

Ekstraksi Fitur Citra Penyakit Kulit Untuk Klasifikasi Penyakit Kulit

Bagus Cipta Pratama, Meilany Dinda Talitha, Andrian Danar Perdana,
Fahmi Shampoerna, Ahmad Sudais
Kelompok 7, Mata Kuliah Pengenalan Pola
Departemen Ilmu Komputer dan Elektronika,
Universitas Gadjah Mada, Yogyakarta, Indonesia

I. PENDAHULUAN

Diagnosis penyakit kulit secara manual oleh dokter kulit sering memakan waktu dan sangat bergantung pada pengalaman klinis individu. Oleh karena itu, sistem otomatis berbasis citra digital dan ekstraksi fitur lanjutan dapat membantu mempercepat proses diagnosis dan meningkatkan konsistensi klasifikasi lesi kulit. Sistem ini mengandalkan pipeline pra-pemrosesan yang memastikan citra seragam, teraugmentasi, dan siap untuk tahap ekstraksi fitur dan pelatihan model klasifikasi. Model klasifikasi yang digunakan melibatkan algoritma pembelajaran mesin seperti Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), dan Decision Tree.

II. DATASET

A. Sumber Data

Dataset yang digunakan diunduh dari Kaggle: “Skin Disease Classification [Image Dataset]” oleh Riya Eliza Shaju. Dataset ini menyediakan ~900 citra klinis yang terbagi menjadi 9 kategori penyakit kulit: Actinic keratosis, Atopic dermatitis, Benign keratosis, Dermatofibroma, Melanocytic nevus, Melanoma, Squamous cell carcinoma, Tinea ringworm candidiasis, dan Vascular lesion.

B. Fitur dan Ukuran

Dataset ini terbagi menjadi:

- Training: 697 citra
- Validation: 181 citra

Citra berformat JPEG atau PNG dengan resolusi bervariasi dan diorganisir dalam folder per kelas untuk memudahkan pipeline preprocessing.

III. METODOLOGI

A. Pra-pemrosesan Data (Teori dan Langkah)

Sebelum dimasukkan ke dalam model klasifikasi, setiap citra melalui serangkaian langkah pra-pemrosesan yang mencakup:

a) *Resizing Citra*: Setiap citra diubah menjadi ukuran 512×512 piksel untuk memastikan dimensi seragam. Ini memastikan bahwa proses ekstraksi fitur berjalan konsisten. Persamaan untuk resize adalah:

$$\text{Resize}(I) = \text{Resize}(I, 512, 512)$$

di mana I adalah citra asli yang akan diubah ukurannya.

b) *Black-Hat Transformation (Penghilangan Rambut)*: Citra diubah menjadi skala abu-abu menggunakan fungsi `rgb2gray`, kemudian diterapkan transformasi Black-Hat untuk menghilangkan artefak seperti rambut yang dapat mengganggu pengolahan lesi kulit. Fungsi Black-Hat untuk penghilangan rambut didefinisikan sebagai:

$$B = I - \text{Closing}(I)$$

di mana I adalah citra input, dan B adalah citra setelah transformasi Black-Hat.

c) *Inpainting dan Gaussian Filtering*: Setelah penghilangan rambut, dilakukan inpainting untuk memperbaiki bagian yang terhapus dengan menggunakan algoritma biharmonik. Kemudian, gambar yang telah diperbaiki dihaluskan menggunakan filter Gaussian untuk mengurangi noise. Proses Gaussian filtering didefinisikan sebagai:

$$G_{\sigma}(I) = \int_{-\infty}^{\infty} I(x) \cdot \exp\left(-\frac{x^2}{2\sigma^2}\right) dx$$

di mana $G_{\sigma}(I)$ adalah citra yang difilter, $I(x)$ adalah citra input, dan σ adalah standar deviasi dari filter Gaussian.

d) *Augmentasi Data Training*: Untuk memperkaya variasi data tanpa menambah gambar baru, dilakukan augmentasi pada set training:

- Zoom Acak: Citra diperbesar atau diperkecil secara proporsional hingga $\pm 20\%$.
- Flip Horizontal dan Vertikal: Citra diputar secara horizontal dan vertikal.
- Rotasi Acak: Citra diputar dalam rentang $\pm 15^\circ$ untuk memberikan variasi orientasi citra.

B. Ekstraksi Fitur

Ekstraksi fitur dilakukan untuk mendeteksi ciri-ciri khas dari setiap citra, yang meliputi:

a) *Fitur Warna*: Fitur warna dihitung untuk setiap saluran warna (R, G, B) dengan menghitung:

$$\text{mean}(C) = \frac{1}{N} \sum_{i=1}^N C_i$$

$$\text{std}(C) = \sqrt{\frac{1}{N} \sum_{i=1}^N (C_i - \mu_C)^2}$$

$$\text{skew}(C) = \frac{1}{N} \sum_{i=1}^N \left(\frac{C_i - \mu_C}{\sigma_C} \right)^3$$

di mana C adalah komponen warna (R, G, atau B), C_i adalah nilai piksel pada citra, N adalah jumlah piksel, μ_C adalah rata-rata, dan σ_C adalah deviasi standar.

b) *Fitur Tekstur (GLCM)*: Menggunakan matriks ko-occurrence abu-abu (GLCM) untuk menghitung:

$$\text{Contrast} = \sum_{i,j} (i - j)^2 P(i, j)$$

$$\text{Correlation} = \frac{\sum_{i,j} (i - \mu)(j - \mu) P(i, j)}{\sigma_X \sigma_Y}$$

di mana $P(i, j)$ adalah probabilitas terjadinya pasangan nilai i dan j pada posisi tertentu, dan μ, σ_X, σ_Y adalah rata-rata dan deviasi standar dari distribusi GLCM.

c) *Local Binary Patterns (LBP)*: Fitur LBP dihitung sebagai berikut:

$$LBP(x, y) = \sum_{p=0}^{P-1} s(I_p - I_c) \cdot 2^p$$

di mana I_p adalah nilai piksel tetangga, I_c adalah nilai piksel pusat, dan $s(x)$ adalah fungsi ambang yang mengembalikan 1 jika $x \geq 0$ dan 0 jika $x < 0$.

C. Modeling dan Evaluasi

Model yang digunakan untuk klasifikasi penyakit kulit mencakup:

- **Logistic Regression**
- **Support Vector Machine (SVM)**
- **K-Nearest Neighbor (KNN)**
- **Decision Tree**

Metrik evaluasi yang digunakan termasuk akurasi, precision, recall, dan f1-score. GridSearchCV digunakan untuk optimasi hyperparameter pada setiap model.

a) *Proses Pelatihan*: Data yang telah dipra-pemrosesan dan diekstraksi fiturnya dibagi menjadi dua set, yaitu training dan validation. Model dilatih dengan menggunakan dataset training, dan hasil evaluasi dilakukan pada dataset validation.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

di mana TP adalah true positives, TN adalah true negatives, FP adalah false positives, dan FN adalah false negatives.

IV. HASIL DAN PEMBAHASAN

A. Perbandingan Akurasi Antara Model

Berikut adalah perbandingan akurasi keempat model yang diuji pada dataset validasi:

TABLE I
PERBANDINGAN AKURASI MODEL KLASIFIKASI

Model	Akurasi (%)
Logistic Regression	52.49
SVM	62.43
KNN	59.12
Decision Tree	57.46

Dari tabel di atas, dapat dilihat bahwa model SVM memiliki akurasi tertinggi (62.43%), diikuti oleh KNN (59.12%), Decision Tree (57.46%), dan Logistic Regression (52.49%). Namun, klasifikasi untuk Melanoma dan Squamous Cell Carcinoma tetap sulit.

B. confusion matrix

Untuk mengevaluasi performa model secara lebih mendalam, confusion matrix menunjukkan bahwa beberapa kelas memiliki tingkat prediksi yang lebih rendah, terutama pada kelas dengan jumlah sampel yang lebih sedikit, seperti Melanoma dan Squamous Cell Carcinoma. Hal ini menunjukkan tantangan dalam mengenali lesi kulit yang memiliki variasi visual yang lebih kompleks.

Confusion Matrix: SVM

True \ Predicted	Actinic keratosis	Atopic Dermatitis	Benign keratosis	Dermatofibroma	Melanocytic nevus	Melanoma	Squamous cell carcinoma	Tinea Ringworm Candidiasis	Vascular lesion
Actinic keratosis	11	0	0	1	0	0	8	0	0
Atopic Dermatitis	1	20	0	0	0	0	0	0	0
Benign keratosis	0	1	16	1	1	1	0	0	0
Dermatofibroma	5	0	0	8	1	1	4	0	1
Melanocytic nevus	0	0	0	0	16	3	1	0	0
Melanoma	0	4	1	0	5	6	4	0	0
Squamous cell carcinoma	7	0	0	1	2	1	9	0	0
Tinea Ringworm Candidiasis	1	4	2	0	0	0	1	12	0
Vascular lesion	1	1	0	2	0	1	0	0	15

Fig. 1. Confusion Matrix SVM

V. KESIMPULAN

SVM meraih akurasi tertinggi 62.43%. Melanoma paling sulit diklasifikasi, sedangkan Atopic Dermatitis paling mudah. Fitur ekstraksi menangkap pola penyakit kulit, namun akurasi terbatas pada 52-62% karena banyaknya klasifikasi. Potensi pengembangan lebih lanjut ada untuk meningkatkan performa, terutama pada Melanoma.