Title: A Comparative Study of Efficiency of Algorithms in Machine Learning in Image Classification task for microscopic pictures with Limited data

Authors: Supawit Marayat, Archirawit Prasom

Supervisors: Acting Sub Lt. Satit Thamkhanta

Mrs. Manatchanok Tamwong

School: Princess Chulabhorn Science High School Chiang Rai, THAILAND

Abstract

The purposes of this study were to study and compare Algorithms used in Machine Learning to create Image Classification, and to examine the process of Data augmentation to increase the limited data. The 5 steps were as follows: step 1) Data preparation was conveyed from the BCCD Data set, step 2) Data augmentation was utilized to increase the data by using Geometric and Color space transformations techniques, step 3) The training process was used to take data from pictures to be Vectors, and each Vector was labeled. And then it was used to train an Algorithm, step 4) Hyperparameter tuning program was used to find the best parameter for the Algorithms, and step 5) Evaluation process was used to predict the test set used with each Algorithm and then Metric scoring was used to save the results.

The results of the study showed that the data augmentation process could increase the amount of data from 352 to 10,028 images. The process of Training and Evaluating showed that an algorithm for image classification had the most performance was the eXtreme gradient boosting at 0.93 with F-measure at 77%, followed by the Light gradient-boosting machine. The co-experiment between Light gradient-boosting machine and gradient boosting was at 0.91 with F-measure at 74% and Cat Boost was at 0.87 with F-measure at 66% respectively.

Keywords: Machine learning, Data augmentation, Hyperparameter tunning, Image Classification