

ABSTRAK

Kanker payudara merupakan salah satu penyebab utama kematian pada perempuan di dunia dengan 2,3 juta kasus dan 685.000 kematian pada tahun 2020. Algoritma Convolutional Neural Network (CNN) telah dikembangkan untuk klasifikasi citra medis. Namun, metode tersebut mengalami *overfitting* karena CNN memerlukan dataset yang besar untuk mendapatkan hasil akurasi yang maksimal. Sementara itu, data citra medis tidak dapat memenuhi kebutuhan dataset besar yang diperlukan oleh CNN. Akibatnya, terjadi *overfitting* pada jaringan dan hasil akurasi yang kurang maksimal. Metode konvensional seperti UNet memiliki keterbatasan dalam generalisasi data, sehingga dibutuhkan ekstrasi fitur yang lebih baik untuk memperbaiki kinerja klasifikasi. Oleh karena itu, ResNet ditambahkan ke dalam UNet untuk memperkaya proses ekstrasi fitur dengan memanfaatkan koneksi residual.

Penelitian ini mengimplementasikan metode ResUNet untuk mengklasifikasi kanker payudara pada citra mammogram. Kedua arsitektur ini memungkinkan data dilatih secara efektif meskipun dengan lapisan yang besar. Metode *waterfall* dengan tahapan studi literatur, pengumpulan data, *pre-processing*, permodelan ResUNet, serta evaluasi model digunakan pada penelitian ini. Data yang digunakan pada penelitian ini adalah “INbreast”. Selain itu, penelitian ini juga menambahkan data *irrelevant image* yang diambil dari Unsplash Random Images untuk memperkaya variasi dataset. Dataset kemudian diseimbangkan, sehingga menghasilkan total citra sebanyak 3.780 untuk *training*, 1.890 untuk *test*, dan 1.890 untuk validasi.

Hasil penelitian menunjukkan bahwa model ResUNet lebih unggul dibandingkan dengan model konvensional CNN, UNet, dan ResNet. Model ResUNet menghasilkan akurasi pelatihan 98% dan akurasi pengujian 96%. Hal ini menunjukkan kemampuan model ResUNet dalam mengatasi *overfitting* dan mampu menggeneralisasi lebih baik pada data baru. Di sisi lain, model CNN, UNet, dan ResNet mengalami keterbatasan pada fase pengujian dengan hasil akurasi CNN 87%, UNet 87%, dan ResNet 76%. Evaluasi model ResUNet dengan rata-rata nilai *precision*, *recall*, dan *f1-score* 0.93 membuktikan bahwa kombinasi UNet dan ResNet secara efektif mampu mengatasi *overfitting*, meningkatkan generalisasi, dan menghasilkan klasifikasi yang lebih akurat pada citra mammogram.

Kata kunci: kanker payudara, ResUNet, klasifikasi, *overfitting*, generalisasi

ABSTRACT

Breast cancer is one of the leading causes of mortality among women worldwide, with 2.3 million cases and 685,000 deaths reported in 2020. Convolutional Neural Network (CNN) algorithms have been developed for medical image classification. However, these methods often face overfitting issues due to the high dataset volume required by CNN to achieve optimal accuracy. Unfortunately, medical imaging datasets typically do not meet these requirements, resulting in overfitting and suboptimal classification performance. Conventional methods, such as U-Net, exhibit limitations in data generalization, necessitating improved feature extraction methods to enhance classification performance. To address this limitation, ResNet is integrated into U-Net to enrich the feature extraction process by utilizing residual connections.

This study implements the ResUNet method for breast cancer classification in mammogram images. The integration of these two architectures enables effective training, even with deep network layers. A waterfall methodology is employed, encompassing literature review, data collection, preprocessing, ResUNet modeling, and model evaluation. The dataset used in this study is "INbreast." Additionally, irrelevant images obtained from Unsplash Random Images are included to enhance dataset variability. The dataset is then balanced, resulting in a total of 3,780 images for training, 1,890 for testing, and 1,890 for validation.

The results indicate that the ResUNet model outperforms conventional models, including CNN, U-Net, and ResNet. The ResUNet model achieves a training accuracy of 98% and a testing accuracy of 96%. These findings demonstrate the model's ability to mitigate overfitting and generalize effectively to new data. In contrast, the CNN, U-Net, and ResNet models exhibit limitations during the testing phase, with accuracies of 87%, 87%, and 76%, respectively. The evaluation of the ResUNet model, with average precision, recall, and F1-score values of 0.93, confirms that the combination of U-Net and ResNet effectively addresses overfitting, enhances generalization, and produces more accurate classifications for mammogram images.

Keywords: breast cancer, ResUNet, classification, overfitting, generalization