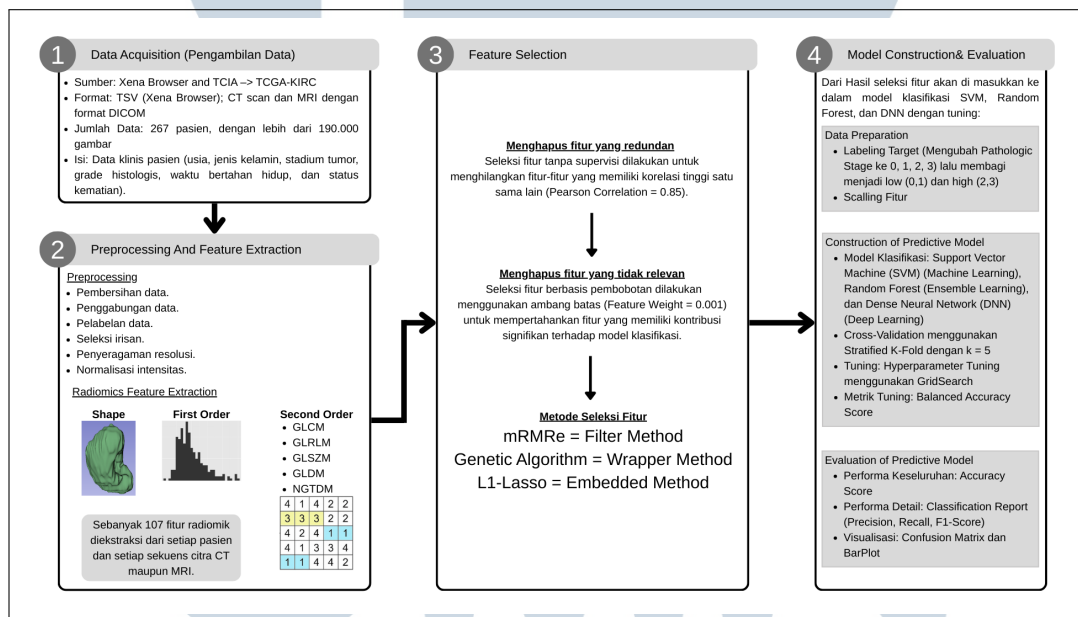


## BAB 3 METODE PENELITIAN

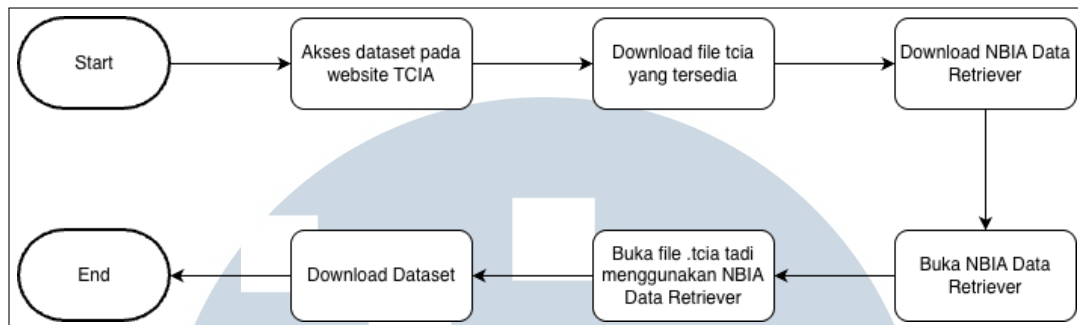
Bab ini menguraikan secara komprehensif desain, strategi, dan prosedur teknis yang diterapkan dalam penelitian untuk mengembangkan sistem klasifikasi stadium kanker ginjal berdasarkan fitur radiomik dari citra CT dan MRI menggunakan dataset TCGA-KIRC [1]. Metodologi penelitian dirancang secara sistematis untuk mengubah citra medis menjadi fitur numerik yang dapat digunakan pada model machine learning. Struktur bab ini mengikuti alur pipeline penelitian, dimulai dari proses akuisisi data, preprocessing, ekstraksi fitur radiomik, seleksi fitur, hingga pembangunan dan evaluasi model. Gambar 3.1 menjelaskan keseluruhan alur penelitian.



Gambar 3.1. Pipeline penelitian radiomik untuk klasifikasi stadium kanker ginjal

### 3.1 Data Acquisition

Penelitian ini menggunakan dataset TCGA-KIRC (Kidney Renal Clear Cell Carcinoma) yang diperoleh dari dua sumber terbuka, yaitu Xena Browser dan The Cancer Imaging Archive (TCIA) [1]. Dua sumber ini menyediakan data klinis dan citra medis yang dibutuhkan untuk proses analisis radiomik. Proses pengambilan data ditunjukkan pada Gambar 3.2.



Gambar 3.2. Proses Pengambilan Data TCGA-KIRC [1]

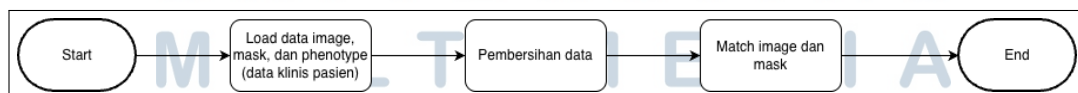
Dataset yang digunakan terdiri dari:

- 267 pasien.
- Lebih dari 190.000 citra CT dan MRI dalam format DICOM.
- Data klinis dalam format TSV berisi usia, jenis kelamin, stadium patologis, grade histologis, status hidup/kematian, serta waktu kelangsungan hidup.

Data klinis dari Xena Browser digabungkan dengan data citra dari TCIA menggunakan Patient ID yang sama sehingga menghasilkan dataset yang lengkap dan dapat dianalisis. Dataset TCGA-KIRC merupakan dataset multi-institusi sehingga memiliki variasi kualitas citra, perbedaan mesin CT/MRI, dan protokol pemindaian yang berbeda, yang menjadi tantangan umum pada penelitian radiomik.

### 3.2 Pre-processing dan Ekstraksi Fitur

Setelah proses pengumpulan data selesai, tahap berikutnya adalah preprocessing untuk menyelaraskan data sebelum diekstraksi fitur radiomik. Tahap ini mencakup proses pembersihan data klinis, normalisasi citra, segmentasi tumor, dan ekstraksi fitur. Gambar 3.3 menunjukkan tahapan preprocessing.



Gambar 3.3. Tahap Preprocessing Data CT dan MRI

#### 3.2.1 Pre-processing Data Klinis

Tahap preprocessing data klinis meliputi:

- Penyelarasan ID antara data klinis dan citra.
- Penggabungan variabel klinis penting seperti stadium.
- Penanganan missing value.
- Seleksi pasien yang memiliki citra lengkap dan label stadium yang valid.

### 3.2.2 Pre-processing Citra Medis

Citra CT dan MRI melalui beberapa tahapan prapemrosesan:

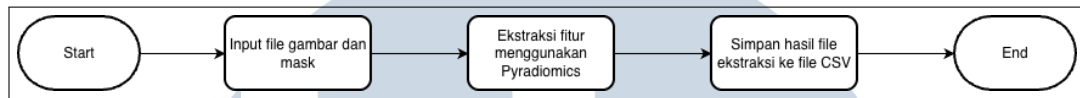
- Seleksi irisan citra yang memuat area tumor.
- Normalisasi intensitas voxel agar konsisten antar-pasien.
- Resampling voxel ke resolusi seragam untuk mengurangi perbedaan perangkat.
- Penghapusan noise dan artefak.

### 3.2.3 Segmentasi Tumor

Segmentasi dilakukan untuk memisahkan region of interest (ROI) atau area tumor dari jaringan sehat di sekitarnya. Proses segmentasi ini dilakukan oleh radiologist berpengalaman, yang memiliki kompetensi klinis dalam mengidentifikasi batas tumor secara akurat pada citra medis. Segmentasi dilakukan menggunakan mask atau annotation tumor yang tersedia pada dataset, baik melalui metode manual maupun semi-otomatis. Dalam penelitian ini, proses segmentasi semi-otomatis dilakukan menggunakan perangkat lunak *3D Slicer* [40], yang menyediakan platform komputasi citra medis yang mendukung visualisasi, anotasi, dan ekstraksi fitur secara efisien. Dengan adanya segmentasi yang dilakukan oleh tenaga ahli menggunakan *3D Slicer*, area tumor yang dianalisis diharapkan merepresentasikan kondisi patologis yang sebenarnya, sehingga dapat meningkatkan keandalan proses ekstraksi fitur radiomik serta akurasi hasil analisis dan pemodelan klasifikasi pada tahap selanjutnya.

### 3.2.4 Ekstraksi Fitur Radiomics

Ekstraksi fitur radiomik digambarkan pada Gambar 3.4.



Gambar 3.4. Proses Ekstraksi Fitur Radiomik

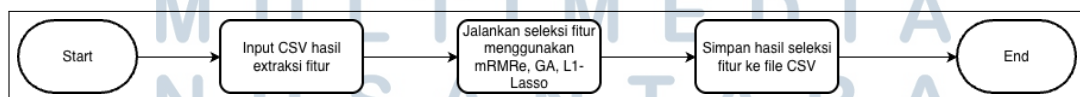
Sebanyak 107 fitur radiomik diekstraksi menggunakan PyRadiomics, meliputi:

- Shape Features
- First-Order Statistics
- Gray Level Co-occurrence Matrix (GLCM)
- Gray Level Run Length Matrix (GLRLM)
- Gray Level Size Zone Matrix (GLSZM)
- Gray Level Dependence Matrix (GLDM)
- Neighborhood Gray Tone Difference Matrix (NGTDM)

Output dari proses ini berupa file CSV berisi seluruh fitur numerik yang siap digunakan untuk tahap seleksi fitur.

### 3.3 Seleksi Fitur

Seleksi fitur bertujuan untuk mengurangi kompleksitas model dengan menghilangkan fitur yang tidak relevan atau redundan. Proses seleksi fitur digambarkan pada Gambar 3.5.



Gambar 3.5. Proses Seleksi Fitur

### 3.3.1 Penghapusan Fitur Redundan

Fitur dengan nilai korelasi tinggi ( $\text{Pearson} > 0.85$ ) dihapus untuk menghindari multikolinearitas yang dapat memengaruhi performa model.

### 3.3.2 Penghapusan Fitur Tidak Relevan

Fitur dengan nilai kontribusi di bawah 0.001 dikeluarkan menggunakan metode berbasis feature importance untuk memastikan hanya fitur signifikan yang disertakan.

### 3.3.3 Metode Seleksi Fitur

Tiga pendekatan seleksi fitur digunakan:

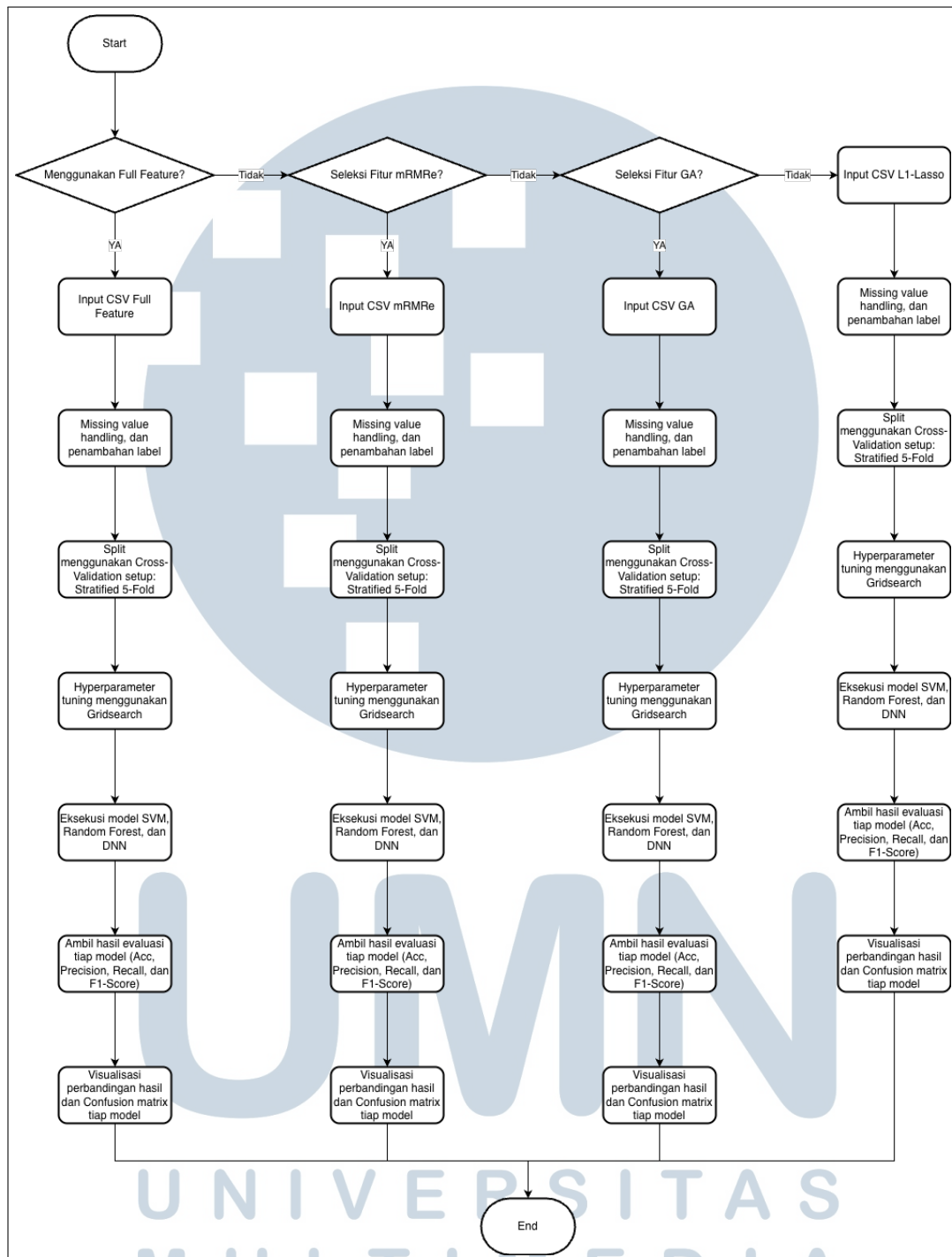
- **mRMRe** (Filter Method)
- **Genetic Algorithm** (Wrapper Method)
- **L1-Lasso** (Embedded Method)

Ketiga metode tersebut menghasilkan subset fitur yang berbeda untuk mengevaluasi konsistensi dan performanya terhadap model.

## 3.4 Pembangunan dan Evaluasi Model

Tahap ini mencakup proses pembangunan dan evaluasi model klasifikasi menggunakan fitur radiomik terpilih untuk memprediksi karakteristik kanker ginjal. Alur proses pembangunan model ditunjukkan pada Gambar 3.6.

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 3.6. Flowchart Pembangunan dan Evaluasi Model

### 3.4.1 Persiapan Data (Data Preparation)

Tahap persiapan data meliputi langkah-langkah berikut:

- **Labeling target:** Mengubah *pathologic stage* menjadi angka 0, 1, 2, 3, kemudian mengelompokkan menjadi kategori low (0,1) dan high (2,3).
- **Skalasi fitur:** Melakukan normalisasi atau standardisasi fitur radiomik untuk memastikan skala yang seragam pada setiap variabel.

### 3.4.2 Pembangunan Model

Model klasifikasi dibangun menggunakan fitur radiomik hasil ekstraksi, baik tanpa seleksi fitur (*full features*) maupun dengan fitur hasil seleksi, untuk mengevaluasi pengaruh reduksi dimensi terhadap performa model. Pembangunan model mencakup tiga pendekatan utama, yaitu *Support Vector Machine* (SVM) sebagai algoritma *machine learning*, *Random Forest* (RF) sebagai algoritma *ensemble learning*, dan *Deep Neural Network* (DNN) sebagai algoritma *deep learning*.

Tahapan pembangunan model meliputi:

- **Hyperparameter tuning:** dilakukan menggunakan GridSearchCV untuk mencari kombinasi parameter optimal pada masing-masing algoritma, termasuk  $C$ ,  $\gamma$ , dan kernel RBF pada SVM, jumlah pohon dan kedalaman maksimum pada RF, serta jumlah lapisan dan neuron pada DNN.
- **Validasi model:** menggunakan Stratified 5-Fold Cross Validation untuk menjaga proporsi kelas yang seimbang pada setiap fold dan memastikan evaluasi performa yang konsisten.
- **Evaluasi performa:** metrik yang digunakan mencakup Balanced Accuracy Score, yang efektif untuk menangani ketidakseimbangan kelas dan memberikan penilaian adil terhadap setiap model.

Pendekatan ini memungkinkan perbandingan performa yang objektif antara SVM, RF, dan DNN dalam klasifikasi ccRCC menggunakan fitur radiomik.

### 3.4.3 Evaluasi Model

Model dievaluasi berdasarkan metrik:

- Accuracy

- Balanced Accuracy
- Precision
- Recall
- F1-Score
- Confusion Matrix

Evaluasi dilakukan pada fitur sebelum seleksi dan setelah seleksi untuk menganalisis dampak seleksi fitur terhadap performa model.

