

**KLASIFIKASI CITRA RETINA
DALAM MENENTUKAN PENYAKIT GLAUKOMA
MENGUNAKAN ARSITEKTUR *VISION TRANSFORMER***

SKRIPSI

**Sebagai Salah Satu Syarat untuk Memperoleh Gelar
Sarjana Sains Bidang Studi Matematika**

Oleh:

Dewi Lestari Dwi Putri

NIM 08011281823103



**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS SRIWIJAYA**

2022

LEMBAR PENGESAHAN

**KLASIFIKASI CITRA RETINA
DALAM MENENTUKAN PENYAKIT GLAUKOMA
MENGUNAKAN ARSITEKTUR *VISION TRANSFORMER***

SKRIPSI

**Sebagai Salah Satu Syarat untuk Memperoleh Gelar
Sarjana Sains Bidang Studi Matematika**

Oleh:

DEWI LESTARI DWI PUTRI

NIM 08011281823103

Pembimbing Kedua



Dr. Yuli Andriani, S.Si., M.Si
NIP. 197207021999032001

**Indralaya, Agustus 2022
Pembimbing Utama**



Anita Desiani, S.Si., M.Kom
NIP. 197712112003122002

**Mengetahui,
Ketua Jurusan Matematika**



Drs. Sugandi Yahdin, M.M
NIP. 195807271986031003

PERNYATAAN KEASLIAN KARYA ILMIAH

Yang bertanda tangan dibawah ini:

Nama Mahasiswa : Dewi Lestari Dwi Putri
NIM : 08011281823103
Fakultas/Jurusan : Matematika dan Ilmu Pengetahuan Alam/Matematika

Menyatakan bahwa skripsi ini adalah hasil karya saya sendiri dan karya ilmiah ini belum pernah diajukan sebagai pemenuhan persyaratan untuk memperoleh gelar kesarjanaan strata satu (S1) dari Universitas Sriwijaya maupun perguruan tinggi lain.

Semua informasi yang dimuat dalam skripsi ini yang berasal dari penulis lain baik yang dipublikasikan atau tidak telah diberikan penghargaan dengan mengutip nama sumber penulis secara benar. Semua isi dari skripsi ini sepenuhnya menjadi tanggung jawab saya sebagai penulis.

Demikianlah surat pernyataan ini saya buat dengan sebenarnya.

Indralaya, 05 Agustus 2022

Penulis



Dewi Lestari Dwi Putri

NIM. 08011281823103

**HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK
KEPENTINGAN AKADEMIS**

Sebagai Civitas Akademik Universitas Sriwijaya, yang bertanda tangan di bawah ini:

Nama Mahasiswa : Dewi Lestari Dwi Putri
NIM : 08011281823103
Fakultas/Jurusan : Matematika dan Ilmu Pengetahuan Alam/Matematika
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada Universitas Sriwijaya “Klasifikasi Citra Retina dalam Menentukan Penyakit Glaukoma Menggunakan Arsitektur Vision Transformer”. Dengan hak bebas royalti non-eksklusif ini Universitas Sriwijaya berhak menyimpan, mengalih, edit/memformat, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir atau skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta. Demikian pernyataan ini saya buat dengan sesungguhnya.

Indralaya, 05 Agustus 2022

Penulis



Dewi Lestari Dwi Putri

NIM. 08011281823103

HALAMAN PERSEMBAHAN

Kupersembahkan skripsi ini untuk:

Yang Maha Kuasa Allah Subhanahu Wa Ta'ala

Orang tuaku tersayang

Saudara-saudaraku tercinta

Keluarga besarku,

Semua guru dan dosenku,

Sahabat-sahabatku,

Almamaterku

Motto

“Apa yang menjadi milikmu, tidak akan pernah melewatkan mu, dan apa yang tidak ditakdirkan untuk mu, tidak akan pernah kamu miliki”

KATA PENGANTAR

Puji syukur kehadiran Allah Subhanahu Wa Ta'ala yang telah memberikan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Klasifikasi Citra Retina dalam Menentukan Penyakit Glaukoma Menggunakan Arsitektur *Vision Transformer*” sebagai salah satu syarat memperoleh gelar sarjana sains bidang studi Matematika di Fakultas MIPA Universitas Sriwijaya.

Penulis menyadari bahwa proses pembuatan skripsi ini merupakan proses pembelajaran yang sangat berharga serta tak lepas dari kekurangan dan keterbatasan. Dengan segala hormat dan kerendahan hati, penulis mengucapkan terima kasih dan penghargaan kepada:

1. Orang tuaku tercinta, Abiku Mas Amin yang tidak pernah berhenti berjuang dan memberikan yang terbaik untukku sebagai putrinya, Bunda tiriku Sahriah yang sabar atas sikapku, terkhusus Almarhumah Amiku Nasriyanti yang sudah melahirkan dan memberikan do'a serta pengajaran terbaik kepadaku selama masa hidupnya. Terima kasih karena tak pernah lelah mendidik, menasehati, membimbing, mendukung dan terus mendo'akan anaknya. Semoga semua yang telah dilakukan menjadi amal ibadah.
2. Bapak Drs Sugandi Yahdin, M.M selaku Ketua Jurusan Matematika FMIPA Universitas Sriwijaya yang telah memberikan arahan dan motivasi kepada penulis selama proses perkuliahan dan Ibu Dr. Dian Cahyawati Sukanda, M.Si selaku Sekretaris Jurusan Matematika FMIPA Universitas Sriwijaya yang telah mengarahkan urusan akademik kepada penulis.

3. Ibu Anita Desiani, S.Si, M.Kom selaku dosen pembimbing utama yang telah bersedia meluangkan waktu, tenaga, pikiran untuk memberikan bimbingan, pengarahan dan didikan berharga selama proses pembuatan skripsi, kompetisi atau program mahasiswa, dan perjalanan perkuliahan ini.
4. Ibu Dr. Yuli Andriani, M.Si selaku dosen pembimbing pendamping yang telah bersedia meluangkan waktu, tenaga, pikiran untuk memberikan bimbingan, pengarahan dan didikan berharga selama proses pembuatan skripsi dan perjalanan perkuliahan ini.
5. Ibu Irmeilyana, M.Si dan Bapak Drs. Endro Setyo Cahyono, M.Si selaku dosen pembahas dan penguji yang telah memberikan tanggapan, kritik serta saran yang sangat bermanfaat untuk perbaikan dan penyelesaian skripsi ini.
6. Ibu Sisca Octarina, S.Si., M.Sc selaku sekretaris tim pelaksana tugas akhir penulis dan Ibu Eka Susanti, M.Sc selaku ketua tim pelaksana tugas akhir dan dosen pembimbing akademik yang telah membimbing dan mengarahkan urusan akademik selama masa perkuliahan penulis.
7. Seluruh Dosen di Jurusan Matematika FMIPA yang telah memberikan ilmu, nasihat, serta bimbingan selama proses perkuliahan dan seluruh guru yang telah memberikan ilmu yang sangat bermanfaat hingga mengantarkan penulis pada pendidikan ini.
8. Pak Irwansyah selaku admin dan Ibu Hamidah selaku pegawai tata usaha Jurusan Matematika FMIPA yang telah membantu penulis selama perkuliahan.
9. Abangku tersayang Kamaludin serta adik-adikku tercinta Ilham Anugrah Tri Putra, Inayah Inas Sakinah, dan Yasmina Humaira alias Rania.

10. Semua sahabat seperjuangan selama masa perkuliahan dan skripsi ini, Duck's Fams, Tydack Cawa, *ViT* Teams, teman-teman komputasi dan teman dari masa SMA ku. Terima kasih sudah menjadi orang-orang baik dan random yang menjadi penyemangat disekeliling penulis yang selalu mendukung, membantu, dan memberikan semangat kepada penulis serta pelajaran luar biasa dari proses pertemanan kita.
11. Kakak-kakak tingkat angkatan 2016 dan 2017 bidang minat komputasi yang telah banyak membantu serta berbagi ilmu selama proses skripsi.
12. Keluarga Matematika 2018, BPH COIN 2019/2020, BPH Himastik Gelora Karya, rekan-rekan perlombaan selama perkuliahan.
13. Serta semua pihak yang tidak dapat penulis sebutkan satu per satu, semoga segala kebaikan yang diberikan mendapat balasan terbaik dari Allah.
14. *Last but not least, I wanna thank me, I wanna thank me for believing in me, I wanna thank me for doing all this hard work, I wanna thank me for having no days off, I wanna thank me for, for never quitting, I wanna thank me for always being a giver, and tryna give more than I receive, I wanna thank me for tryna do more right than wrong, I wanna thank me for just being me at all times.*

Semoga skripsi ini dapat menambah pengetahuan dan bermanfaat bagi mahasiswa/mahasiswi Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya dan pihak yang memerlukan.

Indralaya, Agustus 2022

Penulis

**CLASSIFICATION OF RETINA IMAGE
IN DETERMINING GLAUCOMA DISEASE
USING THE VISION TRANSFORMER ARCHITECTURE**

By:

**Dewi Lestari Dwi Putri
08011281823103**

ABSTRACT

Glaucoma is the second leading cause of blindness in the world. Early detection of glaucoma needs to be done to reduce the risk of blindness. So, we need a technology to predict and classify glaucoma. Therefore, this study performed retinal image classification to determine glaucoma. Glaucoma can be classified to 3 classes, there are advanced glaucoma, early glaucoma, and normal control. Deep learning is often used to perform various tasks in machine learning such as classification. One of the learning methods in classification is Vision Transformer (ViT). The advantage of ViT is its small memory usage with excellent results in performing classification tasks. In this study, retinal image classification was performed to determine glaucoma disease using the ViT architecture. The research stages are data collection, data pre-processing, training, testing, and performance evaluation as well as analysis and interpretation. The study used retinal image data sourced from the Harvard Dataverse with the performance evaluation values for accuracy, sensitivity, specificity, F1-score, and Cohen's kappa were 96.74%, 95.10%, 97.57%, 95.10%, and 92.46%, respectively. Based on the results obtained, it can be concluded that the Vision Transformer architecture is very good in performing the task of classifying retinal images to determine glaucoma disease.

Keyword: Multiclass classification, evaluation, Vision Transformer.

**KLASIFIKASI CITRA RETINA
DALAM MENENTUKAN PENYAKIT GLAUKOMA
MENGUNAKAN ARSITEKTUR *VISION TRANSFORMER***

Oleh:

**Dewi Lestari Dwi Putri
08011281823103**

ABSTRAK

Glaukoma merupakan penyakit penyebab kebutaan terbesar kedua di dunia. Deteksi dini penyakit glaukoma perlu dilakukan untuk mengurangi risiko kebutaan. Sehingga diperlukan adanya teknologi yang mampu memprediksi serta mengklasifikasikan penyakit glaukoma. Maka dari itu, penelitian ini dilakukan untuk mengklasifikasikan citra retina untuk menentukan penyakit Glaukoma. Glaukoma dapat diklasifikasikan menjadi 3 kelas yaitu *advanced glaucoma* (glaukoma lanjut), *early glaucoma* (glaukoma dini), dan *normal control* (normal). *Deep learning* sering digunakan untuk melakukan berbagai tugas dalam pembelajaran mesin seperti klasifikasi. Salah satu metode untuk melakukan pembelajaran dalam klasifikasi adalah *Vision Transformer (ViT)*. Kelebihan dari *ViT* adalah penggunaan memori yang kecil dengan hasil yang sangat baik dalam melakukan tugas klasifikasi. Pada penelitian ini, dilakukan klasifikasi citra retina untuk menentukan penyakit glaukoma menggunakan arsitektur *ViT*. Tahapan penelitian adalah pengumpulan data, *pre-processing* data, *training*, *testing*, dan evaluasi kinerja serta analisis dan interpretasinya. Penelitian menggunakan data citra retina yang bersumber dari Harvard Dataverse dengan nilai hasil evaluasi kinerja akurasi, sensitivitas, spesifisitas, *F1-score*, dan *cohen's kappa* berturut-turut adalah 96,74%, 95,10%, 97,57%, 95,10%, dan 92,46%. Berdasarkan hasil yang diperoleh, dapat disimpulkan bahwa arsitektur *Vision Transformer* sangat baik dalam melakukan tugas klasifikasi citra retina untuk menentukan penyakit glaukoma.

Kata Kunci: Klasifikasi *multiclass*, evaluasi, *Vision Transformer*.

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN.....	ii
HALAMAN PERSEMBAHAN	iii
KATA PENGANTAR.....	iv
ABSTRACT	vii
ABSTRAK	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Perumusan Masalah	5
1.3. Pembatasan Masalah.....	5
1.4. Tujuan	6
1.5. Manfaat	6
BAB II TINJAUAN PUSTAKA.....	7
2.1. Glaukoma.....	7
2.2. Citra Digital	7
2.3. Klasifikasi Citra	8
2.4. Augmentasi	8
2.5. <i>Vision Transformer</i>	9
2.5.1. Patch.....	10
2.5.2. Position and CLS Embedding.....	11
2.5.3. Transformer Encoder	12
2.5.4. Self Attention	13
2.5.5. Multi Head Attention	14
2.6. <i>Multilayer Perceptron (MLP)</i>	15
2.7. <i>Activation Function</i>	17
2.8. <i>Loss Function: Categorical Cross Entropy</i>	17
2.9. <i>Layer Normalization</i>	18
2.10. <i>Optimization Function: Adaptive Moment Estimation (Adam)</i>	19
2.11. <i>Confusion Matrix</i>	20
BAB III METODOLOGI PENELITIAN	23
3.1. Tempat	23
3.2. Waktu.....	23
3.3. Alat.....	23

3.4. Metode Penelitian	23
BAB IV HASIL DAN PEMBAHASAN	27
4.1. Deskripsi Data.....	27
4.2. Augmentasi Data.....	28
4.3. Arsitektur <i>Vision Transformer</i>	29
4.4. Operasi Manual pada <i>Vision Transformer</i>	32
4.5. Proses <i>Training</i>	55
4.6. Proses <i>Testing</i>	58
4.7. Evaluasi.....	60
4.8. Perbandingan Evaluasi Kinerja.....	67
4.9. Analisis dan Interpretasi Hasil.....	67
BAB V KESIMPULAN DAN SARAN	69
5.1. Kesimpulan	69
5.2. Saran	69
DAFTAR PUSTAKA	70

DAFTAR TABEL

Tabel 2.1 Kategori nilai kinerja arsitektur	22
Tabel 4.1. Data sampel citra retina.....	27
Tabel 4.2. Hasil perhitungan rata-rata μ_j untuk setiap <i>mini batch</i>	40
Tabel 4.3. Hasil perhitungan variansi σ_j^2 untuk setiap <i>mini batch</i>	41
Tabel 4.4. Nilai bobot <i>hidden layer</i> dan <i>output</i>	49
Tabel 4.5. <i>Confusion matrix</i> klasifikasi citra retina	58
Tabel 4.6. Perbandingan hasil penilaian evaluasi kinerja model	67
Tabel 4.7. Perbandingan evaluasi kinerja	67

DAFTAR GAMBAR

Gambar 2.1. Model <i>Vision Transformer</i> (Dosovitskiy <i>et al.</i> , 2021).....	9
Gambar 2.2 Ilustrasi <i>Transformer Encoder</i> (Dosovitskiy <i>et al.</i> , 2021).....	12
Gambar 2.3 Ilustrasi <i>Scaled Dot-Product Attention</i> (Vaswani <i>et al.</i> , 2017).....	13
Gambar 2.4 Ilustrasi <i>Multi-Head Attention</i> (Vaswani <i>et al.</i> , 2017).....	14
Gambar 2.5. Ilustrasi <i>Multilayer Perceptron</i>	16
Gambar 4.1. Ilustrasi Augmentasi.....	29
Gambar 4.2. Komponen Utama Arsitektur <i>Vision Transformer</i>	29
Gambar 4.3. Ilustrasi <i>Patching</i> dan <i>Position Embedding</i>	30
Gambar 4.4. Ilustrasi Pemotongan <i>Patch</i>	32
Gambar 4.5. Ilustrasi Perataan Gambar.....	33
Gambar 4.6. Ilustrasi MLP dengan 2 <i>Hidden Layer</i> dan 5 <i>Input</i>	48
Gambar 4.7. Hasil <i>Training</i> Model Arsitektur <i>Vision Transformer</i>	55
Gambar 4.8. Grafik Akurasi <i>Training</i> Arsitektur <i>Vision Transformer</i>	56
Gambar 4.9. Grafik <i>Loss Training</i> Arsitektur <i>Vision Transformer</i>	57

BAB I

PENDAHULUAN

1.1. Latar Belakang

Deep Learning (DL) merupakan sub bidang dari *Machine Learning* (ML) yang menggunakan jaringan saraf tiruan yang terinspirasi oleh struktur dan fungsi otak manusia (Ertam & Aydin, 2017). Model DL telah banyak dikembangkan dan digunakan dalam pembelajaran *Big Data* (Q. Zhang et al., 2018). Salah satu model pembelajaran DL yang saat ini sedang berkembang dan paling banyak digunakan adalah *Convolutional Neural Network* (CNN). CNN adalah pengembangan dari *Multilayer Perceptron* (MLP) atau jaringan saraf tiruan yang didesain untuk mengolah data dua dimensi dan termasuk dalam jenis *Deep Neural Network* karena kedalaman jaringan yang tinggi dan banyak diaplikasikan pada data citra (Putra et al., 2016). CNN berisi beberapa lapisan yang mengubah inputnya dengan operasi konvolusi secara otomatis (Shabanian et al., 2019).

Lapisan yang dimiliki CNN dan operasi konvolusi yang dilakukan dalam proses pembelajaran menyebabkan CNN membutuhkan banyak parameter dalam proses pelatihannya. Sebuah CNN dapat memiliki ratusan hingga jutaan parameter yang masing-masing belajar mendeteksi berbagai gambar (Ilahiyah & Nilogiri, 2018). Banyaknya parameter yang dibutuhkan dalam membangun model CNN memungkinkan ruang desain untuk sistem memori akselerator CNN besar, mencakup semua kemungkinan kombinasi dan hierarki memori aktivasi, memori bobot, dan memori umum (Siu et al., 2018). CNN memiliki kelemahan yaitu proses pelatihan model yang cukup lama dan banyaknya data yang diperlukan untuk

melakukan pelatihan model. Karena keterbatasan pada CNN ini, beberapa ide muncul untuk mengatasi keterbatasan tersebut salah satunya adalah teknik *patching*. Teknik *patching* ini harus dilakukan secara tepat agar kebutuhan CNN terpenuhi.

Beberapa penelitian yang menggunakan metode CNN dan teknik *patching* diantaranya dilakukan oleh Frid-Adar *et al.* (2017) yang menerapkan arsitektur CNN berbasis *Multi-Class Patch* pada deteksi *liver lesions*. Penelitian tersebut melakukan pemotongan *patch* gambar *input* berukuran 50×50 menjadi 32×32 piksel dengan hasil evaluasi kinerja *True Positive Rate* (TPR) sebesar 86,8%. Kim *et al.* (2020) menerapkan metode CNN berbasis *patch* untuk klasifikasi gambar area jalan. Penelitian ini melakukan pemotongan *image patches* dari gambar input dengan menggunakan pergeseran atau *sliding window* dari kiri atas ke kanan bawah dengan hasil kinerja yang diperoleh adalah akurasi sebesar 0,93 dan *IoU* sebesar 0,75. Penelitian lain tentang teknik *patching* juga dilakukan oleh Misra, Crispim-Junior and Tougne, (2020) yang mengidentifikasi spesies pohon dari gambar kulit kayu menggunakan metode *Patch-Based* CNN dengan memotong gambar berukuran 224×224 dengan cara *sliding window*. Hasil evaluasi kinerja dari penelitian tersebut pada model *Squeezenet*, VGG16, dan *MobilenetV2* berturut-turut adalah akurasi sebesar 47,84%, 47,48%, 41,83% dan *IoU F1-score* sebesar 43,63%, 44,09%, 37,22%, namun tidak menggunakan evaluasi kinerja lain.

Berdasarkan beberapa penelitian yang telah disebutkan, dapat dilihat bahwa CNN belum memiliki teknik *patching* secara khusus yang digunakan pada keseluruhan model CNN. Penambahan *patching* pada CNN digunakan untuk

menambah ketersediaan data yang terbatas untuk pemenuhan data CNN yang membutuhkan banyak data. Penambahan teknik *patching* pada proses pembelajaran juga turut menambah waktu yang dibutuhkan untuk melakukan pembelajaran mesin pada *pre-processing* sebelum diterapkannya CNN. *Vision Transformer (ViT)* juga merupakan bagian dari DL. Berbeda dengan CNN, *ViT* sudah memiliki teknik *patching* di dalam arsitekturnya.

Vision Transformer (ViT) adalah model untuk klasifikasi gambar yang menggunakan *attention* dan menghilangkan perulangan serta operasi konvolusi yang dilakukan di atas *patching* sebuah gambar (Dosovitskiy *et al.*, 2021). *ViT* pertama kali diperkenalkan pada tahun 2017 dalam penelitian yang dilakukan oleh Vaswani *et al.* (2017) yang menggunakan arsitektur *ViT* dalam tugas penerjemah bahasa Inggris ke Jerman dan bahasa Inggris ke Prancis dan telah berhasil diterapkan dalam tugas klasifikasi gambar (Zhou *et al.*, 2021). Kelebihan dari *ViT* adalah arsitektur ini tidak menggunakan operasi konvolusi melainkan *attention* sehingga mampu mereduksi parameter yang digunakan dalam CNN (Heo *et al.*, 2021). Model *ViT* telah terbukti lebih baik dalam hal efisiensi dan akurasi komputasi bahkan *ViT* menunjukkan kemampuan pengenalan bentuk yang sebanding dengan sistem visual manusia (Naseer *et al.*, 2021).

Teknik *patching* pada *ViT* langsung memotong gambar menjadi beberapa *patch*, kemudian masing-masing *patch* disusun secara linier, lalu ditambahkan *embedding position*, dan urutan vektor yang dihasilkan akan divektorisasi ke *transformer encoder* (Yuan *et al.*, 2021). Model *ViT* sudah memiliki teknik khusus dalam melakukan *patching* sehingga model ini tidak lagi membutuhkan *patching*

secara terpisah. Kelebihan dari *ViT* adalah penggunaan memori yang tidak terlalu besar untuk melakukan proses pembelajaran karena parameter yang diperlukan dalam arsitektur ini tidak banyak dan tidak ada pengulangan (Jaegle *et al.*, 2021).

Beberapa penelitian untuk pengklasifikasian gambar menggunakan arsitektur *ViT* telah dilakukan oleh Gao *et al.* (2021) yang membandingkan arsitektur *ViT* dan *DenseNet* dalam mengklasifikasikan penyakit COVID dimana diperoleh hasil evaluasi *F1-score* menggunakan *ViT* adalah 0,76. Nilai ini lebih baik dibandingkan *DenseNet* dengan nilai 0,72. Penelitian lain juga dilakukan oleh Bazi *et al.* (2021) yang menghasilkan nilai akurasi sangat baik dengan rata-rata nilai di atas 95% pada klasifikasi pemandangan penginderaan jauh pada *Merced dataset*, *Aerial Image Dataset (AID)* dan *Optimal-31 dataset* menggunakan arsitektur *ViT*. Penelitian lain yang dilakukan oleh Wu *et al.* (2021) juga menghasilkan nilai yang sangat baik pada evaluasi kerja dengan nilai akurasi 91,4%, spesifisitas 97,7%, presisi 92,8%, sensitivitas 92,6%, skor kappa 93,5%, dan AUC 98,6% pada pengenalan penyakit *diabetic retinopathy* menggunakan *ViT*. Berdasarkan beberapa hasil penelitian yang telah dilakukan, dapat dilihat bahwa klasifikasi menggunakan arsitektur *ViT* sudah mendapatkan kinerja yang sangat baik.

Penerapan klasifikasi pada dunia kesehatan salah satunya dilakukan untuk mengklasifikasikan penyakit glaukoma. Glaukoma adalah penyakit mata yang dimulai karena tekanan intraokular yang berlebihan di dalamnya dan menyebabkan kebutaan total pada stadium lanjut (Nawaz *et al.*, 2022). Glaukoma diklasifikasikan menjadi 3 kelas yaitu *advanced glaucoma*, *early glaucoma*, dan *normal control* (Akil *et al.*, 2017). Beberapa penelitian yang telah membahas klasifikasi glaukoma

dilakukan oleh Cahya *et al.* (2021) yang menggunakan arsitektur *AlexNet-CNN* dengan nilai akurasi sudah sangat baik yaitu 98,37%, namun tidak ada evaluasi kinerja lain. Penelitian lainnya dilakukan oleh Ahn *et al.* (2018) yang menggunakan *Inception V3-CNN* dengan hasil akurasi 84,5%, AUROC sebesar 0,93. Penelitian lain juga dilakukan oleh Roshini and Alex, (2022) yang menggunakan model *MultiResUNet* untuk melakukan klasifikasi glaukoma dengan hasil akurasi klasifikasi rata-rata sebesar 97,2%, namun tidak menghitung evaluasi kinerja lain.

Berdasarkan keterbatasan dari CNN pada klasifikasi penyakit glaukoma, maka dalam tugas akhir ini dilakukan penerapan arsitektur *ViT* pada citra retina dalam pengklasifikasian penyakit glaukoma untuk meningkatkan hasil kinerja pada metode sebelumnya. Hasil kinerja pada ViT akan diukur keberhasilannya berdasarkan nilai akurasi, sensitivitas, spesifisitas, *F1-score*, dan *cohen's kappa*.

1.2. Perumusan Masalah

Permasalahan pada penelitian ini membahas bagaimana melakukan klasifikasi penyakit glaukoma pada citra retina menggunakan arsitektur *Vision Transformer* untuk meningkatkan evaluasi kinerja berdasarkan nilai akurasi, sensitivitas, spesifisitas, *F1-score*, dan *cohen's kappa*.

1.3. Pembatasan Masalah

Beberapa pembatasan masalah dalam penelitian ini:

1. Penelitian yang dilakukan hanya membahas klasifikasi penyakit Glaukoma menggunakan citra fundus retina yang terdiri dari 3 kelas yaitu *advanced glaucoma*, *early glaucoma*, dan *normal control*.

2. Pembahasan evaluasi kinerja dari model arsitektur klasifikasi ini hanya akan mengukur kinerja berdasarkan nilai akurasi, sensitivitas, spesifisitas, *F1-score*, dan *cohen's kappa*.
3. Penelitian ini tidak membahas perbaikan citra.

1.4. Tujuan

Tujuan dari penelitian ini adalah untuk mengklasifikasikan penyakit glaukoma pada citra retina menggunakan arsitektur *Vision Transformer* dan mengukur keberhasilannya berdasarkan nilai akurasi, sensitivitas, spesifisitas, *F1-score*, dan *cohen's kappa*.

1.5. Manfaat

Manfaat dari penelitian ini antara lain:

1. Dapat digunakan sebagai model alternatif untuk klasifikasi gangguan penyakit glaukoma.
2. Dapat diterapkan modelnya pada teknologi yang lebih lanjut untuk mesin otomatis pendeteksi penyakit glaukoma.
3. Memberikan referensi baru pada bidang *deep learning*.

DAFTAR PUSTAKA

- Ahn, J. M., Kim, S., Ahn, K. S., Cho, S. H., Lee, K. B., & Kim, U. S. (2018). A deep learning model for the detection of both advanced and early glaucoma using fundus photography. *PLoS ONE*, *13*(11), 1–8. <https://doi.org/10.1371/journal.pone.0207982>
- Akil, H., Huang, A. S., Francis, B. A., Sadda, S. R., & Chopra, V. (2017). Retinal vessel density from optical coherence tomography angiography to differentiate early glaucoma, pre-perimetric glaucoma and normal eyes. *PLoS ONE*, *12*(2), 1–12. <https://doi.org/10.1371/journal.pone.0170476>
- Banerjee, K., Vishak Prasad, C., Gupta, R. R., Vyas, K., Anushree, H., & Mishra, B. (2021). Exploring alternatives to softmax function. *Proceedings of the 2nd International Conference on Deep Learning Theory and Applications, DeLTA 2021*, 81–86. <https://doi.org/10.5220/0010502000810086>
- Bazi, Y., Bashmal, L., Al Rahhal, M. M., Dayil, R. Al, & Ajlan, N. Al. (2021). Vision transformers for remote sensing image classification. *Remote Sensing*, *13*(3), 1–20. <https://doi.org/10.3390/rs13030516>
- Cahya, F. N., Hardi, N., Riana, D., & Hadiani, S. (2021). Klasifikasi penyakit mata menggunakan convolutional neural network (cnn). *Jurnal Sistem Informasi*, *10*, 618–626.
- Claro, M., Veras, R., Santana, A., Araújo, F., Silva, R., Almeida, J., & Leite, D. (2019). An hybrid feature space from texture information and transfer learning for glaucoma classification. *Journal of Visual Communication and Image Representation*, *64*. <https://doi.org/10.1016/j.jvcir.2019.102597>
- Delgado, R., & Tibau, X. A. (2019). Why cohen’s kappa should be avoided as performance measure in classification. *PLoS ONE*, *14*(9), 1–26. <https://doi.org/10.1371/journal.pone.0222916>
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16x16 words: transformers for image recognition at scale. *ICLR 2021*. <http://arxiv.org/abs/2010.11929>
- Ertam, F., & Aydin, G. (2017). Data classification with deep learning using tensorflow. *2nd International Conference on Computer Science and Engineering, UBMK 2017*, 755–758. <https://doi.org/10.1109/UBMK.2017.8093521>
- Fan, H., Xiong, B., Mangalam, K., Li, Y., Yan, Z., Malik, J., & Feichtenhofer, C. (2021). Multiscale vision transformers. *IEEE Xplore*, 6824–6835. <https://doi.org/10.1109/iccv48922.2021.00675>
- Foody, G. M. (2019). Explaining the unsuitability of the kappa coefficient in the assessment and comparison of the accuracy of thematic maps obtained by

- image classification. *Remote Sensing of Environment*, 239(December 2019), 111630. <https://doi.org/10.1016/j.rse.2019.111630>
- Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., & Greenspan, H. (2017). Modeling the intra-class variability for liver lesion detection using a multi-class patch-based cnn. *Springer International*, 129–137. <https://doi.org/10.1007/978-3-319-67434-6>
- Gao, X., Qian, Y., & Gao, A. (2021). Covid-vit: classification of covid-19 from ct chest images based on vision transformer models. *Cornell University*. <https://doi.org/10.48550/arXiv.2107.01682>
- Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y., Yang, Z., Zhang, Y., & Tao, D. (2022). A survey on vision transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–1. <https://doi.org/10.1109/tpami.2022.3152247>
- Hao, W., Yizhou, W., Yaqin, L., & Zhili, S. (2020). The role of activation function in cnn. *Proceedings - 2020 2nd International Conference on Information Technology and Computer Application, ITCA 2020*, 429–432. <https://doi.org/10.1109/ITCA52113.2020.00096>
- Hasnain, M., Pasha, M. F., Ghani, I., Imran, M., Alzahrani, M. Y., & Budiarto, R. (2020). Evaluating trust prediction and confusion matrix measures for web services ranking. *IEEE Access*, 8, 90847–90861. <https://doi.org/10.1109/ACCESS.2020.2994222>
- Hendrycks, D., & Gimpel, K. (2016). Gaussian error linear units (gelus). *Cornell University*, 1–9. <http://arxiv.org/abs/1606.08415>
- Heo, B., Yun, S., Han, D., Chun, S., Choe, J., & Oh, S. J. (2021). Rethinking spatial dimensions of vision transformers. *International Conference on Computer Vision (ICCV)*, 11936–11945. <https://doi.org/10.1109/iccv48922.2021.01172>
- Ilahiyah, S., & Nilogiri, A. (2018). Implementasi Deep Learning Pada Identifikasi Jenis Tumbuhan Berdasarkan Citra Daun Menggunakan Convolutional Neural Network. *JUSTINDO (Jurnal Sistem Dan Teknologi Informasi Indonesia)*, 3(2), 49–56.
- Jaegle, A., Gimeno, F., Brock, A., Zisserman, A., Vinyals, O., & Carreira, J. (2021). Perceiver: general perception with iterative attention. *38th International Conference on Machine Learning*. <http://arxiv.org/abs/2103.03206>
- Jiang, Y., & Han, F. (2017). A hybrid algorithm of adaptive particle swarm optimization based on adaptive moment estimation method. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10361 LNCS, 658–667. https://doi.org/10.1007/978-3-319-63309-1_58
- Jun, T. J., Eom, Y., Kim, D., Kim, C., Park, J. H., Nguyen, H. M., Kim, Y. H., & Kim, D. (2021). TRk-CNN: Transferable Ranking-CNN for image

- classification of glaucoma, glaucoma suspect, and normal eyes. *Expert Systems with Applications*, 182. <https://doi.org/10.1016/j.eswa.2021.115211>
- Juneja, M., Thakur, N., Thakur, S., Uniyal, A., Wani, A., & Jindal, P. (2020). Gc-net for classification of glaucoma in the retinal fundus image. *Machine Vision and Applications*, 31(5), 1–18. <https://doi.org/10.1007/s00138-020-01091-4>
- Kastner, A., & King, A. J. (2020). Advanced glaucoma at diagnosis: current perspectives. *Eye*, 116–128. <https://doi.org/10.1038/s41433-019-0637-2>
- Kim, W. S., Lee, D. H., Kim, Y. J., Kim, T., Hwang, R. Y., & Lee, H. J. (2020). Path detection for autonomous traveling in orchards using patch-based cnn. *Computers and Electronics in Agriculture*, 175(July), 105620. <https://doi.org/10.1016/j.compag.2020.105620>
- Krstinić, D., Braović, M., Šerić, L., & Božić-Štulić, D. (2020). Multi-label classifier performance evaluation with confusion matrix. *Comput Sci Inf Technol*, 01–14. <https://doi.org/10.5121/csit.2020.100801>
- Mishra, S., Vanli, O. A., Huffer, F. W., & Jung, S. (2016). Regularized discriminant analysis for multi-sensor decision fusion and damage detection with Lamb waves. *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2016*, 9803(March), 98032H. <https://doi.org/10.1117/12.2217959>
- Misra, D., Crispim-Junior, C., & Tougne, L. (2020). Patch-based cnn evaluation for bark classification. In A. Bartoli & A. Fusiello (Eds.), *Computer Vision -- ECCV 2020 Workshops* (pp. 197–212). Springer International Publishing.
- Naseer, M., Ranasinghe, K., Khan, S., Hayat, M., Khan, F. S., & Yang, M.-H. (2021). Intriguing properties of vision transformers. *Cornell University, NeurIPS*. <http://arxiv.org/abs/2105.10497>
- Nawaz, M., Nazir, T., Javed, A., Tariq, U., Yong, H., Khan, M. A., & Cha, J. (2022). An efficient deep learning approach to automatic glaucoma detection using optic disc and optic cup localization. *Sensors*. <https://doi.org/https://doi.org/10.3390/s22020434>
- Pogorelov, K., Randel, K. R., Griwodz, C., Eskeland, S. L., De Lange, T., Johansen, D., Spampinato, C., Dang-Nguyen, D. T., Lux, M., Schmidt, P. T., Riegler, M., & Halvorsen, P. (2017). Kvasir: a multi-class image dataset for computer aided gastrointestinal disease detection. *Proceedings of the 8th ACM Multimedia Systems Conference, MMSys 2017*, 164–169. <https://doi.org/10.1145/3083187.3083212>
- Putra, I. wayan S. E., Wijaya, A. Y., & Soelaiman, R. (2016). Klasifikasi citra menggunakan convolutional neural network (cnn) pada caltech 101. *Jurnal Teknik ITS*, 5(1), A65–A69. <http://repository.its.ac.id/48842/>
- Raharja, B. D., & Harsadi, P. (2018). Implementasi kompresi citra digital dengan mengatur kualitas citra digital. *Jurnal Ilmiah Sinus*, 2, 71–77.

- Ramchoun, H., Idrissi, M. A. J., Ghanou, Y., & Ettaouil, M. (2017). New modeling of multilayer perceptron architecture optimization with regularization: An application to pattern classification. *IAENG International Journal of Computer Science*, 44(3), 261–269.
- Roshini, R., & Alex, J. S. R. (2022). Automatic segmentation of optic cup and optic disc using multiresunet for glaucoma classification from fundus image. In M. Saraswat, H. Sharma, & K. V. Arya (Eds.), *Intelligent Vision in Healthcare* (pp. 33–44). Springer Singapore. https://doi.org/10.1007/978-981-16-7771-7_4
- Ruder, S. (2016). An overview of gradient descent optimization algorithms. *Cornell University*, 1–14. <http://arxiv.org/abs/1609.04747>
- Saprudin, Amalia, R., & Rosyani, P. (2021). Klasifikasi citra menggunakan metode random forest dan sequential minimal optimization (smo) image classification using random forest method and sequential minimal optimization (smo). *Jurnal Sistem Dan Teknologi Informasi*, 09(2), 2020–2022. <https://doi.org/10.26418/justin.v9i2.44120>
- Schuster, A. K., Erb, C., Hoffmann, E. M., Dietlein, T., & Pfeiffer, N. (2020). The diagnosis and treatment of glaucoma. *Dtsch Arztebl Int*. <https://doi.org/10.3238/arztebl.2020.0225>
- Shabaniyan, M., Eckstein, E. C., Chen, H., & Devincenzo, J. P. (2019). Classification of neurodevelopmental age in normal infants using 3D-cnn based on brain mri. *Proceedings - 2019 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2019, cCMV*, 2373–2378. <https://doi.org/10.1109/BIBM47256.2019.8983399>
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1). <https://doi.org/10.1186/s40537-019-0197-0>
- Siu, K., Stuart, D. M., Mahmoud, M., & Moshovos, A. (2018). Memory requirements for convolutional neural network hardware accelerators. *2018 IEEE International Symposium on Workload Characterization, IISWC 2018*, 111–121. <https://doi.org/10.1109/IISWC.2018.8573527>
- Solomon, C., & Breckon, T. (2011). *Fundamentals of Digital Image Processing*. John Wiley & Sons, Ltd.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 2017-Decem(Nips)*, 5999–6009.
- Wang, M., Lu, S., Zhu, D., Lin, J., & Wang, Z. (2019). A high-speed and low-complexity architecture for softmax function in deep learning. *2018 IEEE Asia Pacific Conference on Circuits and Systems, APCCAS 2018*, 223–226. <https://doi.org/10.1109/APCCAS.2018.8605654>

- 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>
- Xu, J., Sun, X., Zhang, Z., Zhao, G., & Lin, J. (2019). Understanding and improving layer normalization. *Advances in Neural Information Processing Systems*, 32(NeurIPS), 1–19.
- Yuan, L., Chen, Y., Wang, T., Yu, W., Shi, Y., Jiang, Z., Tay, F. E. H., Feng, J., & Yan, S. (2021). Tokens-to-token vit: training vision transformers from scratch on imagenet. *International Conference on Computer Vision (ICCV)*, 538–547. <https://doi.org/10.1109/iccv48922.2021.00060>
- Zebua, T., & Ndruru, E. (2017). Pengamanan Citra Digital berdasarkan Modifikasi Algoritma RC4. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 4(4), 275–282. <https://doi.org/10.25126/jtiik.201744474>
- Zhang, B., & Sennrich, R. (2019). Root mean square layer normalization. *Advances in Neural Information Processing Systems*, 32(NeurIPS), 1–12.
- Zhang, Q., Yang, L. T., Chen, Z., & Li, P. (2018). A survey on deep learning for big data. *Information Fusion*, 42(November 2017), 146–157. <https://doi.org/10.1016/j.inffus.2017.10.006>
- Zhou, D., Kang, B., Jin, X., Yang, L., Lian, X., Jiang, Z., Hou, Q., & Feng, J. (2021). Deepvit: towards deeper vision transformer. *Cornell University*. <http://arxiv.org/abs/2103.11886>