The Impact of Data Augmentation on the Accuracy of Convolutional Neural Network Models for Skin Cancer Classification Using the Skin Cancer MNIST HAM10000 Dataset

Nadai Pramesti
Science Data Study Program, Faculty of Science and Technology
University of Technology Yogyakarta
Jl. Ringroad Utara Jombor Sleman Yogyakarta
E-mail: email.mahasiswa@gmail.com

ABSTRACT

Skin cancer is a disease that can be identified through dermatoscopic images. This study aims to evaluate the impact of data augmentation on the accuracy of a Convolutional Neural Network (CNN) model in classifying seven types of skin cancer using the HAM10000 dataset. The ResNet50 architecture was employed as the primary model through a transfer learning approach. Three experimental scenarios were conducted: a baseline model without augmentation, Type 1 augmentation (rotation, zoom, and translation), and Type 2 (mild augmentation). To address class imbalance, resampling was performed to ensure each class contained 500 samples. The baseline model achieved an accuracy of 75.48%, though it demonstrated poor performance on minority classes. Type 1 augmentation yielded 71.31% accuracy with improved performance uniformity across classes. Type 2 augmentation resulted in the highest accuracy of 77.37%, along with stable performance across all categories. These findings suggest that data augmentation, particularly mild augmentation techniques, can enhance the generalization capability of CNNs and contribute to more balanced skin cancer classification outcomes.

Keywords: Classification, Skin Cancer, CNN, Data Augmentation, ResNet50, HAM10000.