Note Open Access

Optimizing Artificial Neural Network-Based Models to Predict Rice Blast Epidemics in Korea

Kyung-Tae Lee, Juhyeong Han, and Kwang-Hyung Kim 🆔 *

Department of Agricultural Biotechnology, Seoul National University, Seoul 08826, Korea

To predict rice blast, many machine learning methods have been proposed. As the quality and quantity of input data are essential for machine learning techniques, this study develops three artificial neural network (ANN)-based rice blast prediction models by combining two ANN models, the feed-forward neural network (FFNN) and long short-term memory (LSTM), with diverse input datasets, and compares their performance. The Blast_Weather_FFNN model had the highest recall score (66.3%) for rice blast prediction. This model requires two types of input data: blast occurrence data for the last 3 years and weather data (daily maximum temperature, relative humidity, and precipitation) between January and July of the prediction year. This study showed that the performance of an ANN-based disease prediction model was improved by applying suitable machine learning techniques together with the optimization of hyperparameter tuning involving input data. Moreover, we highlight the importance of the systematic collection of long-term disease data.

Keywords : artificial intelligence, artificial neural network, machine learning, rice blast

*Corresponding author.
Phone) +82-2-880-4672
E-mail) sospicy77@snu.ac.kr
ORCID
Kwang-Hyung Kim
https://orcid.org/0000-0003-3265-5899

Articles can be freely viewed online at www.ppjonline.org.

Rice blast disease, caused by *Pyricularia oryzae* Cavara, is one of the major constraints to rice production, causing significant yield losses worldwide (Katsantonis et al., 2017; Wang et al., 2015). Many studies have been conducted to understand the primary elements of the rice blast epidemic and predict the occurrence of this disease (Chung et al., 2020; Jeon, 2019; Kim et al., 2015). It is well known that low temperature, high humidity, and excessive use of nitrogen fertilizers promote the occurrence of rice blast disease. An integrated disease management (IDM) program based on a proper understanding of rice blast epidemiology is unequivocally the most effective and efficient method for managing rice blast disease in the long term.

IDM is designed to minimize the impact of plant diseases below the level that could cause significant economic damage by deploying all available methods that are optimal and realistic in the context of the rice-growing environments and population dynamics of pathogens. To properly apply various management techniques, it is essential to develop methods that can predict the risk of temporal and spatial occurrences using a plant disease prediction model. In other words, knowing when and to what extent plant diseases occur is crucial for effective plant disease management. This helps determine the optimal timing and order of applying proper disease management methods. In particular, owing to the acceleration of climate change and increasing occurrence of abnormal climate conditions, it is becoming more difficult to predict plant diseases. Therefore, various innovative methods have been proposed to cope with this situation (Juroszek and von Tiedemann, 2011).

Computer modeling has been developed and employed to predict plant disease epidemics using weather, environmental, and agronomic data. Traditional plant disease prediction models find statistical, empirical, and/or mechanical relationships between these data and the occurrence of plant diseases, and simulate the key infection process based on them. These types of prediction models with ei-

ther observed or predicted input data provide information on when and to what extent plant diseases would occur to determine optimal management measures. In particular, when the overall production cost increases due to unnecessary control activities or the control effect decreases owing to ill-timed control, an IDM based on accurate prediction is urgently required. These prediction models are utilized as an important component of IDM along with various plant disease management methods.

Recently, as the amount of agricultural data has exponentially increased, many attempts have been made to model plant diseases using machine learning (Kim and Lee, 2020). Different algorithms, such as support vector machines (SVMs), artificial neural networks (ANNs), and random forest, have been used to create plant disease prediction models based on meteorological variables, including maximum and minimum temperatures, humidity, rainfall, and wind speed. Fenu and Malloci (2019) trained two different models using the ANN and SVM techniques to predict late blight in potatoes. They used 4-year meteorological data (hourly temperature, humidity, rainfall, wind speed, and solar radiation) as input, and classified corresponding disease occurrence data in Southern Sardinia into three risk levels. The SVM model showed better performance for low- and high-risk levels, whereas the ANN model outperformed at the medium-risk level. Interestingly, the ANN model incorrectly classified 40 out of 49 high-risk cases as medium-risk cases, although its overall accuracy was 96%. This was due to the imbalance of the training dataset, where most data were classified as low risk; thus, the overall accuracy was determined by the classification performance for the major class without profoundly considering the minor class. Bhatia et al. (2020) adopted the Extreme Learning Machine algorithm and found proper resampling techniques can solve the problem of highly imbalanced dataset in plant disease occurrence.

Similarly, researchers have attempted to predict rice blast disease using machine learning. Kaundal et al. (2006) developed and compared rice blast prediction models using multiple regression, SVMs, and two ANN algorithms. Malicdem and Fernandez (2015) developed rice blast prediction models using the feed-forward neural network (FFNN) model, the simplest ANN structure that unidirectionally connects the input and output layers without a loop. Recent studies have used the long-short term memory (LSTM) structure to predict rice blast disease occurrence (Kim et al., 2018; Nettleton et al., 2019). LSTM was developed by Hochreiter and Schmidhuber (1996), and is considered to be a high-performance model among recurrent neural networks (RNNs). This is because LSTM solved the prob-

lem of long-term dependency in RNNs by using a variable called the cell state to selectively store information through the input, forget, and output gates (Hochreiter and Schmidhuber, 1996). Using LSTM, Kim et al. (2018) developed region-specific models for the prediction of rice blast in Korea. Three-year rice blast occurrence and weather data (average temperature, relative humidity, and sunshine duration) were used as input data to train the model for rice blast prediction over the next few years. In another study, two machine learning-based rice blast prediction models (M5Rules and LSTM) showed comparable performance to process-based models (Yoshino and WARM) (Nettleton et al., 2019).

Developing a plant disease prediction model using machine learning has many hurdles such as the low quality and quantity of disease occurrence data for training and validation, relatively low performance of the models developed in previous studies, and lack of information on optimal machine learning techniques for plant disease prediction. Although the LSTM sits in the center of interest for plant disease prediction these days, it is well known that LSTM generally performs well for sequences of up to 250-500 timesteps to solve long-term dependency problems (Chemali et al., 2017), indicating it may not outperform the FFNN for rice blast epidemics with a sequence of much shorter timesteps. Therefore, the objectives of this study are to develop ANN-based rice blast prediction models using the limited quality and quantity of rice blast occurrence data available in Korea, and to compare the performance of the FFNN and LSTM models after optimizing hyperparameters of both models. In previous studies, historical rice blast occurrence and weather variables were used as inputs for model training without considering the performance variation depending on hyperparameters (Kim et al., 2018; Nettleton et al., 2019). In machine learning, hyperparameters determine the structure of the learning model, such as the learning rate, number of nodes and layers, and batch size; thus, optimizing them is crucial for model performance (Probst et al., 2019). In this study, the number of observed years for rice blast occurrence, range of observed months, number of timesteps for weather variables, and type and combination of weather variables were included as hyperparameters to examine the variance in model performance.

In the study, historical rice blast occurrence data and weather observation data were used as input data to train rice blast prediction models (Fig. 1). Historical rice blast occurrence data were obtained from the National Crop Pest Management System (NCPMS; https://ncpms.rda.go.kr) of the Rural Development Administration (RDA) of Korea.
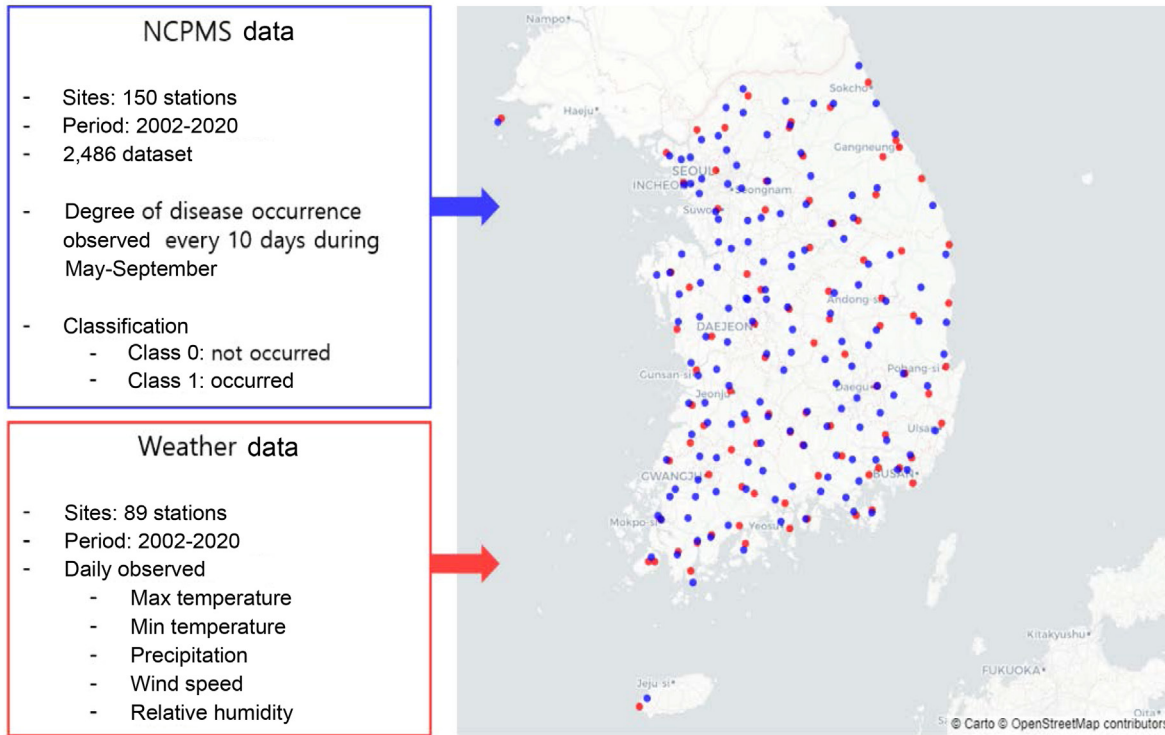
Fig. 1. Distribution map of National Crop Pest Management System (NCPMS) data and weather observation data used in the study. Blue dots indicate the NCPMS data of rice blast occurrence and red dots indicate the weather station sites for weather observation data.

We collected 2,486 occurrence data from 150 RDA rice monitoring plots for 19 years (2002-2020). The rice blast intensity was recorded by measuring the infected leaf area ratio at 10-day intervals from the 20 May to the 20 September (30 September from 2006 to 2008). Excluding 13 missing data points, 656 data points with an infected leaf area ratio more than 0.2% were classified as class 1 (representing blast occurrence), whereas 1,817 data points with a ratio equal to or less than 0.2% were classified as class 0 (representing no blast occurrence).

For weather observation data, we obtained the daily maximum air temperature (°C), minimum air temperature (°C), precipitation (mm), relative humidity (%), and wind speed (m/s) data from 89 weather stations of the Korea Meteorological Administration from 2002 to 2020 (Fig. 1). To avoid biased learning toward specific data due to the difference in scale between input data (Sola and Sevilla, 1997), weather data were min-max normalized when creating the training datasets. To match the weather observation data with the rice blast occurrence data, the nearest weather stations were selected using the haversine formula (Eq. 1). The haversine formula is used to determine the distance between two points using their coordinates (longitude and latitude), assuming the Earth as a sphere (Yang et al.,

2019), as follows:

$$d = 2r \cdot \arcsin$$
$$\left(\sqrt{\sin^2\left(\frac{(\phi_2 - \phi_1)}{2}\right) + \cos(\phi_1)\cos(\phi_2)\sin^2\left(\frac{(\lambda_2 - \lambda_1)}{2}\right)} \right)$$
(1)

, where d is the distance between sites 1 and 2, r is the radius of the earth, ϕ_1 is the latitude of site 1, ϕ_2 is the latitude of site 2, λ_1 is the longitude of site 1, and λ_2 is the longitude of site 2. Among 2,473 data points, seven outliers with distances of more than 36 km between the rice monitoring plot and weather station were eliminated from the sample data.

The models were developed in the following order. First, we constructed a Blast_FFNN model that uses only historical rice blast occurrence data as input for training and conducted hyperparameter tuning for model optimization. Subsequently, we introduced an additional parallel layer using weather input data into the optimized Blast_FFNN model. Consequently, two new models, a Blast_Weather_FFNN model in which weather data go through FFNN layers and a Blast_Weather_LSTM model in which weather data go through LSTM layers, were created and then went

Table 1. Features of the hyperparameters for each model compared in this study

Model	Hyperparameters	Range
Blast_FFNN[a]	Year_size[b]	1-9
	Units[c]	2^2-2^5
	Activation function	Relu, sigmoid, tanh
Blast_Weather_FFNN	Units	2^3-2^7
	Activation function	Relu, sigmoid, tanh
	Months[d]	Jan-Jul to May-Jul
	Period[e]	2-30
	Weather_variables	Combinations of tmax, tmin, wspd, prec, rhum
Blast_Weather_LSTM	Units	2^2-2^6
	Activation function	Relu, sigmoid, tanh
	Months	Jan-Jul to May-Jul
	Period	2-30
	Weather_variables	Combinations of tmax, tmin, wspd, prec, rhum

FFNN, feed-forward neural network; LSTM, long short-term memory.
[a]As indicated by the name, Blast_FFNN uses only historical rice blast occurrence data, while the Blast_Weather_FFNN and Blast_Weather_LSTM models use both blast occurrence and weather data as inputs.
[b]The number of consecutive years of rice blast occurrence observed in the past.
[c]The number of nodes in each hidden layer.
[d]Selected months during which weather variables are used as input.
[e]The number of timesteps during the selected months.

through sequential hyperparameter tuning processes. The hyperparameters considered in the optimization process for each model are listed in Table 1.

In the Blast_Weather_FFNN and Blast_Weather_LSTM models, hyperparameter tuning first determines the number of nodes (units) and activation functions of the parallel layers. Second, the input data of months, period, and weather_variables, referring to the selected months to be included, the number of timesteps during the selected months, and combinations of weather variables, respectively, are determined based on the hyperparameter tuning. For example, in the case of period '4' with the weather_variables of 'tmax' and 'prec' for the months of 'June-July,' the 15-day (61 days for June-July divided by period 4) average values of daily maximum temperature and daily precipitation were used as input data. Selecting target months as input data allowed us to examine whether weather conditions until July (approximately up to 50-70 days after transplanting in most rice cultivation areas in South Korea) have a significant effect on model performance. Additionally, selecting an appropriate period was important for model performance. The period of few sections may dilute the characteristics of the weather conditions affecting rice blast occurrence during the period, whereas the period of too many sections may result in outlier conditions that misrepresent the favorable weather conditions for rice blast. The development

processes for the three models are illustrated in Fig. 2.

Since the number of class 0 (no occurrence) was approximately three times more than class 1 (occurrence), the training and test sets were split according to the same ratio using stratified k-fold cross-validation with k = 10. Additionally, as the rice blast occurrence data used in the study were significantly imbalanced, we increased the number of class 1 samples using the random oversampling method (Batista et al., 2004), which randomly replicates the minority class dataset to a size comparable to that of the major class dataset. Focal loss was used as the loss function (Lin et al., 2017), and Adam optimizer was used as the optimizer setting, with a learning rate of 10^{-3} (Kingma and Ba, 2014). An appropriate number of epochs (100) was determined in the preliminary test. The models were developed using TensorFlow version 2.6.0, an open-source machine-learning library developed by Google (Abadi et al., 2016).

Both accuracy and recall were used as measures to evaluate the performance of the proposed models and optimize the hyperparameters. The prediction results of the classifier were expressed in a confusion matrix, and four classes were defined: true-positive (TP) for correctly predicted class 1 data, false-positive (FP) for incorrectly predicted class 1 data, true-negative (TN) for correctly predicted class 0 data, and false-negative (FN) for incorrectly predicted class 0 data. Accuracy indicates how often the classifier is
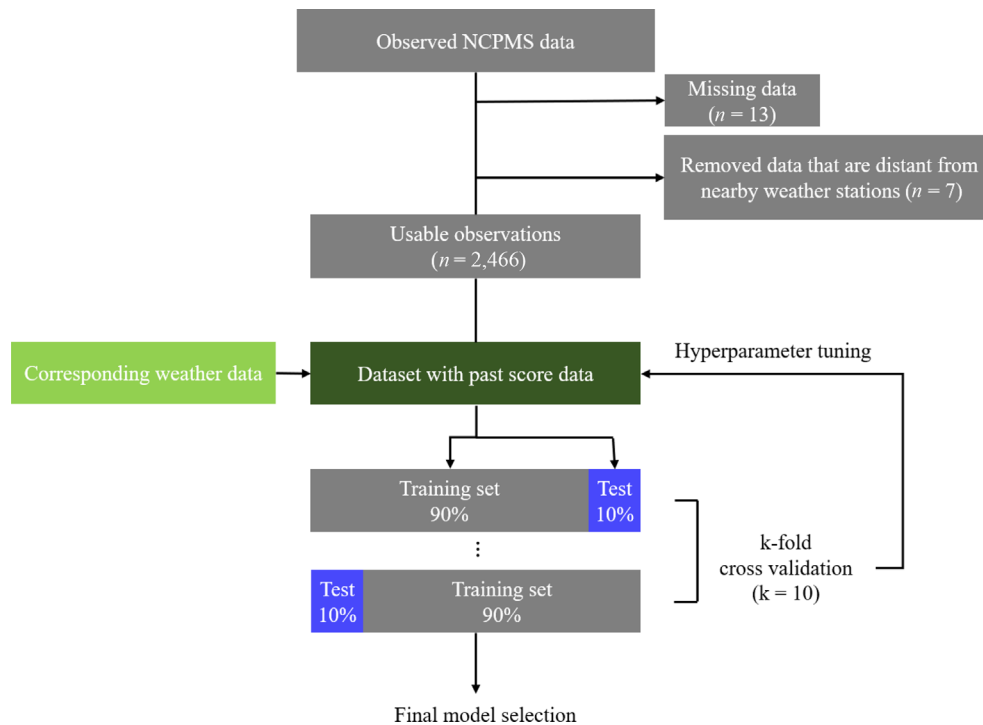
Fig. 2. A flowchart of the development of Blast_Weather_FFNN and Blast_Weather_LSTM models used in this study. Blast_FFNN does not include weather data indicated as the light green box in the figure. FFNN, feed-forward neural network; LSTM, long short-term memory; NCPMS, National Crop Pest Management System.

correct and is calculated as the ratio of (TP + TN)/(TP + FP + FN + TN). Accuracy has some constraints as a performance indicator because model training is biased toward the major class when an ANN model is trained using an imbalanced dataset. In this case, even if the minor class is not predicted well, the overall accuracy can be high, since the major class is generally predicted well. Therefore, we used the recall indicator to evaluate model performance. Recall, also called sensitivity or the TP rate, indicates how often the classifier predicts the actual disease occurrence, calculated as the ratio of TP/(TP + FN). When it comes to plant disease management, it is necessary to predict the actual disease occurrence to inform farmers to implement appropriate disease control measures to reduce potential yield loss. Therefore, we selected hyperparameters with the maximum recall values, as shown in Table 2. Validation of the performance of each model was repeated 10 times, then the average value was obtained.

Experiments for selecting hyperparameters for each of the three models (i.e., Blast_FFNN, Blast_Weather_FFNN, and Blast_Weather_LSTM) verified that the performance of the models depends on the hyperparameters (Table 2). Recall increased as the year_size of the Blast_FFNN model increased from 1 to 3, indicating that a record of rice blast occurrence in recent years helps predict future occurrences.

This is because the amount of initial inoculum of a year results from the overwintered inocula from the epidemics of the previous year(s). Moreover, local specific conditions, such as cultivars, climate, and soil, might influence the inherent disease proneness; thus, rice blast is more likely to occur where it normally occurs. Kim et al. (2018) used data from the past 3 years to examine the feasibility of predicting the occurrence of rice blast. Larger year_size over 3 years in our study reduced the number of training samples due to the presence of missing values. Additionally, unnecessarily old data beyond three years in the past had little impact on prediction and became disruptive to learning. After tuning the remaining hyperparameters with 16 nodes for the hidden layers using the tanh and sigmoid activation functions, the Blast_FFNN model showed the maximum performance with a recall of 55.99%.

Using weather data in addition to rice blast occurrence data as input, the Blast_Weather_FFNN and Blast_Weather_LSTM models showed higher performance with 66.33% and 64.50% recall scores, respectively, compared to the Blast_FFNN model. As shown in previous studies (Fenu and Malloci, 2021; Kim et al., 2018), weather data is necessary to improve the prediction performance of ANN-based rice blast models. We found that the months and periods, based on which the weather data are applied as input,

Table 2. Performances of rice blast prediction models after optimizing each hyperparameter

Model	Hyperparameters	Validation		Optimal values[a]
		Accuracy (%)	Recall (%)	
Blast_FFNN	Year_size	67.93-75.26	40.06-55.77	3
	Units	72.24-73.09	52.78-55.99	16, 16[b]
	Activation function	72.62-73.07	55.19-55.99	Tanh, sigmoid
Blast_Weather_FFNN	Units	69.32-70.68	56.35-65.35	8
	Activation function	70.48-71.31	64.52-65.92	Relu
	Months	70.62-70.87	63.23-65.08	Jan-Jul
	Period	70.39-71.41	62.03-66.41	20
	Weather_variables	69.86-71.59	63.47-66.33	tmax, prec, rhum
Blast_Weather_LSTM	Units	70.99-71.53	61.85-64.32	16
	Activation function	71.18-71.43	57.10-64.77	Relu
	Months	70.25-71.51	63.12-64.03	Mar-Jul
	Period	69.56-70.65	61.40-64.34	24
	Weather_variables	69.80-71.30	58.66-64.50	tmax, prec, rhum

FFNN, feed-forward neural network; LSTM, long short-term memory.
[a]Optimal values were selected to maximize the recall indicator.
[b]The Blast_FFNN model had two optimal values for each of the two hidden layers, while the Blast_Weather_FFNN and Blast_Weather_LSTM models had one hidden layer.

are important determining factors of model performance. The Blast_Weather_FFNN model showed the highest performance with the months between January and July and a period of 20 (approximately 10-day averages), while the Blast_Weather_LSTM model was best parameterized with the months between March and July and a period of 24 (approximately 6-day averages). Both models performed better when the weather data before planting were included as input, probably because it is related to the survival rate of the overwintered inocula from previous epidemics and thus determines the amount of initial inoculum of the prediction year. Among the five weather variables used in this study, both models showed the highest performance when using daily maximum temperature, precipitation, and relative humidity. The optimal numbers of nodes were 8 and 16 for Blast_Weather_FFNN and Blast_Weather_LSTM, respectively. In addition, the rectified linear unit (ReLU) activation function was selected for both models.

As a result, Blast_Weather_FFNN had higher performance (a recall score of 66.33%) than Blast_Weather_LSTM (a recall score of 64.50%). Since LSTM has a complex structure and more parameters compared to other ANN models, the relatively small quantity of NCPMS data for training likely affected the LSTM model to be underfitted. In addition, as there was no clear time-series pattern appearing in weather data for less than a year, there might not be an added value of using LSTM. Thus, we concluded

that it is more appropriate to use FFNN with a limited amount of data and a weak time-series pattern. Furthermore, considering that LSTM requires more computing resources and a longer training time owing to its complex structure and process, other ANN models might be a better starting point to consider.

In this study, long-term NCPMS data were used to develop an ANN-based rice blast prediction model. Government-led data collection in more than 80 locations across the country for two decades has resulted in quality data that are eligible for various machine learning-based studies. Considering that Fenu and Malloci (2021) used only 2-5 years of disease occurrence data, the new models developed in the current study using 19 years of data may show more robust performance in predicting interannual disease variation. Another promising fact is that the amount of data continues to increase with time, as data collection continues even at the moment. The quantity and quality of data are important in data-driven modeling research, such as machine learning-based studies. However, a chronic shortage of disease survey data in most countries has led to very few studies assessing the amount of data required for the reliable prediction of rice blast occurrence using machine-learning approaches. Examining this data requirement aspect requires a significant amount of data that exceeds what we used in the study. One way of securing sufficient data for such analyses is to generate artificial disease occurrence

data using process-based disease epidemiological models considering various environmental, agronomic, and host plant and pathogen factors as input.

Another hurdle in developing ANN-based disease prediction models is imbalanced datasets, which could result in biased model training toward the major class. In particular, for plant disease survey data collected from designated monitoring plots, observations of disease symptoms are very rare. To solve this problem, we used random oversampling and focal loss to increase the number of class 1 samples compared to class 0 samples for model training (Liu et al., 2007). This is because we emphasize reducing false-negative errors over false-positive errors to avoid severe yield losses resulting from no action over actual disease epidemics.

In Korea, unmanned aerial vehicles are commonly utilized in most rice paddies for collaborative disease control (Kim and Jung, 2020). Disease early warnings that use seasonal climate forecasts (SCFs) with a lead time of a few months support collaborative disease controls requiring at least a month before the decision-making of scheduling and preparing the control activities. The Blast_Weather_FFNN model, which showed the best performance in this study, requires weather data from January to July of the prediction year. Considering that rice leaf blast normally occurs between June and August, the model should rely on forecasted weather information from the SCFs. If the model can generate a rice blast alert sometime in May using the SCFs for June to August, the alert information can be applied for planning collaborative disease controls in South Korea. Therefore, follow-up studies should verify the performance of the rice blast prediction model using the SCFs. Unlike observational data, the reliability of the model prediction depends significantly on the predictability of the SCFs. This problem can be overcome by utilizing machine learning techniques, where SCFs are used as input variables to train the prediction model. Thereby, the inherent uncertainty of the SCFs is considered in hyperparameter tuning while training the model.

Conflicts of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viegas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y. and Zheng, X. 2016. TensorFlow: large-scale machine learning on heterogeneous distributed systems. Preprint at https://arxiv.org/abs/1603.04467.

Batista, G. E. A. P. A., Prati, R. C. and Monard, M. C. 2004. A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explor. Newsl.* 6:20-29.

Bhatia, A., Chug, A. and Prakash Singh, A. 2020. Application of extreme learning machine in plant disease prediction for highly imbalanced dataset. *J. Stat. Manage. Syst.* 23:1059-1068.

Chemali, E., Kollmeyer, P. J., Preindl, M., Ahmed, R. and Emadi, A. 2017. Long short-term memory networks for accurate state-of-charge estimation of Li-ion batteries. *IEEE Trans. Ind. Electron.* 65:6730-6739.

Chung, H., Kang, S., Lee, Y.-H. and Park, S.-Y. 2020. Expression patterns of transposable elements in *Magnaporthe oryzae* under diverse developmental and environmental conditions. *Res. Plant Dis.* 26:38-43.

Fenu, G. and Malloci, F. M. 2019. An application of machine learning technique in forecasting crop disease. In: *Proceedings of the 2019 3rd International Conference on Big Data Research*, pp. 76-82. Association for Computing Machinery, New York, NY, USA.

Fenu, G. and Malloci, F. M. 2021. Forecasting plant and crop disease: an explorative study on current algorithms. *Big Data Cogn. Comput.* 5:2.

Hochreiter, S. and Schmidhuber, J. 1996. LSTM can solve hard long time lag problems. In: *Proceedings of the 9th International Conference on Neural Information Processing Systems*, pp. 473-479. MIT Press, Cambridge, MA, USA.

Jeon, J. 2019. Phytobiome as a potential factor in nitrogen-induced susceptibility to the rice blast disease. *Res. Plant Dis.* 25:103-107.

Juroszek, P. and von Tiedemann, A. 2011. Potential strategies and future requirements for plant disease management under a changing climate. *Plant Pathol.* 60:100-112.

Katsantonis, D., Kadoglidou, K., Dramalis, C. and Puigdollers, P. 2017. Rice blast forecasting models and their practical value: a review. *Phytopathol. Mediterr.* 56:187-216.

Kaundal, R., Kapoor, A. S. and Raghava, G. P. S. 2006. Machine learning techniques in disease forecasting: a case study on rice blast prediction. *BMC Bioinformatics* 7:485.

Kim, K.-H., Cho, J., Lee, Y. H. and Lee, W.-S. 2015. Predicting potential epidemics of rice leaf blast and sheath blight in South Korea under the RCP 4.5 and RCP 8.5 climate change scenarios using a rice disease epidemiology model, EPIRICE. *Agric. For. Meteorol.* 203:191-207.

Kim, K.-H. and Jung, I. 2020. Development of a daily epidemiological model of rice blast tailored for seasonal disease early warning in South Korea. *Plant Pathol. J.* 36:406-417.

Kim, K.-H. and Lee, J. 2020. Smart plant disease management using agrometeorological big data. *Res. Plant Dis.* 26:121-133 (in Korean).

Kim, Y., Roh, J.-H. and Kim, H. Y. 2018. Early forecasting of rice blast disease using long short-term memory recurrent neural networks. *Sustainability* 10:34.

Kingma, D. P. and Ba, L. J. 2014. Adam: a method for stochastic optimization. Preprint at https://arxiv.org/abs/1412.6980 .

Lin, T.-Y., Goyal, P., Girshick, R., He, K. and Dollár, P. 2017. Focal loss for dense object detection. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2980-2988. Institute of Electrical and Electronics Engineers, Cambridge, MA, USA.

Liu, A., Ghosh, J. and Martin, C. 2007. Generative oversampling for mining imbalanced datasets. In: *Proceedings of the 2007 International Conference on Data Mining*, eds. by R. Stahlbock, S. F. Crone and S. Lessmann, pp. 66-72. CSREA Press, Las Vegas, NV, USA.

Malicdem, A. R. and Fernandez, P. L. 2015. Rice blast disease forecasting for northern Philippines. *WSEAS Trans. Inf. Sci. Appl.* 12:120-129.

Nettleton, D. F., Katsantonis, D., Kalaitzidis, A., Sarafijanovic-Djukic, N., Puigdollers, P. and Confalonieri, R. 2019. Predicting rice blast disease: machine learning versus process-based models. *BMC Bioinformatics* 20:514.

Probst, P., Boulesteix, A.-L. and Bischl, B. 2019. Tunability: importance of hyperparameters of machine learning algorithms. *J. Mach. Learn. Res.* 20:1-32.

Sola, J. and Sevilla, J. 1997. Importance of input data normalization for the application of neural networks to complex industrial problems. *IEEE Trans. Nucl. Sci.* 44:1464-1468.

Wang, J. C., Correll, J. C. and Jia, Y. 2015. Characterization of rice blast resistance genes in rice germplasm with monogenic lines and pathogenicity assays. *Crop Prot.* 72:132-138.

Yang, I., Jeon, W. H. and Moon, J. 2019. A study on a distance based coordinate calculation method using Inverse Haversine Method. *J. Dig. Contents Soc.* 20:2097-2102.