

BAB 2

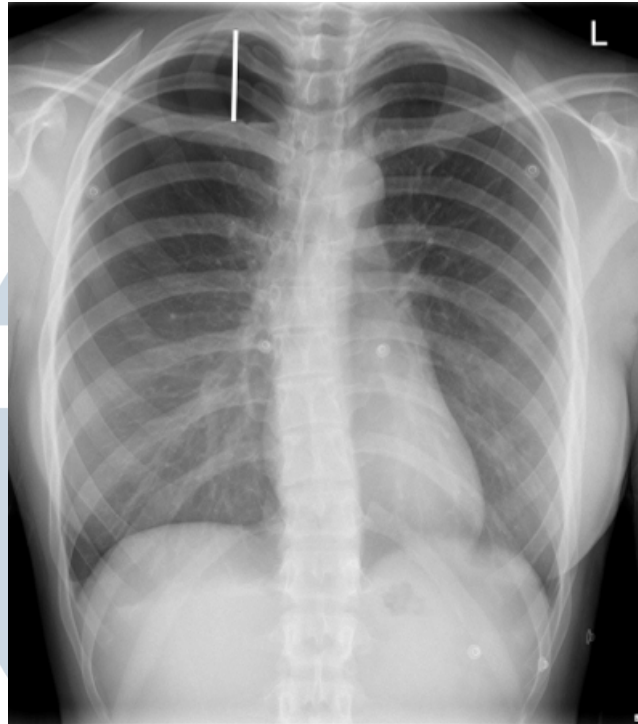
LANDASAN TEORI

Bab ini menguraikan landasan teori yang mendasari penelitian terkait segmentasi *pneumothorax* pada citra *chest X-ray* menggunakan arsitektur U-Net dengan variasi *backbone Convolutional Neural Network* (CNN). Pembahasan dimulai dari tinjauan medis mengenai *pneumothorax* beserta karakteristik visualnya pada citra radiologi, dilanjutkan dengan konsep dasar segmentasi citra medis pada tingkat piksel melalui pendekatan *semantic segmentation*. Selanjutnya, dijabarkan arsitektur U-Net sebagai kerangka segmentasi utama beserta konsep *backbone* sebagai pengekstraksi fitur pada jalur *encoder*, yang mencakup arsitektur ResNet-50, dan DenseNet-169. Pemilihan *backbone* tersebut didasarkan pada kemampuan masing-masing dalam menangani permasalahan ekstraksi fitur pada citra medis sebagaimana telah dibuktikan dalam penelitian-penelitian sebelumnya.

2.1 *Pneumothorax*

Pneumothorax adalah istilah medis yang menggambarkan kondisi kolapsnya bagian paru-paru, yang terjadi ketika gas atau udara masuk ke dalam rongga pleura. Kondisi ini menyebabkan paru-paru gagal mengembang secara maksimal akibat tekanan udara pada rongga pleura [16]. Udara yang berada pada rongga pleura memberikan tekanan pada bagian luar paru-paru, yang menyebabkan kesulitan dalam bernapas [1].

Pneumothorax terbagi menjadi dua, yaitu *Primary Spontaneous Pneumothorax* (PSP) dan *Secondary Spontaneous Pneumothorax* (SSP). *Primary Spontaneous Pneumothorax* (PSP) terjadi tanpa adanya penyakit paru yang tampak klinis, sementara *Secondary Spontaneous Pneumothorax* (SSP) muncul akibat berbagai kondisi penyakit paru yang mendasarinya [6]. *Pneumothorax* dapat disebabkan oleh berbagai kondisi medis, seperti trauma, penyakit paru obstruktif kronis (PPOK), infeksi paru, maupun kelainan jaringan paru.



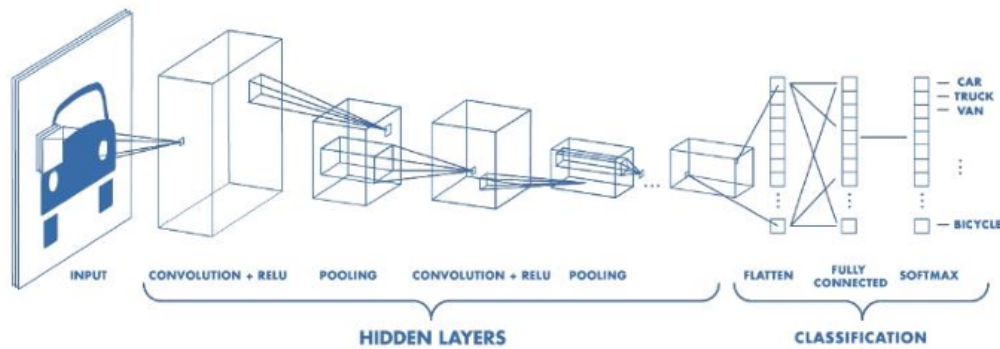
Gambar 2.1. Visualisasi pneumothorax pada citra *chest X-ray*. [1].

Gambar 2.1 memperlihatkan bagaimana karakteristik dari *pneumothorax* pada *chest x-ray*. Indikasi utama dari *pneumothorax* adalah munculnya garis batas paru-paru yang terpisah dari dinding dada bagian dalam. Dalam beberapa kasus, perlengketan jaringan dapat menyebabkan udara yang bocor terperangkap [6]. Selain itu, Kompleksitas identifikasi dapat meningkat karena adanya struktur anatomi yang tumpang tindih, variasi ukuran organ, letak kelainan pada paru-paru, serta posisi pasien saat pengambilan gambar yang berpotensi menyebabkan distorsi citra [8]. Kompleksitas karakteristik visual tersebut menjadi pertimbangan utama dalam pemilihan pendekatan segmentasi berbasis deep learning yang mampu melakukan klasifikasi pada tingkat piksel secara presisi.

2.2 Convolutional Neural Network

Convolution Neural Network (CNN) merupakan salah satu algoritma *deep learning* yang biasa digunakan untuk membantu menyelesaikan permasalahan seperti klasifikasi gambar, melakukan segmentasi serta mendeteksi suatu objek [2]. Di dalam CNN terdapat beberapa serangkaian layer, seperti *convolutional layer* dan *pooling layer* yang berfungsi untuk melakukan ekstraksi fitur relevan dari gambar yang di input. Keunggulan dari CNN adalah dapat menangkap fitur dan

pola spasial pada citra menggunakan arsitektur lapisan hierarkis yang menjalankan operasi konvolusi serta mengekstrak fitur pada berbagai tingkat abstraksi [17].



Gambar 2.2. Arsitektur CNN [2]

Gambar 2.2 memvisualisasikan bagaimana CNN bekerja dalam melakukan ekstraksi dari gambar *input*. *Convolution* layer merupakan lapisan pertama dari CNN yang menjalankan operasi konvolusi terhadap gambar yang di input menggunakan sebuah filter yang merupakan sekumpulan matriks yang berfungsi mengekstrak fitur atau pola spesifik, seperti garis, tepi, atau tekstur. Setelah itu, dilanjut pada *pooling layer* yang digunakan untuk mengurangi dimensi spasial dari *feature map* yang dihasilkan oleh *convolution layer*. Lapisan ini beroperasi pada tiap *feature map* untuk melakukan *downsampling* (penurunan resolusi) dengan mengambil nilai maximum atau rata-rata dari *non-overlapping region* [17].

Kemampuan CNN dalam melakukan ekstraksi fitur secara hierarkis dan otomatis menjadikannya algoritma yang dominan dalam bidang analisis citra medis. Menurut (Mienye et al., 2025) dalam kajian komprehensifnya menyatakan bahwa CNN mampu secara otomatis mempelajari fitur hierarkis dari data pencitraan medis, mulai dari fitur tingkat rendah seperti tepi dan tekstur hingga fitur tingkat tinggi yang lebih kompleks, tanpa memerlukan rekayasa fitur secara manual [18]. Kemampuan *automatic feature learning* ini terbukti menghasilkan terobosan signifikan dalam akurasi diagnostik dan perawatan pasien. Lebih lanjut, CNN telah terbukti mendominasi berbagai modalitas pencitraan medis seperti *X-ray*, *Computed Tomography* (CT), *Magnetic Resonance Imaging* (MRI), dan *Positron Emission Tomography* (PET), dengan aplikasi yang mencakup deteksi tumor, segmentasi organ, hingga peningkatan kualitas citra medis [18].

Bukti bahwa CNN mampu melakukan analisis citra medis telah ditunjukkan secara luas pada berbagai bidang klinis. Pada bidang onkologi, model berbasis

CNN berhasil mencapai akurasi sebesar 99,17% dalam deteksi kanker payudara menggunakan mammogram [18]. Pada bidang neurologi, *multi-modal* CNN yang mengintegrasikan data MRI dan PET mencapai akurasi 98,59% dalam klasifikasi penyakit Alzheimer [18]. Sementara itu, pada bidang pulmonologi, model CNN-LSTM berhasil memprediksi eksaserbasi COPD dengan akurasi mencapai 99,01% [18]. Hasil-hasil tersebut menunjukkan bahwa CNN secara konsisten mampu mengekstraksi fitur spasial yang relevan dari berbagai jenis citra medis dengan performa yang tinggi di berbagai domain klinis.

Menurut penelitian (Abedalla et al., 2021) menyatakan bahwa model deep learning berbasis CNN telah mencapai keberhasilan dalam berbagai tugas computer vision, termasuk klasifikasi citra, deteksi objek, dan segmentasi citra medis. Penelitian tersebut menggunakan U-Net dengan *backbone* CNN *pre-trained* mampu mencapai *Dice Similarity Coefficient* (DSC) sebesar 86,06% dalam melakukan segmentasi terhadap *pneumothorax*. Hasil tersebut menunjukkan bahwa CNN memiliki kemampuan yang baik dalam mengekstraksi fitur spasial pada citra *chest X-ray* sehingga efektif digunakan sebagai *encoder* dalam tugas segmentasi *pneumothorax*. Oleh karena itu, CNN menjadi fondasi utama dalam pengembangan arsitektur U-Net dengan variasi backbone pada penelitian ini [15].

2.3 Semantic Segmentation

Menurut penelitian (Azad et al., 2024) dalam analisis citra medis, pendekatan *semantic segmentation* dipilih karena kemampuannya dalam melakukan klasifikasi pada tingkat piksel yang di mana setiap piksel di dalam sebuah citra diberikan label kategori spesifik, menghasilkan *segmentation mask* dengan dimensi yang sama dengan gambar input. Berbeda dengan klasifikasi gambar yang hanya menghasilkan satu label untuk keseluruhan citra, maupun *instance segmentation* yang berfokus pada deteksi objek dengan batas yang tegas, *semantic segmentation* mampu memberikan lokalisasi anatomi secara presisi melalui klasifikasi per-piksel [14].

Pendekatan ini sangat cocok diterapkan pada segmentasi *pneumothorax* pada citra *X-ray*, mengingat karakteristik visual *pneumothorax* yang memiliki indikator visual utama yang tipis, struktur anatomi yang tumpang tindih, serta letak kelainan pada paru-paru yang berbeda menyebabkan proses segmentasi lebih kompleks [7][8]. Oleh karena itu, pendekatan yang mampu melakukan pelabelan pada setiap piksel diperlukan untuk memperoleh batas area *pneumothorax* secara

lebih akurat.

Menurut penelitian (Abedalla et al., 2021), segmentasi merupakan salah satu metode utama dalam analisis citra medis karena mampu memberikan *pixel-level image understanding* melalui proses pelabelan pada setiap piksel penyusun citra. Kemampuan tersebut memungkinkan model menghasilkan representasi area kelainan secara lebih detail dibandingkan pendekatan klasifikasi yang hanya menentukan keberadaan suatu penyakit tanpa menunjukkan lokasinya [15]. Berdasarkan karakteristik tersebut, pendekatan *semantic segmentation* dipilih dalam penelitian ini karena mampu menghasilkan peta segmentasi yang menggambarkan lokasi serta batas area *pneumothorax* secara rinci pada tingkat piksel.

2.4 U-Net

Menurut studi yang dilakukan oleh (Aranibar Roque & Sebastian., 2025), U-Net merupakan *convolutional neural network* yang pada awalnya dirancang untuk keperluan segmentasi citra biomedis. Arsitektur ini memiliki desain yang menyerupai huruf U dan terdiri dari dua jalur yang simetris, yaitu jalur *encoder* dan jalur *decoder*, yang dihubungkan oleh *skip connections* pada setiap level spasial yang bersesuaian guna mempertahankan detail spasial yang halus (seperti garis tepi dan tekstur) yang berpotensi hilang akibat proses *pooling* [3][19].



berupa garis pleura yang tipis dan memiliki kontras rendah terhadap jaringan di sekitarnya [19].

Selain itu, penelitian tersebut menjelaskan bahwa sifat *fully convolutional* yang dipadukan dengan arsitektur *encoder* dan *decoder* menjadikan U-Net sangat sesuai untuk tugas segmentasi citra medis. Melalui proses *single forward pass*, U-Net mampu menghasilkan *dense pixel-wise classification*, yaitu proses pelabelan pada setiap piksel sehingga setiap piksel pada citra dapat dipetakan ke dalam kategori tertentu, seperti organ, jaringan normal, maupun area kelainan [19].

Dalam kasus Segmentasi *pneumothorax* (Aranibar Roque & Sebastian., 2025) juga menjelaskan bahwa dalam kasus *pneumothorax*, arsitektur ini memiliki keunggulan yang signifikan. Pada citra *X-ray* maupun *CT scan*, karakteristik dari *pneumothorax* sering kali hanya terbatas pada garis pleura yang samar dengan tingkat kontras yang sangat rendah. Fitur *skip connections* pada U-net berfungsi untuk mempertahankan informasi citra tingkat rendah (*low-level*) yang dibutuhkan untuk mendeteksi batas-batas halus tersebut, sedangkan bagian *encoder* menyediakan informasi anatomi secara kontekstual untuk memandu proses penentuan batas area secara akurat [19].

2.5 ResNet-50

Menurut (Abedalla et al., 2021), ResNet (*Residual Network*) merupakan arsitektur CNN yang dikembangkan untuk mengatasi permasalahan *vanishing gradient* pada jaringan yang sangat dalam melalui mekanisme *residual learning*. ResNet memungkinkan gradien mengalir secara langsung melalui *identity shortcut connection*, sehingga proses pembelajaran pada lapisan awal tetap dapat berlangsung secara efektif meskipun jumlah lapisan jaringan semakin bertambah [15]. Salah satu varian yang paling banyak digunakan adalah ResNet-50 yang terdiri dari 50 lapisan dan menawarkan keseimbangan antara kedalaman jaringan, kompleksitas komputasi, dan kemampuan ekstraksi fitur [18].

Mekanisme *residual connection* memungkinkan jaringan mempelajari representasi fitur secara bertahap dari lapisan dangkal hingga lapisan yang lebih dalam tanpa kehilangan informasi penting yang telah dipelajari sebelumnya. Karakteristik ini menjadikan ResNet50 banyak digunakan pada berbagai tugas analisis citra medis yang memerlukan identifikasi pola visual kompleks dan berkontras rendah. Selain itu, kemampuan ResNet50 dalam melakukan ekstraksi fitur hierarkis menjadikannya salah satu backbone yang sering digunakan pada citra

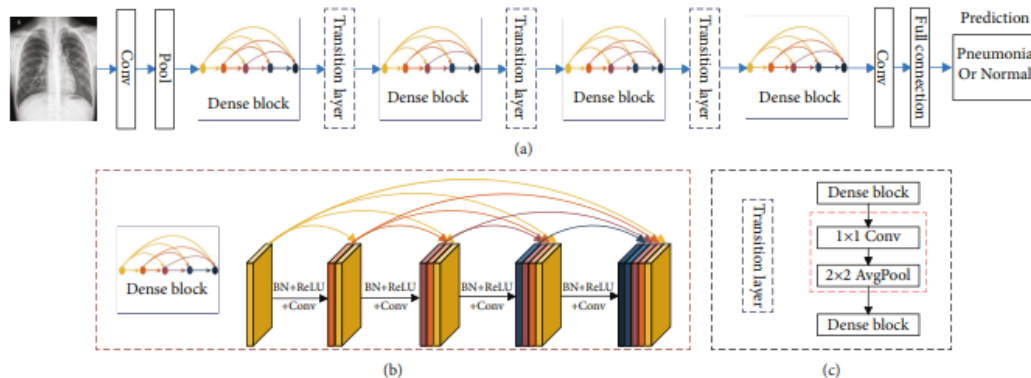
toraks (*chest X-ray*) untuk mendeteksi berbagai abnormalitas paru-paru [8].

Pada kasus *pneumothorax*, (Abedalla et al. 2021) mengintegrasikan ResNet-50 sebagai *encoder* pada arsitektur U-Net untuk melakukan segmentasi area *pneumothorax* pada dataset SIIM-ACR *Pneumothorax Segmentation*. Hasil penelitian tersebut menunjukkan bahwa penggunaan backbone ResNet-50 mampu menghasilkan representasi fitur yang efektif untuk mengidentifikasi area *pneumothorax* yang memiliki batas tipis dan kontras rendah pada citra *chest X-ray* [15]. Selain itu, laporan solusi terbaik pada kompetisi SIIM-ACR *Pneumothorax Segmentation* menunjukkan bahwa beberapa tim peringkat teratas memanfaatkan arsitektur U-Net dengan *encoder* berbasis ResNet, termasuk ResNet34 dan ResNet50, yang menunjukkan bahwa keluarga ResNet merupakan backbone yang terbukti efektif untuk tugas segmentasi *pneumothorax* [15].

Berdasarkan kemampuan *residual learning*, keberhasilannya dalam berbagai tugas analisis citra toraks, serta penggunaannya pada penelitian dan kompetisi segmentasi *pneumothorax* berskala besar, ResNet50 dipilih sebagai salah satu *backbone encoder* yang akan dievaluasi pada penelitian ini.

2.6 DenseNet169

Menurut (Zhou et al., 2022), DenseNet (*Dense Convolutional Network*) merupakan arsitektur *convolutional neural network* (CNN) yang menerapkan mekanisme *dense connection*, yaitu setiap lapisan terhubung secara langsung dengan seluruh lapisan berikutnya melalui proses *feed-forward*. Berbeda dengan arsitektur CNN konvensional, DenseNet menggunakan mekanisme *tensor concatenation* sehingga setiap lapisan menerima *feature map* dari seluruh lapisan sebelumnya sebagai masukan. Mekanisme tersebut memungkinkan terjadinya *feature reuse*, yaitu penggunaan kembali informasi yang telah dipelajari oleh lapisan-lapisan sebelumnya selama proses ekstraksi fitur [4][15].



Gambar 2.4. Arsitektur DenseNet [4]

Gambar 2.4 memvisualisasikan arsitektur DenseNet secara keseluruhan. Bagian (a) menunjukkan proses pengolahan citra yang diawali dari lapisan konvolusi awal, kemudian diteruskan secara berurutan melalui beberapa *dense block* dan *transition layer* hingga menghasilkan keluaran akhir. Bagian (b) memperlihatkan mekanisme *dense connection* di dalam setiap *dense block*, di mana setiap lapisan menerima dan meneruskan *feature map* ke seluruh lapisan berikutnya. Sementara itu, bagian (c) menunjukkan struktur *transition layer* yang terdiri atas konvolusi (1×1) dan *average pooling* (2×2) yang berfungsi untuk mengurangi dimensi *feature map* sebelum diteruskan ke *dense block* berikutnya [4].

Menurut (Abedalla et al., 2021), varian DenseNet169 terdiri atas 169 lapisan yang tersusun dalam beberapa *dense block* yang dihubungkan oleh *transition layer*. Melalui mekanisme *dense connection*, setiap lapisan dapat memanfaatkan informasi dari seluruh lapisan sebelumnya sehingga proses ekstraksi fitur berlangsung secara bertahap dan informasi yang telah dipelajari tidak hilang selama jaringan semakin dalam [15].

Penelitian (Zhou et al., 2022) menjelaskan bahwa mekanisme *feature reuse* pada DenseNet memungkinkan informasi spasial yang diperoleh pada lapisan awal, seperti garis, tepi, dan tekstur, tetap dapat dimanfaatkan oleh lapisan-lapisan yang lebih dalam. Karakteristik tersebut mendukung proses ekstraksi fitur secara hierarkis sehingga model mampu mempelajari representasi visual dari tingkat rendah hingga tingkat tinggi secara berkelanjutan [4].

Lebih lanjut, (Abedalla et al., 2021) menggunakan DenseNet sebagai *backbone encoder* pada arsitektur U-Net untuk melakukan segmentasi *pneumothorax* pada citra *chest X-ray*. Penggunaan mekanisme *dense connection* memungkinkan *feature map* yang dihasilkan *encoder* tetap mempertahankan

informasi spasial dari lapisan-lapisan sebelumnya sehingga dapat diteruskan ke bagian decoder melalui *skip connection*. Dengan demikian, proses rekonstruksi *segmentation mask* dapat memanfaatkan informasi spasial dan semantik secara bersamaan untuk menghasilkan segmentasi yang lebih baik [15].

Berdasarkan penelitian tersebut, DenseNet169 memiliki karakteristik yang sesuai untuk digunakan sebagai *backbone encoder* pada arsitektur U-Net karena mampu mempertahankan informasi dari setiap lapisan melalui mekanisme *dense connection*. Oleh karena itu, DenseNet169 dipilih sebagai salah satu variasi *backbone* pada penelitian ini untuk dianalisis performanya dalam melakukan segmentasi pneumothorax pada citra *chest X-ray*.

2.7 Transfer Learning

Transfer learning merupakan pendekatan yang banyak digunakan dalam pengembangan model *deep learning* untuk citra medis karena keterbatasan data berlabel yang tersedia. Melalui pendekatan ini, model tidak dilatih dari awal (*from scratch*), melainkan memanfaatkan bobot awal yang telah dipelajari dari dataset berskala besar seperti ImageNet, kemudian dilakukan penyesuaian (*fine-tuning*) pada dataset target [20].

Penggunaan *transfer learning* telah banyak diterapkan pada analisis citra *chest X-ray*. (Tang et al., 2022) melakukan evaluasi berbagai strategi *transfer learning* pada arsitektur ResNet-50 dan DenseNet-121 untuk klasifikasi multilabel penyakit paru. Hasil penelitian menunjukkan bahwa ResNet-50 memberikan performa yang lebih baik pada skema inisialisasi langsung menggunakan bobot ImageNet (*transfer learning initialization*), sedangkan DenseNet menunjukkan performa yang lebih baik ketika dikombinasikan dengan pendekatan *co-training* dan *feature concatenation* [21]. Temuan tersebut menunjukkan bahwa efektivitas transfer learning dipengaruhi oleh karakteristik arsitektur backbone yang digunakan.

Pada kasus pneumothorax, (Malhotra et al., 2022) menerapkan *transfer learning* menggunakan *backbone* ResNet pada arsitektur Mask R-CNN untuk melakukan deteksi dan segmentasi *pneumothorax* pada dataset SIIM-ACR. Hasil penelitian menunjukkan bahwa penggunaan bobot prapelatihan dari ImageNet mampu meningkatkan kemampuan ekstraksi fitur serta mempercepat proses konvergensi dibandingkan pelatihan tanpa bobot prapelatihan [8].

Berdasarkan penelitian-penelitian tersebut, *transfer learning* dipilih dalam

penelitian ini karena mampu menyediakan representasi fitur awal yang telah terlatih pada jutaan citra, sehingga model dapat lebih fokus mempelajari karakteristik khusus pneumothorax pada citra *chest X-ray*. Pendekatan ini juga diharapkan dapat meningkatkan kemampuan generalisasi model serta mengurangi risiko *overfitting* selama proses pelatihan.

2.8 Loss Function

Pemilihan *loss function* merupakan salah satu faktor penting yang mempengaruhi performa model segmentasi citra medis, terutama pada dataset yang memiliki ketidakseimbangan kelas (*class imbalance*) seperti SIIM-ACR Pneumothorax Segmentation. Pada dataset ini, jumlah piksel latar belakang jauh lebih dominan dibandingkan piksel lesi *pneumothorax* sehingga penggunaan fungsi kerugian konvensional sering kali menghasilkan model yang bias terhadap kelas mayoritas [22].

Penelitian (Abedalla et al., 2021) menggunakan fungsi kerugian untuk melatih model yang merupakan jumlah dari *Binary Cross-Entropy* (BCE) dan *Dice*. Nilai *Dice loss* didasarkan pada *Dice Similarity Coefficient* (DSC). DSC bekerja secara efisien pada masalah ketidakseimbangan kelas dan biasanya diaplikasikan sebagai metrik dalam tugas segmentasi citra untuk menghitung kesamaan antara piksel yang diprediksi dan *ground truth* yang sesuai [15].

Penelitian oleh Wang dan Lu (2023) menyatakan bahwa terdapat sejumlah besar ketidakseimbangan kelas pada dataset gambar paru-paru SIIM-ACR. Jika hanya fungsi kerugian cross-entropy dan *Dice* yang digunakan, pelatihan akan didominasi oleh area latar belakang yang memiliki lebih banyak piksel, sehingga menyulitkan pembelajaran fitur-fitur dari posisi pneumothorax yang lebih kecil dan mengurangi efektivitas jaringan. Oleh karena itu, penelitian tersebut memperkenalkan fungsi Focal loss untuk menyelesaikan masalah ketidakseimbangan sampel tersebut [23].

Sementara itu, (Yeung et al., 2022) melakukan evaluasi terhadap berbagai fungsi kerugian segmentasi termasuk *Dice Loss*, *Tversky Loss*, dan *Focal Tversky Loss* pada dataset dengan tingkat ketidakseimbangan kelas yang berbeda-beda. Hasil penelitian menunjukkan bahwa Focal Tversky Loss mampu meningkatkan sensitivitas model terhadap objek berukuran kecil karena memberikan penalti yang lebih besar terhadap kesalahan deteksi (*false negative*) [22].

Berdasarkan temuan-temuan tersebut, penelitian ini menggunakan

kombinasi *Binary Cross Entropy* (BCE) dan Dice Loss, yang selanjutnya disebut *BCE-Dice Loss*, sebagai fungsi kerugian pada proses pelatihan model. Kombinasi ini dipilih karena BCE mampu mengoptimalkan klasifikasi piksel secara individual dengan mempertimbangkan probabilitas prediksi pada setiap piksel, sedangkan *Dice Loss* secara langsung mengoptimalkan tingkat kesesuaian (*overlap*) antara area prediksi dan ground truth sehingga lebih tahan terhadap ketidakseimbangan kelas dibandingkan BCE saja. Penggabungan kedua fungsi kerugian tersebut bertujuan untuk memperoleh keseimbangan antara akurasi klasifikasi piksel dan kesesuaian area segmentasi secara keseluruhan.

2.9 Lingkup Pengerjaan

Penelitian ini dilaksanakan menggunakan platform komputasi berbasis cloud Google Colaboratory dengan akselerator perangkat keras GPU NVIDIA A100 40GB. Pemilihan platform ini didasarkan pada tiga pertimbangan teknis: pertama, kapasitas memori GPU A100 yang memadai untuk menangani proses forward propagation dan backpropagation pada resolusi citra 512×512 piksel dengan batch size 16 secara simultan. Penggunaan platform komputasi dapat mendukung terhadap mixed *precision training* (FP16) yang secara signifikan mereduksi konsumsi memori tanpa degradasi akurasi numerik yang berarti. Integrasi langsung dengan Google Drive sebagai penyimpanan persistent untuk checkpoint model, dataset, dan pengerjaan selama siklus pelatihan berlangsung.

