Classification of Japanese Cedar and Japanese Cypress in Sentinel-2/MSI by using Support Vector Machine

Yoichi Hanai1 and Masayuki Matsuoka\*2

1Graduate Student, Graduate School of Engineering, Mie University

1577 Kurimamachiya, Tsu, Mie 514-8507 Japan

Email: 422M523@m.mie-u.ac.jp

2Associate Professor, Graduate School of Engineering, Mie University

1577 Kurimamachiya, Tsu, Mie 514-8507 Japan

Email: matsuoka@info.mie-u.ac.jp

**KEY WORDS:** Classification, Japanese cedar and Japanese cypress, Support vector machine, Spectral band, Permutation importance

**ABSTRACT:** Japanese cedar and Japanese cypress are important species for timber production in Japan. Satellite data can be used to update and correct information on these planted forests. In this study, machine learning based pixel classification of Japanese cedar and Japanese cypress was attempted using multispectral satellite data. Monthly accuracies were compared and seasonal changes in accuracy were assessed. In addition, the spectral wave regions that have a greater influence on the creation of the learning models were investigated by permutation importance. The classification results showed that the most accurate period was at the end of March, with an accuracy of approximately 87%. The results showed the importance of reflectance in the 740 and 1610 nm wavelength ranges in the classification.

# Introduction

## Motivation

Japanese cedar and Japanese cypress account for 70% of the area of planted forests in Japan and are important tree species for timber production (Forestly Agency2023). Artifical forests are managed using compartment maps, forest registration book or geographic information system, but information needs to be updated and corrected, and efficient surveys are needed. Satellite data provide information on forests over a large area, but classification of Japanese cedar and Japanese cypress is difficult due to their similar tree shape and spectral reflectance.

The satellite data show changes in spectral reflectance values at different times of the year. Therefore, the accuracy of machine learning classification also varies with the season. Identification the most effective season for classification is necessary to determine the exact distribution of tree species.

Satellite data often contains multiple spectral data, and some spectral bands are commonly used by many sensors. By identifying spectral bands that have a high importance on the creation of the classification models, it is expected that high classification accuracy can be achieved in multiple satellite data.

## Aims

The main objective of this study was to discriminate between Japanese cedar and Japanese cypress trees in the Sentinel-2 Multispectral Instrument (MSI) imagery using Support Vector Machine (SVM), a machine learning technique. The model was trained and tested using for different seasons to assess the impact of seasonal variations on the classification results. This was done to understand the impact of seasonal vegetation changes on classification accuracy and to identify the optimal classification period. We also investigated which spectral band contribute most to classification when building machine learning models. The permutation importance method was used for the analysis. This evaluation identified the most influential spectral band in the classification of Japanese cedar and Japanese cypress.

## Related work

Pixel-based machine learning classification using satellite imagery has been actively studied for classifying tree species in forests. Polyakova et al. (2023) used the random forest algorithm for the Sentinel-2/the MSI to classify deciduous and coniferous trees. The target area was the Raifa section of Volga-Kama State Reserve in the Tatarstan Republic, Russia. Ten bands and the Normalized Difference Vegetation Index (NDVI) were used for training. The classification results show an accuracy of 98% for the training data and 76% for the test data. Immitzer et al. (2012) used the random forest algorithm to classify ten tree species on WorldView-2 satellite data. The target area was temperate forests in Australia, using eight bands observed in October. The classification accuracy was about 82%. Deur et al. (2020) used two methods, the random forest algorithm and a support vector machine, to classify three tree species on WorldView-3 satellite images. The target area was mixed deciduous forest in central Croatia; in addition to the eight bands, a texture feature by the GLCM was used. The classification accuracies are 95% for random forest and 92% for support vector machine.

# STUDY AREA and data used

The study area covered Mie, Wakayama and Nara prefectures in Japan. Figure 1 shows a true-colour image of the Sentinel-2/MSI in the target area. In order to select pure forests of Japanese cedar and Japanese cypress, national forest data from the National Land Numerical Data System (Ministry of Land, Infrastructure, Transport and Tourism of Japan, 2023) was used. The satellite data were the Sentinel-2 Level 1C which were observed between 2018 to 2023. During this period, data with little or no cloud cover were selected, and cloud mask was applied to remove the cloud pixels in the image. We used at least one data in each month to capture the seasonal trend of the classification accuracy. Ten Sentinel-2 bands were used in the study, excluding bands 1, 9 and 10.



Figure 1. True-colour image of the Sentinel-2/MSI in the target area.

(March 16th, 2022)

# METHODS

## Classification by SVM

SVM is a supervised machine learning technique. It creates a hyperplane that separates multiple data in the feature space by class, and classifies the data according to which side of the boundary plane the data belongs to. For SVM, scikit-learn, an open source machine learning library in Python, was used. Radial basis function (RBF) was used as the kernel of SVM based on the number of variables handled in this study. The gamma value which indicates the complexity of the kernel shape was specified as the “scale” because it is automatically determined from the number of training data and the number of variables.

## Training and Test Datasets

The datasets were constructed from the Sentinel-2/MSI images in pure Japanese cedar and Japanese cypress forest areas based on national forest data. Since we used 10 bands of MSI data, each pixel in image contains 10 variables. If pixels for training and test data were selected from single forest area, the accuracy of classification model would be expected to be unfairly high. To reduce this risk of overfitting, each polygon of the national forest data was assigned either of training or test data.

## Permutation Importance

Permutation importance is a method used in machine learning to assess the importance of the features used (Altmann et al., 2010). This method evaluates how much a feature contributes to improving the prediction accuracy of a model. Specifically, multiple evaluation datasets are generated by randomly sorting the specific feature, and the importance is calculated from the difference between the average accuracy of the feature-sorted dataset and the accuracy of the original dataset. In this study, 10 sorted datasets were generated for each feature and their average importance was calculated.

# Result and discution

## Classification Results

The results of the classification of the forests in March 16th, 2022 into Japanese cedar and Japanese cypress are shown in Figure 2. The green and blue areas represent Japanese cedar and Japanese cypress, respectively, and dark gray is other land area, and bright gray is water.

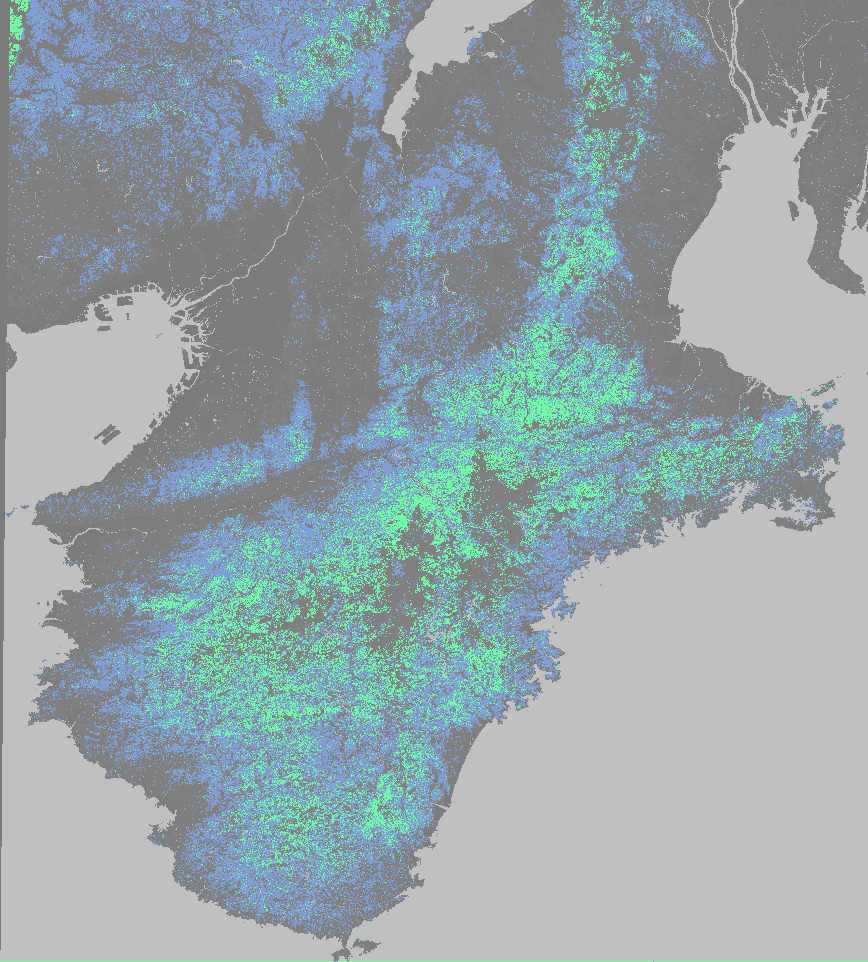


Figure 2. Classified images for the entire forest of the target area only. The green and blue areas represent Japanese cedar and Japanese cypress, respectively.

For the visual assessment of the classification accuracy, comparison with aerial photographs are shown in Figure 3. Figure 3(a) shows an aerial photograph taken on March 4th, 2022. The brown and green areas indicate Japanese cedar and Japanese cypress, respectively. Figure 3(b) is an enlarged image of the same area as Figure 3(a), showing that the shape of the distribution is accurately represented.

|  |  |
| --- | --- |
|  |  |
| (a) Aerial image  4th March 2022 | (b) Classified image  16th March 2022 |

Figure 3. Comparison of classified images with aerial photographs.

Figure 4 shows the time series of classification accuracy, sorted by calendar date. Starting in January, accuracy increased and peaked between March and April, with a range in 86% and 88% over this period. Accuracy then fell to around 81% in May and June. Then, it increased again up to 86% in summer, and slowly declined to winter. The period from January to April was the flowering period of the cedar trees, and it increased the classification accuracy by enlarging the difference of spectral reflectance between Japanese cedar and Japanese cypress. On the other hand, the reason for the decrease in accuracy in May is correctly unclear.

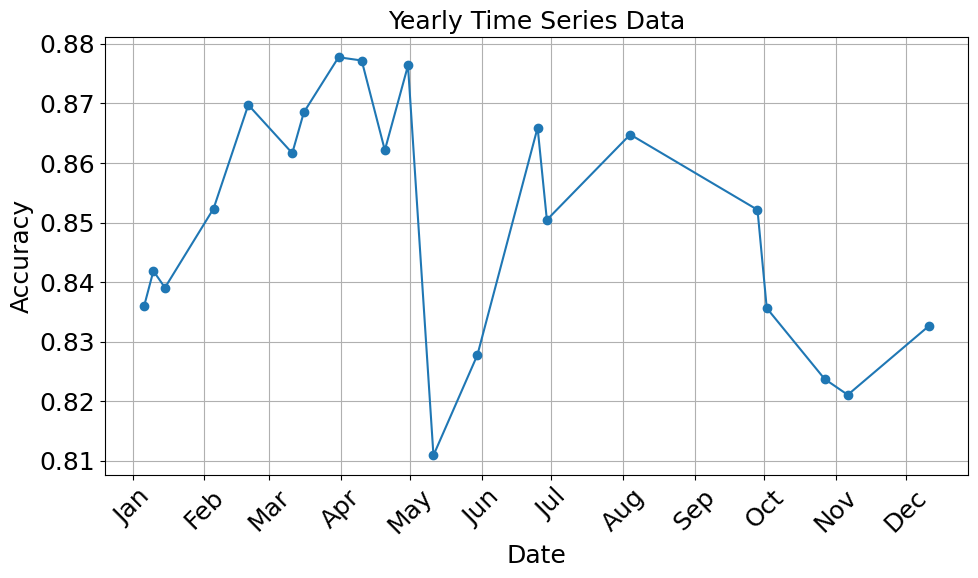


Figure 4.Seasonal change of the classification accuracy.

## Importance of Bands

We calculated the importance of the spectral bands by permutation importance for the most accurate data in March and the least accurate data in May. Figure 5 shows permutation importance scores by band. The two bands with the highest importance for both days were common: bands 6 and 11. Their central wavelengths are 740 nm for band 6 and 1610 nm for band 11, respectively. In the March data, the most important band was band 6 at about 0.49. The second most important band was band 11 with a value of about 0.40.

|  |  |
| --- | --- |
|  |  |
| (a) March 31st, 2021 | (b) May 11th, 2019 |
| Figure 5. Importance of bands calculated by permutation importance. | |

Seasonal changes of the average and the standard deviation of reflectance in bands 3 and 6 for the Japanese cedar and Japanese cypress were shown in Figure 6. In all bands, cypress showed higher reflectance than cedar throughout the year. In band 6, the statistical distribution of Japanese cedar and Japanese cypress had larger distance in spring, when classification had the highest accuracy. Specifically, the each average value was located outside the standard deviation range, unlike other seasons (Figure 6(a)). This suggests that band 6 is of high importance for classification.

|  |  |
| --- | --- |
|  |  |
| (a) Band 6 | (b) Band 3 |
| Figure 6. Annual changes in reflectance of Japanese cedar and Japanese cypress in the Sentinel-2/the MSI data.  The red and blue show cypress and cedar, respectively. Vertical bars indicate standard deviations. | |

# conclusion

In this study, the accuracy of SVM classification of pure Japanese cedar and Japanese cypress forests was verified using Sentinel-2/MSI imagery in Mie, Wakayama and Nara prefecture. Seasonal changes of classification accuracy were examined to assess the influence of vegetation phenology on the discrimination of Japanese cedar and Japanese cypress. We also calculated the importance of spectral bands using permutation importance to evaluate the contribution of each band on classification. The results showed that classification accuracy ranged from 80% to 88% throughout the year, with particularly high accuracy observed in early spring. Furthermore, analysis of the importance of the spectral band showed that band 6 (740nm) and band 11 (1610nm) were particularly important. The results suggest that the use of March-April data is effective in improving the accuracy of Japanese cedar and Japanese cypress classification in this area. It is also clear that reflectance in the 740 nm and 1610 nm wavelength ranges is particularly important. The annual trend in accuracy suggested that the accuracy of the data, particularly in May, was lower than at other times of the year. The reason of this phenomenon will be investigated in detail in future studies.

# References

Altmann, A., Toloşi, L., Sander, O., Lengauer, T., 2010. Permutation importance: a corrected feature importance measure, Bioinformatics, 26(10), 1340, https://doi.org/10.1093/bioinformatics/btq134

Deur, M., Gašparović, M., Balenović, I., 2020. Tree species classification in mixed deciduous forests using very high spatial resolution satellite imagery and machine learning methods, *Remote Sens.* 12(23), 3926. https://doi.org/10.3390/rs12233926

Forestry Agancy, 2023. Data on Japanese cedar and Japanese cypress forests, Available online: https://www.rinya.maff.go.jp/j/sin\_riyou/kafun/data.html (in Japanese) (accessed on March 23, 2023)

Immitzer, M., Atzberger, C., Koukal, T., 2012. Tree species classification with Random forest using very high spatial resolution 8-band worldView-2 satellite data, *Remote Sens.* 4(9), 2661-2693. https://doi.org/10.3390/rs4092661

Ministry of Land, Infrastructure, Transport and Tourism of Japan, 2023. National Land Numerical Data. Available online: https://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-A45.html (in Japanese) (accessed on March 23, 2023)

Polyakova, A., Mukharamova, S., Yermolaev, O., Shaykhutdinova, G., 2023. Automated Recognition of Tree Species Composition of Forest Communities Using Sentinel-2 Satellite Data, *Remote Sens.* 12(2), 329. https://doi.org/10.3390/rs15020329