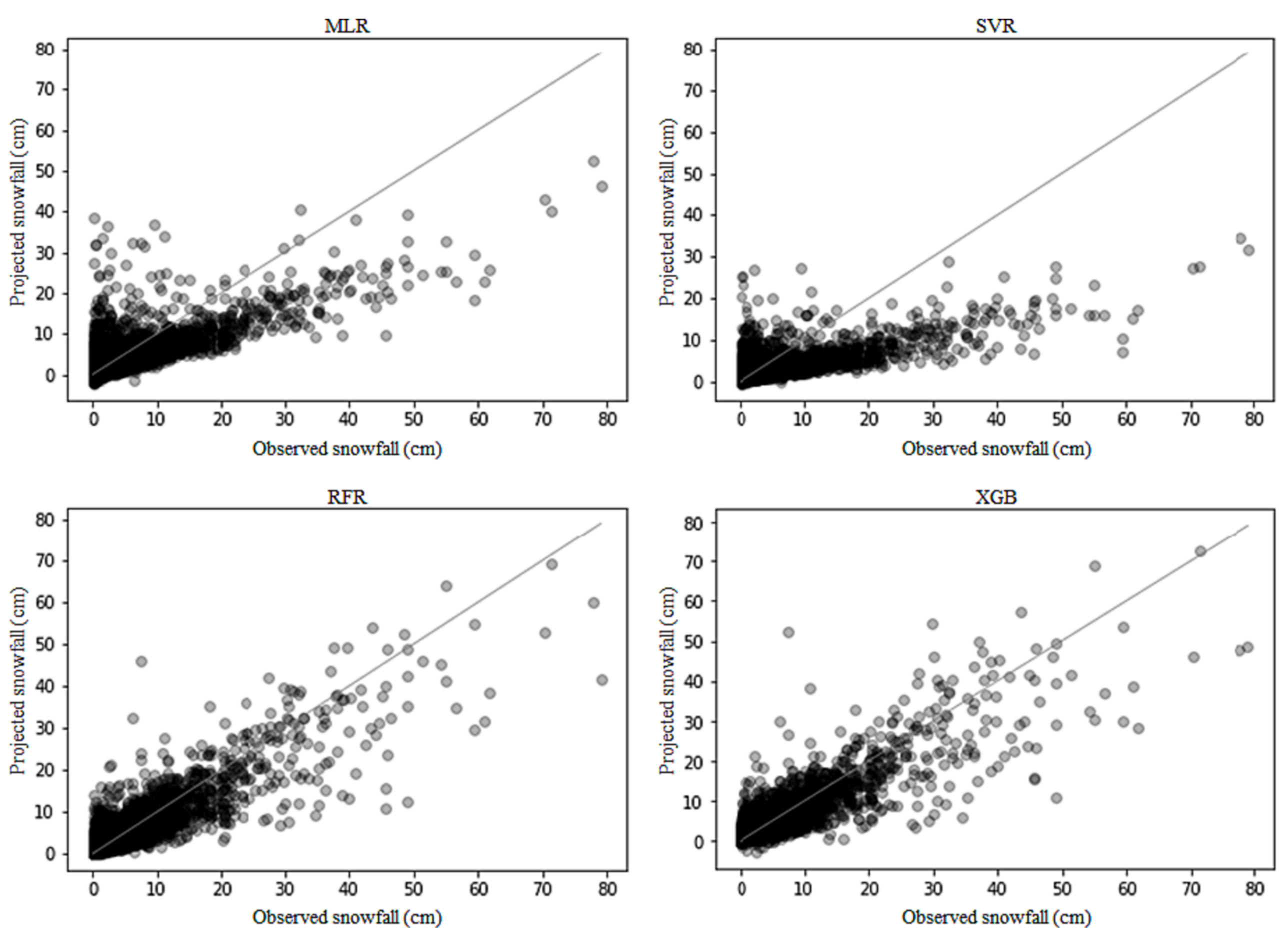

Fig. 3. Correlation of observed and projected snowfall results from (a) MLR, (b) SVR, (c) RFR, (d) XGB.

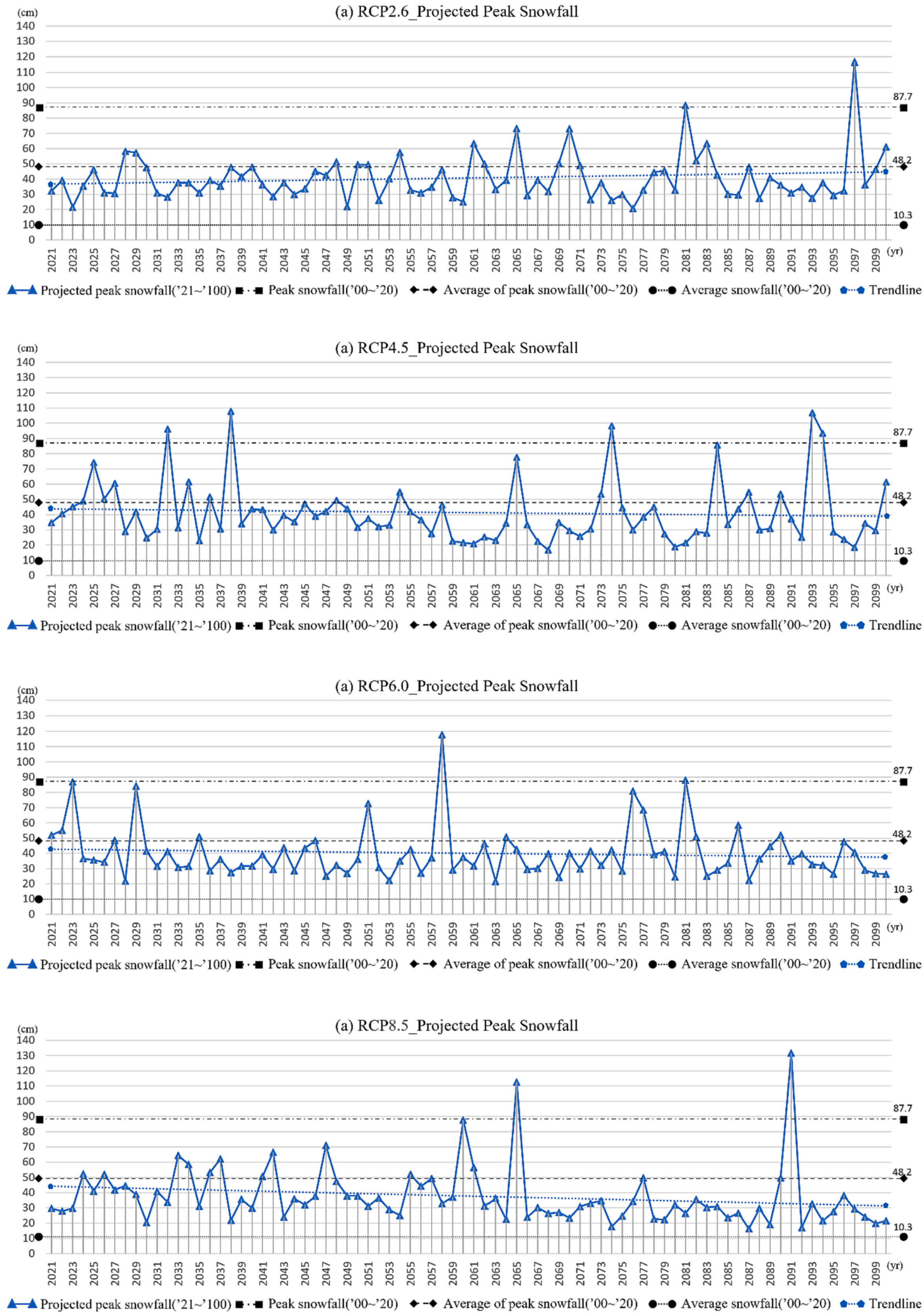

Fig. 4. Projected peak snowfall according to RCP scenarios (a) RCP2.6, (b) RCP4.5, (c) RCP6.0, and (d) RCP8.5.

Image clustering of the areas vulnerable to heavy snowfall

Considering the prediction results for each scenario, extreme snowfall (48.2 cm) occurred a similar number of times. Compared to the complete set of RCPs, RCP8.5 corresponds to the pathway with the highest levels of greenhouse gas emissions (Riahi et al., 2011). Subsequently, clustering analysis was conducted on the 8.5 scenario, which is recognized as the most severe scenario. The 80 annual snowfall projection images were clustered into 5 groups and categorized into 10 levels, aligning with the characteristics of RGB values in GIS. These levels ranged from the annual minimum value of 1.212 cm to the annual maximum value of 131.6 cm. The categorization included standards such as 3 cm for highway snow removal operations, 5 cm for snowfall advisories, and 20 cm for snowfall warnings. Raster images within each group were averaged using GIS spatial analysis to create a representative image for the group. The representative image for each group can be seen on the right side of Figs. 5–9. These images are categorized into 10 levels, ranging from the minimum value of 1.82 cm to the annual maximum value of 56.41 cm.

In Fig. 5, Group (1) was identified as a vulnerable distribution type for Chungcheongbuk-do and Gangwon-do. This group exhibited relatively high projected snowfall (Grades 7 to 8), which corresponds to more than 15 cm of snowfall. This level of snowfall is in line with the criteria for issuing heavy snow warnings in South Korea. In Fig. 6, Group (2) was identified as a vulnerable distribution group in Jeollabuk-do and Gangwon-do. This group exhibited relatively high projected snowfall (Grades 6 to 7), which corresponds to 10 cm or more of snowfall. In Fig. 7, Group (3) was identified as a nationwide distribution type. This group exhibited relatively low projected snowfall (Grades 1 to 4). It contained the largest number of images among the groups. The minimum projected snowfall within this group was observed in Gyeongsangbuk-do, with projected snowfall distribution being less than 3 cm. This aligns with the starting standard for snow removal work in South Korea. In Fig. 8, Group 4 was identified as a vulnerable distribution group in northern Gyeonggi-do. This group exhibited relatively high projected snowfall, falling within Grades 7 to 8. It is noteworthy that significant snowfall is expected in areas outside of Gangwon-do. In Fig. 9, Group 5 was identified as a vulnerable distribution group in Gangwon-do, encompassing the maximum projected snowfall (131 cm) from the scenario predicted for the year 2091. This group represents areas in Gangwon-do with the highest expected snowfall levels, ranging from Grades 8 to 10.

Conclusions

This study was conducted to project the areas and periods vulnerable to heavy snowfall as the damage caused by snowfall increases. A heavy snow projection model using machine learning was developed, and heavy snowfall image clustering was performed for the RCP8.5 scenario using GIS and K-means clustering. The snowfall projection model was developed according to the following steps. Independent variables were selected by analyzing previous studies, and data collection was performed by considering the meteorological and geographic factors collected through the ASOSs in South Korea. Data pre-processing was performed, and the pre-processed data were learned using the MLR, SVR, RFR, and XGB machine learning algorithms. The machine learning algorithms with good performance and widely used algorithms in previous studies were selected as the regression models for projection purposes. The predictive model using the RFR algorithm had the optimal performance ($R^2 = 0.64$) based on a comparison between the observed and projected data. The distribution of snowfall up to 2100 was projected by applying the RCP scenario to the RFR model, selected as the optimal model. Due to examining the heavy snowfall trends for each scenario, RCP2.6 and RCP8.5 showed the highest snowfall (116.6 cm and 131.6 cm) after the 2060 s, and RCP4.5 and RCP6.0 showed the highest snow cover (107.7 cm and 117.8 cm) before the 2060 s. The snowfall trend decreases toward the future, which was analyzed to decrease as the temperature factor used in the projection model increases due to the influence of global warming. In addition, K-means clustering analysis was performed to identify vulnerable areas with the projection model using RCP8.5 and the yearly snowfall images derived through GIS processing. Through image clustering, vulnerable areas were clustered into five groups, including Gangwon-do, Jeollabuk-do, and Gyeonggi-do. It is meaningful because it grouped similar vulnerable areas by year using time-series image data. Understanding each group's clustered year and vulnerable areas and the change in snowfall patterns is imperative in planning a response plan to minimize the risk of heavy snow. Therefore, local governments and public agencies that manage infrastructure in these three vulnerable areas should consider establishing measures to deal with heavy snow disasters.

Snowfall is a nonlinear process in which meteorological and geographic variables such as precipitation, temperature, relative humidity, wind speed, wind direction, latitude and longitude are correlated. Additionally, the projection results may vary depending on the regional research scope and characteristics of the input variable data used for model development. The meteorological factors were provided as daily data when used as input variables in this study. Since the daily average observation data were used as input data for the meteorological factors rather than the data when the heavy snowfall occurred, the performance of the projection model was relatively low (Park et al., 2016). Particularly, when predicting future snowfall using climate change RCP scenario data, improving the predictive power of the model

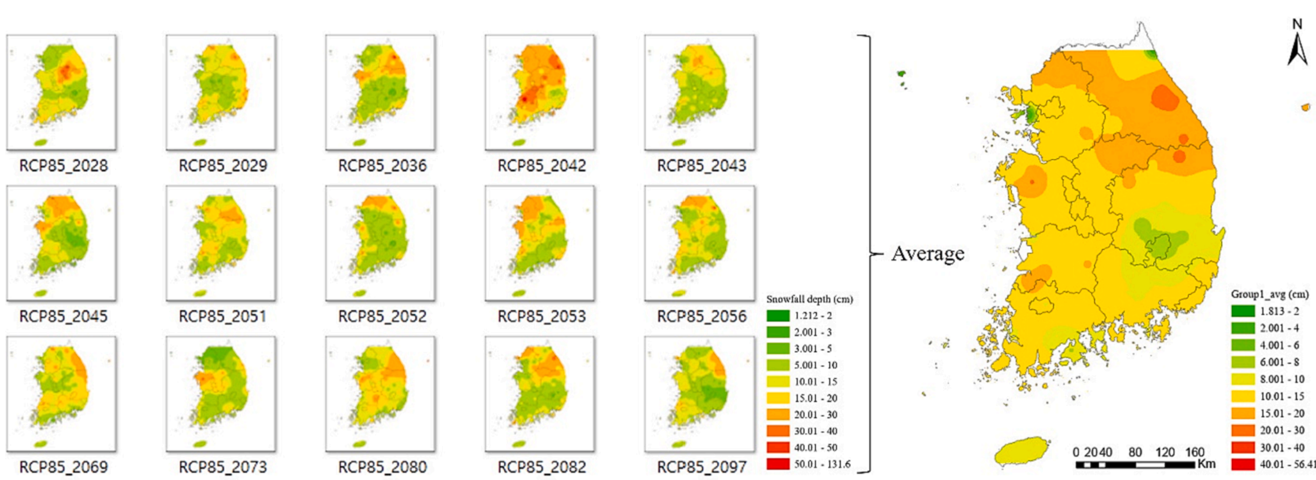

Fig. 5. Group (1): Group of Chungcheongbuk-do and Gangwon-do.

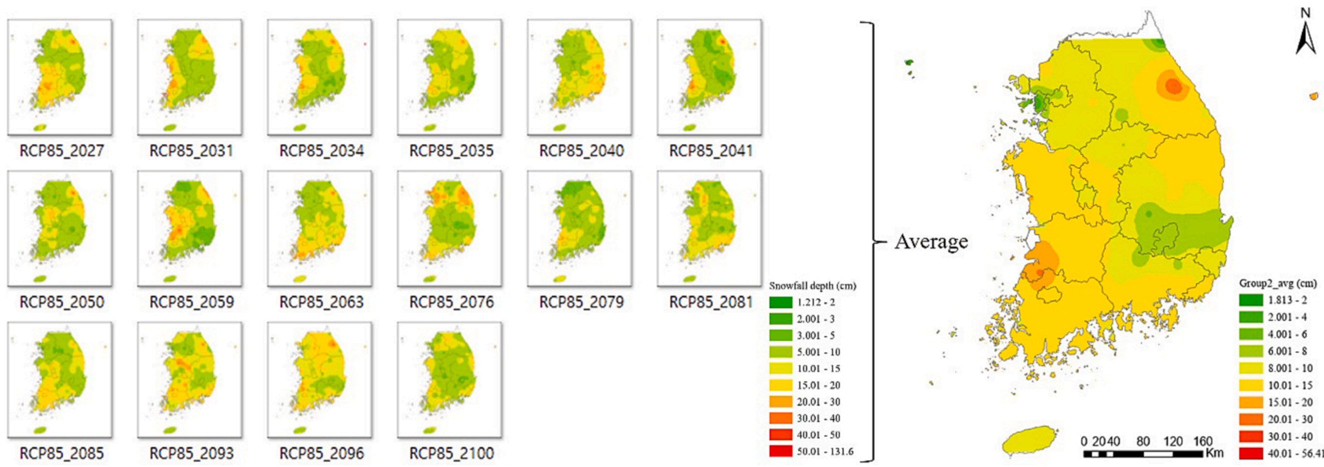

Fig. 6. Group (2): Group of Jeollabuk-do and Gangwon-do.

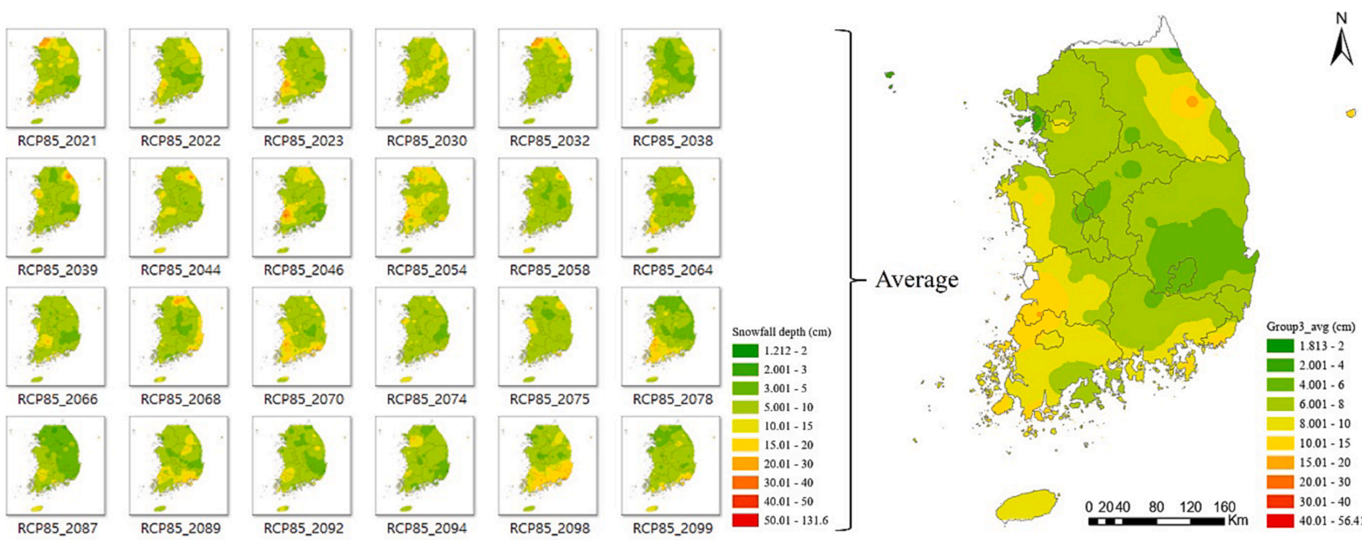

Fig. 7. Group (3): Group of national distribution with relatively low projected snowfall.

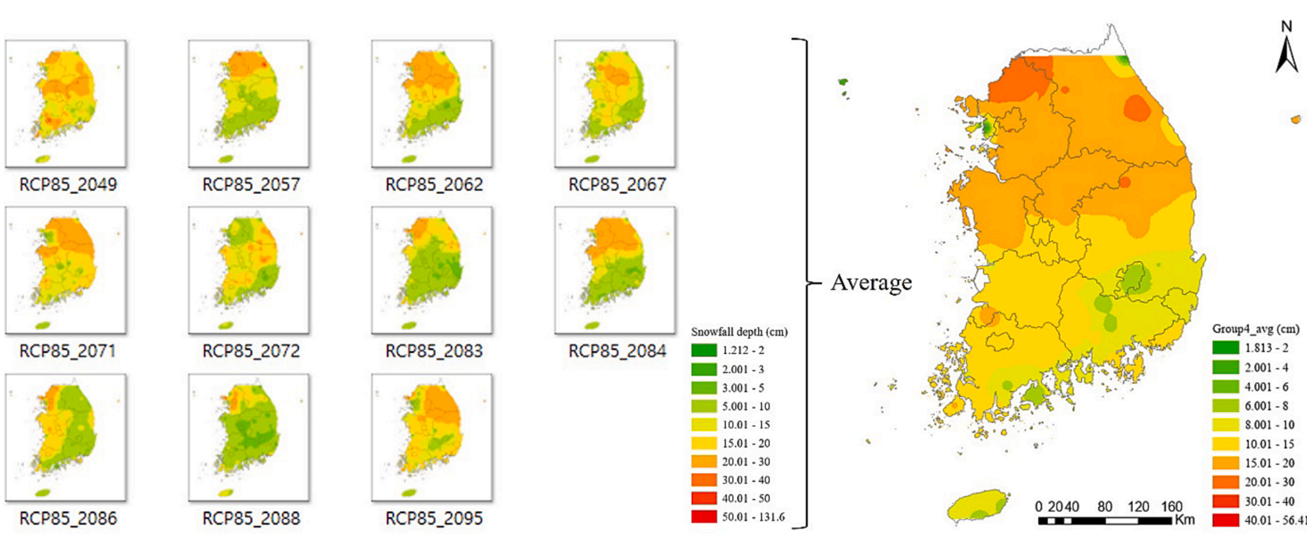

Fig. 8. Group 4: Group of northern Gyeonggi-do.

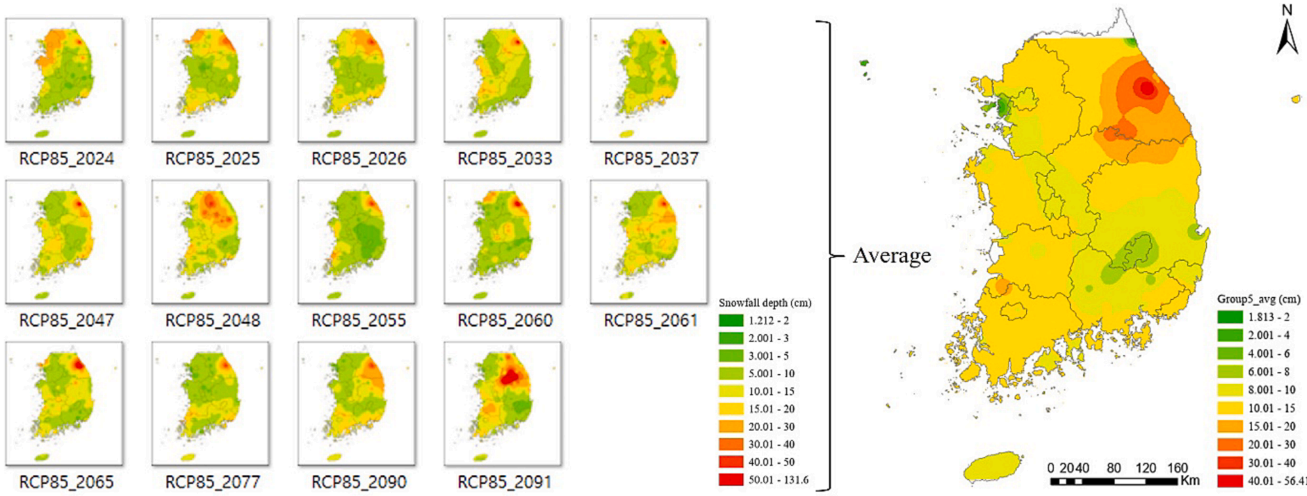

Fig. 9. Group 5: Group of Gangwon-do with the highest projected snowfall.

is difficult considering the uncertainty of the scenario. We recognize the potential benefits of employing multiple RCM models for uncertainty assessment, and we encourage further research in this direction to enhance the reliability of snowfall projections in South Korea. Therefore, the continuous development and validation of predictive models remain of utmost importance. Given the inherent uncertainties associated with climate projection models and RCP scenario data, it becomes imperative to enhance and validate our projection models persistently. This can be achieved through incorporating high-resolution satellite imagery, diversification of learning factors, and advancements in algorithms, all essential for bolstering the reliability of future projection results.

Heavy snowfall disasters adversely affect national infrastructure, such as highways. Thus, it may appear as a problem due to the paralysis of transportation functions in South Korea and is not limited to the vulnerable areas of this study. In the future, this study's results will help establish policies to respond to heavy snow disasters for transportation infrastructures, such as highways, roads, airports, and rail lines.

CRediT authorship contribution statement

Moon-Soo Song: Conceptualization, Methodology, Investigation, Project administration, Supervision, Resources, Data curation, Writing – review & editing. **Jae-Joon Lee:** Methodology, Data curation, Writing – review & editing. **Hong-Sic Yun:** Conceptualization, Methodology, Resources, Validation, Funding acquisition, Writing – review & editing. **Sang-Guk Yum:** Conceptualization, Methodology, Investigation, Resources, Validation, Funding acquisition, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This research was funded by National Research Foundation, Korea (NRF) grant funded by the Korea government (2021R1C1C2010999) and the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2021R1A6A1A03044326). Also this research was supported by Research Grant of Kyungwoon University in 2023.

Appendix 1

Regression metrics

Several criteria were used to evaluate the performance of the regression models. The accuracy of the model was compared and verified using the MAE, RMSE, and R^2 values (Guo et al., 2021). The MAE is the arithmetic mean of the absolute value of the difference between the measured and estimated values. The MAE has high applicability if it has a value close to zero. The low RMSE values demonstrate that the error of the estimation model was small. In this study, it was used to indicate the suitability of estimating high snowfall (Hamidi et al., 2018). R^2 measures the linear relationship between the observed and estimated snowfall and has a value of 0 to 1. An R^2 value close to 1 indicates optimum model applicability. The MAE, RMSE, and R^2 values were calculated using Eqs. (1), (2), and (3).

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|X_i - Y_i| \tag{1}$$

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(X_i - Y_i)^2} \tag{2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(X_i - Y_i)^2}{\sum_{i=1}^{m}(\overline{Y} - Y_i)^2} \quad (3)$$

where X_i is the projected i_{th} value, and Yi is the actual i_{th} value. The regression method predicts the X_i element for the corresponding Y_i element in the observation dataset (Chicco et al., 2021).

References

Abid, N., Shahzad, M., Malik, M.I., Schwanecke, U., Ulges, A., Kovács, G., Shafait, F., 2021. UCL: Unsupervised Curriculum Learning for water body classification from remote sensing imagery. Int. J. Appl. Earth Obs. Geoinf. 105 (102568) https://doi.org/10.1016/j.jag.2021.102568.

Ainiyah, N., Deliar, A., Virtriana, R., 2016. The classical assumption test to driving factors of land cover change in the development region of northern part of west Java. Int. Arch. Photogr. Remote Sens. Spatial Inform. Sci. ISPRS Arch. 41 (July), 205–210. https://doi.org/10.5194/isprsarchives-XLI-B6-205-2016.

Babar, B., Luppino, L. T., Boström, T., & Anfinsen, S. N. (2020). Random forest regression for improved mapping of solar irradiance at high latitudes. *Solar Energy*, *198* (November 2019), 81–92. doi: 10.1016/j.solener.2020.01.034.

Bansal, N., Defo, M., Lacasse, M.A., 2021. Application of support vector regression to the prediction of the long-term impacts of climate change on the moisture performance of wood frame and massive timber walls. Buildings 11 (5). https://doi.org/10.3390/buildings11050188.

Bedi, S., Samal, A., Ray, C., Snow, D., 2020. Comparative evaluation of machine learning models for groundwater quality assessment. Environ. Monit. Assess. 192 (776), 1–23.

Bermolen, P., Rossi, D., 2009. Support vector regression for link load prediction. Comput. Netw. 53 (2), 191–201. https://doi.org/10.1016/j.comnet.2008.09.018.

Bienvenido-Huertas, D., Rubio-Bellido, C., Marín-García, D., Canivell, J., 2021. Influence of the Representative Concentration Pathways (RCP) scenarios on the bioclimatic design strategies of the built environment. Sustain. Cities Soc. 72, 103042 https://doi.org/10.1016/j.scs.2021.103042.

Bienvenido-Huertas, D., Sánchez-García, D., Rubio-Bellido, C., 2022. Influence of the RCP scenarios on the effectiveness of adaptive strategies in buildings around the world. Build. Environ. 208 (108631) https://doi.org/10.1016/j.buildenv.2021.108631.

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32.

Carrera, B., Kim, K., 2020. Comparison analysis of machine learning techniques for photovoltaic prediction using weather sensor data. Sensors (Switzerland) 20 (11). https://doi.org/10.3390/s20113129.

Chaloulakou, A., Grivas, G., Spyrellis, N., 2003. Neural network and multiple regression models for PM10 prediction in athens: A comparative assessment. J. Air Waste Manag. Assoc. 53 (10), 1183–1190. https://doi.org/10.1080/10473289.2003.10466276.

Chattopadhyay, A., Hassanzadeh, P., Pasha, S., 2020. Predicting clustered weather patterns: A test case for applications of convolutional neural networks to spatio-temporal climate data. Sci. Rep. 10 (1), 1–13. https://doi.org/10.1038/s41598-020-57897-9.

Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System Tianqi. Association for Computing Machinery, pp. 785–794.

Chicco, D., Warrens, M.J., Jurman, G., 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput. Sci. 7, 1–24. https://doi.org/10.7717/PEERJ-CS.623.

Claesen, M., De Moor, B., 2015. Hyperparameter Search in Machine Learning. Metaheuristics International Conference 10–14. http://arxiv.org/abs/1502.02127.

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Leaming*, *20*, 273–297. doi: doi: 10.1007/BF00994018.

Feng, T., Zhu, S., Huang, F., Hao, J., Mind'je, R., Zhang, J., & Li, L. (2022). Spatial variability of snow density and its estimation in different periods of snow season in the middle Tianshan Mountains, China. *Hydrological Processes*, *36*(8), 1–15. doi: 10.1002/hyp.14644.

France24. (2021, February 2). *Huge snowstorm blankets US East Coast, halting travel and vaccinations*. France24. https://www.france24.com/en/americas/20210202-huge-snowstorm-blankets-us-east-coast-halting-travel-and-vaccinations.

Ganguly, K. K., Nahar, N., & Hossain, B. M. (2019). A machine learning-based prediction and analysis of flood affected households: A case study of floods in Bangladesh. *International Journal of Disaster Risk Reduction*, *34*(March 2018), 283–294. doi: 10.1016/j.ijdrr.2018.12.002.

Guo, Y., Fu, Y., Hao, F., Zhang, X., Wu, W., Jin, X., Robin Bryant, C., Senthilnath, J., 2021. Integrated phenology and climate in rice yields prediction using machine learning methods. Ecol. Ind. 120, 106935 https://doi.org/10.1016/j.ecolind.2020.106935.

Hamidi, O., Tapak, L., Abbasi, H., Maryanaji, Z., 2018. Application of random forest time series, support vector regression and multivariate adaptive regression splines models in prediction of snowfall (a case study of Alvand in the middle Zagros, Iran). Theor. Appl. Climatol. 134 (3–4), 769–776. https://doi.org/10.1007/s00704-017-2300-9.

Ibrahim, N.S., Osman, M.K., Mohamed, S.B., Abdullah, S.H.Y.S., Sharun, S.M., 2021. The application of UAV images in flood detection using image segmentation techniques. Indonesian J. Electr. Eng. Comput. Sci. 23 (2), 1219–1226. https://doi.org/10.11591/ijeecs.v23.i2.pp1219-1226.

IPCC. (2014). Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. *IPCC*, Geneva, Switzerland, 151 pp. doi: 10.1017/CBO9781139177245.003.

IPCC. (2023). Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. *IPCC*, Geneva, Switzerland, 184 pp., doi: 10.59327/IPCC/AR6-9789291691647.

Jang, D., Park, H., Choi, J., 2015. Selection of Optimum Spatial Interpolation Method to Complement an Area Missing Precipitation Data of RCP Climate Change Scenario. Int. J. Software Eng. Its Applications 9 (8), 179–188.

Kim, Y., Kang, N., Kim, S., Kim, H., 2013. Evaluation for Snowfall Depth Forecasting using Neural Network and Multiple Regression Models. Korean Society of Hazard Mitigation 13 (2), 269–280.

Kim, Y., Kim, S., Kang, N., Kim, T., Kim, H., 2014. Estimation of Frequency Based Snowfall Depth Considering Climate Change Using Neural Network. J. Kor. Soc. Hazard Mitigation 14 (1), 93–107. https://doi.org/10.9798/kosham.2014.14.1.93.

Kim, S., Kim, H., Jung, Y., Heo, J.-H., 2019. Assessment of Frequency Analysis using Daily Rainfall Data of HadGEM3-RA Climate Model. J. Wetlands Res. 21 (spc), 51–60. https://doi.org/10.17663/JWR.2019.21.s-1.51.

Kim, Z., Shim, T., Jin, S., An, K., Jung, J., 2022. Prediction of three-dimensional shift in the distribution of largemouth bass (Micropterus salmoides) under climate change in South Korea. Ecol. Ind. 137 (February), 108731 https://doi.org/10.1016/j.ecolind.2022.108731.

Kodinariya, T.M., Makwana, P.R., 2013. Review on determining of cluster in K-means. International Journal of Advance Research in Computer Science and Management Studies 1 (6), 90–95. https://www.researchgate.net/publication/313554124.

Krasting, J.P., Broccoli, A.J., Dixon, K.W., Lanzante, J.R., 2013. Future changes in northern hemisphere snowfall. J. Clim. 26 (20), 7813–7828. https://doi.org/10.1175/JCLI-D-12-00832.1.

Lee, J., Kim, K., & Kim, K. (2020). *Study on Measures to Secure Safety of Expressway Facilities against Climate Change*.

Liu, S., Zeng, A., Lau, K., Ren, C., Chan, P. wai, & Ng, E. (2021). Predicting long-term monthly electricity demand under future climatic and socioeconomic changes using data-driven methods: A case study of Hong Kong. *Sustainable Cities and Society*, *70* (October 2020), 102936. doi: 10.1016/j.scs.2021.102936.

Luo, G., 2016. A review of automatic selection methods for machine learning algorithms and hyper-parameter values. Network Modeling Analysis in Health Informatics and Bioinformatics 5 (1), 1–16. https://doi.org/10.1007/s13721-016-0125-6.

Mallick, J., Alqadhi, S., Talukdar, S., Alsubih, M., Ahmed, M., Khan, R.A., Kahla, N.B., Abutayeh, S.M., 2021. Risk assessment of resources exposed to rainfall induced landslide with the development of gis and rs based ensemble metaheuristic machine learning algorithms. Sustainability (Switzerland) 13 (2), 1–30. https://doi.org/10.3390/su13020457.

Marta-Almeida, M., Teixeira, J.C., Carvalho, M.J., Melo-Gonçalves, P., Rocha, A.M., 2016. High resolution WRF climatic simulations for the Iberian Peninsula: Model validation. Phys. Chem. Earth 94, 94–105. https://doi.org/10.1016/j.pce.2016.03.010.

Meng, L., Zhu, L., 2023. Statistical modeling of monthly and seasonal Michigan snowfall based on machine learning: A multiscale approach. Artif. Intell. Earth Systems. https://doi.org/10.1175/AIES-D-23-0016.1.

Mitas, L., Mitasova, H., 1999. Spatial Interpolation. Geogr. Information Syst. 481–492 https://doi.org/10.4324/9781351243858-7.

Mois, 2020. 2019 Disaster Yearbook. Mois(ministry of the Interior and Safety).

Murphy, R.R., Curriero, F.C., Ball, W.P., 2010. Comparison of Spatial Interpolation Methods for Water Quality Evaluation in the Chesapeake Bay. J. Environ. Eng. 136 (2), 160–171. https://doi.org/10.1061/(asce)ee.1943-7870.0000121.

Oh, Y., Lee, G., Jun, K.S., Sunwoo, W., Baek, S., Chung, G., 2020. A Study on the Prediction of Daily Snowmelt Depth using Multiple Linear Regression. J. Korean Soc. Hazard Mitigation 20 (6), 311–321. https://doi.org/10.9798/kosham.2020.20.6.311.

Park, HeeSeong, Jeong, S., Chung, G., 2014. Frequency Analysis of Future Fresh Snow Days and Maximum Fresh Snow Depth using Artificial Neural Network under Climate Change Scenarios. J. Korean Soc. Hazard Mitigation 14 (6), 365–377. https://doi.org/10.9798/kosham.2014.14.6.365.

Park, H., Jeong, S., Chung, G., 2016. Frequency Analysis of Future Maximum Fresh Snow Depth using Multiple Regression Model with Interaction. J. Korean Soc. Hazard Mitigation 16 (2), 369–376. https://doi.org/10.9798/kosham.2016.16.2.369.

Park, S., Kim, M., Im, J., 2021. Estimation of Ground-level PM10 and PM2.5 Concentrations Using Boosting-based Machine Learning from Satellite and Numerical Weather Prediction Data. Korean J. Remote Sensing 37 (2), 321–335.

Pratama, I., Permanasari, A.E., Ardiyanto, I., Indrayani, R., 2016. A review of missing values handling methods on time-series data. In: 2016 International Conference on Information Technology Systems and Innovation. https://doi.org/10.1109/ICITSI.2016.7858189.

Press, A., 2018. Waves of Winter Storms Kill at Least 16 in Europe. The Weather Channel. https://weather.com/photos/news/2018-02-27-beast-from-the-east-snow-photos.

Reichler, T., Kim, J., 2008. How well do coupled models simulate today's climate? Bull. Am. Meteorol. Soc. 89 (3), 303–311. https://doi.org/10.1175/BAMS-89-3-303.

Riahi, K., Rao, S., Krey, V., Cho, C., Chirkov, V., Fischer, G., Kindermann, G., Nakicenovic, N., Rafaj, P., 2011. RCP 8.5-A scenario of comparatively high greenhouse gas emissions. Clim. Change 109 (1), 33–57. https://doi.org/10.1007/s10584-011-0149-y.

Ruela, R., Sousa, M.C., DeCastro, M., Dias, J.M., 2020. Global and regional evolution of sea surface temperature under climate change. Global Planet. Change 190 (103190). https://doi.org/10.1016/j.gloplacha.2020.103190.

Shafapourtehrany, M., Yariyan, P., Özener, H., Pradhan, B., Shabani, F., 2022. Evaluating the application of K-mean clustering in Earthquake vulnerability mapping of Istanbul, Turkey. Int. J. Disaster Risk Reduct. 79 (January) https://doi.org/10.1016/j.ijdrr.2022.103154.

Statistics Korea, 2022. [Synoptic weather] Annual/monthly statistics for each station. Statistics Korea. https://kostat.go.kr/.

Tabari, H., Marofi, S., Abyaneh, H.Z., Sharifi, M.R., 2010. Comparison of artificial neural network and combined models in estimating spatial distribution of snow depth and snow water equivalent in Samsami basin of Iran. Neural Comput. & Applic. 19 (4), 625–635. https://doi.org/10.1007/s00521-009-0320-9.

United Press International, 2019. Major winter storm kills 4 in Germany and Austria. Gephardtdaily. https://gephardtdaily.com/national-international/major-winter-storm-kills-4-in-germany-and-austria/.

Vabalas, A., Gowen, E., Poliakoff, E., Casson, A.J., 2019. Machine learning algorithm validation with a limited sample size. PLoS One 14 (11), 1–20. https://doi.org/10.1371/journal.pone.0224365.

Vafakhah, M., Nasiri Khiavi, A., Janizadeh, S., Ganjkhanlo, H., 2022. Evaluating different machine learning algorithms for snow water equivalent prediction. Earth Sci. Inf. https://doi.org/10.1007/s12145-022-00846-z.

Vapnik, V., Golowich, S.E., Smola, A., 1997. Support vector method for function approximation, regression estimation, and signal processing. Adv. Neural Inf. Proces. Syst. 281–287.

Welle, D., 2020. Japan: Heavy snowfall leaves thousands stranded. Deutsche Welle. https://www.dw.com/en/japan-heavy-snowfall-leaves-thousands-stranded/a-55969344.

Zhang, X., Li, X., Li, L., Zhang, S., Qin, Q., 2019. Environmental factors influencing snowfall and snowfall prediction in the Tianshan Mountains, Northwest China. J. Arid. Land 11 (1), 15–28. https://doi.org/10.1007/s40333-018-0110-2.

Zhu, X., Chu, J., Wang, K., Wu, S., Yan, W., Chiam, K., 2021. Prediction of rockhead using a hybrid N-XGBoost machine learning framework. J. Rock Mech. Geotech. Eng. 13 (6), 1231–1245. https://doi.org/10.1016/j.jrmge.2021.06.012.