

Analisis Performa Random Forest dalam Deteksi Diabetes Menggunakan Fitur Manual Citra Retina dan Teknik SMOTE berbasis Website

Galang Adi Saputra¹, Nur Latifah Dwi Mutiara Sari²

^{1, 2}Program Studi Informatika, Universitas Persatuan Guru Republik Indonesia Semarang, Kota Semarang

*Email: ssaputra071@gmail.com¹

Abstract.

The importance of early diabetes detection drives the development of efficient screening technologies. This research implements and analyzes the performance of a Diabetic Retinopathy (DR) detection system using a classical machine learning approach. The system utilizes the Random Forest algorithm trained on a combination of manual features (color, shape, and texture) extracted from retinal fundus images. To address the common issue of imbalanced datasets in medical data, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to the training data. The evaluation results on the test set show excellent performance, with the model achieving an accuracy of 94.82% as well as balanced precision and recall scores of 0.95 for both "Normal" and "Diseased" classes. This finding proves that with appropriate feature engineering and proper data balancing, classical methods like Random Forest can serve as a highly effective and reliable tool for preliminary DR sc lebih akurat dan andal secara klinis.

Keywords: Performance Analysis, Diabetes Detection, Image Classification, Random Forest, Retinal Imagery, Feature Extraction, SMOTE.

Abstrak

Pentingnya deteksi dini diabetes mendorong pengembangan teknologi skrining yang efisien dan mudah diakses. Penelitian ini bertujuan untuk mengimplementasikan dan menganalisis performa sistem deteksi *Diabetic Retinopathy* (DR) berbasis web menggunakan pendekatan *machine learning* klasik. Sistem ini memanfaatkan algoritma Random Forest yang dilatih menggunakan kombinasi fitur manual (warna, bentuk, dan tekstur) dari citra fundus retina. Untuk mengatasi masalah dataset yang tidak seimbang, teknik *Synthetic Minority Over-sampling Technique* (SMOTE) diterapkan pada data latih. Hasil evaluasi pada data tes menunjukkan performa yang sangat baik, dengan model mampu mencapai akurasi sebesar 94.82% serta nilai presisi dan recall yang seimbang (0.95) untuk kelas "Normal" dan "Sakit". Temuan ini membuktikan bahwa dengan rekayasa fitur yang tepat dan penanganan data yang benar, metode klasik seperti Random Forest yang diimplementasikan dalam sistem berbasis web mampu menjadi alat skrining awal DR yang sangat efektif.

Kata Kunci: Analisis Performa, Deteksi Diabetes, Klasifikasi Gambar, Random Forest, Citra Retina, Ekstraksi Fitur, SMOTE, Sistem Berbasis Web.

1. Pendahuluan

Diabetes Mellitus merupakan salah satu tantangan kesehatan global yang paling signifikan di abad ke-21. Salah satu komplikasinya yang paling serius adalah *Diabetic Retinopathy* (DR), sebuah kondisi yang merusak pembuluh darah retina dan menjadi penyebab utama kebutaan pada populasi usia kerja [9]. Karena kerusakan ini seringkali tidak menunjukkan gejala pada tahap awal, deteksi dini menjadi faktor krusial. Pemeriksaan fundus retina menggunakan fotografi adalah metode skrining non-invasif yang efektif, namun analisis manual oleh ahli medis terbatas oleh waktu dan biaya, sehingga pengembangan sistem deteksi otomatis yang cepat dan mudah diakses menjadi sebuah kebutuhan mendesak.

Penelitian modern dalam deteksi DR didominasi oleh pendekatan *Deep Learning* (CNN) yang mampu mencapai akurasi tinggi [8]. Namun, pendekatan ini seringkali membutuhkan

sumber daya komputasi yang besar. Sebagai alternatif, pendekatan *Machine Learning* Klasik yang melibatkan ekstraksi fitur manual menawarkan model yang lebih ringan. Tantangan utama dalam pendekatan ini adalah masalah dataset yang tidak seimbang.

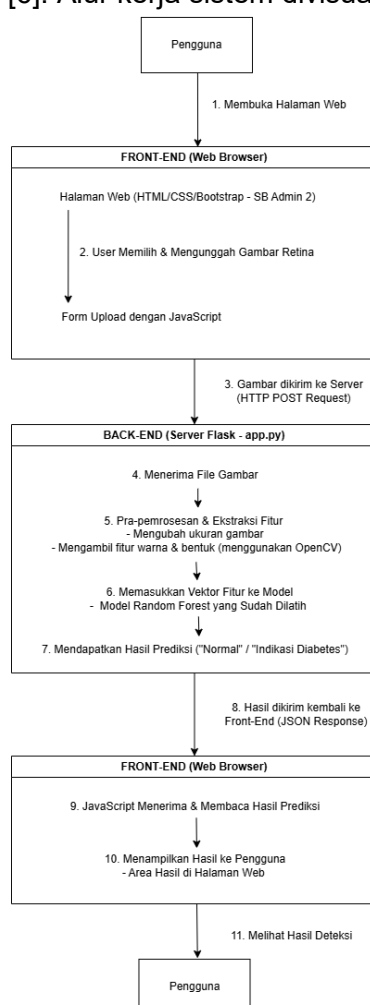
Berdasarkan hal tersebut, penelitian ini berfokus pada analisis performa algoritma Random Forest [3] dengan rekayasa fitur manual yang komprehensif. Penelitian ini secara spesifik mengukur efektivitas teknik penyeimbangan data SMOTE [2] untuk mengatasi masalah dataset tidak seimbang dan diimplementasikan dalam sebuah sistem berbasis web untuk kemudahan akses.

2. Metode

Metodologi penelitian dibagi menjadi beberapa tahapan utama, mulai dari perancangan arsitektur, pengumpulan data, ekstraksi fitur, hingga pelatihan dan evaluasi model.

2.1. Arsitektur Sistem

Sistem yang dibangun mengadopsi arsitektur client-server. Front-End dibangun menggunakan HTML, CSS, dan template SB Admin 2 [7], sementara Back-End menggunakan Python dengan framework Flask [6]. Alur kerja sistem divisualisasikan pada Gambar 1.



Gambar 1. Arsitektur Alur Kerja Sistem

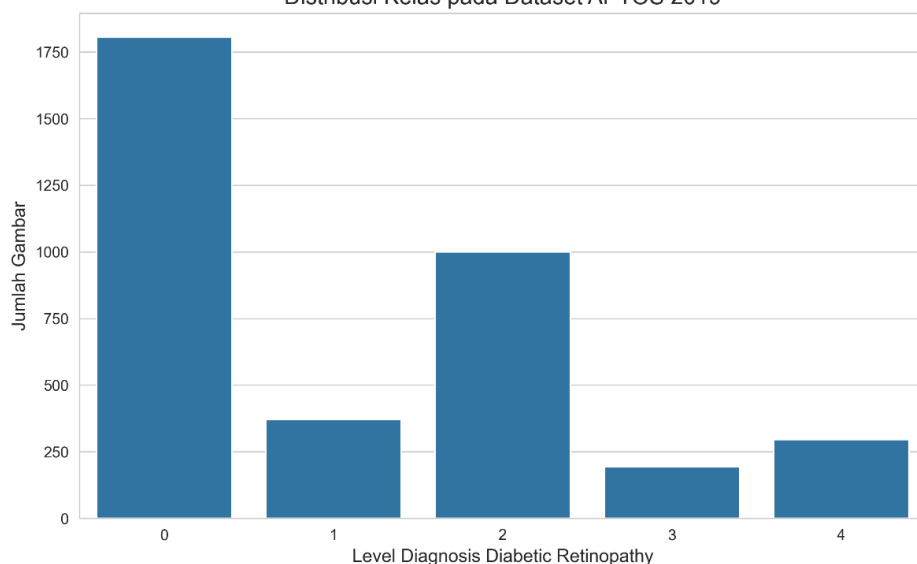
Diagram di atas menunjukkan alur data dari unggahan gambar oleh pengguna di Front-End hingga pemrosesan oleh model di Back-End dan penampilan hasil.

2.2. Dataset dan Pra-pemrosesan

Penelitian ini menggunakan dataset publik "APTOS 2019 Blindness Detection" dari Kaggle [1]. Distribusi kelas pada dataset ini sangat tidak seimbang, seperti yang ditunjukkan pada Gambar 2, yang menjadi justifikasi penggunaan teknik SMOTE. Untuk penelitian ini,

masalah disederhanakan menjadi klasifikasi biner ("Normal" vs. "Terdeteksi Indikasi Diabetes"). Setiap gambar diseragamkan ukurannya menjadi 512x512 piksel.

Distribusi Kelas pada Dataset APTOS 2019



Gambar 2. Grafik Distribusi Kelas.

Dataset Grafik menunjukkan ketidakseimbangan data yang signifikan, di mana kelas 0 (Normal) mendominasi dataset, menjadi justifikasi penggunaan teknik SMOTE.

2.3. Ekstraksi Fitur

Vektor fitur untuk setiap gambar dibangun dari tiga jenis fitur yang diekstrak menggunakan library OpenCV [4], seperti diilustrasikan pada Gambar 3:

1. **Fitur Tekstur (GLCM):** Fitur *contrast*, *homogeneity*, *energy*, dan *correlation* diekstrak dari gambar *grayscale* untuk mengukur tekstur retina.
2. **Fitur Bercak Merah (*Microaneurysms*):** Diekstrak dari kanal hijau gambar, diikuti dengan *adaptive thresholding* dan penyaringan kontur berdasarkan luas area.
3. **Fitur Bercak Kuning (*Exudates*):** Diekstrak dari ruang warna HSV dengan *thresholding* warna.



Gambar 3. Tahapan Proses Ekstraksi Fitur.

Dari kiri ke kanan: citra retina asli, mask hasil deteksi bercak merah (anomali pendarahan), dan mask hasil deteksi bercak kuning (exudates).

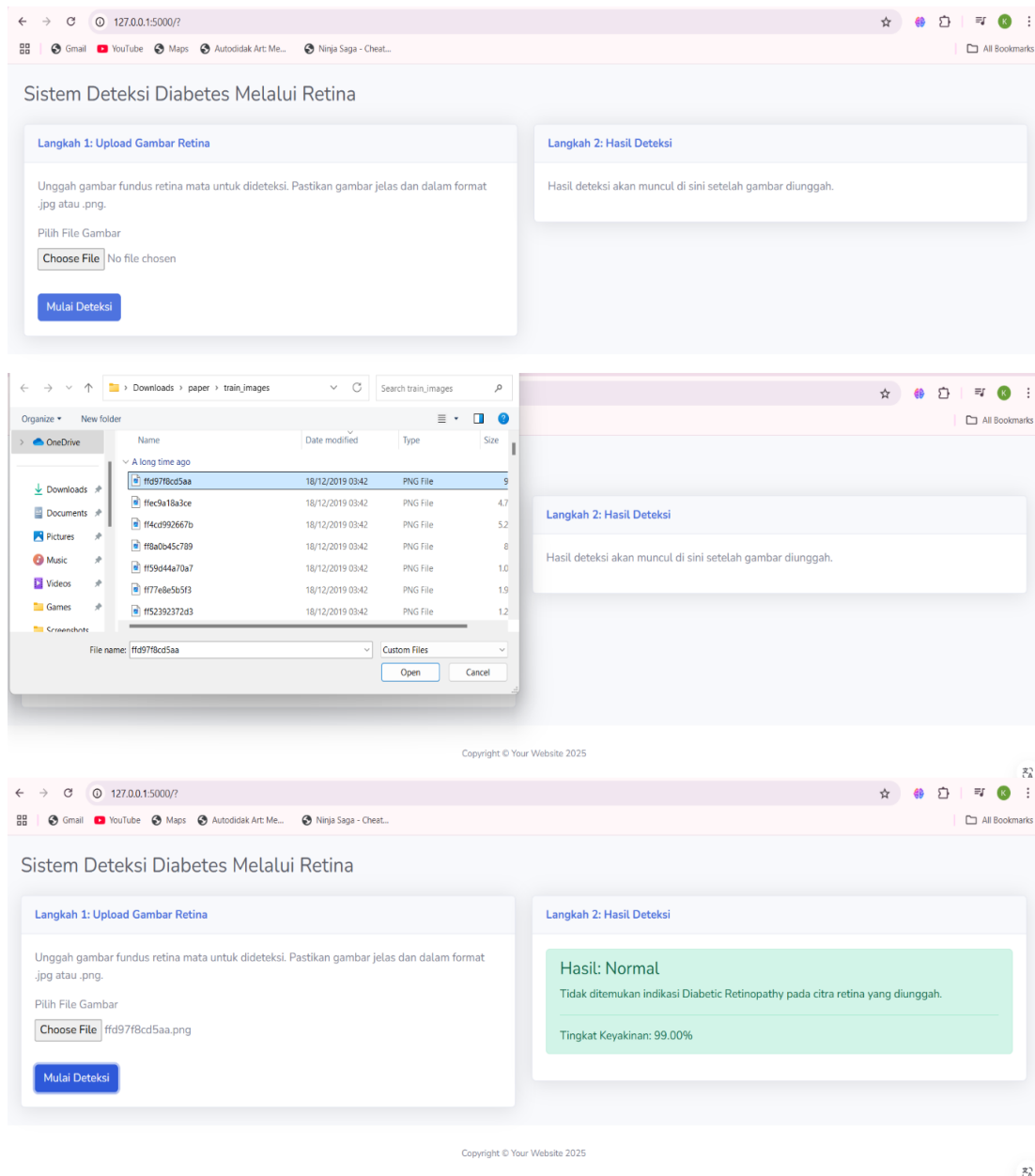
2.4. Klasifikasi dengan Random Forest

Dataset fitur dibagi menjadi 80% data latih dan 20% data tes. Teknik SMOTE diterapkan hanya pada data latih. Model Random Forest dari library Scikit-learn [5] kemudian dilatih menggunakan data latih yang telah seimbang. Performa model dievaluasi pada data tes menggunakan metrik Akurasi, Presisi, Recall, F1-Score, dan Confusion Matrix.

3. Hasil dan Pembahasan

3.1. Hasil Implementasi Sistem

Sistem deteksi diabetes berbasis web telah berhasil diimplementasikan secara fungsional. Pengguna dapat mengunggah gambar dan menerima hasil prediksi melalui antarmuka yang bersih dan responsif. Tampilan antarmuka utama dari sistem disajikan pada Gambar 4.



Gambar 4. Alur Penggunaan Sistem Deteksi Berbasis Web.

Tampilan awal sistem tempat pengguna dapat memilih file. Proses pemilihan file citra retina dari komputer lokal. Tampilan hasil setelah sistem berhasil melakukan prediksi, menunjukkan kelas ("Normal") dan tingkat keyakinan model.

3.2. Hasil Kinerja Model

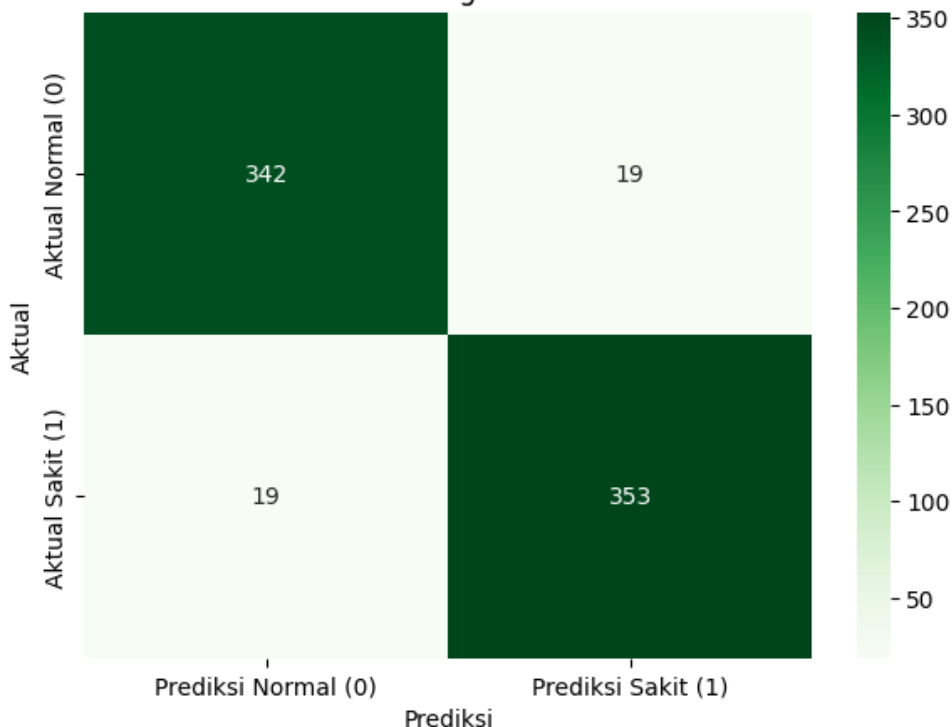
Setelah melalui proses rekayasa fitur yang komprehensif dan penyeimbangan data dengan SMOTE, model Random Forest yang dilatih menunjukkan peningkatan performa yang sangat signifikan. Model final mampu mencapai akurasi keseluruhan sebesar **94.82%** pada data tes. Laporan klasifikasi detail pada Tabel 1 menunjukkan bahwa model tidak hanya akurat, tetapi juga memiliki performa yang sangat seimbang antara kemampuan mendeteksi kelas "Normal" (recall 0.95) dan kelas "Sakit" (recall 0.95).

Tabel 1. Laporan Klasifikasi Model Final

Kelas	Precision	Recall	F1-Score	Support
0 (Normal)	0.95	0.95	0.95	361
1 (Sakit)	0.95	0.95	0.95	372

Visualisasi *Confusion Matrix* pada Gambar 5 lebih lanjut mengkonfirmasi kinerja model yang kuat, dengan jumlah prediksi benar yang sangat dominan.

Confusion Matrix dengan Fitur Tambahan



Gambar 5. Confusion Matrix Model Final.

Matriks menunjukkan performa model pada data tes setelah dilatih dengan fitur lengkap dan teknik SMOTE. Angka pada diagonal menunjukkan jumlah prediksi yang benar.

3.3. Pembahasan

Peningkatan akurasi yang drastis menjadi 94.82% membuktikan efektivitas dari metodologi yang diterapkan. Keberhasilan ini didorong oleh dua faktor kunci. Pertama, penerapan SMOTE berhasil mengatasi masalah bias pada model, memastikannya mampu mempelajari karakteristik kelas minoritas secara efektif. Kedua, dan yang paling signifikan, adalah penambahan fitur tekstur (GLCM) yang melengkapi fitur warna dan bentuk. Kombinasi fitur yang kaya ini memberikan informasi yang cukup bagi algoritma Random Forest untuk membangun model klasifikasi yang andal. Hasil ini menunjukkan bahwa pendekatan *machine learning* klasik, jika didukung oleh rekayasa fitur yang cermat, tetap menjadi alternatif yang sangat valid dan berkinerja tinggi untuk tugas klasifikasi gambar medis.

4. Kesimpulan

Penelitian ini telah berhasil mengimplementasikan dan menganalisis performa sebuah sistem deteksi Diabetes Mellitus berbasis web. Melalui kombinasi rekayasa fitur manual yang komprehensif dan penerapan teknik penyeimbangan data SMOTE, model klasifikasi Random Forest yang dibangun mampu mencapai akurasi yang sangat tinggi sebesar 94.82%. Temuan ini menegaskan bahwa pendekatan *machine learning* klasik dapat menjadi alat skrining awal yang efektif dan andal untuk *Diabetic Retinopathy* jika didukung oleh proses rekayasa fitur dan penanganan data yang kuat.

5. Referensi

- [1]. Kaggle, "APTOS 2019 Blindness Detection," 2019. [Online]. Available: <https://www.kaggle.com/c/aptos2019-blindness-detection>. [Accessed: June 12, 2025].
- [2]. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321-357, 2002.
- [3]. L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5-32, 2001.
- [4]. G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.
- [5]. F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, 2011.
- [6]. M. Grinberg, "Flask Web Development: Developing Web Applications with Python," 2nd ed. O'Reilly Media, 2018.
- [7]. Start Bootstrap, "SB Admin 2 - Bootstrap Admin Theme." [Online]. Available: <https://startbootstrap.com/theme/sb-admin-2>. [Accessed: June 12, 2025].
- [8]. T. M. Aslam, et al., "A review of deep learning models for screening, detection, and diagnosis of diabetic retinopathy," *Eye*, vol. 35, no. 1, pp. 117-130, 2021.
- [9]. A. Sopharak, B. Uyyanonvara, S. Barman, and T. H. Williamson, "Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using computer vision techniques," *Computerized Medical Imaging and Graphics*, vol. 32, no. 8, pp. 720-727, 2008.