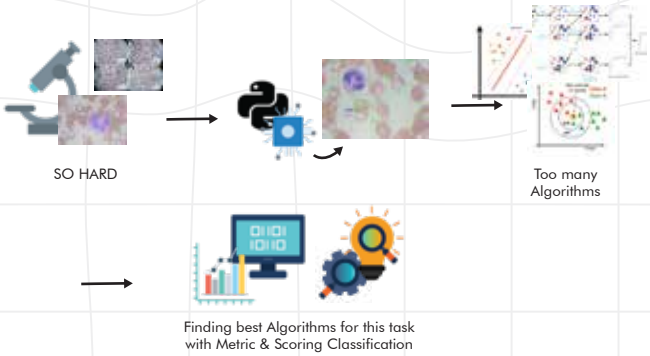
A COMPARATIVE STUDY OF EFFICIENCY OF ALGORITHMS IN MACHINE LEARNING

IN IMAGE CLASSIFICATION TASK FOR MICROSCOPIC PICTURES WITH LIMITED DATA

RESEARCHER : Achirawit prasom Supawit marayat
SUPERVISOR : Satit thamkhanta Manatchanok Tamwong

INTRODUCTION

PROBLEM AND SOLUTION

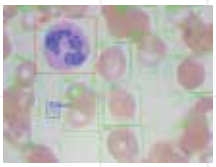


DATASET INTRODUCTION

BCCD

labels :
- RBC (Red Blood Cell)
- WBC (White Blood Cell)
- Platelets

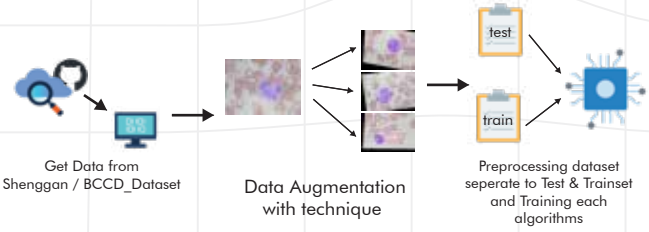
With Full Anotation

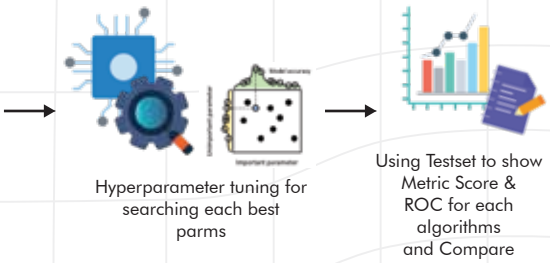


OBJECTIVE

1. compare Algorithms used in Machine Learning to create Image Classification
2. examine the process of Data augmentation to increase the limited data with unbalance problem.

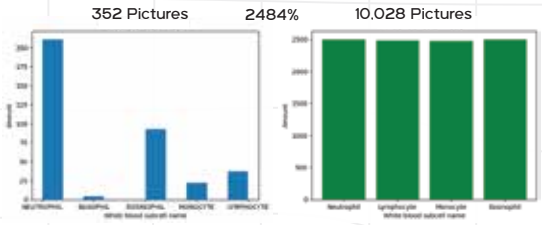
METHOD





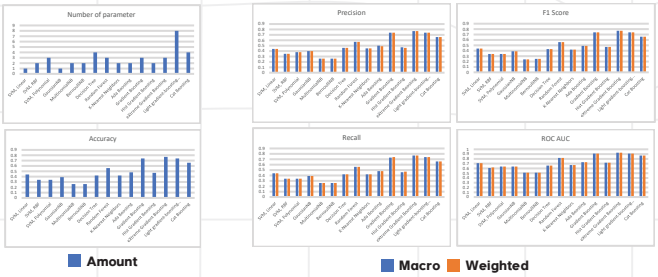
RESULT AND DISCUSSION

DATA AUGMENTATION

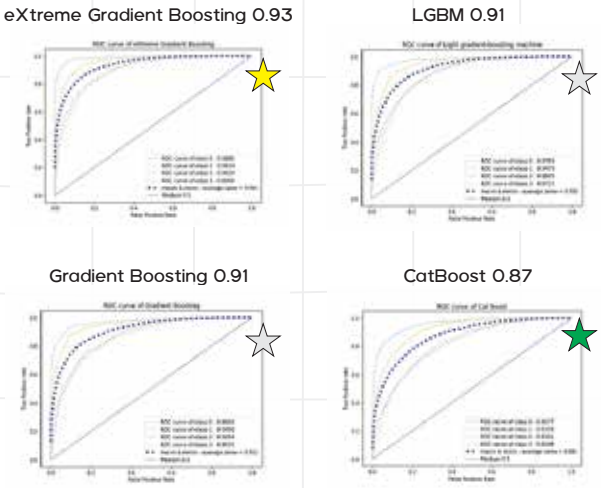


TRAIN AND EVALUATION

OVERVIEW SCORE



TOP 3 SCORE



DISCUSSION

According to the experimental results, using Data Augmentation with Geometric and Color space transformation data can increase data from 352 to 10028 with a ratio of 2484%. Histograms that compare before and after data augmentation and graphs and tables of Metric and Scoring that show Weighted and Macro of each score and each algorithm show with nearly score, indicating that the dataset is now balanced. In 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 Accuracy in 0.77, Precision average is 0.77, Recall average is 0.77, F-measure at 77%, Which performs better than Cat Boost and other algorithms that need more than three parameters and demonstrate that boosting base methods perform better than other algorithms group

CONCLUSION

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.

REFERENCE

- Liu, P., Choo, K.-K. R., Wang, L., & Huang, F. J. S. C. (2017). SVM or deep learning? A comparative study on remote sensing image classification. 21(23), 7053-7065.
- Shorten, C. and T.M.J.J.o.b.d. Khoshgoftaar, A survey on image data augmentation for deep learning. 2019. 6(1): p. 1-48.
- Liashchynskyi, P., & Liashchynskyi, P. J. a. p. a. (2019). Grid search, random search, genetic algorithm: a big comparison for NAS.