BAB 2
TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Penelitian yang dilakukan oleh (Indraswari dkk., 2022) dengan judul Deteksi Penyakit Mata Pada Citra Fundus Menggunakan *Convolutional Neural Network* (CNN). Penelitian ini menggunakan metode CNN dengan arsitektur MobileNetV2 sebagai *base model* yang telah dilatih menggunakan dataset ImageNet. Penelitian ini bertujuan untuk mendeteksi penyakit mata berdasarkan citra fundus dengan dua kelas, yaitu "Normal" dan "Abnormal". Dataset yang digunakan terdiri dari 601 citra fundus retina. Hasil penelitian menunjukkan bahwa model yang diusulkan mampu mencapai akurasi sebesar 72%, precision 72%, recall 72%, dan F1-score sebesar 72%. Selain itu, penelitian ini juga menyimpulkan bahwasannya MobileNetV2 lebih baik daripada metode ResNet50V2, InceptionV3, dan VGG19 dengan memperhatikan performa dan *training time* metode.

Asia dkk. (2022) melaksanakan penelitian dengan judul *Detection of Diabetic Retinopathy in Retinal Fundus Images Using CNN Classification Models*. Penelitian ini menggunakan metode CNN dengan membandingkan beberapa arsitektur, yaitu ResNet-101, ResNet-50, dan VggNet-16. Dataset yang digunakan dalam studi ini diperoleh dari *Xiangya No. 2 Hospital Ophthalmology* (XHO), Changsha, Cina, yang berukuran sangat besar namun memiliki label yang tidak seimbang. Hasil penelitian menunjukkan bahwa arsitektur ResNet-101, berhasil mencapai tingkat akurasi maksimum sebesar 98,88% (0,9888) dengan *training loss* sebesar 0,3499 dan *testing loss* sebesar 0,9882 pada dataset XHO.

Penggunaan metode YOLO V8 menunjukkan bahwa model yang digunakan berhasil mencapai akurasi sebesar 92%, dengan precision 91%, recall 92%, dan F1-score 91% pada penelitian yang dilakukan oleh (Nur dkk., 2023). Penelitian yang berjudul Klasifikasi Penyakit Mata Berdasarkan Citra Fundus Menggunakan YOLO V8 ini melakukan klasifikasi penyakit mata seperti katarak, glaukoma, dan retinopati diabetik berdasarkan citra fundus. Penelitian ini bertujuan untuk membangun sistem deteksi dini penyakit mata secara otomatis.

Usulan metode gabungan VGG16 dengan *XGBoost Classifier*, serta jaringan

DenseNet 121 oleh (Mohanty dkk., 2023) menunjukkan bahwa model jaringan hibrida (VGG16 + XGBoost) berhasil mencapai akurasi sebesar 79,50%, sedangkan model DenseNet 121 berhasil mencapai akurasi sebesar 97,30%. Pada penelitiannya yang berjudul *Using Deep Learning Architectures for Detection and Classification of Diabetic Retinopathy* menggunakan dataset dari kumpulan citra retina *APTOS 2019 Blindness Detection Kaggle Dataset*. Melalui analisis perbandingan dengan metode terdahulu yang menggunakan dataset yang sama, jaringan DenseNet 121 terbukti memiliki performa yang lebih unggul dan sangat efektif dalam klasifikasi citra DR.

Penggunaan metode ViT dengan menerapkan teknik *pre-trained Masked Autoencoders* (MAE) mendapati hasil yang memuaskan pada penelitian oleh (Yang dkk., 2024). Penelitian yang berjudul *Vision Transformer with Masked Autoencoders for Referable Diabetic Retinopathy Classification Based on Large-size Retina Image* memiliki hasil akurasi sebesar 93,42%, *Area Under Curve* (AUC) sebesar 0,9853, sensitivitas tertinggi 0,973, dan spesifisitas tertinggi sebesar 0,9539. Dataset yang digunakan terdiri dari lebih dari 100.000 citra fundus retina publik berukuran lebih besar dari 224 x 224 piksel. Penelitian ini menunjukkan bahwa model ViT dengan *pre-trained* MAE pada citra retina memberikan peningkatan performa yang lebih superior dibandingkan dengan penggunaan bobot dari ImageNet.

Penelitian yang dilakukan oleh (Qodri, 2024) dengan judul Analisis Perbandingan Klasifikasi Tumor Otak Menggunakan *Deep Learning*. Penelitian ini menggunakan beberapa arsitektur, yaitu ResNet50, Xception, dan VGG16. Dataset yang digunakan berupa dataset publik yang terdiri dari 3.618 gambar MRI. Hasil penelitian menunjukkan bahwa model ResNet50 berhasil mendapatkan akurasi sebesar 96% dan arsitektur VGG16 juga mendapatkan akurasi yang sama sebesar 96%.

Penggunaan dataset yang terdiri dari 12.539 citra X-ray yang diambil dari repositori publik di penelitian yang dilakukan oleh (Syakuroh dkk., 2025) dengan judul Analisis Akurasi Model MobileNetV2 dalam Klasifikasi Citra X-Ray untuk Deteksi Kondisi Paru-Paru memberikan hasil yang memuaskan. Hasil penelitian menunjukkan bahwa model MobileNetV2 yang diusulkan mampu mencapai nilai

akurasi yang sangat tinggi sebesar 99,42%, precision 98,87%, recall 98,88%, dan F1-score 98,86%. Penelitian ini menggunakan arsitektur MobileNetV2 dengan menerapkan fine-tuning pada lapisan akhir model.

Perbandingan tiga metode yang diusulkan pada penelitian oleh (Grace & Gunawan, 2025) dengan judul Perbandingan CNN, ResNet50, dan *Vision Transformer* untuk Klasifikasi Kanker Payudara Berbasis Web menunjukkan hasil bahwa model ResNet50 berhasil mencapai akurasi tertinggi sebesar 97%, diikuti oleh arsitektur CNN sebesar 92%, dan *Vision Transformer* (ViT) yang memperoleh akurasi terendah sebesar 71%. Dataset yang digunakan adalah *Curated Breast Imaging Subset of the Digital Database for Screening Mammography* (CBIS-DDSM) dari Kaggle.

Penelitian yang dilakukan oleh (Kawadkar, 2025) dengan judul *Comparative Analysis of Vision Transformers and Convolutional Neural Networks for Medical Image Classification*. Penelitian ini menggunakan metode *deep learning* dengan membandingkan performa antara CNN dan ViTs. Dataset yang digunakan terdiri dari total 8.469 citra medis yang mencakup pneumonia pada X-ray dada, klasifikasi tumor otak, serta deteksi melanoma kanker kulit. Hasil penelitian menunjukkan model ResNet-50 mencapai akurasi tertinggi sebesar 98,37% pada klasifikasi X-ray dada, DeiT-Small unggul dalam deteksi tumor otak dengan akurasi 92,16%, dan EfficientNet-B0 memimpin pada klasifikasi kanker kulit dengan akurasi sebesar 81,84%.

Teknik *transfer learning* menggunakan VGG16 dan ResNet50 nyatanya memberikan hasil yang memuaskan pada penelitian yang dilakukan oleh (Munadi, 2025). Penelitian yang berjudul Penerapan Algoritma *Deep Learning* untuk Deteksi Dini Penyakit dari Citra Medis ini menggunakan dataset yang terdiri dari ribuan gambar medis seperti X-ray, MRI, dan CT scan yang telah diklasifikasikan secara manual oleh para profesional medis. Hasil penelitian menunjukkan bahwa model *transfer learning* ResNet50 memberikan akurasi yaitu 95,3% dan VGG16 memberikan akurasi yaitu 93,7%.

Penggunaan filter pada metode *preprocessing* dapat memberikan pengaruh yang cukup signifikan pada pelatihan metode dalam penelitian oleh (Safitri dkk., 2025). Penelitian berjudul Analisis Perbandingan Metode Preprocessing untuk

Citra Retinopati Diabetik Menggunakan Deep learning ini membandingkan tiga teknik *preprocessing*, yaitu CLAHE, *gamma correction*, dan Retinex, serta menggunakan dua arsitektur, yaitu DenseNet121 dan EfficientNetB0. Dataset yang digunakan adalah data publik *APTOS 2019* yang terdiri dari 5 kelas berdasarkan tingkat keparahan penyakit. Hasil penelitian menunjukkan bahwa kombinasi *gamma correction* dan DenseNet121 menghasilkan akurasi tertinggi sebesar 82,4%, sedangkan akurasi terendah diperoleh dari kombinasi yang menggunakan teknik Retinex. Secara keseluruhan, arsitektur terbaik ditemukan pada EfficientNetB0 yang berhasil mencatatkan rata-rata akurasi sebesar 81,9%.

(Yudono dkk., 2025) dengan judul penelitiannya yaitu Klasifikasi Katarak Berdasarkan Optic Disc Citra Fundus Smartphone - Perbandingan Ekstraksi Ciri Tekstur dan Metode Neural Network menunjukkan bahwa model dengan akurasi terbaik pada GLCM dengan arsitektur PNN dengan tingkat akurasi 100%. Penelitian ini menggunakan metode ekstraksi ciri tekstur seperti Gray-Level Co-occurrence Matrix (GLCM) dan Filter Gabor, serta membandingkan metode klasifikasi Neural Network yaitu Backpropagation Neural Network (BPNN), Probabilistic Neural Network (PNN), dan Radial Basis Function Neural Network (RBFNN). Dataset yang digunakan pada penelitian ini berupa citra fundus retina yang diambil menggunakan kamera smartphone dari pasien di rumah sakit.

Tabel 2.1 Penelitian terdahulu

Peneliti	Metodologi	Hasil
(Indraswari dkk., 2022)	Mendeteksi citra fundus dengan 2 label, yaitu "Normal" dan "Abnormal" menggunakan MobileNetV2.	Model yang diusulkan mencapai akurasi sebesar 72%, precision 72%, recall 72%, dan F1-score sebesar 72%.
(Asia dkk., 2022)	Mendeteksi 2 kelas penyakit retinopati diabetik (positif atau negatif) pada citra fundus menggunakan ResNet-101, ResNet-50, dan VggNet-16.	Arsitektur ResNet-101, mendapatkan akurasi tertinggi sebesar 98,88% dengan *training loss* sebesar 0,3499 dan *testing loss* sebesar 0,9882.
(Nur dkk., 2023)	Klasifikasi 3 penyakit mata, yaitu katarak, glaukoma, dan retinopati diabetik menggunakan model YOLO V8.	Model mendapat akurasi sebesar 92%.
(Mohanty dkk., 2023)	Klasifikasi 5 tingkat keparahan retinopati diabetik menggunakan VGG16 +	Hasil training menunjukkan VGG16 + XGBoost berhasil mencapai akurasi sebesar

Peneliti	Metodologi	Hasil
	XGBoost Classifier dan DenseNet 121.	79,50%, sedangkan model DenseNet 121 berhasil mencapai akurasi sebesar 97,30%.
(Yang dkk., 2024)	Melakukan *pre-trained Masked Autoencoders* (MAE) ViT untuk mendeteksi 2 label citra fundus, yaitu "Normal" dan *"Referable DR"*.	Hasil akurasi didapatkan sebesar 93,42%, *Area Under Curve* (AUC) sebesar 0,9853, sensitivitas tertinggi 0,973, dan spesifisitas tertinggi sebesar 0,9539.
(Qodri, 2024)	Klasifikasi 2 label citra tumor otak, yaitu *"Positive"* dan *"Negative"* menggunakan model ResNet50, Xception, dan VGG16	Hasil terbaik pada model ResNet50 dan VGG16 mendapatkan akurasi yang sama sebesar 96%.
(Syakuroh dkk., 2025)	Deteksi 4 kondisi paru-paru, yaitu normal, *pneumonia*, *cardiomegaly*, dan *pneumothorax* dengan MobileNetV2	MobileNetV2 mencapai nilai akurasi 99,42%, precision 98,87%, recall 98,88%, dan F1-score 98,86%.
(Grace & Gunawan, 2025)	Klasifikasi 2 label penyakit tumor payudara, yaitu *Benign* (jinak) dan *Malignant* (kanker) berdasarkan citra rontgen menggunakan CNN biasa, ResNet50, dan ViT	ResNet50 mendapat akurasi tertinggi sebesar 97%, diikuti oleh CNN sebesar 92%, dan ViT dengan akurasi sebesar 71%.
(Kawadkar, 2025)	Melakukan deteksi terhadap 3 kasus, yaitu X-ray dada untuk deteksi 2 label (normal dan *pneumonia*), deteksi 2 label rontgen otak (*no tumor* atau tumor), dan citra kanker kulit manusia (*not melanoma* atau melanoma). Setiap kasus dideteksi dengan ResNet-50, EfficientNet-B0, ViT dan DeiT	Model ResNet-50 mencapai akurasi tertinggi sebesar 98,37% pada klasifikasi X-ray dada, DeiT-Small unggul dalam deteksi tumor otak dengan akurasi 92,16%, dan EfficientNet-B0 unggul pada klasifikasi kanker kulit dengan akurasi sebesar 81,84%.
(Munadi, 2025)	Klasifikasi banyak jenis penyakit untuk skrining massal menggunakan VGG16 dan ResNet50	ResNet50 memberikan akurasi yaitu 95,3% dan VGG16 memberikan akurasi yaitu 93,7%.
(Safitri dkk., 2025)	Klasifikasi 5 label tingkat keparahan retinopati diabetik menggunakan dataset APTOS 2019 dengan berfokus pada 3 metode *preprocessing,* serta menggunakan model DenseNet121 dan EfficientNetB0	Kombinasi *gamma correction* dan DenseNet121 menghasilkan akurasi tertinggi sebesar 82,4%, sedangkan akurasi terendah diperoleh dari teknik Retinex.

Peneliti	Metodologi	Hasil
(Yudono dkk., 2025)	Klasifikasi 4 label keparahan katarak (Normal, *Mild, Medium, Severe*) menggunakan 3 metode *preprocessing*. Model yang digunakan diantaranya Backpropagation Neural Network (BPNN), Probabilistic Neural Network (PNN), dan Radial Basis Function Neural Network (RBFNN).	Model dengan akurasi terbaik pada GLCM dengan arsitektur PNN dengan tingkat akurasi 100%.

Berdasarkan penelitian terdahulu yang dipaparkan pada Tabel 2.1, penggunaan metode CNN maupun *Transformer* sama-sama telah diterapkan dalam klasifikasi citra medis, namun performa keduanya masih menunjukkan hasil yang bervariasi dan belum ada kesimpulan yang konsisten mengenai arsitektur mana yang secara konsisten lebih unggul. Pada penelitian kali ini, model arsitektur *Transformer*, yaitu ViT-5 dan DeiT III akan dilatih untuk melakukan deteksi retinopati diabetik berdasarkan citra fundus retina. Penelitian ini juga akan menerapkan metode Ben Graham dalam *preprocessing* citra yang menggunakan filter *Gaussian Blur* untuk membantu ekstraksi fitur pada citra. Perbandingan kedua model dilakukan berdasarkan pada akurasi, *recall*, dan *Quadratic Weighted Kappa* (QWK). Selain performa deteksi, perbandingan juga dilakukan dari segi efisiensi sumber daya, yaitu waktu pelatihan, dan waktu inferensi kedua model. Target yang ingin dicapai melalui penelitian ini yaitu model *Transformer* dengan performa deteksi terbaik yang selanjutnya diimplementasikan ke dalam aplikasi Android, dengan komputasi model tetap berada pada komputer yang dihubungkan lewat API, serta tambahan LLM untuk membantu analisis hasil deteksi retinopati diabetik.

2.2 Landasan Teori

Dalam pelaksanaan penelitian ini, dibutuhkan teori-teori yang relevan guna memberikan pemahaman mendasar serta landasan ilmiah yang kuat dalam penyusunan sistem. Berikut pemaparan teori terkait dengan penelitian ini.

2.2.1 Retinopati Diabetik

Retinopati diabetik merupakan komplikasi mikrovaskular yang paling umum terjadi pada penderita diabetes melitus. Penyakit ini menjadi salah satu penyebab

utama hilangnya kemampuan penglihatan atau kebutaan pada populasi usia produktif di dunia. Diagnosis klinis retinopati diabetik umumnya didasarkan pada visualisasi bermanifestasinya lesi-lesi mikrovaskular serta abnormalitas pembuluh darah pada struktur retina. Kadar gula darah tinggi kronis memegang peran penting dalam memicu kerusakan metabolik yang berujung pada kerusakan pembuluh darah retina (W. Wang & Lo, 2018).

Retinopati diabetik memiliki tahapan kerusakan sebagai berikut:

1) *Non-proliferative Diabetic Retinopathy* (NPDR)

   NPDR merupakan tahapan awal dari penyakit retinopati diabetik. Pada tahap ini meliputi munculnya pelebaran kapiler kecil mikroaneurisma, pendarahan retina, dan gejala lainnya. Meskipun kerusakan telah terjadi di dalam retina, pasien pada tahap NPDR sering kali belum merasakan gejala gangguan penglihatan (W. Wang & Lo, 2018).

2) *Proliferative Diabetic Retinopathy* (PDR)

   PDR merupakan tahapan lanjut dari retinopati diabetik. Karakteristik utama yang membedakan PDR dengan tahapan sebelumnya adalah adanya neovaskularisasi, yaitu proses pembentukan pembuluh-pembuluh darah baru yang abnormal pada retina. Pembuluh darah baru ini bersifat sangat rapuh, dimana jika pembuluh tersebut pecah dan pendarahan masuk ke dalam korpus vitreus, atau memicu terjadinya ablasio retina traksional akan berakibat pada penurunan fungsi penglihatan yang masif hingga kebutaan permanen (W. Wang & Lo, 2018).

Berdasarkan tingkat keparahannya, retinopati diabetik diklasifikasikan ke dalam lima tingkatan yang digunakan sebagai label dalam penelitian ini, mengacu pada sistem klasifikasi *International Clinical Diabetic Retinopathy* (ICDR) (Wilkinson dkk., 2003). Setiap tingkatan memiliki karakteristik visual yang dapat diamati pada citra fundus retina sebagai berikut.

1) Tingkat 0 (*No Diabetic Retinopathy*)

   Citra fundus pada tingkat ini menunjukkan kondisi retina yang normal tanpa adanya tanda-tanda kerusakan akibat diabetes. Tidak ditemukan lesi, perdarahan, maupun abnormalitas pembuluh darah pada struktur retina. Tingkatan ini dijadikan sebagai kelas negatif atau referensi dalam proses

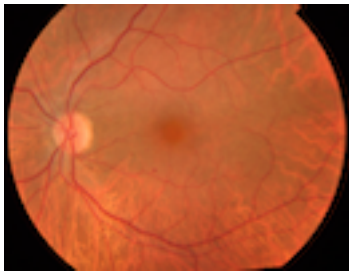
klasifikasi. Gambar 2.1 (Karthik dkk., 2019) memberikan gambaran citra fundus untuk retina tanpa penyakit retinopati diabetik.



Gambar 2.1 Retinopati diabetik tingkat 0

2) Tingkat 1 (*Mild* NPDR)

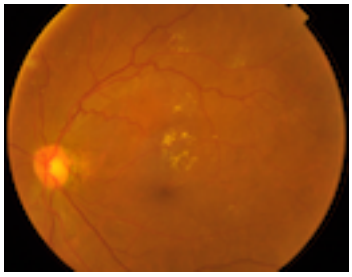
Pada tingkatan ini mulai ditemukan adanya mikroaneurisma, yaitu pelebaran kecil pada kapiler retina yang tampak sebagai titik-titik merah kecil pada citra fundus. Meskipun kerusakan sudah mulai terjadi, perubahan visual pada citra masih sangat minimal dan pasien umumnya belum merasakan gangguan penglihatan. Untuk visual retinopati diabetik tingkat 1 dapat dilihat pada Gambar 2.2 (Karthik dkk., 2019).



Gambar 2.2 Retinopati diabetik tingkat 1

3) Tingkat 2 (*Moderate* NPDR)

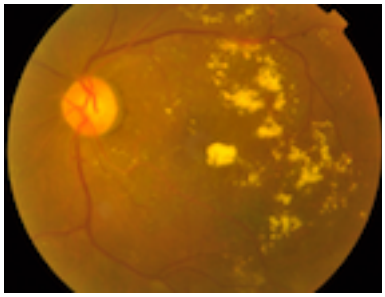
Citra fundus pada tingkatan ini menunjukkan kerusakan yang lebih luas dibandingkan Mild NPDR. Selain mikroaneurisma, mulai tampak perdarahan retina, eksudat keras, serta pelebaran vena retina. Jumlah dan distribusi lesi yang lebih banyak menjadi pembeda utama tingkatan ini dengan tingkatan sebelumnya. Pada Gambar 2.3 (Karthik dkk., 2019) dapat dilihat citra fundus dari retinopati tingkat 2.



Gambar 2.3 Retinopati diabetik tingkat 2

4) Tingkat 3 (*Severe* NPDR)

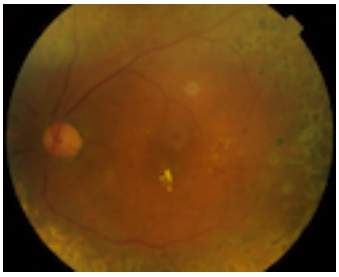
Pada tingkatan ini terjadi peningkatan signifikan dalam jumlah dan keparahan lesi. Citra fundus memperlihatkan perdarahan retina yang lebih masif, anomali mikrovaskular intraretinal (IRMA), serta pelebaran vena yang lebih nyata. Tingkatan ini merupakan tahap kritis sebelum memasuki fase proliferatif dan memiliki risiko tinggi untuk berkembang menjadi PDR. Citra fundus terkait tingkat 3 retinopati diabetik dapat dilihat pada Gambar 2.4 (Karthik dkk., 2019) di bawah.



Gambar 2.4 Retinopati diabetik tingkat 3

5) Tingkat 4 (PDR)

Tingkatan ini merupakan stadium paling parah yang ditandai dengan adanya neovaskularisasi, yaitu tumbuhnya pembuluh darah baru yang abnormal dan rapuh pada permukaan retina. Pada citra fundus, dapat ditemukan jaringan pembuluh darah baru, perdarahan vitreus, hingga tanda-tanda ablasio retina. Kondisi ini berpotensi menyebabkan penurunan fungsi penglihatan yang masif hingga kebutaan permanen apabila tidak segera ditangani. Gambar 2.5 (Karthik dkk., 2019) menunjukkan gambaran penyakit retinopati diabetik yang sudah masuk tingkat 5.



Gambar 2.5 Retinopati diabetik tingkat 5

2.2.2 Deep Learning

Salah satu cabang dari *machine learning* yang menerapkan pendekatan beru untuk melakukan analisis terhadap data yaitu *deep learning*. Pendekatan ini dilakukan dengan cara penekanan pada proses belajar bertahap melalui banyak lapisan jaringan. Pendekatan ini dirancang agar mampu mempelajari representasi data tingkat tinggi melalui arsitektur yang berlapis dan kompleks. Selain itu, *deep learning* juga ditujukan agar hasil akurasi yang bisa didapat menjadi lebih baik dan mempercepat proses komputasi pada berbagai jenis aplikasi.

Terdapat sejumlah tahapan penting dalam penerapan *deep learning*, yaitu (Munappy dkk., 2022):

1) Pengumpulan Data

   Pengumpulan data menjadi tahapan awal dalam membangun model *deep learning*. Dalam tahap ini, kualitas data yang dikumpulkan akan sangat menentukan performa model yang dihasilkan. Data yang dikumpulkan harus sesuai dengan permasalahan yang ingin diselesaikan, baik berupa citra, teks, maupun data numerik. Di samping itu, jumlah data yang cukup dan beragam diperlukan agar proses pembelajaran model dapat berjalan optimal. Apabila jumlah data yang tersedia terbatas, teknik augmentasi dapat diterapkan untuk memperkaya variasi data tanpa mengubah makna informasi aslinya.

2) *Preprocessing* Data

   Preprocessing data dilakukan untuk mempersiapkan data agar dapat dimanfaatkan secara maksimal oleh model. Tahapan ini mencakup pembersihan data dari gangguan (*noise*) dan duplikasi yang berpotensi memengaruhi hasil. Selain itu, dilakukan pula normalisasi atau standardisasi guna menyamakan skala data sehingga pola di dalamnya lebih mudah dikenali oleh model. Data juga harus diubah ke format yang sesuai dengan kebutuhan

arsitektur model, sebagai contoh, penyesuaian dimensi citra atau pengubahan teks menjadi representasi vektor numerik.

3) Pemisahan Data

Data yang telah diproses kemudian dibagi menjadi dua kelompok, yaitu data pelatihan (*training*) yang digunakan untuk mengajarkan pola kepada model, dan data pengujian (*testing*) yang digunakan untuk menilai sejauh mana model mampu melakukan generalisasi terhadap data yang belum pernah dilihat sebelumnya.

4) Pemilihan Model

Pada tahap ini, ditentukan arsitektur yang paling sesuai dengan karakteristik data dan permasalahan yang dihadapi. Sebagai contoh, *Convolutional Neural Network* (CNN) umum digunakan untuk data berbasis citra, sedangkan *Recurrent Neural Network* (RNN) atau *Long Short-Term Memory* (LSTM) lebih sesuai untuk data sekuensial seperti teks dan deret waktu. Terdapat juga arsitektur *Transformer* yang menjadi penemuan terbaru pada *deep learning* dan dapat digunakan baik untuk data berbasis citra maupun teks.

5) Pelatihan Model

Tahap pelatihan merupakan proses di mana model mempelajari pola dari data yang sudah diproses sebelumnya. Proses ini melibatkan dua mekanisme utama, yaitu *forward propagation*, saat data dialirkan melalui jaringan, serta *backpropagation*, yang berfungsi memperbarui bobot jaringan berdasarkan kesalahan yang ditemukan. Proses ini berlangsung berulang dalam beberapa *epoch* untuk meminimalkan nilai *loss function*. Selama pelatihan, parameter seperti *learning rate, batch size,* dan jenis *optimizer* turut disesuaikan agar proses belajar model lebih efektif. Pemantauan nilai *loss* dan akurasi pada setiap *epoch* menjadi hal penting guna menghindari kondisi *overfitting* maupun *underfitting*.

6) Evaluasi Model

Setelah proses pelatihan selesai, dilakukan evaluasi untuk mengukur kemampuan model dalam memprediksi data baru yang belum pernah digunakan sebelumnya. Evaluasi ini memanfaatkan data uji dengan metrik seperti *accuracy, precision, recall, F1-score,* atau *loss function*. Apabila hasil

evaluasi menunjukkan performa yang kurang memuaskan, maka diperlukan penyesuaian pada arsitektur model maupun data yang digunakan.

7) Validasi dan *Hyperparameter Tunning*

Tahap ini bertujuan untuk mengukur kemampuan model dalam mengenali pola pada data yang belum pernah dilihat sebelumnya, menggunakan dataset validasi yang terpisah dari data *training*. Pengukuran ini penting untuk memastikan model benar-benar mempelajari pola secara umum, bukan sekadar menghafal data yang digunakan saat pelatihan. Bersamaan dengan proses ini, dilakukan penyesuaian terhadap *hyperparameter* guna memperoleh kombinasi nilai yang menghasilkan performa terbaik. Berbeda dengan parameter model yang dipelajari secara otomatis selama pelatihan, *hyperparameter* merupakan nilai yang harus ditentukan sebelum proses pelatihan dimulai, seperti *learning rate*, *batch size*, jumlah neuron, maupun jumlah lapisan pada arsitektur.

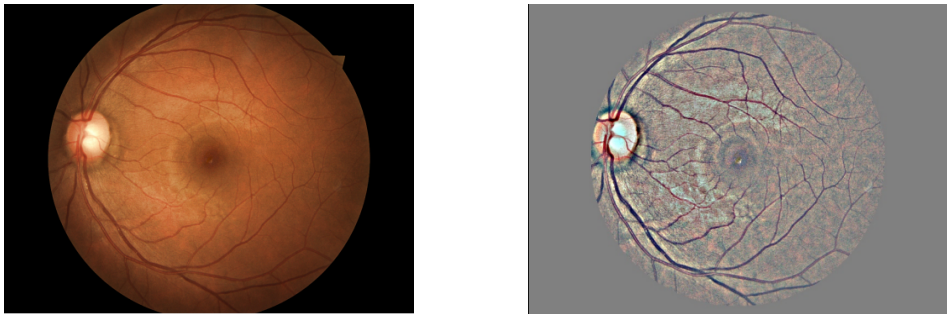
8) Deployment

Setelah melalui tahap pelatihan dan evaluasi, model selanjutnya dipersiapkan agar dapat digunakan secara nyata oleh pengguna pada tahap *deployment*. Tahap ini mencakup proses integrasi model ke dalam aplikasi, dilanjutkan dengan pengujian akhir untuk memastikan seluruh fungsi berjalan sesuai rancangan. Selain itu, performa model perlu terus dipantau secara berkala setelah diterapkan, sehingga apabila ditemukan penurunan kualitas prediksi, langkah perbaikan dapat segera diambil untuk menjaga performa model tetap optimal.

2.2.3 Metode Ben Graham

Metode Ben Graham merupakan teknik mengubah digital citra fundus yang awalnya diperkenalkan oleh Ben Graham dalam kompetisi *Kaggle Diabetic Retinopathy Detection* pada tahun 2015. Masalah utama yang sering dihadapi dalam pengolahan citra fundus retina adalah keragaman yang sangat tinggi pada faktor pencahayaan dan kontras warna akibat perbedaan jenis kamera fundus yang digunakan (Hassan dkk., 2025). Kondisi ini mengurangi, bahkan sampai menghilangkan bias induktif dan menurunkan fungsi *attention* pada model *deep learning* (Yang dkk., 2021). Untuk mengatasi masalah tersebut, metode Ben Graham menerapkan koreksi pencahayaan menggunakan *gaussian blur* (Graham,

2015).



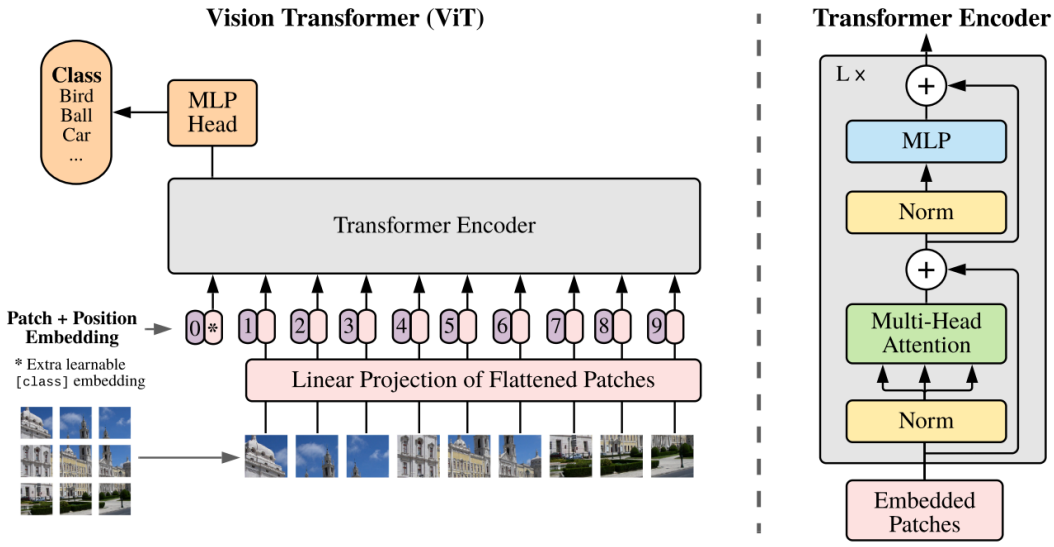
(a) (b)

Gambar 2.6 (a) Citra fundus asli dan (b) Citra fundus setelah pengaplikasian metode Ben Graham

Pada Gambar 2.6 (a) (Graham, 2015), citra tersebut merupakan salah satu dari ragam citra yang memiliki pencahayaan yang berbeda, sehingga perlu dikoreksi pencahayaannya. Koreksi pencahayaan dilakukan dengan filter *gaussian blur* yang berguna untuk mengurangi *noise* dan menyeimbangkan pencahayaan. Selanjutnya, citra tersebut akan dikurang dengan hasil *gaussian blur*-nya sendiri. Proses ini secara efektif menghasilkan gambar dengan latar belakang retina berwarna abu-abu netral yang homogen, sehingga eksudat keras, mikroaneurisma, dan perdarahan yang awalnya samar menjadi sangat kontras dan menonjol, seperti yang terlihat pada Gambar 2.6 (b) (Graham, 2015).

2.2.4 Vision Transformer (ViT)

Vision Transformer (ViT) merupakan salah satu inovasi terkini pada bidang deep learning. Arsitektur ini mengadopsi konsep *Transformer* yang sebelumnya banyak digunakan pada bidang pengolahan bahasa alami, kemudian disesuaikan agar dapat diterapkan pada pengolahan citra. Kombinasi dari tiga komponen utama, yaitu *patch embeddings*, *multi-head attention*, dan *multi-layer perceptron*, memungkinkan ViT untuk menganalisis dan memahami data visual secara lebih mendalam sekaligus efisien. Berkat pendekatan tersebut, informasi visual dapat diolah dengan cara yang jauh lebih optimal dibandingkan metode konvensional yang digunakan sebelumnya.



Gambar 2.7 Arsitektur model ViT

Berikut penjelasan untuk konsep yang diterapkan pada ViT berdasarkan Gambar 2.7 (Dosovitskiy dkk., 2021):

1) *Patch Embeddings*

Arsitektur Vision Transformer mensyaratkan citra masukan agar dapat dipecah menjadi beberapa *patch* berukuran n×n. Melalui proses *embeddings*, tiap *patch* hasil pembagian citra tersebut kemudian diubah menjadi vektor satu dimensi yang memuat nilai piksel dari bagian citra yang bersangkutan.

2) Layer Normalization

Proses pelatihan pada arsitektur *deep learning* umumnya membutuhkan sumber daya komputasi yang besar, sehingga diperlukan teknik yang mampu menekan kebutuhan daya dan waktu selama pelatihan berlangsung. Salah satu teknik yang digunakan untuk tujuan tersebut adalah *layer normalization*. Melalui teknik ini, nilai input pada setiap *layer* dinormalisasi berdasarkan rata-rata dan standar deviasi yang dihitung dari keseluruhan fitur dalam satu sampel data, sehingga proses pelatihan menjadi lebih stabil dan dapat berlangsung lebih cepat. Hasil dari normalisasi ini berupa nilai dengan rata-rata mendekati nol dan standar deviasi mendekati satu, sehingga skala nilai antar *layer* menjadi lebih seragam. Kondisi ini turut mencegah munculnya nilai gradien yang terlalu ekstrem, baik yang terlalu besar maupun terlalu kecil, selama proses pelatihan berlangsung, sehingga model dapat mencapai konvergensi dengan lebih cepat.

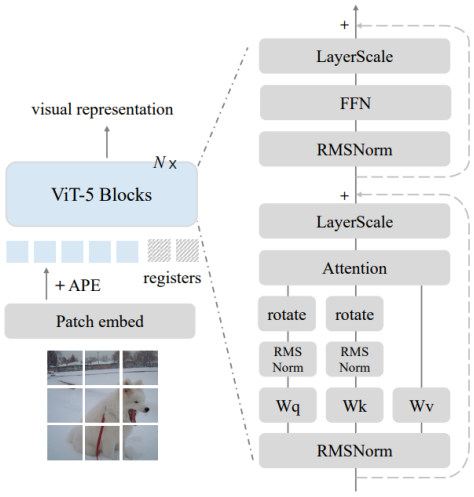
3) *Multi-Head Attention*

Setelah data melewati tahap *layer normalization*, tahap selanjutnya yang dilalui adalah *multi-head attention*. Pada tahap ini, mekanisme *attention* diterapkan agar informasi jangka panjang dapat ditangkap yang terdapat hubungan dari vektor masukan. Secara teknis, vektor masukan pada *multi-head attention* digandakan menjadi tiga komponen, yaitu *query*, *key*, dan *value*, yang selanjutnya diolah bersama melalui metode *scaled dot product*. Proses ini dilakukan berulang sesuai dengan jumlah *head* yang telah ditentukan, sebelum akhirnya seluruh hasilnya digabungkan dan dipadatkan kembali menjadi vektor dengan ukuran yang sudah ditetapkan.

4) *Multi-Layer Perceptron*

Multi-Layer Perceptron (MLP) merupakan salah satu jenis jaringan saraf tiruan yang konsep kerjanya terinspirasi dari sistem saraf manusia. Arsitektur MLP tersusun atas tiga bagian pokok, yaitu lapisan masukan (*input layer*), lapisan tersembunyi (*hidden layer*), dan lapisan keluaran (*output layer*). Setiap neuron dalam MLP memiliki bobot tersendiri, dan seluruh neuron, kecuali yang berada pada lapisan masukan, dilengkapi dengan fungsi aktivasi non-linear.

Penemuan terbaru terkait ViT, yaitu ViT-5 yang diusulkan oleh F. Wang dkk. (2026) melakukan pembaruan beberapa komponen pada ViT, dengan mempertahankan *Multi-Layer Perceptron* yang menjadi ciri khas dari ViT. Pembaruan ini mengadopsi berbagai kemajuan yang sudah terbukti efektif terhadap model selama lima tahun terakhir, tanpa merombak arsitektur dasar ViT.



Gambar 2.8 Arsitektur ViT-5

Berikut penjelasan terkait komponen terbaru yang diimplementasikan ke dalam model ViT-5 berdasarkan arsitektur pada Gambar 2.8 (F. Wang dkk., 2026):

1) LayerScale

Komponen ini berfungsi meningkatkan stabilitas dan performa training pada model ViT. Studi pada ViT-5 menemukan bahwasannya *LayerScale* memberikan hasil yang lebih fleksibel dan memiliki komputasi yang lebih rendah.

2) RMSNorm

ViT-5 menggantikan seluruh lapisan *LayerNorm* dengan *RMSNorm*. Penggantian operasi ini memberikan sedikit peningkatan akurasi sekaligus mengurangi beban komputasi.

3) *Positional Encoding* (APE + 2D RoPE)

ViT-5 menambahkan *Rotary Positional Embedding* (RoPE) namun tetap mempertahankan *Absolute Positional Encoding* (APE) secara bersamaan. Penelitian terkait menunjukkan bahwa ViT-5 memiliki ketahanan yang jauh lebih baik terhadap perubahan resolusi input.

4) Register Tokens

ViT-5 menambahkan *learnable register tokens* yang berfungsi sehingga dapat lebih fokus pada wilayah citra yang relevan. Pada ViT-5, *register tokens* juga diberikan RoPE tersendiri dengan basis frekuensi yang lebih tinggi dibanding token patch, untuk menghindari bias posisi yang tidak diinginkan akibat perbedaan perlakuan rotasi antara register dan patch.

5) QK-Normalization

ViT-5 menerapkan normalisasi tambahan pada vektor *query* dan *key* sebelum proses perhitungan *attention*. Komponen ini memberikan peningkatan performa dan secara signifikan meningkatkan stabilitas training serta mengurangi terjadinya lonjakan *loss*.

2.2.5 Data-efficient Image Transformer (DeiT)

Data-efficient image Transformers (DeiT) merupakan varian arsitektur *Vision Transformer* (ViT) yang dikembangkan untuk mengatasi keterbatasan utama ViT konvensional, yaitu ketergantungan yang tinggi pada dataset berskala besar dan komputasi yang sangat besar. Arsitektur DeiT dirancang agar dapat dilatih

secara optimal dengan dataset yang terbatas tanpa mengurangi performa yang signifikan dengan metode yang menggunakan arsitektur CNN, sehingga meningkatkan efisiensi pelatihan metode *Transformer* dengan sumber daya komputasi yang jauh lebih ringan (Touvron dkk., 2021).

Berikut konsep yang diterapkan dalam DeiT:

1) *Distillation Token*

   Komponen utama yang membedakan DeiT dengan arsitektur ViT standar adalah adanya token baru yang disebut *distillation token*. Token distilasi ini dimasukkan ke dalam susunan *patch embeddings* bersamaan dengan *class token* sebelum masuk ke lapisan *Transformer*. Melalui mekanisme *self-attention*, *distillation token* berinteraksi secara spasial dengan token gambar lainnya. Fungsi utama dari token ini adalah untuk memaksa model mempelajari dan mereproduksi estimasi label atau bias induktif yang dihasilkan oleh model *teacher* yang kuat (Touvron dkk., 2021).

2) *Knowledge Distillation* dan *Hard-Label Distillation*

   Knowledge Distillation merupakan sebuah konsep dalam *deep learning* di mana sebuah model *student* yang lebih kecil atau lebih efisien dilatih untuk meniru perilaku dan mentransfer pengetahuan dari sebuah model *teacher* yang sudah terlatih dan memiliki performa tinggi.

   Dalam proses *knowledge distillation*, DeiT memperkenalkan varian *hard-label distillation* sebagai alternatif dari metode *soft distillation*. Pada metode ini, *hard decision* atau nilai *argmax* dari prediksi model *teacher* diambil dan dianggap langsung sebagai target label nyata untuk dioptimalkan bersama *ground truth*. Penggunaan *hard label* ini terbukti secara signifikan meningkatkan akurasi *transformer student* dibandingkan metode *soft distillation*, serta menyederhanakan kalkulasi karena bersifat *parameter-free* (Touvron dkk., 2021).

Penelitian terbaru terkait DeiT yang disebut dengan DeiT III memberikan sejumlah komponen baru dengan *training recipe* yang terinspirasi dari *training recipe* ResNet-50 dan menghapus konsep *distillation token* dengan bantuan model *teacher*. Hasil dari perubahan ini memberikan efek yang signifikan terhadap akurasi model akhir. Beberapa komponen baru yang diterapkan pada resep training DeiT

III adalah sebagai berikut (Touvron dkk., 2022):

1) *Binary Cross Entropy* (BCE) *Loss*

   DeiT III mengadopsi penggunaan *binary cross-entropy loss* sebagai pengganti cross-entropy (CE) konvensional. Meskipun pada ResNet-50 penggunaan BCE memberikan keuntungan yang terbatas, pada ViT berukuran besar yang dilatih dengan resep DeiT III, BCE memberikan peningkatan performa yang signifikan.

2) *Stochastic Depth* dan *LayerScale*

   Stochastic depth digunakan sebagai metode regularisasi yang efektif untuk *training* jaringan yang dalam, dengan *drop-rate* yang disesuaikan terhadap ukuran model. *LayerScale* dipakai untuk membantu konvergensi *Transformer* yang dalam.

3) *3-Augment*

   Berbeda dengan metode augmentasi otomatis yang umum digunakan pada ViT seperti *RandAugment* atau *Auto-Augment*, *3-Augment* hanya menggunakan tiga jenis transformasi yang dipilih secara acak sebagai berikut:

   a) Grayscale

   b) Solarization

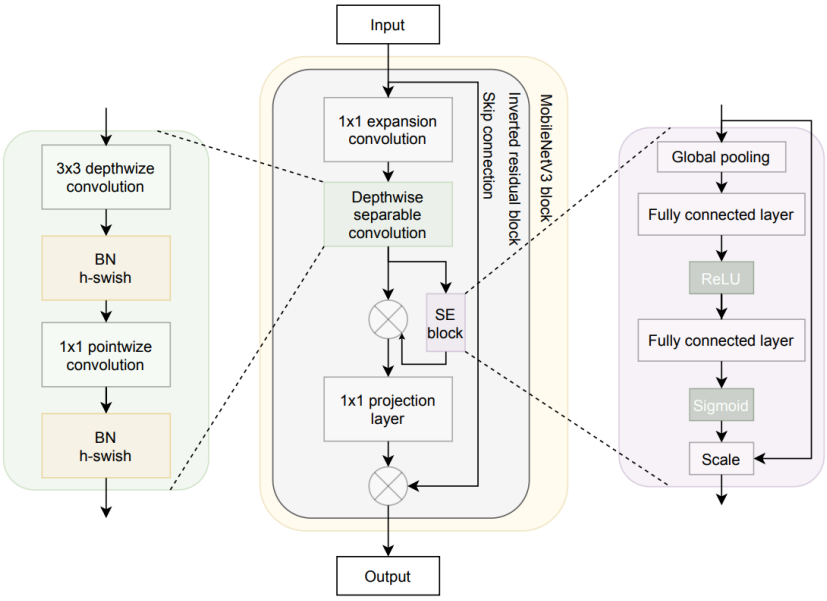
   c) Gaussian Blur

   Selain ketiga transformasi tersebut, ditambahkan juga *color-jitter* dan *horizontal flip* sebagai augmentasi standar. Penelitian yang berkaitan menunjukkan bahwa meskipun lebih sederhana dibanding augmentasi umum ViT, *3-Augment* memberikan performa yang lebih baik pada arsitektur ViT (Touvron dkk., 2022).

2.2.6 MobileNetV3

MobileNetV3 adalah arsitektur CNN yang dirancang untuk perangkat dengan sumber daya terbatas, dengan tujuan menjaga keseimbangan antara akurasi dan kecepatan inferensi (Howard dkk., 2019). Arsitektur ini dibangun dari kombinasi pencarian arsitektur otomatis dan perancangan manual, dan menghasilkan dua varian, yaitu MobileNetV3-*Large* untuk kebutuhan sumber daya yang lebih tinggi dan MobileNetV3-*Small* untuk kebutuhan sumber daya yang lebih rendah. Pada klasifikasi ImageNet, MobileNetV3-Large 3,2% lebih akurat dibandingkan MobileNetV2 dengan latensi 20% lebih rendah, sedangkan MobileNetV3-*Small*

6,6% lebih akurat dibandingkan latensi yang sebanding pada MobileNetV2 (Howard dkk., 2019).



Gambar 2.9 Arsitektur MobileNetV3

Berdasarkan Gambar 2.9, berikut konsep yang diterapkan dalam arsitektur model MobileNetV3 (Abd Elaziz dkk., 2021) :

1) *Inverted Residual Block*

   Struktur utama tempat pemrosesan informasi yang bekerja secara terbalik, yaitu memperluas saluran gambar terlebih dahulu dengan *expansion convolution*, mengekstrak fitur, lalu mengompresinya kembali *projection layer* agar komputasi tetap ringan.

2) *Skip Connection*

   Jalur pintas yang menghubungkan input langsung ke output akhir blok tanpa melewati proses konvolusi di tengahnya, berfungsi menjaga aliran informasi dan kestabilan gradien agar model mudah dilatih.

3) *Depthwise Separable Convolution*

   Teknik konvolusi yang menghemat komputasi hingga 9 kali lipat yang memisahkan proses menjadi dua tahap, yaitu *depthwise convolution* untuk penyaringan spasial, dan *pointwise convolution* untuk pembentukan fitur.

4) *Batch Normalization* dan *h-swish*

   Batch bNormalization (BN) menormalisasi distribusi nilai fitur, sedangkan fungsi aktivasi *hard-swish (h-swish)* sebagai pengganti *swish* yang lebih efisien

dan stabil dihitung pada perangkat seluler.

5) *SE Block (Squeeze-and-Excitation)*

   Modul *attention* ringan yang merangkum fitur gambar lewat *Global pooling*, lalu mempelajari tingkat kepentingan tiap saluran menggunakan *fully connected layer*, *ReLU*, dan *Sigmoid*, untuk kemudian mengalikan bobot penting tersebut ke fitur utama.

2.2.7 Confusion Matrix

Confusion matrix adalah tabel yang dipergunakan untuk melakukan perhitungan akurasi pada data baru yang diinputkan ke sistem yang sudah dilatih. (Heydarian dkk., 2022). Contoh dari C*onfusion matrix* untuk klasifikasi citra 2 label dapat dilihat pada Tabel 2.2 berikut (Heydarian dkk., 2022).

Tabel 2.2 *Confusion matrix*

		Nilai prediksi	
		Positif	Negatif
Nilai sebenarnya	Positif	*True Positive (TP)*	*False Negative (FN)*
	Negatif	*False Positive (FP)*	*True Negative (TN)*

Penjelasan untuk setiap kuadran yang ada pada tabel contoh *confusion matrix* adalah sebagai berikut:

1) *True Positive* (TP) menunjukkan jumlah prediksi label positif yang benar.
2) *True Negative* (TN) menunjukkan jumlah prediksi label negatif yang benar.
3) *False Positive* (FP) menunjukkan jumlah prediksi label positif yang salah.
4) *False Negative* (FN) menunjukkan jumlah prediksi label negatif yang salah.

Berdasarkan hasil prediksi yang diperoleh, perhitungan dapat dilakukan untuk menentukan nilai *accuracy, precision* dan *recall*. Berikut adalah rumus yang digunakan untuk menghitung nilai *accuracy, precision* dan *f1-score* (Heydarian dkk., 2022).

$$Accuracy = \frac{Total\ TP}{Total\ data} \times 100\% \quad (2\text{-}1)$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (2\text{-}2)$$

$$Recall\ (Sensitivity) = \frac{TP}{TP + FN} \times 100\% \quad (2\text{-}3)$$

$$F1\ Score = \frac{2TP}{2TP + FP + FN} \times 100\% \quad (2\text{-}4)$$

1) *Accuracy* digunakan untuk mengukur proporsi jumlah prediksi yang benar terhadap keseluruhan data uji. Metrik ini menunjukkan seberapa baik model dalam melakukan klasifikasi secara umum.
2) *Precision* digunakan untuk menilai ketepatan prediksi positif, yaitu perbandingan antara jumlah prediksi positif yang benar dengan seluruh prediksi positif yang dihasilkan oleh model.
3) *Recall* (*sensitivity*) digunakan untuk mengukur kemampuan model dalam menemukan seluruh kasus positif yang sebenarnya, yaitu perbandingan antara jumlah prediksi positif yang benar dengan seluruh data yang sebenarnya positif..
4) *F1-Score* digunakan untuk menggabungkan presisi dan recall melalui rata-rata harmonik, sehingga memberikan keseimbangan antara keduanya. Metrik ini sangat berguna terutama pada dataset yang tidak seimbang, karena mampu mengevaluasi performa model secara lebih adil pada setiap kelas.

2.2.8 Quadratic Weighted Kappa

Quadratic Weighted Kappa (QWK) merupakan metrik evaluasi yang digunakan untuk mengukur tingkat kesepakatan antara prediksi model dengan label sebenarnya pada data yang bersifat ordinal, seperti tingkat keparahan penyakit. QWK memperhitungkan besarnya jarak antara kelas prediksi dan kelas sebenarnya, sehingga kesalahan prediksi yang jauh dari label asli diberi penalti yang lebih besar dibandingkan kesalahan pada kelas yang berdekatan (Chilukoti dkk., 2024). Perhitungan QWK diawali dengan membangun matriks bobot kuadratik yang merepresentasikan besarnya penalti antara kelas prediksi j dan kelas sebenarnya i, dengan rumus sebagai berikut:

$$w_{i,j} = \frac{(i - j)^2}{(N - 1)^2} \qquad 2\text{-}5$$

N merupakan jumlah kelas sehingga nilai bobot semakin besar seiring bertambahnya jarak antara kelas prediksi dan kelas sebenarnya. Bobot penalti ini selanjutnya digunakan bersama matriks observasi serta matriks untuk menghasilkan nilai akhir QWK, yang merepresentasikan tingkat kesepakatan antara prediksi model dan label sebenarnya setelah memperhitungkan besar-kecilnya jarak kesalahan. Nilai QWK berada pada rentang -1 hingga 1, dengan nilai 1

menunjukkan kesepakatan sempurna antara prediksi model dan label sebenarnya, nilai 0 menunjukkan tingkat kesepakatan yang setara dengan prediksi acak, dan nilai negatif menunjukkan tingkat kesepakatan yang lebih buruk dibandingkan prediksi acak. Skala interpretasi QWK yang diusulkan oleh Landis & Koch (1977) dapat dilihat pada Tabel 2.3 berikut.

Tabel 2.3 Skala interpretasi nilai QWK

Rentang Nilai	Interpretasi
< 0,00	*Poor*
0,00 – 0,20	*Slight*
0,21 – 0,40	*Fair*
0,41 – 0,60	*Moderate*
0,61 – 0,80	*Substantial*
0,81 – 1,00	*Almost Perfect*