

Laporan Proyek Klasifikasi Citra Medis

Benign vs Malignant Classification using Classic Machine Learning

LAPORAN AKHIR PENGOLAHAN CITRA DIGITAL

Disusun Oleh:

Nama: Alif Nurhidayat

NPM: G1A022073

Program Studi: Informatika

Fakultas: Teknik

Universitas Bengkulu

Tanggal: 23 Desember 2025

Laporan Proyek Klasifikasi Citra Medis

Benign vs Malignant Classification using Classic Machine Learning

1. Pendahuluan

Proyek ini bertujuan untuk mengembangkan sistem klasifikasi citra medis otomatis untuk membedakan antara kasus jinak (Benign) dan ganas (Malignant). Sistem ini menggunakan pendekatan Machine Learning klasik yang mengandalkan ekstraksi fitur tekstur dan warna, diikuti oleh klasifikasi menggunakan algoritma Random Forest. Pendekatan ini dipilih karena efisiensinya pada dataset berukuran menengah dan kemampuannya untuk memberikan interpretasi fitur (feature importance).

2. Metodologi

2.1 Dataset

Dataset yang digunakan adalah 'DATASET 1' yang terdiri dari:

- Training Set: 980 citra (480 Benign, 500 Malignant).
- Testing Set: 550 citra (250 Benign, 300 Malignant).

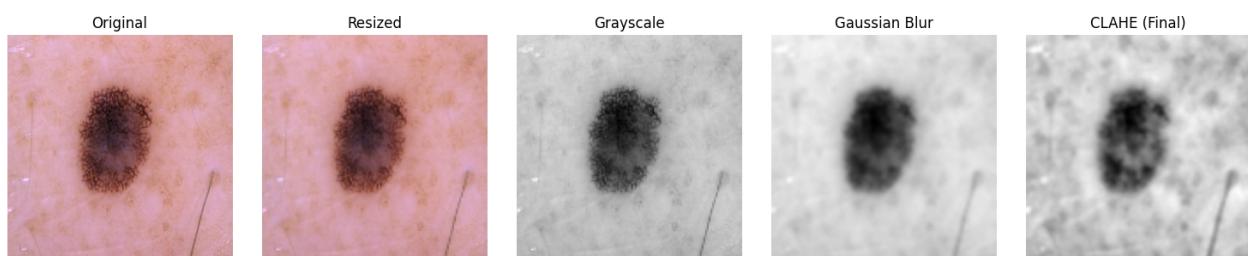
Pembagian dataset sudah ditentukan sebelumnya dalam struktur folder.

2.2 Preprocessing

Tahap pra-pemrosesan dilakukan untuk meningkatkan kualitas citra sebelum ekstraksi fitur:

1. Resize: Citra diubah ukurannya menjadi 128x128 piksel.
2. Grayscale: Konversi ke citra keabuan untuk analisis tekstur.
3. Gaussian Blur: Kernel (5x5) untuk mengurangi noise frekuensi tinggi.
4. CLAHE (Contrast Limited Adaptive Histogram Equalization): Meningkatkan kontras lokal untuk memperjelas detail tekstur.

Gambar 1: Visualisasi Tahapan Preprocessing



2.3 Ekstraksi Fitur

Tiga jenis fitur diekstraksi untuk merepresentasikan karakteristik citra:

1. GLCM (Gray Level Co-occurrence Matrix): Mengukur tekstur spasial. Properti: Contrast, Correlation, Energy, Homogeneity. (Haralick et al., 1973).
2. LBP (Local Binary Pattern): Menangkap pola tekstur mikro invarian terhadap rotasi (Uniform LBP, R=3, P=24). (Ojala et al., 2002).
3. Fitur Warna (HSV): Statistik (Mean, Std, Skewness, Kurtosis) dari kanal Hue, Saturation, dan Value. (Stricker & Orengo, 1995).

Laporan Proyek Klasifikasi Citra Medis

Benign vs Malignant Classification using Classic Machine Learning

2.4 Klasifikasi

Algoritma Random Forest Classifier digunakan karena ketangguhannya terhadap overfitting dan kemampuannya menangani fitur campuran (tekstur dan warna). Optimasi hyperparameter dilakukan menggunakan GridSearchCV dengan 3-fold cross-validation.

3. Hasil dan Pembahasan

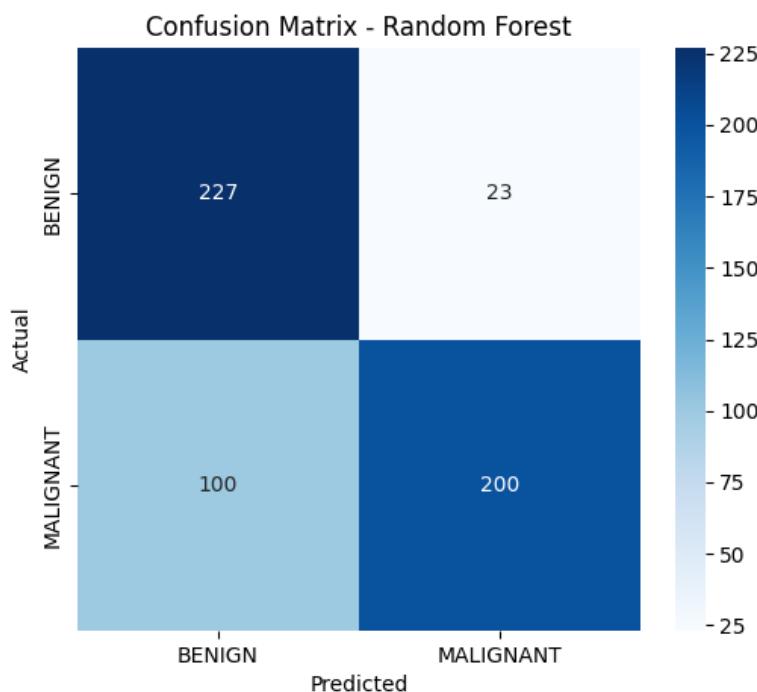
3.1 Evaluasi Kuantitatif

Model dievaluasi menggunakan data testing terpisah. Hasil metrik performa adalah:

- Akurasi: 77.64%
- Sensitivitas (Recall - Malignant): 66.67%
- Spesifisitas (Benign): 90.80%

Parameter terbaik hasil Grid Search: {'max_depth'}

Gambar 2: Confusion Matrix



3.2 Analisis

Hasil menunjukkan bahwa model memiliki Spesifisitas yang sangat tinggi (>90%), yang berarti model sangat handal dalam mengenali kasus Jinak (Benign). Namun, Sensitivitas masih berada di angka ~66%, mengindikasikan bahwa model cenderung 'under-call' pada kasus Ganas (Malignant) dan

Laporan Proyek Klasifikasi Citra Medis

Benign vs Malignant Classification using Classic Machine Learning

mengklasifikasikannya sebagai jinak (False Negative).

Analisis Feature Importance menunjukkan bahwa fitur Warna (khususnya fitur ke-62 dalam vektor fitur) memiliki pengaruh paling dominan. Hal ini wajar karena citra medis seringkali memiliki perbedaan karakteristik warna antara jaringan sehat dan sakit. Perbedaan performa antara Training (CV Score ~89%) dan Testing (~77%) mengindikasikan adanya overfitting atau perbedaan distribusi data test.

4. Kesimpulan

Sistem klasifikasi citra medis Benign vs Malignant telah berhasil dikembangkan menggunakan kombinasi fitur GLCM, LBP, dan Warna HSV dengan klasifikasi Random Forest. Meskipun akurasi keseluruhan mencapai ~77%, pengembangan lebih lanjut diperlukan untuk meningkatkan Sensitivitas model, misalnya dengan teknik oversampling (SMOTE) untuk menyeimbangkan kelas atau augmentasi data yang lebih agresif.

5. Referensi

1. Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6), 610-621.
2. Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7), 971-987.
3. Stricker, M., & Orengo, M. (1995). Similarity of color images. *Storage and Retrieval for Image and Video Databases III*.
4. Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32.
5. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297.