

BAB I

PENDAHULUAN

1.1. Latar Belakang

Kanker prostat adalah jenis kanker organ yang paling umum dan penyebab kematian kedua di Amerika Serikat di kalangan pria [1], [2]. Pada tahun 2019, lebih dari 893.660 kasus kanker tercatat di Amerika Serikat, dengan kanker prostat lebih dari 191.930 kasus dari jumlah tersebut, serta perkiraan jumlah kematian yang disebabkan oleh kanker pada tahun 2020 adalah 321.160 di antaranya 33.310 adalah kanker prostat [3], [4]. Hal ini mungkin disebabkan oleh faktor resiko yang ditemukan pada kanker prostat yang meliputi usia, riwayat keluarga, dan gaya hidup [5]. Selain itu, kanker prostat memiliki beberapa varian biologis yang menunjukkan tingkat keparahan dan respons terhadap pengobatan yang berbeda. Variasi ini dapat memengaruhi jalannya penyakit, mulai dari bentuk yang tumbuh lambat hingga yang lebih agresif dan cepat menyebar.

Kanker prostat, seperti halnya bentuk kanker lainnya, berkembang melalui berbagai tahap yang berpengaruh kepada tingkat keparahan dan metode pengobatan. Tahap ini umumnya dikategorikan ke dalam tahap awal dan akhir ("*early*" dan "*late*"), yang penting untuk keputusan klinis. Kanker prostat stadium awal seringkali bersifat lokal dan lebih responsif terhadap pengobatan, sedangkan kanker stadium akhir cenderung lebih agresif, dengan kemungkinan menyebar yang lebih tinggi [6], [7]. Klasifikasi stadium kanker yang akurat sangat penting untuk mengoptimalkan jalan pengobatan dan meningkatkan keselamatan pasien [8]. Metode konvensional/tradisional klasifikasi kanker, seperti histopatologi, teknik penggambaran, dan tes antigen spesifik prostat (PSA), sudah banyak digunakan. Namun, metode tradisional memiliki keterbatasan dalam mendeteksi perubahan pada skala molekuler halus yang mendahului perkembangan penyakit yang terlihat

[9]. Oleh karena itu, para peneliti beralih ke penanda molekuler dan epigenetik untuk meningkatkan akurasi diagnostik [10]. Dikarenakan keterbatasan metode konvensional dalam mendeteksi perubahan molekuler pada tahap awal kanker, diperlukan pendekatan yang lebih sensitif dan spesifik. Salah satu pendekatan yang belakangan ini banyak digunakan adalah analisis epigenetik, khususnya metilasi DNA, yang dapat memberikan informasi lebih rinci mengenai perubahan biologis yang mendahului perkembangan kanker.

Metilasi DNA adalah modifikasi epigenetik yang mengatur jumlah ekspresi gen tanpa mengubah urutan gen yang telah ada. Metilasi DNA telah muncul sebagai jalan yang menjanjikan untuk klasifikasi kanker [11]. Pola metilasi DNA memengaruhi aktivitas gen dengan menyengap gen penekan tumor, yang dapat bekerja pada perkembangan kanker [10]. Dengan menganalisis pola-pola gen ini, para peneliti dapat mengidentifikasi *biomarker* spesifik yang terkait dengan berbagai stadium kanker. Kemajuan dalam analisis metilasi DNA dalam beberapa tahun terakhir telah membantu meningkatkan kemampuan untuk melakukan deteksi dan klasifikasi kanker, termasuk kanker prostat, serta menawarkan pendekatan diagnostik yang lebih tepat dan lebih awal. Namun, data metilasi DNA memiliki dimensi yang tinggi dan pola yang kompleks, sehingga membutuhkan metode analisis canggih untuk mengidentifikasi pola. Dalam hal ini, metode klasifikasi seperti *Multi Layer Perceptron* datang sebagai solusi atas kemampuannya dalam mengenali pola, sehingga memungkinkan klasifikasi kanker prostat dengan tingkat akurasi yang lebih tinggi [12], [13].

Multi Layer Perceptron (MLP) merupakan sebuah tipe dari jaringan saraf tiruan yang efektif dalam mengenali pola dan mengklasifikasi dalam berbagai bidang termasuk bioinformatika, khususnya untuk diagnostik kanker [14], [15], [16], [17], [18]. MLP mampu menangani data biologis dengan dimensi tinggi dan kompleks, contohnya data metilasi DNA, dengan performa superior dibanding dengan metode statistik tradisional. Berbagai studi telah menunjukkan bahwa MLP dapat digunakan untuk mengidentifikasi pola dalam biomarker spesifik kanker dan

mengklasifikasikan stadium kanker dengan akurasi tinggi [19], [20], [21]. Selain itu, MLP juga dapat meningkatkan konsistensi diagnosis patologis dan berperan sebagai alat pendukung dalam terapi kanker yang terpersonalisasi [22]. Dengan kemampuan tersebut, MLP dapat diintegrasikan ke dalam alur kerja klinis dan membuka potensi besar dalam meningkatkan deteksi dini kanker dan pengambilan keputusan terapeutik yang lebih tepat.

Meskipun *machine learning* telah mendorong kemajuan secara signifikan dalam klasifikasi kanker, sebagian besar model berbasis genomik masih mengabaikan disparitas rasial, khususnya dalam pola metilasi DNA. Abdollahi et al. [] mengusulkan kerangka berbasis radiomic untuk evaluasi prostate cancer menggunakan teknik machine learning dan fitur MRI untuk memprediksi respons terapi, *Gleason score* (GS), dan stadium kanker. Studi ini melibatkan 33 pasien yang menjalani *pre- and post-intensity-modulated radiation therapy* (IMRT) dengan T2-weighted (T2W) dan *apparent diffusion coefficient* (ADC) MRI scans. Model radiomic pasca-T2W menunjukkan performa terbaik (AUC 0.632), diikuti oleh pre-ADC (0.626) dan pre-T2W (0.61). Model berbasis T2W unggul dalam prediksi GS (mean AUC 0.739), sedangkan ADC lebih baik untuk prediksi stadium (mean AUC 0.675). Hartenstein et al. [] mengevaluasi tiga *convolutional neural networks* (CNN) untuk mendeteksi *lymph node infiltration* (LNI) menggunakan citra *contrast-enhanced CT*. Model terbaik (status-balanced CNN) mencapai AUC 0.90, menunjukkan bahwa model memanfaatkan konteks anatomi. Sebaliknya, location-balanced CNN dan *segmentation-masked* (xMask) CNN hanya mencapai AUC 0.858 dan 0.677. Random forest classifiers berbasis nodal volume dan lokasi hanya efektif pada data seimbang status (AUC 0.90), namun gagal pada data seimbang lokasi, menguatkan peran konteks anatomi pada CNN. Eissa et al. [28] memanfaatkan profil data metilasi DNA untuk mengurangi dimensi data dengan teknik metaheuristic, lalu membangun model *Deep Neural Network* (DNN) untuk klasifikasi status keganasan dan tipe kanker. Hasilnya kompetitif, dengan performa tinggi dalam recall, precision, dan accuracy, serta ROC AUC antara 0.85–0.89. Namun, penggunaan methylation data untuk klasifikasi stadium masih jarang,

terutama untuk membedakan antara stadium awal dan lanjut. Selain itu, banyak model tidak memprioritaskan minimasi fitur, yang berdampak pada interpretabilitas dan resiko overfitting. Disparitas rasial dalam pola metilasi DNA juga sering diabaikan oleh penelitian lalu, meskipun jumlah sampel rasial pada dataset publik tidak seimbang. Untuk menjawab kesenjangan ini, kami menggunakan data metilasi DNA dan menerapkan kombinasi DMP dan RFE untuk mengidentifikasi fitur minimal namun berdampak tinggi. Biomarker yang dipilih divalidasi menggunakan *Catalogue Of Somatic Mutations In Cancer* (COSMIC) dan cBioPortal.org untuk memastikan relevansi biologis. Studi ini memperkenalkan model MLP yang sadar ras untuk klasifikasi stadium prostate cancer, khusus dalam membedakan stadium awal dan lanjut, menekankan pentingnya konteks rasial dan minimasi fitur dalam pengembangan alat diagnosis yang lebih akurat dan adil.

1.2. Rumusan Masalah

Masalah yang dijabarkan dalam penelitian ini merupakan:

- Bagaimana signifikansi penggunaan metode seleksi fitur dalam mengidentifikasi biomarker kanker prostat pada data metilasi DNA?
- Bagaimana performa model *Multi-Layer Perceptron* dalam mengklasifikasi stadium kanker prostat?

1.3. Tujuan Penelitian

Tujuan dari penelitian ini meliputi:

- Merancang framework seleksi fitur untuk mengidentifikasi biomarker pada kanker prostat berdasarkan data metilasi DNA, diukur dengan peningkatan akurasi model dan validasi terhadap data referensi gen.
- Membangun model MLP untuk mengidentifikasi stadium kanker prostat berdasarkan data metilasi DNA, dievaluasi dengan metrik klasifikasi seperti akurasi dan AUC.

1.4. Urgensi Penelitian

Penelitian ini memiliki urgensi yang tinggi karena mengintegrasikan berbagai disiplin ilmu penting seperti teknologi kesehatan, genetika, dan bioinformatika. Berbeda dengan pendekatan konvensional yang seringkali bersifat general dan tidak mempertimbangkan variasi individu, pendekatan berbasis kecerdasan buatan menawarkan kemampuan untuk menganalisis data biologis secara mendalam dan mendeteksi pola yang kompleks. Dengan bantuan kecerdasan buatan, identifikasi biomarker spesifik yang terkait dengan alterasi genetik menjadi lebih akurat, sehingga membuka peluang menuju pengobatan yang lebih terpersonalisasi. Hal ini berkontribusi pada pergeseran paradigma dari terapi satu-untuk-semua ke terapi yang disesuaikan dengan karakteristik genetik masing-masing pasien.

1.5. Luaran Penelitian

1. Luaran penelitian ini berupa Artikel Ilmiah dan Prosiding Konferens.
2. *Pipeline* pengembangan untuk deteksi stadium kanker prostat menggunakan data metilasi DNA yang dapat dimanfaatkan oleh pihak rumah sakit.

1.6. Manfaat Penelitian

Manfaat yang didapat dari penelitian ini sebagai berikut;

- Model MLP untuk klasifikasi stadium kanker prostat pada ras spesifik dapat membantu spesialis medis untuk mengklasifikasi kanker prostat.
- Sistem seleksi fitur untuk mengidentifikasi *biomarker* pada data metilasi DNA.

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