KLASIFIKASI CITRA ENDOSKOPI MENGGUNAKAN ARSITEKTUR *CONVNEXT* UNTUK IDENTIFIKASI PENYAKIT *GERD* DAN POLIP

SKRIPSI

Oleh :
MUHAMMAD FAQIH
NIM. 220605110069

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025

KLASIFIKASI CITRA ENDOSKOPI MENGGUNAKAN ARSITEKTUR *CONVNEXT* UNTUK IDENTIFIKASI PENYAKIT *GERD* DAN POLIP

SKRIPSI

Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh :
MUHAMMAD FAQIH
NIM. 220605110069

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025

HALAMAN PERSETUJUAN

KLASIFIKASI CITRA ENDOSKOPI MENGGUNAKAN ARSITEKTUR *CONVNEXT* UNTUK IDENTIFIKASI PENYAKIT *GERD* DAN POLIP

SKRIPSI

Oleh :
MUHAMMAD FAQIH
NIM. 220605110069

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 1 Desember 2025

Pembimbing I,	Pembimbing II,
Okta Qomaruddin Aziz, M.Kom	Ajib Hanani, M.T
NIP. 199110192019031013	NIP. 198407312023211013

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Supriyono, M. Kom
NIP. 19841010 201903 1 012

HALAMAN PENGESAHAN

KLASIFIKASI CITRA ENDOSKOPI MENGGUNAKAN ARSITEKTUR *CONVNEXT* UNTUK IDENTIFIKASI PENYAKIT *GERD* DAN POLIP

SKRIPSI

Oleh :
MUHAMMAD FAQIH
NIM. 220605110069

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 12 Desember 2025

Susunan Dewan Penguji

Ketua Penguji	: Prof. Dr. Suhartono S.Si M.Kom NIP. 196805192003121001	( )
Anggota Penguji I	: Shoffin Nahwa Utama, M.T NIP. 198607032020121003	( )
Anggota Penguji II	: Okta Qomaruddin Aziz, M.Kom NIP. 199110192019031013	( )
Anggota Penguji III	: Ajib Hanani, M.T NIP. 198407312023211013	( )

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Supriyono, M. Kom
NIP. 19841010 201903 1 012

PERNYATAAN KEASLIAN TULISAN

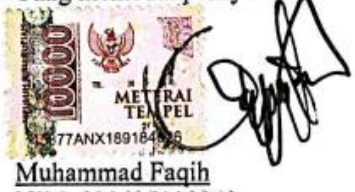
Saya yang bertanda tangan di bawah ini:

Nama	:	MUHAMMAD FAQIH
NIM		: 220605110069
Fakultas / Program Studi		: Sains dan Teknologi / Teknik Informatika
Judul Skripsi		: KLASIFIKASI CITRA ENDOSKOPI BERBASIS ARSITEKTUR *CONVNEXT* UNTUK IDENTIFIKASI PENYAKIT *GERD* DAN POLIP.

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 12 Desember 2025
Yang membuat pernyataan,



Muhammad Faqih
NIM. 220605110069

MOTTO

Not every spark becomes a flame.
But every flame once lived as a spark refusing to die.

「七転び八起き」

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur, karya ini saya persembahkan kepada:

Diriku sendiri. Terima kasih sudah bertahan sejauh ini. Terus belajar, bangkit setiap kali jatuh, dan tetap melangkah walau kadang rasanya berat. Bangga karena tidak menyerah.

Kedua orang tua tercinta, atas doa, dukungan, dan kasih sayang yang selalu jadi alasan untuk terus maju. Terima kasih untuk semuanya.

Seluruh keluarga, yang selalu jadi tempat pulang paling hangat, pemberi semangat, dan pengingat bahwa setiap usaha pasti ada hasilnya.

Para dosen dan pembimbing, yang telah membimbing, mengarahkan, dan membuka cara pandang saya selama proses penyusunan skripsi ini. Terima kasih untuk ilmu dan waktunya.

Sahabat dan teman-teman baik, di kampus maupun di luar, yang selalu hadir dengan obrolan, dukungan, tawa, dan cerita yang bikin perjalanan ini jauh lebih ringan dan menyenangkan. Terima kasih untuk kebersamaannya.

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala limpahan rahmat, taufik, dan hidayah-Nya, sehingga penulis dapat menyelesaikan penyusunan skripsi yang berjudul "Klasifikasi Citra Endoskopi Berbasis Arsitektur *ConvNeXt* Untuk Identifikasi Penyakit *Gerd* Dan Polip" dengan baik dan tepat waktu. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer (S.Kom) pada Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Penyusunan skripsi ini tentu bukanlah hal yang mudah. Prosesnya penuh dengan tantangan, mulai dari mencari ide penelitian, mengumpulkan data, melakukan analisis, hingga menuangkannya dalam bentuk tulisan ilmiah. Namun, berkat bimbingan, dukungan, dan doa dari berbagai pihak, akhirnya penulis dapat melewati setiap tahapan dengan penuh semangat dan ketekunan.

Penulis menyadari bahwa tersusunnya skripsi ini tidak lepas dari bantuan, bimbingan, dan dukungan dari banyak pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. Hj. Ilfi Nur Diana, M.Si., CAHRM., CRMP, selaku Rektor UIN Maulana Malik IBrahim.
2. Dr. H. Agus Mulyono, M.Si., selaku Dekan Fakultas Sains dan Teknologi.
3. Supriyono, M.Kom, selaku Ketua Program Studi Teknik Informatika.
4. Okta Qomaruddin Aziz, M.Kom selaku Pembimbing I dan Ajib Hanani, M.T selaku Pembimbing II, yang dengan kesabaran, ketelitian, dan dedikasi tinggi telah membimbing penulis dalam setiap tahap penyusunan skripsi ini.

Nasihat, arahan, serta ilmu yang diberikan sangat berarti dan menjadi penting dalam penyelesaian penelitian ini.

5. Prof. Dr. Suhartono, M.Kom selaku Ketua Penguji serta Shoffin Nahwa Utama, M.T selaku Anggota Penguji I, yang telah memberikan kritik, masukan, dan saran konstruktif demi menyempurnakan kualitas skripsi ini.
6. Seluruh dosen dan staf Program Studi Teknik Informatika, yang telah memberikan ilmu, pelayanan, dan dukungan administratif selama masa studi. Setiap bimbingan dan fasilitas yang diberikan memiliki kontribusi besar terhadap perkembangan akademik penulis.
7. Kedua orang tua dan keluarga tercinta, atas kasih sayang yang tidak pernah putus, doa yang selalu mengiringi, serta semangat yang diberikan dalam setiap langkah perjalanan pendidikan ini. Pengorbanan dan dukungan mereka menjadi kekuatan utama bagi penulis untuk terus melangkah hingga mencapai titik ini.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun demi perbaikan pada penelitian di masa mendatang. Semoga skripsi ini dapat memberikan manfaat dan menjadi kontribusi kecil bagi pengembangan ilmu pengetahuan, khususnya di bidang Teknik Informatika.

Malang, 15 Desember 2025

Penulis

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

ABSTRAK

Faqih, Muhammad. 2025. **Klasifikasi Citra Endoskopi Berbasis Arsitektur *ConvNeXt* Untuk Identifikasi Penyakit *Gerd* Dan Polip.** Skripsi. Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Okta Qomaruddin Aziz, M.Kom (II) Ajib Hanani, M.T.

.

Kata kunci: *Computer-Aided Diagnosis, ConvNeXt, Convolutional Neural Networks,* Klasifikasi Citra Endoskopi, Pencitraan Medis.

Identifikasi otomatis terhadap kelainan gastrointestinal penting untuk mendukung deteksi dini gastroesophageal reflux disease (*GERD*) dan polip usus. Interpretasi manual citra endoskopi memiliki keterbatasan karena variabilitas antar-pemeriksa dan waktu pemrosesan, sehingga diperlukan sistem diagnosis berbantuan komputer yang andal. Penelitian ini mengusulkan kerangka deep learning berbasis ConvNeXt-Tiny untuk mengklasifikasikan citra endoskopi ke dalam empat kategori, yaitu *GERD*, *GERD* Normal, Polyp, dan Polyp Normal. Dataset yang digunakan adalah GastroEndoNet v3, yang terdiri atas 4.006 citra asli dan 20.030 citra augmentasi. Sebanyak dua belas skenario eksperimen dilakukan untuk mengevaluasi pengaruh augmentasi data, normalisasi, dan ukuran *batch* terhadap kinerja model. Konfigurasi terbaik, yang menggunakan augmentasi aktif dan normalisasi berbasis ImageNet dengan *batch size* 64, mencapai akurasi sebesar 92,94% dan macro *F1-score* sebesar 92,94%. Hasil ini menunjukkan bahwa ConvNeXt-Tiny mampu mengekstraksi pola mukosa secara efektif dengan efisiensi komputasi yang tinggi, sehingga layak diterapkan pada lingkungan klinis. Kerangka yang diusulkan menyediakan baseline yang akurat dan ringan untuk klasifikasi penyakit endoskopi serta menjadi dasar bagi pengembangan lebih lanjut pada analisis video real-time dan validasi multi-senter.

ABSTRACT

Faqih, Muhammad. 2025. **Endoscopic Image Classification Based on ConvNeXt Architecture for Identification of Gerd and Polyps.** Undergraduate Thesis. Informatics Engineering Study Program, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Promotor: I) Okta Qomaruddin Aziz, M.Kom (II) Ajib Hanani, M.T.

Automated identification of gastrointestinal abnormalities is essential for supporting the early diagnosis of gastroesophageal reflux disease (GERD) and intestinal polyps. Manual interpretation of endoscopic images is limited by inter-observer variability and processing time, creating the need for reliable computer-aided diagnostic systems. This study proposes a ConvNeXt-Tiny based deep learning framework for multi-class classification of endoscopic images into four categories: GERD, GERD Normal, Polyp, and Polyp Normal. The experiments used the GastroEndoNet v3 dataset, which includes 4,006 original images and 20,030 augmented images. Twelve experimental scenarios were conducted to evaluate the effects of data augmentation, normalization, and batch size on model performance. The best configuration, which applied active augmentation and ImageNet-based normalization with a batch size of 64, achieved an accuracy of 92.94% and a macro F1-score of 92.94%. These results indicate that ConvNeXt-Tiny effectively captures fine-grained mucosal patterns while maintaining computational efficiency, making it suitable for clinical deployment. The proposed framework provides a lightweight and accurate baseline for automated endoscopic disease classification and forms a foundation for future work on real-time video analysis and multi-center validation.

Keywords: Classification, Computer-Aided Diagnosis, ConvNeXt, Convolutional Neural Networks, Endoscopic Image, Medical Imaging.

مستخلص البحث

فقيه، محمد .ألفان وخمسة وعشرون .تصنيف صور التنظير الداخلي بالاعتماد على معمارية كونفنيكست للتعرّف على مرض الارتجاع المعدي المريئي والسلائل .رسالة جامعية .قسم هندسة المعلوماتية، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم ،الإسلامية الحكومية مالانغ .المشرفان) :الأول (أوكتا قمر الدين عزيز، ماجستير في علوم الحاسوب،)الثاني (أجيب حناني .ماجستير في الهندسة

،الكلمات المفتاحية :التشخيص بمساعدة الحاسوب، كونفنيكست، الشبكات العصبية الالتفافية، تصنيف صور التنظير الداخلي .التصوير الطبي

تُعدّ عملية التعرّف الآلي على اضطرابات الجهاز الهضمي ذات أهمية كبيرة لدعم الكشف المبكر عن مرض الارتجاع المعدي المريئي وسلائل الأمعاء، حيث إن التفسير اليدوي لصور التنظير الداخلي يعاني من عدة قيود، من بينها اختلاف التقييم بين الفاحصين وطول زمن المعالجة، مما يستدعي الحاجة إلى نظام تشخيص موثوق قائم على الحاسوب .تقترح هذه الدراسة إطار عمل للتعلّم العميق يعتمد على نموذج كونفنيكست–تايني لتصنيف صور التنظير الداخلي إلى أربع فئات، وهي :الارتجاع المعدي المريئي، الارتجاع المعدي المريئي الطبيعي، السلائل، والسلائل الطبيعية .تم استخدام مجموعة بيانات غاستروإندونت الإصدار الثالث، التي تتكوّن من أربعة آلاف ،وست صور أصلية وعشرين ألفًا وثلاثين صورة ناتجة عن تقنيات تعزيز البيانات .أُجريت اثنتا عشرة تجربة لتقييم تأثير تعزيز البيانات والتطبيع، وحجم الدفعة على أداء النموذج .وحققت أفضل الإعدادات——باستخدام تعزيز بيانات فعّال وتطبيع قائم على إيميج نت .مع حجم دفعة قدره أربعة وستون——دقة بلغت اثنين وتسعين فاصل أربعة وتسعين في المائة، وقيمة إف واحد الكلية بالمقدار نفسه وتُظهر هذه النتائج أن نموذج كونفنيكست–تايني قادر على استخلاص أنماط الغشاء المخاطي بكفاءة عالية مع الحفاظ على كفاءة حسابية مرتفعة، مما يجعله مناسبًا للتطبيق في البيئات السريرية، كما يوفّر الإطار المقترح خطّ أساس دقيقًا وخفيف الوزن لتصنيف .أمراض التنظير الداخلي، ويشكّل أساسًا لتطويرات مستقبلية تشمل تحليل الفيديو في الزمن الحقيقي والتحقق متعدد المراكز

.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Penyakit *gastroesophageal reflux disease* (*GERD*) dan polip usus merupakan dua kondisi gastrointestinal yang umum terjadi dan memiliki potensi berkembang menjadi komplikasi serius jika tidak terdeteksi secara dini. *GERD* adalah kondisi kronis yang ditandai dengan naiknya isi lambung ke esofagus, menyebabkan gejala seperti nyeri epigastrium (*heartburn*), mual, regurgitasi asam, dan kesulitan menelan. Jika dibiarkan, *GERD* dapat menimbulkan komplikasi seperti esofagitis dan *Barrett's esophagus*, yang berisiko berkembang menjadi kanker esofagus (Syamsu Rijal *et al.*, 2024). Meskipun prevalensinya di Asia tergolong lebih rendah dibandingkan negara-negara Barat, tren kenaikannya cukup mengkhawatirkan; sebagai contoh, di Jepang dan Taiwan, prevalensi *GERD* mencapai 13–15% (Syamsu Rijal *et al.,* 2024).

Sementara itu, polip usus adalah pertumbuhan jaringan abnormal di dinding usus besar yang sering tidak menunjukkan gejala pada stadium awal. Namun, sekitar 35% kasus kanker kolorektal di Indonesia bermula dari polip yang tidak tertangani, dan sebagian besar menyerang usia produktif di bawah 40 tahun (Yasmina Lafau, 2024). Deteksi dini kedua kondisi ini menjadi krusial karena penanganan pada tahap awal terbukti lebih efektif dan dapat mencegah transisi ke fase penyakit yang lebih berat, terutama dalam sistem pelayanan kesehatan yang masih menghadapi tantangan seperti keterbatasan tenaga ahli.

Endoskopi merupakan prosedur medis minimal invasif yang memungkinkan visualisasi langsung saluran pencernaan bagian atas maupun bawah menggunakan kamera fleksibel, sehingga menjadi alat utama dalam diagnosis berbagai penyakit gastrointestinal, termasuk *GERD* dan polip. Dari hasil prosedur ini, diperoleh citra endoskopi yang mengandung informasi visual krusial mengenai kondisi mukosa saluran cerna. Namun, interpretasi citra endoskopi secara manual memiliki keterbatasan signifikan, seperti ketergantungan pada pengalaman klinis dokter, variabilitas subjektif antarpengamat, serta risiko *human error* yang dapat menyebabkan diagnosis kurang akurat atau tertunda (Zhou *et al.*, 2023). Kompleksitas visual, seperti perbedaan warna jaringan yang halus, pencahayaan yang tidak merata, dan keberadaan artefak citra, menambah tantangan dalam penilaian manual (Zhou *et al.*, 2023).

Berkenaan dengan hal ini, teknologi berbasis kecerdasan buatan (*artificial intelligence* / *AI*), khususnya model *deep learning* untuk klasifikasi citra, menawarkan solusi inovatif. Dengan melatih model pada ribuan gambar endoskopi, sistem *AI* mampu mengenali pola-pola patologis secara konsisten dan dalam waktu yang lebih singkat dibandingkan tenaga medis manusia, sehingga berpotensi meningkatkan akurasi diagnosis dan efisiensi layanan kesehatan.

Kecerdasan buatan, khususnya *deep learning* berbasis *convolutional neural network* (*CNN*) seperti *ResNet* atau *EfficientNet*, menawarkan potensi untuk mengatasi tantangan ini melalui analisis citra endoskopi otomatis. Namun, model *CNN* tradisional seperti *ResNet* atau *EfficientNet* sering kali kurang efisien untuk diimplementasikan pada perangkat dengan spesifikasi rendah, yang umum

ditemukan di rumah sakit kecil (Tan & Le, 2019). Untuk mengatasi masalah ini, penelitian ini menggunakan *ConvNeXt*, sebuah arsitektur *CNN* modern yang menggabungkan efisiensi *CNN* dengan prinsip desain *Vision Transformer*. *ConvNeXt* merupakan arsitektur konvolusional modern yang menunjukkan performa tinggi dan efisiensi komputasi yang baik pada berbagai *benchmark* visi komputer umum, seperti klasifikasi *ImageNet*, deteksi objek, dan segmentasi semantik (Liu *et al.,* 2022).

Hasil *benchmark* pada *ImageNet-1K/22K* menunjukkan bahwa berbagai varian *ConvNeXt* secara konsisten melampaui performa *ResNet* dan bersaing dengan atau bahkan melampaui *Vision Transformer* (misalnya *Swin-B*), dengan akurasi lebih tinggi dan *throughput inference* yang lebih baik, sambil tetap mempertahankan efisiensi parameter dan arsitektur *CNN* yang sederhana (Liu *et al.*, 2022). Keunggulan ini menjadikan *ConvNeXt* sangat relevan untuk aplikasi medis, di mana presisi tinggi dibutuhkan namun dalam konteks operasional yang efisien (Esteva *et al., 2021*). Selain itu, struktur *ConvNeXt* yang modular dan tidak terlalu kompleks memungkinkan proses *fine-tuning* lebih fleksibel pada *dataset* medis berskala sedang hingga kecil, sehingga cocok diterapkan untuk sistem klasifikasi otomatis di fasilitas kesehatan (Shin *et al.,* 2016).

Penelitian ini akan menggunakan *dataset* citra endoskopi *GastroEndoNet: Comprehensive Endoscopy Image dataset for GERD and Polyp Detection*, yang tersedia di *Mendeley Data* (Bitto *et al.*, 2025), untuk melatih model *ConvNeXt-Tiny* dalam mendeteksi *GERD* dan polip usus. Performa model akan dievaluasi menggunakan metrik-metrik yang umum digunakan dalam klasifikasi citra medis,

yaitu matriks konfusi (menggambarkan distribusi prediksi benar dan salah untuk setiap kelas), akurasi (persentase prediksi benar dari total data uji), sensitivitas (*recall*, kemampuan mendeteksi kondisi positif seperti *GERD* atau polip), spesifisitas (kemampuan mengenali kondisi normal tanpa salah melabeli sebagai penyakit), dan *F1-score* (keseimbangan antara *precision* dan *recall*) (Esteva *et al.*, 2017).

Penelitian ini bertujuan mengimplementasikan sistem klasifikasi citra endoskopi berbasis *ConvNeXt-Tiny* yang akurat dan efisien, serta mengevaluasi model untuk mendeteksi *GERD* dan polip usus, mendukung diagnosis di rumah sakit tipe C di Indonesia. Dengan demikian, penelitian ini diharapkan dapat meningkatkan efisiensi dan akurasi diagnosis, terutama di wilayah dengan akses terbatas ke spesialis gastroenterologi. Upaya ini sejalan dengan nilai-nilai dalam Islam yang mendorong pencegahan penyakit dan penyelamatan jiwa manusia sebagai bentuk *ihsan* (kebaikan). Allah SWT berfirman:

يٰٓاَيُّهَا الَّذِيْنَ اٰمَنُوْا لَا تُحِلُّوْا شَعَاۤىِٕرَ اللّٰهِ وَلَا الشَّهْرَ الْحَرَامَ وَلَا الْهَدْيَ وَلَا الْقَلَاۤىِٕدَ وَلَآ اٰۤمِّيْنَ الْبَيْتَ الْحَرَامَ يَبْتَغُوْنَ فَضْلًا مِّنْ رَّبِّهِمْ وَرِضْوَانًاۗ وَاِذَا حَلَلْتُمْ فَاصْطَادُوْا ۗ وَلَا يَجْرِمَنَّكُمْ شَنَاٰنُ قَوْمٍ اَنْ صَدُّوْكُمْ عَنِ الْمَسْجِدِ الْحَرَامِ اَنْ تَعْتَدُوْا ۘ وَتَعَاوَنُوْا عَلَى الْبِرِّ وَالتَّقْوٰىۖ وَلَا تَعَاوَنُوْا عَلَى الْاِثْمِ وَالْعُدْوَانِۖ وَاتَّقُوا اللّٰهَۗ اِنَّ اللّٰهَ شَدِيْدُ الْعِقَابِ

"Wahai orang-orang yang beriman! Janganlah kamu melanggar syiar-syiar kesucian Allah, dan jangan (melanggar kehormatan) bulan-bulan haram, jangan (mengganggu) hadyu (hewan-hewan kurban) dan qala'id (hewan-hewan kurban yang diberi tanda), dan jangan (pula) mengganggu orang-orang yang mengunjungi Baitulharam; mereka mencari karunia dan keridhaan Tuhannya. Tetapi apabila kamu telah menyelesaikan ihram, maka bolehlah kamu berburu. Jangan sampai kebencian(mu) kepada suatu kaum karena mereka menghalang-halangimu dari Masjidilharam, mendorongmu berbuat melampaui batas (kepada mereka). Dan tolong-menolonglah kamu dalam (mengerjakan) kebajikan dan takwa, dan jangan tolong-menolong dalam berbuat dosa dan permusuhan. Bertakwalah kepada Allah, sungguh, Allah sangat berat siksaan-Nya." (Q.S. Al-Maidah: 2)

Menurut Tafsir Ibn Kathīr (2008), ayat ini menegaskan kewajiban umat Islam untuk saling bekerja sama dalam kebajikan dan takwa, serta menjauhi kerja sama dalam dosa dan permusuhan. Prinsip ini dapat diaktualisasikan dalam konteks modern, misalnya dengan bekerja sama dalam menjaga kesehatan dan menghindarkan diri dari *mudharat* melalui pengembangan sistem medis berbasis teknologi. Selain itu, terdapat pula sabda Nabi Muhammad ﷺ yang mengandung motivasi untuk berikhtiar dalam menemukan solusi medis bagi setiap penyakit:

ما أَنْزَلَ اللَّهُ داءً إِلَّا أَنْزَلَ له شِفاءً

"Tidaklah Allah menurunkan suatu penyakit, melainkan Dia juga menurunkan obatnya." (HR. Bukhari, no. 5678)

Dalam penjelasan Ibn Ḥajar al-'Asqalānī, hadis ini menunjukkan anjuran untuk melakukan penelitian dan pengobatan, sebab setiap penyakit pasti memiliki solusi yang diciptakan Allah, hanya saja sebagian telah ditemukan dan sebagian lain menunggu untuk diikhtiarkan (Ibn Ḥajar al-'Asqalānī, 2001).

Dengan dasar tersebut, penelitian ini tidak hanya merupakan kontribusi ilmiah dalam bidang teknologi medis, tetapi juga menjadi bentuk pengabdian dalam kerangka etika Islam untuk menjaga kesehatan sebagai bagian dari amanah yang wajib dijaga.

1.2 Rumusan Masalah

Bagaimana membangun dan mengevaluasi model klasifikasi citra endoskopi menggunakan arsitektur *ConvNeXt-Tiny* untuk mendeteksi penyakit

GERD dan polip usus secara otomatis berdasarkan metrik matriks konfusi, akurasi, sensitivitas, dan spesifisitas?

1.3 Batasan Masalah

a. Penelitian ini hanya menggunakan *dataset* citra endoskopi dari GastroEndoNet
b. Deteksi terbatas pada dua jenis kondisi gastrointestinal, yaitu *GERD (gastroesophageal reflux disease)* dan polip usus, tanpa mencakup jenis kelainan lain seperti kanker kolorektal lanjut atau gastritis.
c. Arsitektur yang digunakan terbatas pada *ConvNeXt-Tiny*, tanpa perbandingan dengan arsitektur lain seperti *ResNet, EfficientNet*, atau *ViT*.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk membangun model klasifikasi citra endoskopi menggunakan arsitektur *ConvNeXt-Tiny* dan mengevaluasi performa model berdasarkan metrik akurasi, sensitivitas, spesifisitas, dan matriks konfusi untuk mendeteksi penyakit *GERD* dan polip usus secara otomatis.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat baik secara teoritis maupun praktis, sebagai berikut:

1. Penelitian ini memberikan kontribusi pada pengembangan ilmu pengetahuan di bidang kecerdasan buatan, khususnya dalam penerapan arsitektur *ConvNeXt-Tiny* untuk klasifikasi citra medis.

2. Hasil dari penelitian ini diharapkan dapat mendukung sistem pendukung keputusan medis (*clinical decision support system*), khususnya dalam membantu dokter atau tenaga medis mendiagnosis *GERD* dan polip usus secara lebih cepat dan akurat.

BAB II

STUDI PUSTAKA

2.1 Penelitian Terkait

Dalam beberapa tahun terakhir, pemanfaatan *deep learning*, khususnya *convolutional neural networks* (*CNN*), telah menjadi pendekatan utama dalam klasifikasi citra medis, termasuk dalam analisis citra endoskopi. Deteksi dini penyakit seperti kanker kolorektal maupun abnormalitas gastrointestinal lainnya sangat bergantung pada kemampuan sistem untuk mengenali keberadaan polip atau lesi pada dinding saluran cerna. Sebelumnya, metode deteksi polip banyak bergantung pada fitur buatan (*handcrafted features*) seperti tekstur dan warna. Namun, pendekatan tersebut memiliki keterbatasan, terutama ketika menghadapi variasi bentuk, ukuran, dan pencahayaan dari citra endoskopi.

Seiring berkembangnya *CNN* dan arsitektur mutakhir seperti *YOLO*, *ResNet*, dan *UNet*, performa deteksi dan segmentasi menjadi semakin akurat dan cepat, membuka jalan bagi penerapan sistem *computer-aided diagnosis* (*CAD*) secara *real-time* dalam praktik klinis. Salah satu studi penting dilakukan oleh Jha *et al.* (2021) yang memperkenalkan *ColonSegNet*, sebuah arsitektur *encoder-decoder* yang dirancang untuk segmentasi dan deteksi polip dalam citra kolonoskopi secara *real-time*. Penelitian ini menggunakan *dataset* publik *Kvasir-SEG* dengan 1.072 citra yang telah dianotasi. Melalui *benchmarking* terhadap berbagai arsitektur seperti *YOLOv4*, *RetinaNet*, dan *Faster R-CNN*, *ColonSegNet* menunjukkan performa kompetitif dengan *Dice coefficient* sebesar 0,8206 dan kecepatan

inference mencapai 180 *FPS*, menjadikannya salah satu metode tercepat yang mendekati kebutuhan klinis.

Pendekatan berbeda diusulkan oleh Cao *et al.* (2021), yang fokus pada deteksi polip lambung suatu tantangan tersendiri karena bentuk dan ukuran polip yang lebih kecil serta tekstur mukosa lambung yang kompleks. Dengan memodifikasi *YOLOv3* dan menambahkan modul fusi fitur, mereka berhasil meningkatkan presisi deteksi menjadi 91,6% dan *F1-score* hingga 88,8%, jauh mengungguli *baseline*. Meskipun *dataset* yang digunakan bersifat privat dan khusus untuk polip lambung, studi ini memberikan kontribusi penting dalam hal penanganan objek kecil dan penggabungan fitur multi-level, yang sangat relevan dalam konteks klasifikasi citra endoskopi gastrointestinal atas (Cao *et al.*, 2021; Jha *et al.*, 2021).

Penelitian oleh Chan *et al.* (2023) berfokus pada klasifikasi multi-kelas penyakit *gastroesophageal reflux disease* (*GERD*) menggunakan citra endoskopi berbasis *white light* (*WL*). Dengan menggunakan *dataset* internal dari *Xiangyang Centre Hospital* yang terdiri atas 3.654 citra endoskopi, mereka membagi data berdasarkan klasifikasi *Los Angeles* (*LACS*) menjadi empat kelas: A, B, C, dan D. Metode yang digunakan mencakup pemanfaatan model *CNN pre-trained* seperti *DenseNet121*, *ResNet*, hingga *InceptionResNet*, yang kemudian dikombinasikan dengan teknik *data resampling* dan *attention map*.

Model terbaik diperoleh dari kombinasi *DenseNet121* dengan *oversampling* dan *global attention block* (*GAB*), yang menghasilkan akurasi sebesar 74,69%, *F1-score* sebesar 69,87%, dan *Cohen's kappa* 0,7757. Selain pengembangan model,

peneliti juga merancang antarmuka pengguna berbasis *HuggingFace* untuk memperkuat adopsi klinis. Keunggulan dari studi ini terletak pada pendekatan sistematis terhadap klasifikasi *GERD* secara detail berdasarkan skala *LACS* penuh, serta upaya untuk menangani *imbalance data* dan meningkatkan interpretabilitas melalui *attention map* (Chan *et al.,* 2023).

Perkembangan ini mengindikasikan adanya pergeseran dari pendekatan *CNN* tradisional ke model-model berbasis *Transformer*, terutama dalam aplikasi citra medis yang kompleks. Selain itu, fokus penelitian mulai bergeser dari klasifikasi umum menuju kasus-kasus yang lebih spesifik, seperti *GERD* dan polip, yang merupakan dua kategori penting dalam diagnostik gastrointestinal atas dan bawah (Chan *et al.*, 2023).

Seiring berkembangnya teknologi *deep learning*, pendekatan berbasis arsitektur yang lebih modern mulai dikembangkan untuk mengatasi tantangan klasifikasi citra medis yang kompleks. Salah satu penelitian oleh Li *et al.* (2023) mengusulkan pendekatan hibrida *ConvNeXt* dan *Vision Transformer* (*ViT*) dalam sistem diagnosis berbantuan komputer untuk lesi kulit akibat infeksi virus. Penelitian ini menggunakan model gabungan *ConvNeXt-Small* dan *Swin-T*, serta diterapkan pada *dataset* baru bernama *Skin-CID*, yang mencakup berbagai penyakit kulit termasuk *poxvirus*. Sistem ini dilatih secara *end-to-end* menggunakan strategi penyeimbangan data dan *augmentation* berbasis warna serta tekstur. Hasil evaluasi menunjukkan bahwa model gabungan *ConvNeXt-Small* dengan *Swin-T* berhasil mencapai akurasi tertinggi sebesar 96,03% dan *F1-score* 94,27%, mengungguli model *baseline* seperti *ResNet* dan *DenseNet*. Pendekatan ini membuktikan bahwa

kombinasi *CNN* dan *Transformer* mampu memanfaatkan fitur global dan lokal secara sinergis untuk meningkatkan akurasi diagnosis (Li *et al.*, 2023).

Penelitian terbaru oleh Nergiz (2023) mengkaji penerapan *ConvNeXt* pada klasifikasi lesi kolorektal menggunakan *dataset* MHIST yang berisi 3.152 tile citra histopatologi berlabel *hyperplastic polyp (HP)* dan *sessile serrated adenoma (SSA)*. Peneliti mengevaluasi empat varian *ConvNeXt* (Tiny, Small, Big, Large) pada tiga skenario: full data, k-shot learning, dan gradually increasing difficulty. Hasilnya menunjukkan *ConvNeXt*-L mencapai akurasi 88,9%, *F1-score* 91,21%, AUC 93,91%, serta Cohen's kappa 0,7633, mengungguli arsitektur CNN konvensional seperti ResNet, DenseNet, dan Inception v3. Selain itu, eksperimen few-shot menunjukkan kemampuan generalisasi *ConvNeXt* tetap tinggi meskipun data terbatas, menjadikannya baseline yang menjanjikan untuk tugas klasifikasi medis di domain data terbatas (Nergiz, 2023).

Pendekatan berbasis arsitektur *ConvNeXt* juga diuji secara spesifik dalam konteks klasifikasi penyakit kulit yang menyerupai gejala virus, seperti *monkeypox*. Huan dan Dun (2024) memperkenalkan *MSMP-Net*, sebuah model *multi-scale neural network* yang dibangun di atas *backbone ConvNeXt* untuk klasifikasi lesi kulit akibat virus *monkeypox*. *MSMP-Net* menggabungkan fitur multiskala dari *ConvNeXt* dengan desain *inverse bottleneck* dan *large kernel* untuk meningkatkan ekstraksi fitur spasial.

Penelitian ini menggunakan *dataset MSLD v2.0* dan berhasil mencapai akurasi 87,03%, *F1-score* 86,58%, serta efisiensi tinggi dalam *pipeline end-to-end*. Inovasi utama terletak pada struktur fusi fitur multiskala yang mampu menangkap

perbedaan morfologis halus pada citra kulit. Penelitian ini menandai langkah maju dalam pemanfaatan *ConvNeXt* untuk klasifikasi citra medis nonendoskopik, serta membuka peluang adopsi pada domain yang lebih kompleks seperti endoskopi saluran pencernaan.

Rangkuman dari seluruh penelitian yang telah dibahas dapat dilihat pada Tabel 2.1, yang memuat perbandingan metode, *dataset*, dan hasil utama dari tiap studi.

Tabel 2.1 Ringkasan Penelitian Terkait

No	Peneliti	Tujuan Penelitian	*Dataset*	Metode	Hasil Utama	Kelebihan	Keterbatasan
1	Jha *et al.* (2021)	Deteksi dan segmentasi polip secara real-time	Kvasir-SEG	YOLOv4, ColonSegNet	Dice 0.82; 180 FPS	Real-time, benchmark lengkap	Fokus polip kolon saja
1	Cao *et al.* (2021)	Deteksi polip lambung ukuran kecil	*Dataset* internal	YOLOv3 + feature fusion	*Precision* 91.6%	Deteksi objek kecil efektif	*Dataset* privat, domain sempit
2	Chan *et al.* (2023)	Klasifikasi *GERD* multi-kelas (LACS A–D)	Xiangyang Hosp. (3654 citra)	DenseNet121 + GAB	Akurasi 74.69 %	Fokus *GERD*, interpretabilitas baik	Imbalanced data
4	Li *et al.* (2023)	Klasifikasi infeksi kulit (pox, herpes)	Skin-CID	*ConvNeXt* + Swin-T	Akurasi 96.03 %	Hybrid CNN-ViT, klasifikasi luas	Domain non-endoskopi
5	Nergiz (2023)	*ConvNeXt* dalam klasifikasi lesi kolorektal (SSA vs HP)	MHIST (3.152 tile citra histopatologi)	*ConvNeXt* (Tiny, Small, Big, Large)	Akurasi 88,9%, *F1-score* 91,21 %, AUC 93,91 %, Cohen's	Mengunggul CNN tradisional, kuat pada few-shot learning, hasil stabil pada data sulit	Fokus pada citra histopatologi (bukan endoskopi)

No	Peneliti	Tujuan Penelitian	*Dataset*	Metode	Hasil Utama	Kelebihan	Keterbatasan
					kappa 0,7633		
6	Huan & Dun (2024)	Klasifikasi lesi monkeypox	MSLD v2.0	MSMP-Net (*ConvNeXt* base)	Akurasi 87.03 %	Multi-scale, efisien end-to-end	Bukan *GERD*/polip, non-endoskopi

Berdasarkan enam penelitian yang telah dibahas, terlihat adanya perkembangan signifikan dalam pemanfaatan metode *deep learning* untuk klasifikasi citra medis, khususnya dalam domain endoskopi dan penyakit kulit. Dimulai dari pendekatan *CNN* konvensional yang berfokus pada segmentasi polip atau klasifikasi *GERD*, hingga munculnya model *hybrid* dan *backbone* modern seperti *ConvNeXt* dan *Vision Transformer*, tren penelitian menunjukkan upaya berkelanjutan untuk meningkatkan akurasi, efisiensi, dan generalisasi model.

Sebagian besar penelitian masih terbatas pada klasifikasi satu jenis penyakit secara spesifik, seperti polip atau *GERD* saja, dan belum banyak yang menggabungkan klasifikasi *multi-label* dalam domain endoskopi gastrointestinal. Selain itu, penggunaan *ConvNeXt* dalam domain endoskopi masih minim ditemukan dalam literatur yang ada. Oleh karena itu, penelitian ini hadir dengan kontribusi utama berupa penerapan *ConvNeXt-Tiny* sebagai arsitektur *backbone* untuk klasifikasi citra endoskopi *GERD* dan polip secara bersamaan. Dengan pendekatan ini, penelitian diharapkan dapat menjawab celah riset berupa kebutuhan model ringan, efisien, dan akurat untuk diagnosis *multi-label* berbasis endoskopi. Tabel 2.1 juga menunjukkan bagaimana pendekatan ini berada pada posisi strategis

untuk menjembatani kekosongan antara pendekatan klasik dan arsitektur modern dalam klasifikasi citra medis.

2.2 *Gastroesophageal reflux disease (GERD)*

Gastroesophageal reflux disease (*GERD*) merupakan gangguan saluran cerna kronis yang ditandai dengan refluks isi lambung ke esofagus, yang menimbulkan gejala seperti nyeri epigastrium (*heartburn*), regurgitasi asam, dan disfagia (Rijal *et al.*, 2024). Meskipun prevalensi *GERD* di Asia lebih rendah dibandingkan negara-negara Barat, tren peningkatannya terus berkembang seiring perubahan gaya hidup masyarakat.

Studi oleh Rijal *et al.* (2024) di RS Ibnu Sina Makassar mengungkap bahwa mayoritas pasien *GERD* adalah perempuan usia dewasa muda (20–44 tahun) dengan indeks massa tubuh normal, memiliki riwayat gastritis, dan berprofesi sebagai ibu rumah tangga. Faktor risiko signifikan meliputi obesitas, pola makan tidak sehat, stres, serta kebiasaan seperti merokok dan konsumsi makanan asam atau pedas.

Endoskopi berperan penting dalam diagnosis *GERD*, khususnya dalam mendeteksi komplikasi seperti *Barrett's esophagus*, suatu lesi praneoplastik yang dapat berkembang menjadi adenokarsinoma esofagus jika tidak ditangani. Oleh karena itu, deteksi dini menjadi krusial guna mencegah komplikasi dan mengoptimalkan manajemen klinis, terutama mengingat data yang menunjukkan *GERD* mulai menjangkiti kelompok usia produktif dan berdampak pada kualitas hidup serta beban sistem layanan kesehatan (Rijal *et al.*, 2024).

2.3 Polip Usus

Polip usus merupakan pertumbuhan abnormal pada lapisan mukosa saluran cerna, terutama pada kolon dan rektum, yang dapat bersifat neoplastik maupun non-neoplastik, dengan potensi transformasi menjadi kanker kolorektal, terutama pada jenis adenomatosa dan polip *serrated*. Deteksi dan reseksi dini polip merupakan langkah penting dalam mencegah perkembangan kanker kolorektal, yang secara global merupakan penyebab kematian ketiga tertinggi akibat kanker (Jha *et al.*, 2021).

Meskipun prosedur kolonoskopi telah menjadi metode standar diagnosis dan terapi melalui visualisasi langsung dan polipektomi, keterbatasan berupa tingkat kesalahan deteksi yang mencapai 20% masih menjadi tantangan, khususnya pada polip berukuran kecil dan datar (Cao *et al.*, 2021). Dalam hal ini, pendekatan berbasis *deep learning*, seperti *ColonSegNet* dan model berbasis *YOLOv3*, telah menunjukkan kinerja menjanjikan untuk segmentasi dan deteksi *real-time* citra endoskopi, dengan akurasi tinggi serta kemampuan mendeteksi polip multipel dalam satu gambar sekaligus (Jha *et al.*, 2021; Cao *et al.*, 2021).

Sistem berbasis *CNN* seperti *GastroNet* juga menunjukkan akurasi validasi mencapai 99,2% dalam klasifikasi berbagai kelainan gastrointestinal, termasuk polip, dari citra endoskopi kapsul, memperkuat relevansi integrasi teknologi kecerdasan buatan dalam upaya deteksi dini polip usus secara otomatis dan efisien (Rajkumar *et al.*, 2024). Oleh karena itu, pengembangan sistem klasifikasi citra berbasis *deep learning* tidak hanya mendukung diagnosis yang lebih cepat dan objektif, tetapi juga menjadi fondasi penting dalam sistem *Computer-Aided*

Diagnosis (*CADx*) yang berpotensi meningkatkan kualitas layanan medis gastroenterologi secara luas.

2.4 *Convolutional Neural Network*

Convolutional Neural Network (*CNN*) merupakan salah satu arsitektur jaringan saraf tiruan yang paling dominan dalam bidang visi komputer, terutama untuk tugas-tugas klasifikasi citra. *CNN* dirancang untuk mengenali pola spasial dalam data *grid* seperti citra dua dimensi (*2D*), dengan mengandalkan proses pembelajaran fitur secara otomatis melalui lapisan-lapisan *convolution*, *pooling*, dan *fully connected*. *CNN* memiliki keunggulan karena mampu mengekstraksi fitur lokal secara hierarkis, mulai dari tepi dan tekstur pada lapisan awal hingga bentuk kompleks pada lapisan yang lebih dalam (Nadachowski *et al.*, 2024; Zhao *et al.*, 2020).

Arsitektur dasar *CNN* terdiri dari beberapa komponen utama. Pertama adalah lapisan *convolution*, yang berfungsi menerapkan *kernel* atau *filter* untuk mengekstraksi fitur dari data masukan. Proses ini menghasilkan *feature maps* yang mewakili pola-pola penting dalam citra. Setelah itu, fungsi aktivasi seperti *ReLU* diaplikasikan untuk menambahkan non-linearitas. Kemudian terdapat lapisan *pooling*, umumnya menggunakan metode *max pooling*, yang berfungsi mereduksi dimensi spasial dan menjaga informasi dominan. Lapisan ini juga membantu mengurangi kompleksitas komputasi dan mencegah *overfitting*. Di akhir jaringan, lapisan *fully connected* digunakan untuk pengambilan keputusan klasifikasi, yang sering kali diakhiri dengan fungsi *Softmax* untuk menghitung probabilitas antar kelas (Nadachowski *et al.*, 2024; Zhao *et al.*, 2020).

Dalam pemrosesan citra, *CNN* bekerja dengan cara menerima masukan berupa representasi matriks (seperti citra atau data *DEM*), kemudian secara bertahap mengekstraksi fitur melalui lapisan-lapisan konvolusional. Fitur-fitur ini kemudian digabungkan dalam lapisan *fully connected* untuk menghasilkan keluaran klasifikasi. Proses ini memungkinkan *CNN* untuk belajar langsung dari data mentah tanpa memerlukan rekayasa fitur manual yang kompleks, sebagaimana terlihat pada penerapan *CNN* untuk klasifikasi spektrum *EEG* (Ajra *et al.*, 2022) dan struktur sekunder protein (Zhao *et al.*, 2020).

Salah satu keunggulan utama *CNN* adalah kemampuannya untuk mengatasi keterbatasan pendekatan klasifikasi konvensional dalam pengolahan citra medis, seperti ketergantungan pada fitur yang direkayasa secara manual dan sensitivitas terhadap *noise*. *CNN* dapat mengidentifikasi pola penting bahkan dari citra yang kompleks, sehingga sangat sesuai untuk klasifikasi citra endoskopi yang memiliki variabilitas tekstur tinggi dan perbedaan morfologi halus, seperti pada kasus deteksi *GERD* dan polip. *CNN* juga menawarkan pendekatan yang lebih objektif dan reprodusibel karena tidak tergantung pada subjektivitas interpretasi manusia (Nadachowski *et al.*, 2024; Prommakhot & Srinonchat, 2024).

CNN menjadi fondasi bagi pengembangan arsitektur modern seperti *VGG*, *ResNet*, dan *ConvNeXt*. *VGG*, misalnya, memperkenalkan desain arsitektur yang dalam namun sederhana dengan *filter convolution* 3×3 yang berulang, membuktikan pentingnya kedalaman jaringan dalam meningkatkan akurasi klasifikasi (Nadachowski *et al.*, 2024). Sedangkan *ResNet* menambahkan *residual connection* untuk mengatasi masalah *vanishing gradient* pada jaringan yang sangat

dalam. Arsitektur-arsitektur ini menjadi dasar evolusi ke arah model yang lebih efisien dan canggih, termasuk *ConvNeXt* yang menggabungkan kekuatan *CNN* dengan prinsip desain modern dari *Vision Transformer*. Dengan demikian, memahami *CNN* merupakan langkah awal penting untuk mengaplikasikan arsitektur seperti *ConvNeXt* secara efektif dalam tugas klasifikasi citra medis endoskopi.

2.5 *ConvNeXt*

Kemunculan *ConvNeXt* merupakan respons terhadap dominasi *Vision Transformer* (*ViT*) dalam bidang pengenalan citra visual sejak tahun 2020-an. Meskipun arsitektur konvolusional klasik seperti *ResNet* dan *VGGNet* telah membentuk fondasi kuat untuk visi komputer selama lebih dari satu dekade, pengenalan *ViT* menantang asumsi bahwa konvolusi adalah strategi terbaik untuk pembelajaran fitur visual. *Vision Transformer* menghadirkan keunggulan dalam memodelkan konteks global melalui *self-attention*, namun sering kali mengabaikan *bias* induktif lokal yang menjadi kekuatan utama *ConvNet*. Untuk menjembatani kesenjangan tersebut, Liu *et al.* (2022) mengusulkan *ConvNeXt*, sebuah keluarga arsitektur *ConvNet* yang dimodernisasi dengan mengadopsi prinsip desain dari *Transformer* tanpa sepenuhnya mengabaikan struktur konvolusional tradisional. Tujuannya adalah membuktikan bahwa *ConvNet* murni, dengan pembaruan arsitektur yang cermat, masih mampu bersaing dengan *Transformer* dalam akurasi dan efisiensi.

Arsitektur *ConvNeXt* dibangun di atas struktur *ResNet*, tetapi dimodifikasi secara sistematis untuk meniru perilaku arsitektur *Transformer* sambil tetap

mempertahankan sifat konvolusional. Model ini mempertahankan arsitektur bertingkat (*multi-stage*) dengan resolusi fitur yang menurun secara bertahap, mirip dengan *ConvNet* klasik. Namun, *ConvNeXt* memperkenalkan elemen-elemen baru seperti penggunaan *depthwise convolution*, normalisasi *layer* (*layer normalization*), aktivasi *GELU*, serta *block-block* dengan *inverted bottleneck*. Desain ini memungkinkan *ConvNeXt* untuk menangkap fitur spasial lokal secara efisien sambil juga memperluas jangkauan konteks spasial melalui *kernel* besar. Selain itu, *ConvNeXt* mengadopsi teknik pelatihan modern seperti *optimizer AdamW*, strategi augmentasi data yang intensif, dan regularisasi berbasis *stochastic depth* yang telah terbukti efektif pada *Transformer*.

Detail arsitektur *ConvNeXt* mencakup beberapa komponen kunci. Pertama, penggunaan *depthwise convolution*, teknik yang memungkinkan pemisahan pemrosesan antar kanal sehingga mengurangi jumlah parameter secara signifikan. Teknik ini telah digunakan sebelumnya pada arsitektur seperti *MobileNet*, dan dalam *ConvNeXt* dipadukan dengan *1×1 convolution* untuk memungkinkan pemrosesan spasial dan kanal secara terpisah. Kedua, *ConvNeXt* menerapkan *inverted bottleneck*, yaitu konfigurasi *layer* di mana dimensi fitur diperluas sebelum kembali dipersempit. Pendekatan ini meningkatkan kapasitas representasi tanpa memperbesar jumlah parameter secara drastis. Ketiga, *ConvNeXt* menggantikan aktivasi *ReLU* dengan *GELU*, yang menunjukkan peningkatan performa dalam konteks pembelajaran mendalam.

ConvNeXt juga mengadopsi *kernel convolution* berukuran besar seperti 7×7, berbeda dari pendekatan tradisional yang menggunakan *stacking kernel* kecil

seperti 3×3. Penggunaan *kernel* besar ini memberikan jangkauan reseptif yang lebih luas dan mendekati karakteristik *global attention* pada *Transformer*. *Layer* normalisasi dalam *ConvNeXt* menggunakan *LayerNorm*, berbeda dari *BatchNorm* yang umum dalam *CNN* konvensional, untuk menstabilkan pelatihan terutama ketika ukuran *batch* kecil. Kombinasi dari desain mikro dan makro ini menciptakan blok *ConvNeXt* yang efisien namun tetap kuat dalam ekstraksi fitur, baik untuk citra alami maupun citra medis.

Dibandingkan dengan *CNN* konvensional seperti *ResNet50*, *EfficientNet*, dan *DenseNet*, *ConvNeXt* menunjukkan peningkatan akurasi yang signifikan dalam berbagai tugas klasifikasi dan segmentasi citra. Studi oleh Emegano *et al.* (2025) menunjukkan bahwa *ConvNeXt* mencapai akurasi 98% dalam klasifikasi kanker prostat multikelas, melampaui *ResNet50* (93%) dan bahkan *Swin Transformer* (95%). Keunggulan ini diperoleh tidak hanya dari desain arsitektur yang modern, tetapi juga dari kombinasi efisiensi komputasi dan kemampuannya dalam mengekstraksi fitur lokal maupun global secara efektif. Dalam domain citra medis, sifat tersebut sangat penting mengingat keterbatasan data berlabel dan kebutuhan akan model yang dapat diinterpretasikan.

ConvNeXt dikembangkan dalam beberapa varian yang berbeda dalam jumlah kanal fitur, kedalaman blok, serta kapasitas komputasi. Varian varian ini dirancang agar *ConvNeXt* dapat digunakan pada berbagai skenario, mulai dari aplikasi ringan hingga kebutuhan komputasi besar. Perbedaan utama mencakup jumlah kanal pada tiap stage dan jumlah blok yang digunakan. Liu *et al.* (2022) mendefinisikan lima varian utama yaitu *ConvNeXt Tiny, Small, Base, Large, dan Extra Large.* Semua

varian mempertahankan struktur bertingkat empat *stage* dengan resolusi fitur yang semakin menurun di setiap *stage*, tetapi memiliki kapasitas pemrosesan yang berbeda.

A) ConvNeXt Tiny (ConvNeXt T)

ConvNeXt Tiny merupakan varian dengan kapasitas paling kecil. Model ini menggunakan konfigurasi kanal berturut turut sebanyak 96, 192, 384, dan 768 pada empat *stage,* dengan jumlah blok masing masing 3, 3, 9, dan 3. Total parameter model adalah sekitar 29 juta dengan kebutuhan komputasi 4.5 *GFLOPs.* Meskipun ringan, *ConvNeXt Tiny* mencapai akurasi 82.1 persen pada *ImageNet 1K*. Varian ini cocok digunakan pada skenario dengan sumber daya komputasi terbatas seperti klasifikasi citra endoskopi. Informasi konfigurasi ini dilaporkan oleh Liu *et al.* dalam hasil eksperimen klasifikasi *ImageNet* (2022).

B) ConvNeXt Small (ConvNeXt S)

Varian *Small* memiliki jumlah kanal yang sama dengan *ConvNeXt Tiny* tetapi menambah jumlah blok secara signifikan pada stage ketiga yaitu menjadi 27 blok. Konfigurasi bloknya adalah 3, 3, 27, dan 3 dengan total parameter sekitar 50 juta dan kebutuhan komputasi 8.7 *GFLOPs.* Penambahan kedalaman pada *stage* ketiga bertujuan meningkatkan kapasitas representasi fitur. Model ini mencapai akurasi 83.1 persen pada *ImageNet 1K*. Spesifikasi ini dicantumkan oleh Liu *et al.* (2022) dalam tabel perbandingan performa model.

C) ConvNeXt Base (ConvNeXt B)

ConvNeXt Base merupakan varian yang kapasitasnya sebanding dengan *Swin Transformer Base* serta *ResNet 200.* Varian ini memakai kanal yang lebih

lebar yaitu 128, 256, 512, dan 1024 dengan jumlah blok 3, 3, 27, dan 3. Total parameter mencapai sekitar 89 juta dengan kebutuhan komputasi 15.4 *GFLOPs. ConvNeXt Base* menghasilkan akurasi 83.8 persen pada *ImageNet 1K* dan meningkat hingga 86.8 persen ketika dilakukan *pre training* pada *ImageNet 22K* sebelum *fine tuning*. Seluruh angka ini dilaporkan dalam hasil eksperimen model oleh Liu *et al.* (2022).

D) ConvNeXt Large (ConvNeXt L)

Varian *Large* meningkatkan kanal lebih jauh yaitu 192, 384, 768, dan 1536 dengan jumlah blok tetap sama 3, 3, 27, dan 3. Jumlah parameter mencapai 198 juta dengan kebutuhan komputasi 34.4 *GFLOPs*. Pada pelatihan *ImageNet 1K,* model ini mencapai akurasi 84.3 persen dan meningkat menjadi 87.5 persen ketika dilakukan *pre training* di *ImageNet 22K* lalu *fine tuning* pada resolusi 384 piksel. Hal ini menunjukkan bahwa *ConvNeXt* memiliki kemampuan scaling yang baik ketika kapasitas model diperbesar.

E) ConvNeXt Extra Large (ConvNeXt XL)

ConvNeXt Extra Large merupakan varian dengan kapasitas terbesar. Kanal pada tiap stage adalah 256, 512, 1024, dan 2048 dengan jumlah blok 3, 3, 27, dan 3. Total parameter mencapai sekitar 350 juta dan kebutuhan komputasi 60.9 *GFLOPs* pada resolusi 224 piksel yang meningkat menjadi 179 *GFLOPs* pada resolusi 384 piksel. Varian ini mencatatkan akurasi 87.8 persen pada *ImageNet 22K* setelah *fine tuning* dalam konfigurasi resolusi tinggi. Menurut Liu *et al.* (2022), varian *XL* menunjukkan bahwa *ConvNeXt* mampu bersaing dengan arsitektur *Vision Transformer* besar tanpa memerlukan mekanisme *self attention.*

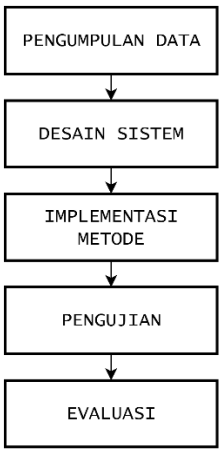
Berdasarkan karakteristik masing-masing varian tersebut, penelitian ini memilih ConvNeXt-Tiny sebagai backbone utama. Pemilihan *ConvNeXt-Tiny* sebagai *backbone* dalam penelitian ini didasarkan pada pertimbangan efisiensi, akurasi, dan keterbatasan sumber daya komputasi yang umum di lingkungan akademik atau laboratorium medis. Sebagai varian terkecil, *ConvNeXt-Tiny* memiliki jumlah parameter yang jauh lebih sedikit dibandingkan varian lain, namun tetap mampu menangani kompleksitas pola dalam citra endoskopi. Model ini telah diaplikasikan dalam penelitian segmentasi video polip oleh Bhattacharya *et al.* (2024) yang menunjukkan performa tinggi bahkan pada skenario citra dengan artefak seperti *motion blur* atau oklusi, dengan tetap mempertahankan kecepatan inferensi *real-time*. Oleh karena itu, *ConvNeXt-Tiny* dinilai tepat untuk digunakan dalam tugas klasifikasi citra endoskopi *GERD* dan polip pada penelitian ini.

BAB III

DESAIN DAN IMPLEMENTASI

3.1 Desain Penelitian

Penelitian dilakukan melalui beberapa tahapan sistematis, yaitu pengumpulan data, pra-pemrosesan data, pelatihan model, pengujian, dan evaluasi model. Data citra endoskopi diperoleh dari sumber yang telah divalidasi, seperti *dataset* publik atau data klinis yang telah dianonimkan. Model *ConvNeXt-Tiny* dipilih karena efisiensi komputasinya yang tinggi dan performa yang baik pada tugas klasifikasi citra, terutama dalam domain medis. Proses pelatihan dilakukan dengan memanfaatkan perangkat keras berbasis *GPU* untuk mempercepat komputasi, diikuti dengan evaluasi menggunakan metrik seperti matriks konfusi, akurasi, presisi, *recall*, dan *F1-score* untuk memastikan reliabilitas model dalam mendeteksi *GERD* dan polip usus. Gambar 3.1 adalah diagram alur penelitian yang menggambarkan tahapan-tahapan secara visual:



Gambar 3.1 Desain Penelitian

3.2 Pengumpulan Data

Penelitian ini menggunakan *dataset GastroEndoNet: Comprehensive Endoscopy Image dataset for GERD and Polyp Detection* yang tersedia di *Mendeley Data* (Bitto *et al.*, 2025). *Dataset* ini merupakan koleksi citra endoskopi *gastrointestinal* berkualitas tinggi yang dirancang untuk mendukung penelitian analisis citra medis, khususnya dalam deteksi dan klasifikasi penyakit *GERD* (*Gastroesophageal Reflux Disease*) dan polip usus. *dataset* ini berisi 24.036 citra dalam format *JPG* dengan resolusi 549×510 piksel, yang dikategorikan ke dalam empat kelas: *GERD*, *GERD Normal*, *Polyp*, dan *Polyp Normal*. Jumlah total citra tersebut berasal dari 4.006 citra primer yang diperluas melalui enam teknik *augmentation* untuk meningkatkan variasi dan jumlah data, sehingga cocok untuk pelatihan model pembelajaran mesin.

Distribusi citra per kelas adalah sebagai berikut:

1. *GERD*: 5.844 citra (974 citra primer × 6 *augmentation*), menampilkan kerusakan esofagus akibat refluks pada pasien yang didiagnosis *GERD* melalui pemeriksaan endoskopi. Citra ini mencakup berbagai tingkat keparahan kerusakan jaringan esofagus.

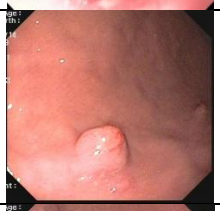
2. *GERD* Normal: 6.618 citra (1.103 citra primer × 6 *augmentation*), menggambarkan saluran *gastrointestinal* sehat tanpa tanda-tanda *GERD*, yang berfungsi sebagai kontrol untuk memastikan model dapat membedakan kondisi patologis dari normal.

3. Polyp: 4.674 citra (779 citra primer × 6 *augmentation*), menunjukkan polip *gastrointestinal* dengan berbagai jenis dan tahap perkembangan, mendukung deteksi dini kondisi yang berpotensi prakanker.

4. Polyp Normal: 6.900 citra (1.150 citra primer × 6 *augmentation*), merepresentasikan kondisi *gastrointestinal* normal tanpa keberadaan polip, yang digunakan untuk perbandingan dalam tugas klasifikasi.

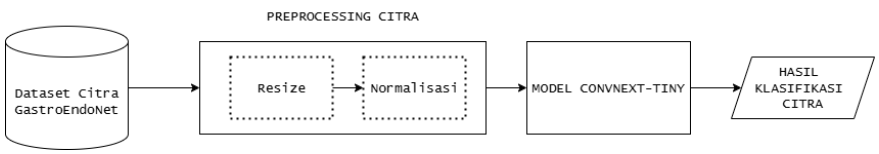
Untuk memberikan gambaran visual tentang karakteristik citra dalam *dataset*, Tabel 3.1 berikut menyajikan contoh citra dari masing-masing kelas. Tabel ini akan diisi dengan citra representatif dari *dataset GastroEndoNet* untuk mendukung analisis visual dan pelatihan model.

Tabel 3.1 Contoh Sampel Data Tiap Kelas

Kelas	**Deskripsi**	**Contoh**
GERD	Citra endoskopi esofagus dengan tanda kerusakan akibat refluks.	
GERD Normal	Citra endoskopi saluran gastrointestinal sehat tanpa tanda *GERD.*	
Polyp	Citra endoskopi dengan polip *gastrointestinal* (berbagai jenis/tahap).	
Polyp Normal	Citra endoskopi saluran *gastrointestinal* normal tanpa polip.	

3.3 Desain Sistem

Desain sistem dalam penelitian ini ditunjukkan pada Gambar 3.2, yang menggambarkan alur klasifikasi citra endoskopi menggunakan model *ConvNeXt-Tiny*. Proses dimulai dari pemanfaatan *dataset* citra *GastroEndoNet* yang berisi berbagai gambar endoskopi saluran pencernaan. Tahap awal dalam pemrosesan data mencakup praproses berupa *resize* citra ke ukuran tertentu yang sesuai dengan *input* model, dilanjutkan dengan normalisasi nilai piksel untuk meningkatkan efisiensi dan akurasi pelatihan model. Setelah melalui tahap praproses, citra-citra tersebut kemudian dimasukkan ke dalam model *ConvNeXt-Tiny*, yang telah diadaptasi untuk tugas klasifikasi medis. Hasil keluaran dari model ini berupa label klasifikasi yang menunjukkan jenis kondisi atau penyakit pada citra endoskopi yang dianalisis.



Gambar 3.2 Desain Sistem

3.3.1 Input Citra

Tahap awal sistem adalah penerimaan citra endoskopi dari *dataset GastroEndoNet*, yang berisi 24.036 citra dalam format JPG dengan resolusi 549 × 510 piksel. Citra tersebut mewakili empat kelas: *GERD* (5.844 citra), *GERD* Normal (6.618 citra), *Polyp* (4.674 citra), dan *Polyp Normal* (6.900 citra). Citra-citra ini diambil dari pemeriksaan endoskopi *gastrointestinal*, mencakup berbagai kondisi patologis dan normal.

3.3.2 *Preprocessing* Citra

Preprocessing bertujuan menyiapkan citra sebelum digunakan dalam model. Tujuannya adalah meningkatkan efisiensi dan akurasi model dengan memastikan data yang diproses konsisten serta lebih mudah dipahami oleh model.

3.3.2.1 *Resize*

Seluruh citra diubah ukurannya menjadi 224 × 224 piksel, yang merupakan resolusi standar untuk pelatihan pada *dataset ImageNet-1K*. Ukuran ini dipilih untuk memastikan kompatibilitas dengan arsitektur *ConvNeXt* yang digunakan. Pada publikasi *ConvNeXt* (Liu *et al.,* 2022), dijelaskan bahwa model dilatih dari awal menggunakan *dataset ImageNet-1K* selama 300 *epoch* dengan resolusi masukan 224 × 224 piksel. Hal tersebut menunjukkan bahwa resolusi ini digunakan sebagai *baseline* pelatihan. Selain itu, citra dalam format *RGB* dengan tiga *channel* merupakan standar pada *dataset ImageNet*, sehingga secara implisit format masukan yang digunakan adalah 224 × 224 × 3. Penyeragaman dimensi ini penting untuk menjaga konsistensi data masukan serta mendukung efisiensi dan kestabilan selama proses pelatihan model.

3.3.2.2 Normalisasi

Setelah citra di-*resize*, langkah praproses selanjutnya adalah normalisasi nilai piksel. Tujuan dari normalisasi ini adalah untuk menyelaraskan distribusi nilai *input* agar lebih stabil dan mudah dipelajari oleh model. Dalam penelitian ini, normalisasi dilakukan berdasarkan statistik global dari *dataset ImageNet*, yaitu dengan menggunakan nilai rata-rata (*mean*) dan deviasi standar (*standard deviation*) pada

masing-masing kanal warna (*R, G, B*). Nilai *mean* yang digunakan adalah [0,485, 0,456, 0,406], sedangkan nilai *standard deviation*-nya adalah [0,229, 0,224, 0,225]. Proses normalisasi ini secara matematis dapat dirumuskan dengan Persamaan 3.1.

$$x_{\text{norm}} = \frac{x/255 - \mu}{\sigma} \tag{3.1}$$

di mana x adalah nilai piksel asli, μ adalah nilai rata-rata (*mean*) untuk tiap kanal, dan σ adalah deviasi standar untuk tiap kanal. Langkah pembagian dengan 255 dilakukan terlebih dahulu untuk mengubah skala piksel dari 0–255 menjadi 0–1, sebelum distandarkan menggunakan nilai statistik. Meskipun tidak disebutkan secara eksplisit dalam *paper ConvNeXt*, proses ini merupakan praktik standar dalam pelatihan model pada *dataset ImageNet* dan juga digunakan dalam implementasi resmi *ConvNeXt*. Dengan melakukan normalisasi ini, model dapat menerima data *input* dengan distribusi yang lebih terpusat, sehingga dapat mempercepat proses pelatihan dan meningkatkan kestabilan konvergensi.

3.4 Implementasi Metode

Pada tahap ini, metode yang digunakan dalam penelitian diimplementasikan untuk membangun model klasifikasi citra endoskopi berbasis arsitektur *ConvNeXt-Tiny*. *ConvNeXt* merupakan pengembangan dari jaringan konvolusional standar (*ConvNet*) yang dimodernisasi dengan mengadopsi prinsip desain *Vision Transformer*, namun tetap mempertahankan sifat sepenuhnya konvolusional. Keunggulan *ConvNeXt* terletak pada kemampuannya menggabungkan efisiensi komputasi *ConvNet* dengan performa tinggi yang kompetitif terhadap arsitektur *Transformer* pada berbagai tugas visi komputer, termasuk klasifikasi citra resolusi tinggi. Dalam studi kasus ini, *ConvNeXt-Tiny* dipilih karena memiliki kompleksitas

komputasi yang relatif rendah, namun mampu mencapai akurasi tinggi pada tugas klasifikasi. Arsitektur ini diadaptasi dan dilatih untuk mendeteksi dua jenis kelainan pada citra endoskopi, yaitu *Gastroesophageal Reflux Disease* (*GERD*) dan polip usus, yang memerlukan pemrosesan fitur visual secara mendetail.

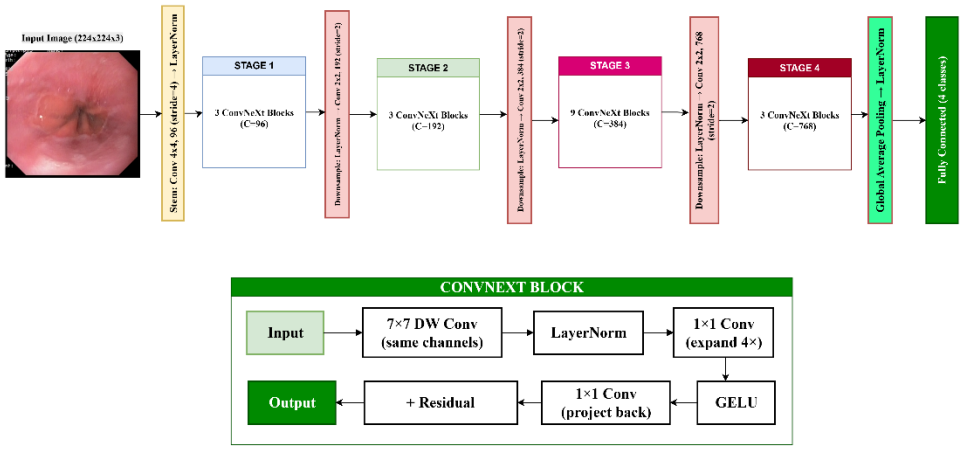
3.4.1 Arsitektur *ConvNeXt-Tiny*

ConvNeXt-Tiny merupakan varian ringan dari keluarga *ConvNeXt* yang dirancang untuk mempertahankan kinerja konvolusional modern sambil mengadopsi prinsip-prinsip desain dari arsitektur *transformer* seperti *Vision Transformer* (*ViT*). Arsitektur ini mengusung pendekatan hierarkis dengan empat tahap (*stage*) yang memiliki resolusi peta fitur berbeda dan jumlah kanal (*channels*) yang meningkat secara progresif, yakni 96, 192, 384, dan 768 kanal. Jumlah blok konvolusional pada masing-masing tahap adalah (3, 3, 9, 3), dengan distribusi beban komputasi 1:1:3:1 sebagaimana disarankan oleh Liu *et al.* (2022).

Setiap blok *ConvNeXt* mengintegrasikan *depthwise convolution* berukuran kernel besar (7 × 7), diikuti *LayerNorm*, dua *pointwise convolution* (1 × 1) yang membentuk struktur *inverted bottleneck*, serta aktivasi nonlinier *GELU*. Selain itu, *ConvNeXt-Tiny* menggunakan mekanisme *LayerScale*, yaitu skalar yang dapat dilatih untuk mengatur kontribusi jalur residual secara dinamis. Kombinasi ini memungkinkan model menangkap konteks spasial yang luas dengan efisiensi komputasi tinggi.

Proses pemrosesan dimulai dari *stem layer* (*patch embedding*) yang mengubah citra masukan beresolusi 224 × 224 piksel menjadi representasi fitur awal, dilanjutkan serangkaian *downsampling layer* untuk menurunkan resolusi

spasial sambil meningkatkan dimensi kanal. Fitur yang dihasilkan kemudian diproses melalui empat tahap blok *ConvNeXt*, dilanjutkan *global average pooling* dan *fully connected layer* untuk menghasilkan keluaran klasifikasi. Ilustrasi lengkap arsitektur *ConvNeXt-Tiny* yang digunakan dalam penelitian ini ditunjukkan pada Gambar 3.3.



Gambar 3.3 Diagram alur arsitektur *ConvNeXt-Tiny*

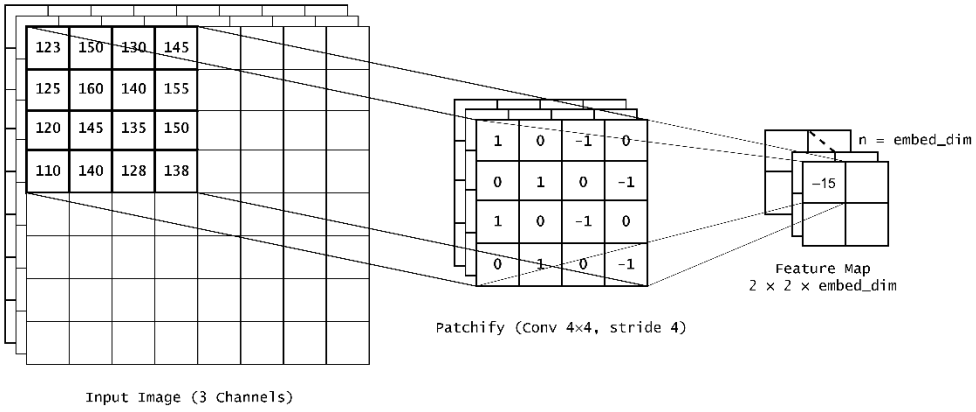
3.4.1.1 *Stem Layer (Patchify)*

Pada tahap awal pemrosesan citra, *ConvNeXt-Tiny* menggunakan *stem layer* atau *patch embedding layer* untuk mengubah citra mentah beresolusi tinggi menjadi representasi fitur awal yang dapat diolah secara efisien oleh jaringan. Berbeda dengan *patch tokenization* pada arsitektur *Vision Transformer*, *ConvNeXt* menggunakan prinsip konvolusi dengan menerapkan *convolutional layer* berukuran *kernel* besar dan *stride* tertentu untuk langsung menurunkan resolusi citra. Operasi pada *stem layer* dapat direpresentasikan dengan proses konvolusi dua dimensi sebagaimana Persamaan 3.2.

$$FM_{(a,b)} = \sum_{h=0}^{k_h-1} \sum_{w=0}^{k_w-1} K_{(h,w)} \times X_{(a+h,b+w)} \quad (3.2)$$

Keterangan:
$FM_{(a,b)}$: output konvolusi pada posisi (a, b)
$K_{(h,w)}$: kernel konvolusi dengan ukuran $k_{\mathrm{h}} \times kw$
$X_{(a+h,b+w)}$: nilai piksel input di posisi $(a + h, b + w)$

Dalam *ConvNeXt-Tiny*, *stem layer* ini menggunakan *kernel* konvolusi berukuran 4×4 dengan *stride* 4, yang berarti bahwa setiap keluaran representasi (*feature map*) mengandung informasi dari area 16 piksel input (4×4).



Gambar 3.4 Contoh sederhana operasi *patchify* pada satu *patch*.

Proses ini mengubah citra RGB berukuran awal 224 × 224 × 3 menjadi representasi awal berukuran 56 × 56 × 96, sebagaimana ditunjukkan pada Gambar 3.4. Transformasi ini dilakukan menggunakan konvolusi dengan *kernel* berukuran 4 × 4 dan *stride* 4, sehingga citra masukan dibagi menjadi potongan *(patch)* berukuran 4 × 4 × 3. Tidak seperti ilustrasi konvolusi sederhana yang hanya menggunakan satu *kernel,* pada tahap ini digunakan sebanyak 96 kernel (sesuai jumlah kanal keluaran yang ditentukan arsitektur *ConvNeXt-Tiny*). Setiap *kernel* menghasilkan satu nilai untuk setiap *patch,* sehingga setiap *patch* diproyeksikan menjadi vektor fitur berdimensi 96. Nilai 96 ini bukan hasil perhitungan langsung dari ukuran *patch*, melainkan jumlah *kernel* yang ditetapkan secara arsitektural

sebagai *embedding dimension* awal. Seluruh hasil proyeksi kemudian disusun kembali membentuk peta fitur awal *(initial feature map)* berukuran 56 × 56 × 96, yang selanjutnya diproses pada tahap berikutnya.

Diilustrasikan pada Gambar 3.4, misalkan diambil sebuah *patch* citra berukuran 4 × 4 pada kanal merah dari citra endoskopi. Nilai intensitas piksel dalam *patch* tersebut, dari kiri atas ke kanan bawah, antara lain 123, 150, 130, 145 pada baris pertama; 125, 160, 140, 155 pada baris kedua; 120, 145, 135, 150 pada baris ketiga; serta 110, 140, 128, 138 pada baris keempat. *Patch* ini kemudian dikalikan secara elemen demi elemen dengan kernel konvolusi berukuran 4 × 4 yang bobotnya terdefinisi sebagai [1, 0, –1, 0] pada baris pertama, [0, 1, 0, –1] pada baris kedua, [1, 0, –1, 0] pada baris ketiga, dan [0, 1, 0, –1] pada baris keempat.

Operasi dilakukan dengan cara sederhana: piksel 123 di kiri atas dikalikan dengan bobot 1 menghasilkan 123, piksel 150 dikalikan dengan 0 menghasilkan 0, piksel 130 dikalikan dengan –1 menghasilkan –130, dan seterusnya. Setelah semua hasil perkalian dijumlahkan, nilai yang diperoleh adalah –15. Dengan demikian, dari satu *patch* citra berukuran 4 × 4, kernel tersebut menghasilkan satu nilai tunggal yang menjadi representasi fitur di posisi tersebut.

Sebagai simulasi, apabila proses serupa dilakukan pada seluruh citra berukuran 8 × 8 dengan stride 4, maka akan terbentuk 4 *patch*, dan masing-masing menghasilkan satu nilai keluaran. Keempat nilai tersebut dapat disusun menjadi feature map berukuran 2 × 2. Perlu dicatat bahwa ilustrasi ini hanya menggambarkan mekanisme dasar *patchify* dengan satu kernel. Pada arsitektur *ConvNeXt-Tiny* yang sesungguhnya, tahap *patchify* menggunakan sebanyak 96

kernel paralel, sehingga setiap *patch* diproyeksikan bukan menjadi satu nilai tunggal, melainkan vektor fitur berdimensi 96. Hal inilah yang menghasilkan representasi awal berukuran 56 × 56 × 96 dari citra masukan 224 × 224 × 3.

3.4.1.2 *Downsampling Layer*

Setiap tahap pada arsitektur *ConvNeXt-Tiny* terdiri atas sejumlah blok konvolusional yang beroperasi pada resolusi spasial tetap. Setelah suatu tahap selesai, diperlukan *downsampling layer* untuk menurunkan resolusi spasial *feature map* sebelum memasuki tahap berikutnya. Berbeda dengan *ResNet* yang mengintegrasikan proses *downsampling* ke dalam blok residual pertama, *ConvNeXt* memisahkannya menjadi lapisan terpisah sehingga proses ekstraksi fitur dan perubahan resolusi dapat dioptimalkan secara independen (Liu et al., 2022).

Secara berurutan, *downsampling layer* pada *ConvNeXt* diawali dengan *Layer Normalization* yang berfungsi menstabilkan distribusi aktivasi antarkanal dan mengurangi pergeseran nilai statistik akibat perubahan resolusi. Setelah itu, dilakukan konvolusi berukuran 2 × 2 dengan *stride* 2, yang secara matematis mengubah dimensi *feature map* dari $(H/s) \times (W/s) \times C$ menjadi $\left(\frac{H}{2s}\right) \times \left(\frac{W}{2s}\right) \times 2C$, dengan H dan W sebagai dimensi citra asli, C jumlah kanal, serta s faktor *downsampling* kumulatif dari tahap sebelumnya.

Sebagai contoh, keluaran tahap pertama dengan ukuran $\left(\frac{H}{4}\right) \times \left(\frac{W}{4}\right) \times 96$ akan berubah menjadi $\left(\frac{H}{8}\right) \times \left(\frac{W}{8}\right) \times 192$ setelah melewati *downsampling layer*. Peningkatan jumlah kanal dari 96 menjadi 192 bertujuan untuk mengompensasi

berkurangnya informasi spasial dengan menambah kapasitas representasi di domain kanal.

Pendekatan ini tidak hanya meningkatkan efisiensi komputasi, tetapi juga memperluas jangkauan pola visual yang dapat ditangkap oleh jaringan. Dalam konteks pengolahan citra medis, strategi ini memungkinkan model untuk menggabungkan informasi lokal dan global secara lebih efektif, yang penting dalam mengidentifikasi pola morfologi atau anomali dengan ukuran bervariasi.

3.4.1.3 *ConvNeXt Block*

ConvNeXt Block merupakan unit bangunan utama dalam arsitektur *ConvNeXt-Tiny*, yang terdiri atas serangkaian operasi konvolusional dan nonlinier yang dirancang untuk meniru efektivitas arsitektur *transformer*, namun tetap berbasis konvolusi murni. Setiap blok dalam *ConvNeXt* memproses representasi fitur dari peta fitur sebelumnya, dengan tujuan menyaring, memperkaya, dan mengekstraksi informasi spasial maupun kontekstual yang lebih dalam. Struktur *ConvNeXt Block* secara umum terdiri atas:

1. *Depthwise Convolution* (7 × 7)
2. *Layer Normalization*
3. *Pointwise Convolution* 1 (4× *Expansion*)
4. *GELU Activation*
5. *Pointwise Convolution* 2 (*Projection*)
6. *Residual Connection* dengan *LayerScale*

Desain ini terinspirasi dari struktur *inverted bottleneck* yang umum digunakan pada arsitektur *mobile* seperti *MobileNetV2*, namun disesuaikan dengan konfigurasi

yang lebih dalam dan stabil. Pemisahan antara konvolusi spasial (*depthwise*) dan konvolusi kanal (*pointwise*) memberikan efisiensi komputasi sekaligus fleksibilitas dalam manipulasi informasi fitur.

Dalam klasifikasi citra endoskopi, *ConvNeXt Block* berfungsi sebagai ekstraktor fitur utama yang mampu mendeteksi pola-pola penting pada permukaan mukosa esofagus atau jaringan usus besar, seperti tekstur kasar akibat *GERD* atau bentuk tonjolan khas dari polip.

1) Depthwise Convolution (7x7)

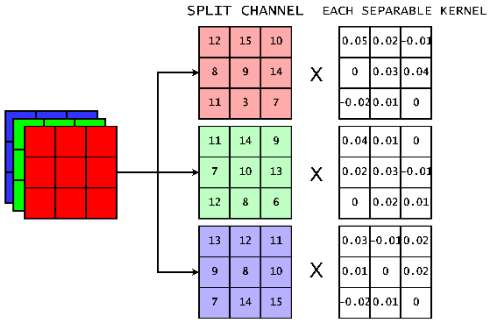
Salah satu inovasi utama dalam *ConvNeXt Block* adalah penggunaan *depthwise convolution* dengan kernel besar berukuran 7 × 7. Berbeda dari konvolusi standar, di mana setiap filter terhubung ke seluruh kanal *input*, *depthwise convolution* hanya melakukan konvolusi secara terpisah pada setiap kanal *input* seperti pada Gambar 3.8. Hal ini memungkinkan penekanan pada fitur spasial per kanal dengan beban komputasi yang jauh lebih ringan. Operasi *depthwise convolution* secara matematis dituliskan sebagaimana Persamaan 3.3.

$$Y_c(a,b) = \sum_{h=0}^{k-1} \sum_{w=0}^{k-1} K_{c(h,w)} \cdot X_{c(a+h,b+w)} \quad (3.3)$$

Keterangan:
$Y_c(a,b)$: output fitur untuk kanal ke-c pada posisi (a,b)
$K_{c(h,w)}$: kernel konvolusi kh × kw untuk kanal c
$X_{c(a+h,b+w)}$: nilai input pada kanal c, posisi (a+h, b+w)
k : panjang sisi kernel, dalam hal ini 7

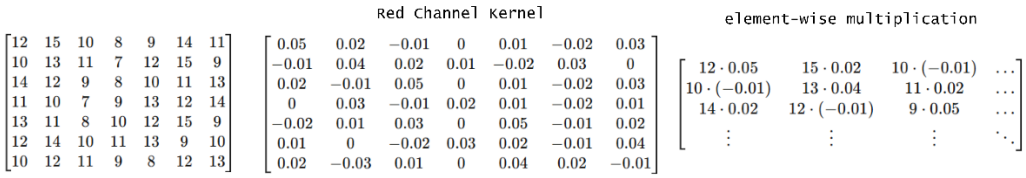
Penggunaan *kernel* besar (7×7) memperluas *receptive field* dari unit konvolusi tanpa menambah kedalaman *layer* secara signifikan. Pada studi kasus citra endoskopi, *receptive field* yang luas berfungsi untuk menangkap pola tekstural

global, seperti perubahan pola mukosa akibat iritasi kronis (*GERD*) atau kontur polip yang menonjol secara halus.



Gambar 3.5 Ilustrasi *Depthwise convolution*

Depthwise convolution juga lebih efisien dibandingkan konvolusi biasa. Jumlah parameter dan komputasi menurun drastis karena *filter* hanya diterapkan pada satu kanal, bukan seluruh *input* seperti pada Gambar 3.5. Hal ini sejalan dengan kebutuhan model ringan seperti *ConvNeXt-Tiny* yang ditujukan untuk penggunaan pada sistem terbatas, termasuk perangkat medis berbasis *edge computing* atau *inference* lokal di rumah sakit.



Gambar 3.6 Contoh Operasi Depthwise Convolution

Sebagai contoh proses *depthwise convolution*, pada gambar 3.6, misalkan diambil sebuah *patch* citra berukuran 7×7 pada kanal merah dari citra endoskopi. Nilai intensitas piksel dalam *patch* tersebut dapat direpresentasikan dalam bentuk matriks sebagaimana ditunjukkan pada Gambar 3.6. Setiap elemen matriks

menyatakan nilai intensitas piksel, misalnya baris pertama berisi [12,15,10,8,9,14,11], baris kedua [10,13,11,7,12,15,9], dan seterusnya hingga baris ketujuh.

Patch citra ini kemudian dikalikan secara elemen-demi-elemen dengan kernel konvolusi berukuran sama, yang juga direpresentasikan dalam bentuk matriks bobot. Sebagai contoh, piksel bernilai 12 pada posisi kiri-atas dikalikan dengan bobot kernel 0.05 menghasilkan 0.6, piksel bernilai 15 dikalikan dengan bobot 0.02 menghasilkan 0.3, dan piksel bernilai 10 dikalikan dengan bobot –0.01 menghasilkan –0.1. Proses perkalian ini dilakukan untuk seluruh 49 pasangan piksel–bobot pada *patch* tersebut.

Hasil dari semua perkalian kemudian dijumlahkan, sehingga menghasilkan satu nilai keluaran, misalnya sebesar 0.6342 untuk posisi tersebut pada kanal merah. Operasi serupa dilakukan secara independen pada kanal hijau dan kanal biru dengan kernel masing-masing, misalnya menghasilkan 0.5821 dan 0.6015. Dengan demikian, pada posisi keluaran yang sama diperoleh vektor [0.6342, 0.5821, 0.6015]. Vektor inilah yang selanjutnya dapat diproses oleh *pointwise convolution* (kernel 1×1) untuk menghasilkan representasi akhir dari *depthwise separable convolution*.

2) Layer Normalization

Setelah proses ekstraksi fitur spasial melalui *depthwise convolution*, keluaran dari setiap kanal fitur dalam *ConvNeXt Block* akan distandarisasi menggunakan *Layer Normalization* (*LayerNorm*). Berbeda dari *Batch Normalization* yang menghitung statistik berdasarkan *mini-batch*, *Layer Normalization* melakukan

normalisasi berdasarkan fitur di dalam satu sampel saja, yaitu di sepanjang dimensi kanal. Hal ini menjadikan *LayerNorm* lebih stabil dan efektif, terutama pada model yang menggunakan *batch* kecil atau *inference* satu per satu kondisi yang sering dijumpai dalam aplikasi medis. Secara matematis, proses normalisasi ini ditulis dalam Persamaan 3.4.

$$\mathrm{LN}(x_i) = \frac{x_i - \mu}{\sqrt{\sigma^2 + \epsilon}} \tag{3.4}$$

Keterangan:
x_i : nilai aktivasi pada dimensi ke-i dari vektor input
μ : rata-rata dari semua nilai dalam satu vektor input
σ^2 : varians dari semua nilai dalam vektor
ε : konstanta kecil untuk mencegah pembagian nol

Posisi *LayerNorm* dalam *ConvNeXt* juga tidak konvensional. Jika arsitektur *ResNet* menempatkan normalisasi setelah aktivasi, *ConvNeXt* sejalan dengan praktik pada *Transformer*, menempatkannya sebelum blok utama (*pre-norm*), yang secara empiris terbukti mempercepat proses konvergensi dan mencegah hilangnya gradien.

Tabel 3.2 Simulasi Perhitungan Layer Normalization dengan Rata-rata (μ) = 0.6059 dan Varians (σ^2) = 0.0005

Kanal	**Nilai awal**	**x - μ**	**(x - μ)²**	**Normalisasi**
1	0.6342	0.0283	0.0008	1.27
2	0.5821	-0.0238	0.0006	-1.05
3	0.6015	-0.0044	0	-0.22

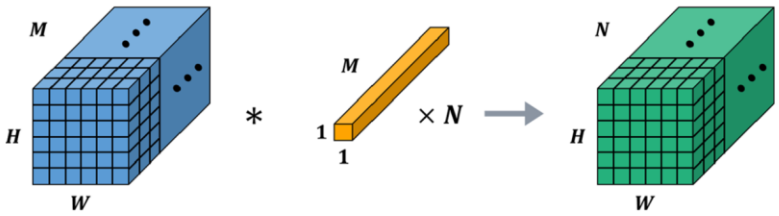
Misalkan sebuah vektor keluaran dari *depthwise convolution* pada satu posisi piksel ditunjukkan pada tabel 3.2 memiliki nilai fitur dari tiga kanal: $[0.6342, 0.5821, 0.6015]$. Proses *Layer Normalization* akan menghitung rata-rata μ dari ketiga nilai ini, yaitu $(0.6342 + 0.5821 + 0.6015)/3 = 0.6059$. Selanjutnya dihitung varians σ^2, misalnya $((0.6342 - 0.6059)^2 + (0.5821 - 0.6059)^2 +$

$(0.6015 - 0.6059)^2)/3 = 0.0005$. Setelah itu, setiap nilai dikurangi dengan rata-rata dan dibagi akar varians ditambah konstanta ϵ. Misalnya, nilai 0.6342 setelah normalisasi menjadi $(0.6342 - 0.6059)/\sqrt{0.0005 + \epsilon} \approx 1.27$. Nilai 0.5821 menjadi $(0.5821 - 0.6059)/\sqrt{0.0005 + \epsilon} \approx -1.05$. Nilai 0.6015 menjadi $(0.6015 - 0.6059)/\sqrt{0.0005 + \epsilon} \approx -0.22$.

Hasil akhirnya adalah vektor $[1.27, -1.05, -0.22]$, yaitu representasi terstandarisasi dari fitur di titik tersebut. Dengan cara ini, LayerNorm memastikan bahwa meskipun intensitas asli antar kanal berbeda, model tetap menerima distribusi fitur dengan rata-rata nol dan varians terkontrol.

3) Pointwise Conv 1 (4× Expansion)

Setelah fitur melewati proses normalisasi, *ConvNeXt Block* melakukan tahap ekspansi dimensi fitur menggunakan *pointwise convolution* berukuran *kernel* 1×1 seperti tampak pada Gambar 3.7. Tahap ini sering disebut *4× Expansion* karena jumlah kanal keluaran diperbesar menjadi empat kali lipat dari jumlah kanal masukan. Misalnya, jika masukan memiliki 96 kanal, maka setelah tahap ini akan menjadi 384 kanal.



Gambar 3.7 Ilustrasi Pointwise Convolution (Sumber: Zhang et al., 2020)

Secara intuitif, proses ini dapat dianalogikan seperti memperluas ruang kerja bagi model. Dengan memperbesar jumlah kanal, model memperoleh kapasitas representasi yang lebih kaya sehingga dapat memproses variasi pola spasial dengan

detail yang lebih halus sebelum tahap kompresi kembali. Secara matematis, *pointwise convolution* dengan *kernel* 1×1 dapat dituliskan sebagai:

$$Y_{i,j,k} = \sum_{c=1}^{C_{\text{in}}} W_{1,1,c,k} \cdot X_{i,j,c} + b_k \tag{3.5}$$

Keterangan:
$(Y_{i,j,k})$: Nilai keluaran pada posisi koordinat spasial (i,j)untuk kanal ke-(k).
$(X_{i,j,c})$: Nilai masukan pada posisi $((i,j))$ untuk kanal ke-(c).
$(W_{1,1,c,k})$: Bobot kernel 1×1 yang menghubungkan kanal masukan ke-(c) dengan kanal keluaran ke-(k).
(b_k) : Nilai bias yang ditambahkan pada kanal keluaran ke$-(k)$.
(C_{in}) : Jumlah total kanal masukan.

Pointwise convolution berbeda dengan konvolusi standar yang menggunakan *kernel* besar; di sini, *kernel* hanya berukuran 1×1 sehingga operasi yang dilakukan murni menggabungkan informasi antarkanal tanpa mengubah dimensi spasial. Hal ini membuatnya sangat efisien secara komputasi, namun tetap memberikan dampak signifikan terhadap kemampuan pemodelan nonlinier jaringan.

$$\underbrace{\begin{bmatrix} 0.5 & -0.2 \\ -0.3 & 0.6 \\ 0.8 & 0.4 \end{bmatrix}^T}_{\text{Weight}} \cdot \underbrace{\begin{bmatrix} 1.27 \\ -1.05 \\ -0.22 \end{bmatrix}}_{\text{Input}} + \underbrace{\begin{bmatrix} 0.1 \\ -0.05 \end{bmatrix}}_{\text{Bias}} = \begin{bmatrix} 0.874 \\ -1.022 \end{bmatrix}$$

Gambar 3.8 Contoh Operasi Pointwise Convolution

Hal ini tampak misalkan pada satu posisi spasial (i,j) setelah melalui *LayerNorm* terdapat tiga kanal fitur dengan nilai $[1.27, -1.05, -0.22]$ yang ditunjukkan oleh gambar 3.8. Pada tahap *pointwise convolution* dengan kernel 1×1, setiap kanal keluaran akan dihitung sebagai kombinasi *linear* dari ketiga nilai ini. Misalnya, untuk kanal keluaran pertama, bobot yang digunakan adalah $[0.5, -0.3, 0.8]$ dengan *bias* 0.1. Perhitungannya menjadi: $(1.27 \times 0.5) +$

$(-1.05 \times -0.3) + (-0.22 \times 0.8) + 0.1 = 0.635 + 0.315 - 0.176 + 0.1 = 0.874$. Untuk kanal keluaran kedua, bobotnya $[-0.2, 0.6, 0.4]$ dengan *bias* -0.05. Maka hasilnya adalah $(1.27 \times -0.2) + (-1.05 \times 0.6) + (-0.22 \times 0.4) - 0.05 = -0.254 - 0.63 - 0.088 - 0.05 = -1.022$. Proses ini dilakukan terus hingga terbentuk empat kanal keluaran (sesuai dengan ekspansi 4× dari tiga kanal masukan). Jika sebelumnya ada 3 kanal, maka setelah ekspansi jumlahnya menjadi 12 kanal; pada arsitektur sebenarnya, 96 kanal akan diekspansi menjadi 384 kanal.

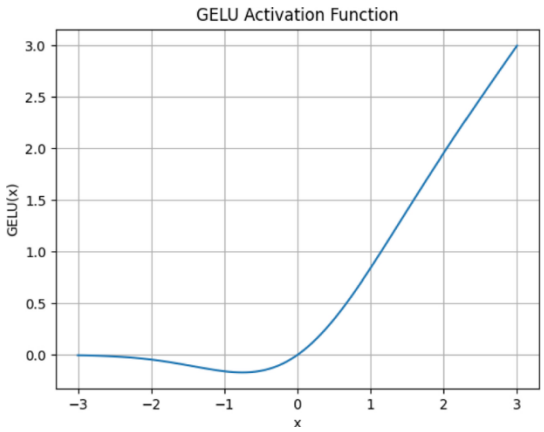
Dengan begitu, pada titik spasial yang sama diperoleh vektor fitur baru berdimensi lebih besar, misalnya $[0.874, -1.022, 0.457, 0.215, \dots]$. Vektor ini kemudian diteruskan ke tahap aktivasi nonlinier berikutnya. Simulasi ini memperlihatkan bahwa meskipun hanya menggunakan kernel 1×1, operasi *pointwise convolution* mampu mencampur informasi antar kanal dan memperkaya representasi fitur.

4) GELU Activation

Setelah tahap ekspansi kanal, keluaran kemudian diproses menggunakan fungsi aktivasi *Gaussian Error Linear Unit* (*GELU*). Fungsi aktivasi ini diperkenalkan untuk memberikan pemetaan nonlinier yang lebih halus dibandingkan fungsi klasik seperti *ReLU*. *GELU* secara probabilistik mempertahankan nilai *input* berdasarkan distribusi *Gaussian*, sehingga transisi antara nilai yang diredam dan dipertahankan menjadi lebih *smooth*. Secara matematis, fungsi aktivasi *GELU* dapat didefinisikan sebagaimana diuraikan dalam Persamaan 3.6.

$$\mathrm{GELU}(x) = x \cdot \Phi(x) \tag{3.6}$$

dengan $\Phi(x)$ merupakan fungsi distribusi kumulatif (*CDF*) dari distribusi normal standar.



Gambar 3.9 Ilustrasi Fungsi Aktivasi GeLU

Namun, dalam implementasi pada jaringan saraf, sering digunakan bentuk pendekatan sebagai berikut:

$$\mathrm{GELU}(x) \approx 0.5x\left[1+\tanh\left(\sqrt{\frac{2}{\pi}}(x+0.044715x^3)\right)\right] \tag{3.7}$$

Perbedaan utama *GELU* dibandingkan *ReLU* adalah sifatnya yang tidak sepenuhnya mematikan nilai negatif, melainkan menimbangnya secara probabilistik. Hal ini membuat pembelajaran menjadi lebih stabil dan memungkinkan jaringan mempertahankan informasi penting yang berada di dekat ambang nol, seperti terlihat pada Gambar 3.9, yang biasanya hilang pada *ReLU*.

Sebagai contoh, misalkan dari hasil *pointwise convolution* di tahap sebelumnya pada satu titik spasial diperoleh empat nilai fitur: $[0.874, -1.022, 0.457, 0.215]$. Masing-masing nilai ini kemudian diproses menggunakan fungsi aktivasi *GELU.* Untuk nilai positif 0.874, fungsi distribusi *Gaussian* akan memberikan probabilitas mendekati 1, sehingga hasil *GELU* kira-kira $\mathrm{GELU}(0.874) \approx 0.874 \times 0.82 =$

0.716. Untuk nilai negatif -1.022, probabilitasnya mendekati 0.16, sehingga $\mathrm{GELU}(-1.022) \approx -1.022 \times 0.16 = -0.164$. Sementara itu, nilai 0.457 akan menghasilkan $\mathrm{GELU}(0.457) \approx 0.457 \times 0.67 = 0.306$, dan nilai 0.215 menjadi $\mathrm{GELU}(0.215) \approx 0.215 \times 0.58 = 0.125$. Sehingga, vektor keluaran setelah melalui *GELU* adalah $[0.716, -0.164, 0.306, 0.125]$. Terlihat bahwa nilai besar positif dipertahankan hampir penuh, nilai kecil positif dilewatkan sebagian, sementara nilai negatif tidak langsung dibuang tetapi tetap diberi kontribusi kecil.

5) Pointwise Conv 2 (Projection)

Lapisan *Pointwise Convolution 2* (*Projection*) merupakan tahap akhir dari bagian *feed-forward* dalam *ConvNeXt Block*, yang berfungsi untuk mengembalikan jumlah kanal (*channel*) fitur ke dimensi semula setelah proses ekspansi pada *Pointwise Convolution 1*. Masukan pada tahap ini adalah peta fitur berdimensi tinggi (*4C*) yang dihasilkan dari *Pointwise Convolution 1* dan telah melalui fungsi aktivasi nonlinier *GELU*. Peta fitur ini memiliki kapasitas representasi yang kaya namun berukuran lebih besar dibandingkan dimensi awal blok.

Untuk memperjelas, misalkan pada satu titik spasial setelah melewati *GELU* terdapat empat kanal hasil ekspansi dengan nilai $[0.716, -0.164, 0.306, 0.125]$. Tahap *Pointwise Convolution 2* bertugas memproyeksikan kembali empat nilai ini menjadi satu kanal (atau lebih umum: dari *4C* kembali ke *C*). Misalnya, bobot kernel 1×1 untuk kanal keluaran adalah $[0.4, -0.2, 0.3, 0.5]$ dengan *bias* sebesar 0.05. Maka perhitungannya menjadi $(0.716 \times 0.4) + (-0.164 \times -0.2) + (0.306 \times 0.3) + (0.125 \times 0.5) + 0.05$. Hasilnya adalah $0.286 + 0.0328 + 0.0918 + 0.0625 + 0.05 = 0.5231$. Sehingga, vektor berdimensi 4 berhasil

dipadatkan menjadi sebuah nilai tunggal 0.5231 pada kanal keluaran. Dalam implementasi nyata, proses ini dilakukan pada setiap posisi spasial dan untuk semua kanal keluaran, sehingga jika awalnya ada 96 kanal masukan yang diekspansi menjadi 384, tahap ini akan mengompresinya kembali ke 96 kanal. Hasil proyeksi ini kemudian siap untuk digabungkan dengan *shortcut connection (residual path).*

6) Residual Connection & LayerScale

Residual connection digunakan untuk mempertahankan aliran informasi dan gradien di dalam jaringan, sehingga pelatihan tetap stabil pada arsitektur yang dalam. Mekanisme ini menambahkan *input* awal (*shortcut connection*) langsung ke keluaran hasil transformasi blok, sehingga jaringan hanya perlu mempelajari fungsi residu alih-alih mempelajari transformasi penuh. Secara matematis, *residual connection* dapat dinyatakan pada Persamaan 3.8.

$$y = F(x, W) + x \tag{3.8}$$

Keterangan:
x : input ke blok
F(x, W) : hasil transformasi non-linear (misalnya depthwise convolution, LayerNorm, pointwise convolution, dan aktivasi)
y : output blok setelah penjumlahan residual

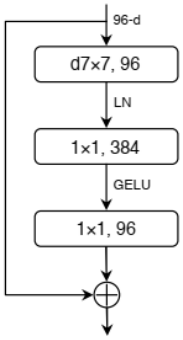
Pada *ConvNeXt*, *residual connection* dikombinasikan dengan *LayerScale*, yaitu parameter penskalaan adaptif (γ) yang diberikan pada setiap kanal output dari fungsi $F(x, W)$ sebelum dilakukan penjumlahan dengan *shortcut connection*. Tujuannya adalah mengatur besarnya kontribusi hasil transformasi terhadap keluaran akhir. Persamaan *LayerScale* tertera pada Persamaan 3.9.

$$y = \left(\gamma \odot F(x, W)\right) + x \tag{3.9}$$

Keterangan:
γ : vektor skalar berukuran sama dengan jumlah kanal, diinisialisasi dengan nilai kecil (misalnya 1e-6)

⊙ : operasi perkalian elemen-per-elemen (element-wise multiplication)

Inisialisasi γ dengan nilai kecil membuat model pada awal pelatihan lebih mengandalkan *shortcut connection*, sehingga mengurangi risiko ketidakstabilan pada iterasi awal. Nilai γ akan menyesuaikan selama pelatihan, memberikan fleksibilitas kontribusi transformasi $F(x, W)$.



Gambar 3.10 Residual Pada *ConvNeXt*

Struktur alur data pada *ConvNeXt Block* yang menerapkan *residual connection* dan *LayerScale* ditunjukkan pada Gambar 3.10. Blok ini dimulai dari *depthwise convolution* (*kernel* 7×7), diikuti *Layer Normalization* (*LN*), dilanjutkan dengan *pointwise convolution* pertama (1×1) yang memperluas dimensi kanal menjadi empat kali lipat, aktivasi *GELU*, lalu *pointwise convolution* kedua (1×1) untuk mengembalikan jumlah kanal ke dimensi semula. Keluaran dari rangkaian transformasi ini kemudian diskalakan menggunakan *LayerScale* sebelum dijumlahkan dengan *shortcut connection*.

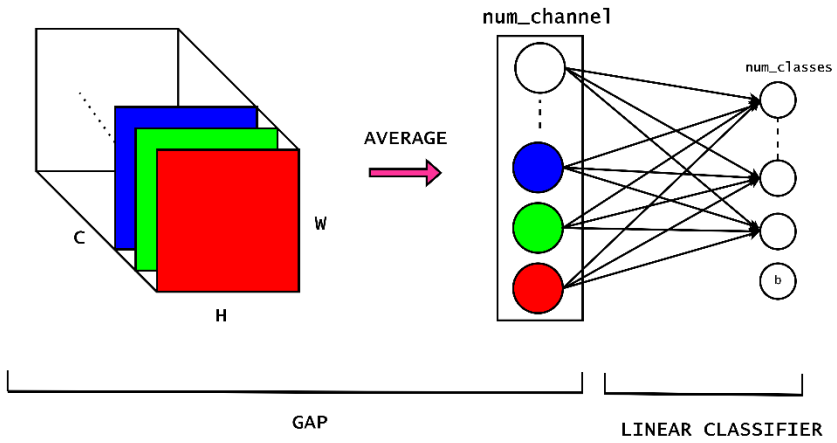
Sebagai simulasi, sebuah fitur berukuran 4×4 piksel dengan 2 kanal masuk ke sebuah blok *ConvNeXt*. Input ini kita sebut x. Setelah melalui *depthwise convolution*, normalisasi, *pointwise convolution*, aktivasi *GELU*, dan *pointwise*

convolution kedua, diperoleh hasil transformasi $F(x, W)$ dengan ukuran sama, yaitu $4 \times 4 \times 2$. Pada tahap *LayerScale*, setiap kanal hasil transformasi dikalikan dengan skalar γ. Misalnya $\gamma = [0.000001, 0.000001]$ di awal pelatihan, maka seluruh nilai pada kanal pertama dan kedua dari $F(x, W)$ dikalikan dengan angka sangat kecil, sehingga nilainya mendekati nol. Dengan demikian, saat dijumlahkan dengan *shortcut connection*, keluaran akhir y hampir identik dengan x. Hal ini membuat model lebih stabil pada iterasi awal karena tidak terjadi perubahan besar terhadap data. Seiring pelatihan, nilai γ menyesuaikan. Misalkan setelah beberapa *epoch*, γ berubah menjadi $[0.8,\ 1.2]$. Artinya kanal pertama dari $F(x, W)$ dikalikan 0.8 (sedikit direduksi kontribusinya), sedangkan kanal kedua dikalikan 1.2 (ditingkatkan kontribusinya). Setelah itu, kedua kanal hasil skala ini dijumlahkan dengan shortcut x. Jadi, untuk setiap posisi piksel (i, j), operasi yang terjadi adalah:

1. Output kanal 1 = $x(i, j, 1) + 0.8 \times F(x, W)(i, j, 1)$
2. Output kanal 2 = $x(i, j, 2) + 1.2 \times F(x, W)(i, j, 2)$

3.4.1.4 Global Average Pooling

Setelah melalui rangkaian *ConvNeXt Block* pada tahap akhir ekstraksi fitur, keluaran jaringan berupa kumpulan *feature map* dengan kedalaman yang setara dengan jumlah kanal pada blok terakhir. Pada arsitektur *ConvNeXt-Tiny*, *feature map* ini tidak langsung diratakan (*flatten*) dan diproses menggunakan *fully connected layer*, melainkan terlebih dahulu melewati tahap *Global Average Pooling* (*GAP*).



Gambar 3.11 Global Average Pooling

GAP bekerja dengan menghitung rata-rata nilai dari setiap *feature map* pada seluruh dimensi spasialnya, sehingga menghasilkan satu nilai representatif untuk setiap kanal. Proses ini ditunjukkan pada gambar 3.11 untuk menjaga keterhubungan langsung antara setiap kanal fitur dengan satu kategori keluaran, sehingga setiap *feature map* dapat diinterpretasikan sebagai peta kepercayaan (*confidence map*) terhadap suatu kelas (Lin et al., 2014). Selain itu, tidak adanya bobot yang perlu dilatih pada *GAP* menjadikannya lebih ringan secara komputasi dan secara alami berperan sebagai *structural regularizer* yang mengurangi risiko *overfitting*.

GAP digunakan sebagai tahap akhir setelah seluruh proses ekstraksi fitur konvolusional, sebelum masuk ke *layer* klasifikasi *softmax*. Integrasi *GAP* menjaga sifat *fully-convolutional* dari arsitektur, sehingga efisien dalam komputasi dan dapat menangani *input* dengan dimensi bervariasi tanpa penyesuaian struktur jaringan (Liu *et al.,* 2022).

Penggunaan *GAP* mempertahankan sifat *fully-convolutional* dari jaringan, memungkinkannya menerima masukan dengan dimensi yang bervariasi tanpa penyesuaian struktur (Liu *et al.*, 2022). Secara matematis, jika $F_k(x, y)$ menyatakan nilai aktivasi pada koordinat (x, y) dari *feature map* ke-k berukuran $H \times W$, maka nilai keluaran *GAP* untuk kanal tersebut diberikan oleh persamaan 3.10

$$g_k = \frac{1}{H \times W} \sum_{x=1}^{H} \sum_{y=1}^{W} F_k(x, y) \quad (3.10).$$

Keterangan:
g_k : Nilai keluaran *Global Average Pooling* untuk kanal *(feature map)* ke-k.
H : Tinggi *(height)* dari *feature map*.
W : Lebar (*width*) dari *feature map*.
x : Indeks posisi piksel pada dimensi tinggi *(height)*.
y : Indeks posisi piksel pada dimensi lebar *(width)*.
$F_k(x, y)$: Nilai aktivasi *(activation value)* pada koordinat (x, y) dari *feature map* ke-k.
$\frac{1}{H \times W}$: Faktor normalisasi untuk menghitung rata-rata seluruh nilai dalam *feature map*.
$\sum_{x=1}^{H} \sum_{y=1}^{W}$: Operasi penjumlahan dua dimensi yang menjumlahkan semua nilai aktivasi pada *feature map*.

Hasil vektor $[g_1, g_2, \ldots, g_K]$ dari proses ini kemudian menjadi masukan bagi *Linear Classifier* pada tahap berikutnya, yang akan mengubahnya menjadi skor prediksi untuk setiap kelas target. Misalkan pada akhir *ConvNeXt-Tiny*, diperoleh *feature map* dengan ukuran 4×4 piksel dan 3 kanal ($K = 3$). Kemudian, kanal pertama (F_1) memiliki nilai piksel [[2, 3, 1, 0], [1, 2, 2, 3], [0, 1, 2, 1], [3, 2, 1, 0]]. Jumlah seluruh elemen = 24. Karena ukuran 4×4 = 16 piksel, maka rata-rata $g_1 = \frac{24}{16} = 1.5$. Kanal kedua ($F_2$) misalnya memiliki jumlah total 32, sehingga rata-rata $g_2 = \frac{32}{16} = 2.0$. Kanal ketiga ($F_3$) misalnya memiliki jumlah total 40, sehingga rata-rata $g_3 = \frac{40}{16} = 2.5$. Maka, vektor hasil *GAP* adalah $[g_1, g_2, g_3] = [1.5, 2.0, 2.5]$. Vektor ini mewakili ringkasan informasi dari seluruh *feature map* tanpa lagi

memiliki dimensi spasial. Selanjutnya, vektor tersebut diteruskan ke *linear classifier* yang mengubahnya menjadi skor probabilitas melalui fungsi *softmax.*

3.4.1.5 Linear Classifier (Fully Connected)

Vektor keluaran dari *Global Average Pooling*, yang memiliki panjang sama dengan jumlah kanal pada *layer* terakhir, kemudian diproses oleh lapisan *Linear Classifier* untuk menghasilkan prediksi akhir. Pada *ConvNeXt-Tiny*, *Linear Classifier* ini direalisasikan sebagai fully connected layer tunggal yang mengubah vektor fitur berukuran K menjadi vektor skor prediksi untuk setiap kelas pada *dataset*. Secara matematis, jika $g \in \mathbb{R}^K$ merupakan vektor hasil *GAP* dan $W \in \mathbb{R}^{C \times K}$ adalah matriks bobot *linear* dengan C adalah jumlah kelas, maka skor prediksi $z \in \mathbb{R}^C$ dapat dihitung sebagaimana persamaan 3.11.

$$z = Wg + b \tag{3.11}$$

di mana $b \in \mathbb{R}^C$ adalah vektor bias. Skor ini kemudian diberikan ke fungsi *softmax* untuk menghasilkan probabilitas prediksi setiap kelas pada persamaan 3.12.

$$p_i = \frac{\exp(z_i)}{\sum_{j=1}^{C} \exp(z_j)}, \quad i = 1,2, \dots, \mathrm{C}\ (\tag{3.12}$$

Keterangan:
p_i : probabilitas prediksi untuk kelas ke-i
z_i : skor keluaran (logit) dari fully connected layer untuk kelas ke-i sebelum fungsi aktivasi softmax
C : jumlah total kelas pada *dataset*
$\exp(\cdot)$: fungsi eksponensial
$\sum_{j=1}^{C} \exp(z_j)$: penjumlahan seluruh skor eksponensial untuk normalisasi agar total probabilitas = 1

Linear classifier pada *ConvNeXt-Tiny* memiliki jumlah parameter yang relatif kecil dibandingkan arsitektur *CNN* konvensional yang menggunakan

beberapa lapisan *fully connected*, karena dimensi *input* ke lapisan ini telah direduksi oleh *GAP*. Pendekatan ini tidak hanya mengurangi kompleksitas model, tetapi juga mempertahankan efisiensi komputasi dan membantu menghindari *overfitting* (Liu *et al.*, 2022).

$$\begin{bmatrix} z_{\text{Sehat}} \\ z_{\text{Tidak Sehat}} \end{bmatrix} = \begin{bmatrix} 0.2 & 0.3 & 0.5 \\ 0.4 & 0.1 & 0.2 \end{bmatrix} \begin{bmatrix} 1.5 \\ 2.0 \\ 2.5 \end{bmatrix} + \begin{bmatrix} 0.10 \\ 0.05 \end{bmatrix} = \begin{bmatrix} 2.25 \\ 1.35 \end{bmatrix}$$

$$\mathbf{p} = \frac{1}{e^{2.25} + e^{1.35}} \begin{bmatrix} e^{2.25} \\ e^{1.35} \end{bmatrix} \approx \frac{1}{13.345} \begin{bmatrix} 9.488 \\ 3.857 \end{bmatrix} = \begin{bmatrix} 0.711 \\ 0.289 \end{bmatrix}.$$

Gambar 3.12 Contoh Perhitungan pada Linear Classifier

Sebagai gambaran, hasil GAP berupa vektor berdimensi tiga [1.5,2.0,2.5]. Vektor ini kemudian masuk ke lapisan *Linear Classifier* yang memiliki bobot dan bias berbeda untuk setiap kelas target seperti ditunjukkan pada gambar 3.12. Sebagai ilustrasi sederhana, misalkan jaringan ini ditujukan untuk mengklasifikasikan citra endoskopi ke dalam dua kelas: “Sehat” dan “Tidak Sehat”. Bobot yang digunakan untuk kelas pertama adalah [0.2,0.3,0.5] dengan *bias* 0.1, sedangkan bobot untuk kelas kedua adalah [0.4,0.1,0.2] dengan *bias* 0.05.

Proses perhitungan dilakukan dengan mengalikan setiap elemen vektor input dengan bobot yang sesuai, lalu menjumlahkannya dengan *bias.* Untuk kelas pertama, hasil perhitungan adalah $(1.5 \times 0.2) + (2.0 \times 0.3) + (2.5 \times 0.5) + 0.1 = 0.3 + 0.6 + 1.25 + 0.1 = 2.25$. Untuk kelas kedua, perhitungannya adalah $(1.5 \times 0.4) + (2.0 \times 0.1) + (2.5 \times 0.2) + 0.05 = 0.6 + 0.2 + 0.5 + 0.05 =$ 1.35. Dua nilai ini, yaitu 2.25 dan 1.35, disebut *logit* atau skor mentah sebelum normalisasi. Agar dapat ditafsirkan sebagai probabilitas, skor tersebut diproses

melalui fungsi *softmax.* Perhitungan *softmax* dilakukan dengan terlebih dahulu menghitung eksponensial dari masing-masing skor: $\exp(2.25) \approx$ $dan\ \exp(1.35) \approx 3.8$. Kemudian, masing-masing nilai dibagi dengan total keduanya $(9.49 + 3.86 = 13.35)$. Hasilnya adalah probabilitas $\frac{9.49}{13.35} \approx 0.71$ untuk kelas pertama dan $\frac{3.86}{13.35} \approx 0.29$ untuk kelas kedua. Dengan demikian, untuk contoh input ini, model memprediksi bahwa citra endoskopi memiliki peluang 71% termasuk ke dalam kelas "Sehat" dan 29% ke dalam kelas "Tidak Sehat".

Selama proses pelatihan, parameter W (bobot) dan b (*bias*) pada *Linear Classifier* diperbarui secara iteratif agar model dapat meminimalkan nilai *loss function*. Pada penelitian ini digunakan fungsi *loss Cross-Entropy*, yang mengukur seberapa jauh distribusi probabilitas prediksi p_i dari label sebenarnya y_i. Secara matematis, *loss* untuk satu sampel dapat dituliskan sebagai:

$$\mathrm{L} = -\sum_{i=1}^{C} y_i \log(p_i) \tag{3.14}$$

Keterangan:
y_i = 1 jika sampel termasuk kelas ke-i, dan 0 untuk kelas lainnya
p_i = probabilitas prediksi untuk kelas ke-i (hasil fungsi *softmax*)
C = jumlah total kelas

Gradien dari fungsi *loss* terhadap setiap parameter Wdan b dihitung melalui proses *backpropagation*. Nilai gradien ini menunjukkan arah perubahan yang harus dilakukan agar *loss* berkurang.

Kemudian, parameter diperbarui menggunakan algoritma *optimizer* Adam yang mengombinasikan konsep momentum dan *adaptive learning rate*. Secara umum, pembaruan bobot dilakukan berdasarkan persamaan:

$$W^{(t+1)} = W^{(t)} - \eta\, \frac{\partial L}{\partial W} \tag{3.14}$$

$$b^{(t+1)} = b^{(t)} - \eta \frac{\partial L}{\partial b} \tag{3.15}$$

Keterangan:
η : *learning rate*
$\frac{\partial L}{\partial W}$, $\frac{\partial L}{\partial b}$ = probabilitas prediksi untuk kelas ke-i (hasil fungsi *softmax*)

Proses ini diulang untuk setiap *batch* data pada seluruh *epoch* pelatihan hingga nilai *loss* konvergen atau mencapai batas *epoch* yang ditentukan. Dengan cara ini, *Linear Classifier* secara bertahap belajar untuk menghasilkan bobot dan bias yang menghasilkan prediksi paling akurat terhadap data endoskopi yang diberikan.

3.5 Evaluasi

Evaluasi kinerja model pada penelitian ini bertujuan untuk menilai sejauh mana arsitektur *ConvNeXt-Tiny* mampu mengklasifikasikan citra endoskopi menjadi kategori *Gastroesophageal Reflux Disease* (*GERD*) atau polip usus. Mengingat pentingnya akurasi dalam diagnosis medis, proses evaluasi tidak hanya berfokus pada satu metrik, melainkan menggunakan *confusion matrix* dan empat metrik turunan, yaitu akurasi, presisi, *recall*, dan *F1-score* seperti yang diusulkan oleh Powers (2011).

Confusion matrix merupakan tabel yang membandingkan hasil prediksi model dengan label sebenarnya, yang tersusun dari empat komponen: *True Positive* (*TP*), yaitu jumlah citra yang benar terprediksi sebagai kelas positif; *True Negative* (*TN*), jumlah citra yang benar terprediksi sebagai kelas negatif; *False Positive* (*FP*), jumlah citra yang salah terprediksi sebagai kelas positif; dan *False Negative* (*FN*), jumlah citra yang salah terprediksi sebagai kelas negatif. Dengan memanfaatkan

confusion matrix, analisis kinerja model dapat dilakukan secara lebih mendalam karena setiap jenis kesalahan dapat diidentifikasi secara spesifik.

Akurasi digunakan untuk mengukur proporsi prediksi yang benar dibandingkan dengan keseluruhan data uji, yang dirumuskan sebagai berikut:

$$\text{Akurasi } (\%) = \left(\frac{\text{TP+TN}}{\text{TP+TN+FP+FN}}\right) \times 100\% \tag{3.14}$$

Meskipun akurasi memberikan gambaran umum kinerja model, metrik ini dapat menyesatkan jika jumlah data pada tiap kelas tidak seimbang. Oleh karena itu, digunakan juga *precision* dan *recall* untuk memberikan perspektif yang lebih komprehensif. *Precision* mengukur tingkat ketepatan prediksi positif, atau seberapa besar proporsi prediksi positif yang benar-benar positif, dengan rumus:

$$\text{Presisi } (\%) = \left(\frac{\text{TP}}{\text{TP+FP}}\right) \times 100\% \tag{3.15}$$

Nilai *precision* yang tinggi menunjukkan bahwa model jarang memberikan prediksi positif yang keliru (*false positive* rendah), yang sangat penting dalam konteks medis agar pasien sehat tidak salah terdiagnosis. Sementara itu, *recall* mengukur kemampuan model dalam menemukan semua sampel positif yang sebenarnya ada, dengan rumus:

$$\textit{Recall } (\%) = \left(\frac{\text{TP}}{\text{TP+FN}}\right) \times 100\% \tag{3.16}$$

Recall yang tinggi berarti *false negative* rendah, sehingga model jarang melewatkan pasien yang sebenarnya sakit. Dalam bidang kesehatan, *recall* sering menjadi metrik prioritas karena kesalahan melewatkan diagnosis dapat berakibat fatal. Untuk mendapatkan keseimbangan antara *precision* dan *recall*, digunakan *F1-score*, yang merupakan rata-rata harmonis keduanya. Rumusnya adalah:

$$\text{F1 Score (\%)} = \left(\frac{2\times\text{TP}}{2\text{TP}+\text{FP}+\text{FN}}\right) \times 100\% \quad (3.17)$$

Nilai *F1-score* yang tinggi menunjukkan bahwa model tidak hanya akurat dalam memberikan prediksi positif, tetapi juga konsisten dalam mendeteksi semua kasus positif. Dengan kombinasi kelima ukuran evaluasi ini, performa model dapat dinilai secara lebih menyeluruh, sehingga hasil yang diperoleh tidak bias terhadap satu jenis metrik saja.

3.6 Skenario Pengujian

Pada tahap ini dilakukan serangkaian pengujian untuk mengevaluasi performa model *ConvNeXt-Tiny* dalam mengklasifikasikan citra endoskopi menjadi tiga kategori, yaitu *GERD*, polip, dan normal. Skenario pengujian dirancang secara sistematis untuk menganalisis pengaruh beberapa faktor terhadap kinerja model, baik dari sisi data maupun konfigurasi pelatihan. Dalam penelitian ini ditetapkan dua faktor utama (skenario mayor) dan satu faktor tambahan (skenario minor). Dua faktor utama adalah penggunaan augmentasi data dan penerapan normalisasi berbasis distribusi *z-score* dengan nilai *mean* dan standar deviasi dari *ImageNet,* sedangkan faktor tambahan adalah variasi ukuran *batch* (*batch size*).

Skenario pertama adalah pengujian dengan dan tanpa augmentasi data. Augmentasi merupakan teknik yang bertujuan meningkatkan keragaman data latih tanpa menambah jumlah data asli, sehingga model diharapkan mampu melakukan generalisasi dengan lebih baik. Bentuk augmentasi yang digunakan dalam penelitian ini antara lain rotasi, *flipping* horizontal, dan penyesuaian kecerahan secara acak. Skenario ini dibagi menjadi dua opsi, yaitu 1a (dengan augmentasi) dan 1b (tanpa augmentasi).

Skenario kedua adalah pengujian dengan dan tanpa normalisasi data. Normalisasi digunakan untuk menyeragamkan distribusi intensitas piksel pada citra agar proses pembelajaran menjadi lebih stabil dan cepat konvergen. Normalisasi dilakukan menggunakan metode *z-score* dengan parameter *mean* dan standar deviasi yang diadopsi dari *dataset ImageNet*. Skenario ini juga dibagi menjadi dua opsi, yaitu 2a (dengan normalisasi) dan 2b (tanpa normalisasi).

Skenario ketiga adalah variasi ukuran *batch* sebagai faktor minor. Ukuran *batch* memengaruhi frekuensi pembaruan bobot, waktu pelatihan, serta penggunaan memori. Tiga opsi *batch size* yang digunakan dalam penelitian ini adalah 16 (3a), 32 (3b), dan 64 (3c). *Batch* yang lebih kecil memberikan pembaruan bobot yang lebih sering sehingga pembelajaran lebih halus, tetapi membutuhkan waktu pelatihan lebih lama. Sebaliknya, *batch* yang lebih besar mempercepat pelatihan namun berpotensi membuat model kurang sensitif terhadap variasi data.

Ketiga skenario tersebut kemudian dikombinasikan untuk membentuk konfigurasi pengujian. Dengan dua opsi augmentasi, dua opsi normalisasi, dan tiga opsi *batch size*, dihasilkan total 12 kombinasi pengujian. Sebelum proses pelatihan dilakukan, *dataset* dibagi menjadi tiga subset dengan perbandingan 7:2:1, yaitu 70% data digunakan untuk pelatihan (*training*), 20% untuk validasi (*validation*), dan 10% untuk pengujian (*testing*). Pembagian ini dilakukan secara acak untuk memastikan distribusi kelas yang seimbang pada setiap subset data, sehingga hasil evaluasi model dapat merepresentasikan performa sebenarnya terhadap data yang belum pernah dilihat. Setiap kombinasi akan dilatih dengan jumlah *epoch* yang sama yaitu 10, serta menggunakan parameter pelatihan lainnya yang ditetapkan

konstan agar perbedaan kinerja yang dihasilkan murni disebabkan oleh variasi pada skenario pengujian. Parameter tetap tersebut meliputi *optimizer* Adam sebagai algoritma pembaruan bobot, fungsi *loss CrossEntropy* untuk klasifikasi multikelas, serta nilai *learning rate* sebesar 0,001. Kombinasi skenario pengujian ditunjukkan pada Tabel 3.5 berikut.

Tabel 3.3 Kombinasi Skenario Pengujian

Nama Skenario	**Kombinasi**	**Deskripsi**
A	1a-2a-3a	Augmentasi aktif, normalisasi aktif, *batch size* 16
B	1a-2a-3b	Augmentasi aktif, normalisasi aktif, *batch size* 32
C	1a-2a-3c	Augmentasi aktif, normalisasi aktif, *batch size* 64
D	1a-2b-3a	Augmentasi aktif, normalisasi tidak dilakukan, *batch size* 16
E	1a-2b-3b	Augmentasi aktif, normalisasi tidak dilakukan, *batch size* 32
F	1a-2b-3c	Augmentasi aktif, normalisasi tidak dilakukan, *batch size* 64
G	1b-2a-3a	Augmentasi tidak dilakukan, normalisasi aktif, *batch size* 16
H	1b-2a-3b	Augmentasi tidak dilakukan, normalisasi aktif, *batch size* 32
I	1b-2a-3c	Augmentasi tidak dilakukan, normalisasi aktif, *batch size* 64
J	1b-2b-3a	Augmentasi tidak dilakukan, normalisasi tidak dilakukan, *batch size* 16
K	1b-2b-3b	Augmentasi tidak dilakukan, normalisasi tidak dilakukan, *batch size* 32
L	1b-2b-3c	Augmentasi tidak dilakukan, normalisasi tidak dilakukan, *batch size* 64

Dengan pengaturan skenario seperti di atas, hasil pengujian diharapkan mampu menunjukkan secara jelas pengaruh penggunaan augmentasi data, penerapan normalisasi, serta variasi *batch size* terhadap performa model *ConvNeXt-Tiny*. Pendekatan ini juga memungkinkan analisis mendalam mengenai kombinasi

parameter yang menghasilkan akurasi terbaik sekaligus mempertahankan efisiensi pelatihan.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Konfigurasi Eksperimen

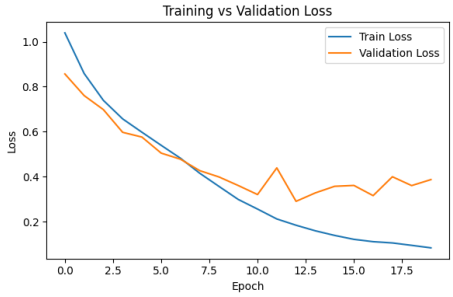
Pelatihan model dilakukan menggunakan *Google Colab* dengan akselerator *GPU NVIDIA Tesla T4* (16 GB). Implementasi dilakukan menggunakan *framework PyTorch* dengan *learning rate* awal sebesar 0,001 dan algoritma optimisasi *Adam*. Fungsi *loss* yang digunakan adalah *Cross-Entropy Loss*, yang sesuai untuk kasus klasifikasi multikelas. Proses pelatihan dijalankan selama maksimum 20 *epoch*, dengan penerapan *early stopping* menggunakan nilai *patience* sebesar 10 *epoch* untuk mencegah *overfitting* dan menjaga efisiensi pelatihan. Seluruh proses menggunakan *random seed* 42 untuk menjaga reprodusibilitas hasil.

Dataset yang digunakan berasal dari *GastroEndoNet* versi 3 (Bitto *et al.*, 2025), yang berisi citra endoskopi lambung dan usus dengan empat kategori: *GERD, GERD* Normal, *Polyp, dan Polyp* Normal. *Dataset* tersebut telah menyediakan dua versi data, yaitu *original* dan *augmented*, sehingga proses augmentasi tidak dilakukan secara manual pada tahap pelatihan, melainkan disesuaikan berdasarkan skenario pengujian. Total citra yang tersedia adalah 4.006 citra asli dan 24.036 citra hasil augmentasi. Seluruh citra berukuran 224×224 piksel dan dibagi menjadi tiga bagian dengan rasio 70% data latih, 20% data validasi, dan 10% data uji, menggunakan pembagian *stratified* agar distribusi antar kelas tetap seimbang.

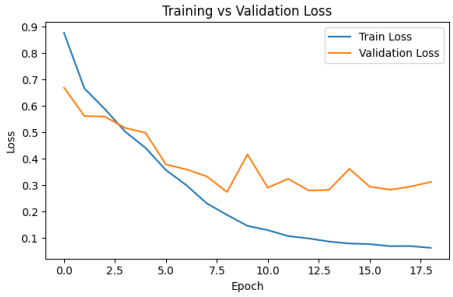
4.2 Hasil Uji Coba

Sebelum membahas performa kuantitatif dan perbandingan antar skenario, perlu dilakukan peninjauan terhadap dinamika proses pelatihan untuk memastikan bahwa model telah mencapai konvergensi dengan stabil. Analisis ini dilakukan melalui visualisasi perubahan *training loss* dan *validation loss* selama 20 epoch pelatihan pada setiap skenario. Pola *loss* memberikan gambaran tentang sejauh mana model mampu menyesuaikan bobot internalnya terhadap data pelatihan tanpa kehilangan kemampuan generalisasi terhadap data validasi.

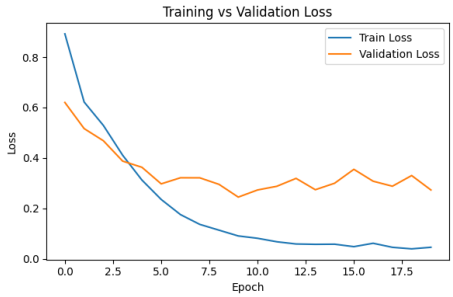
Gambar 4.1 berikut menampilkan visualisasi perubahan *training loss* dan *validation loss* pada enam skenario yang menggunakan augmentasi data (1a–2a–3a hingga 1a–2b–3c). Pola yang tampak menunjukkan tren penurunan yang konsisten pada kedua kurva, baik selama proses pelatihan maupun validasi, yang menandakan bahwa proses optimasi parameter berjalan stabil dan model mampu melakukan generalisasi dengan baik terhadap data validasi. Pada awal pelatihan, *training loss* umumnya bernilai tinggi karena bobot inisialisasi acak belum merepresentasikan pola citra, namun setelah beberapa epoch, kurva mulai menurun tajam dan berangsur mendatar di bawah nilai 0.2 pada epoch ke-15 hingga ke-20. Pola ini sejalan dengan penurunan *validation loss* yang relatif berirama dengan *training loss*, hanya berbeda sedikit pada titik-titik tertentu akibat fluktuasi distribusi *batch*. Tidak terlihat indikasi *overfitting* yang signifikan, karena jarak antara kedua kurva tetap kecil dan tidak terjadi divergensi di akhir pelatihan.



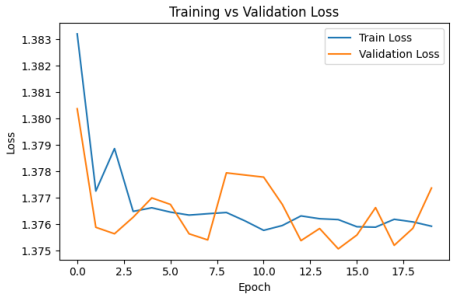
(Skenario A)



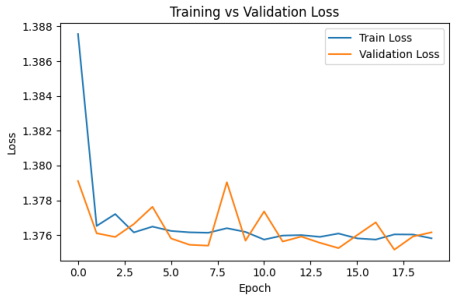
(Skenario B)



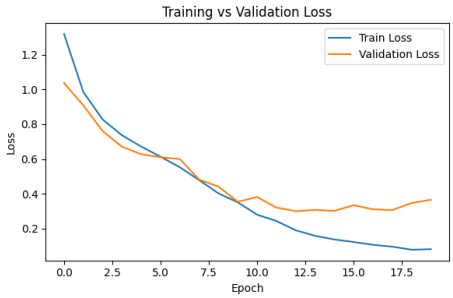
(Skenario C)



(Skenario D)



(Skenario E)



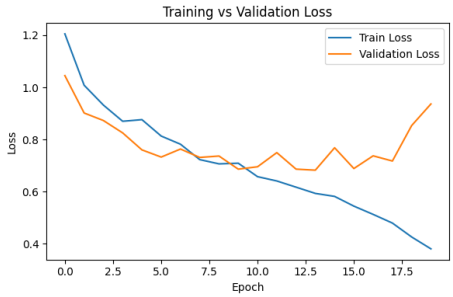
(Skenario F)

Gambar 4.1 Visualisasi *loss* pada data dengan augmentasi (Skenario A-F)

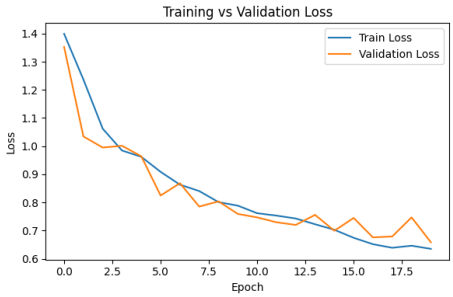
Kestabilan konvergensi ini menunjukkan bahwa penerapan augmentasi memberikan dampak positif terhadap pembelajaran model. Variasi tambahan pada citra pelatihan seperti rotasi, pencahayaan, dan *flipping* memperluas distribusi data, sehingga model belajar mengenali pola tekstur dan bentuk mukosa yang lebih beragam tanpa kehilangan kemampuan generalisasi. Skenario dengan *batch size* besar (misalnya 64) memperlihatkan kurva yang lebih halus dibanding *batch size*

kecil, yang menandakan perataan gradien lebih stabil akibat agregasi sampel yang lebih luas. Selain itu, efek normalisasi turut menjaga kestabilan nilai aktivasi antar *batch*, menjadikan proses pembaruan bobot lebih terkontrol.

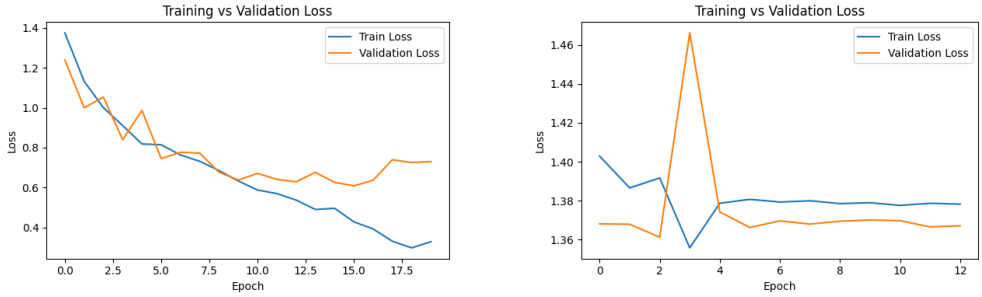
Berbeda dengan kelompok sebelumnya, keenam grafik pada Gambar 4.2 memperlihatkan pola *training loss* dan *validation loss* pada skenario tanpa penerapan augmentasi data. Secara umum, kurva menunjukkan kecenderungan penurunan di awal pelatihan, namun konvergensinya tidak sehalus kelompok dengan augmentasi. Beberapa skenario memperlihatkan fluktuasi cukup tajam pada *validation loss* setelah *epoch* ke-10, bahkan terdapat kecenderungan divergensi di mana nilai *validation loss* meningkat sementara *training loss* terus menurun. Fenomena ini mengindikasikan terjadinya *overfitting*, yaitu kondisi ketika model terlalu menyesuaikan diri terhadap pola data pelatihan dan kehilangan kemampuan generalisasi terhadap data validasi. Penyebab utama gejala tersebut adalah keterbatasan variasi data pelatihan akibat absennya augmentasi, sehingga distribusi citra yang diterima model tidak cukup beragam untuk merepresentasikan kondisi dunia nyata.



(Skenario G)



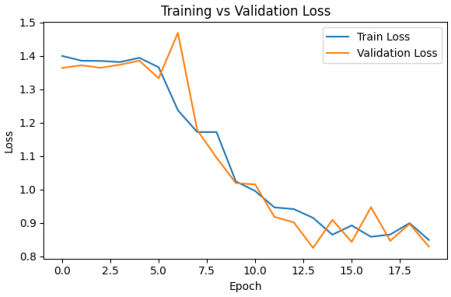
(Skenario H)



(Skenario I)

(Skenario J)

(Skenario K)



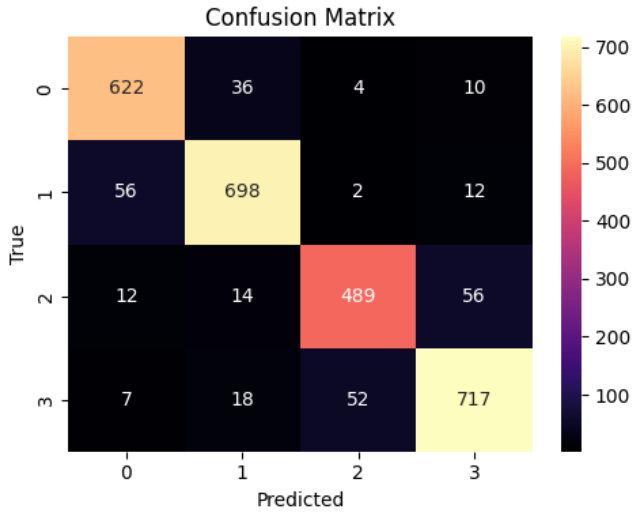
(Skenario L)

Gambar 4.2 Visualisasi *loss* pada *dataset* tanpa augmentasi (Skenario J-L)

Selain itu, pola pelatihan tanpa augmentasi memperlihatkan penurunan *training loss* yang cepat namun tidak diikuti dengan stabilitas pada *validation loss*, menandakan bahwa model belajar terlalu cepat terhadap fitur-fitur dominan tertentu tetapi gagal mempertahankan performa pada variasi minor. Hal ini terlihat pada grafik dengan *batch size* kecil (misalnya 16), di mana fluktuasi validasi cenderung ekstrem karena pembaruan bobot dilakukan pada subset data yang terlalu terbatas. Sebaliknya, pada *batch size* besar (64), fluktuasi sedikit teredam namun tidak menghasilkan peningkatan signifikan pada *validation accuracy*, menandakan bahwa peningkatan ukuran *batch* tidak mampu menggantikan fungsi augmentasi dalam memperkaya distribusi data.

4.2.1 Skenario A

Skenario A merupakan konfigurasi dasar dengan *augmentation* dan *normalization* aktif menggunakan *batch size* 16. Berdasarkan Gambar 4.3, distribusi prediksi relatif seimbang dan mayoritas berada pada diagonal, menunjukkan klasifikasi yang cukup baik. Kesalahan terutama muncul pada pasangan kelas dengan kemiripan visual, seperti *GERD–GERD* Normal dan Polyp–Polyp Normal.



Gambar 4.3 *Confusion matrix* hasil pengujian model *ConvNeXt-Tiny* pada Skenario A.

Tabel 4.1 Hasil evaluasi metrik kuantitatif model pada Skenario A.

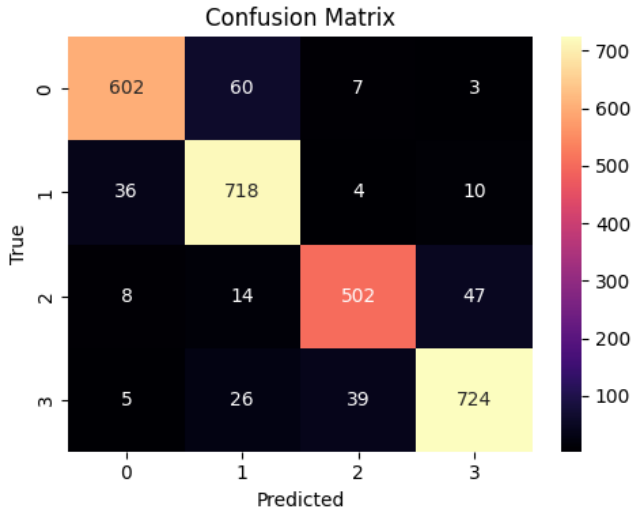
Metrik	**Nilai (%)**
Accuracy	90.05
Precision	89.99
Recall	89.85
F1-score	89.9

Nilai metrik pada Tabel 4.1 menunjukkan *accuracy* 90.05%, *precision* 89.99%, *recall* 89.85%, dan *F1-score* 89.90%. Kedekatan antar metrik menandakan performa yang stabil. *Batch size* kecil memberi variasi gradien yang tinggi sehingga

model menangkap detail tekstur secara baik, namun menyebabkan fluktuasi kecil pada proses validasi. Secara keseluruhan, Skenario A memberikan performa awal yang solid dengan metrik sekitar 0.90.

4.2.1 Skenario B

Pengujian berikutnya pada Skenario B menggunakan konfigurasi yang sama dengan Skenario A, dengan *batch size* yang meningkat menjadi 32. Gambar 4.4 menunjukkan pola diagonal yang lebih kuat dibandingkan Skenario A, dengan peningkatan akurasi terutama pada *GERD* Normal dan *Polyp* Normal. Kesalahan antar kelas lebih kecil dan lebih merata.



Gambar 4.4 *Confusion matrix* hasil pengujian model *ConvNeXt-Tiny* pada Skenario B.

Seperti ditampilkan pada Tabel 4.2, model mencapai *accuracy* 90.77%, *precision* 90.88%, *recall* 90.54%, dan *F1-score* 90.68%. Perbedaan antar metrik kecil (≤0.004), menunjukkan bahwa model semakin stabil dan seimbang. Peningkatan *batch size* terbukti mengurangi fluktuasi *validation loss* dan memperbaiki konsistensi prediksi. Dengan performa seluruh metrik di atas 90%,

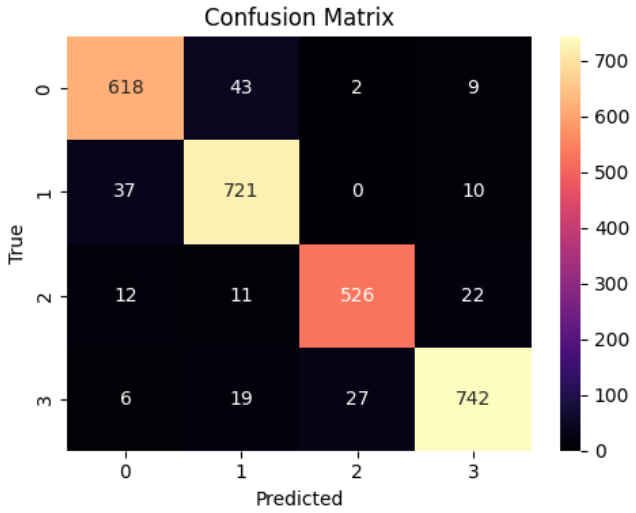
Skenario B dapat dianggap sebagai konfigurasi yang lebih stabil dibandingkan Skenario A.

Tabel 4.2 Hasil evaluasi metrik kuantitatif model pada Skenario B.

Metrik	**Nilai (%)**
Accuracy	90.77
Precision	90.88
Recall	90.54
F1-score	90.68

4.2.1 Skenario C

Skenario C menguji efek *batch size* yang lebih besar, yaitu 64, dengan *augmentation* dan *normalization* tetap aktif. Gambar 4.5 menunjukkan dominasi diagonal yang lebih kuat dibandingkan dua skenario sebelumnya. Kesalahan semakin berkurang, terutama antara *GERD–GERD* Normal, dan konsistensi prediksi meningkat signifikan. Polyp Normal menunjukkan jumlah prediksi benar tertinggi.



Gambar 4.5 *Confusion matrix* hasil pengujian model *ConvNeXt-Tiny* pada Skenario C.

Tabel 4.3 menunjukkan *accuracy* 92.94%, *precision* 93.04%, *recall* 92.85%, dan *F1-score* 92.94%. Keempat metrik sangat berdekatan dan berada di atas 92%, menandakan konvergensi yang stabil serta pembelajaran fitur global yang lebih baik. Kurva *training–validation loss* yang halus (lihat bagian sebelumnya) mendukung bahwa model berlatih tanpa tanda *overfitting*. Dengan hasil tertinggi di antara seluruh konfigurasi, Skenario C ditetapkan sebagai model terbaik untuk analisis lanjutan.

Tabel 4.3 Hasil evaluasi metrik kuantitatif model pada Skenario C.

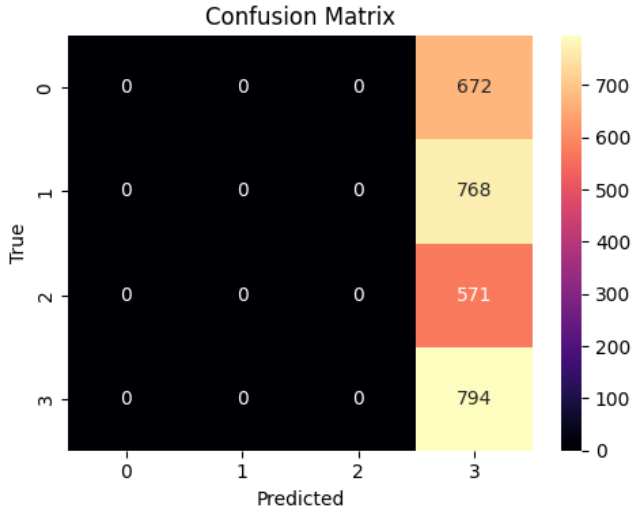
Metrik	**Nilai (%)**
Accuracy	92.94
Precision	93.04
Recall	92.85
F1-score	92.94

Skenario C menunjukkan performa tertinggi dari seluruh konfigurasi dengan augmentasi aktif. Nilai metrik yang konsisten di atas 92% menegaskan bahwa penggunaan *batch size* besar membantu memperhalus gradien dan memperkuat pembelajaran fitur global tanpa kehilangan detail tekstur. Kurva *loss* yang halus dan jarak kecil antara *training* dan *validation loss* (ditampilkan pada bagian sebelumnya) memperkuat bukti bahwa model mencapai konvergensi sempurna tanpa gejala *overfitting*. Dengan demikian, Skenario C ditetapkan sebagai model terbaik secara keseluruhan untuk tahap evaluasi mendalam berikutnya, baik dari sisi kuantitatif maupun kualitatif.

4.2.1 Skenario D

Konfigurasi berikutnya, Skenario D merupakan konfigurasi tanpa *normalization* dengan *augmentation* aktif dan *batch size* 16. Gambar 4.6

menunjukkan bahwa model gagal melakukan klasifikasi: seluruh citra diprediksi sebagai Polyp Normal sehingga diagonal *confusion matrix* hampir seluruhnya kosong. Ketidakmampuan ini mengindikasikan hilangnya stabilitas distribusi fitur akibat absennya normalisasi, sehingga gradien menjadi tidak terkontrol.



Gambar 4.6 *Confusion matrix* hasil pengujian model *ConvNeXt-Tiny* pada Skenario D.

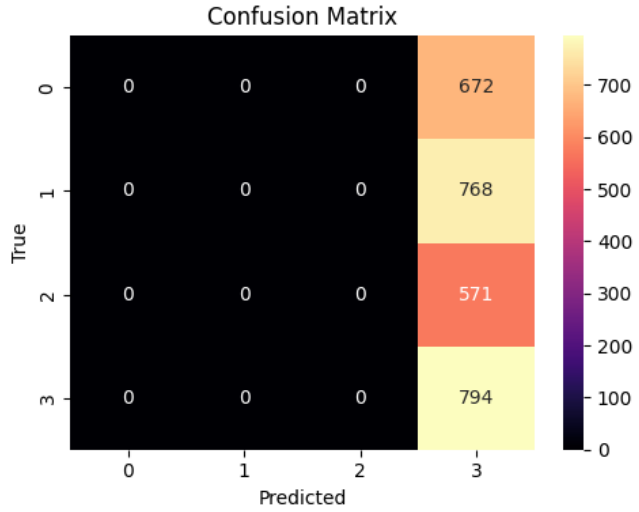
Tabel 4.4 memperlihatkan performa yang sangat rendah: *accuracy* 28.31%, *precision* 07.08%, *recall* 25%, dan *F1-score* 11.03%. Pola ini mencerminkan *mode collapse*, di mana model memilih satu kelas dominan untuk meminimalkan *loss*. Hasil ini menegaskan bahwa normalisasi berperan penting dalam menjaga konsistensi skala fitur dan stabilitas konvergensi.

Tabel 4.4 Hasil evaluasi metrik kuantitatif model pada Skenario D.

Metrik	**Nilai (%)**
Accuracy	28.31
Precision	7.08
Recall	25.00
F1-score	11.03

4.2.1 Skenario E

Pengamatan terhadap dampak ukuran batch yang lebih besar dalam kondisi tanpa *normalization* dilakukan pada Skenario E dengan meningkatkan *batch size* menjadi 32. Tujuannya adalah untuk menilai apakah ukuran batch yang lebih besar dapat mengurangi instabilitas yang muncul tanpa proses normalisasi. Namun, Gambar 4.7 menunjukkan hasil yang identik dengan skenario sebelumnya.



Gambar 4.7 *Confusion matrix* hasil pengujian model *ConvNeXt-Tiny* pada Skenario E.

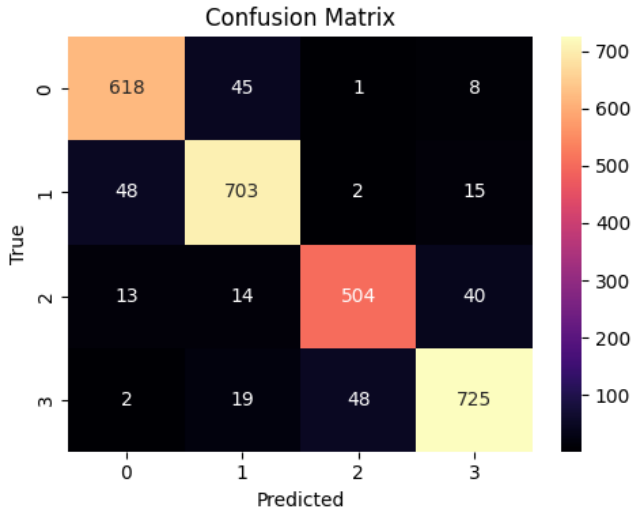
Nilai metrik pada Tabel 4.5 juga sama persis dengan Skenario D (*accuracy* 2831% dan *F1-score* 11.03%), sehingga dapat disimpulkan bahwa peningkatan *batch size* tidak mampu mengimbangi hilangnya proses normalisasi. Tanpa penyeimbangan distribusi fitur antar *batch*, model tetap gagal mempelajari representasi antar kelas meskipun jumlah sampel per *batch* diperbesar.

Tabel 4.5 Hasil evaluasi metrik kuantitatif model pada Skenario E.

Metrik	**Nilai (%)**
Accuracy	28.31
Precision	07.08
Recall	25.00
F1-score	11.03

4.2.1 Skenario F

Berbeda dengan dua skenario sebelumnya, Skenario F menguji *batch size* 64 pada kondisi tanpa *normalization*. Berbeda dengan dua skenario sebelumnya, Gambar 4.8 menunjukkan pemulihan performa yang signifikan. Diagonal *confusion matrix* kembali dominan dan seluruh kelas dapat dikenali dengan baik. Kelas *GERD* Normal dan Polyp Normal memperoleh prediksi benar tertinggi.



Gambar 4.8 *Confusion matrix* hasil pengujian model *ConvNeXt-Tiny* pada Skenario F.

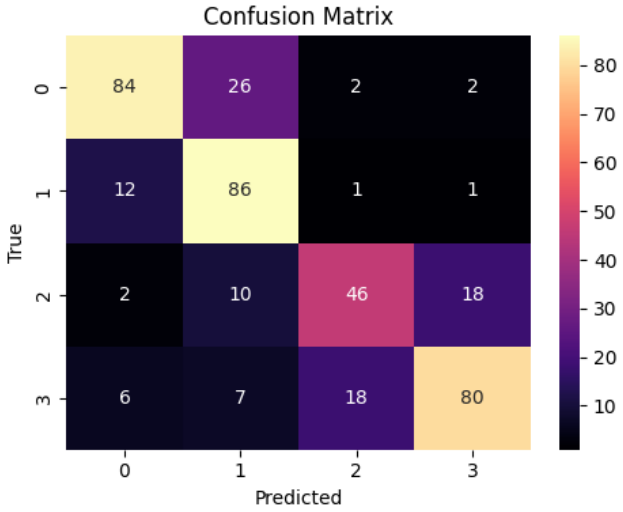
Tabel 4.6 menunjukkan *accuracy* 90.91%, *precision* 90.89%, *recall* 90.77%, dan *F1-score* 90.82%. Konsistensi antar metrik menandakan bahwa *batch size* besar menghasilkan gradien yang lebih stabil meskipun tanpa normalisasi. Namun performanya tetap sedikit di bawah kelompok dengan normalisasi aktif.

Tabel 4.6 Hasil evaluasi metrik kuantitatif model pada Skenario F.

Metrik	**Nilai (%)**
Accuracy	90.91
Precision	90.89
Recall	90.77
F1-score	90.82

4.2.1 Skenario G

Skenario G menguji konfigurasi tanpa *augmentation* dengan *normalization* aktif dan *batch size* 16. Tanpa augmentasi, variasi data menjadi terbatas sehingga kemampuan generalisasi model berpotensi menurun.



Gambar 4.9 *Confusion matrix* hasil pengujian model *ConvNeXt-Tiny* pada Skenario G.

Gambar 4.9 menunjukkan bahwa model masih dapat melakukan klasifikasi dengan cukup baik, namun kesalahan meningkat dibandingkan skenario dengan augmentasi aktif. Kelas *GERD* Normal dan *GERD* memperoleh prediksi benar tertinggi, sedangkan kelas Polyp dan Polyp Normal menunjukkan tumpang tindih besar. Hal ini menandakan bahwa model kesulitan membedakan variasi tekstur yang lebih kompleks.

Tabel 4.7 menunjukkan *accuracy* 73.82%, *precision* 73.83%, *recall* 73.07%, dan *F1-score* 73.00%. Nilai metrik yang konsisten menunjukkan stabilitas, namun penurunan umum pada performa menandakan adanya *underfitting* akibat

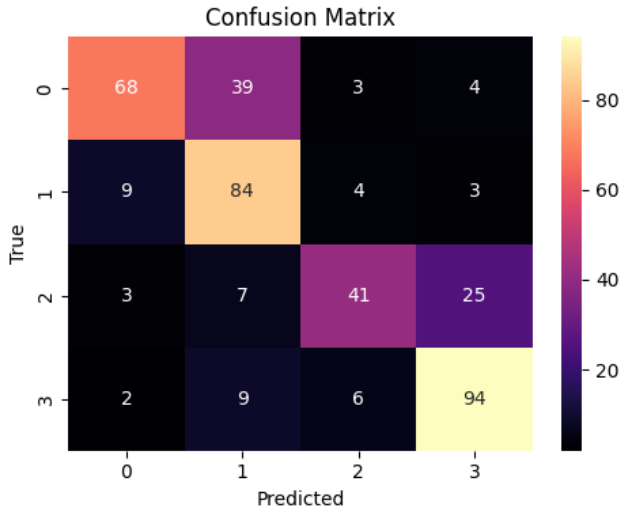
kurangnya variasi visual. Dengan demikian, normalisasi saja tidak cukup untuk mempertahankan performa optimal tanpa dukungan augmentasi.

Tabel 4.7 Hasil evaluasi metrik kuantitatif model pada Skenario G.

Metrik	**Nilai (%)**
Accuracy	73.82
Precision	73.83
Recall	73.07
F1-score	73.00

4.2.1 Skenario H

Pada pengujian berikutnya, Skenario H diperlihatkan Gambar 4.10 menunjukkan bahwa performa menurun dibandingkan Skenario G. *GERD* Normal masih menjadi kelas dengan prediksi benar terbanyak, tetapi kesalahan meningkat pada *GERD* dan Polyp, terutama mis-klasifikasi menuju kelas dengan fitur visual lebih dominan. Hal ini menunjukkan kecenderungan *class bias* ketika variasi citra rendah.



Gambar 4.10 *Confusion matrix* hasil pengujian model *ConvNeXt-Tiny* pada Skenario H.

Tabel 4.8 menunjukkan *accuracy* 71.57%, *precision* 73.47%, *recall* 70.57%, dan *F1-score* 70.52%. Penurunan *recall* mengindikasikan semakin sulitnya model mengenali kelas minor. Secara keseluruhan, peningkatan *batch size* tidak memberikan perbaikan berarti tanpa augmentasi, dan keterbatasan variasi citra tetap menjadi faktor utama yang membatasi kinerja.

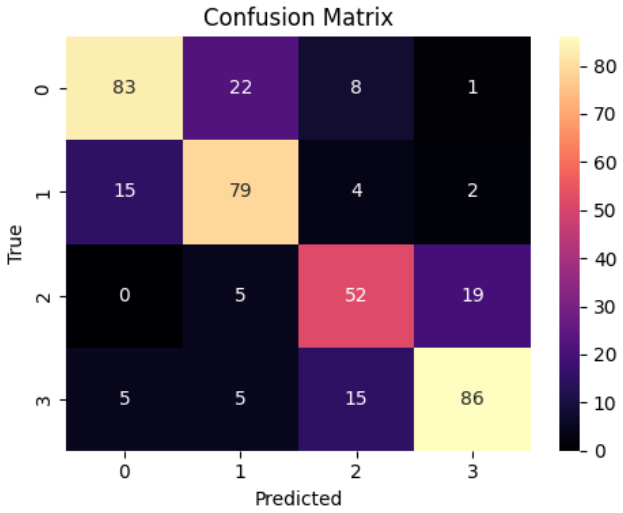
Tabel 4.8 Hasil evaluasi metrik kuantitatif model pada Skenario H.

Metrik	**Nilai (%)**
Accuracy	71.57
Precision	73.47
Recall	70.57
F1-score	70.52

4.2.1 Skenario I

Melanjutkan analisis tersebut, Skenario I pada Gambar 4.11 memperlihatkan adanya peningkatan akurasi di seluruh kelas, dengan *GERD* Normal dan Polyp Normal menjadi kelas yang paling konsisten dikenali. Meskipun tumpang tindih prediksi pada kelas Polyp masih terlihat, dominasi diagonal memperlihatkan adanya stabilisasi yang cukup kuat berkat ukuran *batch* yang besar.

Tabel 4.9 menunjukkan *accuracy* 74.81%, *precision* 74.30%, *recall* 74.43%, dan *F1-score* 74.25%. Perbaikan dibandingkan Skenario G dan H menunjukkan bahwa *batch size* besar membantu menghasilkan gradien yang lebih stabil sehingga performa meningkat meski tanpa augmentasi.



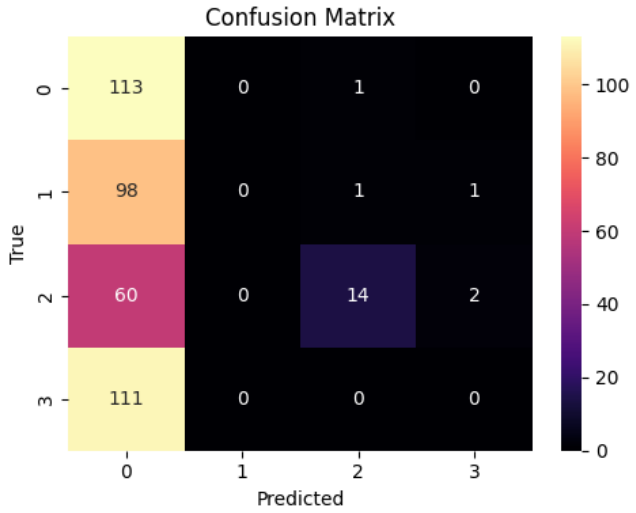
Gambar 4.11 *Confusion matrix* hasil pengujian model *ConvNeXt-Tiny* pada Skenario I.

Tabel 4.9 Hasil evaluasi metrik kuantitatif model pada Skenario I.

Metrik	**Nilai (%)**
Accuracy	74.81
Precision	7.43
Recall	74.43
F1-score	74.25

4.2.1 Skenario J

Berbeda dari Skenario I, Skenario J yang direpresentasikan oleh Gambar 4.12 menunjukkan penurunan performa yang sangat drastis: model hanya memprediksi dua kelas dominan (*GERD* dan Polyp), sementara hampir seluruh diagonal *confusion matrix* bernilai nol. Kelas Polyp hanya memperoleh 14 prediksi benar, dan tiga kelas lainnya didominasi kesalahan. Pola ini menunjukkan *mode collapse* akibat kurangnya stabilitas distribusi fitur.



Gambar 4.12 *Confusion matrix* hasil pengujian model *ConvNeXt-Tiny* pada Skenario J.

Tabel 4.10 menunjukkan *accuracy* 31.67%, *precision* 29.27%, *recall* 29.39% dan *F1-score* 19.00%. Seluruh metrik menunjukkan kegagalan total dalam pembelajaran. Hal ini menegaskan bahwa *augmentation* sebagai sumber keragaman dan *normalization* sebagai pengatur stabilitas merupakan komponen fundamental yang tidak dapat dihilangkan secara bersamaan.

Tabel 4.10 Hasil evaluasi metrik kuantitatif model pada Skenario J.

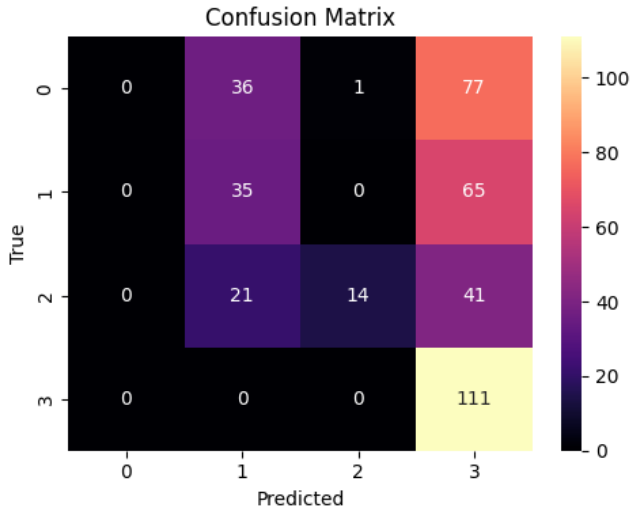
Metrik	**Nilai (%)**
Accuracy	31.67
Precision	29.27
Recall	29.39
F1-score	19.00

4.2.1 Skenario K

Penilaian terhadap kemungkinan peningkatan *batch size* dalam memperbaiki performa dilakukan pada Skenario K dengan memperbesar *batch size* menjadi 32 pada kondisi tanpa *augmentation* dan tanpa *normalization*. Langkah ini

bertujuan untuk menguji apakah ukuran *batch* yang lebih besar dapat meningkatkan performa yang sangat rendah pada Skenario J.

Gambar 4.13 menunjukkan sedikit perbaikan, namun model tetap gagal membedakan empat kelas. Tidak ada prediksi benar pada kelas *GERD*, dan sebagian besar sampel diarahkan ke kelas dominan seperti Polyp Normal. Kelas tersebut menjadi satu-satunya yang menunjukkan akurasi relatif lebih baik (111 prediksi benar), tetapi model masih bergantung pada pola global tanpa mampu menangkap perbedaan tekstur antar kelas.



Gambar 4.13 *Confusion matrix* hasil pengujian model *ConvNeXt-Tiny* pada Skenario K.

Tabel 4.11 memperlihatkan *accuracy* 39.90%, *precision* 42.28%, *recall* 38.36%, dan *F1-score* 30.51%. Walaupun sedikit lebih baik dari Skenario J, nilainya tetap rendah dan menunjukkan ketidakmampuan model untuk belajar secara stabil tanpa normalisasi maupun augmentasi. Hal ini mempertegas bahwa

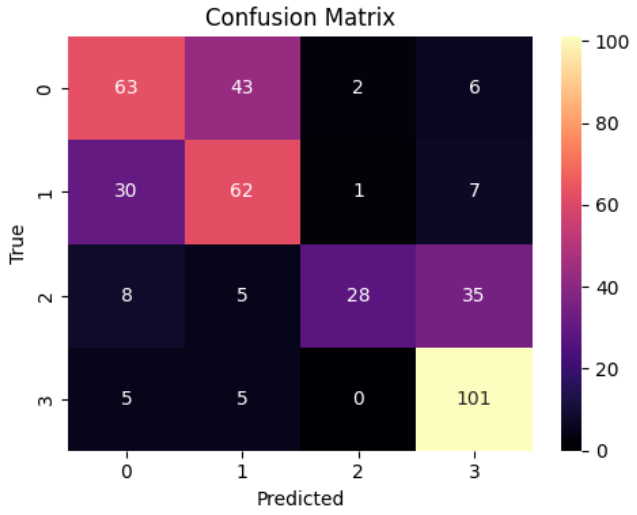
sekadar memperbesar *batch size* tidak cukup untuk mengimbangi hilangnya dua teknik penting tersebut.

Tabel 4.11 Hasil evaluasi metrik kuantitatif model pada Skenario K.

Metrik	**Nilai (%)**
Accuracy	39.90
Precision	42.28
Recall	38.36
F1-score	30.51

4.2.1 Skenario L

Hasil berbeda terlihat pada Skenario L. Gambar 4.14 menunjukkan peningkatan performa dibandingkan Skenario J dan K. Diagonal *confusion matrix* kembali terlihat, dengan Polyp Normal mencapai 101 prediksi benar. Namun kesalahan pada *GERD* dan *GERD* Normal masih tinggi, menandakan bahwa meskipun gradien lebih stabil, absennya normalisasi menyebabkan distribusi fitur antar *batch* tetap tidak konsisten.



Gambar 4.14 *Confusion matrix* hasil pengujian model *ConvNeXt-Tiny* pada Skenario L.

Tabel 4.12 menunjukkan *accuracy* 63.34%, *precision* 67.86%, *recall* 61.27%, dan *F1-score* 61.24%. Nilainya jauh lebih baik daripada Skenario J dan K, tetapi masih jauh dari performa optimal pada konfigurasi dengan augmentasi dan normalisasi aktif.

Tabel 4.12 Hasil evaluasi metrik kuantitatif model pada Skenario L.

Metrik	**Nilai (%)**
Accuracy	63.34
Precision	67.86
Recall	61.27
F1-score	61.24

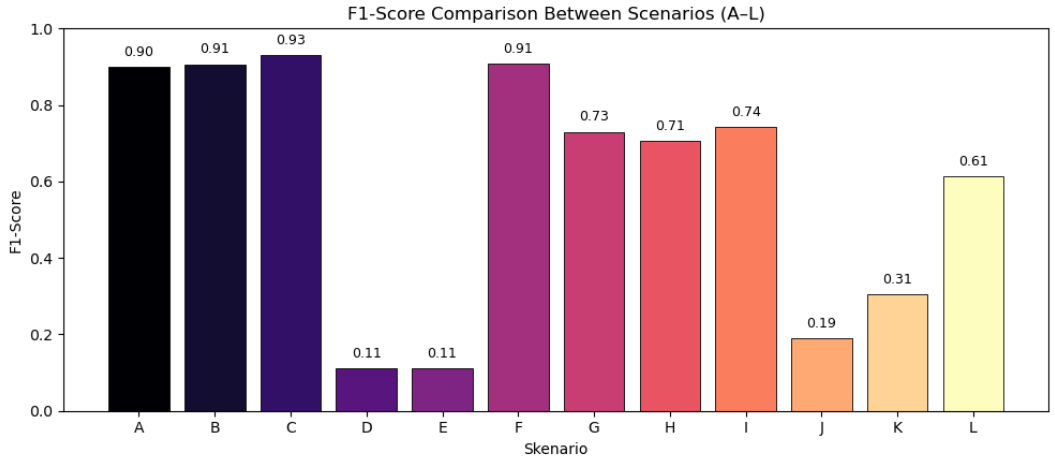
4.3 Analisis dan Pembahasan

Dua belas skenario yang diuji memperlihatkan pola yang konsisten: kombinasi augmentasi aktif + normalisasi aktif memberi kinerja tertinggi dan paling stabil; penghapusan normalisasi merusak pembelajaran hingga titik *mode collapse* (Skenario D–E), namun *batch* besar (64) dapat sedikit menstabilkan (Skenario F). Pada kelompok tanpa augmentasi tetapi normalisasi aktif (G–I), performa berada di tingkat menengah: kurva *loss* lebih stabil daripada tanpa normalisasi, tetapi akurasi dan F1 tetap lebih rendah karena kurangnya keragaman visual saat pelatihan. Kelompok tanpa augmentasi & tanpa normalisasi (J–L) menunjukkan degradasi signifikan; *batch* besar (L) memang membantu, tetapi tetap jauh dari konfigurasi optimal. Secara keseluruhan, Skenario C (1a–2a–3c) menjadi model terbaik dengan F1=0.9294, menunjukkan bahwa gradien yang halus (*batch* 64), keragaman data (augmentasi), dan kestabilan statistik (normalisasi) bekerja sinergis untuk memaksimalkan generalisasi. Tren ini konsisten dengan *confusion matrix* per skenario: kesalahan terbanyak terjadi pada pasangan kelas

berkarakteristik mirip (Polyp vs. Polyp Normal, *GERD* vs. *GERD* Normal), dan berkurang paling jelas pada skenario yang menggabungkan ketiga faktor di atas.

Tabel 4.13 Ringkasan metrik semua skenario (test set, macro-avg)

Skenario	**Kombinasi**	***Accuracy***	***Precision***	***Recall***	***F1-score***
A	1a–2a–3a	0.9005	0.8999	0.899	0.899
B	1a–2a–3b	0.9077	0.9088	0.905	0.9068
C	**1a–2a–3c**	**0.9294**	**0.9304**	**0.929**	**0.9294**
D	1a–2b–3a	0.2831	0.0708	0.25	0.1103
E	1a–2b–3b	0.2831	0.0708	0.25	0.1103
F	1a–2b–3c	0.9091	0.9089	0.908	0.9082
G	1b–2a–3a	0.7382	0.7383	0.731	0.73
H	1b–2a–3b	0.7157	0.7347	0.706	0.7052
I	1b–2a–3c	0.7481	0.743	0.744	0.7425
J	1b–2b–3a	0.3167	0.2927	0.294	0.19
K	1b–2b–3b	0.399	0.4228	0.384	0.3051
L	1b–2b–3c	0.6334	0.6786	0.613	0.6124



Gambar 4.15 Bar chart perbandingan F1 antar skenario

Berdasarkan Tabel 4.13, terlihat bahwa kinerja model sangat dipengaruhi oleh keberadaan normalisasi dan augmentasi. Nilai *accuracy* dan *F1-score* menunjukkan rentang perbedaan yang lebar antar skenario, dari 0.11 pada konfigurasi gagal (D–E) hingga 0.93 pada konfigurasi optimal (C). Tren kenaikan performa dari Skenario A → B → C menunjukkan efek positif dari peningkatan

batch size ketika dua teknik tersebut diaktifkan bersamaan, menandakan bahwa model memperoleh gradien yang lebih stabil dan pembelajaran fitur yang lebih representatif. Sebaliknya, ketika normalisasi dihilangkan (D–E), model kehilangan kemampuan diskriminatif secara total, ditunjukkan dengan *mode collapse* yang tercermin dari nilai metrik di bawah 0.3.

Sementara itu, Gambar 4.15 memperlihatkan perbandingan visual antar skenario dari segi *F1-score*, yang memperjelas kesenjangan performa antar kelompok. Batang tertinggi terdapat pada Skenario C, disusul F dan B, yang semuanya melibatkan augmentasi aktif. Kelompok tanpa augmentasi (G–I) menempati posisi menengah, sedangkan kelompok tanpa normalisasi (D–E, J–K) tampak mendekati dasar grafik, menandakan ketidakstabilan dan kegagalan generalisasi. Pola ini mengonfirmasi bahwa kombinasi augmentasi dan normalisasi memiliki kontribusi sinergis yang krusial dalam memperkuat pembelajaran spasial serta mengurangi *variance* antar *batch*. Dengan demikian, hasil rekapitulasi ini tidak hanya menunjukkan model terbaik secara numerik, tetapi juga memberikan pemahaman empiris tentang hubungan antarvariabel pelatihan terhadap performa akhir model *ConvNeXt-Tiny*.

4.3.1 Evaluasi Perubahan Variabel

Analisis pada bagian ini bertujuan untuk mengevaluasi pengaruh tiga variabel utama yang digunakan dalam seluruh skenario pelatihan, yaitu augmentasi data, normalisasi, dan ukuran *batch*, terhadap performa model *ConvNeXt-Tiny*. Evaluasi dilakukan dengan menghitung nilai rata-rata (*mean*) dan simpangan baku (*standard deviation*) dari *F1-score* pada setiap kelompok yang memiliki

konfigurasi sama untuk satu variabel, sementara dua variabel lainnya bervariasi. Ringkasan statistik tersebut disajikan pada Tabel 4.14, yang menjadi dasar untuk menilai kestabilan serta kontribusi relatif masing-masing variabel terhadap akurasi klasifikasi akhir.

Tabel 4.14 Ringkasan Statistik *F1-score* (Mean ± Std) berdasarkan Variabel Eksperimen

Variabel	Kondisi	Rata-rata F1 (Mean ± Std)	Interpretasi Utama
Augmentasi	Aktif (A–F)	**79.4% ± 31.6%**	Meningkatkan generalisasi model; efek terbesar pada *batch* besar.
	Tidak Aktif (G–L)	**55.4% ± 22.4%**	Model kehilangan variasi fitur, cenderung *underfitting*.
Normalisasi	Aktif (A–C, G–I)	**81.8% ± 9.1%**	Menjaga stabilitas dan konvergensi; performa konsisten tinggi.
	Tidak Aktif (D–F, J–L)	**37.4% ± 29.5%**	Tanpa normalisasi, gradien tidak stabil dan performa turun drastis.
Batch size	16 (A, D, G, J)	**58.2% ± 31.2%**	*Batch* kecil kurang stabil, rawan fluktuasi *loss*.
	32 (B, E, H, K)	**57.9% ± 30.4%**	Kinerja relatif sama; belum cukup menstabilkan tanpa normalisasi.
	64 (C, F, I, L)	**74.2% ± 26.1%**	*Batch* besar memperhalus gradien dan meningkatkan generalisasi.

1) Pengaruh Augmentasi Data

Konfigurasi dengan augmentasi aktif (A–F) menunjukkan rata-rata *F1-score* sebesar 79,4%, sedangkan kelompok tanpa augmentasi (G–L) hanya mencapai 55,4%. Perbedaan sekitar 24 persen ini menunjukkan bahwa augmentasi berperan penting dalam meningkatkan kemampuan generalisasi model, khususnya pada citra endoskopi yang memiliki variasi tinggi pada tekstur, pencahayaan, dan bentuk anatomi. Nilai rata-rata (*mean*) yang tinggi menunjukkan peningkatan

kinerja keseluruhan model, sedangkan simpangan baku (31,6%) yang cukup besar menandakan bahwa efek augmentasi masih bervariasi antar-skenario.

Temuan ini sejalan dengan penelitian Shorten & Khoshgoftaar (2019) yang menjelaskan bahwa *data augmentation* memperluas distribusi sampel melalui *geometric transformation* maupun *photometric transformation*, sehingga model *CNN* dapat belajar dari variasi yang lebih luas dan tidak terjebak pada pola terbatas *dataset* berukuran kecil. Dalam konteks penelitian ini, *data augmentation* membantu model menjadi lebih tahan terhadap variasi pencahayaan, sudut kamera, serta tekstur mukosa yang berbeda.

Namun, nilai simpangan baku yang besar menunjukkan bahwa *data augmentation* bekerja optimal bila disertai *normalization* yang stabil. Beberapa skenario (misalnya D dan E) membuktikan bahwa tanpa *normalization*, proses pelatihan dapat gagal *convergent* meskipun *data augmentation* diaktifkan. Dengan demikian, *data augmentation* efektif dalam memperluas representasi fitur, tetapi tetap memerlukan dukungan *preprocessing* yang konsisten agar hasilnya stabil.

2) Pengaruh Normalisasi (*ImageNet Normalization)*

Normalisasi berbasis statistik *ImageNet* terbukti menjadi variabel paling krusial untuk menjaga stabilitas pelatihan dan distribusi fitur. Kelompok dengan normalisasi aktif (A–C, G–I) memperoleh rata-rata *F1-score* sebesar 81,8%, dengan simpangan baku 9,1%. Sebaliknya, kelompok tanpa normalisasi (D–F, J–L) hanya mencapai rata-rata 37,4%, dengan simpangan baku 29,5%. Nilai rata-rata yang tinggi menunjukkan bahwa normalisasi secara langsung meningkatkan

performa model, sedangkan std yang rendah menandakan kestabilan hasil antar-skenario.

Penurunan performa sekitar 44 persen tanpa normalisasi menunjukkan bahwa model kehilangan kesesuaian distribusi input dengan distribusi saat *pretraining*, sehingga lapisan awal gagal mengekstraksi fitur dengan benar. Secara teoretis, He *et al.* (2016) menegaskan bahwa preprocessing berbasis statistik *ImageNet* (*mean* = [0.485, 0.456, 0.406], *std* = [0.229, 0.224, 0.225]) merupakan bagian penting dari proses *training* dan *inference* pada model *pretrained* modern. Ketidaksesuaian distribusi ini dapat menyebabkan nilai aktivasi ekstrem, gradien tidak stabil, dan kesulitan konvergensi.

Selain itu, simpangan baku yang jauh lebih besar pada kelompok tanpa normalisasi menunjukkan bahwa model menjadi sangat sensitif terhadap kombinasi variabel lain, terutama ukuran *batch* dan augmentasi. Skenario seperti D, E, J, dan K menunjukkan performa rendah bahkan mendekati kegagalan pelatihan. Dengan demikian, normalisasi berfungsi tidak hanya untuk menormalkan skala data, tetapi juga untuk menjaga konsistensi dan kestabilan proses optimasi di seluruh konfigurasi eksperimen.

3) Pengaruh Ukuran *Batch*

Ukuran *batch* memberikan pengaruh moderat namun signifikan terhadap stabilitas dan konvergensi model. Nilai rata-rata *F1-score* yang diperoleh untuk *batch* 16, 32, dan 64 berturut-turut adalah 58,2% ± 31,2%, 57,9% ± 30,4%, dan 74,2% ± 26,1%. Nilai *mean* yang meningkat seiring bertambahnya ukuran *batch* menunjukkan bahwa *batch* besar mampu menghasilkan estimasi gradien yang lebih

stabil, sedangkan simpangan baku yang masih relatif tinggi menunjukkan adanya variasi antarskenario akibat pengaruh variabel lain.

Peningkatan sekitar 16 persen dari *batch* kecil ke *batch* besar memperlihatkan bahwa ukuran *batch* yang besar dapat membantu memperhalus gradien dan mengurangi fluktuasi *loss* selama pelatihan. Efek positif ini paling jelas terlihat pada skenario C dan F, yang mencapai *F1-score* di atas 90%. Penemuan ini konsisten dengan penelitian Goyal *et al.* (2017), yang menunjukkan bahwa *large batch* mampu meningkatkan stabilitas pelatihan dan mempercepat konvergensi jika dikombinasikan dengan pengaturan *learning rate* yang sesuai.

Meskipun demikian, nilai simpangan baku yang masih tinggi menandakan bahwa ukuran *batch* besar tidak dapat sepenuhnya mengompensasi ketidakstabilan distribusi input atau absennya augmentasi dan normalisasi. Pada kelompok tanpa *preprocessing* yang memadai (misalnya skenario J–L), peningkatan *batch size* memang memperbaiki hasil, tetapi performanya tetap jauh di bawah kondisi optimal. Hal ini menegaskan bahwa ukuran *batch* besar hanya efektif jika diintegrasikan dengan augmentasi dan normalisasi yang aktif.

4.3.2 Analisis Kuantitatif dan Kualitatif

Berdasarkan hasil rekapitulasi pada bagian sebelumnya, skenario C (*augmentation* aktif, *normalization* aktif, *batch size* 64) ditetapkan sebagai model dengan performa terbaik dan stabilitas konvergensi tertinggi. Untuk memahami kinerja model secara lebih mendalam, dilakukan dua bentuk analisis: kuantitatif (berdasarkan distribusi *confusion matrix* per kelas) dan kualitatif (berdasarkan visualisasi citra prediksi benar dan salah). Analisis ini bertujuan untuk menilai

kemampuan diskriminatif *ConvNeXt-Tiny* dalam mengenali karakteristik visual antar kelas endoskopi, serta mengidentifikasi pola kesalahan yang masih muncul pada hasil klasifikasi.

Analisis kuantitatif dilakukan dengan menghitung nilai *precision*, *recall*, *F1-score*, dan *accuracy* untuk masing-masing kelas berdasarkan *confusion matrix* pada Tabel 4.15. Nilai *precision* dihitung sebagai rasio prediksi benar terhadap seluruh prediksi kelas tersebut, sedangkan *recall* menggambarkan proporsi prediksi benar terhadap total data aktual pada kelas bersangkutan. Nilai *F1-score* merupakan harmonisasi antara *precision* dan *recall*, sementara *accuracy* dihitung sebagai rasio total prediksi benar terhadap keseluruhan citra uji.

Tabel 4.15 *Confusion matrix* Model Terbaik (Skenario C)

True \ Predicted	***Gerd***	***Gerd* Normal**	**Polyp**	**Polyp Normal**
Gerd	618	43	2	9
***Gerd* Normal**	37	721	0	10
Polyp	12	11	526	22
Polyp Normal	6	19	27	742

Dari hasil perhitungan berbasis matriks di atas, diperoleh hasil sebagai berikut:

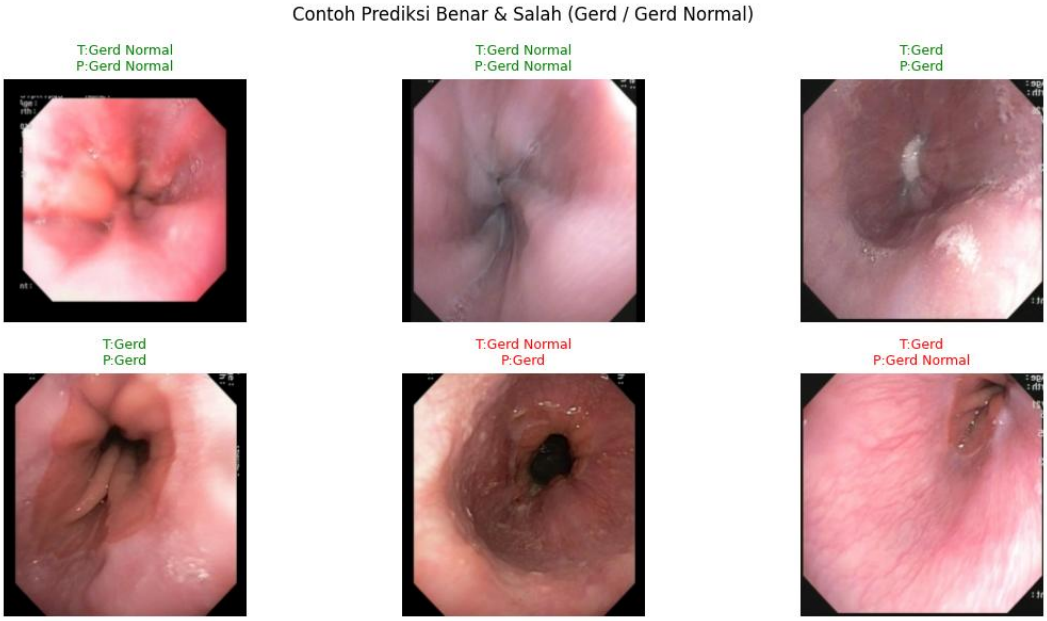
Tabel 4.16 Hasil perhitungan metrik kuantitatif per kelas model *ConvNeXt-Tiny* (Skenario C).

Kelas	***Accuracy***	***Precision***	***Recall***	***F1-score***
GERD	0.9294	0.899	0.92	0.909
***GERD* Normal**		0.912	0.931	0.921
Polyp		0.932	0.931	0.931
Polyp Normal		0.945	0.944	0.944

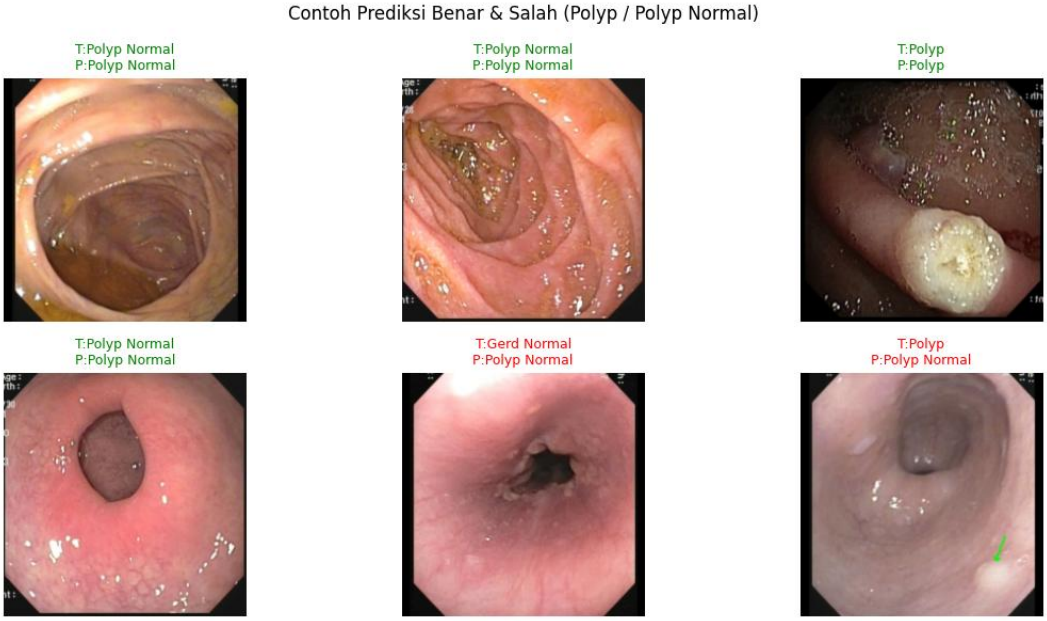
Nilai metrik per kelas pada Tabel 4.16 menunjukkan bahwa model memiliki performa yang relatif seimbang di seluruh kategori, dengan *F1-score* berada pada rentang 0.909–0.944 (atau 91%–94%). Kelas Polyp Normal menunjukkan nilai

tertinggi pada seluruh metrik, menandakan bahwa model paling mudah mengenali pola mukosa sehat dengan tekstur halus dan permukaan teratur. Sebaliknya, kelas *GERD* memiliki *precision* sedikit lebih rendah (0.899 atau 89.9%), yang menunjukkan masih terdapat kesalahan klasifikasi terhadap *GERD* Normal. Hal ini dapat dijelaskan oleh kemiripan visual pada kasus *mild reflux*, di mana dinding esofagus tampak hanya sedikit hiperemik tanpa ulserasi jelas. Meskipun demikian, tidak ditemukan indikasi bias ekstrem terhadap salah satu kelas, yang berarti model memiliki keseimbangan sensitivitas dan ketepatan yang baik (*balanced sensitivity and specificity*).

Analisis kualitatif dilakukan untuk mengamati secara visual hasil prediksi model terhadap citra endoskopi pada dua kelompok utama, yaitu *GERD*/*GERD* Normal dan Polyp/Polyp Normal, sebagaimana ditampilkan pada Gambar 4.16 dan Gambar 4.17. Masing-masing kelompok menampilkan beberapa contoh prediksi benar (*true positive*) dan salah (*false prediction*), yang membantu memahami pola keputusan model dan konteks kesalahan klasifikasinya.



Gambar 4.16 Contoh prediksi benar dan salah model *ConvNeXt-Tiny* pada citra *GERD* dan *GERD* Normal.



Gambar 4.17 Contoh prediksi benar dan salah model *ConvNeXt-Tiny* pada citra Polyp dan Polyp Normal.

Mengacu pada Gambar 4.16, dapat diamati bahwa model mampu membedakan citra *GERD* aktif dari *GERD* Normal dengan cukup baik. Pada

prediksi benar, model tampaknya mengenali ciri khas patologis seperti adanya area hiperemia, erosi mukosa, dan refleksi mukosa tidak teratur di daerah distal esofagus. Sementara pada kasus kesalahan, citra *GERD* dengan inflamasi ringan cenderung diklasifikasikan sebagai *GERD* Normal karena tampilannya menyerupai mukosa sehat dengan perbedaan warna yang sangat tipis. Fenomena ini menegaskan bahwa batas diagnostik visual antara *GERD* ringan dan normal sangat halus bahkan bagi manusia, sehingga wajar bila model masih menunjukkan ambiguitas pada kondisi borderline tersebut.

Sementara itu, pada Gambar 4.17, performa model dalam membedakan *Polyp* dan *Polyp Normal* juga tergolong baik. Citra polip umumnya dikenali berdasarkan pola elevasi mukosa, adanya lesi menonjol dengan tepi tidak rata, atau permukaan berwarna lebih terang akibat pembuluh darah superfisial. Prediksi salah umumnya terjadi pada citra kolon dengan lipatan mukosa besar atau refleksi cairan tinggi, yang menyebabkan struktur menyerupai tonjolan polip padahal bukan. Kesalahan minor ini menandakan bahwa model sensitif terhadap perbedaan topografi lokal, namun masih dapat terkecoh oleh artefak optik akibat pencahayaan dan sudut kamera.

4.3.3 Sintesis dan Implikasi

Hasil penelitian ini menunjukkan bahwa penerapan *ConvNeXt-Tiny* sebagai arsitektur dasar klasifikasi citra endoskopi memberikan performa tinggi dan stabilitas pelatihan yang kompetitif dibandingkan *CNN* konvensional. Nilai *F1-score* makro sebesar 0.9294 pada skenario C membuktikan bahwa desain *modernized convolutional block* dengan *large kernel*, *efficient residual connection*,

serta *adaptive normalization* mampu mengekstraksi fitur morfologis halus pada mukosa gastrointestinal. Kombinasi *augmentation* aktif, *normalization* aktif, dan *batch* besar memperkuat generalisasi model terhadap variasi citra uji yang kompleks. Secara ilmiah, hasil ini menegaskan pergeseran dari pendekatan *handcrafted features* menuju representasi otomatis berbasis *deep learning*. Berbeda dengan studi Jha *et al.* (2021) dan Cao *et al.* (2021) yang masih bergantung pada struktur *CNN* klasik, *ConvNeXt* dengan *large kernel design* dan *Layer Normalization* bergaya *Transformer* mampu menjembatani efisiensi *CNN* dengan kemampuan generalisasi *ViT*.

Penelitian ini juga menegaskan relevansi *ConvNeXt* terhadap arah perkembangan terkini di bidang *medical imaging*. Chan *et al.* (2023) melaporkan *F1-score* 69,87% pada *DenseNet* dengan *attention*. Dalam konteks ini, *ConvNeXt-Tiny* menjadi solusi kompromi ideal setara *ViT* dalam ketepatan klasifikasi, namun lebih efisien dan mudah diintegrasikan ke sistem *CAD*. Dibandingkan pendekatan hibrida seperti Li *et al.* (2023) dan Huan & Dun (2024), varian *Tiny* sudah cukup kuat mencapai akurasi klinis tanpa fusi arsitektur tambahan. Nilai *F1-score* di atas 0.92 pada citra dengan pencahayaan tidak seragam menunjukkan efisiensi representasional yang baik, sejalan dengan Nergiz (2023). Dengan demikian, *ConvNeXt* dapat diposisikan sebagai *benchmark* baru dalam klasifikasi citra medis, menggabungkan efisiensi, presisi, serta potensi penerapan pada perangkat terbatas seperti *embedded GPU* atau sistem *edge* di fasilitas medis kecil.

4.3.4 Integrasi Sains dan Islam

Perkembangan ilmu pengetahuan dan teknologi modern, termasuk penelitian ini yang mengimplementasikan *deep learning* untuk mendeteksi penyakit lambung dan usus, tidak dapat dipisahkan dari prinsip dasar Islam yang menempatkan ilmu sebagai sarana untuk mengenal dan mengabdi kepada Allah SWT. Dalam pandangan Islam, sains bukan sesuatu yang berdiri sendiri dan terpisah dari nilai moral dan spiritual. Sains merupakan bagian dari ibadah intelektual manusia untuk mewujudkan kemaslahatan. Oleh karena itu, integrasi sains dan Islam tidak hanya berarti menyandingkan ayat Al-Qur'an dengan teori ilmiah, tetapi juga menyatukan keduanya pada tingkat pemahaman. Hal ini mencakup keserasian antara ayat *qauliyyah* dan ayat *kauniyyah*. Landasan integrasi ini dapat ditemukan dalam firman Allah SWT pada Surah *Al-Mā'idah* [5]: 32:

مِنْ اَجْلِ ذٰلِكَۚ كَتَبْنَا عَلٰى بَنِيْٓ اِسْرَاۤءِيْلَ اَنَّهٗ مَنْ قَتَلَ نَفْسًاۢ بِغَيْرِ نَفْسٍ اَوْ فَسَادٍ فِى الْاَرْضِ فَكَاَنَّمَا قَتَلَ النَّاسَ جَمِيْعًاۗ وَمَنْ اَحْيَاهَا فَكَاَنَّمَآ اَحْيَا النَّاسَ جَمِيْعًاۗ وَلَقَدْ جَاۤءَتْهُمْ رُسُلُنَا بِالْبَيِّنٰتِ ثُمَّ اِنَّ كَثِيْرًا مِّنْهُمْ بَعْدَ ذٰلِكَ فِى الْاَرْضِ لَمُسْرِفُوْنَ

"Oleh karena itu, Kami menetapkan (suatu hukum) bagi Bani Israil bahwa siapa yang membunuh seseorang bukan karena (orang yang dibunuh itu) telah membunuh orang lain atau karena telah berbuat kerusakan di bumi, maka seakan-akan dia telah membunuh semua manusia. Sebaliknya, siapa yang memelihara kehidupan seorang manusia, dia seakan-akan telah memelihara kehidupan semua manusia. Sungguh, rasul-rasul Kami benar-benar telah datang kepada mereka dengan (membawa) keterangan-keterangan yang jelas. Kemudian, sesungguhnya banyak di antara mereka setelah itu melampaui batas di bumi." (QS. Al-Mā'idah [5]: 32)

Ayat ini menunjukkan bahwa setiap usaha ilmiah yang bertujuan menyelamatkan kehidupan manusia memiliki nilai kemanusiaan yang tinggi.

Dalam penelitian ini, pengembangan sistem klasifikasi citra endoskopi berbasis kecerdasan buatan bertujuan membantu mempercepat dan mempermudah diagnosis penyakit lambung dan polip usus. Upaya ini secara langsung berhubungan dengan penyelamatan jiwa karena deteksi dini penyakit dapat meningkatkan peluang penanganan yang tepat. Tafsir *Ibn Kathir* (2000) menjelaskan bahwa siapa saja yang menjadi sebab tegaknya kehidupan seseorang, baik melalui pemberian makanan, penyelamatan dari kebinasaan, maupun pengobatan, maka ia mendapatkan pahala seakan-akan menyelamatkan seluruh manusia. Penjelasan ini memperlihatkan bahwa kedokteran, farmasi, dan teknologi medis termasuk bidang dengan nilai *fardhu kifayah*, yaitu kewajiban kolektif umat untuk menjamin kelangsungan hidup manusia.

Dalam perspektif *maqāshid al-syarī'ah*, seluruh aktivitas ilmiah yang mendorong kemaslahatan manusia bertujuan menjaga lima aspek utama kehidupan, yaitu agama (*hifz al-dīn*), jiwa (*hifz al-nafs*), akal (*hifz al-'aql*), keturunan atau kehormatan (*hifz al-nasl* atau *al-'ird*), dan harta (*hifz al-māl*). Penelitian ini berkaitan secara langsung dengan tujuan *hifz al-nafs*, yaitu perlindungan jiwa manusia melalui upaya pencegahan dan pengobatan penyakit. Dengan menghadirkan teknologi deteksi dini berbasis kecerdasan buatan untuk penyakit pencernaan, penelitian ini termasuk bentuk ikhtiar ilmiah untuk menjaga kesehatan dan keselamatan manusia sesuai dengan nilai-nilai syariah. Islam juga menekankan kewajiban pencarian ilmu. Nabi Muhammad SAW bersabda:

طَلَبُ الْعِلْمِ فَرِيضَةٌ عَلَى كُلِّ مُسْلِمٍ

"Menuntut ilmu itu wajib bagi setiap Muslim." (HR. *Ibn Mājah*, no. 224, dinilai *hasan* oleh *Al-Albānī* dalam *Ṣaḥīḥ Ibn Mājah*)

Hadis ini menjadi dasar bahwa penelitian ilmiah, termasuk di bidang kedokteran dan teknologi, merupakan bagian dari kewajiban intelektual umat Islam. Menurut pandangan *Al-Ghazali* dalam kitab *Iḥyā' 'Ulūm al-Dīn*, ilmu yang membawa manfaat kepada masyarakat seperti kedokteran, pertanian, dan teknik termasuk kategori *fardhu kifayah* karena ilmu tersebut diperlukan untuk menjaga keberlangsungan hidup manusia. Dengan demikian, penelitian ini yang mengembangkan kecerdasan buatan untuk membantu diagnosis penyakit pencernaan merupakan penerapan nyata dari prinsip *hifz al-nafs* dalam *maqāshid al-syarī'ah*.

Teknologi yang dikembangkan tidak bertujuan menggantikan peran dokter. Teknologi ini berfungsi sebagai pendukung proses klinis dengan memberikan analisis objektif berbasis citra digital. Hal ini sejalan dengan sabda Nabi Muhammad SAW:

...مَا أَنْزَلَ اللَّهُ دَاءً إِلَّا أَنْزَلَ لَهُ دَوَاءً

"Tidaklah Allah menurunkan suatu penyakit kecuali Dia menurunkan pula obatnya; orang yang mengetahuinya mengetahuinya, dan orang yang tidak mengetahuinya tidak mengetahuinya." (HR. *Muslim*, no. 2204)'

Hadis ini menunjukkan bahwa setiap penyakit memiliki solusi yang telah Allah tetapkan dalam hukum alam-Nya (*sunnatullah*). Tugas manusia sebagai *khalīfah fī al-arḍ* adalah berusaha menyingkap pengetahuan tersebut melalui riset dan eksperimen ilmiah. Oleh karena itu, penelitian berbasis *machine learning* di bidang medis seperti ini merupakan bagian dari upaya menemukan *asbāb asy-syifā'*

yang diamanahkan kepada manusia untuk digali melalui akal dan ilmu pengetahuan.

Integrasi sains dan Islam juga mengandung nilai etis yang harus dipegang dalam setiap langkah penelitian. Dalam Islam, ilmu tidak hanya bernilai karena manfaat praktisnya, tetapi juga karena cara dan niat di balik pencapaiannya. Allah SWT berfirman:

وَلَا تَقْفُ مَا لَيْسَ لَكَ بِهٖ عِلْمٌۗ اِنَّ السَّمْعَ وَالْبَصَرَ وَالْفُؤَادَ كُلُّ اُولٰۤىِٕكَ كَانَ عَنْهُ مَسْـُٔوْلًا

"Dan janganlah kamu mengikuti sesuatu yang kamu tidak mempunyai pengetahuan tentangnya. Sesungguhnya pendengaran, penglihatan, dan hati, semuanya itu akan diminta pertanggungjawabannya." (QS. *Al-Isrā'* [17]: 36)

Ayat ini mengajarkan bahwa seorang peneliti tidak boleh menyampaikan kesimpulan tanpa bukti yang jelas. Dalam riset yang melibatkan data medis, hal ini berkaitan dengan kejujuran dalam pengumpulan data, validitas metode penelitian, penyajian hasil secara objektif, serta perlindungan privasi pasien. Nilai-nilai ini menjadi dasar bahwa penelitian harus dilakukan dengan amanah ilmiah dan kesadaran bahwa ilmu adalah titipan Allah yang harus digunakan untuk kebaikan, bukan untuk merugikan pihak lain.

Dengan pemahaman tersebut, inovasi teknologi medis tidak hanya dinilai dari efektivitas algoritmanya, tetapi juga dari kontribusinya terhadap nilai-nilai kemanusiaan. Integrasi sains dan Islam dalam penelitian ini memperlihatkan bahwa ilmuwan Muslim dituntut untuk menggabungkan kecerdasan intelektual dengan kesadaran spiritual, sehingga ilmu pengetahuan kembali pada tujuannya, yaitu memberikan manfaat bagi seluruh manusia.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian dan analisis yang dilakukan, dapat disimpulkan bahwa penerapan arsitektur *ConvNeXt-Tiny* mampu menghasilkan model klasifikasi citra endoskopi yang efektif dan konsisten dalam mengidentifikasi dua kondisi utama saluran cerna, yaitu *Gastroesophageal Reflux Disease (GERD)* dan polip usus. Kinerja terbaik dicapai pada skenario C, yaitu konfigurasi dengan *augmentation* aktif, *normalization* aktif, serta *batch size* 64. Pada pengaturan tersebut, model memperoleh akurasi sebesar 92.94%, *precision* 93.04%, *recall* 92.85%, dan *F1-score* 92.94%, yang menunjukkan keseimbangan sangat baik antara sensitivitas dan spesifisitas.

Secara keseluruhan, kombinasi strategi *augmentation* dan *normalization* memberikan kontribusi paling besar terhadap peningkatan kemampuan generalisasi model. *Augmentation* membantu memperluas variasi data pelatihan sehingga model lebih mampu mengenali perbedaan bentuk dan tekstur mukosa, sementara *normalization* menjaga stabilitas gradien selama proses pelatihan. Selain itu, penggunaan *batch* yang lebih besar berperan dalam memperhalus estimasi gradien dan mempercepat proses konvergensi. Evaluasi kuantitatif menunjukkan performa yang merata di keempat kelas dengan *F1-score* berkisar antara 91–94%, sedangkan analisis kualitatif mengonfirmasi bahwa model dapat mengenali karakteristik visual

patologis seperti area hiperemia, tonjolan mukosa, serta tekstur abnormal secara akurat.

5.2 Saran

Berdasarkan hasil dan keterbatasan penelitian, beberapa saran untuk penelitian dan pengembangan selanjutnya adalah sebagai berikut:

1. Perluasan *dataset* dan variasi kasus klinis

   Dataset yang digunakan dalam penelitian ini masih terbatas pada citra dari *GastroEndoNet* dengan dua kategori utama (*GERD* dan polip). Penelitian selanjutnya disarankan untuk menggunakan *dataset* multi-institusi yang lebih beragam, mencakup variasi kondisi seperti gastritis, ulserasi, atau kanker kolorektal lanjut agar model memiliki kemampuan generalisasi yang lebih luas.

2. Perbandingan antar arsitektur modern

   Untuk memperkuat temuan performa *ConvNeXt*, penelitian berikutnya dapat membandingkan arsitektur ini dengan model-model lain seperti *EfficientNetV2*, *Swin Transformer*, atau *Vision Transformer (ViT)*. Analisis perbandingan tersebut penting untuk menilai efisiensi komputasi, kompleksitas model, serta tingkat interpretabilitas yang paling sesuai untuk aplikasi klinis.

3. Integrasi interpretabilitas model

   Penggunaan *explainable AI (XAI)* seperti *Grad-CAM*, *Score-CAM*, atau *Layer-wise Relevance Propagation* direkomendasikan untuk meningkatkan transparansi prediksi model, sehingga hasil klasifikasi dapat lebih mudah

diverifikasi oleh tenaga medis dan meningkatkan kepercayaan terhadap sistem diagnosis berbantuan *AI*.

4. Optimasi *pipeline* pelatihan dan *inference*

   Penggunaan teknik seperti *mixed precision training*, *adaptive augmentation*, dan *learning rate scheduling* dapat meningkatkan efisiensi pelatihan. Selain itu, konversi model ke format ringan seperti *TensorRT* atau *TFLite* memungkinkan penerapan pada perangkat *edge computing* atau sistem endoskopi portabel.

5. Integrasi ke sistem *CAD* klinis

   Sebagai arah pengembangan aplikatif, model *ConvNeXt-Tiny* dapat diintegrasikan ke dalam sistem *Computer-Aided Diagnosis* real-time untuk membantu dokter dalam mendeteksi lesi atau polip selama prosedur endoskopi. Tahap ini memerlukan pengujian lebih lanjut terhadap aspek *latency*, reliabilitas, dan *user experience* di lingkungan klinis nyata.

Melalui pengembangan lanjutan tersebut, diharapkan hasil penelitian ini dapat berkontribusi pada peningkatan efektivitas pemeriksaan endoskopi, membantu deteksi dini penyakit gastrointestinal, serta mendukung transformasi digital dalam praktik medis modern yang berorientasi pada kemaslahatan dan peningkatan kualitas hidup manusia.

DAFTAR PUSTAKA

Ajra, Z., Xu, B., Dray, G., Montmain, J., & Perrey, S. (2022). Mental arithmetic task classification with convolutional neural network based on spectral-temporal features from EEG. *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. https://doi.org/10.1109/EMBC48229.2022.9870887

Bhattacharya, D., Reuter, K., Behrendt, F., Maack, L., Grube, S., & Schlaefer, A. (2024). PolypNextLSTM: A lightweight and fast polyp video segmentation network using *ConvNeXt* and ConvLSTM. *International Journal of Computer Assisted Radiology and Surgery, 19*, 2111–2119. https://doi.org/10.1007/s11548-024-03244-6

Bitto, A. K., Bijoy, M. H. I., Shakil, K. H., Das, A., Biplob, K. B. B., Mahmud, I., & Hossain, S. M. M. (2025). GastroEndoNet: Comprehensive endoscopy image *dataset* for *GERD* and polyp detection (Version 3). *Mendeley Data*. https://doi.org/10.17632/ffyn828yf4.3

Cao, C., Wang, R., Yu, Y., Zhang, H., Yu, Y., & Sun, C. (2021). Gastric polyp detection in gastroscopic images using deep neural network. *PLOS ONE, 16*(4), e0250632. https://doi.org/10.1371/journal.pone.0250632

Chan, I. N., Wong, T., Wong, P. K., Yan, T., Chan, I. W., Ren, H., & Chan, C. I. (2023). Multi-class gastroesophageal reflux disease classification system using deep learning techniques. *Proceedings of the 2023 10th International Conference on Biomedical and Bioinformatics Engineering (ICBBE)*, Kyoto, Japan. https://doi.org/10.1145/3637732.3637745

Emegano, D. I., Mustapha, M. T., Ozsahin, I., Ozsahin, D. U., & Uzun, B. (2025). Advancing prostate cancer diagnostics: A *ConvNeXt* approach to multi-class classification in underrepresented populations. *Bioengineering, 12*(4), 369. https://doi.org/10.3390/bioengineering12040369

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542*(7639), 115–118. https://doi.org/10.1038/nature21056

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. *Nature medicine*, *25*(1), 24–29. https://doi.org/10.1038/s41591-018-0316-z

Goyal, P., Dollár, P., Girshick, R., Noordhuis, P., Wesolowski, L., Kyrola, A., Tulloch, A., Jia, Y., & He, K. (2017). *Accurate, large minibatch SGD: Training ImageNet in 1 hour*. arXiv. https://arxiv.org/abs/1706.02677

He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 770–778). https://doi.org/10.1109/CVPR.2016.90

Huan, E., & Dun, H. (2024). MSMP-Net: A multi-scale neural network for end-to-end monkeypox virus skin lesion classification. *Applied Sciences, 14*(20), 9390. https://doi.org/10.3390/app14209390

Ibnu Katsir. (2008). Tafsir Ibnu Katsir (Terj. M. Abdul Ghoffar). Jakarta: Pustaka Imam Asy-Syafi'i.

Ibn Ḥajar al-'Asqalānī. (1959). *Fatḥ al-Bārī bi Sharḥ Ṣaḥīḥ al-Bukhārī* (Vol. 10). Beirut: Dār al-Ma'rifah.

Jha, D., Ali, S., Tomar, N. K., Johansen, H. D., Johansen, D., Rittscher, J., Riegler, M. A., & Halvorsen, P. (2021). Real-time polyp detection, localization and segmentation in colonoscopy using deep learning. *IEEE Access, 9*, 40496–40510. https://doi.org/10.1109/ACCESS.2021.3063716

Lafau, Y., Azmi, Z., & Calam, A. (2024). Sistem pakar mendiagnosis polip usus pada manusia menggunakan metode certainty factor. *Jurnal Sistem Informasi TGD, 3*(5), 602–609. https://ojs.trigunadharma.ac.id/index.php/jsi

Lin, M., Chen, Q., & Yan, S. (2014). Network in network. *arXiv*. https://doi.org/10.48550/arXiv.1312.4400

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A ConvNet for the 2020s. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA, 11966–11976. https://doi.org/10.1109/CVPR52688.2022.01167

Nadachowski, P., Łubniewski, Z., Malecha-Łysakowska, A., Trzcińska, K., Wróblewski, R., & Tęgowski, J. (2024). Classification of glacial and fluvioglacial landforms by convolutional neural networks using a digital elevation model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17*, 18549–18561. https://doi.org/10.1109/JSTARS.2024.3470253

Nergiz, M. (2023). Classification of Precancerous Colorectal Lesions via *ConvNeXt* on Histopathological Images. Balkan Journal of Electrical and Computer Engineering, 11(2), 129-137. https://doi.org/10.17694/bajece.1240284

Powers, D. M. W. (2011). *Evaluation: From precision, recall and F-measure to ROC, informedness and markedness*. Journal of Machine Learning Technologies, 2(1), 37–63. https://doi.org/10.48550/arXiv.2010.16061

Prommakhot, A., & Srinonchat, J. (2024). Combining convolutional neural networks for fungi classification. *IEEE Access, 12*, 58021–58032. https://doi.org/10.1109/ACCESS.2024.3391630

Rijal, S., Tayibu, K. M., Musa, I. M., Hapsari, P., & Natsir, P. (2024). Karakteristik penderita gastroesophageal reflux disease. *Fakumi Medical Journal, 4*(5), 402–411. https://fmj.fk.umi.ac.id/index.php/fmj

Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., ... & Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, *dataset* characteristics and transfer learning. *IEEE Transactions on Medical Imaging, 35*(5), 1285–1298. https://doi.org/10.1109/TMI.2016.2528162

Shorten, C., & Khoshgoftaar, T. M. (2019). *A survey on image data augmentation for deep learning*. Journal of Big Data, 6(1), 1–48. https://doi.org/10.1186/s40537-019-0197-0

Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. *Proceedings of the 36th International Conference on Machine Learning (ICML 2019)*, Long Beach, CA, USA, 6105–6114. http://proceedings.mlr.press/v97/tan19a.html

Zhang, J., Li, X., Li, L. *et al.* Lightweight U-Net for cloud detection of visible and thermal infrared remote sensing images. *Opt Quant Electron* **52**, 397 (2020). https://doi.org/10.1007/s11082-020-02500-8

Zhao, Y., Zhang, H., & Liu, Y. (2020). Protein secondary structure prediction based on generative confrontation and convolutional neural network. *IEEE Access, 8*, 199171–199179. https://doi.org/10.1109/ACCESS.2020.3035208

Zhou, S., Hu, X., Fu, Y., Zhou, Y., Li, Y., Ma, H., ... & Hu, H. (2023). Recent advances and trends in endoscopic image analysis for gastrointestinal disease diagnosis using deep learning. *Frontiers in Neuroscience, 17*, 1273686. https://doi.org/10.3389/fnins.2023.1273686