

Penerapan *Machine Learning* Berbasis *Support Vector Machine* dan *Random Forest* untuk Klasifikasi Karies pada Citra Gigi

Bima Variandana Putra¹, Fawaidul Badri², Ngatmari³

^{1,2,3}Fakultas Teknik, Prodi Teknik Elektro, Universitas Islam Malang

variandana.sei@gmail.com, fawaidulbadri@unisma.ac.id, ngatmari@unisma.ac.id

Abstract

This study aims to design a classification program for both algorithms, namely the Random Forest algorithm and the Support Vector Machine. This study uses 3 levels of dental caries severity, namely mild, moderate, and severe. The method uses the Gray-Level Co-occurrence Matrix (GLCM) method which is implemented to extract texture features in images, which include mean, contrast, energy, and homogeneity parameters. The system performance is then tested and evaluated comprehensively through a comparison of the Random Forest algorithm and Support Vector Machine (SVM) based on the Stratified 25-Fold Cross Validation technique. The experimental results prove that the Random Forest algorithm shows superior performance with the highest accuracy of 84.50%, while the Support Vector Machine algorithm produces an accuracy of 83.40%. The characteristics of the contrast and homogeneity features are proven to be effective in distinguishing the severity of dental caries in dental images.

Keywords: Dental Caries Classification, Gray-Level Co-occurrence Matrix, Random Forest, Support Vector Machine, Cross-Validation.

Abstraksi

Penelitian ini bertujuan untuk merancang program klasifikasi kedua algoritma, yakni algoritma *Random Forest* dan *Support Vector Machine*. Penelitian ini menggunakan 3 tingkatan keparahan karies gigi yakni kelas ringan, sedang, dan berat. Untuk Metodenya menggunakan metode *Gray-Level Co-occurrence Matrix (GLCM)* yang diimplementasikan untuk mengekstrak fitur tekstur pada citra, yang meliputi parameter mean, contrast, energy, dan homogeneity. Performa sistem kemudian diuji dan dievaluasi secara komprehensif melalui komparasi algoritma *Random Forest* dan *Support Vector Machine (SVM)* berbasis teknik *Stratified 25-Fold Cross Validation*. Hasil eksperimen membuktikan bahwa algoritma *Random Forest* menunjukkan performa lebih unggul dengan capaian akurasi tertinggi sebesar 84,50%, sedangkan algoritma *Support Vector Machine* menghasilkan akurasi sebesar 83,40%. Karakteristik fitur *contrast* dan *homogeneity* terbukti efektif untuk membedakan tingkat keparahan karies gigi pada citra gigi.

Kata Kunci: Klasifikasi Karies Gigi, Gray-Level Co-occurrence Matrix, Random Forest, Support Vector Machine, Cross-Validation

I. PENDAHULUAN

Di era modern ini, banyak manusia modern yang mengabaikan atau kurang peduli terhadap kesehatan mulut dan giginya. Hal ini dibuktikan dengan tidak sedikit manusia modern yang giginya mengalami infeksi pada jaringan keras gigi yang biasa disebut karies gigi. Hal ini bisa ditandai dengan munculnya lubang dan pembusukan lapisan gigi berwarna hitam pada lapisan gigi yang memiliki tingkat keparahan tertentu tiap orangnya. Tentunya kondisi ini jika dibiarkan lama akan menjadikan kerusakan gigi semakin meluas dan bisa jadi gigi lepas atau hancur secara keseluruhan karena karies tersebut.

Dari Permasalahan tersebut peneliti terfikirkan untuk membuat sebuah system untuk mengklasifikasikan tingkat keparahan karies gigi pada manusia dengan menerapkan *Machine Learning* yang berbasis algoritma *Random Forest* dan *Support Vector Machine*. Selama ini proses diagnosis karies gigi dilakukan dengan cara menggunakan sonde dan kaca mulut untuk mengecek kondisi keparahan karies gigi. Meskipun pada saat ini sudah ada alat yang lebih modern seperti *radiografi periapical*, tetapi metode pemeriksaan ini seringkali terkendala oleh keterbatasan ekonomi dan juga distribusi tenaga medis yang tidak merata di tiap daerah. Yang

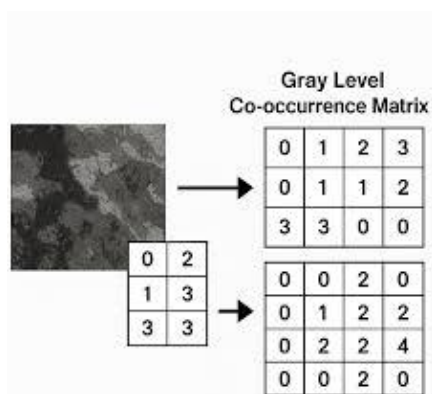
dimana menjadikan salah satu hambatan terbesar orang untuk melakukan pemeriksaan karies gigi.

Oleh karena itu penelitian ini bertujuan membuat sistem klasifikasi sederhana untuk menggolongkan karies gigi termasuk dalam kondisi tingkat keparahan tertentu dengan berbasis *computer vision* dan *Machine Learning*. Sistem ini tidak bertujuan untuk mengungguli alat modern yang sudah ada seperti *radiografi periapical*, tetapi sistem bertujuan untuk mengklasifikasikan kedua algoritma yakni algoritma *Random Forest* dan *Support Vector Machine* dan juga sebagai media alternatif untuk pemeriksaan karies gigi bila penelitian telah dikembangkan sampai tahap tertentu. Sistem ini menggunakan metode *Gray Level Co-occurrence Matrix* (GLCM) dan skema pengujian teknik validasi silang berlapis (*Stratified K-Fold Cross Validation*). Yang dirancang menggunakan coding program *Phyton* di dalam perangkat lunak *PyCharm* untuk mengintegrasikan *pandas library* dan hasil antarmuka akan ditampilkan berbasis web menggunakan *library Streamlit*.

II. METODE PENELITIAN

Metode Penelitian ini menggunakan metode eksperimen dengan pendekatan kuantitatif untuk melakukan klasifikasi tingkat karies gigi menggunakan algoritma *Random Forest* dan *Support Vector Machine* (SVM). Proses penelitian dilakukan secara berurutan mulai dari pengumpulan dataset citra gigi, preprocessing citra, ekstraksi fitur tekstur menggunakan *Gray Level Co-occurrence Matrix* (GLCM). Implementasi sistem dilakukan menggunakan bahasa pemrograman *Python* dengan bantuan *library OpenCV*, *NumPy*, *Scikit-Learn*, *Pandas*, *Streamlit*, dan *Matplotlib*.

Metode penelitian menggunakan *Gray Level Co-occurrence Matrix* (GLCM) karena cocok untuk memetakan kondisi citra sesuai skala abu-abu.



Gambar 1. Gambar Metode GLCM

Pada tahap *preprocessing* dilakukan menggunakan *resize*, normalisasi,. Setelah itu dilakukan ekstraksi fitur tekstur menggunakan metode GLCM untuk memperoleh karakteristik citra berupa *contrast*, *correlation*, *energy*, dan *homogeneity*.

Nilai *contrast* dihitung menggunakan Persamaan (1).

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

Nilai *homogeneity* dihitung menggunakan Persamaan (2).

$$Homogeneity = \sum_{i,j} \frac{P(i, j)}{1 + |i - j|}$$

Keterangan:

$P(i, j)$ merupakan probabilitas pasangan piksel pada matriks GLCM.

Proses klasifikasi dilakukan menggunakan algoritma *Random Forest* dan *Support Vector Machine*. *Random Forest* bekerja menggunakan banyak decision tree untuk meningkatkan akurasi klasifikasi berdasarkan nilai *entropy* dan *information gain*.

Nilai *entropy* dihitung menggunakan Persamaan (3).

$$H(S) = - \sum_{i=1}^c p_i \log_2 p_i$$

Sedangkan *information gain* dihitung menggunakan Persamaan (4).

$$IG(S, A) = H(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} H(S_v)$$

Algoritma SVM digunakan untuk mencari *hyperplane* terbaik dalam memisahkan data antar kelas menggunakan kernel *Radial Basis Function* (RBF).

Kernel RBF dihitung menggunakan Persamaan (5).

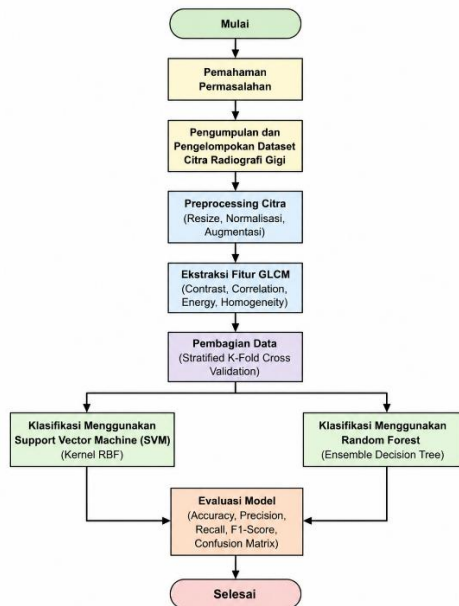
$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$$

Fungsi keputusan SVM ditunjukkan pada Persamaan (6).

$$f(x) = \text{sign} \left(\sum_{i=1}^n \alpha_i y_i K(x_i, x) + b \right)$$

Proses pelatihan dan pengujian model dilakukan menggunakan *stratified 25 k-fold cross validation* untuk menjaga tiap kelas tetap

seimbang pada setiap fold pengujian. Algoritma sistem klasifikasi ditunjukkan pada Algoritma 1.



Gambar 2. Gambar Algoritma Sistem Klasifikasi

1. Pemahaman Permasalahan

Tahap awal dilakukan untuk memahami permasalahan klasifikasi karies gigi serta menentukan tujuan penelitian dan metode yang digunakan.

2. Pengumpulan dan Pengelompokan Dataset

Dataset citra radiografi gigi dikumpulkan dan dikelompokkan berdasarkan tingkat keparahan karies untuk digunakan dalam proses pelatihan dan pengujian model.

3. Preprocessing Citra

Tahap *preprocessing* dilakukan menggunakan *resize*, normalisasi, dan citra untuk meningkatkan kualitas data dan mengurangi risiko *overfitting*.

4. Ekstraksi Fitur GLCM

Ekstraksi fitur dilakukan menggunakan metode *Gray Level Co-occurrence Matrix* (GLCM) untuk memperoleh karakteristik tekstur citra seperti *contrast*, *correlation*, *energy*, dan *homogeneity*.

5. Pembagian Data

Dataset dibagi menggunakan metode *stratified k-fold cross validation* agar distribusi kelas tetap seimbang pada setiap proses pengujian.

6. Klasifikasi Menggunakan SVM

Algoritma **Support Vector Machine (SVM)** digunakan untuk melakukan klasifikasi tingkat karies gigi menggunakan kernel *Radial Basis Function* (RBF).

7. Klasifikasi Menggunakan Random Forest

Algoritma *Random Forest* digunakan untuk melakukan klasifikasi berdasarkan kumpulan *decision tree* guna meningkatkan akurasi prediksi.

8. Evaluasi Model

Hasil klasifikasi dievaluasi menggunakan *confusion matrix*, *accuracy*, *precision*, *recall*, dan *F1-score* untuk mengetahui performa masing-masing metode.

9. Selesai

Tahap akhir penelitian setelah seluruh proses klasifikasi dan evaluasi model selesai dilakukan.

Evaluasi model dilakukan menggunakan *confusion matrix* dengan parameter *accuracy*, *precision*, *recall*, dan *F1-score*. *Accuracy* dihitung menggunakan Persamaan (7).

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

Precision dihitung menggunakan Persamaan (8).

$$Precision = \frac{TP}{TP + FP}$$

Recall dihitung menggunakan Persamaan (9).

$$Recall = \frac{TP}{TP + FN}$$

Sedangkan *F1-score* dihitung menggunakan Persamaan (10).

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Keterangan:

TP merupakan *True Positive*, TN *True Negative*, FP *False Positive*, dan FN *False Negative*.

III. HASIL DAN PEMBAHASAN

1. Hasil Pengujian Model

Pengujian dilakukan menggunakan metode *25-Fold Cross Validation* untuk membandingkan performa algoritma *Random Forest* dan *Support Vector Machine* (SVM) dalam klasifikasi karies gigi. Evaluasi model dilakukan menggunakan *accuracy*, *precision*, *recall*, dan *F1-score*.

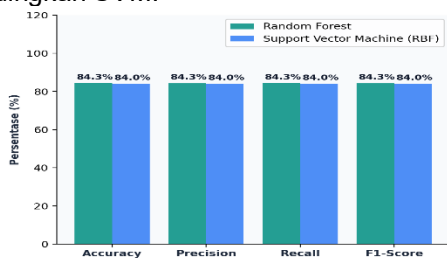
Tabel 1 menunjukkan hasil evaluasi performa kedua algoritma.

Algoritma	Accuracy	Precision	Recall	F1-Score
-----------	----------	-----------	--------	----------

Algoritma	Accuracy	Precision	Recall	F1-Score
Random Forest	84.3%	84.3%	84.3%	84.3%
Support Vector Machine	84.0%	84.0%	84.0%	84.0%

Tabel 1. Tabel Hasil Evaluasi Kedua Algoritma

Berdasarkan hasil pengujian, Random Forest memperoleh performa yang lebih tinggi dibandingkan SVM pada seluruh metrik evaluasi. Random Forest mencapai accuracy sebesar 84.3%, sedangkan SVM memperoleh 84.0%. Selain itu, *precision*, *recall*, dan *F1-score* Random Forest juga menunjukkan hasil yang lebih stabil dibandingkan SVM.

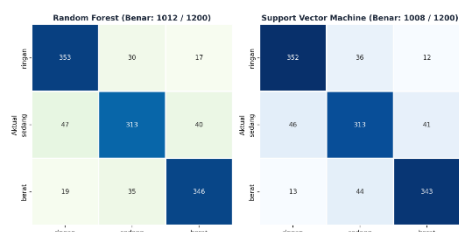


Gambar 3. Gambar Tabel Akurasi Kedua Algoritma

Meskipun selisih performa kedua algoritma relatif kecil, hasil ini menunjukkan bahwa *Random Forest* lebih mampu menangani variasi fitur tekstur citra karies dibandingkan SVM. Hal tersebut terjadi karena *Random Forest* menggunakan pendekatan *ensemble learning* yang lebih stabil terhadap distribusi data yang kompleks dan saling tumpang tindih.

2. Analisis Confusion Matrix

Hasil *confusion matrix* menunjukkan bahwa kedua algoritma mampu mengklasifikasikan data dengan cukup baik, namun masih mengalami kesulitan pada kelas sedang.



Gambar 4. Hasil Confusion Matrix Kedua Algoritma

Pada algoritma *Random Forest*, kelas ringan berhasil diklasifikasikan dengan benar sebanyak 353 data dari total 400 data (dengan *recall* sebesar 0.88). Pada kelas sedang, model berhasil mengklasifikasikan 313 data dengan benar

(dengan *recall* sebesar 0.78), sedangkan pada kelas berat model menghasilkan 346 prediksi benar (dengan *recall* sebesar 0.86).

Sementara itu, pada algoritma *Support Vector Machine*, kelas ringan menghasilkan 352 prediksi benar (dengan *recall* sebesar 0.88). Pada kelas sedang diperoleh nilai yang kembar dengan *Random Forest* yaitu 313 prediksi benar (dengan *recall* sebesar 0.78), sedangkan kelas berat menghasilkan 343 prediksi benar (dengan *recall* sebesar 0.86).

Hasil tersebut menegaskan bahwa kelas sedang menjadi titik utama salah klasifikasi pada kedua algoritma. Kesalahan klasifikasi pada kelas sedang terjadi karena karakteristik tekstur citra berada di fase transisi antara kelas ringan dan berat, sehingga menyebabkan terjadinya tumpang tindih nilai fitur GLCM. Kondisi ini membuat model mengalami hambatan dalam menentukan batas keputusan secara tegas.

Selain itu, kesalahan klasifikasi antar kelas ekstrem (seperti aktual ringan yang diprediksi berat, atau sebaliknya) relatif sangat kecil. Hal ini membuktikan bahwa kedua algoritma tetap memiliki akurasi yang baik dalam membedakan karakteristik tekstur citra yang memiliki perbedaan visual kontras.

3 Analisis Fitur Tekstur GLCM

Kategori Karies	Mean (±Std)	Contrast (±Std)	Energy (±Std)	Homogeneity (±Std)
Ringan	163.05 ± 26.7	40.22 ± 28.7	0.052 ± 0.02	0.387 ± 0.07
Sedang	161.09 ± 19.7	45.22 ± 18.8	0.040 ± 0.01	0.337 ± 0.05
Berat	157.71 ± 20.8	68.43 ± 28.3	0.033 ± 0.01	0.305 ± 0.05

Tabel 2. Tabel Hasil Karakteristik Fitur GLCM per Kategori Karies

Hasil analisis fitur tekstur menunjukkan bahwa tingkat keparahan karies mempengaruhi pola tekstur citra gigi. Nilai *contrast* meningkat pada kelas berat sebesar 68.43 dibandingkan kelas ringan sebesar 40.22. Sebaliknya, nilai *homogeneity* menurun dari 0.3874 pada kelas ringan menjadi 0.3055 pada kelas berat.

Penurunan nilai *homogeneity* menunjukkan bahwa tekstur citra menjadi semakin tidak seragam seiring dengan meningkatnya tingkat kerusakan gigi. Selain itu, nilai *energy* juga mengalami penurunan dari 0.052 pada kelas ringan menjadi 0.033 pada kelas berat, yang menunjukkan bahwa keteraturan pola tekstur

semakin berkurang pada karies dengan tingkat keparahan tinggi.

Hasil tersebut membuktikan bahwa fitur GLCM mampu merepresentasikan perubahan struktur gigi secara kuantitatif. Peningkatan *contrast* dan penurunan *homogeneity* menunjukkan bahwa kerusakan gigi menyebabkan perubahan intensitas piksel dan tekstur citra menjadi lebih kasar serta tidak beraturan.

4 Analisis Waktu Komputasi

Pengujian waktu komputasi dilakukan untuk membandingkan efisiensi kedua algoritma dalam proses pelatihan dan pengujian model.

Algoritma	Total Waktu (detik)	Rata-rata per Fold (detik)
<i>Random Forest</i>	12.1590	0.4488
<i>Support Vector Machine</i>	1.1621	0.0426

Tabel 3. Tabel Waktu Eksekusi Model

Berdasarkan hasil pengujian, SVM memiliki waktu komputasi yang jauh lebih cepat dibandingkan *Random Forest*. SVM hanya membutuhkan total waktu sebesar 1.1621 detik, sedangkan *Random Forest* membutuhkan 12.1590 detik. Dan Nilai rata-rata pembacaan fold untuk SVM sebesar 0.0426 per detik, dan untuk *Random Forest* sebesar 0.4488 per detik. Hal ini menunjukkan bahwa untuk rata-rata pembacaan 25 fold selisih tipis antara kedua algoritma.

Meskipun demikian, *Random Forest* menghasilkan performa klasifikasi yang lebih baik, terutama pada kelas sedang dan berat. Hal ini membuat adanya pertimbangan antara akurasi dan efisiensi waktu komputasi dari tiap algoritma. Jika sistem memprioritaskan kecepatan proses, maka Support Vector Machine lebih optimal. Namun, apabila prioritas utama adalah tingkat akurasi klasifikasi, maka *Random Forest* lebih direkomendasikan.

IV. KESIMPULAN

Rancangan sistem klasifikasi tingkat keparahan karies gigi berbasis citra dalam penelitian ini berhasil dibangun melalui beberapa tahapan utama, yang diawali dengan proses preprocessing (redimensi 128x128 piksel dan grayscale), ekstraksi fitur kuantitatif berupa statistik nilai Mean serta fitur tekstur GLCM (Contrast, Energy, dan Homogeneity), hingga pengujian model menggunakan algoritma Random Forest dan Support Vector Machine (SVM) dengan skenario 25-Fold Cross Validation. Hasil evaluasi perbandingan performa menunjukkan bahwa algoritma Random Forest memiliki efektivitas klasifikasi yang sedikit lebih unggul dan konsisten

dibandingkan SVM pada dataset ini. Random Forest memperoleh akurasi keseluruhan sebesar 84,33% (dengan weighted average Precision 84,60%, Recall 84,33%, dan F1-Score 84,35%), sedangkan SVM memperoleh akurasi sebesar 84,00% (dengan weighted average Precision 84,10%, Recall 84,00%, dan F1-Score 84,00%). Meskipun SVM unggul mutlak dari sisi efisiensi waktu komputasi yang 10 kali lipat lebih cepat (0,0426 detik per fold vs 0,4488 detik per fold milik RF), Random Forest lebih direkomendasikan karena terbukti lebih tangguh meminimalkan misklasifikasi pada area transisi karies kelas sedang.karies gigi.

DAFTAR PUSTAKA

1. Syaripudin, D., Yayat, E., Jayadi, J., & Raharja, A. R. (2025). Klasifikasi Penyakit Periodontal Pada Citra Panoramik Gigi Menggunakan Deep Learning Berbasis ResNet50. *SisInfo : Jurnal Sistem Informasi dan Informatika*, 7(1),
2. Singh, S. B., Laishram, A., Thongam, K., & Singh, K. M. (2024). A Random Forest-based Automatic Classification of Dental Types and Pathologies using Panoramic Radiography Images. *Journal of Scientific & Industrial Research*, 83, 531–543.
3. Siadari, T. S., & Hariyani, Y. S. (2024). Klasifikasi Penyakit Periodontal Pada Citra Gigi Berwarna Menggunakan Arsitektur Berbasis Convolutional Neural Network. *JETT: Jurnal Elektro dan Telekomunikasi Terapan*, 11(2), 13–19.
4. Kristianto, J. (2024). SISTEM IDENTIFIKASI PENYAKIT GIGI MENGGUNAKAN METODE CONVOLUTIONAL NEURAL NETWORK (CNN) BERBASIS MOBILE APPS. (Laporan Tugas Akhir). Universitas Islam Sultan Agung, Semarang.
5. Putra, A. A. (2024). ANALISIS KARIES GIGI MENGGUNAKAN GRAY-LEVEL CO-OCCURENCE MATRIX PADA CITRA RADIOGRAFI PERIAPIKAL. (Tesis). Institut Teknologi Bandung, Bandung.
6. Himel, G. M. S., Islam, M. M., & Hannan, U. H. (2023). *Caries-Spectra: A dataset of Enamel Caries* (Version 2) [Data set]. Mendeley Data.
7. Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6),