

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Menurut data GLOBOCAN tahun 2022, kanker prostat merupakan jenis kanker terbanyak kedua dan menjadi penyebab kematian akibat kanker tertinggi kelima pada pasien laki-laki. Secara global, diperkirakan terdapat sekitar 1,5 juta kasus baru kanker prostat dengan angka kematian mencapai 375 ribu orang [1]. Kanker prostat merupakan salah satu jenis kanker yang terjadi pada pria, ditandai dengan pertumbuhan sel abnormal di kelenjar prostat. Risiko kanker prostat diketahui bervariasi antar kelompok ras [2]. Pria keturunan Afrika-Amerika cenderung memiliki risiko lebih tinggi untuk mengalami kanker prostat, serta memiliki kemungkinan lebih besar untuk terdiagnosis pada stadium lanjut dan memiliki angka kematian yang lebih tinggi dibandingkan ras lainnya [2]. Perbedaan ini mengindikasikan adanya faktor biologis yang dapat mempengaruhi perkembangan dan agresivitas kanker prostat di berbagai ras [1, 2].

Kelenjar prostat sendiri merupakan kelenjar eksokrin dalam sistem reproduksi pria yang berperan dalam menghasilkan sebagian cairan semen yang mendukung kelangsungan hidup sperma. Pada stadium awal (stadium I dan II), kanker prostat berkembang secara lokal di dalam kelenjar prostat. Namun, seiring dengan perkembangan penyakit, sel kanker dapat menyebar ke luar kapsul prostat, terutama ke tulang dan kelenjar getah bening (stadium III dan IV). Penyebaran ini dapat menyebabkan komplikasi serius dan menurunkan kualitas hidup pasien [3]. Diagnosa kanker prostat merupakan faktor krusial dalam meningkatkan peluang kesembuhan dan mengurangi angka kematian. Oleh karena itu, diagnosis yang tepat dan cepat menjadi kunci utama untuk menanggulangi penyakit ini.

Diagnosis kanker prostat dapat dilakukan melalui berbagai rangkaian pemeriksaan, termasuk tes *Prostate-Specific Antigen (PSA)*, *Digital Rectal Exam (DRE)*, serta biopsi untuk mengonfirmasi keberadaan sel kanker [3]. Namun, setiap metode memiliki keterbatasannya masing-masing, baik dari segi sensitivitas, spesifisitas, maupun tingkat invasivitas. Diagnosis dengan menggunakan PSA dilakukan dengan cara mengukur kadar PSA di dalam darah, tetapi metode ini memiliki tingkat akurasi yang cukup rendah. Diagnosis PSA dapat menyebabkan terjadinya *False Positive*. Hal ini dapat terjadi pada lansia dan pasien *benign*

prostatic hyperplasia (BPH) yang juga dapat menyebabkan lonjakan pada kadar PSA [4, 5].

Pemeriksaan *Digital Rectal Exam* (DRE) dilakukan dengan cara memasukkan sebuah jari ke bagian dubur pasien untuk meraba bagian belakang kelenjar prostat. Pemeriksaan ini dilakukan untuk menilai keberadaan kelainan fisik pada prostat yang mencurigakan. Metode ini memiliki keterbatasan dalam ketepatan dan kemampuan mendeteksi kanker prostat. Metode ini kurang efektif dalam mendeteksi kanker pada tahap awal, memiliki keterbatasan dalam menganalisis area tertentu pada prostat, serta sangat bergantung pada keahlian pemeriksa dalam menginterpretasikan hasil pemeriksaan [6].

Biopsi dilakukan dengan bantuan *Magnetic Resonance Imaging* (MRI) untuk mengambil sampel jaringan dari prostat untuk dianalisis guna mendeteksi keberadaan sel kanker. Namun, metode ini memiliki keterbatasan, seperti kemungkinan terjadinya *sampling error*, di mana sampel diambil dari area yang tidak mengandung sel kanker, sehingga dapat menghasilkan hasil *false negative*. Selain itu, tingkat akurasi biopsi juga bergantung pada keahlian dokter dalam menentukan lokasi pengambilan sampel dan menganalisis hasil pemeriksaan [7].

Seiring dengan munculnya teknologi *Next-generation sequencing* (NGS) dalam beberapa tahun terakhir, waktu dan biaya sekuensing genom serta transkriptom telah berkurang secara signifikan. Teknologi ini memungkinkan peneliti mendeteksi perubahan ekspresi gen, variasi jumlah salinan, variasi genomik, modifikasi pasca-transkripsi, dan lain sebagainya [8]. Kemajuan dalam teknologi NGS telah memberikan kontribusi signifikan dalam penelitian yang berkaitan dengan diagnosis kanker, penentuan prognosis, serta pengembangan metode terapeutik. NGS memungkinkan analisis molekuler secara menyeluruh dan mendalam, memberikan gambaran yang lebih detail mengenai kondisi biologis pasien. Dengan pendekatan ini, dapat diidentifikasi pola-pola spesifik yang berkaitan dengan stadium awal maupun lanjut kanker prostat. Dalam penerapannya, teknologi NGS dapat menunjukkan keberadaan mutasi biomarker genetik pada pasien. Biomarker tersebut dapat digunakan untuk menentukan perkembangan kanker [9]. Dengan demikian, dibandingkan dengan metode diagnosis kanker prostat yang konvensional, model diskriminasi stadium berbasis ekspresi gen dan miRNA memiliki potensi besar untuk meningkatkan akurasi penentuan stadium penyakit dan mendukung pemilihan terapi yang lebih tepat [10].

Beberapa penelitian terdahulu menggunakan pendekatan model berbasis *machine learning* yang disertai dengan berbagai metode *Feature Selection* (FS)

untuk menentukan biomarker yang digunakan. Penelitian-penelitian tersebut menggunakan data ekspresi *gen* dan *miRNA* dari hasil teknologi NGS. Salah satu penelitian oleh Kabir et al. [11] menggunakan 550 sampel data *gene expression RNA-seq* dari GDC untuk klasifikasi jaringan tumor dan normal. Metode seleksi fitur yang digunakan dalam penelitian tersebut adalah *Multivariate Correlation*, *Extra-Trees*, dan *Chi-Squared*. Metode seleksi fitur tersebut menghasilkan 12 *gen* utama yang akan digunakan dalam klasifikasi. Algoritma klasifikasi yang digunakan pada penelitian tersebut adalah *Random Forest* (RF), *Decision Tree* (DT), serta *Multilayered Neural Network* (MLP) yang masing-masing menghasilkan akurasi sebesar 93,33%, 90,9% serta 93,93%.

Penelitian oleh Casey et al. [12] menggunakan 546 sampel data *gene expression RNA-seq* dari *The Cancer Genome Atlas (TCGA)* untuk klasifikasi jaringan tumor dan normal. Metode seleksi fitur yang digunakan dalam penelitian tersebut adalah *Mutual Information* (MI) dan *Recursive Feature Elimination* (RFE) yang menghasilkan 20 fitur. Algoritma klasifikasi yang digunakan pada penelitian tersebut adalah *Logistic Regression* yang menghasilkan akurasi 96,1%, presisi 92,6%, recall 94,1%, dan F1-score 91,2%.

Penelitian oleh Singh et al. [13] menggunakan 86 sampel miRNA yang didapatkan dari 86 relawan. *Centre for Genetic Disorders and the Department of Urology at the Institute of Medical Sciences, Banaras Hindu University* bekerja sama dengan para peneliti untuk mendapatkan sampel tersebut. Penelitian ini bertujuan untuk membedakan antara *benign prostatic hyperplasia* (BPH) dan kanker prostat. Seleksi fitur dilakukan menggunakan tes *Mann-Whitney U* yang menghasilkan 4 miRNA. Algoritma klasifikasi yang digunakan pada penelitian tersebut adalah *random forest* yang menghasilkan akurasi 77,42%, presisi 86,21%, recall 71,43%, dan F1-score 78,12% .

Penelitian-penelitian yang telah dilakukan tersebut menjadi dasar untuk penelitian ini. Berbeda dengan berbagai penelitian sebelumnya, fokus klasifikasi pada penelitian ini diarahkan untuk membedakan antara stadium II dan stadium III kanker prostat. Stadium-stadium ini memiliki perbedaan klinis yang signifikan, termasuk dalam hal ukuran tumor, proses penyebaran kanker, serta kompleksitas terapi yang dibutuhkan [14]. Berbeda dengan stadium I yang umumnya memiliki prognosis yang sangat baik dan dapat ditangani dengan intervensi minimal, serta stadium IV yang ditandai oleh metastasis dan karakteristik penyakit yang lebih heterogen, stadium II dan III menawarkan konteks klinis yang lebih menantang namun masih dapat diintervensi secara efektif. Penelitian akan menggunakan

dataset dari *The Cancer Genome Atlas Prostate Adenocarcinoma* (TCGA-PRAD). Dataset ini mencakup informasi *gene expression RNA-Seq* dan *miRNA expression* yang telah banyak digunakan dalam berbagai studi serta penelitian bioinformatika dan kanker. Untuk meningkatkan akurasi model klasifikasi, diterapkan metode seleksi fitur berbasis statistik seperti *Lasso* dan RFE, serta metode bioinformatika *Differentially Expressed Genes* (DEG) menggunakan pendekatan *limma*. Model klasifikasi yang digunakan akan dikembangkan menggunakan algoritma *Logistic Rgression* dan *Random Forest*, yang telah terbukti efektif dalam studi terdahulu. Penekanan pada stadium II dan III dipilih karena keduanya menunjukkan perbedaan karakteristik klinis yang penting dalam konteks perkembangan kanker prostat. Penekanan pada pasien berkulit putih dilakukan untuk meminimalkan potensi bias genetik dan meningkatkan homogenitas data. Penelitian ini diharapkan dapat membantu dalam mengidentifikasi biomarker potensial yang tidak hanya mendukung proses klasifikasi stadium kanker prostat, tetapi juga membuka peluang untuk pengembangan metode diagnostik yang lebih akurat dan terpersonalisasi.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, berikut merupakan rumusan masalah pada penelitian ini:

1. Bagaimana membedakan antara stadium II dan stadium III kanker prostat menggunakan data ekspresi gen dan miRNA dari dataset TCGA-PRAD?
2. Metode seleksi fitur berbasis apa yang efektif dalam mengidentifikasi gen dan miRNA yang relevan terhadap perkembangan stadium kanker prostat?
3. Bagaimana performa algoritma *Random Forest* dan *Logistic Regression* dalam membangun model klasifikasi stadium kanker prostat berdasarkan fitur-fitur terpilih?

1.3 Batasan Permasalahan

Untuk mengarahkan penelitian agar lebih terfokus dan terukur, maka penelitian ini memiliki beberapa batasan sebagai berikut:

1. Data yang digunakan dalam penelitian ini terbatas pada data ekspresi gen dan ekspresi miRNA dari dataset publik *The Cancer Genome Atlas - Prostate Adenocarcinoma* (TCGA-PRAD).

2. Klasifikasi stadium kanker prostat difokuskan pada pembeda antara stadium II dan stadium III, sesuai dengan kriteria yang tersedia dalam dataset.
3. Sampel yang digunakan dalam penelitian ini dibatasi pada pasien dengan label ras *White* untuk menjaga homogenitas data dan mengurangi potensi bias yang dapat muncul akibat variasi genetik antar etnis.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dirumuskan sebelumnya, berikut merupakan rumusan tujuan penelitian ini:

1. Mengembangkan model klasifikasi untuk membedakan antara stadium II dan stadium III kanker prostat menggunakan data ekspresi gen dan miRNA dari dataset TCGA-PRAD.
2. Mengidentifikasi metode seleksi fitur yang dapat menentukan gen dan miRNA yang relevan terhadap perkembangan stadium kanker prostat.
3. Mengevaluasi kinerja algoritma *Random Forest* dan *Logistic Regression* dalam membangun model klasifikasi stadium kanker prostat berdasarkan fitur-fitur yang telah diseleksi.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan kontribusi secara teoritis maupun praktis dalam bidang bioinformatika dan deteksi kanker. Adapun manfaat yang dapat diperoleh dari penelitian ini antara lain:

1. Memberikan kontribusi terhadap pengembangan ilmu pengetahuan, khususnya dalam bidang bioinformatika, kanker prostat, dan analisis data transkriptomik, melalui penerapan kombinasi metode seleksi fitur statistik dan DEG
2. Menambah wawasan ilmiah mengenai pengaruh ekspresi gen dan miRNA terhadap perkembangan stadium kanker prostat.
3. Menyediakan referensi ilmiah terkait efektivitas metode seleksi fitur dan algoritma klasifikasi dalam membangun model diagnosis kanker berbasis data transkriptomik.

4. Menghasilkan pendekatan berbasis machine learning yang dapat meningkatkan akurasi dan efisiensi diagnosis kanker prostat, terutama dalam membedakan stadium II dan III.
5. Membantu dalam identifikasi biomarker potensial berupa gen atau miRNA yang relevan pada kanker prostat

1.6 Sistematika Penulisan

Sistematika penulisan laporan adalah sebagai berikut:

1. Bab 1 PENDAHULUAN
Berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian yang ingin dicapai serta berbagai penelitian terdahulu yang mendukung penelitian ini.
2. Bab 2 LANDASAN TEORI
Berisi landasan teori dan konsep-konsep yang mendukung penelitian ini.
3. Bab 3 METODOLOGI PENELITIAN
Berisi penjelasan mengenai tahapan penelitian, mulai dari pengumpulan data, pengolahan data, proses dan metode pemilihan fitur, model klasifikasi yang digunakan, serta berbagai skenario eksperimen dan evaluasi yang dilakukan.
4. Bab 4 HASIL DAN DISKUSI
Berisi pemaparan hasil eksperimen, analisis performa model, serta pembahasan terkait efektivitas metode yang diusulkan dalam membedakan tahap kanker prostat.
5. Bab 5 KESIMPULAN DAN SARAN
Berisi kesimpulan dari hasil penelitian yang telah dilakukan, serta saran untuk pengembangan lebih lanjut.