AI modeling for characterization of paddy rice yields under extreme weather conditions using remote sensing and geospatial data

Jihye Ahn[1] and *Yangwon* Lee[2#]

[1]*Pukyong National University, Research Institute for Geomatics, 45 Yongso-ro, Nam-gu, Busan, South Korea*
[2]*Pukyong National University, Department of Spatial Information Engineering,*
45 Yongso-ro, Nam-gu, Busan, South Korea
[#]*Corresponding author: modconfi@pknu.ac.kr*

ABSTRACT

Climate change, such as increase in CO_2 levels and rising temperatures, can have a significant impact on paddy rice production and increase the uncertainty of yield forecasts. This study aims to employ AI modeling for forecasting paddy rice yield and present the findings of a quantitative analysis to determine its ability to generate stable forecasts under extreme weather conditions, such as heatwaves, low temperatures, and heavy rainfall. Vegetation growth indices from the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite product were utilized. These indices include the Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Leaf Area Index (LAI), Fraction of Absorbed Photosynthetically Active Radiation (FPAR), and Near-Infrared Reflectance of vegetation (NIRv). Meteorological variables such as downward solar radiation flux, daily temperature difference, precipitation, relative humidity, and temperature were also used. Over 23 years of experimentation (2000-2022), yields under extreme weather conditions did not exhibit a significant difference from the normal period, with a Mean Absolute Error (MAE) ranging from 0.30 to 0.33 ton/ha, representing a 4-5% error of the average yield. This study presents an AI modeling methodology that enables stable predictions of paddy rice yields, even under extreme weather conditions. Future work should focus on refining input data and optimizing the model by analyzing cases of extreme weather.

Keywords: paddy rice yields; AI modeling; extreme weather.

1. Introduction

Global temperatures are expected to continue rising in the 21st century, and increased carbon dioxide concentrations and warmer temperatures could affect paddy rice growth. Despite adaptation efforts such as adjusting planting timing and improving crop varieties, extreme climate changes may surpass the biological capacity of paddy rice. Despite this uncertainty, there are only a few studies on paddy rice yield forecasting using geospatial data. Hong et al. (2012) and Na et al. (2020) employed linear modeling of paddy rice yield using Moderate Resolution Imaging Spectroradiometer (MODIS) satellite imagery and meteorological data. Subsequently, with the introduction of Artificial Intelligence (AI) modeling, Ma et al. (2017) utilized AutoEncoder, and Cho et al. (2021) enhanced the prediction accuracy through Random Forest (RF) modeling. Although previous studies conceptually expect to maintain predictive power under extreme weather conditions, there is a lack of quantitative analysis on whether paddy rice yield models in Korea exhibit stable predictive power under such conditions where nonlinearity must be considered, except for international instances (Schierhorn et al., 2021).

Therefore, this study employed AI-based paddy rice yield modeling using remote sensing and geospatial data, and analyzed whether the model retains similar predictive performance as during normal conditions under extreme weather conditions such as heatwaves, low temperatures, and heavy rainfall. The study period is 23 years from 2000 to 2022, and experiments were conducted at the city and county level for paddy rice yield.

2. Study Area

The study area is located in Korea between 33.5˚N to 38.5˚N and 125.0˚E to 130.0˚E. Fig. 1 shows a map of East Asia including Korean Peninsula, where the blue rectangle indicates the study area. For the paddy rice yield modeling, 129 cities and counties located in this study area were selected based on the data provided by the Korean Statistical Information Service (KOSIS), which served as the true data.

Figure 1. A map of East Asia including Korean Peninsula, where the blue rectangle indicates the study area (33.5–38.5˚N and 125.0–130.0˚E).

3. Data and Methods

3.1. Data

Meteorological data were collected from the Climatic Research Unit-JMA Reanalysis (CRU-JRA) for the months of July, August, and September from 2000 to 2022. CRU-JRA ver. 2.4 is an ensemble of the UK Climatic Research Unit (CRU) and the Japanese Meteorological Agency (JMA) Japanese Reanalysis (JRA), providing global 0.5° gridded data at 6-hour intervals. Downward solar radiation flux, daily temperature difference, precipitation, relative humidity, and temperature variables are aggregated by month and municipality. Relative humidity is calculated using air temperature, specific humidity, and barometric pressure.

In addition, Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Leaf Area Index (LAI), Fraction of Absorbed Photosynthetically Active Radiation (FPAR), and Near-Infrared Reflectance of vegetation (NIRv), which are highly indicative of vegetation growth among MODIS satellite products, were aggregated monthly at the city and county level.

Meteorological data and satellite images in raster format were allocated to each city and county polygon through a zonal operation, taking into account the weighted average based on the pixel area overlapping with the polygon. NDVI is calculated from the red and near-infrared light reflected by vegetation. Healthy vegetation absorbs mostly red light and reflect near-infrared light. As NDVI may not sensitive to high values in very dominant vegetation, EVI is employed as a vegetation index to address this limitation by factoring in canopy background effects. LAI quantifies the leaf area within a specified area, while FPAR represents the fraction of solar radiation absorbed by vegetation leaves for photosynthesis.

To conduct paddy rice yield modeling by integrating meteorological data and satellite images, we utilized the true data as the paddy rice yield provided by the KOSIS. Paddy rice yield data per year at the city and county level were collected from 2000 to 2023. We collected data on paddy rice production (in tons) and cultivated area (in hectares), and converted them into yield per unit area (tons/ha), resulting in the construction of 2,967 dataset. Table 1 summarizes the data used in this study.

Table 1. Summary of the data used in study

Data	Variables	Use	Spatial Resolution	Temporal Resolution
Yield	Paddy rice yield [ton/ha]	Target variable	Point	Yearly
CRU-JRA	Solar radiation, daily temperature difference, precipitation, relative humidity, temperature	Explanatory variable	0.5˚	6-hour
MODIS	NDVI, EVI, LAI, FPAR, NIRv	Explanatory variable	1km	Monthly

3.2. Analysis Methods

The paddy rice yield model was constructed using Automated Machine Learning (AutoML) technique. AutoML enhances accuracy by conducting hyperparameter optimization for various AI models such as RF, Gradient Boosting Machine (GBM), and Deep Neural Network (DNN), and subsequently ensembling the top N models based on the leader board. In this study, AutoML modeling was executed using the AutoKeras library, which supports high-quality model development through efficient algorithm tuning (Fig. 2).

The performance evaluation of the AutoML paddy rice yield model was conducted using 5-fold cross-validation. 5-fold cross-validation entails shuffling the entire dataset and partitioning it into five groups. One group is designated as the validation set, while the model is trained on the data from the remaining four groups. Subsequently, blind evaluation is performed on the validation set. This process is repeated for a total of five rounds to obtain accuracy statistics for different combinations, which are then aggregated for overall performance evaluation. Accuracy statistics were obtained using Mean Bias Error (MBE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Correlation Coefficient (CC).

Following the cross-validation of the AutoML model for a total of 2,967 cases in 129 cities and counties over 23 years, accuracy statistics under extreme weather conditions such as heatwave, low temperature, and heavy rainfall were separately calculated. These statistics were then compared to the overall statistics to evaluate the reliability of the model's accuracy under extreme weather conditions.

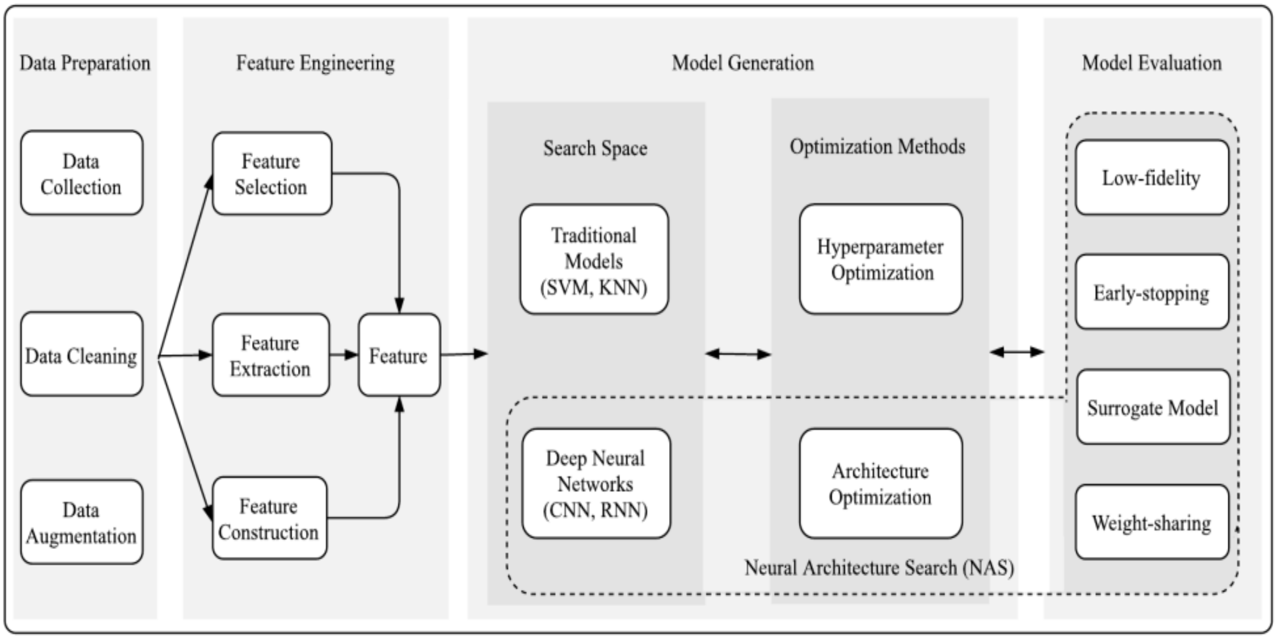

Figure 2. Overview of AutoML pipeline covering with 4 sections (data preparation, feature engineering, model generation, and model evaluation) (He et al., 2021).

4. Results and Discussion

The AutoML cross-validation of a total of 2,967 cases from 129 cities and counties over 23 years showed an MAE of approximately 0.312 ton/ha and a CC of around 0.725. These results confirm that the model can closely simulate real-world paddy rice yields. During heatwaves, low temperatures, and heavy rainfall, the MAEs ranged from 0.302 to 0.326 ton/ha, which is nearly identical to the MAE of the full dataset. Furthermore, these MAEs represent only 5% of the average yield, demonstrating that the model's prediction of paddy rice yield under extreme weather conditions is nearly as accurate as under normal conditions.

Paddy rice grows in water-filled paddies, but it can also be damaged by water, especially due to heavy rains or typhoons, can cause rice lodging, leading to significant yield losses. Moreover, heavy rains during September, the peak of the rice growing season, can keep the paddy rice wet, promoting damage by disease and pest and potentially reducing yields. Abnormally low temperatures during the planting season in May and June can impede the development of crucial plant organs such as leaves, stems, and roots, as well as the differentiation of young ears, consequently affecting overall growth and yield. Severe heatwaves can hinder grain filling and reduce the size of rice grains. If these heatwaves persist into September, it may delay the harvest, further impacting yields.

As shown in Table 2 and Fig. 3, the estimation error for paddy rice yield across all 2,967 cases is 0.312 ton/ha. However, for 258 cases of heatwaves, 115 cases of low temperatures, and 248 cases of heavy rainfall, the estimation error ranges from 0.302 to 0.326 ton/ha. These results suggest that the AutoML paddy rice yield model maintains similar accuracy under extreme weather conditions as it does under normal conditions. With an average paddy rice yield of 6.5 tons/ha over 23 years, this error is relatively small, approximately 5% of the yield. In particular, when analyzing the error during heavy rainfall, there was a notable tendency for overestimation in years with medium to large typhoon events accompanied by strong winds and heavy rainfall. Consequently, we believe that enhancing the model to incorporate the impact of typhoons, particularly after August and September when typhoon mainly occur, would improve model accuracy.

5. Conclusions

While previous studies have conceptually explored the impact of extreme weather on yield, this study establishes an AutoML paddy rice yield model that maintains stable predictive power under extreme weather conditions such as heatwaves, low temperatures, and heavy rainfall in Korea. Through experiments conducted on paddy rice yield at 129 city and county level for 23 years from 2000 to 2022, cross-validation of the entire dataset showed an MAE of approximately 0.312 ton/ha. This confirms that the AutoML model produces predictions similar to the actual yield. Additionally, during heatwaves, low temperatures, and heavy rainfall, the MAE ranged from approximately 0.302 to 0.326 ton/ha, which accounts for about 5% of the average yield. These results indicate that the predictive power of the model is well-maintained even under extreme weather conditions. It is expected that the methodology presented in this study can be utilized for paddy yield characterization in the future by refining the input data and model.

Table 2. Accuracy statistics of paddy rice yield model

Type	N	MBE	MAE	RMSE	CC
All	2967	-0.144	0.312	0.395	0.725
Heatwaves	258	0.209	0.326	0.402	0.649
Low temperatures	115	0.178	0.302	0.377	0.678
Heavy rainfall	248	0.084	0.317	0.421	0.612

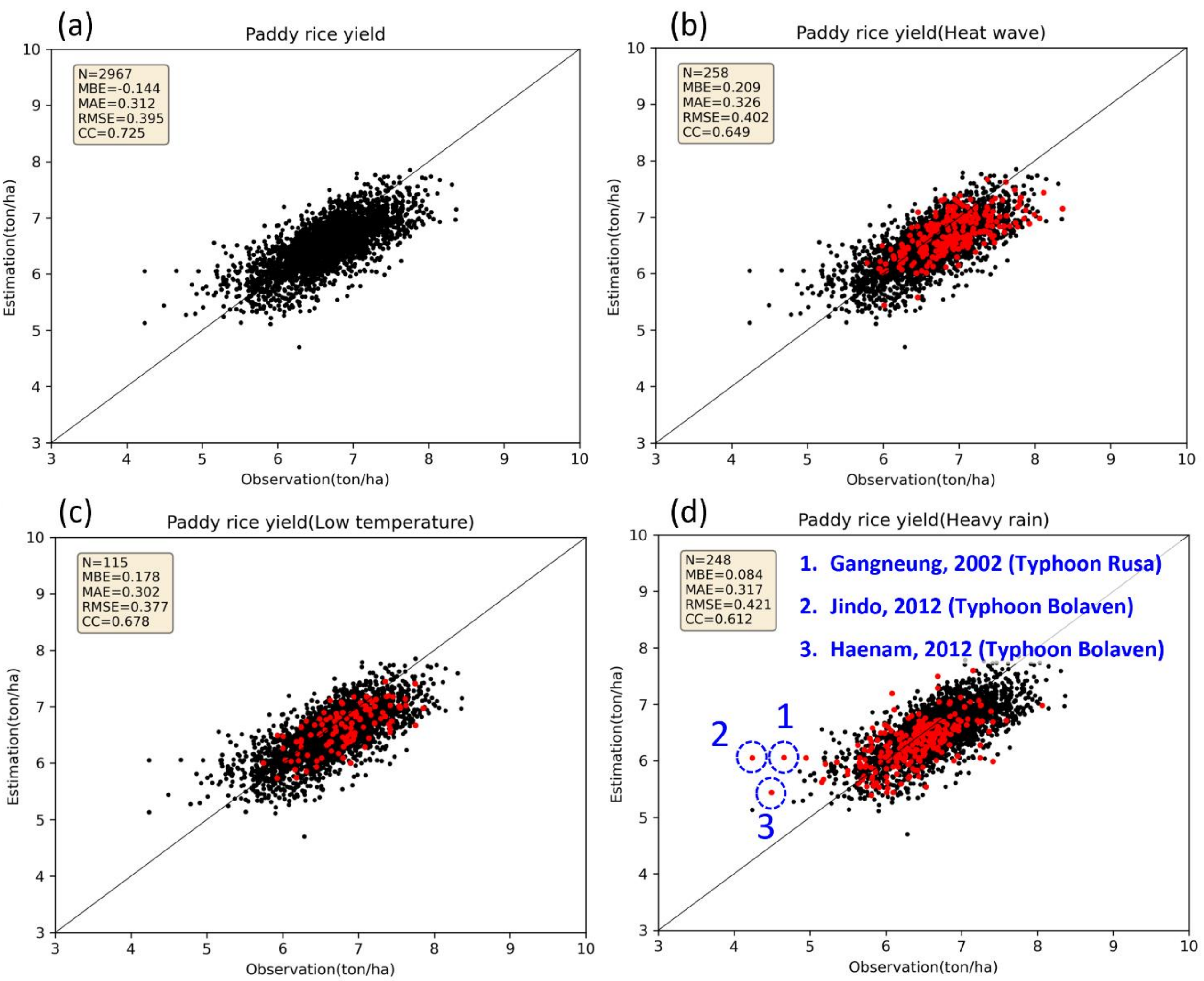

Figure 3. Scatter plots for the observed and estimated paddy rice yield for (a) entire cases, (b) heatwaves, (c) low temperature, and (d) heavy rainfall cases. The red dots indicate extreme weather cases.

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. RS-2023-00276052).

References

Cho, S., Youn, Y., Kim, S., Jeong, Y., Kim, G., Kang, J., Kim, K., Cho, J., and Lee, Y. "A Comparative Evaluation of Multiple Meteorological Datasets for the Rice Yield Prediction at the County Level in South Korea", Korean J. Remote Sens, 37(2), pp. 337-357, 2021. https://doi.org/10.7780/kjrs.2021.37.2.12

He, X., Zhao, K., and Chu, X. "AutoML: a survey of the state-of-the-art", Knowledge-Based Systems, 212, 106622, 2021. https://doi.org/10.1016/j.knosys.2020.106622

Hong, S., Hur, J., Ahn, J., Lee, J., Min, B., Lee, C., Kim, Y., Lee, K., Kim, S., Kim, G., and Shim, K. "Estimating Rice Yield Using MODIS NDVI and Meteorological Data in Korea", Korean J. Remote Sens, 28(5), pp. 509-520, 2012. http://dx.doi.org/10.7780/kjrs.2012.28.5.4

Ma, J., Lee, K., Choi, K., and Heo, J. "Rice Yield Estimation of South Korea from Year 2003-2016 Using Stacked Sparse AutoEncoder", Korean J. Remote Sens, 33(5), pp. 631-640, 2017. https://doi.org/10.7780/kjrs.2017.33.5.2.3

Na, S., Ahn, H., Park, C., Hong, S., So, K., and Lee, K. "Effect of the Application of Temporal Mask Map on the Relationship between NDVI and Rice Yield", Korean J. Remote Sens, 36(5), pp. 725-733, 2020. https://doi.org/10.7780/kjrs.2020.36.5.1.6

Schierhorn, F., Hofmann, M., Gagalyuk, T., Ostapchuk, I., and Müller, D. "Machine learning reveals complex effects of climatic means and weather extremes on wheat yields during different plant developmental stages", J. Clim. Change, 169(39). 2021. https://doi.org/10.1007/s10584-021-03272-0