



INTISARI

Balanced-MixUp on Manifold for Highly Imbalanced Breast Cancer Classification

Oleh

Muhammad Fadlurrohman

22/501256/PPA/06383

Klasifikasi kanker payudara menghadapi tantangan besar akibat ketidakseimbangan yang tinggi pada dataset mamogram, di mana jumlah kasus jinak jauh lebih banyak dibandingkan kasus ganas. Ketimpangan ini dapat menyebabkan model *deep learning* menjadi bias terhadap kelas mayoritas, sehingga mengurangi efektivitasnya dalam mendeteksi kasus kanker. Untuk mengatasi masalah ini, kami mengusulkan sebuah kerangka augmentasi *multi-level* yang baru bernama *Balanced-MixUp on Manifold*, yang menggabungkan *MixUp* berbasis *class* dan berbasis *instance* pada level input dengan interpolasi di ruang laten melalui *Manifold Mixup*. *Balanced-MixUp* menghasilkan sampel sintetik dengan pengaturan nilai *hyperparameter* α menggunakan distribusi *Beta* ($\alpha, 1$) untuk meningkatkan representasi kelas minoritas, sementara *Manifold Mixup* melakukan interpolasi fitur pada lapisan tersembunyi untuk menghasilkan batas keputusan yang lebih halus. Metode ini dievaluasi pada dataset mamogram EMBED C-view menggunakan empat arsitektur *CNN*: *MobileNet-V2*, *VGG-19*, *DenseNet-121*, dan *ResNeXt-50*. Validasi silang stratifikasi 5-fold digunakan, dengan metrik evaluasi meliputi *Matthews Correlation Coefficient (MCC)*, *Balanced Accuracy (B-ACC)*, dan *Macro-F1*. Hasil menunjukkan bahwa metode yang kami usulkan secara signifikan meningkatkan performa klasifikasi, khususnya pada kasus kanker yang kurang terwakili.

Kata kunci: Klasifikasi Kanker Payudara, Ketidakseimbangan Tinggi, Mamogram C-view, *Balanced-MixUp*, *Manifold Mixup*



ABSTRACT

Balanced-MixUp on Manifold for Highly Imbalanced Breast Cancer Classification

By

Muhammad Fadlurrohman

22/501256/PPA/06383

Breast cancer classification faces significant challenges due to the highly imbalanced nature of mammography datasets, where benign cases vastly outnumber malignant cases. This imbalance can lead deep learning models to become biased toward the majority class, limiting their effectiveness in detecting cancerous cases. To address this, we propose a novel multi-level augmentation framework called Balanced-MixUp on Manifold, which combines class-based and instance-based MixUp at the input level with latent space interpolation via Manifold Mixup. Balanced-MixUp generates synthetic samples with controlled α -hyperparameter using a Beta ($\alpha, 1$) distribution to improve minority class representation, while Manifold Mixup interpolates features within hidden layers to encourage smoother decision boundaries. We evaluate our method on the EMBED C-view mammogram dataset using four architectures CNN: MobileNet-V2, VGG-19, DenseNet-121, and ResNeXt-50. Stratified 5-fold cross-validation is performed, and evaluation metrics include Matthews Correlation Coefficient (MCC), Balanced Accuracy (B-ACC), and Macro-F1. Results demonstrate that our propose method significantly improves classification performance on underrepresented cancer cases.

Keywords: Breast Cancer Classification, Highly Imbalanced, C-view Mammogram, Balanced-MixUp, Manifold Mixup