

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Karsinoma sel ginjal (*Renal Cell Carcinoma* atau RCC) masih menjadi salah satu tantangan utama dalam penanganan kanker urologi secara global. Subtipe yang paling umum adalah *clear cell renal cell carcinoma* (ccRCC), yang menyumbang sekitar 80% dari seluruh kasus RCC di dunia [1]. Tingginya prevalensi ccRCC menjadikan proses identifikasi yang akurat sangat penting dalam menentukan strategi pengobatan yang tepat serta pengelolaan klinis pasien. Diagnosis yang andal juga dapat membantu meningkatkan akurasi prediksi luaran klinis sekaligus mengurangi risiko *overtreatment* maupun tindak lanjut yang kurang memadai [2,3].

Meskipun pencitraan medis seperti *Computed Tomography* (CT) dan *Magnetic Resonance Imaging* (MRI) merupakan metode utama dalam proses diagnosis, interpretasi citra tersebut masih sangat bergantung pada penilaian manusia. Ketergantungan ini dapat menimbulkan variasi interpretasi antar pengamat (*inter-observer variability*), sehingga hasil penilaian antara satu dokter dengan dokter lainnya dapat berbeda. Walaupun teknologi pencitraan medis memiliki peran penting dalam deteksi dini tumor, citra konvensional sering kali belum mampu menangkap seluruh karakteristik biologis tumor secara menyeluruh. Variasi karakteristik tumor yang berkaitan dengan pertumbuhan, agresivitas, serta respons terhadap terapi sering kali hanya terlihat sebagian melalui pengamatan visual [4].

Dalam beberapa tahun terakhir, pendekatan radiomik (*radiomics*) berkembang sebagai metode yang mampu mengekstraksi informasi kuantitatif dari citra medis. Radiomik mengubah citra medis menjadi kumpulan fitur numerik yang merepresentasikan berbagai karakteristik tumor, seperti tekstur, bentuk, intensitas, dan heterogenitas jaringan [5-7]. Perangkat lunak seperti PyRadiomics dan 3D Slicer telah banyak digunakan dalam penelitian radiomik karena memungkinkan proses ekstraksi fitur secara sistematis dan terstandarisasi dari citra kanker [7,8]. Namun demikian, fitur radiomik yang dihasilkan umumnya memiliki dimensi yang sangat tinggi dan sering kali mengandung informasi yang redundan atau kurang relevan. Kondisi ini dapat menurunkan stabilitas dan performa model pembelajaran mesin apabila tidak dilakukan proses seleksi fitur yang tepat.

Untuk mengatasi permasalahan dimensi tinggi tersebut, teknik *feature selection* sering diterapkan untuk memilih fitur yang paling informatif sekaligus mengurangi redundansi. Metode seperti mRMRe (*Minimum Redundancy Maximum Relevance*) banyak digunakan karena mampu memilih fitur yang memiliki relevansi tinggi terhadap variabel target sekaligus meminimalkan redundansi antar fitur. Selain itu, pendekatan berbasis optimasi seperti *Genetic Algorithm* (GA) dapat mengeksplorasi berbagai kombinasi fitur melalui mekanisme evolusi untuk menemukan subset fitur yang memberikan performa prediksi terbaik. Metode lain seperti LASSO (*Least Absolute Shrinkage and Selection Operator*) melakukan seleksi fitur secara terintegrasi dengan proses pelatihan model melalui regularisasi L1 yang mendorong koefisien fitur kurang informatif menuju nilai nol, sehingga menghasilkan model yang lebih sederhana dan mudah diinterpretasikan [9-11].

Dalam proses analisis citra medis, metode pembelajaran mesin (*machine learning*) memiliki kemampuan untuk menemukan pola kompleks yang sulit dikenali melalui pengamatan visual manusia. Algoritma seperti *Support Vector Machine* (SVM) dan XGBoost mampu memodelkan hubungan kompleks antar fitur radiomik untuk meningkatkan akurasi klasifikasi. SVM dikenal efektif untuk dataset berdimensi tinggi seperti radiomik karena mampu membangun hiperbidang pemisah optimal meskipun jumlah fitur jauh lebih besar dibandingkan jumlah sampel [12-14]. Di sisi lain, metode *gradient boosting* seperti XGBoost memiliki kemampuan yang baik dalam menangkap hubungan nonlinier serta interaksi antar fitur melalui proses pembelajaran berbasis pohon keputusan yang dilakukan secara iteratif. Selain itu, *Deep Neural Network* (DNN) mampu mempelajari representasi nonlinier yang bersifat hierarkis dari data, sehingga sangat cocok digunakan dalam permasalahan klasifikasi biomedis yang kompleks.

Dalam perkembangan terbaru, pendekatan *stacking ensemble* semakin banyak digunakan karena mampu menggabungkan prediksi dari beberapa model dasar (*base learners*). Pendekatan ini sering menghasilkan performa yang lebih stabil dan akurat dibandingkan penggunaan satu model klasifikasi saja, karena memanfaatkan keunggulan komplementer dari berbagai algoritma [15,16]. Dalam arsitektur *stacking*, model dasar berfungsi untuk menangkap berbagai perspektif pola pada data, sementara *meta-learner* bertugas menggabungkan hasil prediksi dari model-model tersebut untuk menghasilkan keputusan akhir. *Deep Neural Network* sangat cocok digunakan sebagai *meta-learner* karena mampu memodelkan hubungan nonlinier yang kompleks dari probabilitas prediksi yang dihasilkan oleh model dasar.

Berdasarkan perkembangan tersebut, penelitian ini mengusulkan pendekatan klasifikasi ccRCC berbasis radiomik dan *stacking ensemble*. Proses penelitian diawali dengan segmentasi tumor secara manual untuk menentukan wilayah tumor pada citra medis. Selanjutnya dilakukan ekstraksi fitur radiomik untuk memperoleh berbagai karakteristik kuantitatif tumor. Fitur-fitur yang diperoleh kemudian melalui proses seleksi menggunakan beberapa metode untuk mendapatkan subset fitur yang paling informatif. Subset fitur tersebut selanjutnya digunakan sebagai masukan dalam sistem pembelajaran berlapis (*stacking*), di mana model SVM dan XGBoost berperan sebagai *base learners*, sedangkan *Deep Neural Network* digunakan sebagai *meta-learner* untuk menggabungkan probabilitas prediksi dari kedua model tersebut. Pendekatan ini diharapkan dapat memberikan metode klasifikasi ccRCC yang lebih akurat, objektif, dan bersifat non-invasif sehingga dapat mendukung pengambilan keputusan klinis secara lebih baik.

## 1.2 Rumusan Masalah

Berdasarkan judul penelitian ini, permasalahan yang ingin dipecahkan adalah:

1. Bagaimana membangun model klasifikasi berbasis radiomik menggunakan pendekatan *meta-learning* untuk mengklasifikasikan *clear cell renal cell carcinoma* (ccRCC) secara akurat berdasarkan fitur radiomik dari citra *Computed Tomography* (CT) dan *Magnetic Resonance Imaging* (MRI)?
2. Bagaimana kombinasi model *base learner* seperti *Support Vector Machine* (SVM) dan XGBoost, serta *meta-learner Deep Neural Network* (DNN) dalam kerangka *stacking ensemble* dapat meningkatkan performa klasifikasi?
3. Sejauh mana berbagai metode seleksi fitur (*filter*, *wrapper*, dan *embedded*) dapat meningkatkan performa dan stabilitas model klasifikasi dibandingkan dengan penggunaan seluruh fitur radiomik?

## 1.3 Batasan Permasalahan

Batasan permasalahan dalam penelitian ini bertujuan untuk memperjelas ruang lingkup serta menjaga fokus penelitian agar tetap sesuai dengan tujuan yang

telah ditetapkan. Adapun batasan-batasan dalam penelitian ini adalah sebagai berikut:

1. Dataset yang digunakan berasal dari *The Cancer Imaging Archive* (TCIA), khususnya koleksi *TCGA-KIRC*, dengan jumlah data yang terbatas pada 42 pasien yang memiliki anotasi segmentasi tumor oleh radiolog.
2. Data citra yang digunakan hanya terbatas pada *Computed Tomography* (CT) dan *Magnetic Resonance Imaging* (MRI) yang telah memiliki anotasi segmentasi tumor oleh radiolog.
3. Klasifikasi yang dilakukan dibatasi pada dua kelas, yaitu *low stage* (stadium I–II) dan *high stage* (stadium III–IV).
4. Fitur yang digunakan dalam penelitian ini adalah fitur radiomik yang diekstraksi menggunakan PyRadiomics, meliputi fitur *first-order*, *shape*, serta fitur tekstur dan *high-order*.
5. Metode seleksi fitur yang digunakan dibatasi pada tiga pendekatan, yaitu *Minimum Redundancy Maximum Relevance* (mRMR), *Genetic Algorithm* (GA), dan L1-LASSO.
6. Model klasifikasi yang digunakan adalah pendekatan *stacking ensemble* dalam kerangka *meta-learning*, dengan *Support Vector Machine* (SVM) dan XGBoost sebagai *base learners*, serta *Deep Neural Network* (DNN) sebagai *meta-learner*.
7. Evaluasi model dilakukan menggunakan *stratified 5-fold cross-validation* dengan metrik *ROC-AUC*, akurasi, presisi, *recall*, dan F1-score.
8. Penelitian ini tidak membahas aspek implementasi klinis secara langsung serta tidak melakukan validasi eksternal pada dataset lain.

#### 1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Membangun model klasifikasi berbasis radiomik menggunakan pendekatan *meta-learning* untuk mengklasifikasikan *clear cell renal cell carcinoma* (ccRCC) berdasarkan citra *Computed Tomography* (CT) dan *Magnetic Resonance Imaging* (MRI).

2. Mengimplementasikan metode *stacking ensemble* dengan *Support Vector Machine* (SVM) dan XGBoost sebagai *base learners*, serta *Deep Neural Network* (DNN) sebagai *meta-learner*, guna meningkatkan performa klasifikasi.
3. Menganalisis pengaruh berbagai metode seleksi fitur, yaitu *Minimum Redundancy Maximum Relevance* (mRMR), *Genetic Algorithm* (GA), dan L1-LASSO, terhadap performa dan stabilitas model klasifikasi.

### 1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Manfaat Teoritis: Memberikan kontribusi dalam bidang radiomik dan *machine learning*, khususnya dalam penerapan pendekatan *meta-learning* untuk klasifikasi *clear cell renal cell carcinoma* (ccRCC) berbasis citra medis.
2. Manfaat Metodologis: Memberikan gambaran mengenai pengaruh metode seleksi fitur seperti mRMR, *Genetic Algorithm*, dan L1-LASSO terhadap performa model klasifikasi dalam kerangka *stacking ensemble*.
3. Manfaat Jangka Panjang: Menjadi dasar pengembangan sistem *artificial intelligence* yang lebih *robust* dan *generalizable* dalam bidang diagnosis kanker berbasis radiomik.

### 1.6 Sistematika Penulisan

Sistematika penulisan laporan adalah sebagai berikut:

1. Bab 1 PENDAHULUAN  
Menjelaskan latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, serta sistematika penulisan.
2. Bab 2 LANDASAN TEORI  
Menguraikan landasan teori yang mendukung penelitian, meliputi konsep radiomik, *machine learning*, *meta-learning*, serta metode-metode yang digunakan dalam penelitian.

3. Bab 3 METODOLOGI PENELITIAN

Menjelaskan metodologi penelitian yang digunakan, termasuk tahapan pengumpulan data, *pre-processing*, segmentasi tumor, ekstraksi fitur, seleksi fitur, serta pembangunan model klasifikasi.

4. Bab 4 HASIL DAN DISKUSI

Menyajikan hasil penelitian dan pembahasan, meliputi evaluasi performa model pada berbagai skenario seleksi fitur serta analisis terhadap hasil yang diperoleh.

5. Bab 5 SIMPULAN DAN SARAN

Berisi simpulan dari hasil penelitian yang telah dilakukan serta saran untuk pengembangan penelitian selanjutnya.

