

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Kanker paru-paru merupakan salah satu penyebab utama kematian akibat kanker di seluruh dunia. Berdasarkan data dari *Global Cancer Observatory* tahun 2020, kanker paru-paru menempati posisi teratas sebagai penyebab kematian akibat kanker secara global, dengan mencatatkan sekitar 1,8 juta kematian dari total 2,2 juta kasus baru [1]. Di Indonesia, penyakit ini juga menjadi perhatian serius, berada pada peringkat ketiga tertinggi dalam jumlah kasus baru, yaitu sebanyak 34.783 kasus, dan angka kematian mencapai 30.843 jiwa pada tahun 2020 [1]. Laporan dari Kementerian Kesehatan Republik Indonesia memperkuat temuan ini, menunjukkan bahwa prevalensi kanker paru-paru terus meningkat dan menjadi fokus utama dalam program pengendalian kanker nasional [2]. Tingginya angka insiden dan mortalitas ini menandakan adanya kebutuhan mendesak untuk strategi penanganan yang lebih efektif dan inovatif.

Tingginya angka kematian ini sebagian besar disebabkan oleh sulitnya deteksi dini serta terbatasnya pilihan pengobatan yang efektif. Metode pengobatan konvensional seperti kemoterapi dan radioterapi, meskipun sering digunakan, memiliki efek samping yang signifikan dan tidak jarang menimbulkan resistansi pada sel kanker. Oleh karena itu, pencarian alternatif terapi yang lebih aman, murah, dan efektif menjadi sangat penting. Salah satu pendekatan yang semakin berkembang adalah pemanfaatan senyawa alami dari tanaman. Salah satu pendekatan yang semakin berkembang adalah pemanfaatan senyawa alami dari tanaman. Indonesia, sebagai negara dengan keanekaragaman hayati yang melimpah, memiliki potensi besar sebagai sumber tanaman obat yang belum banyak dieksplorasi.

Berbagai penelitian menunjukkan bahwa tanaman mengandung senyawa bioaktif seperti flavonoid, alkaloid dan terpenoid yang memiliki aktivitas antikanker [3]. Dalam konteks kanker paru-paru, penelitian terkini bahkan berhasil mengidentifikasi senyawa alami potensial melalui pendekatan *virtual screening* dan pemodelan molekuler berbasis *machine learning*, yang dapat menghambat perkembangan *non-small cell lung cancer* [4]. Pendekatan ini membuka jalan untuk penemuan obat yang lebih cepat dengan biaya yang lebih rendah dibandingkan

dengan metode *skrining* tradisional.

Penggunaan teknologi kecerdasan buatan, khususnya algoritma *machine learning* seperti *Random Forest*, memberikan terobosan baru dalam dunia biomedis. Algoritma ini memiliki kemampuan dalam mengenali pola dari data yang kompleks dan berjumlah besar. Dalam bidang bioinformatika dan penemuan obat, *machine learning* telah diterapkan untuk memprediksi aktivitas antikanker senyawa berbasis tanaman secara akurat [5, 6]. Kemajuan ini juga didukung oleh penggunaan metode seperti QSAR (*Quantitative Structure–Activity Relationship*) dan *molecular docking* untuk menganalisis efektivitas senyawa [7, 8]. Metode-metode ini memungkinkan analisis hubungan antara struktur kimia suatu senyawa dengan aktivitas biologisnya, sehingga proses identifikasi kandidat obat menjadi lebih terarah. Pendekatan ini terbukti sangat efektif, bahkan digunakan untuk melakukan virtual screening dalam rangka mengidentifikasi kandidat inhibitor EGFR generasi baru untuk mengatasi resistansi obat pada kanker paru-paru [9].

Beberapa studi di Indonesia telah mengembangkan sistem klasifikasi berbasis web untuk mengelompokkan tanaman obat menggunakan algoritma *Random Forest* [10], yang menunjukkan potensi besar dalam pengembangan sistem pendukung keputusan medis. Selain itu, buku-buku ilmiah terbaru menunjukkan bahwa teknik pembelajaran mesin, termasuk metode *few-shot* dan *zero-shot learning*, memiliki peranan penting dalam mempercepat proses *screening* awal kandidat obat dari senyawa alami [11].

Dari uraian di atas, penelitian ini bertujuan untuk mengidentifikasi dan mengklasifikasikan senyawa dari tanaman yang berpotensi sebagai agen penyembuh kanker paru-paru. Pendekatan ini menggabungkan data kimiawi senyawa—seperti jumlah *hydrogen bond donor* (HBD), *hydrogen bond acceptor* (HBA), massa molekul, logP dan Molecular Fingerprints jenis ECFP4 dengan representasi 1024 bit, yang dihitung dari struktur SMILES senyawa dengan kemampuan algoritma *Random Forest* dalam melakukan klasifikasi. Diharapkan hasil dari penelitian ini dapat memberikan kontribusi signifikan terhadap pengembangan terapi berbasis tanaman yang lebih terstruktur, efisien, dan mudah diakses.

1.2 Rumusan Masalah

Dari latar belakang di atas, maka didapatkan rumusan masalah sebagai berikut:

1. Bagaimana mengidentifikasi senyawa yang berpotensi menyembuhkan

penyakit kanker paru-paru dari tanaman-tanaman tersebut.

2. Bagaimana cara kerja Algoritma Random Forest dalam implementasi sistem klasifikasi tersebut?
3. Seberapa efektif algoritma Random Forest dalam memberikan hasil klasifikasi ?

1.3 Batasan Permasalahan

Penelitian ini memiliki beberapa batasan yang ditetapkan untuk menjaga fokus dan ruang lingkup kajian agar tetap relevan dengan tujuan penelitian. Batasan tersebut adalah sebagai berikut:

1. Data penelitian bersumber dari database publik ChEMBL dan terbatas pada senyawa yang telah diuji aktivitas biologisnya (IC) terhadap target protein Human EGFR (ID: ChEMBL203).
2. Senyawa diklasifikasikan ke dalam dua kategori biner: Potensial (IC kurang dari 10.000 nM) dan Tidak Potensial (IC lebih dari 10.000 nM).
3. Fitur yang digunakan untuk pemodelan adalah parameter HBD, HBA, massa molekul, dan logP dan Molecular Fingerprints jenis ECFP4 dengan representasi 1024 bit, yang dihitung dari struktur SMILES senyawa.
4. Algoritma machine learning yang digunakan terbatas pada Random Forest.
5. Evaluasi model utama menggunakan metode 5-Fold Cross-Validation.
6. Fokus kanker yang dikaji adalah kanker paru-paru.
7. Aplikasi yang dikembangkan bersifat prototipe dan tidak terhubung secara dinamis ke database.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah dan batasan yang telah diuraikan sebelumnya, tujuan dari penelitian ini adalah sebagai berikut:

1. Mengidentifikasi senyawa yang berpotensi menyembuhkan kanker paru-paru dari tanaman berdasarkan parameter HBD, HBA, massa molekul, logP dan Molecular Fingerprints jenis ECFP4 dengan representasi 1024 bit, yang dihitung dari struktur SMILES senyawa.
2. Mengetahui cara kerja Algoritma Random Forest pada sistem klasifikasi yang akan dibangun.
3. Mengetahui seberapa efektif Algoritma Random Forest pada sistem rekomendasi yang akan dibangun.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat bagi peneliti, peneliti lain, dan masyarakat, yaitu sebagai berikut:

1. Memberikan referensi awal dalam mencari senyawa tanaman untuk obat kanker paru-paru. Manfaat ini secara langsung ditujukan kepada para peneliti di bidang farmasi dan kedokteran untuk mempercepat tahap eksplorasi awal.
2. Menjadi contoh penerapan algoritma Random Forest dalam bidang bioinformatika. Penelitian ini dapat menjadi studi kasus bagi mahasiswa dan akademisi informatika yang tertarik pada aplikasi kecerdasan buatan di dunia kesehatan.
3. Menyediakan dasar untuk pembuatan sistem prediksi obat berbasis tanaman. Prototipe yang dihasilkan dapat dikembangkan lebih lanjut menjadi sebuah platform yang komprehensif untuk penemuan obat herbal.

1.6 Sistematika Penulisan

Sistematika penulisan laporan adalah sebagai berikut:

- Bab 1 PENDAHULUAN

Bab ini berisi latar belakang masalah yang menjelaskan urgensi penelitian tentang kanker paru-paru dan potensi senyawa tanaman sebagai terapi alternatif, rumusan masalah, batasan masalah, tujuan, manfaat, serta sistematika penulisan laporan.

- Bab 2 LANDASAN TEORI

Bab ini menguraikan teori-teori dan penelitian terdahulu yang relevan, seperti konsep dasar kanker paru-paru, senyawa bioaktif dalam tanaman, parameter kimiawi senyawa (HBD, HBA, massa molekul, logP dan Molecular Fingerprints jenis ECFP4 dengan representasi 1024 bit, yang dihitung dari struktur SMILES senyawa), serta penjelasan mengenai algoritma Random Forest dan penerapannya dalam bioinformatika.

- Bab 3 METODOLOGI PENELITIAN

Bab ini menjelaskan langkah-langkah yang dilakukan dalam penelitian, meliputi pengumpulan dan seleksi data senyawa tanaman, proses ekstraksi fitur kimia, penerapan algoritma Random Forest untuk klasifikasi, serta pengembangan prototipe sistem berbasis web.

- Bab 4 HASIL DAN DISKUSI

Bab ini menyajikan hasil dari proses klasifikasi senyawa, analisis performa model Random Forest, serta interpretasi terhadap senyawa-senyawa yang teridentifikasi sebagai berpotensi menyembuhkan kanker paru-paru. Diskusi juga mencakup keterbatasan penelitian dan saran pengembangan ke depan.

- Bab 5 SIMPULAN DAN SARAN

Bab terakhir ini memuat simpulan dari hasil penelitian serta saran untuk pengembangan penelitian selanjutnya, baik dari sisi data, metode, maupun penerapan sistem klasifikasi senyawa tanaman untuk penyakit lain.

U M M N
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