

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Kanker merupakan masalah sosial, kesehatan, dan ekonomi yang besar di abad ke-21 dengan peringkat ketiga penyebab utama kematian dini pada orang dewasa di 177 dari 183 negara [1]. Pada tahun 2022, kanker bertanggung jawab atas kematian 9,7 juta di seluruh dunia, dengan kanker paru sebagai penyebab terbanyak kematian terkait kanker sekitar 1,8 juta [1], [2]. Profil Kesehatan Indonesia 2024 oleh Kementerian Kesehatan Republik Indonesia dan data GLOBOCAN 2022 menunjukkan bahwa kanker paru tercatat sebagai salah satu penyebab utama kematian akibat kanker di Indonesia, khususnya pada pria di mana penyakit ini menyumbang sekitar 15,4% dari total kematian akibat kanker [3], [4]. Tingginya angka kematian ini diperparah karena tantangan infrastruktur kesehatan dan tingginya perokok, di mana lebih dari 80% kasus baru kanker paru di Indonesia terdiagnosis pada stadium lanjut [5]. Meskipun tersedia intervensi standar seperti kemoterapi, penghambat tirosin kinase, dan imunoterapi, prognosis klinis untuk kanker paru tetap buruk karena deteksi pada stadium lanjut dan munculnya resistensi obat yang tak terhindarkan [6]. Oleh karena itu, terdapat kebutuhan mendesak dan penting untuk menemukan agen terapeutik baru yang efektif dan kurang beracun, khususnya melalui eksplorasi senyawa alami sebagai kandidat obat potensial [6], [7].

Sehubungan dengan itu, berbagai pendekatan sedang dieksplorasi, termasuk penggunaan tanaman herbal. Sebagai salah satu negara dengan megabiodiversitas terbesar di dunia, Indonesia memiliki ekosistem yang kaya akan spesies dan sumber daya genetik yang dibentuk oleh kondisi geologis dan biologisnya yang unik [8], [9]. Kekayaan botani ini menawarkan potensi signifikan untuk dieksplorasi melalui pendekatan farmakologi modern. Sebagai contoh, pemanfaatan Pegagan (*Centella asiatica*) telah dieksplorasi efektivitasnya dalam penanganan hipertensi [10]. Lebih lanjut, pendekatan komputasional *in silico* juga telah berhasil memetakan aktivitas farmakologis Sambiloto (*Andrographis paniculata*) sebagai agen terapi yang efektif dalam penanganan asma melalui analisis molecular docking [11].

Secara khusus dalam bidang onkologi, beberapa tanaman herbal terbukti memiliki aktivitas antikanker paru yang menjanjikan. Misalnya, senyawa *panduratin A* dari *Boesenbergia pandurata* (temu kunci) menunjukkan sifat antikanker yang luar biasa [12]. Selain itu, *Annona muricata* (sirsak) telah terbukti efektif menginduksi apoptosis pada sel kanker paru melalui jalur yang dimediasi mitokondria, dengan tingkat toksisitas yang lebih rendah pada sel normal dibandingkan kemoterapi standar [13], [14]. Lebih lanjut, *asiatic acid*, unsur aktif dari *Centella asiatica* (pegagan), menunjukkan efek penghambatan yang

kuat pada pertumbuhan sel kanker paru baik secara *in vitro* maupun *in vivo* [15]. Eksplorasi serupa pada Bawang Dayak (*Eleutherine bulbosa*) mengonfirmasi bahwa tanaman ini memiliki efek antitumoral dan antimetastatik yang terbukti secara *in silico* [16]. Hal yang sama juga terlihat pada Lempuyang Gajah (*Zingiber zerumbet*), di mana senyawa utamanya, *zerumbone*, tidak hanya terbukti secara *in vitro* mampu menghambat proliferasi sel kanker paru [17], tetapi interaksi target molekulernya juga telah divalidasi secara ekstensif melalui pendekatan komputasional [18]. Meskipun tanaman herbal telah terbukti menjanjikan, pemahaman ilmiah yang komprehensif tentang mekanisme molekuler yang tepat dan interaksi target obat spesifiknya masih sangat terbatas [10].

Secara historis, upaya mengungkap interaksi obat-target kompleks sangat bergantung pada skrining *in vitro* dan *in vivo* secara tradisional, yang dikenal sangat mahal, memakan waktu lama, dan memiliki tingkat kegagalan yang tinggi. Studi sistematis menunjukkan bahwa rata-rata biaya kapitalisasi pengembangan obat baru dapat mencapai miliaran dolar, dengan obat onkologi tercatat sebagai kategori terapi yang membutuhkan biaya pengembangan tertinggi, berkisar antara \$944 juta hingga \$4,54 miliar [19]. Pendekatan komputasi *in silico* berbasis *machine learning* (ML) menjadi salah satu solusi utama untuk mengatasi kendala tersebut. Secara kuantitatif, implementasi ML dalam fase penemuan awal (*early drug discovery*) dilaporkan mampu memangkas waktu penapisan konvensional dari rata-rata 4 hingga 5 tahun menjadi hanya dalam hitungan bulan [20], [21]. Selain akselerasi waktu, penapisan virtual berbasis algoritma ini juga berpotensi menekan biaya riset praklinis hingga 40% sampai 50% dengan mengeliminasi pengujian eksperimental *in vitro* yang tidak perlu pada kandidat senyawa yang diprediksi gagal [22], [23].

Sebagai instrumen krusial dalam bioinformatika, model prediksi *Drug-Target Interaction* (DTI) menggeser paradigma ini dengan memungkinkan penapisan ribuan senyawa secara komputasional dengan efisiensi tinggi [24], [25]. Melalui pendekatan DTI ini, kandidat molekul aktif baru dapat diidentifikasi dan dievaluasi potensinya secara prediktif sebelum memasuki tahap pengujian klinis yang memakan waktu dan biaya besar [26]. Pendekatan DTI telah banyak diterapkan pada tanaman herbal, seperti pada analisis *Network Pharmacology* untuk tanaman Kemuning (*Murraya paniculata*) serta eksplorasi senyawa herbal untuk terapi penyakit metabolik [10], [27].

Untuk menangani kompleksitas data biologis yang heterogen, arsitektur berbasis *tree* tingkat lanjut menjadi pilihan utama karena kemampuannya yang tangguh dalam memproses matriks fitur molekuler. Hal ini didukung oleh berbagai penelitian yang menunjukkan bahwa algoritma XGBoost terbukti mengungguli model jaringan saraf tiruan (RPropMLP) secara konsisten pada 6 dataset prediksi toksisitas yang berbeda [28], sementara Random Forest mendominasi sebagai algoritma prediksi terbaik dengan metrik akurasi dan skor $F1 \geq 0,8$ pada 13 dari 35 model yang dievaluasi [29]. Performa ini dapat

dioptimalkan melalui arsitektur *deep learning* berbasis pohon seperti *Cascade Deep Forest* (CDF). Sebagai contoh ketangguhan dasar arsitektur ini, literatur terbaru menunjukkan bahwa pengembangan dari CDF (seperti model LGBMDF) mampu memprediksi interaksi obat-target dengan nilai AUC mencapai 0,9844 [30]. Studi terbaru juga mengonfirmasi keunggulan model *ensemble* dalam memperkirakan nilai *binding affinity* dan menyaring ruang kimiawi unik dari produk tanaman alami. Secara spesifik, XGBoost menghasilkan skor AUC 0,8636 dalam memprediksi interaksi produk alami-protein [31], serta mencapai akurasi prediksi *binding affinity* dengan nilai R^2 hingga 0,92 pada dataset berskala besar [32]. Penggunaan rangkaian metode ini sejalan dengan kerangka *Model-Informed Drug Development* (MIDD) yang diakui secara regulasi sebagai pendekatan efektif untuk mengefisienkan penyaringan dan mempercepat pengambilan keputusan dalam riset farmasi modern [33].

Meskipun pendekatan *machine learning* terbukti efektif dalam memprediksi kemungkinan interaksi senyawa bioaktif dengan target secara cepat, model tersebut umumnya beroperasi menggunakan deskriptor numerik dan tidak secara langsung mempertimbangkan aspek struktural tiga dimensi yang mendasari interaksi fisik ligan-protein [34]. Oleh karena itu, *pipeline* sekuensial yang menggabungkan penyaringan berbasis ML dengan validasi menggunakan *molecular docking* diperlukan untuk mengevaluasi kelayakan struktural serta stabilitas interaksi senyawa-protein secara komputasional [24]. Menurut beberapa studi, *pipeline* sekuensial yang menggabungkan penyaringan komputasi berkecepatan tinggi dengan evaluasi struktural tiga dimensi telah terbukti efektif dalam mengidentifikasi kandidat inhibitor baru secara presisi, baik untuk target onkologi maupun dalam eksplorasi senyawa bahan alam untuk berbagai penyakit sistemik [26], [34], [35].

Meskipun demikian, sebagian besar penelitian terdahulu berfokus pada pustaka senyawa umum atau senyawa sintesis yang tersedia secara global. Kajian komprehensif yang mengeksplorasi potensi senyawa bioaktif tanaman herbal Indonesia sebagai kandidat anti-kanker paru masih relatif terbatas. Selain itu, penerapan dan evaluasi komparatif berbagai algoritma *ensemble machine learning* untuk memprediksi interaksi antara senyawa herbal dan protein target kanker paru juga masih jarang dilaporkan. Oleh karena itu, penelitian ini mengusulkan pendekatan *in silico* yang mengintegrasikan data senyawa bioaktif tanaman herbal dari repositori botani seperti KNApSack [36] serta data protein target terkait kanker paru dari basis data biomedis seperti OMIM [37], MalaCards [38], dan UniProt [39]. Pustaka senyawa dan data target yang diperoleh kemudian disaring menggunakan beberapa algoritma *ensemble machine learning* untuk memprediksi interaksi obat-target. Model dengan performa terbaik selanjutnya digunakan untuk melakukan *virtual screening*, sementara kandidat senyawa dengan skor prediksi tertinggi dievaluasi lebih lanjut melalui simulasi *molecular docking* guna mengamati stabilitas dan afinitas

pengikatan pada tingkat struktural. Integrasi ekstraksi data dari basis data publik, *ensemble machine learning*, dan simulasi tiga dimensi ini diharapkan dapat menjadi pendekatan ilmiah yang efisien dan hemat biaya untuk mempercepat identifikasi kandidat obat kanker paru berbasis kekayaan hayati Indonesia.

1.2 Rumusan Masalah

Berdasarkan latar belakang dan batasan masalah yang telah dipaparkan, maka permasalahan utama dalam penelitian ini dapat dirumuskan dalam bentuk pertanyaan ilmiah sebagai berikut:

1. Senyawa bioaktif apa saja dari tanaman herbal Indonesia yang berpotensi sebagai kandidat potensial anti-kanker paru, beserta signifikansi jalur biologisnya, berdasarkan pendekatan komputasional secara *in silico*?
2. Bagaimana performa model *machine learning* dalam memprediksi interaksi antara senyawa bioaktif tanaman herbal Indonesia dengan protein target kanker paru?
3. Bagaimana hasil evaluasi interaksi senyawa-protein yang diprediksi menggunakan simulasi molecular docking?

1.3 Batasan Permasalahan

Penelitian ini memiliki beberapa batasan yang ditetapkan untuk mempertajam fokus kajian sesuai dengan rumusan masalah dan tujuan penelitian. Batasan-batasan tersebut adalah sebagai berikut:

1. Penelitian dibatasi pada pemanfaatan data senyawa bioaktif dari tanaman herbal Indonesia serta data protein target kanker paru yang diperoleh melalui basis data publik.
2. Data latih untuk pengembangan model *machine learning* sepenuhnya berasal dari basis data sekunder yang telah terkurasi, tanpa melibatkan proses pengumpulan data primer.
3. Jalur biologis dibatasi pada penggunaan anotasi *Gene Ontology* (GO) dan analisis *Kyoto Encyclopedia of Genes and Genomes* (KEGG).
4. Evaluasi interaksi senyawa-protein dibatasi pada pendekatan komputasional menggunakan simulasi molecular docking.
5. Penelitian ini tidak mencakup tahapan validasi eksperimental di laboratorium, baik *in vitro* pada kultur sel maupun *in vivo* pada hewan uji.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dikemukakan, penelitian ini bertujuan untuk mencapai beberapa hal sebagai berikut:

1. Mengidentifikasi senyawa-senyawa bioaktif dari tanaman herbal Indonesia yang berpotensi sebagai kandidat inhibitor protein target kanker paru
2. Mengaplikasikan model *machine learning* untuk memprediksi interaksi antara senyawa bioaktif tanaman herbal dengan protein target kanker paru.
3. Mengevaluasi interaksi senyawa bioaktif terpilih dengan protein targetnya menggunakan metode simulasi molecular docking.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Memberikan daftar kandidat senyawa bioaktif dari tanaman herbal Indonesia yang berpotensi sebagai inhibitor protein target kanker paru berdasarkan pendekatan *in silico*.
2. Membantu mempercepat proses penelitian lanjutan di laboratorium (*in vitro* maupun *in vivo*) melalui penyediaan hasil penyaringan awal (*virtual screening*) yang lebih efisien.
3. Memberikan kontribusi dalam pengembangan metode penemuan obat berbasis komputasi (*Computer-Aided Drug Design*) dengan mengintegrasikan ekstraksi data dari basis data publik, *ensemble machine learning*, dan simulasi molecular docking.
4. Menjadi referensi awal bagi penelitian selanjutnya terkait pemanfaatan kekayaan hayati Indonesia dalam pengembangan kandidat obat kanker paru.

1.6 Sistematika Penulisan

Berisikan uraian singkat mengenai struktur isi penulisan laporan penelitian, dimulai dari Pendahuluan hingga Simpulan dan Saran.

- Bab 1 PENDAHULUAN

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

- Bab 2 LANDASAN TEORI

Membahas tinjauan pustaka yang meliputi teori pendukung dan penelitian terdahulu terkait senyawa bioaktif, kanker paru, tanaman herbal Indonesia, *drug-target interaction*, metode *machine learning*, serta simulasi molecular docking.

- Bab 3 METODOLOGI PENELITIAN

Menjelaskan metodologi penelitian yang digunakan, meliputi pengumpulan dan ekstraksi data dari basis data publik, tahap pra-pemrosesan data, *feature engineering*, pemodelan prediksi DTI menggunakan *machine learning*, serta prosedur evaluasi menggunakan molecular docking.

- Bab 4 HASIL DAN DISKUSI

Menyajikan hasil penelitian dan pembahasan, meliputi hasil prediksi interaksi senyawa-protein, evaluasi performa model, serta analisis hasil simulasi molecular docking.

- Bab 5 KESIMPULAN DAN SARAN

Berisi kesimpulan yang diperoleh dari penelitian serta saran untuk pengembangan penelitian selanjutnya.

