

ABSTRAK

Klasifikasi multikelas lesi kulit berbasis citra dermoskopi masih menghadapi tantangan *class imbalance* karena dapat membuat model lebih dominan mengenali kelas mayoritas dan kurang optimal dalam mengenali kelas minoritas. *Class imbalance* merupakan kondisi ketika jumlah data pada setiap kelas tidak terdistribusi secara seimbang. Penelitian ini bertujuan untuk menganalisis pengaruh distribusi data dan fungsi *loss* dalam penanganan *class imbalance* pada klasifikasi multikelas lesi kulit menggunakan EfficientNetB0. Dataset yang digunakan pada penelitian ini adalah HAM10000 yang diperoleh dari Kaggle. HAM10000 terdiri atas 10.015 citra dermoskopi dari tujuh kelas diagnosis lesi kulit. Eksperimen dirancang dalam 16 skenario eksperimen yang diperoleh dari kombinasi empat skenario distribusi data, yaitu *original*, *oversampling*, *undersampling*, dan *over-undersampling*, serta empat fungsi *loss*, yaitu *categorical cross-entropy*, *weighted cross-entropy*, *focal loss*, dan *focal loss* dengan *class weight*. Evaluasi performa model dilakukan menggunakan *balanced accuracy* sebagai metrik utama, serta *G-mean*, *macro precision*, *macro-F1* dan *overall accuracy* sebagai metrik pendukung. Hasil penelitian menunjukkan bahwa perubahan distribusi data lebih berpengaruh terhadap performa model dibandingkan perubahan fungsi *loss*. Hal ini terlihat dari rentang rata-rata *balanced accuracy* antar skenario distribusi data sebesar 0,0590, lebih besar dibandingkan rentang antar fungsi *loss* sebesar 0,0196. Skenario terbaik pada data validasi adalah *over-undersampling* dengan *focal loss*, yang memperoleh *balanced accuracy* sebesar 0,7126. Dua skenario dengan performa validasi terdekat adalah *undersampling* dengan *weighted cross-entropy* sebesar 0,7102 dan *over-undersampling* dengan *categorical cross-entropy* sebesar 0,7082. Pada evaluasi akhir menggunakan data uji, skenario *over-undersampling* dengan *focal loss* memperoleh *balanced accuracy* sebesar 0,7116, sehingga performanya relatif konsisten terhadap hasil validasi. Hasil ini menunjukkan bahwa distribusi data berperan lebih besar dalam penanganan *class imbalance*, sedangkan fungsi *loss* tetap memengaruhi performa ketika dikombinasikan dengan komposisi data latih yang sesuai.

Kata kunci : *class imbalance*, distribusi data, EfficientNetB0, fungsi *loss*, HAM10000, klasifikasi lesi kulit.