

ABSTRAK

Penyakit Alzheimer merupakan gangguan neurodegeneratif progresif yang diagnosis dan evaluasi dininya banyak memanfaatkan citra *Magnetic Resonance Imaging* (MRI). Seiring berkembangnya teknologi kecerdasan buatan, deep learning berbasis *Convolutional Neural Network* (CNN) telah banyak digunakan untuk membantu klasifikasi penyakit Alzheimer dari citra MRI dengan tingkat *accuracy* yang tinggi. Namun, model CNN sering dianggap sebagai *black box* karena kurang transparan dalam menjelaskan proses pengambilan keputusan secara klinis. Penelitian ini bertujuan membandingkan kinerja arsitektur EfficientNet-B0 dan ResNet50 dalam klasifikasi empat stadium penyakit Alzheimer serta mengevaluasi interpretabilitas model menggunakan metode *Gradient-weighted Class Activation Mapping* (Grad-CAM). Dataset yang digunakan terdiri atas 6.400 citra MRI yang dibagi ke dalam kelas *Non-Demented*, *Very Mild Demented*, *Mild Demented*, dan *Moderate Demented*. Pengujian dilakukan pada beberapa skenario, yaitu data original, augmentasi data, penerapan *class weight*, serta kombinasi keduanya antara augmentasi data dan penerapan *class weight*. Hasil penelitian menunjukkan bahwa EfficientNet-B0 memiliki stabilitas dan kemampuan generalisasi yang lebih baik dengan *validation accuracy* sebesar 96,5% dan *recall* 95% pada skenario augmentasi. Sementara itu, ResNet50 mencapai *validation accuracy* tertinggi sebesar 97% pada data original, namun cenderung mengalami *overfitting* dan fenomena *shortcut learning*. Visualisasi Grad-CAM menunjukkan bahwa EfficientNet-B0 lebih konsisten memfokuskan perhatian pada area anatomi otak yang relevan dibandingkan ResNet50. Berdasarkan hasil tersebut, EfficientNet-B0 dinilai lebih unggul dalam menjaga keseimbangan antara performa klasifikasi, kemampuan generalisasi, dan interpretabilitas klinis.

Kata Kunci: Alzheimer, CNN, EfficientNet-B0, ResNet50, Grad-CAM