

INTISARI

EVALUASI MODEL BERBASIS CONCATENATED FITUR UNTUK DIAGNOSIS MULTIKELAS PENYAKIT PARU-PARU PADA DATASET YANG TIDAK SEIMBANG

Oleh

Wahyu Ajitomo

23/529905/PPA/06703

Penyakit paru-paru seperti pneumonia, tuberkulosis, dan COVID-19 merupakan tantangan kesehatan global, terutama dalam klasifikasi citra X-ray yang sering menghadapi distribusi kelas tidak seimbang. Ketidakseimbangan ini menyebabkan model cenderung bias terhadap kelas mayoritas sehingga sensitivitas kelas minoritas menurun. Oleh karena itu, diperlukan pendekatan yang tidak hanya fokus pada akurasi, tetapi juga menekan false negative agar hasil lebih relevan secara klinis. Penelitian ini mengusulkan model gabungan berbasis concatenated CNN dengan Focal Loss Multiclass ($\gamma = 2$, α berbasis distribusi kelas) untuk meningkatkan sensitivitas pada dataset X-ray multikelas berjumlah 7.135 citra (normal, pneumonia, tuberkulosis, COVID-19). Kombinasi arsitektur seperti DenseNet121 + VGG16 dan VGG16 + ResNet50 dirancang agar fitur global dan lokal saling melengkapi, sehingga representasi lebih komprehensif. Hasil eksperimen menunjukkan bahwa model gabungan mencapai akurasi 87% dan sensitivitas 87% dengan AUC-ROC 0,97, lebih baik dibanding model tunggal (78–84% akurasi, 78–84% sensitivitas, AUC 0,94–0,96). Visualisasi Grad-CAM memperlihatkan aktivasi lebih konsisten pada area anatomi paru yang relevan. Walaupun waktu komputasi meningkat, manfaat berupa sensitivitas lebih tinggi dan diskriminasi probabilistik yang lebih baik menjadikan pendekatan ini potensial sebagai sistem pendukung keputusan klinis berbasis citra medis.

Kata Kunci: Dataset tidak seimbang, Klasifikasi Citra X-ray dada, Diagnosis penyakit paru-paru, Gabungan CNN, Focal Loss multikelas, Interpretabilitas Grad-CAM

ABSTRACT

EVALUATION OF A FEATURE-CONCATENATED MODEL FOR MULTICLASS DIAGNOSIS OF PULMONARY DISEASES ON AN IMBALANCED DATASET

By

Wahyu Ajitomo

23/529905/PPA/06703

Lung diseases such as pneumonia, tuberculosis, and COVID-19 remain a global health challenge, particularly in chest X-ray classification tasks that often face class imbalance. This imbalance causes models to be biased toward majority classes, thereby reducing sensitivity for minority classes. Therefore, an approach is needed that not only focuses on accuracy but also reduces false negatives to produce results that are more clinically relevant. This study proposes a concatenated CNN model combined with Focal Loss Multiclass ($\gamma = 2$, α based on class distribution) to enhance sensitivity on a multiclass chest X-ray dataset of 7,135 images (normal, pneumonia, tuberculosis, COVID-19). Architectural combinations such as DenseNet121 + VGG16 and VGG16 + ResNet50 were designed so that global and local features complement each other, resulting in more comprehensive representations. Experimental results show that the combined models achieved 87% accuracy and 87% sensitivity with an AUC-ROC of 0.97, outperforming single models (78–84% accuracy, 78–84% sensitivity, AUC 0.94–0.96). Grad-CAM visualization demonstrated more consistent activation on relevant lung anatomical regions. Although computational time increased, the benefits of higher sensitivity and improved probabilistic discrimination make this approach a promising candidate for clinical decision-support systems based on medical imaging.

Keywords: *Imbalanced dataset, Chest X-ray classification, Lung disease diagnosis, CNN concatenation, Focal loss multiclass, Grad-CAM interpretability*