

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Kanker lambung atau *gastric cancer* merupakan salah satu jenis kanker dengan tingkat mortalitas yang tinggi di dunia. Berdasarkan data *Global Cancer Observatory (GLOBOCAN)*, kanker lambung masih termasuk dalam lima besar penyebab kematian akibat kanker secara global, terutama di negara berkembang [1]. Tingginya angka kematian ini menunjukkan bahwa kanker lambung masih menjadi permasalahan kesehatan yang serius dan membutuhkan upaya penanganan yang lebih efektif, khususnya pada tahap diagnosis dan penentuan stadium penyakit.

Salah satu faktor utama yang berkontribusi terhadap tingginya mortalitas kanker lambung adalah keterlambatan diagnosis serta ketidakakuratan dalam penentuan stadium penyakit pada fase awal. Stadium kanker memiliki peranan krusial dalam menentukan strategi terapi, seperti pemilihan tindakan bedah, kemoterapi, maupun terapi target, serta dalam memprediksi prognosis pasien. Kesalahan dalam klasifikasi stadium berpotensi menyebabkan pemilihan terapi yang kurang tepat, yang pada akhirnya dapat menurunkan peluang keberhasilan pengobatan.

Secara klinis, penentuan stadium kanker lambung umumnya dilakukan menggunakan sistem TNM (*Tumor, Node, Metastasis*) yang mengevaluasi ukuran tumor primer, keterlibatan kelenjar getah bening, dan keberadaan metastasis jauh [2]. Meskipun sistem ini telah menjadi standar internasional, pendekatan tersebut memiliki beberapa keterbatasan, antara lain bersifat invasif, memerlukan prosedur diagnostik yang kompleks dan berbiaya tinggi, serta sangat bergantung pada interpretasi subjektif tenaga medis. Variasi penilaian antar pengamat (*inter-observer variability*) juga dapat menimbulkan ketidakkonsistenan dalam penentuan stadium, terutama pada kasus dengan karakteristik klinis yang ambigu.

Seiring dengan perkembangan teknologi biologi molekuler dan *high-throughput sequencing*, analisis data ekspresi gen memungkinkan pemahaman yang lebih mendalam terhadap perubahan molekuler yang mendasari progresi kanker. Data ekspresi gen mencerminkan aktivitas biologis sel dan telah terbukti berkorelasi dengan stadium, agresivitas, serta prognosis berbagai jenis kanker, termasuk kanker lambung [3, ?]. Oleh karena itu, pemanfaatan data molekuler berpotensi menjadi

pendekatan alternatif yang lebih objektif dalam membantu penentuan stadium kanker.

Namun, data ekspresi gen memiliki karakteristik berdimensi tinggi, di mana jumlah fitur gen jauh melebihi jumlah sampel. Kondisi ini menimbulkan permasalahan analisis seperti meningkatnya risiko *overfitting*, beban komputasi yang tinggi, serta kesulitan dalam mengidentifikasi gen-gen yang benar-benar relevan [4]. Tanpa metode analisis yang tepat, informasi penting yang terkandung dalam data ekspresi gen tidak dapat dimanfaatkan secara optimal.

Pendekatan *machine learning* menawarkan solusi untuk mengatasi permasalahan tersebut dengan kemampuan dalam menangani data berdimensi tinggi, mempelajari pola kompleks, serta menghasilkan model prediktif yang akurat. Dengan mengombinasikan teknik seleksi fitur dan algoritma klasifikasi, *machine learning* dapat digunakan untuk mengklasifikasikan stadium kanker lambung berdasarkan profil ekspresi gen secara lebih objektif dan konsisten [5]. Pendekatan ini juga mendukung paradigma *precision medicine*, yang menekankan penggunaan data molekuler untuk pengambilan keputusan klinis yang lebih personal dan berbasis data [6].

Berdasarkan permasalahan tersebut, penelitian ini difokuskan pada pengembangan model klasifikasi stadium kanker lambung ke dalam kategori stadium awal dan stadium lanjut menggunakan data ekspresi gen dengan pendekatan *machine learning*. Klasifikasi ini penting untuk membantu membedakan pasien dengan karakteristik progresi penyakit yang berbeda, sehingga dapat mendukung proses penentuan terapi dan prognosis secara lebih akurat. Selain itu, penelitian ini diharapkan mampu mengidentifikasi gen-gen kunci yang berpotensi digunakan sebagai biomarker molekuler dalam mendukung penentuan stadium kanker lambung.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana melakukan praproses dan seleksi fitur pada data ekspresi gen agar dapat merepresentasikan karakteristik biologis kanker lambung secara optimal?
2. Bagaimana membangun model *machine learning* untuk mengklasifikasikan

stadium kanker lambung menjadi *early stage* dan *late stage* berdasarkan data ekspresi gen?

3. Bagaimana kinerja model yang dikembangkan berdasarkan metrik evaluasi seperti ROC-AUC, akurasi, dan F1-score?
4. Sejauh mana model yang dihasilkan berpotensi digunakan sebagai sistem pendukung keputusan klinis berbasis *biomarker molekuler*?

1.3 Batasan Penelitian

Untuk menjaga fokus dan ruang lingkup penelitian, batasan penelitian yang diterapkan adalah sebagai berikut:

1. Data yang digunakan berupa data ekspresi gen pasien kanker lambung.
2. Klasifikasi dilakukan secara biner, yaitu *early stage* (stadium I dan II) dan *late stage* (stadium III dan IV).
3. Metode yang digunakan terbatas pada algoritma *machine learning* konvensional.
4. Evaluasi model dilakukan menggunakan metrik ROC-AUC, akurasi, dan F1-score.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Melakukan praproses dan seleksi fitur pada data ekspresi gen pasien kanker lambung.
2. Mengembangkan model *machine learning* untuk klasifikasi stadium kanker lambung menjadi *early stage* dan *late stage*.
3. Mengevaluasi kinerja model menggunakan metrik evaluasi klasifikasi yang relevan.
4. Mengidentifikasi gen-gen penting yang berkontribusi dalam klasifikasi stadium kanker lambung.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Manfaat Akademik

Menambah literatur ilmiah terkait penerapan *machine learning* dalam analisis data ekspresi gen untuk klasifikasi stadium kanker.

2. Manfaat Praktis

Memberikan dasar pengembangan sistem pendukung keputusan klinis berbasis *biomarker molekuler*.

3. Manfaat Teknologis

Mendukung pengembangan pendekatan *precision medicine* berbasis data molekuler di bidang onkologi.

1.6 Sistematika Penulisan

Sistematika penulisan laporan ini adalah sebagai berikut:

- Bab 1 PENDAHULUAN

Membahas latar belakang penelitian, rumusan masalah, tujuan, batasan, manfaat, serta sistematika penulisan.

- Bab 2 TINJAUAN PUSTAKA

Menguraikan tinjauan pustaka dan landasan teori yang berkaitan dengan kanker lambung, data ekspresi gen, seleksi fitur, dan *machine learning*.

- Bab 3 METODE PENELITIAN

Menjelaskan metodologi penelitian, meliputi pengumpulan data, praproses data, seleksi fitur, pembangunan model, dan evaluasi.

- Bab 4 HASIL DAN PEMBAHASAN

Menyajikan hasil eksperimen, evaluasi performa model, serta pembahasan hasil penelitian.

- Bab 5 SIMPULAN SARAN

Berisi kesimpulan dan saran untuk pengembangan penelitian selanjutnya.