ABSTRACT

COMPARISON OF DATA BALANCING TECHNIQUES BETWEEN AUGMENTATION, CLASS WEIGHT, AND A COMBINATION OF AUGMENTATION AND CLASS WEIGHT IN MULTI-CLASS SKIN LESION CLASSIFICATION BASED ON DERMOSCOPIC IMAGES USING SWIN TRANSFORMER

By

Fatur Rozak

Skin lesion classification based on dermoscopic images is an important approach to support the early detection of skin cancer. However, imbalanced data often becomes a major challenge in multi-class classification, as it can degrade model performance, particularly in recognizing minority classes. This study aims to analyze the effect of data balancing techniques on the performance of a Transformer-based skin lesion classification model using the HAM10000 dataset, which consists of seven skin lesion classes. The model employed is the Swin Transformer with four training scenarios: (1) without any balancing technique, (2) using data augmentation, (3) using class weights, and (4) a combination of data augmentation and class weights.The experimental results show that the model using the combined balancing technique of data augmentation and class weights achieves the best and most balanced performance across evaluation metrics, with an accuracy of 88.77%, precision of 83.28%, sensitivity of 83.42%, specificity of 97.07%, F1-score of 83.10%, and an AUC value of 99.04%. These results indicate that the combination of data variation and penalty through weighting enables the Swin Transformer model to learn the data effectively while minimizing errors compared to training without balancing, with data augmentation only, or with class weights only. Furthermore, this study demonstrates that the choice of data balancing techniques significantly affects model performance, especially for multi-class imbalanced datasets such as HAM10000.

Keywords: Skin Lesion, Classification, Augmentation, Class Weight, Comparison, Balancing Data.

ABSTRAK

PERBANDINGAN *BALANCING DATA* ANTARA AUGMENTASI, *CLASS WEIGHT* DAN KOMBINASI AUGMENTASI + *CLASS WEIGHT* DALAM KLASIFIKASI MULTI-KELAS LESI KULIT BERBASIS CITRA DERMATOSKOPIK MENGGUNAKAN *SWIN TRANSFORMER*

Oleh

Fatur Rozak

Klasifikasi lesi kulit berbasis citra dermatoskopi merupakan pendekatan penting dalam mendukung deteksi dini kanker kulit. Namun, *imbalanced data* sering menjadi tantangan utama dalam klasifikasi multi-kelas karena dapat menurunkan performa model, khususnya dalam mengenali kelas minoritas. Penelitian ini bertujuan untuk menganalisis pengaruh teknik *balancing data* terhadap performa model klasifikasi lesi kulit berbasis *Transformer* menggunakan dataset HAM10000 yang terdiri dari tujuh kelas lesi kulit. Model yang digunakan adalah *Swin Transformer* dengan empat skenario pelatihan, yaitu: (1) tanpa *teknik balancing*, (2) menggunakan augmentasi data, (3) menggunakan *class weight*, dan (4) kombinasi augmentasi dan *class weight*. Hasil eksperimen menunjukkan bahwa model dengan teknik *balancing* gabungan antara augmentasi + *class weight* memberikan performa terbaik dan seimbang untuk metrik evaluasi nya dengan *accuracy* mencapai 88.77%, *precision* 83,28%, *sensitivity* 83,42%, *specifisity* 97,07%, *f1-score* 83,10% dan nilai AUC sebesar 99,04%. Hasil tersebut menunjukkan kombinasi variasi dan pinalti oleh bobot menjadikan model *Swin Transformer* mampu mempelajari data dengan baik dengan meminimalisir kesalahan dibanding tanpa *balancing*, dengan augmentasi saja dan dengan *class weight* saja. Hasil ini juga membuktikan bahwa pemilihan teknik *balancing* dapat mempengaruhi performa model terutama untuk data *imbalance* multi-kelas seperti data HAM10000.

Kata-kata kunci: Lesi Kulit, Klasifikasi, Augmentasi, *Class Weight*, Perbandingan, *Balancing Data*.