ICAI 2023
12th International Conference on Applied Informatics
Eger, Hungary, 2023

Cell Image Classification using Ensemble Approaches

Ahiya Ahammed[a], **Balazs Harangi**[b]

Faculty of Informatics, University of Debrecen
evan@mailbox.unideb.hu
harangi.balazs@inf.unideb.hu

Abstract

Since most of the real-life datasets are imbalanced, classification for those datasets is a difficult research challenge in machine learning. Emerging machine learning algorithms improve classification accuracy by correctly classifying the majority of classes; while incorrectly classifying the minority of classes. But in real-life implementations, minority class instances are more interested in representing the principle than the majority class instances. For classifying those unbalanced datasets, multiple approaches focused on sampling procedures such as (under-sampling of the majority class/over-sampling of the minority class), Synthetic Minority Oversampling Technique (SMOTE) and Class-weights and a few cost-sensitive learning methods have currently been used in textbooks in various researches [4].

In this paper, we introduced some Ensemble methods using a customized loss function called Weighted Categorical Cross-Entropy with several data balancing methods such as Under-sampling and Class-weight and built some more powerful classifiers for the unbalanced cell image datasets. Three types of design model architectures have been used in this research such as *ResNet50*, *VGG16* and *MobileNetV2*. We used a highly unbalanced binary-class cell image dataset where the target values are *Healthy* and *Unhealthy* containing *8,738* and *2,134* cell images respectively, and the imbalance ratio of this dataset is *4.09* which is used for model training. Afterwards, we used *2,184* Healthy cell images and *533* Unhealthy cell images for testing, in order to make the evaluation of the model consequently [2].

We tested the performances of the above-mentioned data-balancing techniques on those three model architectures using the classification *Accuracy*, *Precision*, *Recall*, *F1-score* of the minority class and *Confusion Matrix*, which are calculated

considering the weighted average [1]. Hence, we got the Ensemble method's *Average Data Combine & Weighted Majority Voting* (using the *F1-score*) results on our test set [3]. When the dataset is highly unbalanced; the classification *Accuracy* or *Error-rate* are not the best parameters to decide the performance of that particular classifier. Therefore, we emphasized on the *F1-score* value of the minority class (Unhealthy) and the Confusion Matrix generated by each model. It is proved that our proposed methods have boosted each classifier in order to classify minor class instances more, rather than only focusing on the majority class (Healthy) images, which is more important with a view to classifying correctly these kinds of unbalanced cell images.

Keywords: Imbalanced Classification, Imbalance Ratio, Convolutional Neural Network (CNN), Sampling, Class-weight, Loss function and Ensemble.

References

[1] A. Ahammed, A. Hajdu, B. Harangi: *Hybrid AdaBoost and Naive Bayes Classifier for Supervised Learning* (2020), pp. 1–18, doi: https://ceur-ws.org/Vol-2874/.

[2] K. Balaji, K. Lavanya: *Chapter 5 - Medical Image Analysis With Deep Neural Networks*, in: Deep Learning and Parallel Computing Environment for Bioengineering Systems, ed. by A. K. Sangaiah, Academic Press, 2019, pp. 75–97, isbn: 978-0-12-816718-2, doi: https://doi.org/10.1016/B978-0-12-816718-2.00012-9.

[3] D. R. Patil, J. B. Patil: *Malicious URLs Detection Using Decision Tree Classifiers and Majority Voting Technique*, Cybernetics and Information Technologies 18.1 (2018), pp. 11–29, doi: doi:10.2478/cait-2018-0002, url: https://doi.org/10.2478/cait-2018-0002.

[4] F. Rayhan, S. Ahmed, A. Mahbub, R. Jani, S. Shatabda, D. M. Farid: *CUSBoost: Cluster-Based Under-Sampling with Boosting for Imbalanced Classification*, in: 2017 2nd International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS), 2017, pp. 1–5, doi: https://doi.org/10.1109/CSITSS.2017.8447534.