

# Improvement Performance Deep Learning-based Multi-class ECG Classification Model with Limited Medical Dataset

Sanghoon Choi<sup>1</sup>, Segyeong Joo<sup>1</sup>

<sup>1</sup>Department of Biomedical engineering, Asan Medical Institute of Convergence Science and Technology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea.

## ABSTRACT

**Background and Objective:** Real-world medical data, such as electrocardiogram (ECG), often exhibits imbalanced data distributions, which presents a considerable difficulty for classification tasks. To address this issue, numerous AI-based augmentation methods have been proposed to balance the classes. However, the use of augmented data remains controversial in the clinical field, as it is not real and may introduce biases into the training data learned by the generative model. In this study, we proposed a method for achieving optimal results without relying on augmentation techniques.

**Methods:** We conducted three experiments to verify the effects of different approaches for addressing imbalanced data in classifying a six-class, 12-lead ECG dataset collected from the Seoul Asan Medical Center (IRB 2021-1259). The experiments included using focal loss as the loss function (Experiment A-1), using class weight (Experiment A-2), balancing all classes to match the smallest class (Experiment B), and configuring subclasses (Experiment C). We employed the ResNet deep learning model for multi-class classification.

**Results:** Experiment C had an equal number of data but performed worse with a score of 0.86 compared to the Imbalance state. Meanwhile, Experiment A used either focal loss or class weight and achieved high scores of 0.95 and 0.93 respectively. Focal loss had the best performance among the two methods.

**Conclusion:** We developed a method to improve the performance of an ECG classifier with limited data. Results show that properly weighting the loss function, specifically using focal loss, in a deep learning model is more effective than altering the amount of data to solve the issue of imbalanced data. Future studies should focus on developing an optimal classifier for limited medical environments. The highest F1 scores achieved were 0.95 for Inception net with focal loss, and 0.86 in the limited data environment with the same ratio.

Table 1. Amount of dataset in three Experiments

Class	Name of class	Imbalance dataset	Balance dataset	Subclass dataset
1	1st degree AV block or 2nd degree AV block (Mobits type 1)	279	279	885
2	PSVT	307	279	
3	High degree or complete AV block	299	279	1451
4	Irregular narrow QRS tachyarrhythmia	1451	279	
5	Sinus tachyarrhythmia	1132	279	1132
6	VPCs	2382	279	2382

Table 2. Results for Micro-Average score in each Experiment

Experiment	Precision	Recall	F1-score	Accuracy	
Imbalance	0.9	0.86	0.88	0.9	
Focal loss	<b>0.95</b>	<b>0.95</b>	<b>0.95</b>	<b>0.95</b>	
Class weight	0.93	0.91	0.93	0.93	
Balance	0.86	0.86	0.86	0.86	
Subclass	1st	0.96	0.95	0.96	0.96
	2nd	0.84	0.83	0.85	0.84

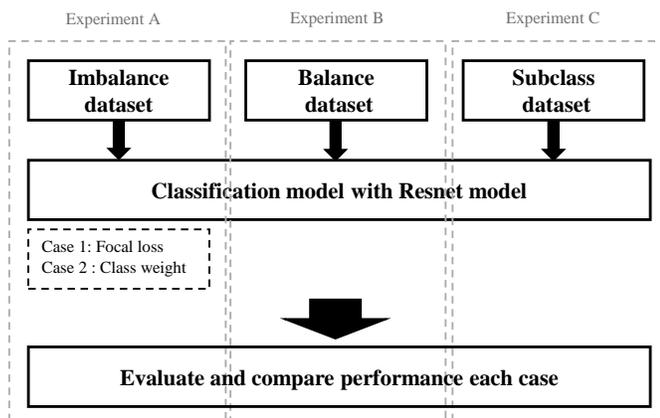


Figure 1. overview of proposed method.