Establishing the Potential Rice Loss Prediction Model of Climate and Nature Disaster Factors Based on Machine Learning Theory

Yun-Ju Hsu, Yu-Keng Chen, and Chao-Heng Tseng

1National Taipei University of Technology, Taipei, Taiwan

Abstract. The United Nations points out that extreme climate events are frequent and widespread in the 21st century and have become a global security issue. Artificial intelligence and machine learning have attracted much attention in environmental applications. This study aims at applying machine learning (ML) to rice disaster prediction, and uses SPSS to analyze environmental impact factors. After model training and evaluation, four models are provided, among which short-term prediction results show high accuracy on a single event, which are suitable for water damage, cold damage, and plant diseases and insect pests respectively. In terms of long-term prediction, using future meteorological prediction values to predict potential rice losses is better, especially within a specific time period. Ultimately, relevant units such as the Council of Agriculture or the Agri-Food and Food Administration can choose a suitable model based on different purposes (short-term or long-term forecasting).

Keywords: Climate change, artificial intelligence, machine learning, rice loss, Statistical Software for Social Sciences (SPSS)

1.Introduction

In recent years, the global economy has been booming, energy and fuels have been used extensively, and carbon dioxide emissions have increased rapidly, causing global warming to become increasingly serious. Increased surface temperatures have brought many negative impacts, such as melting icebergs, rising sea levels, and flooding of coastal areas [1] and natural disasters such as erosion [2]. The IPCC points out that extreme climates will become more frequent and widespread. Especially in the past two years, Climate Emergency has replaced Climate Change, and the world has declared a climate emergency [4, 5]. In order to respond promptly to the various climate crises brought about by the climate emergency,
countries must adopt more progressive methods, such as artificial intelligence [6] and machine learning [7, 8].

Artificial intelligence (AI) is a technology that presents human intelligence through computer programs and is dedicated to solving common cognitive problems related to human intelligence, such as learning, problem solving, and pattern recognition. Provide data for AI to learn from past experiences and machine learning solutions. In recent years, AI has begun to be combined with the Internet of things (IoT) [13-15], allowing AI to analyze huge amounts of data faster, and the data can include Sources, places, objects and event information that have never been touched before.

Therefore, machine learning (ML) theory under the concept of artificial intelligence (AI) is used to build models. Machine learning models can solve the complexity and accuracy of big data [16-19], and predict the future climate and possibilities. Natural disasters suffered are set as impact parameters to predict rice losses in the agricultural industry chain, and predict long- and short-term impacts, and then introduce adaptation policy planning for the rice industry chain in response to climate change, so as to prepare rice-related industries for disasters. Advance prediction and preparation can reduce avoidable losses of people and property, and can also provide industries with early preventive measures against climate change disasters.

2. Methods and materials

This study will use Statistical Software for Social Sciences (SPSS) to build the model. First, through two screening modes: independent sample T test and neural network-like, historical climate data, natural disaster data and agricultural related data from 1946 to 2019 were used to classify variables and screen important variables, and select the significant or important variables. Variables with higher reliability are selected, and then the model is trained using two different decision tree growth methods: Chi-square Automatic Interaction Detection (CHAID) and Classification and Regression Tree (CART), and the ROC analysis curve is used to determine whether the model has Predictive value.

The prediction part is divided into long and short-term predictions. The short-term prediction is the prediction month data, and the month with the greater rice loss among the five disaster items is selected as the prediction case, a total of five; the long-term prediction is the prediction year data, and the first part is the prediction for 2020. The actual 2020 observations and MOHC used HadGEM2-ES to simulate the future climate monthly data of IPCC and the TCCIP statistical downscaled monthly data prediction data of each county and city in the future to estimate the amount of potential rice losses in the future, and Compare the difference with the actual amount of rice losses in 2020; the second part is to predict the past three years (2017-2019). Meteorological factors and natural disaster factors from 2017 to 2019 are put into the best model to calculate the amount of potential rice losses. Forecast and compare the difference with the average of the actual rice loss amount in the three years from 2017 to 2019.

Finally, through the theory of machine learning, we completed the construction of a prediction model for the potential loss amount of rice affected by meteorological factors and natural disaster factors, and then judged whether the model was suitable for short-term prediction or long-term prediction, and its practical practicability. The research architecture is shown in Figure 3-1.

Data related to historical meteorology and natural disasters are taken from the Central Meteorological Administration’s Climate Statistics Annual Report - Ground Data and data from the Statistics Network of the Ministry of Transport.

In terms of agricultural-related data, because the historical data collection period is relatively long, some data are unrecorded. In order to reduce the errors caused by missing
values in the research data, this study uses Excel as the calculation tool to process the missing values, and based on Calculation of data in the "Council of Agriculture Statistical Annual Report" provided by the Council of Agriculture. If it is impossible to make an inference, assumptions will be made.

Since there may be more than one natural disaster in each month, each loss item also has more than one number. Therefore, this part of this study uses the average value for calculation. The data collected in this study will be imported into SPSS for machine learning model construction.

(1) Dependent variable
The calculated amount of potential rice disaster loss is the part of the causal variable. Then, using "whether rice loss is caused" as the "category", it is judged based on the amount of potential rice disaster loss. If there is a potential rice disaster loss amount, it is regarded as 1; If there is no potential rice disaster loss amount, it is regarded as 0.

(2) Independent variables
From the data related to meteorological factors and natural disaster factors from 1946 to 2019, the independent variables are summarized, including meteorological factors and natural disaster factors.

Table 1. Yield impact coefficient of rice disaster

<table>
<thead>
<tr>
<th>NO.</th>
<th>Disaster type</th>
<th>Yield impact coefficient</th>
<th>references</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Flooding injury</td>
<td>10%</td>
<td>[1]</td>
</tr>
<tr>
<td>2</td>
<td>Typhoon</td>
<td>17%</td>
<td>[2]</td>
</tr>
<tr>
<td>3</td>
<td>Chilling injury</td>
<td>3%</td>
<td>[3]</td>
</tr>
<tr>
<td>4</td>
<td>Drought injury</td>
<td>7%</td>
<td>[4]</td>
</tr>
<tr>
<td>4</td>
<td>Insect injury</td>
<td>15%</td>
<td>[5]</td>
</tr>
</tbody>
</table>

Table 2. Independent variable

<table>
<thead>
<tr>
<th>No.</th>
<th>variable name</th>
</tr>
</thead>
<tbody>
<tr>
<td>X01</td>
<td>average temperature</td>
</tr>
<tr>
<td>X02</td>
<td>precipitation(mm)</td>
</tr>
<tr>
<td>X03</td>
<td>Maximum daily precipitation(mm)</td>
</tr>
<tr>
<td>X04</td>
<td>maximum wind speed(m/s)</td>
</tr>
<tr>
<td>X05</td>
<td>sunshine hours(hour)</td>
</tr>
<tr>
<td>X06</td>
<td>Relative humidity(%)</td>
</tr>
</tbody>
</table>
Among the natural disaster factors, in addition to the environmental impact on rice growth, sudden natural disasters, such as typhoons and heavy rains, will directly cause flooding and lodging of rice, increasing agricultural losses. The area of agricultural disaster losses caused by natural disasters from 1946 to 2009 was, in order, typhoon 10,145 thousand hectares (67%), heavy rain (13%), drought (10%), pests and diseases (5%), and cold damage (4%) and others (1%) [331]. This part is not far from the data collected by this study. However, this study only investigated the damage area of rice, so there are slight differences. This study collected rice from 1946 to 2019. Damaged area results

![Typhoon Flood Drought](image1)

**Fig1.** Area of rice lost due to natural disasters from 1946 to 2019

![ROC curve](image2)

**Fig2.** ROC curve
3 Machine learning

After processing the data, the receiver uses the independent sample T test and the multi-layer perceptron (MLP) in the neural network (ANN) to screen important variables, and then puts the screened important variables into the decision tree. Carry out model training and evaluation, and finally use the Receiver Operating Characteristic (ROC) and the Area Under the Curve of ROC (AUC) to evaluate the accuracy and predictive value of the model to complete the construction of the model. The software used is IBM SPSS 22.

Decision tree is used for training evaluation. The training decision tree growth method in this study is divided into two parts, namely Chi-square automatic interaction detection method (CHAID) and classification regression tree (CART). The model verification setting part mainly focuses on splitting samples. After different testing and training calibration ratios, the most appropriate training and calibration ratio is selected as the basis for data segmentation. This study uses whether rice losses are caused for decision tree model training, and trains with training-test ratios of 60:40, 70:30, and 80:20. The input independent variables are independent samples T-test and neural network-like screening, important variables.

In the decision tree segmentation sample verification part, SPSS will display the overall correct percentage. The presentation effect is equivalent to the concept of a confusion matrix. You can see the overall correct percentage during training and testing respectively, and choose a better sample segmentation ratio. Also evaluate which model algorithm is best.

The model performance indicator is the Receiver Operating Characteristic (ROC), which has recently developed well in the fields of machine learning and data mining, and is also a relatively approachable diagnostic value indicator.

In signal detection theory, the ROC curve graphically presents the performance of a binary classifier system under a specific classification or threshold. The vertical axis (y-axis) of the graph is the true positive rate (TPR), also known as sensitivity; the horizontal axis (x-axis) is the false-positive rate (FPR), expressed as 1 – specificity, while sensitivity is the probability of correctly judging the result as positive, and specificity is the probability of correctly judging the result as negative or negative.

4. Results and Discussion

The trends of meteorological factors and natural disaster factors from 1946 to 2019 were analyzed separately, and the relationship between the two factors for rice was discussed.

The trend charts of six meteorological factors from 1946 to 2019 are shown in Figure 4.1 to Figure 4.6

According to Figure 4-1, it can be found that the average temperature is increasing year by year, which is consistent with the global climate warming trend. In May 2020, the global surface temperature was 0.95°C higher than the 20th century average temperature, breaking the 140-year record. In Taiwan, the annual average temperature has continued to reach new highs in the past five years. From 2015 to 2019, it was among the top six with the highest average temperatures, and the annual average temperature after 2015 has never been lower than 24.2°C. In terms of global trends, in the past 30 years, the average temperature has increased by 0.2°C every 10 years. Since 1977, the global average temperature has been higher than the centennial climate value for 43 consecutive years, and the top eight hottest places in the past century have all occurred in nearly 10 years.

As can be seen from Figure 4-2, Taiwan's precipitation had stabilized until 2006, but it suddenly spiked in 2005. It is speculated that because 2003 to 2004 was the Year of the Holy Child, the precipitation in Taiwan was less than in previous years, and 2005 Year.
In Figure 4-3, the maximum daily rainfall trend from 1946 to 2019 fluctuates up and down with no definite pattern. The average value falls at 184.72 mm per day. According to the Central Meteorological Administration’s definition of heavy rain, Taiwan almost often heavy rains, but because such climate conditions are suitable for the growth of rice, the monthly rainfall is about 100~200 mm.

In the maximum wind speed section of Figure 4-4, the overall range is large and irregular. The maximum wind speed falls on an average of 61.04 (m/s). In other words, the annual maximum wind speed in Taiwan causes a reduction of about 40% to 60% in rice yields.

As for the relative humidity in Figure 4-5, the overall trend is stable and remains at around 76%.

Figure 4-6 For the sunshine hours, the average value falls at 145.57 (hr), and the overall trend is a slow decline. According to relevant studies, there has been a significant downward trend in sunshine rate since 1975. It is speculated that this is caused by the increase in the number of suspended particles in the world after the industrial revolution and the increase in cloud condensation nuclei, resulting in an increase in cloud cover. Sunshine rate is defined as the ratio of actual sunshine hours to astronomical sunshine hours, expressed as a percentage. Various reasons point to the impact of suspended particles in the air, which will lead to a downward trend in sunshine rate and sunshine hours. Comparison of PM 2.5 removal efficiency and CADR at the 60th minute of the
The relationship between meteorological factors, natural disaster factors and rice

Calculate the total damaged area of rice from 1946 to 2019, the damaged area of rice in the first period, the damaged area of rice in the second period, the total potential loss amount of rice, the potential loss amount of rice in the first period and the potential loss amount of rice in the second period calculated in this study. Six natural

Fig5. Trend chart of maximum daily rainfall (mm) over the years

Fig6. Rainfall (mm) trend chart over the years

Fig7. Maximum wind speed (m/s) trend chart over the years
disaster factors and six meteorological factors were used to draw XY scatter diagrams, and the coefficient of determination of the linear trend (Coefficient of determination, recorded as $R^2$)

In the model construction, two growth methods of decision tree, chi-square automatic interaction detection method (CHAID) and classification and regression tree (CART) are used, coupled with two variable screening modes, independent sample T test and neural network, and three segmented sample ratios of 60:40, 70:30, and 80:20 for prediction. The results obtained are that the results of each model judged by the ROC curve and AUC have predictive value, and the models are respectively It has the best performance under the conditions of $T$-test + Decision Tree(CHAID_70:30), $T$-test + Decision Tree(CART_80:20), ANN + Decision Tree(CHAID_70:30) and ANN + Decision Tree(CART_70:30) Performance.

5. Conclusion

Use the machine learning (ML) software SPSS to do disaster prediction and climate simulation analysis of environmental impact factors on the industrial chain, and then make

![Fig8. Maximum wind speed (m/s) trend chart over the years](image1)

![Fig9. Relationship between total rice damaged area and average temperature](image2)
corresponding implementation strategies based on the analysis results to provide the industry with early preventive measures against climate change disasters and address the development of climate adaptation policies, such as reducing crop impacts and establishing early warning system mechanisms.

Through the construction results of this rice potential loss prediction model, it can be seen that the amount of potential rice losses can be estimated by meteorological factors and natural disaster factors, and the meteorological prediction data of IPCC_HADGEM2-ES_RCP4.5 has also been applied to obtain quite good results, which is very accurate. In summary, it is successful to build a rice potential loss prediction model using machine learning theory, and it also represents the feasibility of introducing artificial intelligence into the agricultural industry chain.

Long-term monitoring, tracking management and early warning systems are the focus of future agricultural adjustment policies, integrating existing new technologies, such as using artificial intelligence tools for automated monitoring and intelligent management, to improve the resistance to climate crises caused by climate change. Ability to deal with adversity.

5.1 Model research results

Using machine learning to build a potential rice loss prediction model: This study uses the decision tree algorithm in machine learning theory and historical data from 1946 to 2019 to combine two decision tree growth methods and two variable screening modes. And three different split sample ratios, model training and testing were carried out with SPSS software, and then the ROC curve and AUC area chart were used to judge the prediction results of the entire model. The overall results showed that each model had a considerable degree of accuracy and predictive value.

5.2 Model training and evaluation

The judgment models of this study are respectively in T-test + Decision Tree (CHAID_70:30), T-test + Decision Tree (CART_80:20), ANN + Decision Tree (CHAID_70:30) and ANN + Decision Tree (CART_70:30). Best performance.

5.3 Directions in which the model can improve

In the IPCC_HADGEM2-ES_RCP4.5 prediction value processing part, this study is based on the historical proportions of each district, county and city. If the meteorological prediction values can be processed with more standard downscaling technology, the overall prediction results will be more accurate.

5.4 Model recommendations

This research model was built and trained based on the potential amount of rice losses, but the prediction results were not ideal. It is recommended that future research can use the actual amount of rice losses to train the model to reduce the error between it and the actual situation.
References

16. H.-C. Li, Y.-C. Che, and M.-J. Guo, "The Development and Application of Taiwan Typhoon Loss Assessment System (TLAS)," 2013.
