

Hyperspectral Image Analysis Pipeline

**Busov I.D., Genaev M.A., Komyshev E.G., Zykova T.E.,
Afonnikov D.A.**

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

E-mail: i.busov@g.nsu.ru

The analysis of hyperspectral images is of great interest in the problems of plant research. Hyperspectral technologies are used to assess the concentration of mineral elements in wild barley grain (Herzig P. et al., 2019), to classify viable and non-viable wheat seeds (Zhang T. et. al., 2018), to identify barley germinated at an early stage (Arngren M. et al., 2011). This work presents a hyperspectral image analysis pipeline. The Python pipeline includes: basic statistical analysis, visualization of a multichannel hyperspectral image, solving a classification problem using classical machine learning methods. Methods for constructing a confidence interval for the difference in sample means, checking the equality of distributions of two classes of hyperspectral images based on Pearson's chi-square test, visualization in two-dimensional space using PCA and Umap, classification using logistic regression or ridge regression, random forest and catboost were implemented. The developed pipeline was tested in the task of determining the pigment composition of barley grains. The results of the analysis showed significant differences in the studied target groups.

The work was supported by the Russian Science Foundation, the project 22-74-00122, <https://rscf.ru/project/22-74-00122/>