Classification of tree species from reflectance spectra

Stident name: Jincheng Sun
Supervisor name: KRAUSS.FRANK, DONOGHUE.DANIEL N.M, BROWN. ANTHONY.M, MORLE.PETER.J
Submitted as part of the degree of M.Sc. MISCADA to the Board of Examiners in the Department of Computer Science, Durham University
Code link: https://github.com/jddk23/Sunjincheng.git

1.Abstract

Context: Classification of tree species can be a tough work for spending too much time, the use of remote sensing can simplify this. One solution is using machine learning methods to deal with reflectance spectra, which is hyperspectral data.
Aim: Identify a small number of spectral bands with the full bands combined discriminating power between a large number of relevant species, at a scale and with an accuracy by a most suitable classifier model. The planned reduction of data dimensionality greatly reduces noise, and it allows future development of bespoke systems that enable non-specialists to identify tree species using lightweight multi-band cameras.
Method: Tree species classification on reflectance spectra data can be solved with the dimension reduction methods and classification models. For the datasets coffee tree and mangroves tree, the first step is choosing the suitable wavelength and remove the noisy data. The second step is using dimension reduction methods like PCA, ICA, LLE to improve the dataset. The third step is using different classification models like Naive Bayes, Support Vector Machine(SVM), Neural Network, Tree models, to compare the classification effects and identify the best algorithm. The last step is optimizing the parameters to improve the best algorithm.
Results: This is a multiple classification task so should consider both accuracy and kappa. After the optimization, ICA-Xgboost model was found with best classification effect, where the test accuracy and kappa can reach around 0.97 and 0.96 on the two datasets respectively. Meanwhile the training time of this identified best method ICA-Xgboost only need within 1 seconds. After considering different range of datasets, it seems no need to use the full range of wavelength to do the classification, since the effect of different wavelength range almost same under ICA-Xgboost model.
Conclusions: Machine learning approaches are very appropriate for the tree species classification problem. The SVM model and neural network seems difficulty to design for the parameters and hidden layers. In tree models, the Xgboost is a good classifier model for all range of wavelength, especially when reduced the dimension with ICA or LLE methods. This model can classify the samples in each classes well for the high value of kappa and easy to improve the parameters of the model with grid search.
Keywords: Tree Species Classification, Wavelength, PCA, ICA, LLE, t-SNE, Machine Learning, Naive Bayes, SVM, Decision Tree, Neural Network, Random Forest, GBDT, Xgboost, Boosting, Bagging, Accuracy, Kappa, F-score, Confuse Matrix.

2.Introduction

Plant pigments influence the reflectance spectra of leaves, their greenness, in the visible part of the electromagnetic spectrum while cellular structure and foliar moisture dominate leaf reflectance in the near infrared. Researchers will collect a large number of samples of reflectance spectra data of the identified species in order to classify the tree species. The spectra will be collated in an open-source database, containing hundreds of individual spectra of the same species, their averages and variabilities, which aims to be the largest database for key species of critical importance for the UK forestry sector and woodland conservation organisations. Many tree species are vulnerable and to drought and disease and dieback, so the ability to identify the location of different species is critical to management and conservation efforts. Using machine learning techniques can develop methods that enable automated species identification based on their spectral characteristics.

2A. Background of Tree Classification with Hyperspectral Data

Previously, number of studies focusing on tree species classification by sensor types has constantly increased in the past 35 years, especially the cases observed for the hyperspectral data or imaging spectroscopy. A detailed summary of earlier research exploring the influence of spatial and temporal dynamics on reflectance properties of trees was found to provide much information(Hesketh, 2012). In the study of spectral variation among tree species, the leaves was found performed better than other tissues like woody branches(bark) for the full range(Matthew L. Clark, 2012). In reflection spectrum, researches found that the wavebands at 780nm, 790nm, 800nm, 1480nm, 1530nm and 1550nm were identified as the most useful bands for mangrove species classification(LE WANG, 2009). The curse of dimensionality (e.g., Dalponte et al., 2013) is another problem, in order to avoid it, dimension reduction method like PCA are widely used. Remote sensing-based tree species classification evolved with methodological developments in the domain of statistical learning. In resent studies most widely used classification techniques included Neural Networks, RF and SVM (Immitzer et al., 2012; Pant et al., 2013) instead of simple models like tree and logistic models. Moreover, a detailed review on classification approaches in a remote sensing context is for example was provided by Lu and Weng in 2007.

2B. Project Objective and Achievements

For this project, there was two major objectives, the first objective is to collect the machine learning methods of classification and compare their effect, in order to define a effective tree species classification system by machine learning approaches and to see whether it can do well in the classification of these trees species with hyperspectral wavelength data of mangroves and coffee trees. Since the mangroves tree dataset and the coffee tree dataset are all labeled, the supervised classifier

models instead of unsurpervised classifier models like k-nearest neighbors algorithm (KNN) should be used. This objective is to compare the effect of different dimension reduction methods on the kinds of classifier models to find the most effective combination and define whether dimension reduction methods can improve the classification effect. Ensemble methods like bagging and boosting are also used to obtain better predictive performance. After defining the suitable tree species classification system, the grid search method was used to adjust parameters in order to get a better classification effect. The second objective is to see whether different range of wavelength will affect the classification results. In mangroves tree dataset, consider two different range, 400nm-900nm and 400nm-2400nm to compare their accuracy, kappa, F-score and confuse matrix.

3.Related work

Machine learning models are widely used in the hyperspectral data classification in resent works and many solutions to classification problem have achieved success. To classify the tree species with reflectance spectra, which is the hyperspectral data, must be analysed, and features extracted. In this section, an overview of the use of different kinds of techniques for classification in other fields will be shown.

3A. Features Extracted and dimension Reduction Methods

In the projects of classification, some drawbacks of hyperspectral data are put forward and listed(G. Camps-Valls, 2005): (1) curse of dimensionality: this means the dimension of the samples is too high, while there are not so many labeled samples for the reason of money and time. (2) large spatial variability of spectral signature, this means there are noisy data in the spectrum, especially in the wavelength of the beginning and the end. To deal with the curse of dimensionality, some dimension reduction methods were proposed. Most common one is principal component analysis(PCA). The PCA method force on the largest variance in the projected dimensions. Range of researchers consider this method to reduce dimension, but this method may be easily effected by the noise and less important data. Method of independent component analysis(ICA) was used to reduce dimension and after comparing with PCA, ICA provide advantages over the PCA(Abdulhamit Subasi, 2010). The benefit of nonlinear dimension reduction methods like LLE is that it can find the meaningful low dimensional structure hidden in high dimensional data, since it can keep more characteristics. T-SNE is another nonlinear dimension reduction method which has a good effect on visualization.

3B. Classifier Model

Many machine learning models were used and proved successful for classification in other fields, except tree species. Models like Tree model, SVM model, Naive Bayes model and Neural Network model are frequently used. Each classifier model has its

own drawbacks and needs to be chosen with regard to the specific requirements of the application. To deal with these weaknesses, methods like dimension reduction methods, ensemble methods are used to improve the classification effect.

Naive Bayes Model

Previous studies have demonstrated the usefulness of Naive Bayes classifier in solving various classification problems. The CC-ICA-Naive-Bayes model(Liwei Fan,2009) was put forward to do the classification of microarray data on medical decision making, which showed feature selection and dimension reduction method ICA can improve classification effect. In addition, the class-conditional independent component analysis (CC-ICA) technique(Vitria, 2007) is a improvement based on ICA method. Ensemble methods like boosting (Kearns, 1988), was also added to Naive Bayes model aiming to improve the classification effect. Boosting Naive Bayes model(Stijn Viaene, 2004) was applied on claim fraud diagnosis, however it found the boosting method can just improved the classification effect a little.

SVM Model

Support Vector Machine(SVM) is a classification method which is based on the use of the kernel function that allows an optimal separation of data. SVM model(L. Kanaan, 2011) was use to classify the electrocardiograph and found that dimension reduction methods like KPCA and PCA combined with SVM model perform better than the SVM model, which can rise the test accuracy about 20 %. In addition, among the kernels of linear, poly and Gaussian, the Gaussian had the best effect. Except using dimension reduction methods to improve the classification effect of SVM model, ensemble methods are also used. The Multiple SVM-RFE(Gaussian kernel) on mammogram data(Sejong Yoon, 2009) used the boosting method to improve the effect, which performed better than SVM model in test accuracy, especially on the mass dataset. Another ensemble method bagging is also used (Pan Huang, 2020) on the common non-normal walking, mainly compared the method SVM model, Relief SVM method(Relief is a method for future selection) and the Relief Bagging SVM method. The test accuracy rise around 3% after using the bagging method on the Relief SVM model.

Neural Network

Unlike CNN method which is mainly used in the classification of hyperspectral image, the neural network is often used on hyperspectral data. Deep learning methods have been applied for classification tasks in various fields like computer vision, financial and so on. The DNN model was found performs better compared with SVM model, 7.9% accuracy rate higher on the whole-brain resting-state functional connectivity patterns of schizophrenia(Junghoe Kim, 2015). The DNN model was also compared with other models like Logistic Regression, Naive Bayes(Selçuk BAYRACI, 2019) in application to loan default prediction. The experimental results indicated that, the DNN model significantly improves the performance of a credit scoring system relative to LR, Naive Bayes and SVM models in terms of balanced accuracy.

Ensemble tree models

The Decision Tree model is one of the common classification method in previous work. The tree model was proved suitable for the spectral information of typical surface features about fly ash information(Jinfa Dong, 2012). However, classification effect of Decision Tree is not good enough for most of time, some ensemble methods were used to improve its effect, like bagging(random forest) and boosting. In the spatial prediction of flood occurrence in the Quannan area, a comparison of Decision Tree, Random Forest and Naive Bayes Tree was done and found the Random Forest performed best with the highest accuracy(Wei Chen, 2019). The researches were also found that reducing dimension can improve the classification effect of random forest. The ICA method was used to reduce the dimension in the study of alcohol dependence patients and healthy controls(Xi Zhu, 2018), which found that 90% feature elimination performed better than 50% feature elimination and No feature elimination, 10% higher of accuracy than other conditions. The boosting method was also used to improve the Tree Model like GBDT, based on GPS signal reception(Rui Sun, 2019), which found that the GBDT performed better than Decision Tree and KNN, especially on the multipath classification. Extreme Gradient Boosting Decision Tree(Xgboost) is a improvement of GBDT(Tianqi Chen, 2014). Some reaches were proved this model Xgboost possesses a stronger generalization ability, the performance in the star/galaxy classification is better than the other algorithms like GBDT, Random Forest, Adaboost, not only on test accuracy but also on training time(LI Chao, 2019). The reduction dimension method like PCA was combined with Xgboost on Stroke Detection(N. Fitriah, 2017) and found that PCA could increase the accuracy in the same number of channel.

4. Solution

Previously, there are little work force on comparing different kinds classifier models combined with dimension reduction and ensemble methods, especially on the field of tree species with reflectance spectra. Therefore, four main kinds of classification methods, Naive Bayes, SVM-REF, DNN and Ensemble Tree models were considered to apply on the coffee tree and mangrove tree dataset by using sklearn in python. The dimension reduction and ensemble methods were also taken into consideration in order to improve the effect and define the most effective classification system.

In this work, the ICA-Xgboost performs best on both coffee and mangrove tree datasets, so there would be a simple explanation of ICA and Xgboost method. In addition, a comparison of ICA-Xgboost with other classifiers will also be taken by the different classifier/ensemble packages in sklearn and “xgboost” package in python.

4A. Introduce on the tree species data

The mangrove tree data and coffee tree data was collected by Frank Krauss, Anthony Brown, Danny Donoghue and Pete Morley(Durham University). In addition the mangroves tree data was collected with the ASD FieldSpec3 and the the coffee tree

data was collected with spectrometer FLMS03901. These below pictures in Figure 1-4 showed some of the samples of mangroves tree and the machine used.

Figure 1: white mangroves leaf

Figure 2: red mangroves leaf

Figure 3: black mangroves leaf

Figure 4: picture of ASD FieldSpec3

There are 850 samples on coffee tree data and 120 of them are useless for the reason of unlabeled. The wavelength range of greenness detected is from 187nm-1027nm with 7 classes of categories, “health good” class, “rust” class, “infected green canopy” class, “geisha” class, “geisha canopy” class, “young rust yellow” class, “young rust green leaf” class. There are 841 labeled samples with 4 classes of different colors, “white”, “black”, “mud” and “red” on the mangroves tree data, and its wavelength range of greenness is from 350nm-2500nm. On the wavelength of coffee data and mangroves tree data, there are some noisy data, especially in the beginning and end of the the wavelength, therefore it is needed to remove these these noisy data. For coffee data, leave the wavelength 325nm-865nm with 1400 dimensions and for mangroves tree data, leave the wavelength 400nm-2400nm with 2000 dimensions. However for the mangroves tree data, the wavelength 1800nm-2000nm is quite noisy, so the new wavelength range 400nm-1800nm and 2000nm-2400nm with 1800 dimensions was used. In order to compare the classification effect of different wavelength range, another mangroves tree range of 400nm-900nm with 500 dimensions was used.

4B. Dimension Reduction method ICA

In this section, a effective dimension reduction method Independent Component Analysis(A. Hyvarinen, 2000) would be introduced. The explanation of algorithms of ICA method is not the main task, so here is just a roughly introduce of the main idea of this algorithm, which has two steps:

1. The first step is to perform whitening processing to make the output signal uncorrelated and with the same variance.
2. The second step is to find a rotation (that is, orthogonal transformation) to make

the output signal not only irrelevant, but also statistically independent.
ICA is implemented in scikit-learn using the Fast ICA algorithm. Since the ICA model does not include noise term, for the model to be correct, whitening must be applied.

Introduce of Parameters

In the scikit-learn of python, the FastICA package can be used to fulfil the ICA dimension reduction. There are two main parameters in ICA, the n_component and random_state. Whenever randomization is part of the scikit-learn algorithm, a random_state parameter may be provided to control the random number generator used. The random_state is a random seed, because many algorithms in sklearn contain random factors, in order to perform repeatable training and comparing each classification effect of different models, a random_state need to be fixed. In the coffee tree data, set the random_state equals to 40 and in mangroves tree data, set the random_state equals to 100. In fact, this random_state can be changed to any number. The other more important parameter is the n_component, which is the number of features that a transformer should transform the input into. A too large n_component may be slow in training or almost have no improvement in the classification effect compared with smaller n_component. Meanwhile a too small n_component may loss some characteristic of the wavelength, which may cause a worse classification effect. Therefore it is important to design a suitable value of n_component.
In order to find the suitable value of n_component, compare and find the best test accuracy under different components from 2 to 20 in coffee tree data and from 2 to 25 in mangroves tree data. The 20 and 25 can be changed to any larger value, but in fact they are enough to do the comparison in this project.

4C. Xgboost model

The Xgboost(Tianqi Chen, 2014) is a boosting(Kearns, 1988) method, which improves the effect by adding many models with suitable weight. Furthermore, Xgboost model is a lifting algorithm, which uses the second order Taylor expansion for loss function derivative information in the optimization. Compared with the traditional model GBDT, Xgboost model not only converges more quickly but also add a regular term in the loss function of XGBoost to control the complexity of the model, which can provide over fitting. The practical derivation process of the XGBoost algorithm(LI Chao, 2019) is not the main work in this project, here is just a simple explanation.

Main Idea of Xgboost

The basic idea of the XGBoost algorithm is to continuously perform feature splitting to grow a tree, and learning a tree in each round is actually to fit the residual between the predicted value and the actual value of the previous round of the model. When finished training and get k trees, the score of a sample need to be predicted. According to the characteristics of this sample, each tree will fall to a corresponding leaf node, and each leaf node corresponds to a score, and finally add up the scores

corresponding to each tree to get the predicted value of the sample.

Objective of Xgboost

Each round of training a tree is to minimize the loss function, which aims to minimize the residual between the predicted value and the actual value. In order to minimize each objective function, the derivative can be set to 0. Therefore the most predictive score of each leaf node and the function of tree models can be obtained. A detail of the prove of mathematics method can be found in article "XGBoost: A Scalable Tree Boosting System"(Tianqi Chen, 2016), which is not the main contest in this report.

$$
\begin{aligned}
L(y, \hat{y}) &= \sum_{i=1}^{M} L(y_i, \hat{y}_i) + \Omega(f) \\
&= \sum_{i=1}^{M} L((y_{i,0}, y_{i,1}, \ldots y_{i,N-1}), (\hat{y}_{i,0}, \hat{y}_{i,1}, \ldots \hat{y}_{i,N-1})) + \Omega(f) \\
&= -\sum_{i=1}^{M} \sum_{n=0}^{N-1} y_{i,n} \log\left(\frac{e^{\hat{y}_{i,n}}}{\sum_{c=0}^{N-1} e^{\hat{y}_{i,c}}}\right) + \Omega(f)
\end{aligned}
$$

Equation 1: the loss function(objective function) of Xgboost model

M is the number of samples, N is the samples of classes, yi,t**(i is the order of sample, t is the class) is the actual possibility of sample i in class t, which equals to 1 or 0.**

$\widehat{y_{i,t}} = \sum_{1}^{N} Stjl$ **(j is n_estimators) is the predicted possibility of sample i in class t.**

In objective function, first term is the loss and second term is the regularization term, which is used to avoid over fitting. In order to calculate **f**mj(decision tree model), the predicted score in each class, the package "xgboost" with python can be used. But indeed, the sklearn and "xgboost" package in python can directly get the predicted results of each sample without any calculation.

Simple Process of Xgboost Model

In last section, it is known that using the package "xgboost", the decision tree and score can be calculated, this section will show the way of using the predicted scores, which in the leaf nodes, to get the possibility of each of the sample in each class .

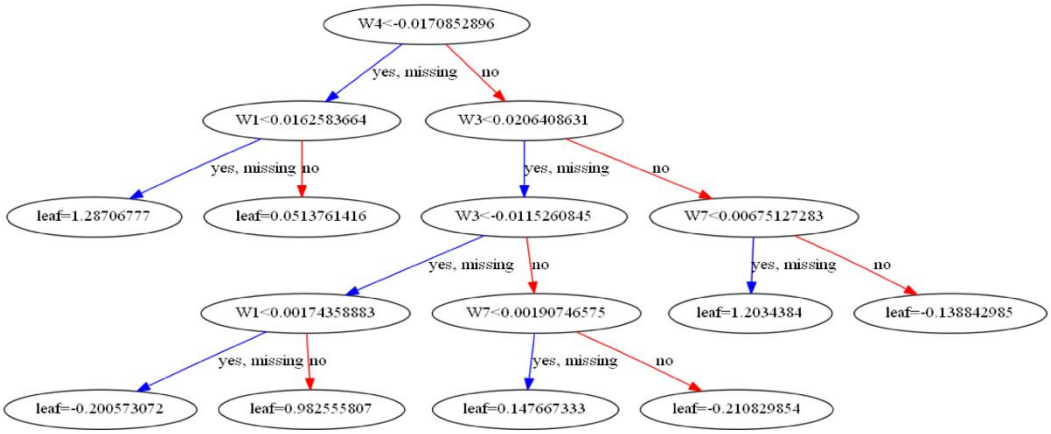

Figure 5: example of decision tree for the coffee data(W1-W10 are input's features)

In Figure 5, each tree model splits for many times, but the splitting algorithm of is

not the main task here, which can be seen in Tianqi Chen's easy. There are many tree models with inputs and outputs, each input data(sample/residual) will get an output, which is the corresponding score of leaf. Except the tree in first round, all the input in round t are the residual in last term, which can be calculated by the equation 2.

$$y_i - f_1(x_i) - f_2(x_i) - \ldots\ldots - f_t(x_i)$$

Equation 2: the residual in t round, ft **is the tree model in round t**

Take a example of coffee tree data(7 classes) in order to show the way of using each input's output got from the tree to get the predicted possibility of each class, each sample should be calculated for 7 times, which is the number of classes. Considering class 1, the Xgboost model is consisted of tree models from Tree 1 to Tree (n-1)+1, where n is number of rounds. For sample i, each of the tree model's leaf node gives a score **S**1ji, where j is number of rounds, **S**1i equals to the sum of **S**1ji, similarly scores of **S**2i, **S**3i, **S**4i, **S**5i, **S**6i, **S**7i are get. Therefore the probability of sample i in class t, where the range from 1 to 7, equal to possibility **Pi**(m)=**EXP**(**S**ti)/sum(**EXP**(**S**ti)). So, each sample was defined which classes they belongs to by largest possibility. In fact, in python code, predicted class label can be get directly without any calculating by XgboostClassifier, this section just aims to simply explain this classification process.

4D. Fulfil and Parameters Adjust of Xgboost

The "xgboost" package can be integrated with scikit-learn for Python users to fulfil the Xgboost classification. It is easy to training the the datasets with XgboostClassifier, however, in order to make it performs better, the parameters should be adjusted by python. Before considering on adjusting the parameters of Xgboost model, the Cross Validation method, which is frequently used in this process, should be introduced.

Cross Validation

When evaluating different settings for estimators on the experiment, there is still a risk of over fitting on the test set because the parameters can be tweaked until the estimator performs optimally. The knowledge about the test set cause the model and evaluation metrics no longer report on generalization performance. In order to solve this problem, the validation dataset was used to evaluate while it may lead to the reduction of the number of samples used on classifier models.

Cross-validation(CV) is a solution to deal with this problem. In this method, the training set is split into k smaller sets, the validation set can be removed and the test set should be kept to evaluate the classification effect. A model is trained for k times, each time the folds except the **k**th fold is used as the training set, the **k**th fold is used as the validation set, then perform the evaluation on the test set. The performance measure reported by cross-validation is the average values computed in the loop. To fulfil this method in python, the cross_val_score in package of model_selection can be used with sklearn.

Parameters introduction

In the classifier model of sklearn, there are some important parameters which define

the classifier model. In order to optimize the classifier model, these paremeter also need to be adjusted. The below table is the simple introduction of some important parameters of Xgboost model , which would be adjusted in the experiment.

Table 1: Introduction of Important Parameters of Classifier Model Xgboost

Max depth	Maximum tree depth for base learners	reg_lambda	L2 regularization term on weights
n_jobs	Number of parallel threads used to run xgboost.	Subsample	Subsample ratio of the training instance.
Num_class	Number of classes	n_estimators	Number of estimator
seed	Random state seed	Learning rate	Boosting learning rate
gamma	Minimum loss reduction required to make a further partition on a leaf	Min_child weight	Minimum sum of instance weight(hessian) needed in a child
objective	Specify the corresponding learning objective or a custom objective function to be used	Colsample bytree	Subsample ratio of columns when constructing each tree.

These parameters in Xgboost can be adjusted by GridSearchCV of model selection package in sklearn with python, which used the test accuracy as the index of the classification evaluation in this project. After adjusting the parameters to best values, a suitable Xgboost model can be found to train and evaluate the datasets.

4E. Summary and Steps of Experiment

From above sections, the best classifier ICA-Xgboost with suitable parameters can be found with the sklearn and package of “xgboost” in python. Then other classifier models and dimension reduction methods will be applied by classifier/ensemble packages in sklearn, which will compare the classification effects with ICA-Xgboost.

Table 2: A summary of experiment steps

Step 1	Remove the noise data and select suitable bands(wavelength) to study.
Step 2	Reduce the dimension with ICA using FastICA with sklearn
Step 3	Identify the suitable dimension by comparing the test accuracy under the Xgboost model with package of “xgboost” and sklearn.
Step 4	Adjust the parameters of Xgboost on considering higher value of test accuracy with GridesearchCV method by sklearn and package of “xgboost”
Step 5	Identify the suitable Xgboost model and evaluate its performance by seeing the test accuracy, kappa, F-score and confuse matrix with sklearn.
Step 6	Compare the classification effect of ICA-Xgboost model with other classifier models and dimension reduction methods with sklearn in python.

Table 2 is a summary of this experiment and then repeating this work in different bands of datasets to get their classification effect. So, this experiment can not only show the outstanding of ICA-Xgboost but also define suitable bands for classification.

5.Results

In this section, to enable the evaluate of the results of the tree species classification solution, we will produce the index of accuracy, Kappa, F-score and confuse matrix.

Table 3: Equation of index

accuracy	(True positive+False negative)/Total samples		
kappa	K= $(Po - Pe)/(1 - Pe)$	F-score	F=2*(precision recall)/(precision + recall) (average of each class with weighting)

Po is the empirical probability of agreement on the label assigned to any sample and Pe is the expected agreement when both annotators assign labels randomly.

5A. Coffee Tree Date

1. Dataset

Before fitting our models, first dealing with coffee tree dataset to remove the noisy data. The edge of wavelength 187nm-300nm and 900nm-1027nm are quite noisy so removed them and choose the wavelength 325nm-865nm in order to keep 1400 dimensions. There are 850 samples but 120 samples are unlabeled, so in this classification only 730 samples are needed indeed.

2. Classification Effect

The combinations of different dimension reduction methods, PCA, ICA, LLE and t-SNE and different machine learning models, Naive Bayes(NB), Boost Naive Bayes (BoostNB), SVM, Neural Network, Decision Tree, Boost Decision Tree, GBDT, Xgboost were tried. Each model combination was trained upon the dataset, which contained 730 useful samples with 1400 dimensions. Throughout the training, a standard 7:3 training/testing split was adhered to. The results of Naive Bayes(NB), Boost Naive Bayes (BoostNB), SVM, Neural Network combined with dimension reduction methods are showed below.

Table 4: the best dimension reduction methods based on some classifier models

	ICA-NB	ICA-BoostNB	t-SNE-SVM	DNN	ICA-DNN
accuracy	0.84	0.91	0.93	0.98	0.94
kappa	0.79	0.88	0.91	0.97	0.92
F-score	0.84	0.91	0.93	0.98	0.94

For different models, they may suitable for different dimension reduction methods. For coffee data, for example, independent components analysis(ICA) is suitable for Naive Bayes model and Boost Naive Bayes model. However, it seems for Neural Network, we needn't to reduce dimension. In this table, the Neural Network provides the largest F-score(average), Kappa and accuracy of 0.98, 0.97, 0.98 respectively.

Decision Tree model is another kind of classification method, bagging and boosting are the improvement methods it, which can get some new models like, Random Forest(RF), Adaboost, Gradient Boosting Decision Tree(GBDT), Extreme Gradient Boosting Decision Tree (Xgboost). The finding is among PCA, ICA and LLE, the ICA seems the best choice for all these models after applying different kinds of dimension reduction methods, which showed in Table 5.

Table 5: dimension reduction method based on ensemble tree models

	ICA-Tree	ICA-RF	ICA-Adaboost	ICA-GBDT	ICA-Xgboost
accuracy	0.89	0.96	0.96	0.965	0.97
kappa	0.86	0.95	0.95	0.96	0.96
F-score	0.89	0.96	0.96	0.965	0.97
	PCA-Tree	PCA-RF	PCA-Adaboost	PCA-GBDT	PCA-Xgboost
accuracy	0.89	0.92	0.91	0.88	0.90
kappa	0.87	0.90	0.89	0.85	0.87
F-score	0.89	0.92	0.91	0.88	0.90
	LLE-Tree	LLE-RF	LLE-Adaboost	LLE-GBDT	LLE-Xgboost
accuracy	0.89	0.91	0.95	0.92	0.94
kappa	0.86	0.88	0.94	0.90	0.92
F-score	0.89	0.91	0.95	0.92	0.93

The Table 5 shows for all these tree models, the ICA is the superior dimension reduction method for solving coffee tree classification problem, with the almost largest F-score(average), Kappa and accuracy, which are the statistics that provide the best summary of performance. This table also shows the performance of Xgboost model is the best, its accuracy is around 97.5%, the kappa is around 97% and the F-score is around 97.5%.

3. Speed and Parameter adjust

In order to compare the effects, first consider the speed these two models. Seeing the time of classification for Neural Network, 18.1 seconds, seems a little slow. If reduce the dimension with ICA method, the classification time can drop to 1.75 seconds but the classification effect also dropped. Then see the classification speed of ICA-Xgboost model, it only need 0.39 second, with accuracy 97%, kappa 96%, and F-score 97%. Now compare these two models we can find the classification effect of Neural Network is a little better than ICA-Xgboost, but the running speed of Neural Network is much slower. When training more samples or wide wavelength, this drawback will be quite obvious. Therefore improving the classification effect of ICA-Xgboost through adjusting the parameters is the next step of the work.

First for dimension of ICA , adjust the n_components in FastICA to a suitable value.

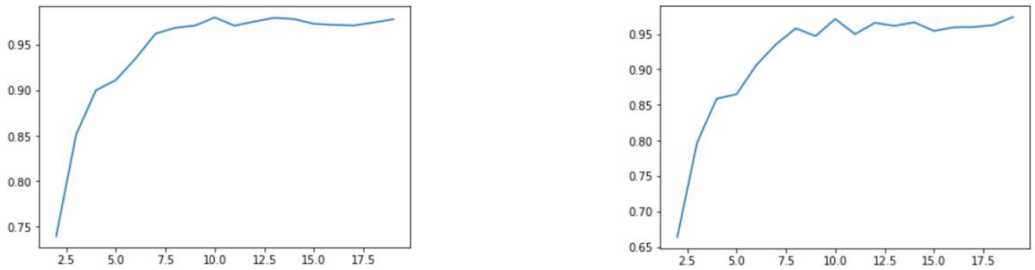

Figure 6: left plot 1 tested by average of split_train_test, right plot 2 tested by CV.

Plot 1 use the average of train_test_split method and Plot 2 use the CV method, both of them show the dimension 10 is a good choice. Since for dimension below 10, the test accuracy keep rising, while after 10 dimensions, test accuracy fluctuates. The

plots show the test accuracy in different dimensions of ICA-Xgboost model.
Then for Xgboost, there are some important parameters to be adjusted in order to improve the classification effect. Table 6 shows test accuracy effect after adjusting.

Table 6: classification effect of parameters adjust(before and after)

		Number of components	Max Depth/ Child Weight	Gamma	Subsample/ Colsample Bytree	Learning Rate
Test Acc	Change	From 30 to 86	Max Depth = 4 / Child Weight = 1 (no change)	0.3 - 0.1	Subsample: 0.8 - 0.6 / Colsample Bytree: 0.8 - 0.7	0.4 (no change)
	Before Change	96.8%	97.26%	97.26%	97.4%	97.8%
	After Change	97.26%	97.26%	97.4%	97.8%	97.8%

From Table 6, the test accuracy rising 1% after changing the parameters. Now the effect of ICA-Xgboost is almost the same as Neural Network, but the running speed is much faster than the Neural Network. Therefore the ICA-Xgboost is the suitable classification model which is defined to use.

4. Confuse Matrix

Aims to see the classification effect of ICA-Xgboost model in each class better, then consider the test confuse matrix of coffee tree data. The below Table 7 shows theconfuse matrix of ICA-Xgboost method for coffee data.

Table 7: Confuse Matrix of Coffee Tree Data

	Predict Health good	Predict Rust	Predict Infected Green Canopy	Predict Geisha	Predict Geisha Canopy	Predict Young Rust Yellow	Predict Young Rust Green
Actual Health good	53	0	2	0	0	0	0
Actual Rust	0	57	0	0	0	0	0
Actual Infected Green Canopy	1	0	31	2	0	0	0
Actual Geisha	0	0	0	45	0	0	0
Actual Geisha Canopy	0	0	0	0	17	0	0
Actual Young Rust Yellow	0	0	0	0	0	10	0
Actual Young Rust Green	0	0	0	0	0	0	1

There are 7 different classes with total 730 samples which are “health good” class, “rust” class, “infected green canopy” class, “geisha” class, “geisha canopy” class, “young rust yellow” class, “young rust green leaf” class. From this table, the three classes “Health Good”, “Infected Green Canopy”, “Geisha” are easier to make mistake, while other classes can classify well. In order to improve the classification effect, more samples may needed.

5B. Mangrove Tree

The main aim of training this dataset is to test the classification effect of ICA-Xgboost model, in order to know whether this method is suitable for the tree species classification task. Meanwhile, for this dataset, consider the classification effect for different range of wavelength to see if full range of wavelength can perform better.

1. Dataset

In the dataset of mangrove trees, first need to remove the noisy data. Wavelength 350nm-400nm and 2400nm-2500nm are the edge of wavelength, so remove them to avoid the noise. Then the below plots are the mean and the variance of all these samples.

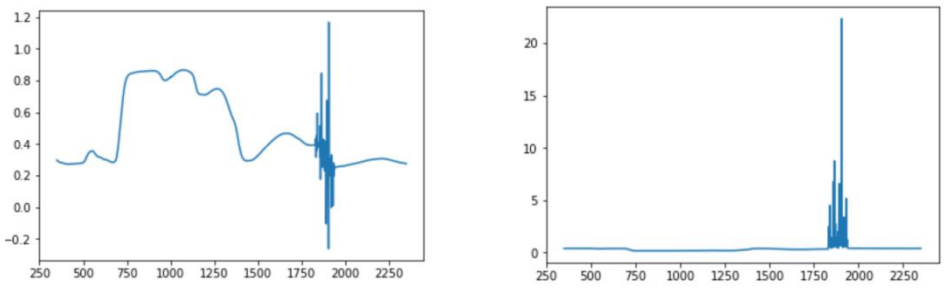

Figure 7: left plot is the mean of mangroves, right plot is the variance of mangroves

Consider the range of 400nm-900nm and 400-2400nm, the range 1800nm-2000nm is very noise for the reason of mud class. So this range of the data was removed in the work and then the two ranges are chosen as below:

range 1(R1): 400nm-900nm range 2(R2): 400-1800nm, 200nm-2400nm

For range 1, there are 841 samples with 500 dimensions, and for range 2, there are 841 samples with 1800 dimensions.

2. Classification Effect

The combinations of different dimension reduction methods, PCA, ICA, LLE , and different machine learning models, Naive Bayes(NB), Boost Naive Bayes(BoostNB), SVM, Neural Network, Decision Tree, Boost Decision Tree, GBDT, Xgboost were still tried for the dataset. Each model combination was trained upon the dataset, which contained 841 samples. Throughout training, a standard 7:3 training/testing split was adhered to.

For range 1, two tables are shown below in order to see the benefits of dimension reduction methods. Table 8 is the classification effect of different classifier model before reducing dimension, Table 9 is the classification effect of different classifier models after using the best dimension reduction methods.

Table 8: classification effect of different models(R1)

	NB	Boost NB	SVM	DNN	DT	Boost DT	GBDT	RF	Xgboost
accuracy	0.58	0.64	0.79	0.93	0.76	0.81	0.78	0.75	0.76
kappa	0.36	0.47	0.69	0.90	0.65	0.72	0.68	0.63	0.66
F-score	0.51	0.59	0.78	0.93	0.76	0.80	0.78	0.75	0.76

Table 9: combine of best dimension reduction method with different models(R1)

	LLE-NB	ICA-Boost NB	PCA-SVM	ICA-DNN	ICA-DT	ICA-Boost DT	ICA-GBDT	ICA-RF	ICA-Xgboost
accuracy	0.78	0.93	0.75	0.96	0.80	0.95	0.92	0.87	0.97
kappa	0.66	0.89	0.64	0.94	0.70	0.94	0.88	0.80	0.96
F-score	0.78	0.92	0.75	0.96	0.80	0.95	0.92	0.87	0.97

The Table 8 and Table 9 shows the dimension reduction can largely improve the classification effect in terms of accuracy, kappa and F-score, except the Neural Network and SVM. Some models like Nave Bayes and ensemble tree models even improved 20% in these three index.

Then consider range 2, seeing another two tables, Table 10 and Table 11, which is the effects of wavelength 400nm-1800nm, 2000nm-2400nm for different models. Then, analysis whether using larger wavelength can classify the samples better than smaller wavelength and whether the suitable reducing dimension methods can remove the classification effect gap of two different wavelength range.

Table 10； classification effect of different models(R2)

	NB	Boost NB	SVM	DNN	DT	Boost DT	GBDT	RF	Xgboost
accuracy	0.71	0.92	0.95	0.97	0.86	0.92	0.90	0.87	0.96
kappa	0.59	0.89	0.93	0.96	0.79	0.88	0.85	0.81	0.95
F-score	0.71	0.92	0.95	0.97	0.86	0.92	0.90	0.87	0.96

Table 11: combine of best dimension reduction method with different models(R2)

	ICA-NB	ICA-Boost NB	LLE-SVM	ICA-DNN	LLE-DT	ICA-Boost DT	ICA-GBDT	LLE-RF	ICA-Xgboost
accuracy	0.88	0.90	0.91	0.95	0.88	0.95	0.94	0.93	0.97
kappa	0.84	0.85	0.87	0.92	0.83	0.93	0.91	0.90	0.96
F-score	0.87	0.90	0.91	0.94	0.88	0.95	0.94	0.93	0.97

From Table 10 and Table 11, reducing dimension can improve the classification effect for most of the models, except Neural Network, Boost Naive Bayes and SVM, which is almost the same as the the smaller wavelemngth range 400nm-900nm. In addition, ICA-Xgboost is still a good classifier model for mangroves tree samples although

dimension reduction method can just improve the classification effect a little, especially compared with the range 1(400nm-900nm).

Compare the Table 8 and Table 10, if not reduce the dimension, the larger range of wavelength can classify the mangrove tree species better, while seeing the Table 9 and Table 11, the dimension reduction method can reduce the gap of classification effect between the two range, especially under the ensemble tree models. In addition, force on the Xgboost model, and consider the classification effect after adjusting the parameters from the wavelength range of 400nm-900nm to the wavelength range of 400nm-2400nm, the accuracy rise from 76% to 96%, the kappa rise from 66% to 95% and the F-score rise from 76% to 96%. While if using dimension reduction methods like ICA to reduce the dimensions to a suitable number, like 23 or others, the accuracy, kappa and F-score almost have no change under the two different ranges, with 97%, 96% and 97% respectively.

3. Confuse Matrix

Just like the coffee tree data, after adjusting the parameters to suitable value, the confuse matrix shows the classification results in each class. The mangroves tree samples have 4 different classes of colors, the “Black” class, the “Mud” class, the “Red “ class and the “White” class with 841 samples and 23 dimensions for each sample after applying the ICA method. Confuse matrix can show as the classification effect of the test samples, where the ratio of train/test is 7/3. In order to see whether there are some difference between the two ranges with the ICA-Xgboost, the two tables Table 12 and Table 13 of the confuse matrix are shown below.

Table 12: confuse matrix of range 400nm-900nm

	Predict Black	Predict Mud	Predict Red	Predict White
Actual Black	93	0	0	0
Actual Mud	0	11	0	1
Actual Red	1	1	83	2
Actual White	1	0	2	58

After adjust the parameters: Accuracy=96.8%,Kappa=95.4%,F-score=96.8% time=0.5s

Table 13: confuse matrix of range 400nm-1800nm and 2000nm-2400nm

	Predict Black	Predict Mud	Predict Red	Predict White
Actual Black	96	0	0	1
Actual Mud	1	8	0	0
Actual Red	1	0	78	1
Actual White	1	0	1	65

After adjust the parameters: Accuracy=97.6%,Kappa=96.5%,F-score=97.6%,time=0.5s

The Table 12 and Table 13 show the effect of confuse matrix almost the same, the larger wavelength range just performs a little better(less tan 1%) than the small wavelength range. Both of the two confuse matrix showed that the classification of these four classes can make some mistakes, but there are no any two classes more likely to make the mistake. Therefore the ICA-Xgboost model for mangrove trees is a quite good after adjust the parameters, since not only the code runs fast, within 1 second, but also classification effect is good with accuracy around 97%, kappa around 96% and F-score around 97%. Meanwhile, with ICA-Xgboost model, both the

large and small wavelength range have the similar classification effect. Although the classification can sometimes make some mistakes, adding more samples to each class may improve the effect.

5C. Summary

After classifying the coffee and mangroves tree datasets, there are some main findings. Firstly, it is necessary to remove the noisy data in edge of wavelength and reduce its dimension. Dimension reduction method, especially the ICA method, can indeed improve the classification effect for different models, except the Neural Network. Secondly, lager range of wavelength performed better than the smaller range of wavelength, but when reducing the dimension(especially same dimensions), the gap of the classification effect will be reduced. Thirdly, the Xgboost is a good classifier model after applying the ICA dimension reduction method, the accuracy can reach 97%, kappa can reach 96% and the F-score can reach 97% within 1 second of training time for both of the mangroves tree and coffee tree data.

6.Evaluate

In this section, an evaluation of solution is given by using the results presented in the previous section. The strength and the weakness of the solution and results are discussed and some new ideas of other possible classifier models are put forward.

6A. Strength

1.Benefits of ICA-Xgboost model
The results in last section proves that Xgboost is a very appropriate model combined with ICA dimension reduction method for tree species classification. The ICA method can provide some independent components, which is useful to remove the effect of noisy data and reduce the training time. The only model can be compared with ICA-Xgboost is Neural Network, but after considering the train time, Xgboost showed its benefits. The other largest benefit of ICA-Xgboost model is that the adjust of parameters of Xgboost model is much easier than other models by using GridSearch.
2.Benefits of the design
The datasets of tree species must have some noisy data, especially in the begin and the end of the wavelength, which be seen in the coffee tree samples clearly. The smaller wavelength range 400nm-900nm seems have same classification effect compared with the full wavelength range 400nm-2500nm after reducing noise data and applying suitable dimension reduction methods like ICA. Therefore smaller wavelength range can be used to reduce the workload.
3.Adjust of models
In this work, consider four kinds of models, SVM, Naive Bayes model, Neural Network model and Tree model. The finding is that the classification effect of Naive Bayes model and Tree model is not good enough, so the ensemble methods like bagging

and boosting were used to improve, which can rise the accuracy, kappa and F-score for over 5% on both mangrove tree and coffee tree data .

6B. Weakness

1.Choose of Dataset
The first limitation is the lack of enough datasets to prove the conclusion to evaluate the classification effect of ICA-Xgboost. The second limit is that there are some classes in the coffee data with few valid samples in the coffee tree data.
2.Weakness of Classifier Model Design
There are still some weakness of the design of models. The first problem is about the feature reducing. Many dimension reduction methods were considered, but still miss many other methods, like LDA, SVD, Isomap and so on. Therefore maybe there are some other better dimension reduction methods to apply. The second problem is about the parameters of the different models, especially the SVM model and the Neural Network. For SVM model, the kernel is different to design, which has four main kinds of kernels, linear, sigmoid, Gaussian and polynomial. Although most of the time, the Gaussian kernel has a better effect, there is still no enough evidence to prove polynomial kernel is worse than the Gaussian kernels in tree species datasets. The other problem of the SVM model is the suitable range of parameter gamma. A suitable range of gamma should be set when using the grid search to define the best value. If the value of gamma is too large, the impact of the sample is very small, which may cause a meaningless model. For the Neural Network, the hidden layer can be difficulty to design since unable to use the grid search to find suitable parameters.
The last limitation of the classifier models is that not try enough methods to improve the classifier models, especially for Xgboost, whose improvement called LightBGM (Guolin Ke, 2017). The LightBGM can significantly outperform Xgboost in terms of computational speed and memory consumption. In addition, the SVM tree model showed better performance as the group size becomes larger compared with SVM ensemble methods(Shaoning Pang, 2005).

6C. Improvement

If given a chance to repeat the project, the extension objective of this project, some other range of wavelength(band) can be considered. In addition, the design of SVM model and DNN should be more careful on their parameters. The new classifier methods like LightBGM and SVM tree can be taken into consideration to see whether it performs better than ICA-Xgboost model.
Another idea to improve the classification effect is using both spectral information and geographical ancillary data such as the shape index, spatial location information, etc to fulfil the work. Decision tree classification was used by Jinfa Dong in the year of 2012, which found if only use the spectral information, the test accuracy is only 50% while using both the spectral information and the geographical ancillary data, the test accuracy up to 70%.

7.Conclusion

In this project, by successfully fulfilling the above objectives, this work found that the ICA dimension reduction method combined with Xgboost performs well. In the coffee tree dataset, the accuracy, kappa, F-score reach 97%, 96% and 97% respectively. In the mangrove tree dataset, the larger range of wavelength after removing the noisy data performs better than smaller range, while after applying suitable dimension reduction methods like ICA on Xgboost model, the classification effect is almost same for these two different wavelength range, which means dimension reduction method can remove the gap classification effect. Both the range of 400nm-900nm and 400nm-2400nm for accuracy, kappa, F-score equals about 97%, 96% and 97% respectively. In this project, the way of fulfilling this method would be shown and comparing its classification effect with other classifier models.

In terms of fulfilling the aims, the project was an overall success. The tree species classification by using ICA-Xgboost approach was implemented and verified with coffee tree and mangroves tree dataset. The comparing with other machine learning models proved that ICA-Xgboost is best. Moreover, ICA largely improve classification effect, up to of test accuracy round 20% and reduced the training time. Therefore it is useful to reduce the dimension of the given wavelength, the only problem is that the noisy data should be removed before using ICA. The other successful finding is that although larger range of wavelength performed better, machine learning methods like ICA-Xgboost can reduce the gap of classification effect. There is no need to detect whole wavelength range with much the workload. Meanwhile, the edge of detected wavelength are often found noisy, removing them and selecting a suitable wavelength range before classification are needed.

One suitable direction for future work is to change the Xgboost to some improved classifier models like LightBGM, and compare the effects of different wavelength ranges to define a better band. Another suitable direction is using both geographical ancillary data and wavelength data for tree classification. This project allows future development of bespoke systems that enable non-specialists to identify tree species using lightweight multi-band cameras which can be fitted to drones for field surveys.

Reference

[1]. Fabian Ewald Fassnacht, Hooman Latififi, Krzysztof Stereńczak, Aneta Modzelewska, Michael Lefsky, Lars T. Waser, Christoph Straub, Aniruddha Ghosh, 2016, Review of studies on tree species classifification from remotely sensed data. Remote Sensing of Environment.

[2]. Le Wang, Wayne P. Sousa, 2009, Distinguishing mangrove species with laboratory measurements of hyperspectral leaf reflectance. International Journal of Remote Sensing, pp1267-1281.

[3]. Matthew L. Clark, Dar A. Roberts, 2012, Species-Level Differences in Hyperspectral Metrics among Tropical Rainforest Trees as Determined by a Tree-Based Classifier. Remote Sensing.

[4]. https://scikit-learn.org/stable/

[5]. Yushi Chen, Zhouhan Lin, Xing Zhao, Gang Wang, Yanfeng Gu, 2014, Deep Learning-Based Classification of Hyperspectral Data.
[6]. N. Fitriah, S. K. Wijaya, M. I. Fanany, C. Badri, M. Rezal, 2017, EEG channels reduction using PCA to increase XGBoost’s accuracy for stroke detection
[7]. Liwei Fan, Kim-Leng Poh, Peng Zhou, 2009, A sequential feature extraction approach for naïve bayes classification of microarray data. Expert Systems with Applications
[8]. Junghoe Kim, Vince D. Calhoun,Eunsoo Shim, Jong-Hwan Lee, 2015, Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia. NeuroImage. pp127-146
[9]. Rui Sun, Guanyu Wang, Wenyu Zhang , Li-Ta Hsu, Washington Y. Ochieng, 2020, A gradient boosting decision tree based GPS signal reception classification algorithm, Applied Soft Computing Journal.
[10]. Selçuk BAYRACI, Orkun SUSUZ, 2019, A Deep Neural Network (DNN) based classification model in application to loan default prediction, Theoretical and Applied Economics, pp75-84
[11]. Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, Tie-Yan Liu, 2017, LightGBM: A Highly Efficient Gradient Boosting Decision Tree
[12]. Stijn Viaene, Richard A. Derrig, and Guido Dedene, 2004, A Case Study of Applying Boosting Naive Bayes to Claim Fraud Diagnosis
[13]. LI-MM WANG, SEN-MIA YUAN, LING LIZ, HAI-JUN LIZ, 2004, BOOSTING NAIVE BAYES BY ACTIW LEARNING, Proceedmgs of the Third International Conference on Machine Learning and Cybernetics, pp26-29.
[14]. Sejong Yoon and Saejoon Kim, 2009, Multiple SVM-RFE using Boosting for Mammogram Classification, International Joint Conference on Computational Sciences and Optimization, pp740-742.
[15]. L. Kanaan, D. Merheb, M. Kallas, C. Francis, H. Amoud, P. Honeine, 2011, PCA AND KPCA OF ECG SIGNALS WITH BINARY SVM CLASSIFICATION, pp344-348.
[16]. Jinfa Dong, Qingsheng Liu, Gaohuan Liu, Wenming Shen, Dan Huang, 2012, Extracting Fly Ash Site Information Using Decision Tree Classification, International Congress on Image and Signal Processing, pp429-433
[17]. Carlos F. Navarro and Claudio A. Perez, 2019,Color－Texture Pattern Classification Using Global－Local Feature Extraction, an SVM Classifier, with Bagging Ensemble Post-Processing. applied sciences.
[18]. Pan Huang , Yanping Li, Xiaoyi Lv, Wen Chen and Shuxian Liu, 2020, Recognition of Common Non-Normal Walking Actions Based on Relief-F Feature Selection and Relief-Bagging-SVM, sensors
[19]. Jing Wang, Chein-I Chang, 2006,Independent Component Analysis-Based Dimensionality Reduction With Applications in Hyperspectral Image Analysis.
[20]. A.Hyvarinen and E.Oja, 2000, Independent Component Analysis: Algorithms and Applications, Neural Networks
[21]. Abdulhamit Subasi, M. Ismail Gursoy，2010，EEG signal classification using PCA, ICA, LDA and support vector machines，Expert Systems with Applications
[22]. LI Chao, ZHANG Wen-hui, LIN Ji-ming, Study of Star/Galaxy Classification Based on the XGBoost Algorithm, Chinese Astronomy and Astrophysics, pp539－548