

RINGKASAN

Telah dilakukan penelitian Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma *Convolutional Neural Network (CNN)* dan *Random Forest* pada Citra Mammogram berbasis Python. Dalam upaya klasifikasi penyakit kanker payudara, analisis dilakukan berdasarkan citra mammogram. Algoritma *Convolutional Neural Network (CNN)* dan *Random Forest* dipilih menjadi metode utama dalam analisis ini melalui penerapan dengan Python. Penelitian ini bertujuan untuk menerapkan dan membandingkan algoritma *Convolutional Neural Network (CNN)* dan *Random Forest* dalam hal *accuracy*, *precision*, *recall*, dan *F1-score* dalam klasifikasi penyakit kanker payudara berdasarkan dataset citra mammogram. Salah satu langkah awal yang dilakukan dalam pengumpulan data dengan menggunakan platform terbuka yaitu Kaggle, yang mencakup 2 kategori dalam citra mammogram payudara yaitu tumor jinak (*benign*) dan tumor ganas (*malignant*). Data yang diperoleh kemudian di *preprocessing* dengan cara pembersihan data untuk menghilangkan noise atau gangguan yang mungkin ada pada citra mammogram, dilanjutkan dengan normalisasi intensitas piksel. Ekstraksi fitur juga dilakukan untuk membedakan ciri khas setiap jenis kanker payudara. Seluruh dataset kemudian diubah ke format yang sesuai dan dikonversi menjadi format warna yang diperlukan untuk memastikan fokus pada fitur dalam citra. Augmentasi data untuk menyeimbangkan data yang tidak seimbang, dengan menggunakan pemrograman Python, khususnya melalui syntax yang sesuai, proses visualisasi menjadi lebih efisien dan efektif. Interpretasi yang mendalam dan jelas mengenai perbedaan visual antara setiap jenis kanker, memudahkan dalam analisis dan penentuan diagnosis. Model CNN dan *Random Forest* dilatih dan dievaluasi berdasarkan matriks yang digunakan. Proses pembagian dataset menjadi data latih dan data uji dilakukan dengan parameter *random_state=42* untuk memastikan model memiliki kemampuan yang baik. Hasil evaluasi menunjukkan akurasi yang tinggi. Model CNN mencapai *accuracy* sebesar 97%, *precision* 90%, *recall* 94%, dan *F1-score* 92%. Sementara itu, model *Random Forest* mencatat hasil dengan *accuracy* 95%, *precision* 86%, *recall* 91%, dan *F1-score* 88%. Model CNN yang dikembangkan memiliki performa yang lebih baik dalam mengklasifikasikan jenis-jenis kanker payudara dengan *accuracy* yang tinggi. Penelitian ini menawarkan solusi diagnostik yang potensial untuk mendukung penanganan penyakit kanker payudara.

SUMMARY

A research has been conducted on breast cancer classification using Convolutional Neural Network (CNN) and Random Forest algorithms applied to mammogram images with Python implementation. In an effort to classify breast cancer, analysis was performed based on mammogram images. Convolutional Neural Network (CNN) and Random Forest algorithms were selected as the primary methods in this analysis through implementation with Python. This research aimed to apply and compare CNN and Random Forest algorithms in terms of accuracy, precision, recall, and F1-score in breast cancer classification based on mammogram image datasets. One of the initial steps involved data collection using the open platform Kaggle, which included two categories of breast mammogram images: benign tumors and malignant tumors. The acquired data underwent preprocessing through data cleaning to eliminate potential noise or disturbances in mammogram images, followed by pixel intensity normalization. Feature extraction was also performed to differentiate the distinctive characteristics of each type of breast cancer. The entire dataset was then converted to an appropriate format and transformed into the required color format to ensure focus on the features within the images. Data augmentation was implemented to balance uneven data distributions. Using Python programming, particularly through appropriate syntax, the visualization process became more efficient and effective. In-depth and clear interpretation of visual differences between each cancer type facilitated analysis and diagnosis determination. CNN and Random Forest models were trained and evaluated based on established metrics. The dataset was divided into training and testing data using a `random_state` parameter of 42 to ensure the models possessed good capabilities. Evaluation results demonstrated high accuracy. The CNN model achieved an accuracy of 97%, precision of 90%, recall of 94%, and F1-score of 92%. Meanwhile, the Random Forest model recorded results with an accuracy of 95%, precision of 86%, recall of 91%, and F1-score of 88%. The developed CNN model exhibited better performance in classifying breast cancer types with high accuracy. This research offers a potential diagnostic solution to support breast cancer treatment.