

BAB III

METODE PENELITIAN

3.1. Metode Penelitian

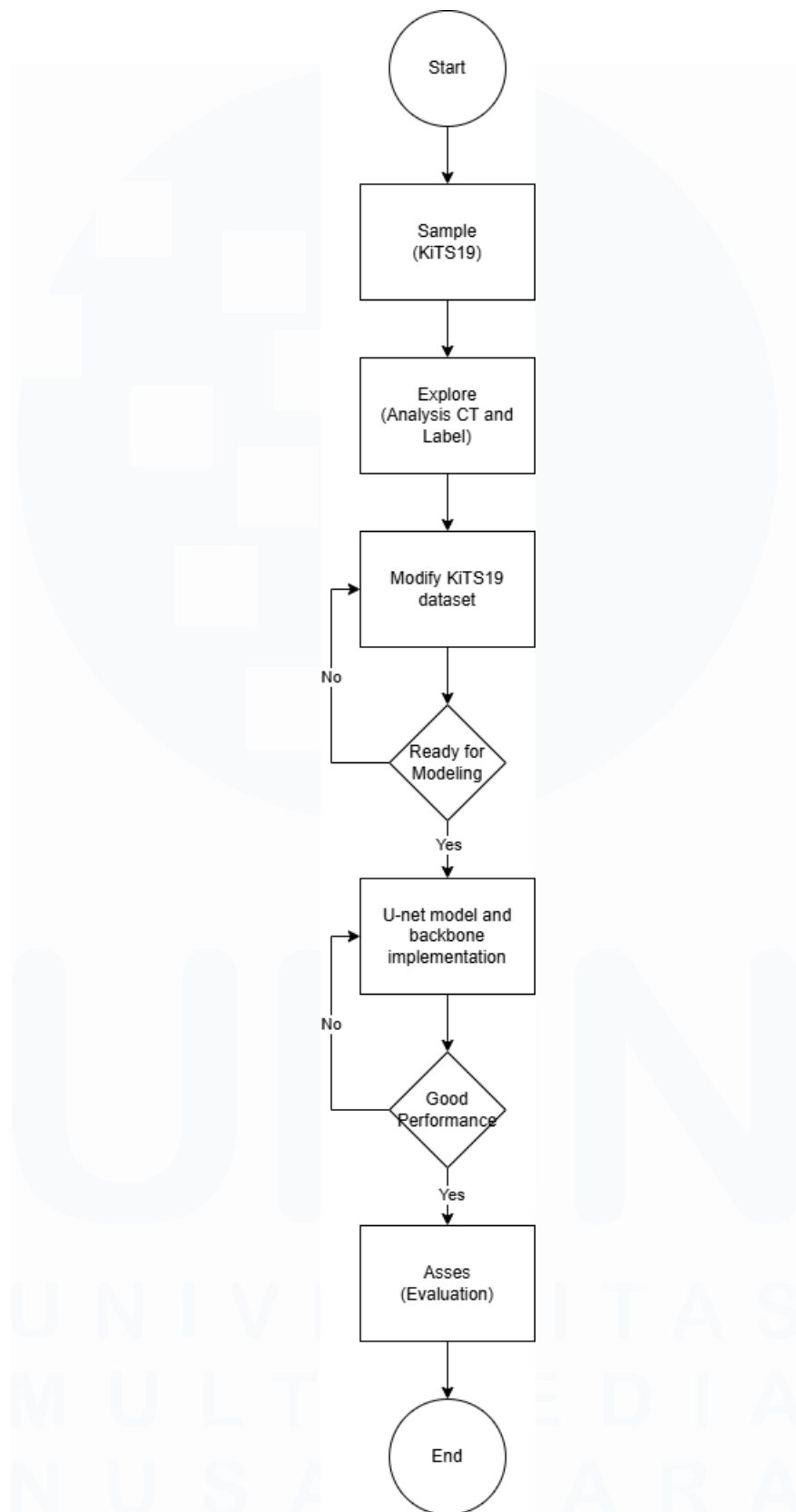
Penelitian ini menggunakan pendekatan SEMMA (Sample, Explore, Modify, Model, Assess) yang dikembangkan oleh SAS Institute. SEMMA merupakan metodologi analisis data yang berorientasi pada eksplorasi dan pemodelan data skala besar, sehingga sangat cocok diterapkan dalam studi yang melibatkan segmentasi citra medis, seperti deteksi ginjal dan tumor pada data CT Scan.

Metode SEMMA terdiri dari lima tahapan utama, yaitu:

- 1) Sample: Pengambilan dan pemilihan data yang relevan untuk penelitian.
- 2) Explore: Eksplorasi karakteristik data untuk memahami pola, distribusi, dan potensi tantangan.
- 3) Modify: Proses transformasi dan rekayasa fitur agar data siap digunakan dalam pemodelan.
- 4) Model: Tahapan pembangunan model pembelajaran mesin atau deep learning.
- 5) Assess: Evaluasi performa model menggunakan metrik yang sesuai.

Pendekatan SEMMA dipilih karena memberikan struktur kerja yang sistematis, fleksibel, serta memungkinkan iterasi berulang guna meningkatkan akurasi model. Dalam konteks segmentasi medis, di mana data cenderung kompleks, metode ini sangat berguna untuk memastikan bahwa seluruh proses, mulai dari akuisisi data hingga evaluasi model, dilakukan secara terkontrol dan dapat direplikasi.

3.1.1 Alur Penelitian



Gambar 3.1 Flowchart Alur Penelitian

Alur penelitian secara umum digambarkan dalam Gambar 3.1 Flowchart Penelitian, yang menunjukkan proses terintegrasi dari tahap awal hingga akhir penelitian.

Penelitian dimulai dengan tahap Sample, di mana data diambil dari dataset Kidney Tumor Segmentation Challenge 2019 (KiTS19). Dataset ini dipilih karena memiliki struktur data lengkap, terdiri dari citra CT Scan abdomen pasien yang telah dilakukan nephrectomy serta label segmentasi manual oleh ahli radiologi.

Tahap berikutnya adalah Explore, yaitu proses eksplorasi data. Pada tahap ini dilakukan pemeriksaan menyeluruh terhadap kualitas citra, format data, distribusi kelas segmentasi (background, ginjal, tumor), serta potensi permasalahan seperti artefak gambar, inkonsistensi label, atau ketidakseimbangan kelas.

Kemudian dilanjutkan ke tahap Modify, yaitu proses modifikasi dan rekayasa data. Proses ini meliputi konversi data dari format asli .nii.gz (3D volumetric) menjadi irisan citra 2D berformat .png, pewarnaan label segmentasi untuk membedakan kelas secara visual, normalisasi nilai piksel ke rentang $[0,1]$, resizing citra menjadi resolusi 256x256 piksel, serta augmentasi data untuk meningkatkan keragaman dan generalisasi model.

Tahap Model merupakan inti dari penelitian ini, di mana dilakukan pelatihan model segmentasi citra berbasis U-Net, dipadukan dengan lima jenis backbone berbeda yaitu VGG19, MobileNetV2, ResNet50, DenseNet121, dan EfficientNetB7. Proses pelatihan memanfaatkan platform Google Colab, yang menyediakan GPU untuk percepatan komputasi, sehingga proses eksperimen dapat dilakukan secara efisien tanpa bergantung pada perangkat keras lokal.

Jika model telah selesai dilatih, masuk ke tahap Assess, yaitu evaluasi performa model menggunakan metrik Dice Coefficient, terutama dice khusus kelas tumor, yang menjadi fokus utama penelitian ini. Evaluasi dilakukan secara kuantitatif dan visual, dengan membandingkan prediksi model terhadap label ground truth. Apabila performa model belum memenuhi standar, dilakukan iterasi ulang dari tahap Modify atau Model hingga didapatkan model yang optimal.

Proses ini berlanjut hingga diperoleh hasil evaluasi yang memadai, dan penelitian dapat dinyatakan selesai.

3.2 Tahapan Penelitian

Tahapan SEMMA dalam konteks penelitian ini dijelaskan sebagai berikut:

1) Sample (Pengambilan Sampel)

Tahap awal ini melibatkan pemilihan data dari sumber yang relevan untuk tujuan penelitian. Dataset yang digunakan adalah Kidney Tumor Segmentation Challenge 2019 (KiTS19), yang terdiri dari data CT scan abdomen dari 210 pasien penderita tumor ginjal, lengkap dengan label segmentasi yang menunjukkan area ginjal dan tumor. Dataset ini dipilih karena bersifat terbuka, lengkap, dan sering digunakan dalam penelitian segmentasi medis.

2) Explore (Eksplorasi Data)

Pada tahap ini dilakukan analisis awal terhadap struktur data, kualitas citra, serta distribusi kelas pada segmentasi. Data dalam format .nii.gz dipelajari untuk memahami nilai intensitas piksel, keberadaan artefak, dan karakteristik anatomi yang ditandai dalam label segmentasi. Hasil eksplorasi menunjukkan bahwa segmentasi mencakup tiga kelas utama: latar belakang (background), ginjal (parenkim), dan tumor.

3) Modify (Modifikasi Data)

Tahap ini mencakup proses preprocessing dan augmentasi data. Format asli .nii.gz dikonversi secara manual ke dalam bentuk .png agar kompatibel dengan pipeline deep learning berbasis Keras. Proses ini dilakukan secara manual untuk mendapatkan data gambar slice yang mencakup segmentasi ginjal (medulla) dan tumor. Segmentasi dikodekan dengan pewarnaan tertentu: ungu untuk background, biru untuk ginjal, dan kuning untuk tumor. Gambar kemudian diubah ukurannya ke resolusi 256×256 piksel, dinormalisasi ke rentang nilai $[0,1]$, dan augmentasi citra dilakukan secara on-the-fly menggunakan data generator untuk meningkatkan generalisasi model terhadap berbagai variasi data.

4) Model (Pemodelan)

Pada tahap ini dilakukan proses pelatihan model segmentasi citra menggunakan arsitektur U-Net yang dibuat dengan lima jenis backbone dalam *notebook* yang berbeda, yaitu VGG19, MobileNetV2, ResNet50, DenseNet121, dan EfficientNetB7. Setiap model dilatih secara terpisah dan dievaluasi berdasarkan Dice Coefficient khusus untuk kelas tumor. Proses pelatihan dilakukan menggunakan platform Google Colab untuk memanfaatkan GPU yang tersedia dan mempercepat proses komputasi. Pemilihan backbone dilakukan untuk mengeksplorasi kombinasi arsitektur yang paling optimal dalam menangani segmentasi ginjal dan tumor.

5) Assess (Evaluasi Model)

Evaluasi dilakukan terhadap performa model pada data validasi, menggunakan metrik Dice Coefficient per kelas dan Mean Dice Coefficient (MDC) untuk mengukur akurasi segmentasi. Model terbaik dipilih berdasarkan skor tertinggi untuk kelas tumor, yang merupakan fokus utama penelitian ini. Evaluasi ini juga menjadi dasar untuk melakukan tuning lebih lanjut atau merevisi arsitektur apabila diperlukan.

3.3 Teknik Pengumpulan Data

Teknik pengumpulan data dalam penelitian ini dilakukan melalui pemanfaatan dataset publik yang tersedia secara daring, yaitu dataset Kidney Tumor Segmentation Challenge 2019 (KiTS19). Dataset ini dipilih karena memiliki kelengkapan data, bersifat open-source, serta telah digunakan secara luas dalam penelitian terkait segmentasi tumor ginjal. KiTS19 menyediakan data citra medis dalam format volumetrik .nii.gz (Neuroimaging Informatics Technology Initiative), yang merupakan standar format file untuk data pencitraan medis tiga dimensi.

Dataset KiTS19 terdiri dari 210 data pasien penderita kanker ginjal yang telah menjalani tindakan nephrectomy atau pengangkatan ginjal. Data ini mencakup dua komponen utama, yaitu citra CT Scan abdomen tanpa kontras, dan label segmentasi manual yang menunjukkan area ginjal, parenkim ginjal (medulla), serta lesi tumor

yang telah ditandai oleh ahli radiologi profesional. Data ini diunduh secara lengkap dari situs resmi penyelenggara, yaitu <https://kits19.grand-challenge.org>, dengan terlebih dahulu melakukan registrasi akun dan persetujuan penggunaan data untuk keperluan akademis.

Setelah data berhasil diunduh, dilakukan proses verifikasi integritas file untuk memastikan tidak ada kerusakan atau ketidaksesuaian data. Selanjutnya, data dalam format .nii.gz dikonversi ke dalam format .png agar kompatibel dengan pipeline deep learning berbasis TensorFlow dan Keras. Proses konversi ini dilakukan secara manual menggunakan library Python seperti nibabel dan PIL, di mana setiap irisan axial dari citra CT diproses menjadi gambar dua dimensi, serta diberikan pewarnaan spesifik untuk membedakan kelas-kelas segmentasi. Pewarnaan tersebut menggunakan standar visual, yaitu warna ungu untuk latar belakang (background), biru untuk area ginjal, dan kuning untuk area tumor.

Selain itu, dilakukan proses resizing citra ke resolusi 256x256 piksel untuk standarisasi ukuran input ke model, serta normalisasi nilai intensitas piksel ke dalam rentang [0,1] guna meningkatkan stabilitas pelatihan. Proses modifikasi data ini juga mencakup augmentasi citra, seperti rotasi, flipping, dan zoom, yang diterapkan secara on-the-fly saat pelatihan untuk meningkatkan variasi data, memperkaya representasi fitur, dan meningkatkan generalisasi model terhadap berbagai kondisi citra medis. Dengan demikian, seluruh data siap digunakan dalam proses pelatihan model segmentasi citra ginjal dan tumor.

3.4 Teknik Analisis Data

Teknik analisis data dalam penelitian ini dilakukan dengan menerapkan pendekatan deep learning, khususnya arsitektur U-Net yang dimodifikasi dengan berbagai jenis backbone Convolutional Neural Network (CNN). Arsitektur U-Net dipilih karena terbukti efektif dalam tugas segmentasi citra, terutama pada bidang medis, berkat kemampuannya mengekstraksi fitur spasial secara detail melalui struktur encoder-decoder simetris. Penelitian ini mengimplementasikan lima variasi backbone yang berbeda, yaitu EfficientNetB7, VGG19, MobileNetV2, ResNet50,

dan DenseNet121, untuk mengevaluasi pengaruh arsitektur backbone terhadap akurasi segmentasi area ginjal dan tumor.

Seluruh proses analisis dilakukan di platform Google Colab, yang menyediakan lingkungan komputasi berbasis cloud dengan dukungan GPU, sehingga memungkinkan proses pelatihan model dilakukan secara efisien tanpa memerlukan perangkat keras lokal yang mahal atau spesifikasi tinggi. Penggunaan Google Colab juga memberikan fleksibilitas dalam melakukan eksperimen, iterasi kode, serta penyimpanan hasil yang terintegrasi dengan Google Drive.

Tahapan analisis diawali dengan proses preprocessing data, meliputi konversi data dari format .nii.gz ke dalam citra .png, normalisasi nilai intensitas piksel, resizing citra ke resolusi standar, serta augmentasi data untuk meningkatkan keragaman dan generalisasi model. Setiap warna pada citra segmentasi mewakili kelas tertentu, yaitu ungu untuk latar belakang, biru untuk area ginjal (parenkim), dan kuning untuk area tumor. Data yang telah diproses kemudian dibagi menjadi data pelatihan dan data validasi, sehingga model dapat diuji tidak hanya berdasarkan data yang dilatih, tetapi juga kemampuan generalisasi terhadap data yang belum pernah dilihat sebelumnya.

Proses pelatihan dilakukan dengan menerapkan fungsi loss tertimbang (weighted loss), yang memberi bobot lebih besar pada kelas tumor untuk mengatasi permasalahan ketidakseimbangan data antara area latar belakang, ginjal, dan tumor. Hal ini dilakukan karena dalam citra CT Scan, area tumor cenderung memiliki ukuran yang jauh lebih kecil dibandingkan latar belakang, sehingga model tanpa penyesuaian bobot cenderung bias terhadap prediksi latar belakang. Oleh karena itu, digunakan kombinasi fungsi categorical crossentropy dengan tambahan custom loss berbasis Dice Coefficient khusus untuk kelas tumor.

Evaluasi model dilakukan menggunakan metrik Dice Coefficient, yang merupakan ukuran kesesuaian antara prediksi model dengan label ground truth pada tiap kelas segmentasi. Fokus utama penelitian adalah peningkatan Dice Coefficient untuk kelas tumor, mengingat deteksi area tumor memiliki implikasi klinis yang

krusial. Selain evaluasi kuantitatif, dilakukan juga evaluasi visual terhadap hasil segmentasi dengan cara membandingkan citra prediksi dengan label segmentasi asli, sehingga dapat diamati apakah model mampu mendeteksi area ginjal dan tumor secara akurat.

Melalui teknik analisis ini, penelitian bertujuan untuk mengidentifikasi kombinasi arsitektur backbone terbaik dalam meningkatkan akurasi segmentasi tumor ginjal menggunakan U-Net, sekaligus memvalidasi efektivitas pipeline deep learning berbasis cloud yang dapat diterapkan dalam studi segmentasi citra medis lainnya di masa mendatang.