PENERAPAN ALGORITMA RANDOM FOREST DAN SUPPORT VECTOR MACHINE UNTUK KLASIFIKASI DATA MEDIS DALAM DIAGNOSA GLAUKOMA

Nurfitri N. Lasida[1], Sigit Soijoyo[2], Adelina Ibrahim[3]
nurulfitry391@gmail.com[1], sigit_soijoyo@yahoo.com[2], adelinaibrahim07@gmail.com[3]
Universitas Muhammadiyah Maluku Utara

ABSTRAK

Glaukoma merupakan salah satu penyebab kebutaan permanen yang memerlukan deteksi dini untuk mencegah kerusakan saraf optik. Penelitian ini bertujuan untuk membandingkan algoritma Random Forest dan Support Vector Machine (SVM) dalam klasifikasi glaukoma menggunakan data tabular pasien. Dataset yang digunakan terdiri dari 477 data rumah sakit serta dataset pembanding eksternal untuk menguji kemampuan generalisasi model. Atribut yang digunakan meliputi usia, jenis kelamin, visus mata kanan, visus mata kiri, serta fitur turunan berupa selisih dan rata-rata visus. Hasil pengujian pada dataset rumah sakit menunjukkan bahwa algoritma Support Vector Machine memiliki performa terbaik dengan akurasi sebesar 98,96%, precision 98,44%, recall 100%, dan F1-score 99,21%, sedangkan Random Forest memperoleh akurasi sebesar 97,92%. Namun, pada pengujian menggunakan dataset pembanding, performa kedua algoritma mengalami penurunan. SVM memperoleh akurasi sebesar 72,31%, sedangkan Random Forest sebesar 68,46%. Penurunan performa ini menunjukkan adanya perbedaan distribusi data dan keterbatasan atribut klinis yang memengaruhi kemampuan generalisasi model. Penelitian ini menyimpulkan bahwa SVM lebih optimal pada dataset internal, namun diperlukan pengembangan lebih lanjut untuk meningkatkan kemampuan generalisasi model sebelum diterapkan secara luas.
Kata Kunci: Glaukoma, Random Forest, Support Vector Machine, Klasifikasi.

ABSTRACT

Glaucoma is one of the leading causes of permanent blindness and requires early detection to prevent optic nerve damage. This study aims to compare the Random Forest and Support Vector Machine (SVM) algorithms in glaucoma classification using patients' tabular data. The dataset consists of 477 hospital records and an external comparative dataset to evaluate the model's generalization capability. The attributes used include age, gender, right eye visual acuity, left eye visual acuity, and derived features such as visual acuity difference and average visual acuity. The results on the hospital dataset show that the Support Vector Machine achieved the best performance with an accuracy of 98.96%, precision of 98.44%, recall of 100%, and F1-score of 99.21%, while Random Forest achieved an accuracy of 97.92%. However, performance decreased when tested on the external dataset. SVM achieved an accuracy of 72.31%, whereas Random Forest achieved 68.46%. This decline indicates differences in data distribution and limitations of clinical attributes affecting the model's generalization capability. This study concludes that SVM is more optimal for the internal dataset, however, further development is required to improve generalization performance before wider clinical implementation.
Keywords: *Glaucoma, Random Forest, Support Vector Machine, Classification.*

PENDAHULUAN

Perkembangan teknologi informasi dan ilmu komputer pada era modern telah memberikan dampak yang signifikan terhadap kemajuan di bidang kesehatan, terutama dalam aspek pengolahan data medis. Pemanfaatan teknologi ini berperan penting dalam meningkatkan efektivitas proses diagnosis dan pengambilan keputusan medis secara lebih cepat, akurat, serta berbasis data. Saat ini, penerapan sistem cerdas berbasis data menjadi kebutuhan utama dalam upaya mendukung tenaga medis untuk memberikan pelayanan

kesehatan yang efisien dan berkualitas. Salah satu penerapan pentingnya dapat ditemukan pada proses diagnosis penyakit kronis dan kompleks, seperti glaukoma.

Glaukoma merupakan penyakit mata yang menyerang saraf optik, umumnya disebabkan oleh peningkatan tekanan intraokular. Penyakit ini menjadi salah satu penyebab kebutaan terbesar di dunia, menempati posisi kedua setelah katarak menurut data World Health Organization (2023). Karakteristik glaukoma yang tidak menunjukkan gejala pada tahap awal menjadikan deteksi dini sebagai langkah kunci untuk mencegah kerusakan penglihatan permanen (Putri et al., 2022). Oleh karena itu, diperlukan pengembangan metode diagnosis yang efisien dan akurat guna menekan angka kebutaan akibat penyakit ini.

Secara konvensional, diagnosis glaukoma dilakukan melalui observasi manual oleh dokter spesialis dengan menilai citra fundus retina dan mengukur tekanan bola mata. Walaupun metode ini telah digunakan secara luas, proses tersebut masih menghadapi beberapa kendala, seperti keterbatasan waktu pemeriksaan, faktor subjektivitas dokter, serta potensi inkonsistensi hasil diagnosis. Permasalahan tersebut menunjukkan perlunya teknologi pendukung diagnosis berbasis komputer yang mampu melakukan klasifikasi kondisi pasien secara otomatis, objektif, dan konsisten. Kehadiran data medis dalam format digital saat ini membuka peluang besar bagi penerapan pembelajaran mesin (machine learning) sebagai sarana bantu dalam proses diagnosis medis.

Dalam konteks pembelajaran mesin, algoritma Random Forest (RF) dan Support Vector Machine (SVM) menjadi dua metode yang banyak digunakan untuk klasifikasi data medis. SVM bekerja dengan menentukan hyperplane optimal yang dapat memisahkan kelas data secara efektif, sehingga unggul dalam menangani data berdimensi tinggi dan kompleks (Pertiwi & Utomo, 2023). Sementara itu, Random Forest merupakan metode ensemble learning yang mengombinasikan sejumlah decision tree untuk menghasilkan keputusan yang lebih stabil dan akurat, sekaligus meminimalkan risiko overfitting pada model tunggal (Rahman & Susilo, 2023). Keunggulan dari kedua algoritma ini menjadikannya sangat potensial untuk digunakan dalam proses klasifikasi data medis dengan karakteristik yang bervariasi.

Berbagai penelitian sebelumnya telah menunjukkan efektivitas kedua algoritma tersebut dalam klasifikasi penyakit, , termasuk glaukoma. Pertiwi dan Utomo (2023). menunjukkan bahwa SVM mampu mencapai tingkat akurasi yang tinggi dalam mendeteksi glaukoma berdasarkan analisis citra fundus retina. Sementara itu, Taufik (2023) melaporkan bahwa algoritma Random Forest memberikan hasil prediksi yang stabil dan akurat pada data medis yang kompleks. Meskipun demikian, penelitian yang secara langsung membandingkan performa Random Forest dan SVM dalam mendiagnosis glaukoma masih tergolong terbatas. Kondisi ini membuka peluang penelitian lebih lanjut untuk mengevaluasi efektivitas relatif kedua algoritma tersebut dalam konteks diagnosis medis berbasis data digital.

Penelitian ini dilakukan untuk mengidentifikasi secara empiris keunggulan dan keterbatasan masing-masing algoritma dalam mendiagnosis glaukoma melalui pendekatan kuantitatif dengan metode eksperimen komparatif. Dataset yang digunakan mencakup data medis dari sumber terbuka seperti Kaggle dan data dari rumah sakit, dengan memperhatikan etika penelitian. Kinerja model akan dievaluasi menggunakan metrik pengukuran seperti akurasi, presisi, recall, dan F1-score guna memperoleh gambaran komprehensif mengenai efektivitas kedua algoritma tersebut.

Hasil penelitian ini diharapkan dapat memberikan kontribusi nyata dalam pengembangan sistem diagnosis medis otomatis yang dapat membantu tenaga medis dalam pengambilan keputusan klinis. Dengan mengetahui algoritma yang lebih unggul dalam

klasifikasi data glaukoma, proses diagnosis dapat dilakukan secara lebih cepat, efisien, dan akurat. Penerapan sistem semacam ini juga berpotensi mengurangi beban kerja tenaga medis serta meningkatkan kualitas layanan kesehatan secara keseluruhan. Lebih jauh, penelitian ini memiliki nilai praktis yang tinggi tidak hanya bagi tenaga medis, tetapi juga bagi pasien yang membutuhkan diagnosis dini agar dapat memperoleh penanganan tepat waktu. Integrasi teknologi machine learning dalam analisis data medis diharapkan mampu mempercepat deteksi glaukoma dan mencegah komplikasi serius, sekaligus mempertegas peran penting inovasi teknologi dalam peningkatan kualitas pelayanan kesehatan modern.

Berdasarkan uraian yang telah dipaparkan di atas, maka penulis berinisiatif membuat sistem "Penerapan Algoritma Random Forest dan Support Vector Machine untuk Klasifikasi Data Medis dalam Diagnosa Glaukoma" sebagai upaya untuk merancang sistem klasifikasi yang efektif, akurat, dan bermanfaat dalam mendukung proses diagnosis penyakit glaukoma secara digital.

METODOLOGI

Penelitian ini dilaksanakan dengan menggunakan metode eksperimen kuantitatif melalui pendekatan komparatif yang bertujuan untuk menilai dan membandingkan kinerja dua algoritma pembelajaran mesin, yaitu Random Forest dan Support Vector Machine, dalam melakukan klasifikasi data medis untuk diagnosa glaukoma.

Pendekatan komparatif merupakan teknik penelitian yang melibatkan perbandingan antara dua atau lebih perlakuan atau kondisi guna mengidentifikasi perbedaan dan kelebihan masing-masing variabel independen terhadap variabel dependen. Dalam penelitian ini, variabel independen adalah jenis algoritma yang digunakan yaitu Random Forest dan Support Vector Machine, sementara variabel dependen adalah hasil klasifikasi berupa diagnosis glaukoma positif atau negatif.

tahap awal melibatkan pengumpulan dataset, yang diikuti oleh normalisasi data, validasi data, dan tahap klasifikasi. Dua algoritma klasifikasi yang digunakan dalam penelitian ini adalah Random Forest dan Support Vector Machine. Hasil dari tahap klasifikasi adalah sebuah model klasifikasi, dan evaluasi dilakukan dengan menggunakan akurasi berdasarkan confusion matri

HASIL DAN PEMBAHASAN

Bab ini menyajikan tentang hasil implementasi dari sistem yang dirancang, data-data hasil percobaan dan analisa. Hasil dari penelitian ini terdiri dari beberapa tahap untuk melakukan identifikasi glaukoma dengan menggunakan metode-metode yang telah diusulkan pada bab sebelumnya.

A. Alur Kerja Sistem

Alur kerja sistem dimulai ketika pengguna menjalankan aplikasi melalui browser. Setelah aplikasi terbuka, sistem akan menampilkan halaman dashboard, sebagai tampilan awal.

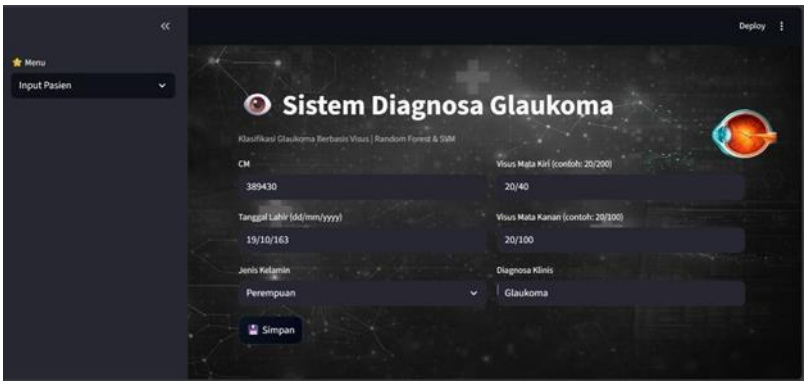
Pengguna dapat mengakses fitur-fitur utama melalui menu navigasi yang tersedia, antara lain:

1. Dashboard



Gambar 1 Tampilan Dashboard

2. Input Data Pasien



Gambar 2 Input Data Pasien

Data pasien yang dimasukkan akn diproses dan disimpan dalam basis data, selanjutnya, data tersebut digunakan sebagai input dalam proses pelatihan model klasifikasi. Model yang telah dilatih kemudian digunakan untuk melakukan prediksi terhadap data pasien baru.

B. Pre Processing

Tahap preprocessing merupakan langkah penting sebelum proses pelatihan model dilakukan. Pada penelitian ini, preprocessing diterapkan pada data tabular hasil ekstraksi fitur, bukan pada data citra secara langsung. Dalam citra fundus telah diolah sebulumnya sehingga direpresentasikan dalam bentuk nilai numerik.

Tahapan preprocessing yang dilakukan diawali dengan seleksi fitur yang relevan, yaitu usia jenis kelamin, visus mata kanan, visus mata kiri, serta fitur turunan berupa selisih visus dan rata-rata visus. Fitur-fitur tersebut dipilih karena memiliki keterkaitan dengan kondisi penglihatan pasien yang berhubungan dengan glaukoma.

Selanjutnya dilakukan proses feature engineering dengan membentuk fitur selisih visus dan rata-rata visus. Fitur selisih visus digunakan untuk menggambarkan perbedaan tingkat ketajaman penglihatan antara mata kanan dan mata kiri, sedangkan fitur rata-rata visus digunakan untuk mempresentasikan kondisi penglihatan pasien secara keseluruhan.

Data kategorikal berupa jenis kelamin dikonversi kedalam bentuk numerik melalui proses encoding agar dapat diproses oleh algoritma, selain itu, dilakukan standarisasi data menggunakan StandardScaler pada algoritma Support Vector Machine. Standarisasi ini bertujuan untuk menyamakan skala antar fitur sehingga proses pembelajar model dapat berjalan lebih optimal. Pada algoritma Random Forest, proses standarisasi tidak dilakukan karena algoritma tersebut tidak sensitif terhadap perbedaan skala data.

Pembagian data dilakukan menggunakan metode stratified train-test split dengan proporsi 80% data latih dan 20% data uji. Metode ini digunakan untuk menjaga

keseimbangan proporsi kelas glaukoma dan normal pada data latih maupun data uji. Selain itu, dilakukan validasi silang menggunakan Stratified K-Fold dengan jumlah lipatan sebanyak lima untuk mengukur kestabilan dan konsistensi performa model.

Berdasarkan tahapan preprocessing yang telah dilakukan, data yang digunakan telah siap untuk diproses oleh algoritma Random Forest dan Support Vector Machine sehingga dapat menghasilkan model klasifikasi dengan performa yang optimal.

C. Hasil Pelatihan Model pada Dataset Rumah Sakit

Pengujian pertama dilakukan menggunakan dataset rumah sakit yang berjumlah 477. Proses pelatihan model dilakukan menggunakan dua algoritma klasifikasi, yaitu Random Forest dan Support Vector Machine.

1. Hasil Random Forest

	Akurasi	Precision	Recall	F1-Score
Random Forest	97,92%	100,00%	96,83%	98,39%

a) Confusion Matriks dan Evaluasi

Evaluasi performa model klasifikasi glaukoma dilakukan menggunakan confusion matrix yang bertujuan untuk mengetahui kemampuan model dalam mengklasifikasikan data pasien ke dalam kelas Normal dan Glaukoma secara benar.

Berdasarkan hasil pengujian model, diperoleh confusion matrix sebagai berikut:

Kelas Aktual	Prediksi Normal	Prediksi Glaukoma
Normal	33	0
Glaukoma	2	61

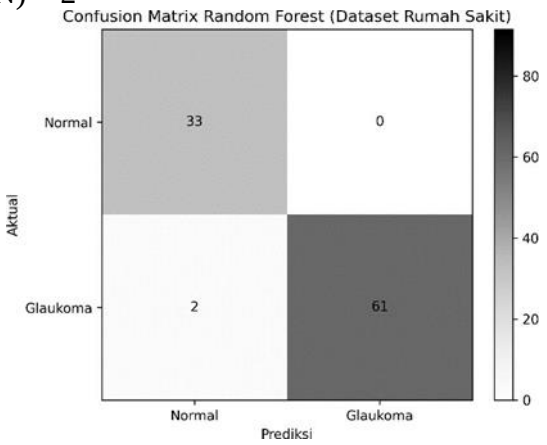
Berdasarkan tabel confusion matriks tersebut, diperoleh nilai sebagai berikut:

True Positive (TP) = 61

True Negative (TN) = 33

False Positive (FP) = 0

False Negative (FN) = 2



Gambar 3 Confusion Matriks Random Forest pada Dataset Rumah Sakit

b) Perhitungan Metrik Evaluasi

Accuracy = $\frac{TP+TN}{TP+TN+FP+FN} = \frac{61+33}{61+33+0+2}$ = 0,9792 = 97,92%

Precision = $\frac{TP}{TP+FP} = \frac{61}{61+0}$ = 1 = 100%

Recall $= \frac{TP}{TP+FN} = \frac{61}{61+2} =$ 0,9683 =96,83%

F1= $2x \frac{\text{Precision x Recall}}{\text{Precision + Recall}} = 2x \frac{1{,}00 \text{ x } 0{,}9683}{1{,}00 + 0{,}9683}$ = 0,9839 =98,39%

2. Support Vector Machine

Berdasarkan hasil pelatihan menggunakan algoritma Support Vector Machine diperoleh performa sebagai berikut:

	Akurasi	Precision	Recall	F1-Score
Support Vector Machine	98,96%	98,44%	100,00%	99,21%

a) Confusion Matriks dan Evaluasi

Evaluasi performa model klasifikasi glaukoma dilakukan menggunakan confusion matrix yang bertujuan untuk mengetahui kemampuan model dalam mengklasifikasikan data pasien ke dalam kelas Normal dan Glaukoma secara benar.

Berdasarkan hasil pengujian model, diperoleh confusion matrix sebagai berikut:

Kelas Aktual	Prediksi Normal	Prediksi Glaukoma
Normal	32	1
Glaukoma	0	63

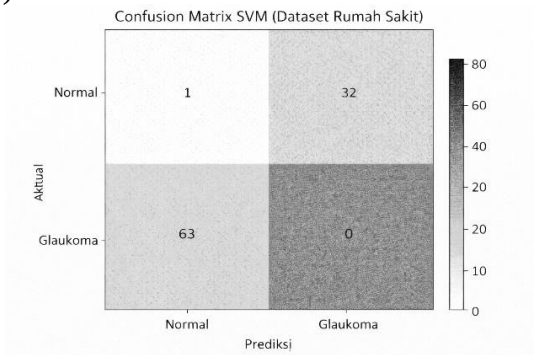
Berdasarkan tabel confusion matriks tersebut, diperoleh nilai sebagai berikut:

True Positive (TP) = 63

True Negative (TN) = 32

False Positive (FP) = 1

False Negative (FN) = 0



Gambar 4 Confusion Matriks SVM pada Dataset Rumah Sakit

b) Perhitungan Metrik Evaluasi

Accuracy $= \frac{TP+TN}{TP+TN+FP+FN} = \frac{63+32}{63+32+0+1} = 0{,}9896 = 98{,}96\%$

Precision $= \frac{TP}{TP+FP} = \frac{63}{63+1} = 0{,}9844 = 98{,}44\ \%$

Recall $= \frac{TP}{TP+FN} = \frac{63}{63+0} = 1 = 100\%$

F1$= 2x \frac{\text{Precision x Recall}}{\text{Precision + Recall}} = 2x \frac{0{,}9844\ x\ 1{,}00}{0{,}9844\ +\ 1{,}00} = 0{,}9921 = 99{,}21\%$

D. Hasil Pengujian Model pada Dataset Pendamping

Pengujian selanjutnya dilakukan menggunakan dataset pembanding **(**eksternal) yang berasal dari sumber berbeda dengan dataset rumah sakit. Dataset ini digunakan untuk menguji kemampuan generalisasi model dalam mengklasifikasikan glaukoma pada data di luar lingkungan pelatihan.

Dataset eksternal yang digunakan memiliki keterbatasan atribut klinis, sehingga dilakukan penyesuaian struktur fitur agar sesuai dengan model yang telah dilatih. Pengujian dilakukan tanpa pelatihan ulang model.

1. Hasil Random Forest

Berdasarkan hasil pelatihan menggunakan algoritma Random Forest, diperoleh performa sebagai berikut:

	Akurasi	Precision	Recall	F1-Score
Random Forest	68,46%	41,03%	47,06%	43,84%

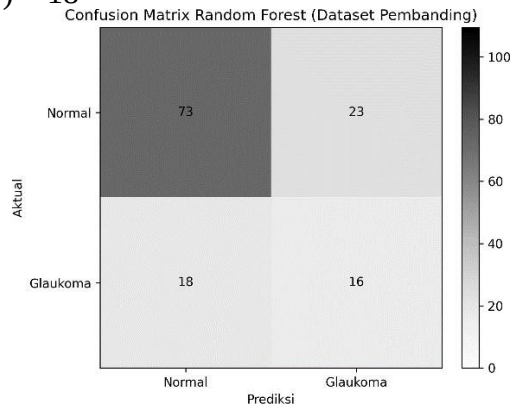
a) Confusion Matriks dan Evaluasi

Evaluasi performa model klasifikasi glaukoma dilakukan menggunakan confusion matrix yang bertujuan untuk mengetahui kemampuan model dalam mengklasifikasikan data pasien ke dalam kelas Normal dan Glaukoma secara benar.

Berdasarkan hasil pengujian model, diperoleh confusion matrix sebagai berikut:

Kelas Aktual	Prediksi Normal	Prediksi Glaukoma
Normal	73	23
Glaukoma	18	16

Berdasarkan tabel confusion matriks tersebut, diperoleh nilai sebagai berikut:
True Positive (TP) = 16
True Negative (TN) = 73
False Positive (FP) = 23
False Negative (FN) = 18



Gambar 5 Confusion Matriks Random Forest pada Dataset Pembanding

b) Perhitungan Metrik Evaluasi

Accuracy $= \frac{TP+TN}{TP+TN+FP+FN} = \frac{16+73}{16+73+23+18}$ = 0,6846 = 68,46%

Precision $= \frac{TP}{TP+FP} = \frac{16}{16+23}$ = 0,4103= 41,03%

Recall $= \frac{TP}{TP+FN} = \frac{16}{16+18}$ = 0,4706 = 47,06%

F1$= 2x \frac{\text{Precision x Recall}}{\text{Precision + Recall}} = 2x \frac{0,4103 \text{ x } 0,4706}{0,4103 + 0,4706}$ = 0,4384 =43,84%

2. Support Vector Machine

Berdasarkan hasil pelatihan menggunakan algoritma Support Vector Machine diperoleh performa sebagai berikut:

	Akurasi	Precision	Recall	F1-Score
Support Vector Machine	72,31%	48,00%	70,59%	57,14%

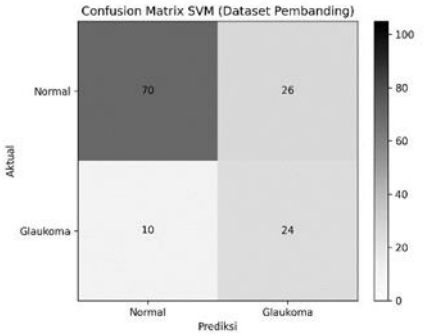
a) Confusion Matriks dan Evaluasi

Evaluasi performa model klasifikasi glaukoma dilakukan menggunakan confusion matrix yang bertujuan untuk mengetahui kemampuan model dalam mengklasifikasikan data pasien ke dalam kelas Normal dan Glaukoma secara benar.

Berdasarkan hasil pengujian model, diperoleh confusion matrix sebagai berikut

Kelas Aktual	Prediksi Normal	Prediksi Glaukoma
Normal	70	26
Glaukoma	10	24

Berdasarkan tabel confusion matriks tersebut, diperoleh nilai sebagai berikut:
True Positive (TP) = 24
True Negative (TN) = 70
False Positive (FP) = 26
False Negative (FN) = 10



Gambar 6 Confusion Matriks SVM pada Dataset Pembanding

b) Perhitungan Metrik Evaluasi

Accuracy $= \frac{TP+TN}{TP+TN+FP+FN} = \frac{24+70}{24+70+26+10} = 0{,}7231 = 72{,}31\%$

Precision $= \frac{TP}{TP+FP} = \frac{24}{24+26} = 0{,}48 = 48{,}00\ \%$

Recall $= \frac{TP}{TP+FN} = \frac{24}{24+10} = 0{,}7059 = 70{,}59\%$

F1$= 2x \frac{\text{Precision x Recall}}{\text{Precision + Recall}} = 2x \frac{0{,}48\ x\ 0{,}7059}{0{,}48\ +\ 0{,}7059} = 0{,}5714 = 57{,}14\%$

E. Perbandingan Performa Random Forest dan Support Vector Machine

Berdasarkan hasil pengujian pada dataset rumah sakit, algoritma Random Forest dan Support Vector Machine sama-sama menunjukkan performa yang sangat tinggi dalam mengklasifikasikan glaukoma. Random Forest memperoleh akurasi sebesar 97,92% dengan nilai precision 100,00%, recall 96,83%, dan F1-score 98,39%. Sementara itu, Support Vector Machine menunjukkan performa yang sedikit lebih baik dengan akurasi sebesar 98,96%, precision 98,44%, recall 100,00%, dan F1-score 99,21%.

Perbedaan performa ini menunjukkan bahwa Support Vector Machine memiliki kemampuan yang lebih baik dalam membentuk batas keputusan yang optimal pada data dengan fitur yang telah melalui proses standarisasi. Hal ini sesuai dengan karakteristik algoritma SVM yang sensitif terhadap skala data dan mampu bekerja secara optimal pada data berdimensi rendah dengan distribusi yang relatif terpisah dengan jelas.Sebaliknya, Random Forest menunjukkan performa yang sangat baik dalam mengklasifikasikan data dengan tingkat kesalahan yang rendah, terutama pada kesalahan false positive. Hal ini menunjukkan bahwa Random Forest mampu melakukan pemodelan hubungan non-linear antar fitur dengan baik dan relatif stabil terhadap variasi data.

Pada dataset pembanding, performa kedua algoritma mengalami penurunan yang cukup signifikan. Random Forest memperoleh akurasi sebesar 68,46%, sedangkan Support Vector Machine memperoleh akurasi sebesar 72,31%. Meskipun SVM memiliki akurasi yang lebih tinggi, Random Forest menunjukkan karakteristik yang lebih stabil dalam menghadapi perbedaan distribusi data. Hasil ini mengindikasikan bahwa pemilihan algoritma yang optimal sangat bergantung pada karakteristik dataset yang digunakan.

F. Analisis Kemampuan Generalisasi Model

Pengujian pada dataset pembanding dilakukan untuk mengevaluasi kemampuan generalisasi model dalam mengklasifikasikan data glaukoma dari sumber yang berbeda dengan data pelatihan. Hasil pengujian menunjukkan adanya penurunan performa yang signifikan dibandingkan dengan pengujian pada dataset rumah sakit.

Penurunan performa ini disebabkan oleh beberapa faktor, antara lain perbedaan distribusi data antara dataset rumah sakit dan dataset pembanding, keterbatasan atribut klinis yang tersedia pada dataset eksternal, serta perbedaan prosedur pengambilan data.

Dataset pembanding tidak memiliki kelengkapan fitur klinis yang sama dengan dataset pelatihan, sehingga informasi yang digunakan oleh model menjadi terbatas.

Hasil ini menunjukkan bahwa model cenderung menyesuaikan diri dengan karakteristik data pelatihan (domain-specific). Oleh karena itu, meskipun model memiliki performa yang sangat baik pada data internal, diperlukan penyesuaian lebih lanjut atau pelatihan ulang model agar dapat digunakan secara optimal pada data dari sumber yang berbeda. Temuan ini penting untuk menjadi pertimbangan sebelum sistem diterapkan secara luas dalam lingkungan klinis yang beragam.

G. Analisis Kesalahan Klasifikasi

Berdasarkan confusion matrix pada dataset pembanding, kesalahan klasifikasi yang terjadi sebagian besar berupa false negative dan false positive, khususnya pada algoritma Random Forest. False negative menunjukkan bahwa beberapa kasus glaukoma diklasifikasikan sebagai normal, sedangkan false positive menunjukkan bahwa data normal diklasifikasikan sebagai glaukoma.

Dalam konteks medis, kesalahan false negative memiliki dampak yang lebih serius karena dapat menyebabkan keterlambatan dalam penanganan pasien yang sebenarnya menderita glaukoma. Kesalahan ini kemungkinan disebabkan oleh keterbatasan fitur klinis yang digunakan serta perbedaan karakteristik data antara dataset pelatihan dan dataset pengujian eksternal.

Pada algoritma Support Vector Machine, jumlah false negative lebih rendah dibandingkan Random Forest, namun masih ditemukan kesalahan klasifikasi. Hal ini menunjukkan bahwa meskipun SVM memiliki kemampuan generalisasi yang sedikit lebih baik, model tetap membutuhkan penyesuaian fitur atau penambahan data pelatihan yang lebih beragam agar tingkat kesalahan dapat diminimalkan.

KESIMPULAN

Berdasarkan hasil penelitian mengenai penerapan algoritma Random Forest dan Support Vector Machine dalam klasifikasi penyakit glaukoma menggunakan data tabular pasien, maka dapat disimpulkan sebagai berikut :

1. Penelitian ini berhasil membangun model klasifikasi penyakit glaukoma berbasis machine learning menggunakan atribut usia, jenis kelamin, visus mata kanan, dan visus mata kiri untuk mengelompokkan data pasien ke dalam dua kelas, yaitu Normal dan Glaukoma.
2. Berdasarkan hasil pengujian pada dataset rumah sakit, algoritma Random Forest menghasilkan performa yang sangat baik dengan nilai:
    a. Accuracy sebesar 97,92%
    b. Precision sebesar 100%
    c. Recall sebesar 96,83%
    d. F1-score sebesar 98,39%
3. Pada dataset rumah sakit, algoritma Support Vector Machine (SVM) menghasilkan performa yang lebih tinggi dibandingkan Random Forest dengan nilai:
    a. Accuracy sebesar 98,96%
    b. Precision sebesar 98,44%
    c. Recall sebesar 100%
    d. F1-score sebesar 99,21%
4. Pengujian menggunakan dataset pembanding sebagai data uji eksternal menunjukkan adanya penurunan performa pada kedua algoritma.
5. Pada dataset pembanding, algoritma Random Forest memperoleh nilai:

   a. Accuracy sebesar 68,46%
   b. Precision sebesar 41,03%
   c. Recall sebesar 47,06%
   d. F1-score sebesar 43,84%
6. Pada dataset pendamping, algoritma Support Vector Machine (SVM) memperoleh nilai:
   a. Accuracy sebesar 72,31%
   b. Precision sebesar 48,00%
   c. Recall sebesar 70,59%
   d. F1-score sebesar 57,14%
7. Penurunan performa pada dataset eksternal menunjukkan adanya perbedaan karakteristik distribusi data serta keterbatasan atribut yang digunakan sehingga memengaruhi kemampuan generalisasi model.
8. Secara keseluruhan, model yang dibangun memiliki potensi sebagai alat bantu skrining awal glaukoma, namun masih memerlukan pengembangan lebih lanjut melalui penambahan atribut klinis, peningkatan jumlah dataset, dan penggunaan data yang lebih beragam.

Saran

Berdasarkan hasil penelitian yang telah dilakukan, maka saran yang dapat diberikan adalah sebagai berikut:

1. Penelitian selanjutnya disarankan menggunakan dataset dengan jumlah yang lebih besar dan berasal dari berbagai institusi pelayanan kesehatan agar kemampuan generalisasi model menjadi lebih baik.
2. Penelitian selanjutnya disarankan menambahkan atribut klinis yang lebih lengkap seperti tekanan intraokular, cup-to-disc ratio (CDR), hasil pemeriksaan lapang pandang, serta data citra fundus retina agar model dapat menghasilkan klasifikasi yang lebih akurat.
3. Perlu dilakukan proses preprocessing data yang lebih optimal seperti normalisasi data, seleksi fitur, serta penyeimbangan distribusi kelas terutama pada penggunaan dataset eksternal.
4. Penelitian selanjutnya disarankan melakukan optimasi parameter pada algoritma Random Forest dan Support Vector Machine untuk memperoleh performa model yang lebih stabil.
5. Penelitian selanjutnya dapat mengembangkan penggunaan algoritma klasifikasi lain seperti k-Nearest Neighbor (k-NN), Naïve Bayes, Logistic Regression, serta algoritma boosting seperti XGBoost, LightGBM, dan CatBoost yang memiliki performa baik pada data tabular.
6. Apabila jumlah dataset semakin besar, pendekatan deep learning seperti Artificial Neural Network (ANN) dapat digunakan untuk menangkap hubungan antar fitur yang lebih kompleks.

DAFTAR PUSTAKA

American Optometric Association. (2022). Glaucoma. American Optometric Association.

Arifiyanti, A. A., & Wahyuni, S. (2020). Evaluasi kinerja algoritma klasifikasi menggunakan confusion matrix. Jurnal Teknologi Informasi, 14(2), 85–92.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

Chawla, N. V. (2022). Data imbalance, sampling techniques, and evaluation metrics. ACM Computing Surveys, 54(6), 1–36.

Chevrier, R., Foufi, V., Gaudet-Blavignac, C., Robert, A., & Lovis, C. (2019). Use and

understanding of anonymization and de-identification in biomedical research. Journal of Biomedical Informatics, 91, 103112.
Fallo, S. I., & Nay, F. A. (2024). Klasifikasi support vector machine dan random forest pada data biomedis: Aplikasi dalam analisis data penyakit diabetes. Asimtot: Jurnal Kependidikan Matematika, 6(1), 63–73
Hasdyna, N., & Dinata, R. (2020). Matthew correlation coefficient untuk evaluasi klasifikasi data tidak seimbang. Jurnal Ilmu Komputer, 8(2), 101–109.
Kasanah, U., Sari, N. R., & Putra, A. (2019). Analisis akurasi algoritma klasifikasi data medis. Jurnal Informatika Kesehatan, 7(1), 22–30.
Lingga, R. A., Sembiring, J., & Sinaga, K. (2017). Penerapan support vector machine dalam klasifikasi data. Jurnal Teknologi dan Sistem Komputer, 5(3), 120–127.
Lutz, M. (2013). Learning Python (5th ed.). O'Reilly Media.
Meurers, D., et al. (2025). Risks of re-identification in anonymized medical datasets. BMC Medical Ethics, 26(4), 1–12.
Novaković, J., Strbac, P., & Bulatović, D. (2017). Toward optimal feature selection using ranking methods. Applied Soft Computing, 20, 305–316.
Nugroho, A. S., Witarto, A. B., & Handoko, D. (2003). Support vector machine: Teori dan aplikasinya. Bandung: Informatika.
NumPy Developers. (2013). NumPy user guide. NumPy.org.
Nurfajar, A., Prasetyo, E., & Wibowo, A. (2022). Deteksi glaukoma menggunakan LBP dan support vector machine. Jurnal Teknologi Informasi dan Ilmu Komputer, 9(3), 455–463.
Pedregosa, F., et al. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.
Pertiwi, A. A., & Utomo, A. N. (2023). Implementasi metode support vector machine (SVM) pada citra fundus retina mata untuk deteksi penyakit glaukoma. Jurnal Rekayasa Informasi, 12(1).
Rahman, A., & Susilo, B. (2023). Perbandingan random forest dan SVM pada klasifikasi data medis. Jurnal Ilmu Komputer dan Informasi, 11(2), 95–104.
Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. IBM Journal of Research and Development, 3(3), 210–229.
Sepas, R., et al. (2022). Medical data anonymization techniques and ethics. Health Informatics Journal, 28(1), 1–15.
Siringoringo, H. (2018). Pemodelan dan evaluasi sistem klasifikasi. Yogyakarta: Andi Offset.
Streamlit Inc. (2021). Streamlit documentation. Streamlit.io.
Sunata, M., Hidayat, R., & Kurniawan, D. (2020). Analisis performa random forest pada data medis. Jurnal Informatika, 14(1), 55–63.
Syahrudin, M., & Kurniawan, R. (2018). Pemrograman Python untuk data science. Bandung: Informatika.
Tempola, F., Muhammad, M., & Khairan, A. (2018). Kinerja random forest dalam klasifikasi data besar. Jurnal Teknologi Informasi, 6(1), 12–20.
Wibisono, A., & Fahrurozi, I. (2019). Implementasi random forest pada klasifikasi supervised learning. Jurnal Komputer Terapan, 5(2), 77–85.
Wibowo, A. (2016). Penerapan random forest dalam data mining. Yogyakarta: Deepublish.
World Health Organization. (2023). World report on vision. WHO Press.
Zareapoor, M., & Shamsolmoali, P. (2015). Application of credit card fraud detection based on bagging ensemble classifier. Procedia Computer Science, 48, 679–685.