

BAB 2

LANDASAN TEORI

Bab ini menguraikan teori, konsep, dan penelitian terdahulu yang menjadi landasan dalam pelaksanaan penelitian. Landasan teori diperlukan untuk memberikan pemahaman mengenai permasalahan yang diteliti serta menjadi dasar dalam menentukan metode yang digunakan. Pembahasan pada bab ini mencakup kanker lambung, karakteristik citra histopatologi, histopatologi digital, dan dataset GasHisSDB. Selain itu, bab ini menjelaskan konsep deep learning, Convolutional Neural Network (CNN), transfer learning, fine-tuning, serta arsitektur EfficientNet-B0 dan EfficientNet-B1 sebagai model yang dibandingkan dalam penelitian. Bab ini juga membahas preprocessing, augmentasi citra, fungsi loss dan optimisasi, evaluasi klasifikasi, serta Explainable Artificial Intelligence (XAI) menggunakan Grad-CAM dan Grad-CAM++. Pada bagian akhir, dipaparkan penelitian terdahulu yang relevan, celah penelitian, dan posisi penelitian ini terhadap penelitian sebelumnya.

2.1 Kanker Lambung

Kanker lambung merupakan keganasan yang berkembang pada jaringan lambung dan memiliki karakteristik histologis serta molekuler yang beragam. Secara global, penyakit ini tetap memberikan beban kesehatan yang besar karena jumlah kasus dan kematiannya masih tinggi. Perubahan beban kanker dipengaruhi oleh pertumbuhan dan penuaan penduduk, variasi faktor risiko, serta ketimpangan dalam akses deteksi dan pengobatan [1, 2]. Kondisi tersebut menempatkan diagnosis yang akurat sebagai bagian penting dalam penentuan tata laksana dan prognosis pasien.

Sebagian besar kanker lambung berasal dari sel epitel mukosa dan berkembang melalui perubahan morfologi yang dapat diamati pada jaringan. Pemeriksaan jaringan membantu memastikan keberadaan keganasan, menilai tingkat diferensiasi, mengenali pola pembentukan kelenjar, serta mengamati invasi dan heterogenitas antardaerah. Pada citra histopatologi, jaringan abnormal dapat memperlihatkan susunan kelenjar yang tidak teratur, peningkatan kepadatan inti, variasi bentuk dan ukuran inti, perubahan stroma, serta hilangnya keteraturan arsitektur jaringan.

Perkembangan computational pathology menunjukkan bahwa informasi morfologi pada citra histopatologi dapat dipelajari secara komputasional. Penelitian berbasis whole-slide image telah menggunakan pola morfologi untuk subklasifikasi karsinoma, prediksi karakter molekuler, dan stratifikasi luaran klinis [15, 16, 17]. Meskipun demikian, hubungan antara pola citra dan label yang ditemukan model merupakan hubungan statistik. Temuan tersebut tidak dapat langsung dianggap sebagai mekanisme biologis atau dasar keputusan klinis tanpa validasi oleh ahli dan pengujian eksternal.

2.2 Histopatologi dan Pewarnaan H&E

Histopatologi adalah pemeriksaan mikroskopis terhadap jaringan untuk mengidentifikasi perubahan struktur sel dan jaringan yang berkaitan dengan penyakit. Sampel jaringan melalui proses fiksasi, pemrosesan, penanaman dalam parafin, pemotongan tipis, dan pewarnaan sebelum diamati oleh patolog. Dalam diagnosis kanker, pemeriksaan histopatologi digunakan untuk membedakan jaringan normal dan abnormal serta menilai pola pertumbuhan tumor [12].

Hematoksilin dan eosin (H&E) merupakan kombinasi pewarnaan yang paling umum digunakan. Hematoksilin menonjolkan komponen basofilik, terutama inti sel, dengan warna biru hingga keunguan. Eosin mewarnai komponen eosinofilik, seperti sitoplasma, kolagen, dan matriks ekstraseluler, dengan warna merah muda hingga kemerahan. Kombinasi kedua warna tersebut menghasilkan kontras yang membantu pengamatan bentuk inti, kepadatan sel, susunan kelenjar, dan hubungan antara epitel dengan stroma [3].

Walaupun protokol H&E telah digunakan secara luas, citra dapat memperlihatkan variasi warna akibat perbedaan laboratorium, konsentrasi pewarna, ketebalan irisan, pemindai, dan kondisi penyimpanan. Variasi ini perlu dipertimbangkan dalam pengembangan model karena CNN dapat memanfaatkan pola warna yang tidak selalu berkaitan dengan perubahan morfologi. Oleh karena itu, preprocessing, augmentasi warna yang terbatas, dan validasi pada data yang berbeda menjadi bagian penting dalam analisis citra histopatologi.

2.3 Histopatologi Digital

Histopatologi digital mengubah preparat kaca menjadi citra digital beresolusi tinggi melalui proses pemindaian. Citra hasil pemindaian dapat disimpan

sebagai whole-slide image (WSI) atau dibagi menjadi patch berukuran lebih kecil. Digitalisasi memungkinkan pengarsipan, pembagian data, pengukuran kuantitatif, serta penerapan metode kecerdasan artifisial untuk membantu penyaringan dan klasifikasi [3].

WSI mempertahankan konteks jaringan pada area yang luas, tetapi memiliki ukuran data yang sangat besar. Pengolahannya biasanya membutuhkan tiling, pemilihan area jaringan, multiple-instance learning, atau agregasi prediksi antarpatch. Sebaliknya, patch dapat diproses dengan kebutuhan memori lebih rendah dan sesuai untuk eksperimen klasifikasi terkontrol, tetapi konteks spasial yang lebih luas menjadi terbatas. Benchmarking computational pathology memperlihatkan bahwa performa pipeline dipengaruhi oleh level anotasi, strategi sampling, pembentukan patch, dan metode agregasi [4, 3].

Penelitian ini berada pada level patch. Dengan demikian, keluaran model menunjukkan kemampuan membedakan citra patch normal dan abnormal, bukan keputusan diagnosis pada tingkat slide, pasien, staging, atau prognosis. Pembatasan tingkat analisis ini perlu dinyatakan agar hasil tidak ditafsirkan melebihi karakteristik data yang digunakan.

2.4 Dataset GasHisSDB

GasHisSDB atau Gastric Histopathology Sub-size Image Database dikembangkan sebagai basis data publik untuk mengevaluasi metode computer-aided diagnosis pada citra histopatologi lambung. Dataset lengkap berisi 245.196 patch H&E yang berasal dari 600 citra patologi berukuran 2048 x 2048 piksel. Patch tersedia dalam ukuran 80 x 80, 120 x 120, dan 160 x 160 piksel serta dikelompokkan menjadi kelas Normal dan Abnormal [10].

Kelas Abnormal berisi patch dengan area kanker yang dominan, sedangkan kelas Normal berasal dari jaringan normal. Penyediaan beberapa ukuran patch memungkinkan pengujian pengaruh skala terhadap kemampuan klasifikasi. Namun, penggunaan patch juga berarti bahwa model tidak mengakses keseluruhan konteks jaringan pada slide. Oleh sebab itu, evaluasi harus dipahami sebagai klasifikasi patch dan tidak dapat langsung disetarakan dengan diagnosis klinis.

Eksperimen perbandingan pada GasHisSDB menghasilkan rentang performa yang berbeda antarmetode. Perbandingan dari algoritma regresi hingga visual Transformer menunjukkan bahwa skala data dan karakteristik model memengaruhi hasil klasifikasi [18]. Pendekatan deep ensemble dapat meningkatkan performa,

tetapi menambah kompleksitas pelatihan dan inferensi [19]. Penelitian ini menggunakan subset terkontrol sebanyak 5.000 citra dari sub-database 160 x 160 piksel agar perbandingan EfficientNet-B0 dan EfficientNet-B1 dapat dilakukan dengan konfigurasi data yang sama.

2.5 Deep Learning

Deep learning merupakan bagian dari machine learning yang mempelajari representasi data melalui susunan lapisan parametrik. Pada pendekatan konvensional, fitur sering dirancang secara manual sebelum digunakan oleh classifier. Deep learning mengintegrasikan pembelajaran fitur dan proses klasifikasi dalam satu model sehingga representasi yang relevan dapat dipelajari langsung dari data [3].

Pada citra, lapisan awal umumnya mempelajari pola sederhana seperti tepi, arah, warna, dan tekstur. Lapisan yang lebih dalam menggabungkan pola tersebut menjadi representasi yang lebih kompleks. Dalam histopatologi, representasi tersebut dapat berkaitan dengan kepadatan inti, susunan kelenjar, tekstur stroma, dan perubahan arsitektur jaringan. Kemampuan ini menjadikan deep learning sesuai untuk citra yang memiliki pola visual kompleks dan sulit dirumuskan seluruhnya melalui fitur buatan tangan [4, 3].

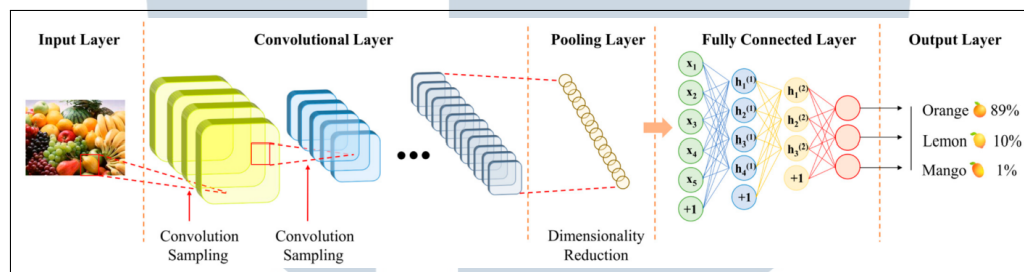
Kinerja deep learning sangat bergantung pada jumlah dan keragaman data, kualitas label, strategi pembagian data, serta kesesuaian antara data pelatihan dan data pengujian. Model dengan kapasitas tinggi dapat mempelajari pola yang berguna, tetapi juga dapat menghafal artefak atau karakteristik spesifik dataset. Karena itu, evaluasi perlu dilakukan pada data yang tidak digunakan selama pelatihan dan dilengkapi pengendalian overfitting [4, 9].

2.6 Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) adalah jenis jaringan saraf yang dirancang untuk mengolah data berbentuk grid, termasuk citra. CNN menggunakan prinsip local connectivity dan weight sharing, yaitu kernel yang sama digunakan pada berbagai posisi citra untuk mendeteksi pola lokal. Dengan mekanisme tersebut, CNN dapat mempelajari fitur visual secara lebih efisien dibandingkan jaringan fully connected pada citra berukuran besar.

Secara sederhana, proses konvolusi pada CNN dilakukan dengan menggeser

kernel pada citra atau feature map untuk menghasilkan representasi fitur baru. Kernel berfungsi sebagai filter yang menangkap pola tertentu, seperti tepi, tekstur, warna, atau bentuk lokal. Hasil dari proses tersebut membentuk feature map yang merepresentasikan informasi visual dari citra masukan. Fungsi aktivasi digunakan untuk memperkenalkan non-linearitas sehingga model dapat mempelajari hubungan fitur yang lebih kompleks. Selain itu, pooling atau convolution dengan stride digunakan untuk mengurangi resolusi spasial dan memperluas receptive field. Pada bagian akhir jaringan, global average pooling atau flattening mengubah feature map menjadi representasi fitur yang kemudian dipetakan ke logit kelas oleh fully connected layer [20]. Skema umum CNN ditunjukkan pada Gambar 2.1.



Gambar 2.1. Skema umum Convolutional Neural Network

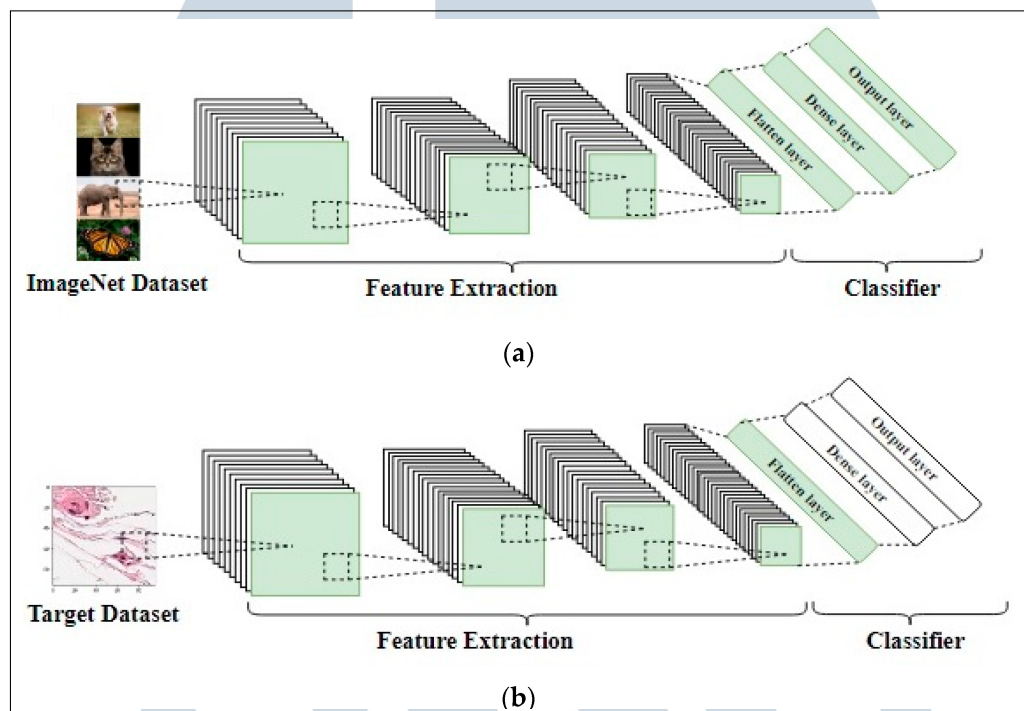
Sumber: [21]

Keunggulan CNN terletak pada kemampuannya mempelajari hierarki fitur secara otomatis. Meskipun demikian, kemampuan generalisasi tetap dipengaruhi ukuran dan variasi dataset. Studi mengenai ukuran data histopatologi menunjukkan bahwa penambahan data pelatihan dapat meningkatkan akurasi, tetapi keuntungan marginal dapat menurun dan generalisasi tetap bergantung pada keragaman sumber data [22].

2.7 Transfer Learning

Transfer learning memanfaatkan pengetahuan yang telah dipelajari pada domain sumber untuk menyelesaikan tugas pada domain target. Dalam klasifikasi citra, model yang telah dilatih pada dataset besar seperti ImageNet digunakan sebagai inisialisasi. Lapisan konvolusi model membawa fitur umum seperti tepi, warna, bentuk, dan tekstur yang dapat digunakan kembali pada tugas baru [23, 24]. Dalam penelitian ini, EfficientNet menggunakan bobot pralatih ImageNet. Seluruh feature extractor dibekukan, sedangkan classifier terakhir diganti untuk menghasilkan dua kelas target. Strategi ini mengurangi jumlah parameter trainable

dan risiko overfitting pada data terbatas. Karena domain citra alami berbeda dari citra histopatologi H&E, transfer learning selanjutnya dilanjutkan dengan fine-tuning agar fitur tingkat tinggi dapat beradaptasi [7, 8, 24]. Panel (a) pada Gambar 2.2 menunjukkan model sumber yang dilatih menggunakan ImageNet, sedangkan panel (b) menunjukkan feature extractor yang dibekukan dan classifier baru yang dilatih pada dataset target.



Gambar 2.2. Mekanisme transfer learning menggunakan model pralatih ImageNet

Sumber: [24]

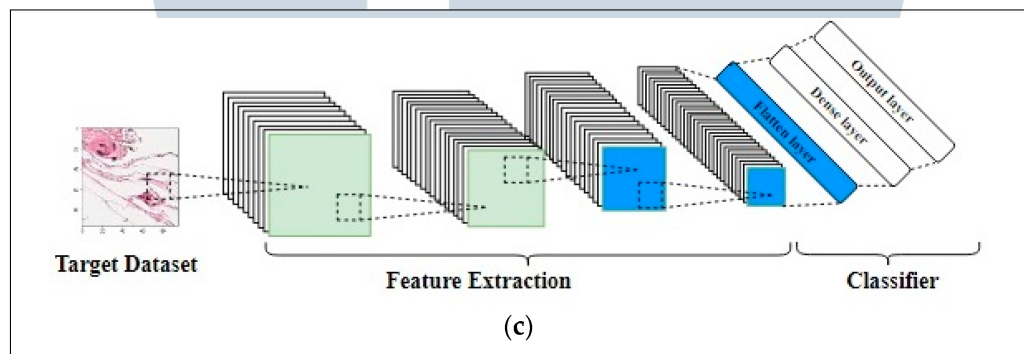
Pemanfaatan transfer learning untuk histopatologi telah ditunjukkan melalui penggunaan EfficientNet pada citra H&E dan transfer lintas kanker [7, 8]. Walaupun membantu ketika data berlabel terbatas, performa tetap bergantung pada kesamaan domain, strategi pembekuan lapisan, dan kualitas data target.

2.8 Fine-Tuning

Fine-tuning merupakan tahap lanjutan transfer learning dengan membuka sebagian lapisan model pralatih agar parameternya dapat disesuaikan dengan domain target. Lapisan awal umumnya mempertahankan fitur visual yang bersifat umum, sedangkan lapisan akhir memuat representasi yang lebih spesifik terhadap tugas sumber. Oleh karena itu, penelitian ini mempertahankan lapisan awal dalam

kondisi beku dan membuka dua bagian terakhir feature extractor bersama classifier agar fitur tingkat tinggi dapat beradaptasi terhadap citra histopatologi [25, 24].

Fine-tuning menggunakan learning rate yang lebih kecil daripada tahap transfer learning agar bobot prelatih tidak berubah secara terlalu agresif. Learning rate yang terlalu besar dapat menyebabkan catastrophic forgetting, sedangkan nilai yang terlalu kecil dapat memperlambat adaptasi. Jumlah lapisan yang dibuka juga perlu dikendalikan karena semakin banyak parameter trainable, semakin besar kebutuhan komputasi dan risiko overfitting [25]. Pada panel (c) Gambar 2.3, lapisan berwarna hijau menunjukkan bagian yang tetap dibekukan, lapisan biru menunjukkan bagian yang dilatih kembali, dan lapisan putih menunjukkan classifier untuk dataset target.



Gambar 2.3. Mekanisme fine-tuning sebagian lapisan feature extractor

Sumber: [24]

Dalam penelitian ini, fine-tuning ditempatkan sebagai tahap adaptasi setelah transfer learning. Pembukaan sebagian lapisan akhir digunakan untuk menyesuaikan representasi model terhadap citra histopatologi, sedangkan rincian jumlah blok yang dibuka, learning rate, dan durasi pelatihan dijelaskan pada Bab 3. Dengan pemisahan tahap tersebut, kontribusi transfer learning dan fine-tuning dapat dievaluasi secara terpisah.

2.9 Arsitektur EfficientNet

EfficientNet dikembangkan untuk meningkatkan akurasi sekaligus menjaga efisiensi komputasi. Arsitektur ini menggunakan compound scaling, yaitu peningkatan kedalaman jaringan, lebar kanal, dan resolusi masukan secara terkoordinasi. Berbeda dari pendekatan yang hanya memperbesar satu dimensi, compound scaling berupaya menjaga keseimbangan kapasitas pada seluruh dimensi

model [26].

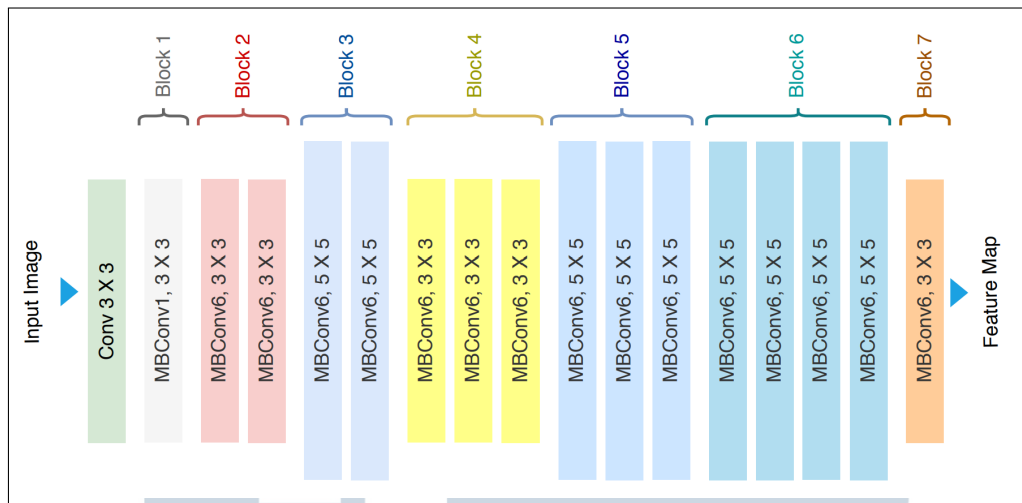
Secara konseptual, compound scaling menaikkan kedalaman, lebar, dan resolusi model menggunakan koefisien skala yang seimbang. Prinsip ini membuat peningkatan kapasitas model dilakukan secara lebih terkontrol sehingga biaya komputasi tidak meningkat hanya pada satu sisi arsitektur. Blok utama EfficientNet adalah Mobile Inverted Bottleneck Convolution (MBConv). Blok ini mencakup ekspansi kanal, depthwise convolution, squeeze-and-excitation, dan proyeksi kanal. Depthwise convolution memproses setiap kanal secara terpisah sehingga biaya komputasi lebih rendah daripada konvolusi penuh. Squeeze-and-excitation mempelajari bobot antar-kanal agar fitur yang relevan dapat ditekankan secara adaptif [27, 28].

EfficientNet juga menggunakan skip connection pada blok yang memiliki bentuk masukan dan keluaran sesuai. Kombinasi MBConv, squeeze-and-excitation, dan compound scaling menghasilkan keluarga model B0 sampai B7. Varian yang lebih tinggi umumnya memiliki kedalaman, lebar, resolusi nominal, serta jumlah parameter yang lebih besar.

2.9.1 EfficientNet-B0 dan EfficientNet-B1

EfficientNet-B0 merupakan baseline dari keluarga EfficientNet. EfficientNet-B1 dibentuk melalui peningkatan skala dari B0 sehingga memiliki pengulangan blok dan kapasitas representasi yang lebih besar. B0 lebih ringan dan berpotensi memiliki waktu inferensi lebih singkat, sedangkan B1 menyediakan kapasitas tambahan untuk mempelajari pola yang lebih kompleks [27, 29].

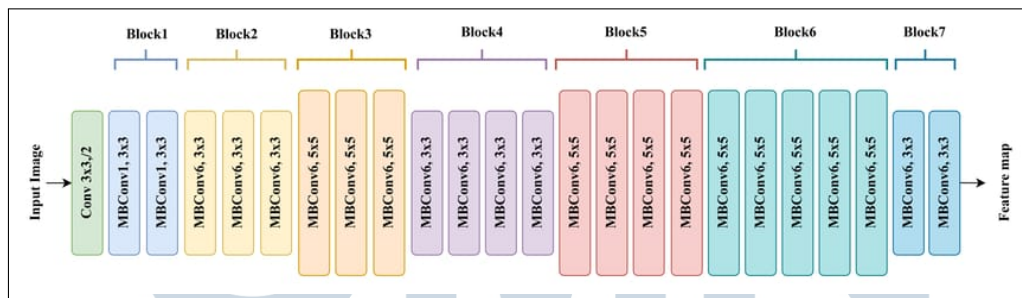
U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A



Gambar 2.4. Arsitektur EfficientNet-B0

Sumber: [30]

Gambar 2.4 memperlihatkan susunan EfficientNet-B0 yang dimulai dari convolutional stem, dilanjutkan rangkaian MBConv dengan kernel dan jumlah kanal yang berbeda hingga menghasilkan feature map. Meskipun gambar sumber menerapkan B0 pada data histopatologi kolon, struktur backbone yang ditampilkan merepresentasikan susunan umum EfficientNet-B0.



Gambar 2.5. Arsitektur EfficientNet-B1

Sumber: [31]

Berdasarkan Gambar 2.5 EfficientNet-B1 mempertahankan prinsip blok yang sama dengan EfficientNet-B0, tetapi beberapa tahap memiliki pengulangan lebih banyak. Dalam konfigurasi standar ImageNet, resolusi nominal EfficientNet-B0 adalah 224 x 224 piksel dan B1 adalah 240 x 240 piksel. Dalam penelitian ini, kedua model menerima masukan 224 x 224 piksel agar kondisi perbandingan tetap identik. Penyeragaman ini membantu mengisolasi pengaruh kapasitas arsitektur, meskipun tidak menggunakan resolusi nominal EfficientNet-B1 [27, 29].

Tabel 2.1. Perbandingan Karakteristik EfficientNet-B0 dan EfficientNet-B1

Karakteristik	EfficientNet-B0	EfficientNet-B1
Posisi dalam keluarga	Baseline	Varian lebih besar dari B0
Resolusi nominal ImageNet	224×224	240×240
Resolusi pada penelitian	224×224	224×224
Parameter total setelah classifier dua kelas	Sekitar 4,01 juta	Sekitar 6,52 juta
Dimensi fitur classifier	1.280	1.280
Kecenderungan komputasi	Lebih ringan	Lebih besar dan lebih berat
Tujuan perbandingan	Baseline efisien	Menilai manfaat kapasitas tambahan

Jumlah parameter pada Tabel 2.1 mengacu pada model penelitian setelah classifier 1.000 kelas diganti menjadi dua kelas. Kapasitas B1 berpotensi menangkap representasi yang lebih kaya, tetapi peningkatan performa belum tentu sebanding dengan biaya komputasinya. Penelitian terdahulu menunjukkan bahwa EfficientNet efektif pada klasifikasi H&E dan dapat dikombinasikan dengan interpretasi berbasis class activation mapping [28, 20]. Oleh karena itu, pemilihan antara B0 dan B1 perlu mempertimbangkan performa serta efisiensi.

2.10 Preprocessing Citra

Preprocessing citra merupakan proses penyiapan data sebelum diberikan kepada model. Tujuannya adalah menyelaraskan dimensi, format, jumlah kanal, dan rentang nilai citra dengan kebutuhan arsitektur. Proses yang diterapkan secara konsisten membantu mencegah perbedaan masukan yang tidak berkaitan dengan kelas [4, 3].

Citra GasHisSDB berukuran 160×160 piksel diubah menjadi 224×224 piksel agar sesuai dengan masukan yang digunakan pada eksperimen. Citra kemudian dikonversi menjadi tensor. Kanal RGB dinormalisasi menggunakan mean [0,485; 0,456; 0,406] dan standard deviation [0,229; 0,224; 0,225] dari ImageNet. Normalisasi menjaga skala masukan mendekati distribusi yang digunakan ketika bobot pralatih dibentuk [23].

Resize dan normalisasi diterapkan pada data training, validation, dan test.

Namun, transformasi acak hanya diterapkan pada training. Pemisahan tersebut penting agar data validation dan test merepresentasikan evaluasi yang konsisten. Preprocessing tidak dimaksudkan untuk menghasilkan sampel baru, melainkan menyesuaikan representasi data dengan kebutuhan model [4, 9].

2.11 Augmentasi Citra

Augmentasi citra menghasilkan variasi sampel pelatihan melalui transformasi yang tetap mempertahankan label. Tujuannya adalah mengurangi ketergantungan model pada orientasi atau kondisi visual tertentu dan meningkatkan kemampuan generalisasi. Augmentasi dilakukan secara acak saat data training dimuat, sehingga citra yang sama dapat muncul dengan transformasi berbeda pada epoch yang berbeda [4].

Augmentasi yang digunakan meliputi horizontal flip dan vertical flip dengan probabilitas 0,5, rotasi acak dari -15 sampai 15 derajat, serta perubahan brightness dan contrast sebesar 0,1. Transformasi geometri sesuai untuk patch histopatologi karena orientasi jaringan pada preparat tidak memiliki arah absolut. Color jitter diterapkan secara terbatas agar model lebih tahan terhadap variasi pewarnaan tanpa mengubah karakter H&E secara ekstrem.

Augmentasi yang terlalu agresif dapat menghasilkan citra yang tidak realistis dan menghilangkan ciri morfologi penting. Oleh karena itu, parameter augmentasi perlu disesuaikan dengan karakteristik domain. Dalam penelitian ini, konfigurasi yang sama diterapkan pada B0 dan B1 agar perbedaan hasil tidak disebabkan oleh perbedaan variasi data pelatihan [4].

2.12 Fungsi Loss dan Optimisasi

Pelatihan model dilakukan dengan meminimalkan fungsi loss melalui proses optimisasi. Fungsi loss mengukur ketidaksesuaian antara prediksi dan label sebenarnya, sedangkan optimizer menentukan cara parameter diperbarui. Learning rate, scheduler, dan early stopping mengendalikan besar serta durasi pembaruan selama pelatihan [32].

2.12.1 Cross-Entropy Loss

Cross-entropy loss digunakan untuk klasifikasi dua kelas dengan keluaran berupa logit. Setelah logit diubah menjadi probabilitas melalui softmax, loss

memberikan penalti yang lebih besar ketika probabilitas untuk kelas sebenarnya rendah. Untuk satu sampel dengan label one-hot y_k dan probabilitas p_k , cross-entropy dituliskan sebagai berikut.

Rumus 2.1 menunjukkan perhitungan *Cross-Entropy Loss*.

$$L = - \sum_{i=1}^C y_i \log(\hat{y}_i) \quad (2.1)$$

Cross-entropy tidak hanya mempertimbangkan benar atau salah, tetapi juga tingkat keyakinan model. Prediksi yang benar dengan probabilitas rendah masih menghasilkan loss lebih tinggi daripada prediksi benar dengan probabilitas

2.12.2 Optimizer Adam

Adam merupakan optimizer adaptif yang menggabungkan informasi momentum gradien dan skala kuadrat gradien untuk memperbarui parameter model [33]. Mekanisme adaptif tersebut membuat Adam relatif stabil pada berbagai skala gradien dan banyak digunakan dalam pelatihan maupun fine-tuning model deep learning. Meskipun demikian, penggunaan Adam tidak menghilangkan kebutuhan untuk menentukan learning rate yang sesuai. Learning rate tetap berperan penting karena menentukan besar perubahan parameter selama proses optimisasi.

2.12.3 Learning Rate

Learning rate menentukan besar perubahan parameter pada setiap langkah optimisasi. Nilai yang terlalu besar dapat menyebabkan pembaruan melewati daerah minimum dan membuat pelatihan tidak stabil. Nilai yang terlalu kecil dapat memperlambat konvergensi dan membuat model berhenti sebelum memperoleh representasi yang optimal.

Pada transfer learning, learning rate biasanya dapat dibuat lebih besar karena parameter yang diperbarui hanya terdapat pada classifier. Pada fine-tuning, learning rate umumnya dibuat lebih kecil agar bobot prelatih tidak berubah secara terlalu agresif. Nilai learning rate yang digunakan dalam eksperimen dijelaskan pada Bab 3.

2.12.4 ReduceLROnPlateau

ReduceLROnPlateau adalah scheduler yang menurunkan learning rate ketika metrik pemantauan tidak membaik dalam sejumlah epoch. Scheduler ini umumnya memantau validation loss atau metrik validasi lain untuk menentukan kapan laju pembelajaran perlu diturunkan.

Scheduler membantu model melakukan pembaruan yang lebih halus ketika pelatihan mendekati plateau. Penurunan learning rate tidak menjamin peningkatan performa, tetapi dapat mengurangi osilasi dan memberi kesempatan bagi optimizer untuk mencapai daerah loss yang lebih rendah. Konfigurasi faktor penurunan, patience, dan batas minimum learning rate dijelaskan pada Bab 3.

2.12.5 Early Stopping

Early stopping menghentikan pelatihan apabila metrik validasi tidak menunjukkan perbaikan setelah sejumlah epoch. Mekanisme ini mengurangi komputasi yang tidak diperlukan dan membantu membatasi overfitting.

Bobot model dapat disimpan ketika validation loss mencapai nilai terbaik, kemudian dimuat kembali untuk evaluasi test. Dengan demikian, evaluasi tidak harus menggunakan bobot pada epoch terakhir apabila performanya telah menurun. Early stopping merupakan mekanisme pengendalian pelatihan, bukan bagian dari optimizer.

2.13 Evaluasi Klasifikasi

Evaluasi klasifikasi dilakukan pada data test yang tidak digunakan untuk memperbarui parameter. Penggunaan beberapa metrik diperlukan karena satu nilai saja tidak selalu menggambarkan pola kesalahan model. Confusion matrix menunjukkan jumlah prediksi pada setiap kombinasi kelas, sedangkan accuracy, precision, recall, dan F1-score merangkum kinerja pada threshold tertentu. ROC curve dan AUC menilai kemampuan diskriminasi pada berbagai threshold [32].

2.13.1 Confusion Matrix

Confusion matrix menyajikan hubungan antara label sebenarnya dan kelas hasil prediksi. Untuk kelas yang ditetapkan sebagai positif, true positive (TP) adalah sampel positif yang diprediksi positif, false positive (FP) adalah sampel negatif yang

diprediksi positif, false negative (FN) adalah sampel positif yang diprediksi negatif, dan true negative (TN) adalah sampel negatif yang diprediksi negatif. Matriks ini membantu mengidentifikasi jenis kesalahan yang mungkin tidak terlihat dari accuracy.

2.13.2 Accuracy

Accuracy mengukur proporsi seluruh prediksi yang benar terhadap jumlah sampel. Rumusnya adalah sebagai berikut.

Rumus 2.2 menunjukkan cara perhitungan *Accuracy*.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.2)$$

Accuracy mudah dipahami dan sesuai digunakan ketika distribusi kelas seimbang. Namun, metrik ini dapat menutupi ketimpangan performa antarkelas, sehingga tetap perlu dilengkapi dengan precision, recall, dan F1-score.

2.13.3 Precision

Precision mengukur proporsi prediksi positif yang benar. Nilai tinggi menunjukkan bahwa model menghasilkan sedikit false positive.

Rumus 2.3 menunjukkan cara perhitungan *Precision*.

$$Precision = \frac{TP}{TP + FP} \quad (2.3)$$

Dalam klasifikasi multikelas atau biner yang dinilai per kelas, precision dapat dihitung untuk setiap kelas. Macro average memberikan bobot yang sama pada setiap kelas, sedangkan weighted average memberi bobot sesuai jumlah sampel kelas.

2.13.4 Recall

Recall atau sensitivity mengukur proporsi sampel positif yang berhasil dikenali. Nilai rendah menunjukkan banyak sampel positif yang terlewat sebagai false negative.

Rumus 2.4 menunjukkan cara perhitungan *Recall*.

$$Recall = \frac{TP}{TP + FN} \quad (2.4)$$

Pada penggunaan sebagai alat bantu penyaringan, recall kelas abnormal penting karena false negative dapat berarti citra abnormal diprediksi normal. Walaupun demikian, penelitian pada tingkat patch tidak dapat langsung digunakan untuk menyimpulkan risiko klinis pada tingkat pasien.

2.13.5 F1-Score

F1-score merupakan rata-rata harmonik precision dan recall. Metrik ini tinggi apabila keduanya memiliki nilai yang seimbang.

Rumus 2.5 menunjukkan cara perhitungan *F1-Score*.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2.5)$$

Macro F1-score menghitung F1 setiap kelas lalu merata-ratakannya dengan bobot sama. Weighted F1-score menggunakan jumlah sampel tiap kelas sebagai bobot. Pada data yang seimbang, nilai macro dan weighted cenderung berdekatan, tetapi keduanya tetap dapat dilaporkan untuk transparansi.

2.13.6 ROC Curve

Receiver Operating Characteristic (ROC) curve menggambarkan hubungan antara true positive rate (TPR) dan false positive rate (FPR) pada berbagai threshold. TPR setara dengan recall, sedangkan FPR dirumuskan sebagai $FP/(FP+TN)$. Kurva yang semakin mendekati sudut kiri atas menunjukkan kemampuan pemisahan kelas yang lebih baik.

Dalam evaluasi ROC, kelas positif harus ditetapkan secara konsisten agar interpretasi kurva tidak keliru. Pada konteks klasifikasi citra kanker, kelas abnormal umumnya dapat diprioritaskan sebagai kelas positif, tetapi penggunaan kelas positif lain tetap valid apabila dinyatakan secara jelas dan digunakan secara konsisten dalam perhitungan.

2.13.7 AUC

Area Under the Curve (AUC) adalah luas di bawah ROC curve dan merangkum kemampuan model mengurutkan sampel positif lebih tinggi daripada sampel negatif. Nilai 0,5 mendekati pemisahan acak, sedangkan nilai yang mendekati 1 menunjukkan diskriminasi yang semakin baik.

AUC tidak terikat pada satu threshold keputusan. Oleh karena itu, dua model dapat memiliki AUC yang hampir sama tetapi accuracy atau F1-score berbeda pada threshold yang digunakan. AUC juga tidak menunjukkan lokasi perhatian model dan tidak memiliki hubungan langsung dengan Grad-CAM atau Grad-CAM++.

2.14 Explainable Artificial Intelligence (XAI)

Explainable Artificial Intelligence (XAI) mencakup metode yang menyediakan informasi mengenai faktor yang memengaruhi prediksi model. Dalam patologi digital, XAI dapat digunakan untuk mengaudit kesalahan, mengidentifikasi kemungkinan shortcut learning, dan menilai apakah model memperhatikan jaringan atau artefak. XAI tidak mengubah model menjadi sepenuhnya transparan, tetapi memberikan bukti tambahan untuk inspeksi.

Explainability perlu dibedakan dari causability. Penjelasan teknis dapat menunjukkan sensitivitas model terhadap suatu area, tetapi belum membuktikan bahwa area tersebut merupakan penyebab biologis atau alasan klinis yang benar [14]. Penelitian berorientasi pengguna juga menunjukkan explainability paradox: visualisasi yang sederhana dapat mudah dipahami, tetapi berisiko meningkatkan kepercayaan secara berlebihan apabila keterbatasan dan ketidakpastian tidak dijelaskan [13].

Karena itu, visualisasi XAI pada penelitian ini diposisikan sebagai inspeksi kualitatif. Heatmap dibandingkan berdasarkan lokasi hotspot, luas area aktif, dan konsistensi antarmodel. Interpretasi tidak menyatakan bahwa area panas merupakan lesi yang telah divalidasi oleh patolog. Integrasi AI yang bertanggung jawab tetap membutuhkan kolaborasi manusia dan mesin [34].

2.14.1 Grad-CAM

Gradient-weighted Class Activation Mapping atau Grad-CAM merupakan salah satu metode Explainable Artificial Intelligence (XAI) yang digunakan untuk menampilkan area penting pada citra yang berkontribusi terhadap keputusan model.

Metode ini bekerja dengan memanfaatkan informasi gradien dari skor kelas terhadap feature map pada lapisan konvolusi terakhir atau lapisan target tertentu. Dengan cara tersebut, Grad-CAM dapat menghasilkan heatmap yang menunjukkan bagian citra yang paling berpengaruh terhadap prediksi model [35, 36].

Pada citra histopatologi, Grad-CAM dapat membantu melihat apakah model memperhatikan area jaringan yang relevan ketika melakukan klasifikasi. Heatmap yang dihasilkan biasanya ditumpangkan pada citra asli, sehingga area dengan warna lebih panas menunjukkan kontribusi relatif yang lebih kuat terhadap kelas target. Namun, warna panas pada heatmap tidak dapat langsung diartikan sebagai lokasi lesi secara klinis. Heatmap hanya menunjukkan area yang dianggap penting oleh model berdasarkan proses komputasi, bukan bukti pasti bahwa area tersebut merupakan jaringan abnormal.

Grad-CAM memiliki kelebihan karena dapat diterapkan pada model CNN tanpa mengubah struktur utama model. Metode ini juga mudah digunakan untuk membandingkan pola perhatian visual antara model yang berbeda. Meskipun demikian, Grad-CAM memiliki keterbatasan karena hasil visualisasinya cenderung kasar. Hal ini disebabkan oleh penggunaan feature map dari lapisan konvolusi yang umumnya memiliki resolusi lebih rendah dibandingkan citra masukan. Selain itu, pemilihan lapisan target juga dapat memengaruhi hasil heatmap. Lapisan yang lebih awal memiliki resolusi lebih tinggi, tetapi fitur yang ditangkap masih bersifat umum, sedangkan lapisan akhir lebih spesifik terhadap kelas tetapi memiliki resolusi spasial yang lebih rendah[36].

2.14.2 Grad-CAM++

Grad-CAM++ merupakan pengembangan dari Grad-CAM yang dirancang untuk menghasilkan visualisasi perhatian model yang lebih rinci. Metode ini memperbaiki cara pemberian bobot terhadap area penting pada feature map, sehingga dapat memberikan lokalisasi yang lebih baik terutama ketika terdapat beberapa area yang berkontribusi terhadap kelas yang sama [36].

Secara konseptual, Grad-CAM++ tidak hanya melihat kontribusi umum dari suatu feature map, tetapi juga memberikan perhatian lebih rinci terhadap lokasi spasial yang memiliki pengaruh terhadap prediksi kelas. Oleh karena itu, heatmap Grad-CAM++ dapat menghasilkan area perhatian yang lebih tajam atau lebih terpisah dibandingkan Grad-CAM. Dalam konteks klasifikasi citra histopatologi, hal ini dapat membantu melihat apakah model memberikan perhatian pada pola

jaringan tertentu yang berhubungan dengan kelas normal atau abnormal[36].

Meskipun Grad-CAM++ dapat memberikan visualisasi yang lebih detail, hasilnya tetap harus ditafsirkan secara hati-hati. Heatmap yang lebih tajam tidak selalu berarti lebih benar secara klinis, terutama jika penelitian tidak menggunakan anotasi lokasi lesi sebagai ground truth. Hasil visualisasi juga dapat dipengaruhi oleh pemilihan lapisan target, ukuran feature map, preprocessing citra, serta karakteristik dataset yang digunakan[36].

Dalam penelitian ini, Grad-CAM dan Grad-CAM++ digunakan sebagai alat bantu interpretasi visual terhadap hasil klasifikasi EfficientNet-B0 dan EfficientNet-B1. Penggunaan sampel citra yang sama pada kedua model diperlukan agar perbandingan heatmap dapat dilakukan secara lebih konsisten. Dengan demikian, analisis XAI tidak digunakan untuk membuktikan kebenaran klinis secara langsung, tetapi untuk memahami pola perhatian visual model dalam membedakan citra histopatologi normal dan abnormal.

2.15 Penelitian Terdahulu

Penelitian terdahulu menunjukkan perkembangan klasifikasi histopatologi dari benchmark dataset, CNN, ensemble, Transformer, hingga arsitektur hibrida. Ringkasan berikut mempertahankan penelitian yang menjadi dasar Bab 2 dan menempatkan keterbatasannya terhadap fokus penelitian ini.

Tabel 2.2. Ringkasan penelitian terdahulu

Penelitian	Data/Tugas	Pendekatan dan temuan utama	Keterbatasan terhadap penelitian ini
Hu, Li, et al. (2022)	GasHisSDB, 245.196 patch histopatologi lambung yang terdiri atas 97.076 citra abnormal dan 148.120 citra normal pada ukuran 160 × 160, 120 × 120, dan 80 × 80 piksel.	Mempublikasikan dataset GasHisSDB dan melakukan benchmark menggunakan metode machine learning klasik serta deep learning. Model yang diuji mencakup VGG16, Inception-V3, ResNet50, dan Vision Transformer. Akurasi terbaik deep learning dilaporkan sebesar 96,47%, sedangkan metode machine learning klasik terbaik mencapai 86,08%.	Belum berfokus pada perbandingan EfficientNet-B0 dan EfficientNet-B1 secara terkendali serta belum menambahkan analisis XAI.

Penelitian	Data/Tugas	Pendekatan dan temuan utama	Keterbatasan terhadap penelitian ini
Hu, Chen, et al. (2022)	GasHisSDB sub-database A 160×160 , B 120×120 , dan C 80×80 piksel untuk klasifikasi biner normal dan abnormal.	Membandingkan fitur citra seperti color histogram, luminance histogram, HOG, LBP, dan GLCM dengan classifier LR, kNN, Naive Bayes, ANN, SVM, dan Random Forest. Penelitian ini juga membandingkan VGG16, Inception-V3, ResNet50, dan Vision Transformer. VGG16 mencapai akurasi 96,47% pada ukuran 120×120 , ResNet50 mencapai 96,09% pada ukuran 160×160 , dan ViT meningkat sampai 94,59% setelah eksperimen epoch tambahan.	Fokus penelitian terlalu luas pada banyak classifier, bukan trade-off dua varian EfficientNet yang berdekatan.
Fu et al. (2022)	Public gastric pathological image dataset dengan 4 kelas, yaitu normal tissue, tubular adenocarcinoma, mucinous adenocarcinoma, dan papillary adenocarcinoma; validasi tambahan dilakukan pada BreakHis dan Endometrium dataset.	Mengusulkan StoHisNet, yaitu model hibrida CNN-Transformer multi-scale. Transformer digunakan untuk menangkap fitur global, sedangkan CNN digunakan untuk menangkap fitur lokal. Pada dataset utama gastrik, model mencapai akurasi 94,69%, F1-score 94,96%, recall 94,95%, dan precision 94,97%. Pada BreakHis akurasi mencapai 91,64%, sedangkan pada Endometrium mencapai 81,74%.	Arsitektur lebih kompleks dan tugasnya multi-kelas, sehingga tidak mengisolasi kontribusi EfficientNet-B0 dan EfficientNet-B1 pada klasifikasi biner GasHisSDB.
Yong et al. (2023)	GasHisSDB untuk deteksi kanker lambung pada sub-database 160×160 serta resolusi 534×400 dan 1067×800 piksel.	Menguji berbagai model CNN berbasis transfer learning, termasuk MobileNet, EfficientNet, DenseNet, Inception, dan Xception, kemudian menggabungkan model terbaik menggunakan ensemble seperti majority voting, unweighted averaging, weighted averaging, dan stacking. Ensemble top-5 mencapai akurasi 99,20% pada sub-database 160×160 ; pada dataset 534×400 dan 1067×800 , akurasi tertinggi masing-masing mencapai 96,47% dan 97,99%.	Performa tinggi dicapai melalui ensemble beberapa model, tetapi biaya komputasi dan interpretasi model menjadi lebih kompleks dibandingkan penggunaan satu model EfficientNet.

Penelitian	Data/Tugas	Pendekatan dan temuan utama	Keterbatasan terhadap penelitian ini
Jang, Song, et al. (2021)	Whole-slide image TCGA-STAD dan dataset internal kanker lambung untuk subklasifikasi histologi gastric carcinoma.	Menggunakan deep learning patch-level untuk membedakan differentiated vs undifferentiated serta non-mucinous vs mucinous tumor pada WSI. Performa dilaporkan menggunakan AUC, yaitu AUC 0,932 untuk differentiated/undifferentiated dan AUC 0,979 untuk non-mucinous/mucinous.	Unit analisis berupa WSI dan tugasnya adalah subklasifikasi histologi, bukan klasifikasi patch biner normal-abnormal pada GasHisSDB.
Shi et al. (2022)	897 WSI spesimen endoskopi dari dua rumah sakit untuk klasifikasi negative for dysplasia, low-grade dysplasia, dan high-grade dysplasia/ intramucosal invasion neoplasia.	Mengembangkan DLDA, yaitu deep learning dengan domain adaptation untuk diagnosis displasia lambung lintas rumah sakit. Macro-averaged AUC pada validasi internal mencapai 0,97. Pada validasi eksternal, domain adaptation meningkatkan AUC dari 0,67 menjadi 0,82.	Fokus pada displasia endoskopi dan domain adaptation WSI, bukan klasifikasi patch biner GasHisSDB.
Veldhuizen et al. (2023)	TCGA-GC/TCGA-STAD untuk pelatihan, serta cohort eksternal KCCH Jepang dan KIEL Eropa untuk subtype Lauren intestinal dan diffuse.	Menggunakan attention-based multiple instance learning dengan feature extractor ResNet-50 berbasis RetCCL self-supervised learning. Validasi internal 5-fold pada TCGA menghasilkan mean AUROC $0,93 \pm 0,07$. Model juga digunakan untuk stratifikasi survival 5 tahun dan menghasilkan hazard ratio yang informatif pada cohort eksternal.	Tujuan penelitian adalah subtyping dan prognosis survival, bukan klasifikasi normal-abnormal berbasis patch.
Munien & Viriri (2021)	ICIAR 2018 BACH breast histology dataset dengan 4 kelas, yaitu normal, benign, in situ carcinoma, dan invasive carcinoma.	Menggunakan transfer learning EfficientNet-B0 sampai EfficientNet-B7 dengan stain normalization Reinhard dan Macenko. EfficientNet-B2 memberikan hasil terbaik dengan accuracy dan sensitivity 98,33% menggunakan Reinhard, serta accuracy dan sensitivity 96,67% menggunakan Macenko.	Organ yang digunakan adalah payudara, sehingga karakteristik data dan kelas berbeda dari histopatologi lambung.

Penelitian	Data/Tugas	Pendekatan dan temuan utama	Keterbatasan terhadap penelitian ini
Kallipolitis et al. (2021)	BreakHis breast cancer dataset dengan 7.909 citra pada magnifikasi 40x, 100x, 200x, dan 400x; serta dataset kolon berisi 100.000 citra dengan 9 kelas jaringan.	Menguji EfficientNetB0–B7, InceptionV3, Xception, VGG16, dan ResNet152V2, kemudian membuat ensemble EfficientNetB0–B2 dan EfficientNetB1–B3. Pada klasifikasi breast cancer, ensemble EfficientNetB0–B2 mencapai accuracy 99,25% dan AUC 0,9985. Pada dataset kolon, model mencapai accuracy 99,46% dan AUC 0,9991. Grad-CAM dan Guided Grad-CAM digunakan untuk interpretasi area penting.	Bukan dataset lambung dan bukan perbandingan terkendali EfficientNet-B0 dengan EfficientNet-B1 pada data serta konfigurasi pelatihan yang sama.
Springenberg et al. (2023)	Lima dataset histopatologi publik, yaitu PCam breast lymph-node metastasis 327.680 patch, BreaKHis 40x 2.013 citra, IDC breast 53.153 patch, GasHisSDB(A) gastric 33.284 patch, dan MHIST colorectal 3.152 citra.	Membandingkan ResNet-BiT, BotNet-50, InceptionV3, ConvNeXt-T/L, ViT-T/L, Swin-T/L, GasHis hybrid, CTransPath, RetCCL, dan ResNet-18 self-supervised. ConvNeXt-L menjadi model terbaik pada sebagian besar dataset, dengan accuracy 90,31% dan AUC 0,9722 pada PCam, accuracy 90,12% dan AUC 0,9545 pada IDC, accuracy 99,20% dan AUC 0,9996 pada GasHisSDB(A), serta accuracy 88,56% dan AUC 0,9539 pada MHIST. Pada BreaKHis 40x, GasHis hybrid memperoleh accuracy 93,00% dan AUC 0,9966.	Studi merupakan benchmark multi-arsitektur dan robustness, bukan fokus pada EfficientNet-B0/B1 serta bukan evaluasi dua tahap transfer learning dan fine-tuning.
Khayatian et al. (2025)	GasHisSDB ukuran 80×80 , 120×120 , dan 160×160 piksel.	Menggunakan pipeline hybrid deep feature extraction dan classifier gradient boosting. Beberapa pretrained network diuji untuk ekstraksi fitur, lalu fitur diklasifikasikan menggunakan classifier. Metode akhir menggabungkan EfficientNetV2B0 sebagai feature extractor dan CatBoost sebagai classifier. Accuracy yang dilaporkan adalah 89,7% pada 80×80 , 93,1% pada 120×120 , dan 93,9% pada 160×160 . Precision, recall, dan F1-score berada di atas 0,90. Grad-CAM digunakan untuk visualisasi area diskriminatif.	Pipeline menggunakan ekstraksi fitur dan CatBoost, sehingga berbeda dari fine-tuning end-to-end EfficientNet-B0/B1.

Penelitian	Data/Tugas	Pendekatan dan temuan utama	Keterbatasan terhadap penelitian ini
Plass et al. (2023)	Artikel konseptual pada konteks digital pathology, bukan eksperimen pada dataset klasifikasi tertentu.	Membahas perbedaan explainability dan causability dalam patologi digital serta menekankan bahwa heatmap XAI harus dievaluasi berdasarkan kegunaan bagi manusia, bukan hanya tampilan visual. Penelitian ini tidak melatih classifier dan tidak melaporkan accuracy.	Tidak melakukan eksperimen klasifikasi maupun perbandingan arsitektur EfficientNet.
Ren et al. (2025)	GasHisSDB, TCGA-STAD, dan NCT-CRC-HE-100K untuk klasifikasi histopatologi dan validasi eksternal.	Mengusulkan hybrid deep learning CNN–Transformer. CNN digunakan untuk mengekstraksi fitur lokal, sedangkan Transformer encoder digunakan untuk menangkap konteks global. Grad-CAM digunakan untuk interpretabilitas. Pada GasHisSDB, model mencapai accuracy 99,2%, macro F1-score 0,991, dan AUC 0,996.	Arsitektur hybrid lebih kompleks dan tidak mengisolasi perbandingan EfficientNet-B0 dengan EfficientNet-B1.
Xia et al. (2025)	Systematic scoping review terhadap 22 studi deep learning pada pathology image analysis kanker lambung; dataset yang dibahas mencakup GasHisSDB, TCGA, BOT gastric biopsy database, dan data klinis self-collected.	Memetakan tugas deteksi, klasifikasi, segmentasi, prognosis, serta kendala translasi klinis pada deep learning untuk patologi kanker lambung. Review ini melaporkan bahwa beberapa model deteksi kanker lambung mencapai accuracy di atas 95%, tetapi tidak menguji satu algoritma baru dan tidak memiliki satu angka accuracy utama.	Kajian literatur, bukan eksperimen langsung pada GasHisSDB atau perbandingan EfficientNet-B0/B1.

Berdasarkan Tabel 2.2 secara umum, studi pada GasHisSDB banyak membandingkan sejumlah algoritma, membangun ensemble, atau mengusulkan arsitektur hibrida. Pendekatan tersebut penting untuk memperoleh performa tinggi, tetapi tidak selalu memperlihatkan secara terisolasi manfaat peningkatan kapasitas dari satu varian EfficientNet ke varian yang berdekatan. Selain itu, hasil akhir sering dilaporkan tanpa memisahkan performa sebelum dan sesudah fine-tuning.

2.16 Celah Penelitian dan Posisi Penelitian

Kajian terdahulu memperlihatkan dua kecenderungan. Pertama, penelitian histopatologi lambung berkembang menuju WSI, prediksi biomarker, subtype

histologis, ensemble, Transformer, dan model hibrida lintas dataset. Kedua, penelitian EfficientNet dan XAI banyak dilakukan pada organ lain atau sebagai bagian dari gabungan beberapa model [37, 12]. Kondisi tersebut menyisakan ruang untuk evaluasi yang lebih sederhana, terkendali, dan mudah direplikasi pada GasHisSDB.

Celah utama penelitian terletak pada belum terfokusnya perbandingan langsung EfficientNet-B0 dan EfficientNet-B1 menggunakan data split, preprocessing, augmentasi, optimizer, dan prosedur evaluasi yang sama. Selain itu, kontribusi transfer learning dan fine-tuning perlu dipisahkan agar peningkatan performa dapat dikaitkan dengan tahap adaptasi model. Pemilihan model praktis juga memerlukan analisis jumlah parameter, parameter trainable, epoch efektif, dan waktu pelatihan, bukan hanya accuracy.

Dari sisi kualitas data, kemiripan atau duplikasi citra lintas split dapat menyebabkan estimasi performa terlalu optimistis. Pelaporan pemeriksaan duplikat pada penelitian klasifikasi patch masih terbatas. Dari sisi interpretabilitas, heatmap tidak selalu disertakan atau dapat ditafsirkan secara berlebihan. Kajian mutakhir menempatkan generalisasi, validasi eksternal, interpretabilitas, dan reproduisibilitas sebagai kendala translasi deep learning pada patologi kanker lambung [12].

Tabel 2.3. Pemetaan celah dan posisi penelitian

Aspek	Kecenderungan penelitian terdahulu	Posisi penelitian ini
Perbandingan arsitektur	Banyak model, ensemble, atau arsitektur hibrida.	B0 dan B1 dibandingkan langsung dalam kondisi identik.
Tahap pembelajaran	Sering hanya melaporkan hasil akhir.	Transfer learning dan fine-tuning dievaluasi terpisah.
Efisiensi	Fokus utama pada accuracy atau performa prediksi.	Parameter, epoch efektif, dan waktu pelatihan dianalisis.
Kualitas data	Pelaporan duplikat lintas split terbatas.	Average hash digunakan untuk penyaringan sebelum pelatihan.
Interpretabilitas	Heatmap tidak selalu disertakan atau hanya satu metode.	Grad-CAM dan Grad-CAM++ diterapkan pada sampel yang sama.
Kehati-hatian klinis	Heatmap berisiko dianggap sebagai bukti lesi.	Interpretasi dibatasi sebagai inspeksi kualitatif tanpa anotasi patolog.

Berdasarkan Tabel 2.3 dengan demikian, posisi penelitian ini adalah evaluasi komparatif yang sederhana tetapi terkendali. Penelitian tidak mengusulkan arsitektur baru, melainkan mengisolasi pengaruh kapasitas EfficientNet-B0 dan EfficientNet-B1, mengukur manfaat fine-tuning, mendokumentasikan pemeriksaan duplikat lintas split, menilai biaya komputasi, dan membandingkan Grad-CAM serta Grad-CAM++ pada sampel yang sama. Posisi tersebut melengkapi penelitian berarsitektur kompleks dengan baseline yang lebih mudah direplikasi dan lebih jelas untuk menilai trade-off model.