



Klasifikasi Penyakit Mata Menggunakan Random Forest Dengan Optimasi Hyperparameter RandomSearchCV

Muh. Fatkhi Alexander, Virgiafan Rido Taufik Adrian, Levi Renov Esprayenduo, Muhammad Naufal*

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Semarang, Indonesia

Email: ¹11202214537@mhs.dinus.ac.id, ²11202214530@mhs.dinus.ac.id, ³11202214512@mhs.dinus.ac.id,

^{4,*}m.naufal@dsn.dinus.ac.id

(* : coressponding author)

Abstrak—Penyakit mata seperti katarak, retinopati diabetik, dan glaukoma merupakan penyebab utama gangguan penglihatan dan kebutaan di seluruh dunia, sehingga deteksi dini melalui analisis citra medis sangat penting untuk mencegah komplikasi dan kehilangan penglihatan permanen. Perkembangan teknologi kecerdasan buatan dan machine learning memberikan peluang besar dalam membantu tenaga medis melakukan diagnosis secara lebih cepat, akurat, dan efisien. Penelitian ini memiliki tujuan untuk mengembangkan model klasifikasi penyakit mata menggunakan algoritma Random Forest dengan optimasi hyperparameter untuk membedakan empat kondisi mata, yaitu katarak, retinopati diabetik, glaukoma, dan normal. Dataset yang digunakan bersumber dari publik dan terdiri dari citra fundus mata yang telah melalui proses preprocessing dan ekstraksi fitur untuk meningkatkan kualitas data. Data dibagi menjadi training dan testing, kemudian dilakukan pelatihan model Random Forest dengan optimasi hyperparameter menggunakan RandomizedSearchCV sebanyak 20 iterasi dan 5-fold cross-validation untuk memperoleh kombinasi parameter terbaik. Model terbaik mencapai akurasi sebesar 80,92% pada data testing dengan nilai ROC-AUC macro sebesar 0,9422, dimana performa terbaik diperoleh pada klasifikasi retinopati diabetik dengan precision 99,54%, recall 99,09%, dan ROC-AUC 1,0000. Selain itu, specificity macro mencapai 93,66%, menunjukkan kemampuan model yang baik dalam mengidentifikasi kasus negatif secara tepat. Hasil penelitian menunjukkan bahwa pendekatan Random Forest dengan optimasi hyperparameter memiliki performa yang sangat baik untuk klasifikasi penyakit mata dan berpotensi untuk diimplementasikan sebagai sistem pendukung diagnosis medis berbasis kecerdasan buatan di fasilitas pelayanan kesehatan.

Kata Kunci: Random Forest; Klasifikasi Penyakit Mata; Citra Medis; Deteksi Dini; RandomizedSearchCV

Abstract—*Eye diseases such as cataracts, diabetic retinopathy and glaucoma are the leading causes of visual impairment and blindness worldwide, so early detection through medical image analysis is essential to prevent complications and permanent vision loss. The development of artificial intelligence and machine learning technology provides great opportunities to help medical personnel carry out diagnoses more quickly, accurately and efficiently. This research aims to develop an eye disease classification model using the Random Forest algorithm with hyperparameter optimization to differentiate four eye conditions, namely cataract, diabetic retinopathy, glaucoma, and normal. The dataset used is sourced from the public and consists of eye fundus images that have gone through preprocessing and feature extraction to improve data quality. The data was divided into training and testing, then the Random Forest model was trained with hyperparameter optimization using RandomizedSearchCV for 20 iterations and 5-fold cross-validation to obtain the best parameter combination. The best model achieved an accuracy of 80.92% on testing data with a macro ROC-AUC value of 0.9422, where the best performance was obtained in the classification of diabetic retinopathy with a precision of 99.54%, recall of 99.09%, and ROC-AUC of 1.0000. In addition, macro specificity reached 93.66%, indicating the model's good ability to identify negative cases correctly. The research results show that the Random Forest approach with hyperparameter optimization has excellent performance for eye disease classification and has the potential to be implemented as an artificial intelligence-based medical diagnosis support system in health care facilities.*

Keywords: *Random Forest; Classification of Eye Diseases; Medical Images; Early Detection; RandomizedSearchCV*

1. PENDAHULUAN

Kemajuan teknologi yang berkembang pesat membawa pengaruh besar pada banyak aspek kehidupan masyarakat, termasuk bidang medis. Salah satu area yang mengalami perkembangan pesat adalah penggunaan teknologi untuk menunjang proses diagnosis penyakit melalui pengolahan data medis. Namun demikian, tantangan yang dihadapi fasilitas kesehatan adalah keterbatasan sistem yang dapat menghasilkan informasi strategis untuk mendukung keputusan medis, padahal kebutuhan akan sistem yang kompetitif dan efisien sangat mendesak. Pentingnya pengelolaan data medis dalam skala besar tidak bisa diabaikan, terutama dalam konteks diagnosis di rumah sakit dan klinik mata, karena hal ini berdampak pada akumulasi informasi medis yang substansial [1]. Oleh karena itu, diperlukan implementasi teknologi yang mampu memfasilitasi ekstraksi dan analisis informasi baru dari kumpulan data medis yang belum teridentifikasi, sehingga dapat meningkatkan kemampuan fasilitas kesehatan dalam menegakkan diagnosis berdasarkan data [1].

Gangguan penglihatan merupakan permasalahan kesehatan yang memiliki dampak global dan signifikan. Data dari World Health Organization (WHO) menunjukkan bahwa terdapat sekitar 2,2 miliar individu di seluruh dunia yang mengalami berbagai tingkat gangguan penglihatan, dengan lebih dari 1 miliar kasus yang sebenarnya dapat dihindari atau belum mendapatkan penanganan yang tepat. Berbagai kondisi seperti katarak, penyakit mata akibat diabetes, dan glaukoma berkontribusi sebagai penyebab utama terjadinya kebutaan.

Mengingat besarnya jumlah data gambar mata yang tersedia, proses diagnosis tidak dapat lagi mengandalkan cara konvensional sebab memerlukan durasi yang panjang serta berpotensi menghasilkan diagnosis yang tidak akurat serta tidak konsisten. Kondisi saat ini menuntut adanya sistem diagnosis otomatis yang dapat memberikan hasil dengan kualitas optimal. Namun, sebagian besar fasilitas kesehatan masih menggunakan pendekatan manual dalam menganalisis data



medis pasien sebelum memberikan diagnosis. Hal ini tentunya belum optimal mengingat perlunya teknologi informasi yang dapat mengelola dan menganalisis data medis berukuran besar untuk menghasilkan informasi tambahan yang dapat meningkatkan nilai diagnosis. Oleh karena itu, dibutuhkan sebuah metode yang dapat mengekstrak dan mengolah data secara efektif untuk menghasilkan output informasi yang lebih akurat dan berkualitas, sejalan dengan peningkatan volume data yang tersedia.

Untuk mengatasi tantangan tersebut, teknologi Data Mining diterapkan sebagai solusi yang menjanjikan. Data Mining merupakan kumpulan proses yang digunakan untuk mengumpulkan data dalam jumlah besar dengan tujuan menemukan pola atau hubungan yang mungkin terlewatkan sebelumnya [2]. Tahapan-tahapan dalam Data Mining meliputi ekstraksi nilai tambahan dari dataset yang ada, mencakup berbagai informasi seperti klasifikasi, estimasi, prediksi, asosiasi, dan pengelompokan data.

Dalam konteks klasifikasi penyakit mata, pengolahan data dapat dilakukan menggunakan pendekatan klasifikasi berbasis Data Mining. Metode ini memungkinkan pengembangan model prediktif yang sistematis dengan membangun model klasifikasi yang memasukkan data terkait ke dalam sistem. Setiap algoritma dirancang dengan mekanisme tersendiri untuk mengidentifikasi model terbaik yang dapat menggambarkan korelasi antara atribut dan label kelas pada dataset yang digunakan [2].

Dalam ranah data mining, terdapat beragam metode klasifikasi yang sudah banyak diterapkan secara luas untuk keperluan sistem informasi pendukung keputusan medis. Sejumlah metode yang kerap diaplikasikan dalam konteks prediksi meliputi Naive Bayes Classifier, Decision Tree, Neural Network, serta K-Nearest Neighbour. Dari sekian banyak metode yang ada, Random Forest menjadi salah satu yang paling banyak digunakan dalam ranah klasifikasi. Studi [3] mengungkapkan bahwa keunggulan utama dari algoritma ini ada pada kapabilitasnya mengelola data yang memiliki dimensi tinggi dan resistensinya terhadap permasalahan overfitting. Algoritma Random Forest menunjukkan performa optimal ketika diterapkan pada data dunia nyata yang bersifat kompleks [4]. Random Forest tergolong dalam kategori algoritma ensemble dengan performa yang dapat menyamai Decision Tree maupun Neural Network.

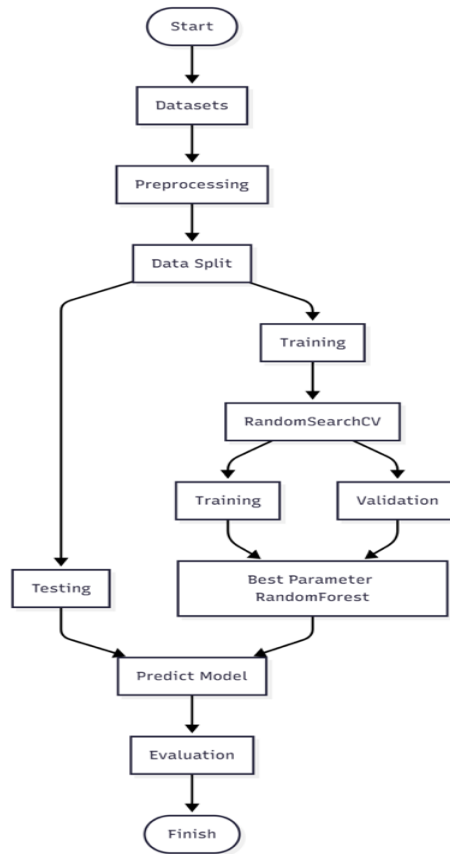
Sejumlah riset telah mendemonstrasikan keefektifan algoritma Random Forest dalam berbagai domain aplikasi. Riset yang dikerjakan oleh [5] memusatkan perhatian pada klasifikasi untuk keperluan diagnosis penyakit, dengan kesimpulan yang menyebutkan bahwa implementasi Random Forest mampu memberikan bantuan medis dalam proses diagnosis sehingga dapat menghasilkan peningkatan kualitas pelayanan yang lebih optimal. Riset sebelumnya yang berkaitan dengan prediksi penyakit diabetes mengimplementasikan algoritma Random Forest menunjukkan hasil bahwa penerapan klasifikasi menggunakan Random Forest menghasilkan tingkat akurasi yang tetap baik baik pada dataset berukuran besar maupun kecil [6]. Implementasi algoritma Random Forest juga telah dilakukan dalam klasifikasi untuk penyakit mata [7]. Algoritma Random Forest telah banyak diterapkan dalam berbagai sistem pendukung keputusan, di antaranya: sistem pendukung keputusan untuk menangani penyakit kronis [8], pengklasifikasian kanker [9], sistem diagnosis yang berbasis pada gejala penyakit [10], serta sistem prediksi yang memanfaatkan data medis sebagai basis [11].

Berdasarkan kajian penelitian-penelitian terkait yang ditemukan, sebagian besar penelitian terfokus pada satu penyakit mata saja dan optimasi hyperparameter seringkali tidak dilakukan secara sistematis. Penelitian ini menerapkan teknik Data Mining pada dataset penyakit mata dengan menggunakan supervised learning melalui algoritma Random Forest. Maka penelitian ini adalah sebagai bahan pertimbangan fasilitas kesehatan dalam pengambilan keputusan ketika melakukan screening awal penyakit mata sehingga dapat mengatur prioritas pasien dengan harapan agar deteksi dini penyakit mata dapat meningkat dan kasus kebutaan dapat dikurangi.

2. METODOLOGI PENELITIAN

2.1 Alur Metodologi Penelitian

Keseluruhan metodologi penelitian dari tahap awal hingga evaluasi akhir dapat divisualisasikan dalam bentuk *flowchart* yang menggambarkan alur sistematis dan sekuensial dari setiap proses. *Flowchart* ini memberikan gambaran komprehensif mengenai arsitektur sistem klasifikasi penyakit mata yang dikembangkan dalam penelitian.

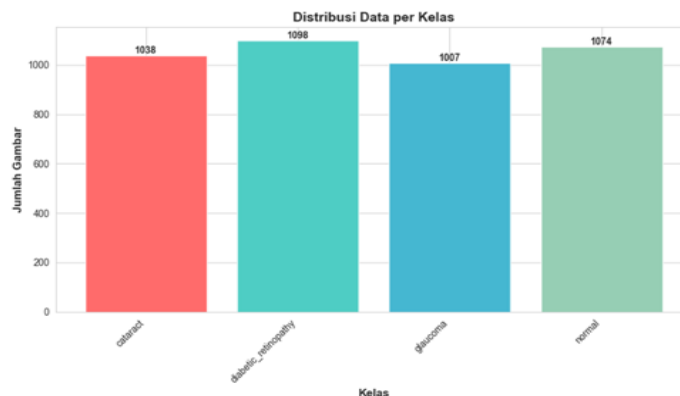


Gambar 1. Alur Diagram Penelitian

Alur dimulai dari pengumpulan dataset penyakit mata empat kategori, dilanjutkan *preprocessing* (*resize, normalisasi, flattening*) untuk mengubah citra menjadi vektor fitur numerik. Data dibagi 80:20 untuk *training* dan *testing*. *Data training* masuk ke *RandomizedSearchCV* dengan loop iteratif *training-validation* untuk mencoba 20 kombinasi *hyperparameter* menggunakan *5-fold cross-validation*. Kombinasi terbaik dipilih dan digunakan untuk memprediksi *data testing*. Hasil dievaluasi dengan berbagai metrik (*accuracy, precision, recall, F1-score, specificity, ROC-AUC*) untuk mengukur performa model secara komprehensif. Penelitian ini menggunakan bahasa pemrograman Python dengan spesifikasi hardware processor Apple M4 Chip, memory 16 GB RAM, dan sistem operasi macOS 15.4 (Sequoia) pada platform Mac mini.

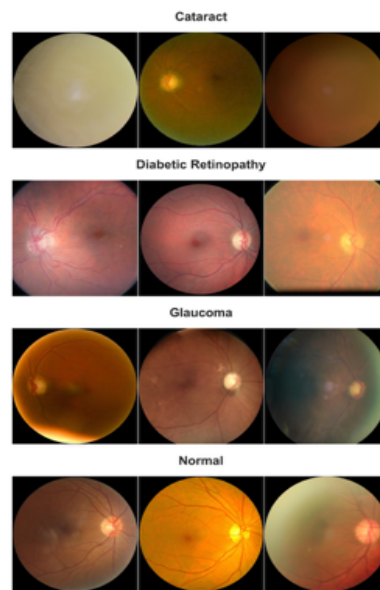
2.2 Dataset Penelitian

Penelitian ini menggunakan dataset citra penyakit mata yang terdiri dari empat kategori yaitu *cataract* (katarak), *diabetic retinopathy* (retinopati diabetik), *glaucoma* (glaukoma), dan normal (mata sehat) [12]. Dataset diperoleh dari repositori publik yang berisi gambar fundus mata dan foto klinis mata dengan berbagai kondisi patologis. Untuk memvisualisasikan komposisi dataset, dibuat bar chart yang menampilkan jumlah gambar pada setiap kategori. Visualisasi ini memberikan gambaran jelas mengenai proporsi data dan mengkonfirmasi bahwa tidak terjadi ketidakseimbangan ekstrem yang dapat mempengaruhi performa model secara signifikan dapat dilihat pada Gambar 2.



Gambar 2. Distribusi Data Setiap Kelas

Grafik pada Gambar 2 menggambarkan total dataset yang digunakan berjumlah 4217 gambar dengan distribusi sebagai berikut: kategori cataract sebanyak 1038 gambar (24.61%), diabetic retinopathy sebanyak 1098 gambar (26.04%), glaucoma sebanyak 1007 gambar (23.88%), dan normal sebanyak 1074 gambar (25.47%). Distribusi data yang relatif seimbang ini sangat menguntungkan karena dapat mengurangi bias dalam proses pembelajaran model dan memastikan setiap kategori memiliki representasi yang memadai. Setiap gambar dalam dataset memiliki dimensi yang bervariasi dan menampilkan karakteristik visual yang berbeda untuk setiap jenis penyakit. Gambar cataract menunjukkan kekeruhan pada lensa mata yang tampak sebagai area putih atau abu-abu. Gambar diabetic retinopathy menampilkan pendarahan retina, eksudat, atau mikroaneurisma yang merupakan tanda kerusakan pembuluh darah. Gambar glaucoma memperlihatkan perubahan pada diskus optikus seperti cup-to-disc ratio yang meningkat. Sedangkan gambar normal menunjukkan struktur mata yang sehat tanpa tanda-tanda patologi. Untuk memberikan gambaran visual konkret tentang karakteristik setiap kategori penyakit, ditampilkan sample gambar dari masing-masing kelas ditampilkan pada Gambar 3 berikut.



Gambar 3. Sample Gambar Dari Setiap Kelas

2.3 Prapemrosesan Data

Preprocessing merupakan tahapan krusial dalam *pipeline machine learning* berbasis citra untuk mengubah data mentah menjadi format optimal [13]. Gambar-gambar dalam dataset memiliki dimensi heterogen sehingga dilakukan *resize* menjadi 64x64 piksel menggunakan interpolasi *bilinear*, dengan ukuran ini dipilih berdasarkan *trade-off* antara preservasi informasi visual dengan efisiensi komputasi. Nilai piksel yang awalnya dalam rentang 0-255 dinormalisasi menjadi 0-1 dengan pembagian terhadap 255, dimana normalisasi ini mempercepat konvergensi pembelajaran dan membuat model lebih robust terhadap variasi pencahayaan.

Random Forest bekerja dengan data tabular satu dimensi sehingga diperlukan transformasi *flattening* untuk mengubah gambar 3D menjadi vektor 1D. Setiap gambar 64x64 piksel dengan 3 channel RGB (dimensi $64 \times 64 \times 3$) di-flatten menjadi vektor 1D sepanjang 12288 elemen, dimana setiap elemen merepresentasikan satu piksel dari satu channel warna yang diperlakukan sebagai fitur independen oleh *Random Forest*. Selain itu, dilakukan *label encoding* untuk mengubah label kategori tekstual menjadi representasi numerik dengan *mapping*: *cataract*=0, *diabetic_retinopathy*=1, *glaucoma*=2, dan *normal*=3.

Selanjutnya dilakukan tahapan pembagian data dimana dataset yang telah melalui *preprocessing* dibagi menjadi dua subset yaitu data *training* dan data *testing* dengan proporsi 80:20 menggunakan teknik *stratified splitting* yang memastikan proporsi setiap kelas pada data *training* dan *testing* tetap sama dengan proporsi dataset keseluruhan. Dari total 4217 gambar, sebanyak 3373 gambar (80%) dialokasikan sebagai data training yang digunakan untuk melatih model mempelajari pola dan karakteristik setiap kategori penyakit, sedangkan sisanya 844 gambar (20%) dijadikan data testing untuk mengevaluasi performa model pada data yang belum pernah dilihat selama proses training.

2.4 Modeling

2.4.1 Algoritma Random Forest

Random Forest merupakan algoritma *ensemble learning* yang membuat multiple decision tree selama proses training dan menghasilkan prediksi berdasarkan voting mayoritas dari seluruh tree untuk masalah klasifikasi [14], [15]. Setiap decision tree dalam forest dibangun menggunakan subset data yang berbeda melalui bootstrap sampling, dan pada setiap node splitting, hanya subset acak dari fitur yang dipertimbangkan untuk menentukan split terbaik.

2.4.2 RandomizedSearchCV

Hyperparameter tuning dilakukan menggunakan RandomizedSearchCV untuk menemukan kombinasi optimal [16]. Berbeda dengan Grid Search yang mencoba semua kombinasi, Randomized Search melakukan sampling secara acak dari distribusi hyperparameter, lebih efisien untuk hyperparameter space yang besar. Parameter-parameter yang dituning meliputi: *n_estimators*: Jumlah decision tree dalam forest [50, 100, 150, 200]. Semakin banyak tree meningkatkan performa dengan mengurangi variance, tetapi meningkatkan waktu komputasi secara linear. *max_depth*: Kedalaman maksimum setiap tree [10, 20, 30, None]. Parameter ini mengontrol kompleksitas tree dimana None berarti tree berkembang tanpa batasan sampai leaf pure. Nilai terlalu besar dapat menyebabkan overfitting.

min_samples_split: Jumlah minimum sampel untuk split pada internal node [2, 5, 10]. Nilai lebih besar membuat tree lebih konservatif, berfungsi sebagai regularization untuk mencegah overfitting. *min_samples_leaf*: Jumlah minimum sampel di leaf node [1, 2, 4]. Parameter ini memaksa setiap leaf memiliki sampel minimum, mencegah prediksi berdasarkan sampel terlalu sedikit. *max_features*: Jumlah fitur untuk best split di setiap node ['sqrt', 'log2']. 'sqrt' berarti $\sqrt{12288} \approx 110$ fitur, 'log2' berarti $\log_2(12288) \approx 13$ fitur. Pembatasan ini meningkatkan diversity antar tree. *bootstrap*: Boolean untuk bootstrap sampling [True, False]. True menggunakan sampling with replacement untuk training set setiap tree, False menggunakan seluruh dataset.

2.4.3 K-Fold Cross Validation

RandomizedSearchCV dikonfigurasi untuk mencoba 20 kombinasi *hyperparameter* dengan *5-fold cross-validation*. Data *training* dibagi menjadi 5 *fold* dengan *stratified splitting*. Proses *training-validation* dilakukan 5 kali dimana 4 *fold* untuk training dan 1 *fold* untuk validasi, dengan *fold* validasi dirotasi. *Cross-validation score* dihitung sebagai rata-rata *accuracy* dari 5 *validation fold*. Kombinasi dengan score tertinggi dipilih sebagai konfigurasi optimal. Total terdapat 20 kombinasi $\times 5 \text{ fold} = 100$ model fit. Parameter tambahan: *n_jobs* = -1 untuk *parallel processing* dan *random_state* = 42 untuk *reproducibility*.

2.5 Evaluasi Model

Model terbaik hasil hyperparameter tuning dievaluasi menggunakan data testing yang terpisah dan belum pernah dilihat selama proses training maupun validation. Evaluasi dilakukan menggunakan berbagai metrik untuk mengukur performa model dari berbagai aspek dan memberikan gambaran komprehensif tentang kemampuan model.

2.5.1 Confusion Matrix

Confusion matrix dibuat untuk melihat detail distribusi prediksi model pada setiap kelas [17]. Matrix ini menampilkan jumlah prediksi untuk setiap kombinasi *actual* label dan *predicted* label, memberikan *insight* tentang pola kesalahan klasifikasi dan kelas mana yang sering tertukar satu sama lain. Diagonal utama merepresentasikan *True Positive* (prediksi positif yang benar) untuk masing-masing kelas, sedangkan elemen off-diagonal merepresentasikan kesalahan klasifikasi dimana instance kelas tertentu salah diprediksi sebagai kelas lain. Struktur confusion matrix untuk klasifikasi 4 kelas ditunjukkan pada Gambar 4.

		Predicted Class			
		C ₁	C ₂	...	C _N
Actual Class	C ₁	C _{1,1}	FP	...	C _{1,N}
	C ₂	FN	TP	...	FN

	C _N	C _{N,1}	FP	...	C _{N,N}

Gambar 4. [2] Confusion Matrix Diagram

Dari confusion matrix, didapat *True Positive* (TP) = prediksi positif yang benar, *True Negative* (TN) = prediksi negatif yang benar, *False Positive* (FP) = prediksi positif yang salah, dan *False Negative* (FN) = prediksi negatif yang salah. yang digunakan untuk menghitung berbagai metrik klasifikasi dasar untuk setiap kelas maupun secara keseluruhan menggunakan formula berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

Accuracy mengukur proporsi prediksi benar dari total prediksi, selanjutnya *precision* mengukur akurasi prediksi positif, yaitu proporsi prediksi positif yang benar, metrik ini penting untuk mengetahui seberapa akurat model ketika memprediksi suatu penyakit. *Recall (Sensitivity)* mengukur sensitivitas model dalam menemukan kelas positif. Metrik ini krusial dalam aplikasi medis untuk memastikan tidak ada kasus penyakit yang terlewat. F1-Score merupakan *harmonic mean* dari *precision* dan *recall* yang memberikan *single metric* untuk keseimbangan keduanya. Metrik-metrik ini dihitung per kelas untuk melihat performa pada setiap kategori penyakit, serta dihitung *macro average* (rata-rata sederhana antar kelas) dan *weighted average* (rata-rata berbobot berdasarkan jumlah sampel per kelas) untuk performa keseluruhan. Specificity menunjukkan kapabilitas model dalam mengidentifikasi kelas negatif dengan akurat, atau dengan kata lain kemampuan model dalam menghindari *false alarm*. Perhitungan Specificity dilakukan untuk setiap kelas menggunakan *confusion matrix* dengan formula:

$$Specificity = \frac{TN}{TN+FP} \tag{5}$$

Dimana TN (*True Negative*) merupakan total kelas negatif yang berhasil diprediksi dengan tepat, dan FP (*False Positive*) merupakan total kasus negatif yang keliru diprediksi sebagai positif. Metrik ini memiliki peranan penting dalam konteks *medical screening* karena *false positive* dapat memicu tingkat kecemasan yang tidak perlu (*overdiagnosis*, *unnecessary treatment*, dan *psychological distress* pada pasien. Specificity yang tinggi mengindikasikan bahwa model dapat dipercaya ketika memberikan prediksi negatif.

2.5.2 ROC-AUC

Receiver Operating Characteristic (ROC) curve digunakan untuk setiap kelas dengan memvisualisasikan *True Positive Rate* terhadap *False Positive Rate* pada berbagai nilai ambang batas klasifikasi [18]. *Area Under Curve (AUC)* dihitung sebagai metrik teragregasi yang merangkum kinerja model *across all possible thresholds*. ROC-AUC mengukur kemampuan model dalam membedakan antar kelas secara keseluruhan, tidak tergantung pada threshold spesifik. Nilai AUC berkisar 0-1 dimana 0.5 menunjukkan *random classifier* dan 1.0 menunjukkan *perfect classifier*. Metrik ini *robust* terhadap *class imbalance* dan memberikan evaluasi yang lebih komprehensif dibandingkan *accuracy*.

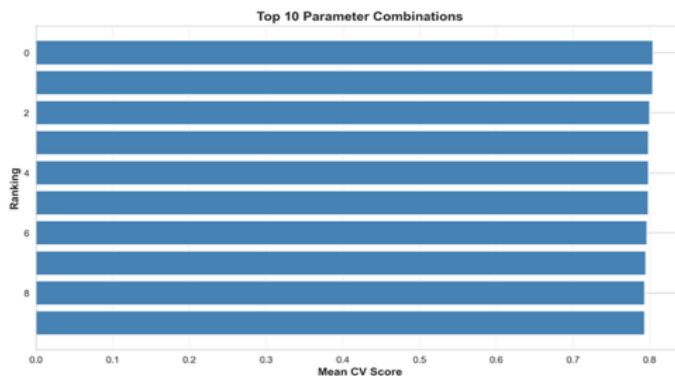
3. ANALISA DAN PEMBAHASAN

3.1 Preprocessing

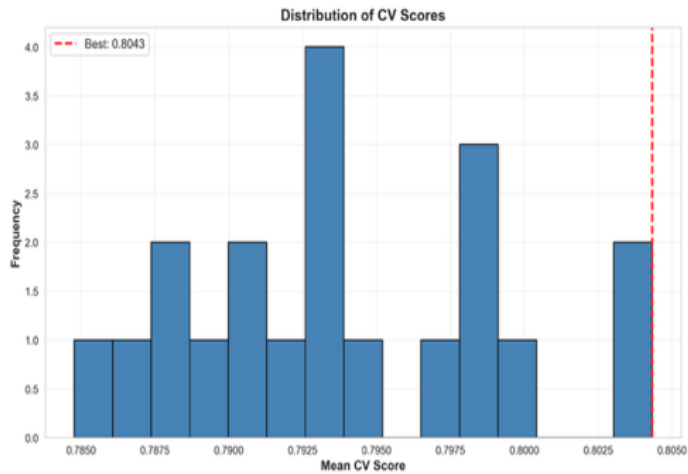
Proses preprocessing berhasil mengolah 4217 gambar menjadi ukuran uniform 64x64x3 piksel melalui resize, normalisasi (0-1), dan flattening. Hasil akhir adalah matriks fitur X berukuran (4217, 12288) dan vektor label y (4217,). Dataset dibagi menjadi training set 3373 sampel (80%) dan testing set 844 sampel (20%) dengan stratified split. Label encoding: cataract=0, diabetic_retinopathy=1, glaucoma=2, normal=3.

3.2 Hyperparameter Tuning

RandomizedSearchCV dengan 20 iterasi dan 5-fold cross-validation menghasilkan 100 model fit. Dua visualisasi dibuat untuk menganalisis hasil: *bar chart ranking* top 10 kombinasi parameter dan histogram distribusi CV scores.



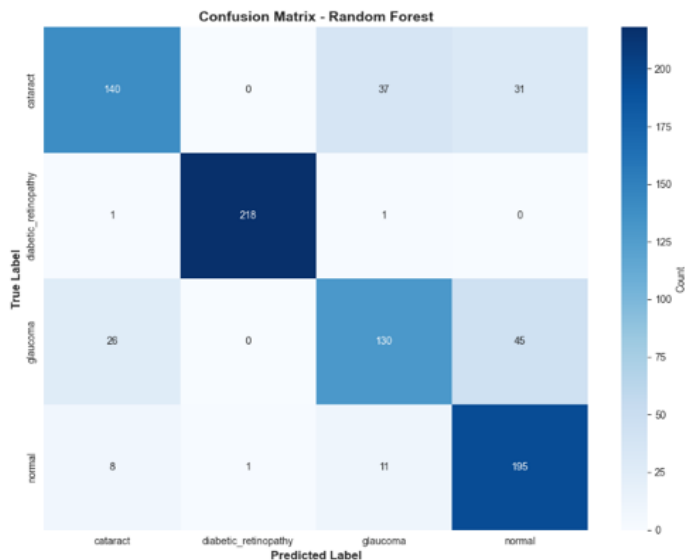
Gambar 5. Top 10 Parameter Combinations



Gambar 6. *Distribution of CV Scores*

3.3 Testing

Model terbaik dilakukan pada data test sejumlah 844 sampel kemudian dievaluasi menggunakan confusion matriks dan kurva ROC-AUC. Confusion matrix menampilkan distribusi prediksi dan pola kesalahan dalam bentuk heatmap.



Gambar 7. *Confusion Matrix Diagram*

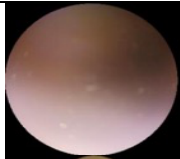
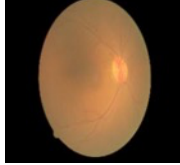
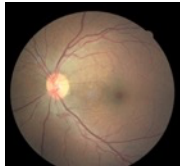

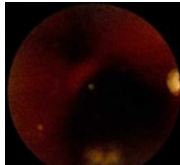

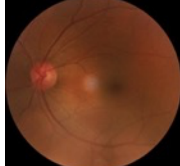

Cataract: 140 benar, confusion terbesar dengan glaucoma (28 kasus) karena kemiripan visual. Diabetic Retinopathy: 218 dari 220 benar, hampir sempurna karena ciri-cirinya sangat khas. Glaucoma: 130 benar, confusion tinggi dengan cataract (32) dan normal (23) karena perubahan yang halus. Normal: 195 dari 215 benar, recall tinggi mencegah false negative penyakit. Kemudian dari nilai *confusion matrix* dihitung nilai akurasi, presisi, sensitivity, specificity dan F1 seperti pada tabel 1 berikut.

Tabel 1. Hasil Evaluasi Model per Kelas

Class	Precision	Sensitivity	Specificity	F1
Cataract	0.8000	0.6730	0.9449	0.7310
Diabetic Retinopathy	0.9954	0.9909	0.9983	0.9931
Glaucoma	0.7262	0.6467	0.9237	0.6842
Normal	0.7195	0.9069	0.8791	0.8024
Accuracy			0.8092	
Macro-avg	0.8103	0.8044	0.9365	0.8027
Weighted-avg	0.8128	0.8092	0.9370	0.8064

Untuk mengevaluasi performa model secara kualitatif, dilakukan pengujian pada beberapa sampel data testing dari setiap kelas penyakit mata. Tabel 1 menunjukkan hasil prediksi model Random Forest terhadap 8 sampel gambar yang dipilih secara acak dari dataset testing, lengkap dengan label dan distribusi probabilitas untuk keempat kelas penyakit.

Tabel 2. Hasil Prediksi Model Berdasarkan Skor Probabilitas dan Label Sebenarnya per Kelas

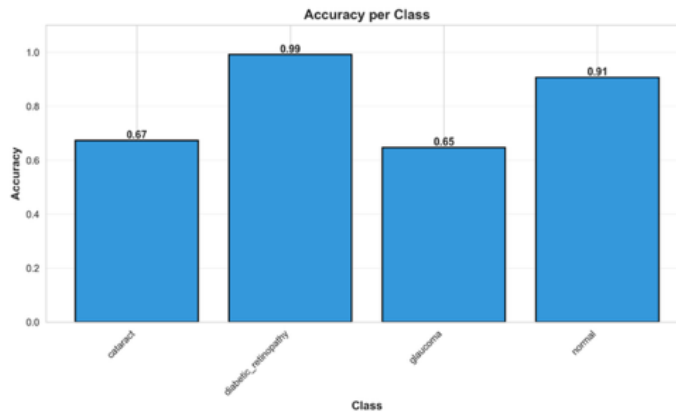
Gambar	Label	Probabilitas
	Cataract	Cataract: 43.00% Glaucoma: 29.00% Normal: 17.00% Diabetic Retinopathy: 11.00%
	Cataract	Cataract: 72.00% Glaucoma: 20.00% Diabetic Retinopathy: 4.00% Normal: 4.00%
	Diabetic retinopathy	Diabetic Retinopathy: 96.00% Normal: 2.00% Cataract: 1.00% Glaucoma: 1.00%
	Diabetic retinopathy	Diabetic Retinopathy: 42.00% Glaucoma: 33.00% Cataract: 19.00% Norma: 16.00%
	Glaucoma	Glaucoma: 47.00% Cataract: 29.00% Normal: 15.00% Diabetic Retinopathy: 9.00%
	Glaucoma	Glaucoma: 44.00% Normal: 27.00% Cataract: 22.00% Diabetic Retinopathy: 7.00%
	Normal	Normal: 38.00% Glaucoma: 28.00% Cataract: 27.00% Diabetic Retinopathy: 7.00%
	Normal	Normal: 31.00% Cataract: 29.00% Glaucoma: 26.00% Diabetic Retinopathy: 14.00%

Berdasarkan hasil prediksi pada Tabel 2, model Random Forest menunjukkan performa yang bervariasi pada setiap kelas. Model menunjukkan probabilitas tinggi untuk kelas Diabetic Retinopathy yang memiliki karakteristik visual distingtif. Namun, untuk kelas Glaucoma dan Normal, model menunjukkan probabilitas lebih rendah dengan distribusi yang lebih merata. Evaluasi empat metrik kunci pada Gambar 8 mengungkapkan disparitas performa notable antar kategori penyakit mata. *Diabetic Retinopathy* mendemonstrasikan superioritas dengan skor mencapai 0.99 pada seluruh metrik, sementara *Glaucoma* memperlihatkan tantangan substansial dengan *Sensitivity* 0.65. *Cataract* dan Normal menunjukkan karakteristik intermediate dengan *trade-off* berbeda, di mana *Cataract* unggul dalam Precision (0.80) namun Normal superior dalam Sensitivity (0.91).



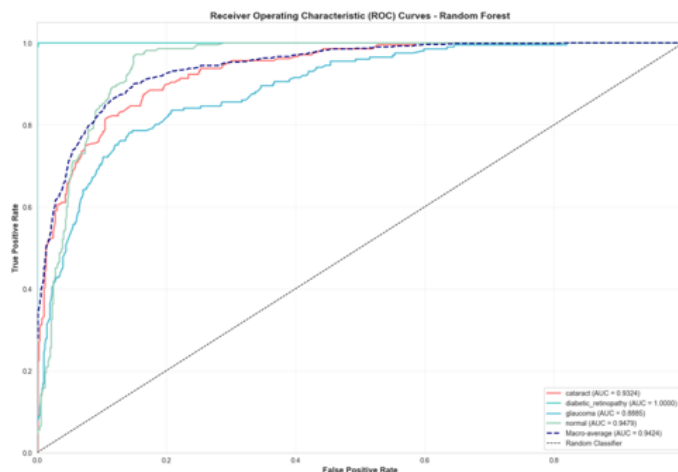
Gambar 8. Performance Metrics per Class

Visualisasi pada Gambar 9 menunjukkan heterogenitas kemampuan model dengan *accuracy* yang bervariasi: *Diabetic Retinopathy* (0.99), *Normal* (0.91), *Cataract* (0.67), dan *Glaucoma* (0.65). Disparitas ini mengindikasikan bahwa *Diabetic Retinopathy* memiliki *discriminative features* paling menonjol, sedangkan *Glaucoma* dan *Cataract* menghadapi ambiguitas lebih tinggi dalam *feature space*. Variabilitas performa mengsuggestikan kebutuhan untuk optimasi yang tertarget seperti *class-weighted loss functions* atau *specialized feature engineering*.



Gambar 9. Accuracy per Class

Specificity mengukur kemampuan model dalam mengidentifikasi sampel yang bukan anggota suatu kelas dengan benar (rumus $TN/(TN+FP)$). Hasil menunjukkan *specificity* yang tinggi: *cataract* (94,50%), *diabetic retinopathy* (99,84%), *glaucoma* (92,38%), normal (87,92%), dengan rata-rata makro 93,66%. *Diabetic retinopathy* hampir sempurna, yang berarti hampir tidak ada *false positive*. *Specificity* terendah pada kelas normal (87,92%) terjadi karena beberapa kasus penyakit diprediksi sebagai normal, namun ini merupakan *trade-off* yang dapat diterima dalam skrining medis di mana *sensitivity* lebih krusial. Nilai *specificity* yang tinggi mengindikasikan bahwa model mempunyai daya pembeda yang baik sebagai alat skrining. ROC-AUC mengukur kemampuan diskriminatif model pada semua ambang batas dengan membuat plot *True Positive Rate* terhadap *False Positive Rate*. Hasil per kelas: *cataract* (0,9324), *diabetic retinopathy* (1,0000), *glaucoma* (0,8885), normal (0,9479), dengan rata-rata makro 0,9422 dan rata-rata tertimbang 0,9435.



Gambar 10. ROC Curves



Pada Gambar 10 menunjukkan kurva yang sempurna (AUC 1,0) yang langsung naik vertikal ke TPR=1,0, mengonfirmasi pemisahan sempurna dari kelas lain. *Glaucoma* memiliki AUC terendah (88,85%) namun masih di atas garis diagonal, sesuai dengan hasilnya sebagai kelas yang paling menantang. Kurva rata-rata makro (garis putus-putus navy) menunjukkan performa agregat yang sangat baik. Skor ROC-AUC yang tinggi (>0,88) mengindikasikan skor probabilitas terkalibrasi dengan baik dan bermanfaat untuk memprioritaskan kasus berdasarkan tingkat keyakinan dalam aplikasi klinis.

3.4 Performa Hyperparameter Tuning

RandomizedSearchCV efektif menemukan konfigurasi optimal (skor CV 80,43%, akurasi uji 80,92%), menunjukkan generalisasi yang sangat baik. Iterasi sebanyak 20 kali memberikan keseimbangan antara eksplorasi dan efisiensi, dengan distribusi skor CV yang terkonsentrasi menunjukkan *Random Forest* tangguh terhadap variasi *hyperparameter*. Kombinasi $max_features=log_2'$ (13 fitur) dengan $bootstrap=False$ mencapai keragaman murni melalui randomisasi fitur yang agresif, memanfaatkan ruang fitur berdimensi tinggi (12.288 fitur).

3.5 Analisis Performa per Kelas

Diabetic Retinopathy memiliki performa yang baik (*precision* 99,54%, *recall* 99,09%, AUC 1,0) dijelaskan oleh manifestasi visual yang khas: perdarahan (bintik merah), eksudat keras (endapan kuning-putih), mikroaneurisma (titik merah kecil). Pola warna yang khas dan lesi berkontras tinggi memberikan sinyal diskriminatif yang kuat yang mudah ditangkap oleh *Random Forest*, sehingga menciptakan batas keputusan yang jelas. *Cataract* dan *Glaucoma* memiliki hasil performa yang moderat (*cataract* F1 73,11%, *glaucoma* F1 68,42%) mencerminkan tantangan yang melekat pada kedua kelas ini. *Cataract* tahap awal menunjukkan *opacity* yang halus dengan spektrum keparahan yang luas.

Glaucoma menantang karena *optic disc cupping* yang halus, variabilitas anatomi antar individu, dan *fundus image* saja memiliki informasi yang terbatas (memerlukan multimodal tests). Kebingungan antara keduanya (28 *cataract*→*glaucoma*, 32 *glaucoma*→*cataract*) terjadi karena keduanya dapat *co-exist* dan *opacity* dapat mengaburkan tampilan *optic disc*. Sedangkan kelas Normal memiliki Recall yang tinggi (90,70%) namun precision yang moderat (71,96%) menunjukkan model bersifat konservatif dalam memprediksi penyakit, lebih memilih alarm palsu daripada melewatkan penyakit hal ini diinginkan untuk skrining medis. False positives (20 kasus) kemungkinan merupakan penyakit tahap awal atau kasus borderline.

3.6 Perbandingan Hasil dengan Penelitian Sebelumnya

Berdasarkan tabel perbandingan yang terdapat pada Tabel 3, metode yang diusulkan dalam penelitian ini, yaitu Random Forest dengan optimasi hyperparameter menggunakan *RandomizedSearchCV* (RF-*RandomSearchCV*), mampu menghasilkan performa yang jauh lebih baik dibandingkan penelitian sebelumnya oleh Ayyu yang menggunakan metode Sobel Segmentation dan Linear SVC. Penelitian sebelumnya hanya memperoleh akurasi sebesar 44,3%, sedangkan metode yang diusulkan pada penelitian ini mencapai akurasi 80,92%. Hasil tersebut menunjukkan bahwa penerapan algoritma Random Forest dengan proses optimasi hyperparameter lebih efektif dalam melakukan klasifikasi penyakit mata dibandingkan pendekatan berbasis segmentasi Sobel dan Linear SVC. Peningkatan akurasi yang signifikan ini juga mengindikasikan bahwa metode yang diusulkan memiliki kemampuan generalisasi dan klasifikasi yang lebih baik terhadap data citra penyakit mata.

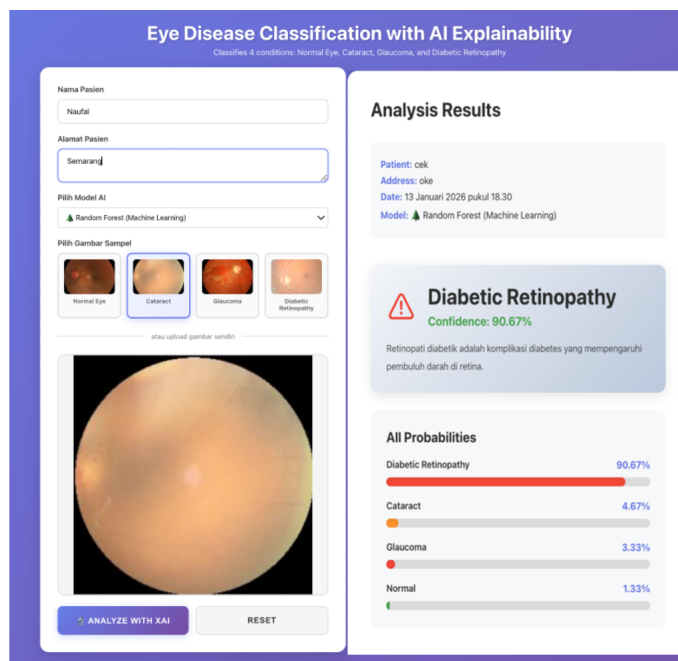
Tabel 3. Perbandingan dengan Hasil Penelitian Sebelumnya

Penelitian	Metode	Akurasi
Ayyu, 2024 [19]	Sobel Segmentation + Linear SVC	44,3%
Penelitian kami	RF- <i>RandomSearchCV</i>	80,92%

4. IMPLEMENTASI

Implementasi sistem pada penelitian ini dilakukan dalam bentuk aplikasi berbasis web yang bertujuan untuk mempermudah proses klasifikasi penyakit mata secara interaktif dan real-time. Aplikasi dikembangkan menggunakan platform web dan di-deploy melalui Netlify sehingga dapat diakses secara online tanpa memerlukan instalasi tambahan pada perangkat pengguna. Sistem menggunakan model machine learning Random Forest yang telah dioptimasi menggunakan *RandomizedSearchCV* untuk melakukan klasifikasi empat kondisi mata, yaitu normal, katarak, glaukoma, dan retinopati diabetik.

Pada antarmuka aplikasi, pengguna dapat memasukkan identitas pasien, memilih model kecerdasan buatan yang digunakan, serta mengunggah citra fundus mata atau memilih contoh gambar yang telah disediakan oleh sistem. Implementasi ini dirancang agar mudah digunakan oleh pengguna umum maupun tenaga medis sebagai sistem pendukung diagnosis awal penyakit mata, dimana contohnya dapat dilihat pada Gambar 11 berikut.



Gambar 11. Sample Prototype Sistem

Pada prototype yang dilampirkan pada Gambar 11 pengguna melakukan proses input data dan memilih citra fundus mata, sistem akan menjalankan proses analisis menggunakan model Random Forest untuk menghasilkan prediksi klasifikasi penyakit mata. Output sistem ditampilkan secara otomatis pada halaman web dalam bentuk hasil prediksi kategori penyakit beserta tingkat keyakinan model terhadap hasil klasifikasi tersebut.

5. KESIMPULAN

Penelitian ini berhasil mengembangkan sistem klasifikasi penyakit mata (*cataract*, *diabetic retinopathy*, *glaucoma*, normal) menggunakan *Random Forest* dengan *RandomizedSearchCV*. *Preprocessing* meliputi resize ke 64x64 piksel, normalisasi 0-1, *flattening* menjadi 12.288 fitur, dengan pembagian *stratified* 80:20 (3.373 training, 844 testing). *Hyperparameter tuning* (20 iterasi, 5-fold CV) menghasilkan konfigurasi optimal: *n_estimators*=150, *max_depth*=30, *max_features*=log2, *bootstrap*=False, dengan skor CV 80,43% dan akurasi uji 80,92%. Performa bervariasi per kelas: *diabetic retinopathy* luar biasa (*precision* 99,54%, *recall* 99,09%, ROC-AUC 1,0), *glaucoma* terendah (F1 68,42%, ROC-AUC 88,85%), *cataract* dan normal moderat (F1 73,11% dan 80,25%). Metrik agregat: *precision* 81,03%, *recall* 80,44%, *F1-score* 80,27%, *specificity* 93,66%, ROC-AUC 94,35%. Penelitian ini, menunjukkan model memiliki potensi kuat untuk skrining *diabetic retinopathy*, dan sebagai alat bantu untuk *cataract* dan *glaucoma*. Keterbatasan: resolusi rendah, fitur piksel mentah, dataset tunggal, tanpa metadata klinis. Penelitian mendatang disarankan dengan menggunakan resolusi lebih tinggi, feature engineering/transfer learning, validasi eksternal, pendekatan multimodal.

REFERENCES

- [1] W. He, "Medical Data Large-scale Processing Technology Based on March Algorithm," *Procedia Computer Science*, vol. 259, pp. 572–580, 2025, doi: 10.1016/j.procs.2025.04.006.
- [2] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J Big Data*, vol. 8, no. 1, p. 53, Mar. 2021, doi: 10.1186/s40537-021-00444-8.
- [3] C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning," *Electron Markets*, vol. 31, no. 3, pp. 685–695, Sep. 2021, doi: 10.1007/s12525-021-00475-2.
- [4] P. Probst, M. N. Wright, and A. Boulesteix, "Hyperparameters and tuning strategies for random forest," *WIREs Data Min & Knowl*, vol. 9, no. 3, p. e1301, May 2019, doi: 10.1002/widm.1301.
- [5] M. A. Ganaie, M. Hu, A. K. Malik, M. Tanveer, and P. N. Suganthan, "Ensemble deep learning: A review," *Engineering Applications of Artificial Intelligence*, vol. 115, p. 105151, Oct. 2022, doi: 10.1016/j.engappai.2022.105151.
- [6] L. Yang *et al.*, "Study of cardiovascular disease prediction model based on random forest in eastern China," *Sci Rep*, vol. 10, no. 1, p. 5245, Mar. 2020, doi: 10.1038/s41598-020-62133-5.
- [7] L. Novita, W. Fuadi, and K. Kurniawati, "Cataract Eye Disease Diagnosis Using the Random Forest Method," *Int. J. Eng. Scie. and Inform. Technology*, vol. 5, no. 2, pp. 33–41, Jan. 2025, doi: 10.52088/ijesty.v5i2.777.
- [8] M. Agrawal, N. Mohan, and V. Jain, "Chronic Kidney Disease Prediction Using Random Forest, Decision Tree and Ada Boost Classifier," in *2023 4th International Conference on Smart Electronics and Communication (ICOSEC)*, Trichy, India: IEEE, Sep. 2023, pp. 1589–1593. doi: 10.1109/ICOSEC58147.2023.10276324.



- [9] G. Pallavi and K. Vidhya, "Brain tumor detection with high accuracy using random forest and comparing with thresholding method," presented at the Fifth International Conference On Applied Sciences: ICAS2023, Baghdad, Iraq, 2024, p. 020079. doi: 10.1063/5.0198189.
- [10] A. Meiliana, N. M. Dewi, and A. Wijaya, "Artificial Intelligent in Healthcare," *Indones Biomed J*, vol. 11, no. 2, pp. 125–35, Aug. 2019, doi: 10.18585/inabj.v11i2.844.
- [11] M. Naufal, H. Al Azies, and R. M. Brilianto, "Enhanced Brain Tumor Classification through Gamma Correction in Deep Learning," *SISTEMASI*, vol. 13, no. 6, p. 2348, Nov. 2024, doi: 10.32520/stmsi.v13i6.4474.
- [12] G. V. Doddi, "Eye diseases classification dataset." 2021. [Online]. Available: <https://www.kaggle.com/datasets/gunavenkatdoddi/eye-diseases-classification>
- [13] F. Alzami, S. Winarno, M. Naufal, and H. Al Azies, "Enhancing Driver Drowsiness Detection through GMM-Optimized CLAHE," in *2024 International Seminar on Application for Technology of Information and Communication (iSemantic)*, Semarang, Indonesia: IEEE, Sep. 2024, pp. 212–217. doi: 10.1109/iSemantic63362.2024.10762529.
- [14] R. Septia, Junadhi, Susi Erlinda, and Wirta Agustin, "Heart Failure Disease Classification Using Random Forest Algorithm with Grid Search Cross Validation Technique," *ijcs*, vol. 14, no. 2, Apr. 2025, doi: 10.33022/ijcs.v14i2.4765.
- [15] R. Anitha and D. Siva Sundhara Raja, "Development of computer-aided approach for brain tumor detection using random forest classifier," *Int J Imaging Syst Tech*, vol. 28, no. 1, pp. 48–53, Mar. 2018, doi: 10.1002/ima.22255.
- [16] J. Gao, J. Ren, and Z. Wen, "Research on Diabetes Prediction Based on the Randomized Search CV Method," in *2025 2nd International Conference on Electronic Engineering and Information Systems (EEISS)*, Nanjing, China: IEEE, May 2025, pp. 1–4. doi: 10.1109/EEISS65394.2025.11086023.
- [17] F. Alzami, M. Naufal, H. A. Azies, S. Winarno, and M. A. Soeleman, "Time Distributed MobileNetV2 with Auto-CLAHE for Eye Region Drowsiness Detection in Low Light Conditions," *ijacsa*, vol. 15, no. 11, 2024, doi: 10.14569/IJACSA.2024.0151146.
- [18] Ş. K. Çorbacıoğlu and G. Aksel, "Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value," *Turkish Journal of Emergency Medicine*, vol. 23, no. 4, pp. 195–198, Oct. 2023, doi: 10.4103/tjem.tjem_182_23.
- [19] F. T. A. Sayyidul Laily, "Feature Extraction and Classification of Retinal Images Using Sobel Segmentation and Linear SVC," *ijaimi*, vol. 2, no. 2, pp. 136–149, Nov. 2024, doi: 10.56705/ijaimi.v2i2.153.