Manuscript Number: NHESS-2022-118

We sincerely thank the Reviewer 2 for the time in effort on reviewing our manuscript with many insightful comments. We believe that we have addressed each of the comments carefully and properly, while improving the quality of paper significantly. We hope that the changes listed below are acceptable for publication. In addition, we have made significant changes in all the relevant main text body, which could be aligned well with our responses to the comments in the revised manuscript. The changes made in the revised manuscript are highlighted in blue to facilitate their identification.

General Comment: The authors compared four machine learning algorithms and derived an optimal model to predict heavy snowfall; however, the method and analysis process used are general and lack novelty.

Response: First, thank you for positively viewing our research ideas, with very insightful comments listed below.

The snowfall prediction model of this study used four machine learning algorithms (MLR, SVM, RF, XGB) to learn the weather factors and geographic factors collected through the 102 ASOSs. This model will be used for GIS-based predicted snowfall distribution according to future RCP climate change scenarios. The four machine learning algorithms were selected as regression models for the purpose of prediction, not for identifying the cause of the heavy snowfall, and grid search and k-fold cross-validation techniques were used to improve learning performance. In addition, it is meaningful that geographic factors (latitude, longitude, altitude) as input data that were not considered in the study of snowfall prediction in Korea.

We hope that this is acceptable and reasonable. Thank you.

Comment 1

In Chapter 1. Introduction, the authors only mentioned previous research, but did not describe this research's superiority such as contributions and novelty of the research.

Response: The key of this study is to secure the accuracy of heavy snowfall predicting for the vulnerability assessment to future heavy snow disasters. Therefore, in this study, four algorithms with good performance were selected and used through analysis of previous studies. In addition, the most used input factors were synthesized and classified into meteorological factors and geographic factors and used as input data. South Korea was designated as the case study area, and observation data from an actual weather station was used, not remote sensing data, considering the spatiotemporal resolution. As mentioned in 'Chapter 4 Discussion and Conclusion' of the manuscript, the predictive model derived through this study can be used for GIS-based predicted

snowfall distribution according to future RCP climate change scenarios.

We have re-written the Chapter 1 to clearly present this research's superiority such as contributions and novelty, as follows:

Lines 95-110

Prediction of snowfall in previous studies is a non-linear process in which precipitation, temperature, relative humidity, and geographic variables are variously related. Various machine learning techniques that can take this non-linear process have shown good results in predicting the amount of snowfall. This is because nonlinear activation functions (Sigmoid and Tanh) are used in machine learning algorithms to explain the nonlinear relationship between input factors(Tabari et al., 2010). However, the prediction results may vary greatly depending on the regional research scope and the characteristics of the input variable data used for model development. In this study, South Korea as the study area, input variables not applied in existing domestic studies from previous studies were synthesized and heavy snowfall prediction was performed using an excellent machine learning algorithm. In addition, the predictive model derived through this study can be used for GIS-based predicted snowfall distribution according to future RCP climate change scenarios and heavy snowfall disaster management.

Comment 2

In Chapter 2. Materials and methods, it should have mentioned why the four machine learning algorithms (MLR, SVR, RFR, and XGB) were chosen among the other various machine learning algorithms.

Response: This study is to evaluate the applicability of various regression machine learning methods for predicting heavy snowfall in South Korea. Many machine learning techniques for regression and prediction have been proposed in the literature. Among the machine learning algorithms used in previous studies, models with good regression results were selected and applied to this study. In this paper, we focused on three leading approaches that have proven to be most effective in a wide variety of applications, i.e., support vector regression (SVR), random forest regressor (RFR) and extreme gradient boosting(XGB).

SVR most widely used in various prediction fields and have also been applied to prediction of snowfall and frost occurrence. RFR shows the best performance in similar previous studies and has the advantage of being able to check the importance scores of various variables. XGB is more recent, and its application for prediction is emerging. In the case of MLR, as mentioned in the manuscript, it was selected for comparison with the other three regression models.

The table below shows machine learning algorithms and compares the performance of

each model through the coefficient of determination used in previous studies. Random Forest and SVM show high performance.

References	Machine learning algorithms	Performance(R^2)
Tabari et al., 2013	MLR, ANN, NNGA	NNGA
Park et al., 2014	ANN, MLR	ANN> MLR
Ahn et al., 2015	SR	-
Park et al., 2016	MLR	-
Oh et al., 2017	MLR	-
Hamidi et al., 2018	RF, SVM, MARS	RF>SVM>MARS
Ki et al., 2018	RF, GLM, NNET	RF>GLM≈NNET
Zhang et al., 2019	RF, MLR, RNN	RF>RNN>MLR
Hu et al., 2021	ANN, SVM, RF	RF>SVM>ANN

We mentioned the Chapter 1 to explain why the four machine learning algorithms (MLR, SVR, RFR, and XGB) were chosen among the other various machine learning algorithms in this research, as follows:

Line 95-98

Among the machine learning algorithms used in previous studies, models with good regression results were selected and applied to this study. In this paper, we focused on three leading approaches that have proven to be most effective in a wide variety of applications, i.e., support vector regression (SVR), random forest regressor(RFR) and extreme gradient boosting(XGB) (Bedi et al., 2020).

Comment 3

In Chapter 4. Discussion and Conclusions, specific discussions on applicability of the optimal model, which is connected with RCP scenario and heavy snowfall disaster management, should be presented. Furthermore, there were a few sentences to confuse readers (for example, lines 269-272). The authors need to clarify those sentences in the manuscript and proofread the manuscript before submission.

Response: We are going to modify the ‘Chapter 4. Discussion and Conclusions’ to add

the specific explanation of RCP scenario and heavy snowfall disaster management.

Line 309-313

In the future, the proposed model can be used as an estimation model to obtain the distribution of the predicted daily snowfall in South Korea using the RCP climate change scenario data. Predicted daily snowfall data with RCP scenario can aid in establishing response and management strategies for heavy snowfall disasters in road facilities and transportation sectors by providing long-term snowfall prediction data until 2100.

Also, we fix the sentences (lines 269-272 / 274-277(revised version)) to clarify the results of the MLR and SVR models underestimating snowfall.

Line 278-281

The MLR and SVR models partially interpreted the variance in snowfall. In the case of field observation data, there is a lack of datasets for high snowfall and there are a lot of datasets for low snowfall. The imbalance of datasets was analyzed as a result of underestimating the MLR and SVR models(Park et al., 2021).

We have undergone proofreading twice for this thesis, but we will have it proofread once more before final submission.

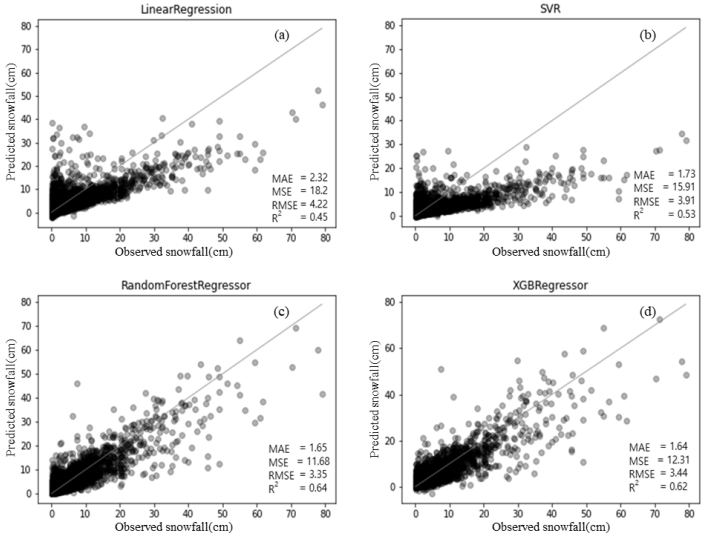
Comment 4

Furthermore, there are additional explanation of the a, b, c, and d of figure 5 should be added to make clear to the readers.

Response: We added the explanation of (a) MLR, (b) SVR, (c) RFR, and (d) XGB.

Line 273-275

The predicted snowfall values obtained using the MLR, SVR, RFR, and XGB models and the observed snowfall values are shown in Figures 5 through scatter plots of (a) MLR, (b) SVR, (c) RFR, and (d) XGB



Comment 5

Please add more relevant literature review with up-to-date.

Response: We added 7 more relevant literate review.

1. 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.
2. Associated Press. (2018). *Waves of Winter Storms Kill at Least 16 in Europe*. The Weather Channel.
3. Bedi, S., Samal, A., Ray, C., & Snow, D. (2020). Comparative evaluation of machine learning models for groundwater quality assessment.
4. Deutsche Welle. (2020). *Japan: Heavy snowfall leaves thousands stranded*. Deutsche Welle.
5. France24. (2021). *Huge snowstorm blankets US East Coast, halting travel and vaccinations*.
6. Mallick, J., Alqadhi, S., Talukdar, S., Alsubih, M., Ahmed, M., Khan, R. A., Kahla, N. Ben, & 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.
7. United Press International. (2019). *Major winter storm kills 4 in Germany and Austria*. Gephardtdaily.