**Supplementary Note Data processing process of this research**

1. **Hyperspectral indices extraction**

The programs using for hyperspectral indices extraction of 533 *O. sativa* accessions

is same as that in the published article (<https://github.com/fenghuifh2006/Maize-RGB-CT-HSI-program>)

1. **Metabolite levels extraction**

The metabolites data of 533 *O. sativa* accessions recorded were processed with LabSolutions 5.91 software. The data matrix was *log2* transformed.

1. **Pearson correlation coefficient calculation between hyperspectral indices and metabolites**

The Pearson correlation coefficient calculation between hyperspectral indices and metabolites was calculated using the built-in function ‘cor’ in the R4.2.1 environment. The method “pearson” was selected and the parameter “pairwise.complete.obs” was used for missing value processing.

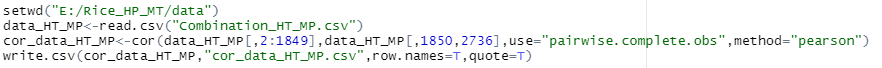


Figure 1 The code of calculating pearson correlation coefficient

1. **Eight machine learning models**

Eight machine methods including PLSR, LGBM, LASSO, RR, CNN, SVM, RF, SLR Regression were used for metabolites prediction. The codes of these methods were established in Python 3.8 environment based on sklearn module.

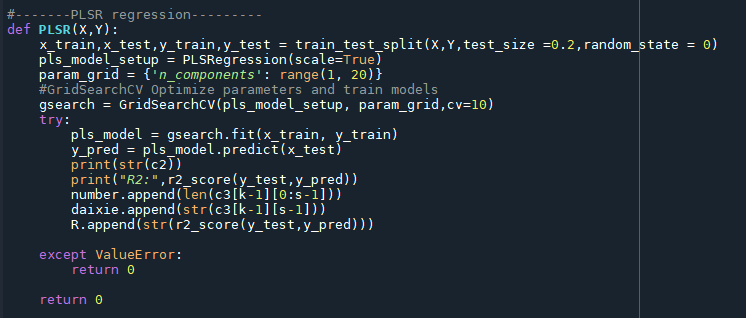


Figure 2 The function of PLSR regression

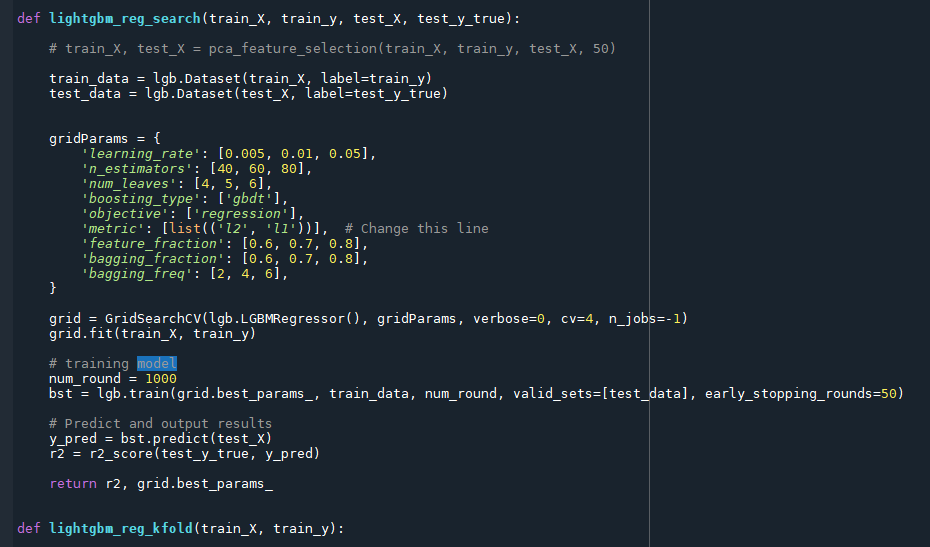


Figure 3 The function of LGBM regression

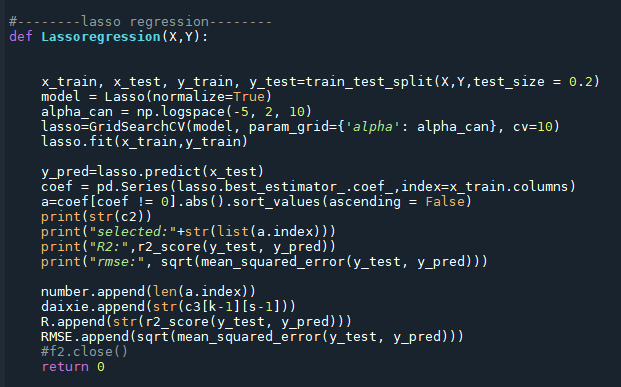


Figure 4 The function of LASSO regression

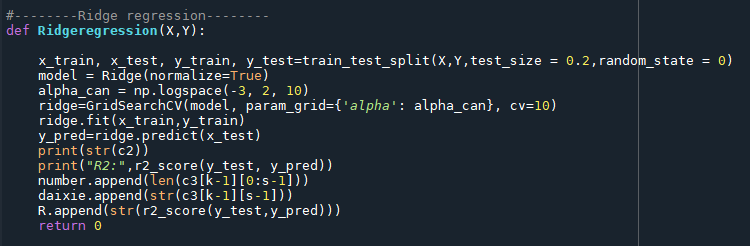


Figure 5 The function of RR regression

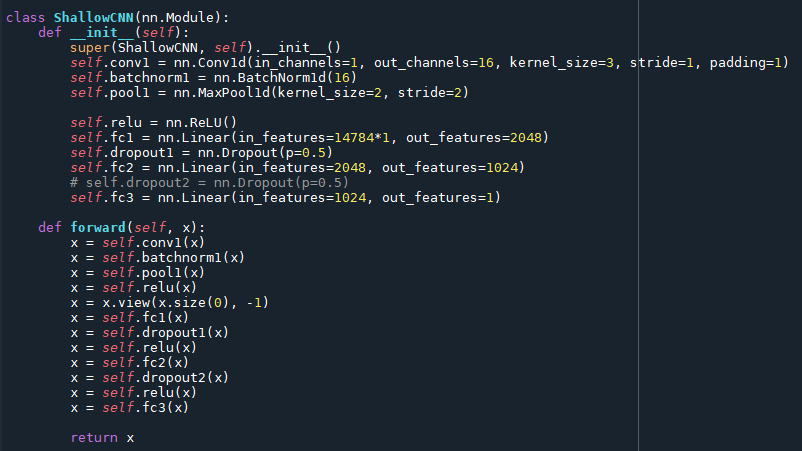


Figure 6 The function of CNN regression

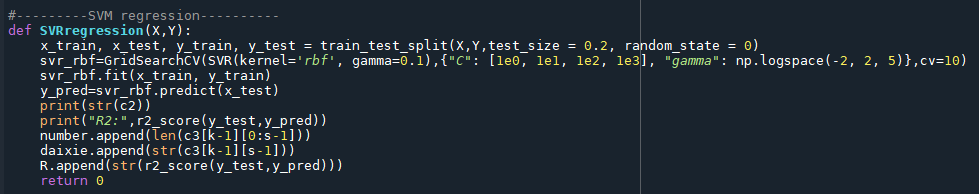


Figure 7 The function of SVM regression

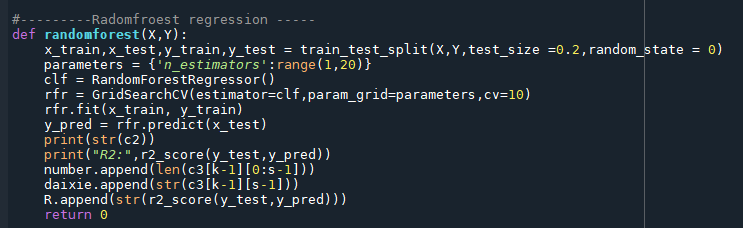


Figure 8 The function of RF regression

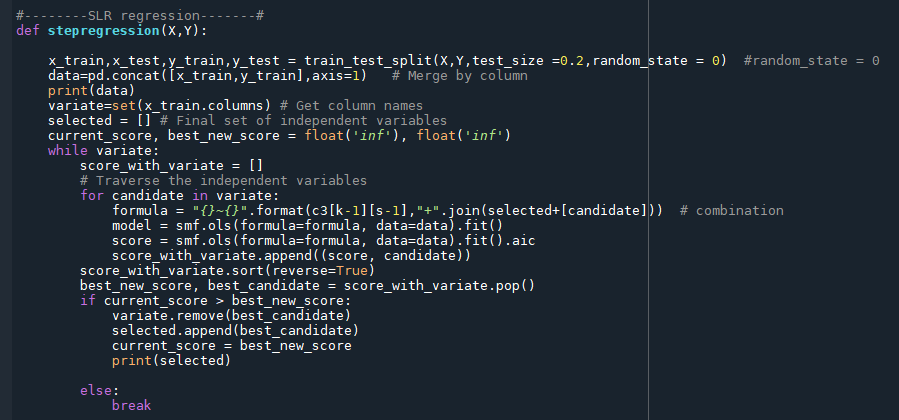


Figure 9 The function of SLR regression

1. **Co-localization analysis**

Co-localization analysis was performed with the screened locus of hGWAS and mGWAS according to the distance of the leading SNPs, while the threshold was 300kB within the chromosome.



Figure 9 The Python script of colocalization analysis

1. **Gene expression pattern selection**

The co-localized locus with expression level higher than 200 in heading stage of rice panicles or 7-21 days after pollination of rice endosperms were reserved for further analysis.

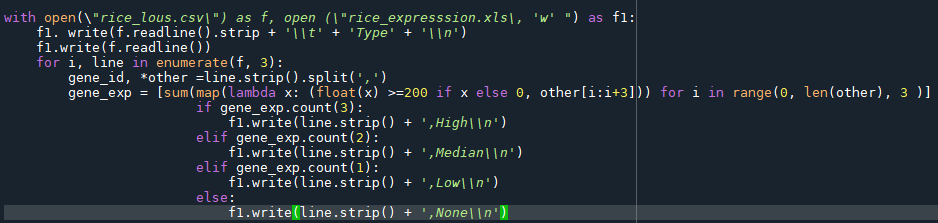


Figure 10 The Python script of Gene expression pattern selection

1. **KEGG keywords mapping**

Several keywords were attached to each metabolite group, and the co-localized locus of corresponding metabolites having these keywords were screened.

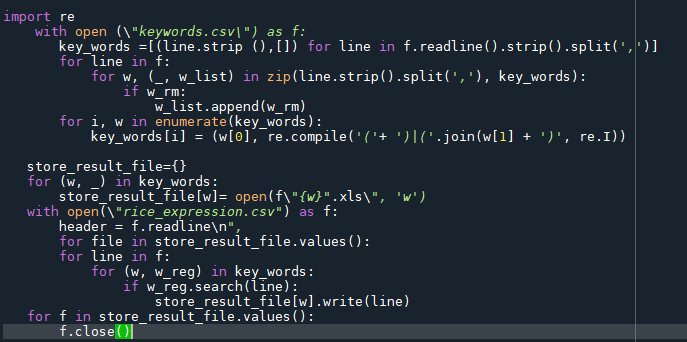


Figure 11 The Python script of KEGG keywords mapping