e-ISSN 2828-0490
JURNAL MEDIA COMPUTER SCIENCE

The Implementation Of Computer Vision For Eye Disease Classification Using Vision Transformer Architecture On Fundus Images

Penerapan Computer Vision Untuk Klasifikasi Penyakit Mata Menggunakan Arsitektur Vision Transformers Pada Citra Fundus

Indra Yustiana[1)]; Irvan Yudistiansyah [2)]; M. Ikhsan Thohir [3)]
[1),2),3)]***Prodi Teknik Informatika, Fakultas Teknik Komputer dan Desain, Universitas Nusa Putra***
Email: [1)] indra.yustiana@nusaputra.ac.id ;[2] irvan.yudistiansyah_ti21@nusaputra.ac.id
[3)] ikhsan.thohir@nusaputra.ac.id

How to Cite :
Yustiana, I., Yudistiansyah, I., Thohir, M, I. (2026). The Implementation Of Computer Vision For Eye Disease Classification Using Vision Transformer Architecture On Fundus Images. Jurnal Media Computer Science, 5(1)

ARTICLE HISTORY
Received [30 Juli 2026]
Revised [20 Januari 2026]
Accepted [25 Januari 2026]

KEYWORDS
Vision Transformer, Fundus Image, Eye Disease Classification, Deep Learning, Computer Vision.

ABSTRAK

Penyakit mata seperti katarak, glaucoma, dan diabetic retinopathy merupakan penyebab utama kebutaan yang dapat dicegah jika dideteksi secara dini, penelitian ini bertujuan untuk mengembangkan sistem klasifikasi penyakit mata menggunakan arsitektur vision transformer menggunakan citra fundus, data yang digunakan terdiri dari 4.028 citra fundus yang terbagi secara seimbang kedalam empat kelas yaitu, katarak, glaucoma, diabetic retinopathy, dan normal. Model vision transformer dilatih melalui tahapan preprocessing, augmentasi, fine-tuning, dan evaluasi menggunakan metrik seperti akurasi, presisi, recall, dan f1-score. Hasil pengujian menunjukan bahwa model vision transformer mampu mengklasifikasikan penyakit mata dengan akurasi tinggi dan performa stabil di seluruh kelas. Model ini juga mampu mengenali ciri khas masing-masing penyakit secara efektif, membuktikan keunggulan vision transformer dalam memahami konteks visual global citra medis. Dari penelitian ini menyatakan bahwa vision transformer dapat digunakan secara efektif dalam sistem deteksi otomatis penyakit mata menggunakan citra fundus, meskipun masih memerluka optimalisasi lebih lanjut untuk penggunaan perangkat dengan sumber daya terbatas. Sistem ini diharapkan dapat membantu proses deteksi dini dan mengurangi beban kerja tenaga medis.

ABSTRACT

Eye diseases such as cataracts, glaucoma, and diabetic retinopathy are leading causes of blindness that can be prevented if detected early. This study aims to develop an eye disease classification system using a vision transformer architecture using fundus images. The data used consisted of 4,028 fundus images evenly divided into four classes: cataracts, glaucoma, diabetic retinopathy, and normal. The vision transformer model underwent preprocessing, augmentation, fine-tuning, and evaluation using metrics such as accuracy, precision, recall, and f1-score. Test results showed that the vision transformer model was able to classify eye diseases with high accuracy and stable performance across all classes. This model was also able to effectively recognize the characteristics of each disease,

demonstrating the superiority of the vision transformer in understanding the global visual context of medical images.This study suggests that the vision transformer can be effectively used in an automated eye disease detection system using fundus images, although further optimization is still needed for use with devices with limited resources. This system is expected to facilitate early detection and reduce the workload of medical personnel.

PENDAHULUAN

Penyakit mata merupakan salah satu masalah kesehatan yang berdampak signifikan terhadap kualitas hidup masyarakat di seluruh dunia (Adjeng dkk., 2024). Penyakit seperti retinopati diabetik, glaukoma, dan katarak sering kali berkembang tanpa gejala awal yang jelas, sehingga diagnosis dini menjadi sangat penting guna mencegah terjadinya kebutaan permanen (Dana, 2020). Berdasarkan data Badan Pusat Statistik (BPS) tahun 2022, tercatat sekitar 8 juta kasus gangguan penglihatan di Indonesia, dengan 272 ribu kasus kebutaan dan sekitar 700 ribu kasus kesulitan penglihatan yang serius (Wirawan, 2024). Salah satu metode utama dalam mendeteksi penyakit mata adalah dengan menggunakan citra fundus, yaitu gambaran visual bagian belakang bola mata (retina) yang memberikan informasi penting mengenai kondisi kesehatan mata (Retina, 2022).

Namun, proses pemeriksaan dan interpretasi citra fundus secara manual membutuhkan tenaga medis yang memiliki keahlian tinggi, serta waktu yang tidak sedikit (Jatmoko & Lestiawan, 2024). Terbatasnya jumlah dokter spesialis mata, khususnya di wilayah terpencil, menjadi tantangan tersendiri dalam pemeriksaan mata secara berkala dan menyeluruh (Bintang & Imaduddin, 2024). Seiring berkembangnya teknologi kecerdasan buatan, khususnya dalam bidang *computer vision*, berbagai metode otomatis telah dikembangkan untuk membantu proses deteksi penyakit mata melalui citra fundus (Putri & Rakasiwi, 2025). Salah satu pendekatan terbaru yang menunjukkan hasil menjanjikan adalah Vision Transformer (ViT), yang menggunakan mekanisme *self-attention* untuk menangkap hubungan spasial dan semantik antarbagian citra secara lebih menyeluruh (Wu dkk., 2021).

Penelitian-penelitian sebelumnya telah menunjukkan bahwa model deep learning dapat memberikan hasil yang cukup akurat. (Bintang & Imaduddin, 2024) melakukan perbandingan arsitektur CNN seperti ResNet152V2, Xception, DenseNet201, dan InceptionV3 untuk klasifikasi retinopati diabetik, dan melaporkan akurasi mencapai 96%. Sementara itu, (Wu dkk., 2021) menggunakan model Vision Transformer murni untuk klasifikasi tingkat keparahan retinopati diabetik dan memperoleh akurasi hingga 91,4%. Penelitian lainnya oleh (Yang dkk., 2024) menerapkan Vision Transformer dengan pendekatan *Masked Autoencoders* (MAE) pada dataset fundus berskala besar dan berhasil mencapai akurasi sebesar 93,42%, menunjukkan efektivitas pretraining dengan data domain-spesifik.

Meskipun telah banyak dilakukan penelitian terkait klasifikasi penyakit mata menggunakan citra fundus, namun sebagian besar masih berfokus pada satu jenis penyakit, seperti retinopati diabetik. Belum banyak penelitian yang mengeksplorasi klasifikasi multi-penyakit seperti glaukoma dan katarak secara bersamaan dalam satu sistem. Selain itu, sebagian besar penelitian juga menggunakan dataset global dan belum banyak yang menggunakan data dari Indonesia yang memiliki karakteristik populasi dan kondisi pencitraan berbeda. Di sisi lain, pemanfaatan arsitektur Vision Transformer dalam konteks lokal juga masih terbatas, serta belum banyak dilakukan penelitian terkait optimalisasi model ini untuk digunakan di wilayah dengan keterbatasan sumber daya.

Oleh karena itu, penelitian ini hadir dengan tujuan untuk mengembangkan sistem klasifikasi otomatis penyakit mata berdasarkan citra fundus dengan pendekatan Vision Transformer yang mampu mengenali berbagai jenis penyakit mata secara sekaligus. Keunggulan dari arsitektur Vision Transformer dalam memahami konteks visual secara menyeluruh diharapkan mampu meningkatkan akurasi dan efisiensi dalam proses deteksi penyakit mata, khususnya dalam konteks

pelayanan kesehatan di Indonesia. Inovasi utama dari penelitian ini terletak pada penerapan arsitektur ViT pada klasifikasi multi-penyakit mata, pemanfaatan dataset lokal, serta pengembangan sistem yang dapat diadopsi pada wilayah dengan keterbatasan tenaga medis.

LANDASAN TEORI

Klasifikasi

Klasifikasi merupakan proses pengelompokan data berdasarkan kesamaan karakteristik atau fitur yang dimiliki. Entitas atau objek yang memiliki ciri-ciri serupa dikelompokkan ke dalam satu kelas, sedangkan entitas yang berbeda ditempatkan pada kelas yang berbeda. Proses ini bertujuan untuk mengidentifikasi pola dalam data dan melakukan prediksi terhadap kelas suatu data baru berdasarkan hasil pelatihan model sebelumnya (Chandana dkk., 2024; Hadiprakoso & Buana, 2021). Secara umum, proses klasifikasi terdiri dari dua tahap utama, yaitu *training* (pelatihan) dan *testing* (pengujian). Pada tahap pelatihan, sistem akan mempelajari pola dari data yang sudah diketahui kelasnya menggunakan algoritma tertentu. Sedangkan pada tahap pengujian, model yang telah dilatih akan digunakan untuk memprediksi kelas dari data baru yang belum diketahui kelasnya (Sarker dkk., 2019). Hasil dari proses klasifikasi ini kemudian dievaluasi melalui metrik seperti akurasi, presisi, dan recall untuk mengukur kinerja model dalam mengelompokkan data secara tepat (Rifqi, 2024).

Mata

Mata adalah salah satu indra penglihatan manusia yang berfungsi untuk menerima dan memproses rangsangan cahaya dari lingkungan (Budiarti, 2023). Mata secara otomatis menyesuaikan diri terhadap intensitas cahaya, memfokuskan penglihatan pada objek yang berada pada jarak dekat maupun jauh, serta menghasilkan citra visual yang kemudian dikirimkan ke otak untuk diinterpretasikan. Struktur mata terdiri atas bagian luar dan dalam, seperti kelopak mata, alis, kornea, retina, dan pupil, yang masing-masing memiliki fungsi tersendiri dalam mendukung proses penglihatan (Sepe & Stefanus Stanis, 2023).

Penyakit Mata

Penyakit mata merupakan kondisi gangguan kesehatan yang menyerang organ mata dan mengakibatkan terganggunya fungsi penglihatan. Gejala umum yang sering muncul meliputi mata merah, gatal, perih, gangguan penglihatan seperti rabun dekat maupun rabun jauh, serta dalam kasus yang parah dapat menyebabkan kebutaan. Beberapa jenis penyakit mata yang umum terjadi antara lain adalah katarak, glaukoma, dan penyakit retina (*retinal disease*) (Muhlashin & Stefanie, 2023).

Deep Learning

Deep Learning merupakan salah satu cabang dari *machine learning* yang berfokus pada pemrosesan data menggunakan jaringan saraf tiruan (*Artificial Neural Network* atau ANN) dengan struktur lapisan (layer) yang sangat dalam (Jamil & Pulukadang, 2025). Metode ini mampu mengidentifikasi pola-pola tersembunyi dalam data, kemudian mempelajarinya untuk melakukan klasifikasi dan prediksi secara otomatis guna menghasilkan output yang akurat (Nurhakiki & Yahfizham, 2024). Keunggulan dari deep learning terletak pada kemampuannya untuk melakukan pembelajaran fitur secara bertingkat atau hierarkis, sehingga semakin dalam lapisan yang digunakan, semakin kompleks pola yang dapat dikenali oleh model (Tamba, 2024).

Dalam proses pembelajaran, *deep learning* memungkinkan sistem komputer untuk belajar dari pengalaman secara mandiri tanpa diprogram secara eksplisit (Nugraha dkk., 2023). Proses ini mencakup analisis data melalui berbagai lapisan, mulai dari input mentah hingga ekstraksi fitur kompleks yang kemudian menghasilkan keputusan (Dhamayanti dkk., 2021). *Deep learning* juga

dinilai efisien dalam waktu pemrosesan untuk data berskala besar, karena dapat melakukan generalisasi data secara optimal dengan minim intervensi manusia (Hutagalung & Sitompul, 2023).

Vision Transformers

Vision Transformer (ViT) adalah salah satu model berbasis arsitektur *transformer* yang diadaptasi untuk pemrosesan citra visual. Berbeda dengan pendekatan konvensional seperti *Convolutional Neural Network* (CNN), ViT membagi gambar input ke dalam potongan-potongan kecil yang disebut *patch*, kemudian memproses patch tersebut secara paralel menggunakan mekanisme *self-attention* (Dosovitskiy dkk., 2020). Proses ini memungkinkan model untuk memahami fitur global dari gambar tanpa harus bergantung pada struktur spasial lokal seperti pada CNN.

ViT memiliki keunggulan dalam menangani gambar beresolusi tinggi dan dataset berukuran besar karena arsitekturnya yang skalabel. Namun, kelemahan dari Vision Transformer terletak pada tingginya kebutuhan komputasi dan memori, terutama karena kompleksitas perhitungan *self-attention* yang meningkat seiring bertambahnya jumlah patch (Sacadibrata dkk., 2025). Meskipun demikian, ViT dinilai efektif dalam mengenali pola visual secara menyeluruh, serta mampu mencapai performa yang kompetitif dibandingkan model CNN dalam berbagai tugas pengenalan citra. Berbeda dengan CNN dan RNN yang mengandalkan konvolusi dan urutan waktu, arsitektur transformer tidak memerlukan kedua proses tersebut. Sebaliknya, ViT menggunakan *embedding* dari patch citra yang dikombinasikan dengan posisi dan token kelas untuk membentuk representasi global dari citra secara keseluruhan (Pokhrel, 2024). Hal ini menjadikan Vision Transformer sebagai pendekatan baru yang inovatif dalam pengolahan citra berbasis deep learning.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif yang berfokus pada pengembangan sistem klasifikasi otomatis penyakit mata berbasis citra fundus dengan menerapkan arsitektur *Vision Transformer* (ViT). Metode ini dipilih karena mampu mengolah data citra secara efisien dan mendalam melalui mekanisme *self-attention*, serta memungkinkan evaluasi kinerja model secara objektif menggunakan metrik kuantitatif seperti akurasi, presisi, recall, dan F1-score. Proses penelitian mencakup tahapan perancangan model deep learning, pelatihan menggunakan dataset citra fundus, pengujian model, serta analisis hasil untuk mengukur keakuratan dalam mengklasifikasikan jenis penyakit mata. Dengan pendekatan ini, penelitian diharapkan menghasilkan sistem klasifikasi yang tidak hanya akurat namun juga dapat diandalkan sebagai solusi pendukung dalam diagnosis dini penyakit mata.

Penelitian ini dilaksanakan melalui serangkaian tahapan yang dirancang secara sistematis dan terstruktur untuk mencapai tujuan penelitian secara tepat dan efisien. Setiap tahapan disusun sebagai panduan pelaksanaan penelitian agar proses yang dilakukan menjadi lebih terarah, mulai dari perencanaan awal hingga evaluasi akhir terhadap model klasifikasi yang dibangun. Tahapan pertama dimulai dengan identifikasi masalah dan perumusan tujuan penelitian, yang dilanjutkan dengan studi literatur guna memperoleh landasan teoritis yang relevan. Selanjutnya, dilakukan proses pengumpulan data dengan mengunduh dataset citra fundus dari sumber terbuka (Kaggle) yang telah diberi label sesuai dengan jenis penyakit mata, seperti mata normal, katarak, glaukoma, dan diabetic retinopathy. Setelah data diperoleh, dilanjutkan dengan tahapan preprocessing data, yang meliputi penyesuaian ukuran gambar, normalisasi, dan augmentasi data jika diperlukan. Tahap ini bertujuan untuk memastikan bahwa data siap digunakan dalam proses pelatihan model. Berikutnya adalah tahapan inti, yaitu pembangunan model klasifikasi menggunakan arsitektur Vision Transformer (ViT). Pada tahap ini, model dilatih menggunakan data latih yang telah diproses, kemudian dilakukan pengujian model dengan data uji untuk mengevaluasi performa menggunakan metrik seperti akurasi, presisi, recall, dan F1-score. Tahapan terakhir adalah analisis hasil dan evaluasi kinerja model. Peneliti akan menganalisis efektivitas model dalam mengklasifikasikan

berbagai jenis penyakit mata berdasarkan citra fundus. Hasil dari evaluasi ini menjadi dasar dalam menarik kesimpulan dan memberikan rekomendasi untuk pengembangan penelitian lebih lanjut.

Metode pengumpulan data dalam penelitian ini dilakukan melalui teknik dokumentasi gambar, yaitu dengan memanfaatkan dataset citra medis fundus mata yang tersedia secara publik di platform Kaggle. Dataset tersebut terdiri dari gambar retina (fundus) yang telah diberi label berdasarkan kategori penyakit mata, yakni normal, katarak, glaukoma, dan diabetic retinopathy. Data yang digunakan berbentuk file digital dengan format JPG atau PNG dan memiliki resolusi yang bervariasi. Dalam pemilihannya, data harus memenuhi kriteria tertentu, seperti legalitas penggunaan, kualitas gambar yang baik, serta label yang jelas dan akurat mencerminkan jenis penyakit mata. Dataset yang telah dikumpulkan disimpan dalam Google Drive dan diintegrasikan dengan Google Colab untuk proses pelatihan model menggunakan bantuan GPU. Sebelum digunakan dalam pelatihan dan pengujian model Vision Transformer, seluruh gambar terlebih dahulu diproses dalam tahap *preprocessing* untuk memastikan keseragaman dan kualitas data. Dengan metode pengumpulan data ini, peneliti dapat mengakses data secara efisien dan legal guna mendukung pengembangan sistem klasifikasi penyakit mata berbasis citra fundus.

Teknik analisis data dalam penelitian ini dilakukan melalui proses evaluasi kuantitatif terhadap performa model klasifikasi berbasis arsitektur Vision Transformer (ViT). Analisis dilakukan dengan membandingkan hasil prediksi model terhadap label sebenarnya untuk menilai efektivitas model dalam mengidentifikasi berbagai jenis penyakit mata berdasarkan citra fundus. Evaluasi performa model menggunakan beberapa metrik standar klasifikasi, yaitu akurasi, presisi, recall, dan F1-score, yang dihitung secara numerik. Akurasi digunakan untuk mengukur proporsi prediksi yang benar secara keseluruhan, sementara presisi dan recall digunakan untuk mengevaluasi ketepatan dan sensitivitas model dalam mengklasifikasikan masing-masing kelas penyakit. F1-score, sebagai rata-rata harmonis dari presisi dan recall, memberikan gambaran menyeluruh tentang keseimbangan performa model. Selain evaluasi numerik, dilakukan juga visualisasi hasil dalam bentuk confusion matrix, yang memberikan informasi detail mengenai jumlah prediksi benar dan salah untuk setiap kategori penyakit mata. Grafik tambahan juga disajikan untuk memperjelas distribusi hasil klasifikasi dan tingkat performa model pada tiap label. Melalui teknik analisis ini, peneliti dapat menafsirkan sejauh mana model Vision Transformer mampu mengklasifikasikan penyakit mata dengan akurat, serta mengidentifikasi potensi perbaikan pada model di masa depan.

HASIL DAN PEMBAHASAN

Dataset

Dalam penelitian ini, proses pelatihan dan pengujian model dilakukan menggunakan platform *Google Colaboratory (Colab).* Pemilihan *Colab* didasarkan pada kemampuannya menyediakan akses GPU/TPU secara gratis serta integrasi yang mudah dengan Google Drive sebagai media penyimpanan dataset. Dataset citra fundus retina yang digunakan dalam penelitian ini disimpan di *Google Drive* dan diakses langsung dari lingkungan *Colab* melalui potongan kode tertentu. Setelah kode dijalankan, pengguna akan diminta memberikan otorisasi akses ke akun *Google* untuk memungkinkan sinkronisasi antara *Google Drive* dan *Colab*. Hal ini mempermudah proses pelatihan karena pengguna tidak perlu mengunggah ulang dataset setiap kali runtime dimulai ulang.

Dataset citra fundus yang digunakan diperoleh dari sumber terbuka yaitu platform Kaggle. Dataset ini terdiri dari 4.028 citra yang terbagi ke dalam empat kelas kategori penyakit mata, yaitu normal, cataract, glaucoma, dan diabetic retinopathy. Masing-masing kelas memiliki jumlah data yang seimbang, yaitu sebanyak 1.007 citra, sehingga mendukung pelatihan model yang adil dan bebas bias kelas. Untuk keperluan pelatihan model, dataset ini dibagi menjadi tiga bagian utama, yaitu 70% untuk data training (sebanyak 2.816 gambar), 15% untuk data validation (604 gambar), dan 15% untuk data testing (608 gambar). Pembagian ini dilakukan secara proporsional dan merata untuk setiap kelas, guna memastikan bahwa setiap kategori penyakit mata terwakili secara seimbang dalam seluruh proses pelatihan, validasi, dan pengujian. Dataset ini menjadi fondasi

utama dalam mengembangkan dan mengevaluasi performa model klasifikasi penyakit mata menggunakan arsitektur Vision Transformer.

Preprocessing dan Augmentasi Data

Preprocessing data merupakan tahap awal yang dilakukan untuk menyesuaikan format, ukuran, serta distribusi nilai piksel agar sesuai dengan kebutuhan input dari model Vision Transformer. Proses ini bertujuan untuk menstandarkan dimensi dan format citra fundus sehingga dapat diproses secara optimal oleh jaringan saraf tiruan. Tahapan preprocessing dilakukan dengan mengubah ukuran seluruh citra menjadi 224x224 piksel menggunakan metode resize, yang merupakan standar input pada arsitektur Vision Transformer. Setelah itu, citra dikonversi menjadi format tensor menggunakan fungsi ToTensor(), yang sekaligus melakukan normalisasi nilai piksel ke dalam rentang [0,1]. Proses ini penting untuk memastikan bahwa data dapat diolah secara efisien dan konsisten oleh model.

Setelah konversi ke tensor, dilakukan proses normalisasi menggunakan nilai mean dan standar deviasi dari dataset ImageNet, yaitu: Mean = [0.485, 0.456, 0.406], Std = [0.229, 0.224, 0.225]

Normalisasi ini bertujuan untuk menyesuaikan distribusi data masukan agar sesuai dengan data yang digunakan pada saat pelatihan awal model (pretraining), khususnya ketika menggunakan bobot pretrained. Hal ini menjadi krusial karena distribusi input yang tidak selaras dapat menurunkan performa model secara signifikan. Selain preprocessing, dilakukan pula proses augmentasi data yang bertujuan untuk memperkaya variasi citra dalam data latih. Teknik augmentasi yang digunakan meliputi Random Horizontal Flip, yaitu pembalikan citra secara horizontal secara acak dan Random Rotation, yaitu rotasi citra secara acak hingga 15 derajat.

Augmentasi ini dirancang untuk meningkatkan kemampuan generalisasi model terhadap variasi kondisi nyata, seperti perbedaan orientasi mata atau sudut pengambilan gambar. Seluruh transformasi preprocessing dan augmentasi ini disusun menggunakan fungsi transforms.Compose dan diterapkan langsung pada proses pemuatan data melalui DataLoader. Ilustrasi potongan kode preprocessing dan augmentasi data ditampilkan berikut:

```
[ ]  # Get automatic transforms from pretrained ViT weights
     pretrained_vit_transforms = pretrained_vit_weights.transforms()
     print(pretrained_vit_transforms)
```

```
ImageClassification(
    crop_size=[224]
    resize_size=[256]
    mean=[0.485, 0.456, 0.406]
    std=[0.229, 0.224, 0.225]
    interpolation=InterpolationMode.BILINEAR
)
```

Gambar 1. Potongan Kode Preprocessing dan Augmentasi Data

Dengan menerapkan strategi preprocessing dan augmentasi ini, model tidak hanya belajar dari citra asli, tetapi juga dari variasi hasil augmentasi. Hal ini membuat model lebih tangguh (robust) terhadap noise, perbedaan pencahayaan, dan perubahan orientasi, sehingga mampu mengklasifikasikan berbagai jenis penyakit mata dari citra fundus secara lebih akurat.

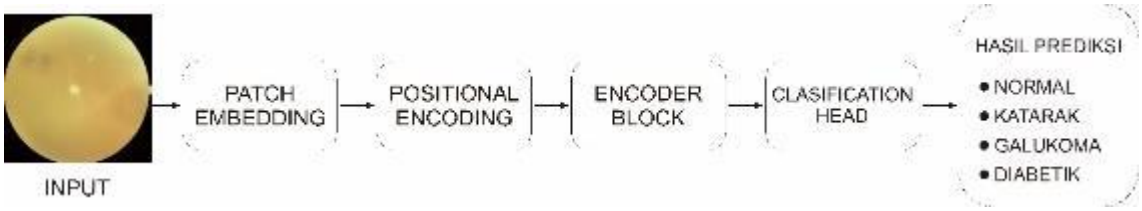
Arsitektur Model Vision Transformer

Alur arsitektur Vision Transformer (ViT) yang digunakan dalam penelitian ini adalah model yang diimplementasikan adalah *ViT-B16 (Vision Transformer Base dengan patch size 16).* Pemilihan model ini didasarkan pada kemampuannya dalam menangkap hubungan spasial secara menyeluruh di seluruh area citra, yang sangat relevan dalam analisis citra fundus mata. Citra fundus memiliki karakteristik kompleks, seperti pembuluh darah, bercak pendarahan, dan diskus optik,

e-ISSN 2828-0490
JURNAL
MEDIA COMPUTER SCIENCE

sehingga membutuhkan model yang mampu memahami konteks visual secara global. Arsitektur ViT-B16 terdiri dari beberapa komponen utama12 lapisan encoder, yang memproses representasi vektor citra secara berurutan melalui mekanisme self-attention, memungkinkan model untuk fokus pada bagian-bagian penting dari citra secara dinamis. 16 attention heads, yang memungkinkan perhatian model terdistribusi ke berbagai aspek atau area dari citra secara paralel Dimensi patch embedding sebesar 768, yaitu dimensi vektor fitur hasil dari proyeksi linear setiap patch citra berukuran 16x16 piksel.

Sebelum melalui tahap transformasi, citra fundus dengan resolusi 224x224 piksel dibagi menjadi 196 patch kecil, masing-masing berukuran 16x16 piksel. Setiap patch tersebut kemudian diproyeksikan ke dalam ruang berdimensi 768 melalui lapisan patch embedding. Pada proses ini juga ditambahkan token CLS (classification token) serta positional encoding untuk mempertahankan informasi urutan dan posisi relatif antar patch, yang penting dalam menjaga konteks spasial.



Gambar 2. Diagram Alur Arsitektur Vision Transformer

Model dilatih menggunakan framework TensorFlow/Keras dengan pengaturan hyperparameter antara lain Optimizer: Adam, Learning Rate: 0.0001 (1e-4), Fungsi Aktivasi (pada layer output): Softmax. Penggunaan fungsi aktivasi softmax pada lapisan keluaran bertujuan untuk menghasilkan nilai probabilitas terhadap empat kelas target, yaitu normal, katarak, glaukoma, dan retinopati diabetik. Arsitektur Vision Transformer ini dirancang untuk mengenali dan membedakan pola-pola visual spesifik pada citra retina secara efisien, sehingga dapat digunakan dalam proses klasifikasi penyakit mata berbasis citra secara akurat dan efektif.

Pembuatan Model Vision Transformer

Setelah tahap preprocessing selesai dilakukan, langkah selanjutnya dalam penelitian ini adalah membangun model klasifikasi citra menggunakan arsitektur *Vision Transformer* (ViT) dengan pendekatan *transfer learning*. Pendekatan ini dipilih untuk memanfaatkan pengetahuan yang telah dipelajari oleh model dari dataset berskala besar seperti ImageNet, guna meningkatkan performa klasifikasi pada dataset yang relatif kecil. Model yang digunakan adalah varian ViT-B/16 dari pustaka torchvision.models, yang telah dilatih sebelumnya menggunakan dataset ImageNet yang terdiri atas lebih dari satu juta gambar dari 1.000 kelas.

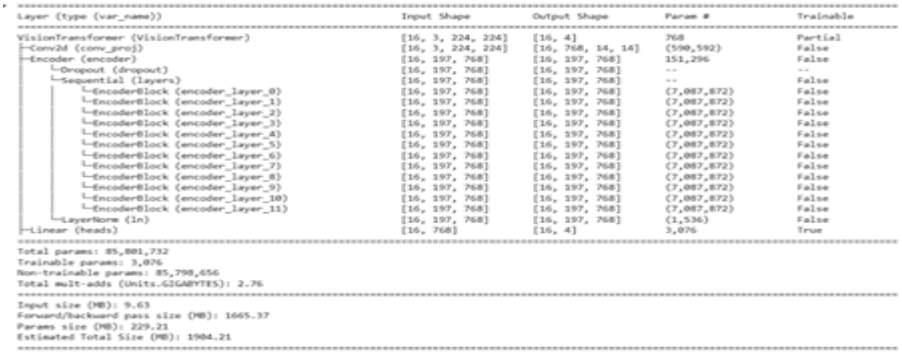
```
# 1. Get pretrained weights for ViT-Base
pretrained_vit_weights = torchvision.models.ViT_B_16_Weights.DEFAULT

# Setup device-agnostic code
device = "cuda" if torch.cuda.is_available() else "cpu"
print(f"Using device: {device}")

# 2. Setup a ViT model instance with pretrained weights
pretrained_vit = torchvision.models.vit_b_16(weights=pretrained_vit_weights).to(device)

# 3. Freeze the base parameters
for parameter in pretrained_vit.parameters():
    parameter.requires_grad = False

class_names = ['cataract','diabetic','glaucoma','normal']

set_seeds()
pretrained_vit.heads = nn.Linear(in_features=768, out_features=len(class_names)).to(device)
```

Gambar 3. Pembuatan Model VIT

Implementasi model ViT-B/16 dilakukan untuk menyelesaikan tugas klasifikasi penyakit mata berbasis citra fundus. Pendekatan transfer learning ini memungkinkan pemanfaatan fitur visual umum yang telah dipelajari oleh model dari jutaan gambar, sehingga mampu memberikan hasil klasifikasi yang baik meskipun jumlah data dalam penelitian ini terbatas. Proses dimulai dengan memuat bobot pralatih (*pre-trained weights*) dari model ViT dan menempatkannya pada perangkat komputasi yang tersedia, baik CPU maupun GPU. Seluruh parameter pada bagian *backbone* atau *encoder* dibekukan (frozen), artinya bobot pada bagian tersebut tidak diperbarui selama pelatihan berlangsung. Strategi ini digunakan untuk mempertahankan kemampuan representasi visual yang telah dipelajari sebelumnya. Hanya bagian akhir model, yaitu *classifier head*, yang dilatih ulang agar dapat menyesuaikan diri dengan jumlah kelas dalam dataset citra fundus, yang terdiri dari empat kategori: normal, katarak, glaukoma, dan diabetic retinopathy.

Modifikasi dilakukan pada layer *classifier* dengan mengganti layer linear terakhir menjadi layer yang memiliki 4 output neuron sesuai jumlah kelas target. Input model berupa citra RGB berukuran 224x224 piksel diubah menjadi 196 patch berukuran 16x16 piksel. Setiap patch kemudian direpresentasikan dalam bentuk *vector embedding* berdimensi 768, yang kemudian digabungkan dengan *classification token* dan *positional embedding*. Proses ini memungkinkan model memahami konteks spasial secara menyeluruh melalui 12 blok encoder transformer yang mengandalkan mekanisme *self-attention*.

Struktur arsitektur model divisualisasikan menggunakan pustaka torchinfo, untuk memastikan bahwa seluruh parameter encoder telah dibekukan dan hanya bagian classifier yang dapat dilatih. Visualisasi ini juga memberikan informasi jumlah parameter, ukuran input dan output, serta struktur jaringan yang digunakan. Hasil ringkasan tersebut menunjukkan bahwa model telah dikonfigurasi secara efisien dan siap untuk proses pelatihan dan evaluasi. Secara keseluruhan, pendekatan transfer learning menggunakan arsitektur Vision Transformer memberikan sejumlah keunggulan, antara lain efisiensi dalam proses pelatihan, kestabilan performa model, dan kemampuan generalisasi yang baik terhadap data citra fundus. Dengan hanya melatih bagian akhir dari model dan tetap memanfaatkan fitur visual yang telah dipelajari sebelumnya, sistem klasifikasi ini mampu menghasilkan performa yang optimal dalam mengidentifikasi berbagai jenis penyakit mata.



Gambar 4. Ringkasan Arsitektur Model Vision Transformer

Untuk memperjelas struktur dari model ViT yang digunakan, berikut adalah penjabaran masing-masing komponen utama dalam arsitektur ViT-B/16:

1. Layer Conv2D (conv_proj)
   Layer ini bertanggung jawab dalam membagi citra input berukuran 224x224 piksel menjadi patch kecil, kemudian mengubahnya menjadi *vector embedding* berdimensi 768 melalui proses konvolusi 2D. Layer ini merupakan bagian dari model pralatih, sehingga tidak dilatih ulang.
2. Encoder
   Encoder terdiri dari beberapa komponen utama, yaitu:
   a. *Dropout* dan *Positional Embedding* yang digunakan untuk menambahkan informasi posisi pada setiap patch dan mengurangi risiko overfitting.

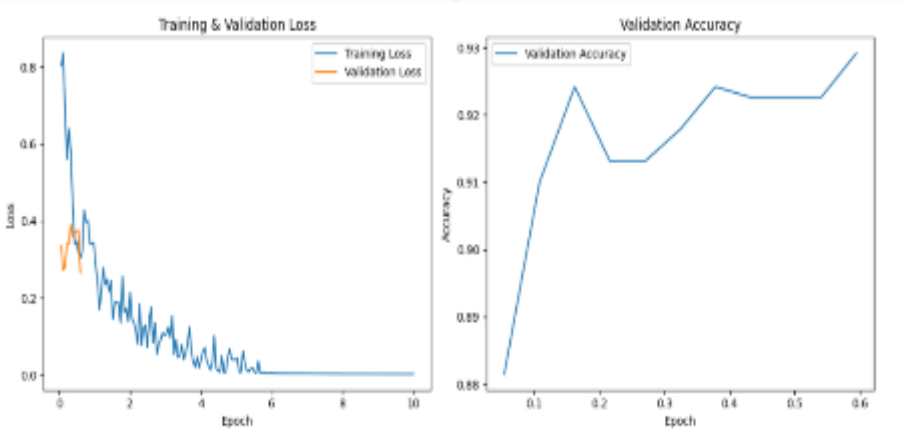
b. 12 buah *Encoder Block* yang berfungsi untuk memproses hubungan spasial dan kontekstual antar patch. Semua blok ini dibekukan selama pelatihan.
c. *Layer Normalization (LayerNorm)* yang dilakukan secara menyeluruh pada output terakhir encoder untuk menstabilkan distribusi data.

3. Linear Layer (head)
Layer ini adalah satu-satunya bagian yang dilatih ulang selama proses pelatihan. Linear layer bertugas melakukan klasifikasi berdasarkan output *CLS token* dari encoder. Modifikasi dilakukan untuk menghasilkan 4 neuron output, masing-masing mewakili kelas target: normal, katarak, glaukoma, dan diabetic retinopathy. Total parameter yang dilatih pada layer ini adalah sebanyak 3.076.

Hasil Pelatihan Model

Proses pelatihan model dilakukan dengan menggunakan arsitektur vision transformer yang diimplementasikan untuk klasifikasi citra fundus retina. Dataset dibagi menjadi tiga bagian yaitu data training, validation, dan testing yang masing-masing bedara dalam folder training, validation, dengan proporsi masing-masing 70%, 15%, 15%, dan testing sebagaimana telah ditunjukan pada struktur direktori sebelumya. Model dilatih selama 15 epoch menggunakan optimizer adam dengan learning rate sebesar 0.001 dan batch size sebesar 16. Untuk mencegah overfitting dan meningkatkan generalisasi, dilakukan teknik augmentasi pada data pelatihan seperti horizontal flip, rotasi acak, dan color jitter. Sementara itu, data validasi dan testing hanya melalui proses normalisasi dan resize ke ukuran 224x224 piksel.

Kurva akurasi dan loss

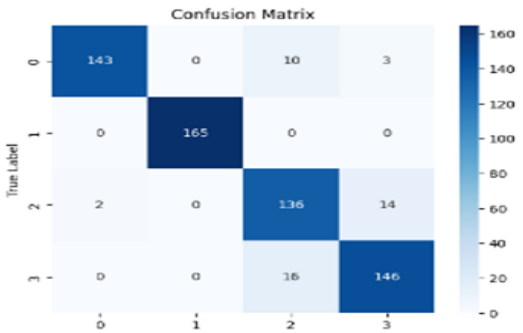


Gambar 5. Kurva Pembelajaran Model

Gambar 5 menunjukan kurva pembelajaran model selama proses pelatihan. Pada grafik sebelah kiri ditampilkan training loss mengalami penurunan yang signifikan dari awal pelatihan hingga mendekati nol, yang menandakan bahwa model berhasil mempelajari pola dari data pelatihan dengan baik. Validation loss juga menunjukan tren penurunan di awal, hal ini mengindikasikan bahwa model tidak mengalami overfitting secara signifikan. Sementara itu, grafik disebelah kanan menunjukan peningkatan akurasi validasi secara konsisten dari sekitar 88% hingga mencapai lebih dari 93%. Peningkatan akurasi ini menunjukan bahwa model memiliki kemampuan generalisasi yang baik terhadap data yang belum pernah dilihat sebelumnya. Secara keseluruhan kurva tersebut menggambarkan proses pelatihan yang stabil dan efektif.

Confusion matrix

Untuk mengetahui performa model pada data uji, dilakukan evaluasi menggunakan confusion matrix berikut :



Gambar 6. Confusion Matrix

Gambar 6 menunjukan gambar confusion matrix. Confusion matrix menampilkan distribusi hasil klasifikasi dari model vision transformer vit-b16 terhadap empat kelas penyakit mata, yaitu katarak, glaucoma, diabetic retinopathy,cdan normal, confusion matrix masing-masing direpresentasikan dengan label 0 hingga 3. Confusion matrix ini berguna untuk mengevaluasi seberapa akurat model dalam memprediksi label sebenarnya dari data uji. Sebagian besar nilai pada matriks terletak pada diagonal utama, yang merupakan indikasi bahwa model berhasil mengklasifikasikan Sebagian besar data dengan benar. Hal ini mencerminkan bahwa model memiliki kemampuan klasifikasi yang tinggi dan stabil. Secara rinci model mampu mengklasifikasikan 143 dari 156 citra berlabel katarak dengan benar, namun masih terdapat 10 citra yang salah diprediksi sebagai glaucoma, dan 3 citra sebagai normal.Sebaliknya, kelas glaucoma menunjukan tantangan yang cukup besar dalam klasifikasi. Dari 152 citra glaukom, hanya 136 citra yang berhasil diklasifikasikan dengan benar, sementara 14 citra salah dikenali sebagai normal, dan 2 citra sebagai katarak. Kesalahan ini menunjukan bahwa ada tumpeng tindih atau kemiripan fitur visual antara kondisi glaucoma dengan kelas lainnya, khususnya pada citra fundus stadium awal yang mungkin belum menunjukan perbedaan signifikan secara visual. Hal yang sama terjadi pada kelas normal, dimana 146 dari 162 citra diklasifikasikan dengan benar, sementara 16 citra lainnya salah diklasifikasikan sebagai glaucoma ini menunjukan bahwa model masih memiliki keterbatasan dalam membedakan dua kondisi yang memiliki kemiripan dalam pola distribusi pembuluh darah dan struktur retina secara umum.

Evaluasi matrix klasifikasi

Setelah model vision transformer (ViT B16) dilatih, dilakukan proses pengujian terhadap data uji untuk mengevaluasi kinerja model dalam mengklasifikasikan citra fundus kedalam empat kelas yaitu : normal, katarak, glaukoma, dan diabetic retinopati. Evaluasi dilakukan menggunakan metrik :

1. Accuracy: untuk mengukur persentase prediksi yang benar dari seluruh prediksi
2. Precision: untuk mengukur ketepatan model dalam memprediksi kelas positif
3. Recall: untuk mengukur sensitivitas model dalam menangkap kelas positif
4. F1-Score: harmonisasi antara precision dan recall

Tabel 1. Laporan Hasil Klasifikasi Report

Kelas	Precision (%)	Recall (%)	F1-Score (%)
Katarak	99.0	92.0	95.0
Diabetic retinopathy	100.0	100.0	100.0
Glaucoma	84.0	89.0	87.0
Normal	90.0	90.0	90.0
Rata-rata	93.3	92.8	93.0

Tabel 1 menunjukan laporan hasil klasfifikasi dari model vision transformer menggunakan vit-b16. Berdasarkan hasil laporan klasifikasi menunjukan performa klasifikasi yang baik pada keempat kelas penyakit mata yaitu katarak, glaucoma, diabetic retinopathy, dan normal. Kelas diabetic

retinopathy mencatatkan hasil yang sempurna dengan nilai precision, recall, dan f1-score masing-masing sebesar 100%, yang menunjukan bahwa model mampu mengenali dan memprediksi seluruh data pada kelas ini tanpa kesalahan. Pada kelas kelas katarak, model meraih nilai precision sebesar 99% dan recall sebesar 92%, menghasilkan f1-score sebesar 95%, yang mencerminkan akurasi tinggi meskipun terdapat sedikit kesalahan prediksi ke kelas lain.

Sementara itu, performa pada kelas glaucoma cenderung sedikit lebih rendah dibandingkan kelas lainnya, dengan precision sebesar 84%, recall sebesar 89%, dan f1-score sebesar 87%. Nilai-nilai ini mengindikasikan bahwa model masih mengalami kesulitan dalam membedakan glaucoma dari kelas lain, khususya dari kelas normal yang memiliki fitur visual yang mirip pada citra fundus. Pada kelas normal, model memperoleh nilai precision dan recall sebesar 90%, dengan f1-score yang juga berada pada angka 90%, menandakan bahwa model mampu mengenali kondisi mata normal dengan akurasi yang baik. Secara keseluruhan, rata-rata performa model berada pada angka 93.3% untuk precision, 92.8% untuk recall, dan 93.0% untuk f1-score, yang menunjukan bahwa model vision transformer memiliki kinerja klasifikasi yang konsisten dankuat pada keempat kelas yang diuji. Tingginya nilai f1-score pada seluruh kelas membuktikan bahwa model tidak hanya akurat dalam prediksi positif, tetapi juga mampu menangkap sebagian besar data actual dari setiap kelas dengan baik. Hasil ini menunjukan bahwa arsitektur vision transformer sangat efektif dalam menangani tugas klasifikasi citra fundus, bahkan pada kondisi yang secara visual dibedakan seperti antara glaucoma dan normal.

Visualisasi Hasil (sudah)

Untuk memberikan Gambaran yang lebih jelas mengenai performa model vison transformer dalam mengklasifikasikan citra fundus, ditampilkan beberapa contoh visualisasi hasil prediksi dari model. Visualisasi ini menyajikan citra fundus yang diinputkan kedalam model, label sebenarnya, hasil prediksi model, serta probabilitas keyakinan (confidence score) dari prediksi tersebut.

Tabel 2. Contoh Visualisasi Hasil Klasifikasi Model ViT

Gambar Citra Fundus	Label Sebenarnya	Prediksi Model	Probabilitas (%)
	Katarak	Katarak	96%
	Glaukoma	Glaukoma	99%
	Diabetik	Diabetik	99%
	Normal	Normal	98%

Visualisasi ini menunjukkan bahwa model memiliki Tingkat kepercayaan yang tinggi terhadap prediksi yang dihasilkan. Probabilitas yang mendekati 100% menunjukkan bahwa model mampu mengklasifikasikan citra dengan keyakinan tinggi dan konsisten, sesuai dengan label ground truth.

Selain itu, klasifikasi yang tepat pada keempat kela (normal, katarak, glaucoma, dan retinopati diabetik) juga memperkuat validitas model dalam menghadapi variasi data nyata. Hal ini menunjukkan bahwa model tidak hanya bekerja pada metrik evaluasi numerik, tetapi juga berfungsi baik secara visual dan interpretative yang penting untuk implementasi di aplikasi nyata.

Implementasi sistem Deteksi Penyakit Mata

Model klasifikasi penyakit mata berbasis Vision Transformer yang telah dilatih kemudian diintegrasikan kedalam sebuah aplikasi berbasis web. Tujuan dari implementasi ini adalah untuk menyediakan sistem deteksi dini penyakit mata yang dapat diakses secara mudah oleh pengguna non-teknis, seperti Masyarakat umum maupun tenaga Kesehatan di daerah dengan keterbatasan akses terhadap fasilitas diagnostik. Aplikasi ini dibangun menggunakan framework Streamlit, yang memungkinkan pengembangan antarmuka web secara cepat, interaktif, dan ringan. Sistem ini dirancang agae dapat menampilkan hasil klasifikasi secra real-time berdasarkan masukan berupa citra fundus retina. Fitur utama dari aplikasi ini meliputi :

1. Unggah gambar citra fundus : pengguna dapat mengunggah file gambar retina dalam format .jpg atau .png
2. Prediksi penyakit mata : Setelah gambar diunggah, sistem akan menampilkan prediksi penyakit mata yang terdeteksi oleh model (normal, katarak, glaucoma, atau retinopati diabetik)
3. Visualisasi probabilitas kelas : Sistem juga menampilkan skor probabilitas dari setiap kelas prediksi dalam bentuk tabel sehingga pengguna dapat memahami seberapa yakin model terhadap hasil klasifikasinya.

Gambar 7. Diagram Alur Website

Terdapat beberapa fitu utama dalam perancangan website klasifikasi penyakit mata, diantaranya:

1. Halaman Awal (Home)
   Halaman awal menyajikan penjelasan mengenai apa itu penyakit mata, contoh gambar penyakit mata, dan tujuan aplikasi. Halaman ini juga dilengkapi dengan tombol “mulai deteksi sekarang” yang mengarahkan pengguna ke tahap selanjutnya.
2. Upload
   Pada halaman ini, pengguna dapat mengunggah gambar fundus dalam format JPG, JPEG, atau PNG yang ingin di identifikasi. Setelah gambar dipilih, secara otomatis akan langsung memunculkan hasil identifikasi.
3. Hasil Klasifikasi Penyakit Mata
   Sistem akan menampilkan hasil klasifikasi dalam bentuk teks yang menyebutkan nama penyakit beserta nilai kepercayaan (confidance) dalam presentasi. Selain itu, sistem juga menyajikan visualisasi dalam bentuk grafik batang horizontal untuk menunjukan probabilitas masing-masing kelas.

e-ISSN 2828-0490
JURNAL MEDIA COMPUTER SCIENCE

Aplikasi deteksi penyakit mata berbasis citra fundus ini memberikan kemudahan bagi para pengguna untuk mengidentifikasi kondisi mata secara cepat dan mandiri.dengan fitur unggah, hasil klasifikasi otomatis, dan tampilan visual beruma nilai kepercayaan dan grafik probabilitas, aplikasi ini efektif untuk membantu dalam mendeteksi penyakit mata seperti katarak, glaucoma, diabetic retinopathy. Seluruh proses klasifikasi dilakukan menggunakan arsitektur vision transformer yang mampu mengekstraksi fitur visual secara mendalam dan akurat, sehingga meningkatkan performa deteksi pada citra fundus dengan tingkat presisi yang tinggi. Adapun teknologi yang digunakan dalam pengembangan website ini yaitu:

1. Bahasa pemrograman : Python
2. Framework : Streamlit
3. Model klasifikasi : Vision transformer (ViT-B16, Tensorflow/keras)
4. Lingkungan deployment : Lokal (localhost) dan dapat dikembangkan lebih lanjut untuk cloud.

Untuk memberikan gambaran lebih jelas mengenai implementasi sistem, berikut ditampilkan antarmuka dari website klasifikasi penyakit mata:

Halaman awal website yang menampilkan penjelasan mengenai penyakit mata

Gambar 8. Halaman Awal Website

Pada halaman upload atau deteksi ini, pengguna dapat mengunggah gambar citra fundus yang ingin dilakukan klasifikasi, setelah gambar dipilih lalu secara otomatis akan memunculkan hasil klasifikasi.

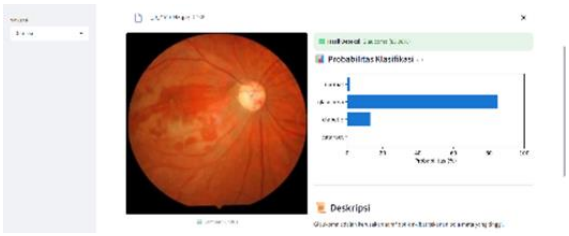
Gambar 9. Halaman Upload

Pada tampilan hasil menampilkan hasil klasifikasi berupa nama penyakit mata nilai kepercayaan, dan juga menunjukan probabilitasnya.

Gambar 10. Halaman Hasil Klasifikasi Penyakit Mata Katarak

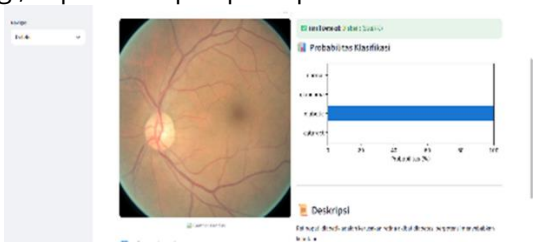
Gambar diatas menunjukan hasil deteksi klasifikasi penyakit mata berdasarkan analisis model, citra tersebut terklasifikasi sebagai katarak dengan tingkat probabilitas sebesar 96,37%, yang

berarti model sangat yakin terhadap prediksi tersebut. Citra tampak buram, sesuai dengan ciri khas katarak, yaitu kekeruhan lensa mata yang menyebabkan penglihatan kabur.



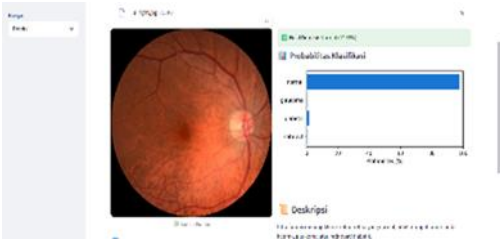
Gambar 11. Halaman Hasil Klasifikasi Penyakit Mata Glaukoma

Gambar menunjukan hasil klasifikasi citra fundus mata yang terdeteksi sebagai glaucoma dengan probabilitas 85,08%. Grafik menunjukan bahwa kelas glaucoma memiliki probabilitas tertinggi dibandingkan kelas lainnya. Glaucoma ditandai dengan kerusakan saraf optic akibat tekanan bola mata yang tinggi, seperti tampak pada perubahan struktur saraf optic pada citra.



Gambar 12. Halaman Hasil Klasifikasi Penyakit Mata Diabetic Retinopathy

Gambar menunjukan hasil klasifikasi citra fundus mata yang terdeteksi sebagai diabetic retinopathy dengan probabilitas sangat tinggi, yaitu 99,87%. Grafik klasifikasi memperlihatkan dominasi penuh pada kelas diabetic, sementara lainnya tidak signifikan. Diabetic retinopathy ditandai dengan kerusakan pada pembuluh darah retina akibat diabetes, yang dapat dilihat dari bercak atau perubahan pola pembuluh darah pada citra.



Gambar 13. Halaman Hasil Klasifikasi Mata Normal

Gambar menunjukan hasil klasifikasi citra fundus mata yang terdeteksi sebagai normal dengan probabilitas 97,68%. Grafik klasifikasi menunjukan dominasi kelas normal, sementara kelas penyakit mata lainnya rendah. Citra fundus tampak memiliki struktur retina yang sehat tanpa adanya indikasi katarak, glaucoma, maupun diabetic retinopathy.Berdasarkan hasil implementasi model vision transformer (ViT) terbukti sangat efektif dalam mengklasifikasikan penyakit mata berdasarkan citra fundus retina. Model ini mampu memberikan hasil prediksi dengan akurasi tinggi dan performa yang stabil disemua kelas penyakit. Jika dibandingkan dengan model CNN (convolutional neural network) yang banyak digunakan sebelumnya, vision transformer memiliki beberapa keunggulan :

1. Akurasi lebih tinggi dari CNN pada data yang sama
2. Mampu menangkap pola global dalam citra tanpa perlu pooling, sehingga bisa memahami struktur retina secara menyeluruh

3. Lebih tahan terhadap variasi warna dan pencahayaan, yang sering terjadi pada citra fundus

Namun arsitektur vision transformer juga memiliki kekurangan yaitu membutuhkan komputasi yang besar. Proses pelatihan dan prediksi membutuhkan perangkat keras (GPU) yang cukup kuat, sehingga kurang cocok untuk diterapkan langsung di perangkat dengan sumber daya terbatas seperti hp atau sistem real-time tanpa optimasi khusus. Oleh karena itu, meskipun arsitektur vision transformer sangat menjanjikan untuk klasifikasi penyakit mata, ke depannya perlu dilakukan pengembangan lebih lanjut agar model dapat berjalan lebih ringan dan efisien, misalnya dengan menggunakan varian transformer yang lebih ringan atau teknik optimasi model. Secara keseluruhan, arsitektur vision transformer dapat menjadikan solusi alternatif yang lebih cerdas dan akurat dibandingkan CNN, dan memiliki potensi besar untuk diterapkan dalam sistem pendukung diagnosis penyakit mata di masa depan.

KESIMPULAN DAN SARAN

Kesimpulan

Penelitian ini berhasil mengembangkan sistem klasifikasi penyakit mata berbasis citra fundus menggunakan arsitektur Vision Transformer (ViT). Model yang dibangun mampu mengklasifikasikan empat jenis kondisi mata, yaitu normal, katarak, glaucoma, dan retinopati diabetik dengan tingkat akurasi yang tinggi. Evaluasi performa menunjukkan hasil yang stabil di seluruh kelas dengan prediksi yang sangat baik. Visualisasi hasil klasifikasi menguatkan bahwa model mampu mengenali karakteristik masing-masing penyakit dengan tingkat keyakinan tinggi. Dibandingkan dengan pendekatan konvensional seperti CNN, vision Transformer menunjukkan keunggulan dalam memahami konteks visual secara menyeluruh. Meskipun demikian, model ini memerlukan sumber daya komputasi yang besar, sehingga perlu diperhatikan dalam penerapan di sistem real-time atau perangkat terbatas.

Saran

Untuk pengembangan selanjutnya, disarankan agar dilakukan optimasi model agar lebih ringan dan efisien, misalnya dengan menggunakan varian Vision Transformer yang lebih kecil atau menerapkan teknik kompresi model. Pengujian juga dapat diperluas dengan menggunakan dataset yang lebih kompleks dan bervariasi agar model lebih robust terhadap kondisi nyata. Pengembangan antarmuka sistem berbasis web atau mobile juga sangat penting agar sistem ini dapat di akses lebih mudah oleh masyarakat umum maupun tenaga medis. Selain itu, integrasi sistem dengan perangkat keras berdaya rendah seperti Raspberry Pi atau smartphone dapat menjadi alternatif penerapan dalam sistem deteksi dini berbasis edge computing. Terakhir, validasi hasil klasifikasi perlu melibatkan tenaga medis professional untuk memastikan kesesuaian prediksi sistem dengan diagnosis klinis sebenarnya.

DAFTAR PUSTAKA

Adjeng, A. N. T., Himayani, R., Graharti, R., Adrifianie, F., & Oktoba, Z. (2024). Deteksi Dini Ulkus Kornea yang Mengancam Penglihatan dan Menurunkan Kualitas Hidup Masyarakat Pekon Kedaung Pringsewu. *Jurnal Kreativitas Pengabdian Kepada Masyarakat (PKM)*, *7*(8), Article 8.

Bintang, Y. K., & Imaduddin, H. (2024). Pengembangan model deep learning untuk deteksi retinopati diabetik menggunakan metode transfer learning. *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, *9*(3), 1442–1455.

Budiarti, I. S. (2023). *Indra Penglihatan; Mata*. Bumi Aksara.

Chandana, M. H., Dorasanamma, G., Kiran, S., & Kumar, A. A. (2024). A Review on Feature Extraction Techniques using Machine Learning. *Macaw International Journal of Advanced Research in Computer Science and Engineering*, *10*(1), Article 1.

Dana, M. M. (2020). Gangguan Penglihatan Akibat Kelainan Refraksi yang Tidak Dikoreksi. *Jurnal Ilmiah Kesehatan Sandi Husada*, *9*(2), Article 2. https://doi.org/10.35816/jiskh.v12i2.451

Dhamayanti, R., Rohmah, M. F., & Zahara, S. (2021). Penggunaan Deep Learning dengan Metode Convolutional Neural Network Untuk Klasifikasi Kualitas Sayur Kol Berdasarkan Citra Fisik. *SUBMIT: Jurnal Ilmiah Teknologi Infomasi Dan Sains*, *1*(1), Article 1.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., & Gelly, S. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Hadiprakoso, R. B., & Buana, I. K. S. (2021). Performance comparison of feature extraction and machine learning classification algorithms for face recognition. *The IJICS (International Journal of Informatics and Computer Science)*, *5*(3), 250.

Hutagalung, E. F. S., & Sitompul, P. (2023). Implementasi Deep Learning Menggunakan Metode Cnn Untuk Klasifikasi Jenis Ulos Batak Toba. *Student Scientific Creativity Journal*, *1*(4), 01–19.

Jamil, J., & Pulukadang, S. (2025). Application of Deep Learning Method in Learning. *Formosa Journal of Sustainable Research*, *4*(6), Article 6. https://doi.org/10.55927/fjsr.v4i6.308

Jatmoko, C., & Lestiawan, H. (2024). Prediksi Penyakit Mata Menggunakan Convolutional Neural Network. *Semnas Ristek (Seminar Nasional Riset Dan Inovasi Teknologi)*, *8*(01), Article 01.

Muhlashin, M. N. I., & Stefanie, A. (2023). Klasifikasi Penyakit Mata Berdasarkan Citra Fundus Menggunakan YOLO V8. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(2), Article 2.

Nugraha, S. N., Pebrianto, R., & Fitri, E. (2023). Penerapan Deep Learning Pada Klasifikasi Tanaman Paprika Berdasarkan Citra Daun Menggunakan Metode CNN. *INFORMATION SYSTEM FOR EDUCATORS AND PROFESSIONALS : Journal of Information System*, *8*(2), 133–142.

Nurhakiki, J., & Yahfizham, Y. (2024). Studi Kepustakaan: Pengenalan 4 Algoritma Pada Pembelajaran Deep Learning Beserta Implikasinya. *Pendekar : Jurnal Pendidikan Berkarakter*, *2*(1), 270–281.

Pokhrel, S. (2024). Digital Technologies in Physics Education: Exploring Practices and Challenges. *Teacher Education Advancement Network Journal*, *15*(1), Article 1.

Putri, C. A., & Rakasiwi, S. (2025). Diagnosis Dini Penyakit Mata: Klasifikasi Citra Fundus Retina dengan Convolutional Neural Network VGG-16. *Edumatic: Jurnal Pendidikan Informatika*, *9*(1), 208–216.

Retina, C. F. (2022). *Klasifikasi Penyakit Mata Berdasarkan Citra Fundus Retina Menggunakan Dimensi Fraktal Box Counting Dan Fuzzy K-Means*.

Rifqi, M. (2024). *Klasifikasi Penyakit pada Daun Kopi Menggunakan Metode Vision Transformer (Vit)*. 9

Sacadibrata, S., Rahman, T., & Anggai, S. (2025). Perbandingan Convolutional Neural Network dan Vision Transformer Untuk Klasifikasi Penyakit Daun Pada Tomat. *PROKASDADIK:Prosiding Kecerdasan Artifisial, Sains Data,dan Pendidikan Masa Depan*, *3*.

Sarker, I. H., Kayes, A. S. M., & Watters, P. (2019). Effectiveness analysis of machine learning classification models for predicting personalized context-aware smartphone usage. *Journal of Big Data*, *6*(1), 57. https://doi.org/10.1186/s40537-019-0219-y

Sepe, F. Y. & Stefanus Stanis. (2023). *Buku Ajar Anatomi Fisiologi Manusia*. Zahir Publishing.

Tamba, M. (2024). Analisis dan Implementasi Teknologi Deep Learning dalam Pengolahan Citra Digital. *Circle Archive*, *1*(6), Article 6. https://circle-archive.com/index.php/carc/article/view/290

Wirawan, N. A. (2024). *Angka Kebutaan dan Gangguan Penglihatan di Indonesia Tembus 8 Juta Kasus*. GoodStats Data.

Wu, J., Hu, R., Xiao, Z., Chen, J., & Liu, J. (2021). Vision Transformer-based recognition of diabetic retinopathy grade. *Medical Physics*, *48*(12), 7850–7863. https://doi.org/10.1002/mp.15312

Yang, Y., Cai, Z., Qiu, S., & Xu, P. (2024). Vision transformer with masked autoencoders for referable diabetic retinopathy classification based on large-size retina image. *PLOS ONE*, *19*(3), e0299265. https://doi.org/10.1371/journal.pone.0299265