BAB III
METODOLOGI PENELITIAN

3.1 Objek Penelitian

Objek penelitian pada studi ini adalah citra digital mata manusia yang merepresentasikan tiga kategori kondisi kesehatan mata, yaitu:

1. *Cataract* (Katarak) → ditandai dengan kekeruhan pada lensa mata yang mengakibatkan penurunan ketajaman penglihatan.
2. *Diabetic Retinopathy* → kerusakan pembuluh darah retina akibat komplikasi diabetes mellitus yang ditandai dengan perdarahan, mikroaneurisma, dan eksudat.
3. *Glaucoma* (Glaukoma) → ditandai dengan kerusakan progresif pada saraf optik yang umumnya berkaitan dengan peningkatan tekanan intraokular. Citra mata glaukoma menunjukkan pelebaran cekungan diskus optik (*optic disc cupping*) dan penipisan lapisan serabut saraf retina.

Dataset yang digunakan dalam penelitian ini diperoleh dari *repositori publik Kaggle* dengan tautan: https://www.kaggle.com/datasets/gunavenkatdoddi/eye-diseases-classification.

Dataset yang digunakan dalam penelitian ini terdiri dari 900 citra mata yang terbagi menjadi tiga kategori, yaitu 300 citra kategori *Cataract*, 300 citra kategori *Diabetic Retinopathy*, dan 300 citra kategori *Glaucoma*. Seluruh citra tersedia dalam format RGB dengan resolusi yang bervariasi, sehingga pada tahap *preprocessing resolusi* gambar akan diseragamkan untuk memastikan konsistensi input model. Setiap citra telah dilengkapi dengan label yang disediakan oleh penyedia dataset berdasarkan hasil diagnosis medis, sehingga data siap digunakan untuk proses pelatihan dan evaluasi model klasifikasi.
Pemilihan dataset ini didasarkan pada beberapa pertimbangan:

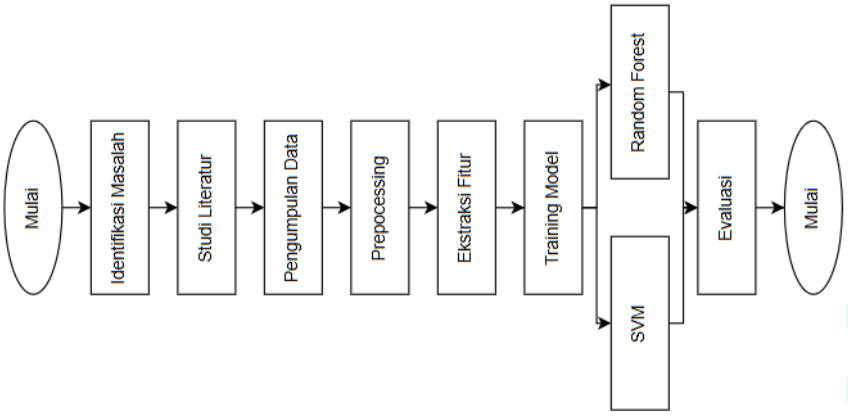
1. Relevansi : dataset memuat jenis penyakit mata yang menjadi fokus penelitian.

2. Kualitas Gambar : citra memiliki resolusi yang memadai untuk analisis visual dan ekstraksi fitur.
3. Keseimbangan Kelas : jumlah citra pada tiap kategori seimbang sehingga dapat meminimalkan bias dalam pelatihan model.
4. Ketersediaan Terbuka : dataset bersifat *open access* sehingga dapat digunakan, diolah, dan direplikasi oleh peneliti lain.

Dalam penelitian ini, citra digunakan sebagai input untuk proses *preprocessing*, ekstraksi fitur dengan metode *Histogram of Oriented Gradients* (HOG), dan klasifikasi menggunakan algoritma *Support Vector Machine* (SVM) serta *Random Forest*. Hasil klasifikasi diukur menggunakan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score untuk mengetahui performa model pada objek penelitian ini.

3.2 Prosedur Penelitian

Penelitian ini melalu beberapa tahapan yaitu identifikasi masalah, studi literatur, pengumpulan data, *preprocessing data*, *feature extraction*, implementasi algoritma *Support Vector Machine*, dan *evaluation*.



Gambar 3. 1 Alur Penelitian

3.2.2Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh dari situs web Kaggle.com melalui tautan:

https://www.kaggle.com/datasets/gunavenkatdoddi/eye-diseases-classification.

Dataset tersebut terdiri dari total 900 citra mata yang telah terbagi secara seimbang ke dalam tiga kelas, yaitu:

1. *Cataract* (Katarak) : 300 citra
2. *Diabetic Retinopathy* : 300 citra
3. *Glaucoma* : 300 citra

Seluruh citra berformat RGB dengan resolusi yang bervariasi. Setiap data gambar telah disertai label sesuai dengan kondisi mata, sehingga memudahkan proses klasifikasi menggunakan algoritma *machine learning*.

Pemilihan *dataset* ini didasarkan pada beberapa pertimbangan, antara lain:

1. Relevansi: Dataset secara langsung berhubungan dengan tujuan penelitian, yaitu klasifikasi penyakit mata.
2. Keseimbangan kelas: Jumlah data yang seimbang pada masing-masing kategori mendukung performa model yang lebih stabil.
3. Kualitas data: Gambar memiliki kualitas yang cukup baik untuk mendukung proses preprocessing, pelatihan, dan pengujian.
4. Ketersediaan label: Setiap gambar telah dilengkapi label yang valid, sehingga dapat mempercepat proses analisis.
5. Aksesibilitas: Dataset bersifat terbuka sehingga mendukung replikasi dan validasi hasil penelitian oleh peneliti lain di masa mendatang.

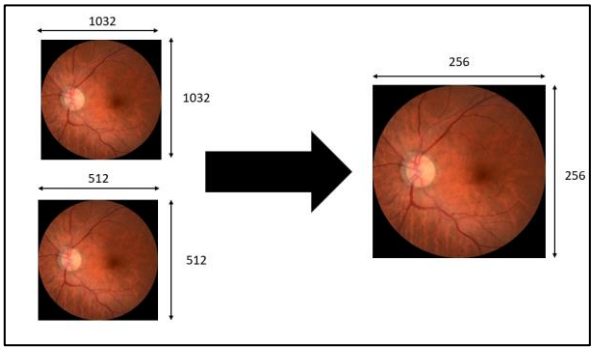
3.2.3 *Preprocessing*

Tahap preprocessing merupakan langkah awal dalam pengolahan data citra sebelum digunakan pada proses pelatihan model. *Preprocessing* bertujuan untuk menyamakan karakteristik data agar sesuai dengan kebutuhan algoritma machine learning serta meningkatkan kualitas input.

Beberapa tahapan *preprocessing* yang dilakukan pada penelitian ini adalah sebagai berikut:

1. ***Resizing Citra***

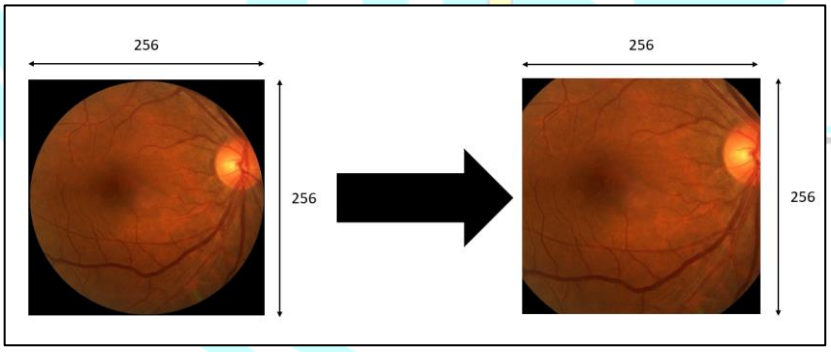
   Seluruh citra diubah ukurannya menjadi 256 × 256 piksel. Normalisasi ukuran ini diperlukan untuk memastikan konsistensi dimensi gambar sehingga mempermudah proses ekstraksi fitur dan pelatihan model.



Gambar 3. 2 Proses *Resizing* Data Gambar

2. **Augmentasi Data (*Zooming/Scaling*)**

   Setelah proses *resizing*, dilakukan augmentasi data dengan teknik *zooming*. Pada tahap ini, citra diperbesar secara proporsional dengan faktor 1,2 kali. Tujuan dari augmentasi adalah untuk menghasilkan variasi data yang lebih beragam sehingga dapat mengurangi risiko *overfitting* serta meningkatkan kemampuan generalisasi model.



Gambar 3. 3 Proses *Augmenation zoom*

3.2.4 Ekstraksi Fitur

Setelah tahap *preprocessing*, langkah berikutnya adalah melakukan ekstraksi fitur guna mendapatkan representasi numerik dari citra. Pada penelitian ini digunakan metode *Histogram of Oriented Gradients* (HOG).

Tahapan ekstraksi fitur HOG adalah sebagai berikut:

1. **Konversi ke *Grayscale***

   Citra berwarna (RGB) diubah ke dalam skala abu-abu (*grayscale*) agar lebih sederhana namun tetap mempertahankan informasi tepi dan kontur.

2. **Pembagian ke dalam Sel**

   Gambar dibagi menjadi sel-sel kecil berukuran tertentu.

3. **Perhitungan Gradien Orientasi**

   Gradien orientasi pada setiap sel dihitung untuk mendeteksi pola tepi dan arah kontur pada citra.

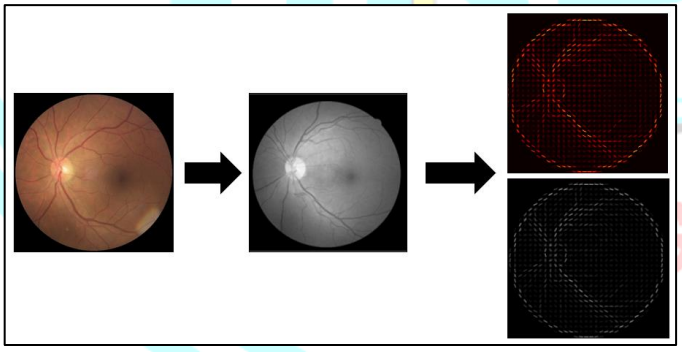
4. **Pembentukan Histogram**

   Nilai gradien orientasi pada masing-masing sel disusun dalam bentuk histogram.

5. **Vektor Fitur HOG**

   Histogram-histogram tersebut kemudian digabungkan sehingga menghasilkan vektor fitur HOG, yang merepresentasikan distribusi gradien orientasi dari keseluruhan citra.

Hasil ekstraksi fitur ini digunakan sebagai representasi yang lebih informatif dan diskriminatif untuk proses klasifikasi penyakit mata.



Gambar 3. 4 Proses *Ektrasi* Fitur Menggunakan HOG

Sebelum membangun model, dataset dibagi menjadi dua subset, yaitu:

1. Training set (80%) → digunakan untuk melatih model.
2. Testing set (20%) → digunakan untuk mengevaluasi performa model.

Proporsi ini dipilih agar model memiliki cukup banyak data untuk belajar, sekaligus dapat diuji pada data yang belum pernah dilihat, sehingga akurasi hasil klasifikasi dapat diukur secara objektif.

1. ***Training Support Vectore Machine (SVM)***

Setelah tahap *preprocessing* dan ekstraksi fitur selesai, langkah berikutnya adalah membangun model klasifikasi menggunakan algoritma *Support Vector*

Machine (SVM).

Pada penelitian ini, proses pelatihan model dilakukan dengan bantuan *GridSearchCV*, yang berfungsi untuk mencari kombinasi *hyperparameter* terbaik. Beberapa *hyperparameter* yang dieksplorasi meliputi:

1. **C** → parameter regulasi yang mengontrol *trade-off* antara margin yang lebar dan kesalahan klasifikasi.
2. **Gamma (γ)** → parameter kernel RBF yang menentukan seberapa jauh pengaruh satu data pelatihan.
3. **Kernel** → fungsi kernel yang digunakan untuk memetakan data ke ruang berdimensi lebih tinggi (misalnya *linear, RBF, polynomial*).

Untuk meningkatkan reliabilitas hasil, digunakan metode cross validation sebanyak 5 kali (5-*fold cross validation***)**. Pendekatan ini membagi data latih menjadi lima bagian, kemudian model dilatih dan divalidasi secara bergantian, sehingga dapat mengurangi bias dan memberikan estimasi performa yang lebih akurat.

2. ***Training Algoritma Random Forest***

Setelah tahap *preprocessing* data selesai, langkah berikutnya adalah membangun model *Random Forest*. Dalam penelitian ini, digunakan pendekatan penyesuaian *hyperparameter* untuk memperoleh performa terbaik. Beberapa *hyperparameter* yang dieksplorasi meliputi jumlah pohon keputusan (*n_estimators*), jumlah fitur yang dipertimbangkan pada setiap pemisahan (*max_features*), dan kedalaman maksimum pohon (*max_depth*).

Proses pelatihan dilakukan dengan metode *bootstrap sampling*, di mana setiap pohon dibangun dari subset data latih yang diambil secara acak dengan pengembalian (*with replacement*). Pemisahan pada setiap node ditentukan menggunakan *Gini Index*:

Selain itu, dilakukan validasi silang (*cross-validation*) sebanyak 5 kali untuk mendapatkan estimasi kinerja model yang lebih reliabel, dengan evaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score.

Selain algoritma *Support Vector Machine (SVM)*, penelitian ini juga membangun model klasifikasi dengan menggunakan algoritma *Random*

Forest. Random Forest merupakan metode *ensemble learning* yang membangun sejumlah pohon keputusan (*decision trees*) dan menggabungkan hasil prediksi dari masing-masing pohon untuk memperoleh keputusan akhir yang lebih stabil dan akurat.

Pada penelitian ini, proses pelatihan model *Random Forest* dilakukan melalui penyesuaian *hyperparameter* dengan tujuan memperoleh performa terbaik. *Hyperparameter* yang dieksplorasi meliputi:

1. ***n_estimators*** → jumlah pohon keputusan yang digunakan.
2. ***max_features*** → jumlah fitur yang dipertimbangkan pada setiap pemisahan node.
3. ***max_depth*** → kedalaman maksimum pohon keputusan.

Pelatihan dilakukan menggunakan metode *bootstrap sampling*, di mana setiap pohon dibangun dari subset data latih yang diambil secara acak dengan pengembalian (*with replacement*). Proses pemisahan node pada setiap pohon ditentukan menggunakan kriteria Gini Index, yaitu ukuran tingkat ketidakmurnian suatu node.

Untuk menguji keandalan model, digunakan metode *cross validation* sebanyak 5 kali (5-*fold cross validation*). Dengan metode ini, dataset latih dibagi ke dalam lima bagian, kemudian model dilatih dan divalidasi secara bergantian. Evaluasi model dilakukan menggunakan metrik:

1. Akurasi
2. Presisi
3. Recall
4. F1-score

3.2.6 Evaluasi

1. *Confusion Matrix*

Merupakan sebuah metode untuk menampilkan jumlah data uji yang akan di klasifikasikan untuk mendapatkan hasil yang benar dan salah, untuk memudahkan evaluasi sebuah akurasi pada suatu sistem klasifikasi. *Confusion matrix* merupakan

metode yang mudah dan juga efektif untuk mengukur sebuah kinerja

sistem klasifikasi. Tujuan utama metode ini adalah menilai performa atau akurasi pada sebuah sistem klasifikasi untuk mengklasifikasikan data uji.

Akurasi merupakan ketepatan antara nilai prediksi dengan nilai aktual.

Rumus untuk menghitung nilai akurasi dapat dilihat pada Persamaan 8.

$$Akurasi = \frac{(TP+TN)}{(TP+FP+FN+TN)} \quad (10)$$

Presisi merupakan tingkat keberhasilan model dalam memberikan jawaban dengan tepat kepada pengguna. Rumus untuk menghitung nilai presisi dapat dilihat pada Persamaan 10.

$$Presisi = \frac{TP}{TP+FP} \quad (11)$$

Recall merupakan tingkat keberhasilan model dalam menemukan kembali informasi dengan benar. Rumus untuk menghitung nilai recall dapat dilihat pada Persamaan 11.

$$Recall = \frac{TP}{TP+FN} \quad (12)$$

F1-Score atau F-Measure. F1- Score merupakan hasil perbandingan antara nilai presisi dan recall. Rumus untuk menghitung nilai f1-Score dapat dilihat pada Persamaan 12.

$$F1 - Score = \frac{2 \text{ x } Presisi \text{ x } Recall}{Presisi + Recall} \quad (13)$$

Keterangan:

TP : True Positive

TN : True Negative

FP : False Positive

FN : False Negative

Tabel 3. 1 Confusion Matrix

Confusion Matrix		Prediksi		
		Positif	*Non Aspek*	*Negatif*
Aktual	*Positif*	*Tpospos*	*PFnon*	*PFNeg*
	Non Aspek	*NonFP*	*TNonNon*	*NonFNeg*
	Negatif	*NegFP*	*NegFNon*	*TNegNeg*

Seperti yang di tunjukan pada tabel 3.1 di atas *confusion matrix*, yaitu sebuah tabel yang digunakan untuk mengevaluasi kinerja model klasifikasi. *Confusion matrix* menampilkan perbandingan antara hasil prediksi model dengan kondisi aktual yang sebenarnya. Pada tabel ini terdapat tiga kelas utama, yaitu *Positif, Non Aspek, dan*

Negatif, baik pada sisi aktual maupun pada sisi prediksi. Melalui *confusion matrix*, dapat diketahui jumlah prediksi yang benar (*true*) maupun yang salah (*false*) pada setiap kategori. Informasi ini sangat penting karena tidak hanya menunjukkan tingkat akurasi model secara keseluruhan, tetapi juga memberikan gambaran detail mengenai seberapa baik model mampu membedakan tiap kelas. Dengan demikian, *confusion matrix* menjadi dasar dalam menghitung berbagai metrik evaluasi, seperti *precision, recall, f1-score, dan accuracy*, yang digunakan untuk menilai kualitas performa model.

UBP
KARAWANG