

BAB 2

LANDASAN TEORI

2.1 Leukemia

Leukemia adalah keganasan hematologis yang ditandai oleh proliferasi abnormal dan tidak terkendali dari sel-sel darah putih di sumsum tulang dan darah tepi, yang mengganggu proses hematopoiesis normal serta menyebabkan akumulasi sel-sel blastik imatur [24]. Berdasarkan laju progresi penyakit dan tipe sel yang terlibat, leukemia diklasifikasikan menjadi empat subtype utama, yaitu leukemia limfoblastik akut (ALL), leukemia mieloid akut (AML), leukemia mieloid kronik (CML), dan leukemia limfositik kronik (CLL) [8]. ALL merupakan subtype yang paling sering ditemukan pada anak-anak dan remaja, ditandai oleh proliferasi cepat limfoblas imatur di sumsum tulang; sementara AML adalah subtype akut yang paling umum pada dewasa dan ditandai oleh akumulasi mieloblas maligna yang menghambat fungsi sumsum tulang normal [25]. CML, yang didefinisikan oleh adanya kromosom Philadelphia akibat translokasi gen *BCR-ABL1*, bersifat kronik namun dapat mengalami transformasi menjadi krisis blast yang bersifat fatal; sedangkan CLL adalah tipe leukemia paling umum secara keseluruhan dan umumnya terjadi pada populasi usia lanjut, ditandai oleh akumulasi progresif limfosit B klonal yang matur [24, 25].

Pada tingkat molekuler, leukemogenesis diinisiasi oleh serangkaian perubahan genetik dan epigenetik pada sel hematopoietik, termasuk aktivasi onkogen, inaktivasi gen supresor tumor, serta disregulasi modifikasi epigenetik yang mengubah pola ekspresi gen [10]. Jalur sinyal utama yang terlibat dalam patogenesis leukemia mencakup jalur PI3K/AKT/mTOR, NF- κ B, JAK/STAT3, Wnt/ β -catenin, dan Notch, yang secara kolektif mengatur proliferasi, kelangsungan hidup sel, dan hambatan diferensiasi [11]. Leukemia stem cells (LSCs) memainkan peran sentral dalam inisiasi tumor, pemeliharaan klon leukemik, serta resistansi terhadap terapi melalui aktivasi jalur *self-renewal* seperti Hedgehog dan STAT3 [11]. Gangguan mekanisme apoptosis, terutama melalui overekspresi protein antiapoptotik keluarga BCL-2 (BCL-2, MCL-1, BCL-XL), merupakan ciri khas leukemia yang menyebabkan sel kanker lolos dari kematian sel terprogram dan berkontribusi terhadap resistansi kemoterapi [13]. Resistansi multidrug (MDR) yang dimediasi oleh pompa efluks seperti P-glikoprotein (P-gp) serta

heterogenitas tumor juga menjadi hambatan utama dalam keberhasilan terapi leukemia konvensional [12].

2.2 Algoritma Centrality

Dalam analisis jaringan dan teori graf, *centrality* merupakan ukuran yang digunakan untuk mengevaluasi tingkat kepentingan suatu node berdasarkan posisinya di dalam suatu jaringan [4]. Nilai *centrality* memberikan gambaran mengenai kontribusi relatif setiap node terhadap struktur jaringan secara keseluruhan. Berbagai ukuran *centrality* telah dikembangkan untuk merepresentasikan karakteristik jaringan dari sudut pandang yang berbeda. Penelitian ini menggunakan empat ukuran utama, yaitu *degree centrality*, *betweenness centrality*, *closeness centrality*, dan *eigenvector centrality* [4, 5].

Degree centrality menggambarkan banyaknya hubungan langsung yang dimiliki suatu node dengan node lain. *Betweenness centrality* menunjukkan seberapa besar peran sebuah node sebagai penghubung yang menghubungkan node-node lain melalui jalur terpendek. Sementara itu, *closeness centrality* mengukur tingkat kedekatan suatu node terhadap seluruh node dalam jaringan berdasarkan rata-rata jarak terpendek. Berbeda dengan ketiga ukuran tersebut, *eigenvector centrality* tidak hanya mempertimbangkan jumlah koneksi, tetapi juga memperhitungkan tingkat kepentingan node-node yang terhubung sehingga mampu menggambarkan pengaruh relatif suatu node dalam jaringan [4, 5].

Secara matematis, keempat ukuran *centrality* tersebut dirumuskan sebagai berikut.

Degree Centrality (DC) pada rumus 2.1 menyatakan proporsi jumlah koneksi langsung yang dimiliki suatu node terhadap jumlah koneksi maksimum yang mungkin terbentuk dalam jaringan [26]:

$$DC(v) = \frac{\deg(v)}{n-1} \quad (2.1)$$

dengan $\deg(v)$ menyatakan jumlah *edge* yang dimiliki node v , sedangkan n merupakan jumlah seluruh node pada jaringan.

Betweenness Centrality (BC) pada rumus 2.2 menyatakan tingkat keterlibatan suatu node pada jalur terpendek yang menghubungkan pasangan node lain di dalam jaringan [27, 28]:

$$BC(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \quad (2.2)$$

dengan σ_{st} merupakan jumlah seluruh jalur terpendek dari node s menuju node t , sedangkan $\sigma_{st}(v)$ menyatakan jumlah jalur terpendek tersebut yang melewati node v .

Closeness Centrality (CC) pada rumus 2.3 digunakan untuk mengukur kedekatan suatu node terhadap seluruh node lain berdasarkan total jarak terpendeknya [26]:

$$CC(v) = \frac{n-1}{\sum_{u \neq v} d(v, u)} \quad (2.3)$$

dengan $d(v, u)$ menunjukkan panjang jalur terpendek antara node v dan node u .

Eigenvector Centrality (EC) pada rumus 2.4 mengukur tingkat pengaruh suatu node dengan mempertimbangkan kualitas node-node yang terhubung dengannya. Sebuah node akan memperoleh nilai yang lebih tinggi apabila terhubung dengan node-node yang juga memiliki tingkat pengaruh tinggi [29]:

$$EC(v) = \frac{1}{\lambda} \sum_{u \in N(v)} EC(u) \quad (2.4)$$

dengan $N(v)$ menyatakan himpunan node tetangga dari v , sedangkan λ merupakan nilai *eigenvalue* terbesar dari matriks ketetanggaan (*adjacency matrix*).

Pada penelitian berbasis jaringan interaksi protein-protein (*protein-protein interaction/PPI*), analisis *centrality* banyak dimanfaatkan untuk menentukan protein yang memiliki posisi strategis dalam jaringan biologis [6]. Protein dengan nilai *degree centrality* yang tinggi umumnya berinteraksi dengan lebih banyak protein lain sehingga sering dikaitkan dengan fungsi biologis yang esensial dan berpotensi menjadi target terapeutik [4, 6]. Selain itu, kombinasi beberapa ukuran *centrality* memungkinkan proses penentuan prioritas protein menjadi lebih komprehensif karena setiap ukuran merepresentasikan aspek topologi jaringan yang berbeda. Dalam penelitian ini, pendekatan tersebut digunakan untuk mengidentifikasi protein target utama yang berperan penting pada patogenesis leukemia sehingga dapat diprioritaskan pada tahapan analisis berikutnya, termasuk proses *enrichment*

analysis dan *molecular docking* [5, 7].

2.3 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) adalah teknik analisis multivariat tak tersupervisi yang digunakan untuk mereduksi dimensi data tinggi menjadi sejumlah kecil komponen utama (*principal components*) yang merepresentasikan sebagian besar variasi dalam data [30]. PCA bekerja dengan cara mengidentifikasi arah variasi terbesar dalam data multidimensi dan memproyeksikannya ke ruang berdimensi lebih rendah tanpa memerlukan asumsi teoritis yang kuat, sehingga banyak digunakan sebagai metode eksplorasi data biologis yang kompleks [30]. Dalam bidang bioinformatika dan penemuan obat, PCA digunakan untuk analisis data ekspresi gen, visualisasi data omik, identifikasi pola biologis, serta karakterisasi sifat fisikokimia senyawa kimia [31].

Secara intuitif, PCA dapat dipahami sebagai proses mencari sumbu (arah) baru dalam ruang data yang mampu menangkap sebanyak mungkin informasi (variasi) dari data asli dengan jumlah variabel yang lebih sedikit. Pada penelitian ini, keempat parameter centrality (Degree, Betweenness, Closeness, dan Eigenvector Centrality) yang dimiliki setiap protein diperlakukan sebagai empat variabel berdimensi tinggi. PCA mentransformasikan keempat variabel tersebut menjadi dua variabel baru yang saling bebas (independen), yaitu Principal Component 1 (PC1) dan Principal Component 2 (PC2). PC1 merupakan kombinasi linear dari keempat parameter centrality yang menjelaskan proporsi variasi data terbesar, sehingga dapat digunakan sebagai skor komposit tunggal untuk memeringkat protein berdasarkan tingkat sentralitasnya secara keseluruhan, tanpa harus membandingkan keempat parameter tersebut satu per satu.

Dalam penelitian farmakologi jaringan berbasis herbal, PCA berperan dalam mengidentifikasi pola hubungan antar variabel dan membantu menentukan kandidat protein prioritas berdasarkan beberapa parameter analisis jaringan. Dengan mereduksi sejumlah variabel menjadi beberapa komponen utama, PCA mampu mempertahankan sebagian besar informasi penting sehingga proses interpretasi data menjadi lebih sederhana [30].

Sebelum dilakukan PCA, seluruh variabel dinormalisasi menggunakan metode standardisasi (*Z-score*) agar setiap fitur memiliki rata-rata nol dan simpangan baku satu. Standardisasi dilakukan menggunakan persamaan berikut.

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \quad (2.5)$$

dengan:

- x_{ij} adalah nilai fitur ke- j pada observasi ke- i ,
- μ_j adalah rata-rata fitur ke- j ,
- σ_j adalah simpangan baku fitur ke- j ,
- z_{ij} adalah nilai hasil standardisasi.

Setelah seluruh data berada pada skala yang sama, hubungan antar variabel direpresentasikan dalam bentuk matriks kovarian.

$$C = \frac{1}{n-1} Z^T Z \quad (2.6)$$

dengan:

- C adalah matriks kovarian,
- Z adalah matriks data hasil standardisasi,
- n adalah jumlah observasi.

Selanjutnya dilakukan dekomposisi eigen terhadap matriks kovarian untuk memperoleh eigenvalue dan eigenvector.

$$Cv = \lambda v \quad (2.7)$$

dengan:

- C adalah matriks kovarian,
- v adalah eigenvector,
- λ adalah eigenvalue.

Eigenvector digunakan sebagai arah pembentukan komponen utama, sedangkan eigenvalue menunjukkan besarnya variasi data yang mampu dijelaskan oleh masing-masing komponen.

$$T = ZW \quad (2.8)$$

dengan:

- T adalah matriks skor principal component,
- Z adalah matriks data hasil standardisasi,
- W adalah matriks eigenvector.

Besarnya variasi data yang mampu dijelaskan oleh setiap komponen utama dihitung menggunakan Explained Variance Ratio (EVR).

$$EVR_i = \frac{\lambda_i}{\sum_{j=1}^m \lambda_j} \quad (2.9)$$

dengan:

- EVR_i adalah proporsi variasi yang dijelaskan oleh komponen ke- i ,
- λ_i adalah eigenvalue komponen ke- i ,
- m adalah jumlah seluruh principal component.

Nilai *Explained Variance Ratio* digunakan untuk mengetahui seberapa besar informasi yang mampu dipertahankan oleh setiap komponen utama setelah proses reduksi dimensi. Semakin besar nilai EVR, semakin besar kontribusi komponen tersebut dalam merepresentasikan variasi data asli.

2.4 Enrichment Analysis

Enrichment analysis adalah metode komputasional yang digunakan untuk mengidentifikasi proses biologis, fungsi molekuler, komponen seluler, dan jalur sinyal yang secara statistik diperkaya (over-represented) dalam suatu kelompok gen atau protein target [32]. Dua pendekatan utama yang paling banyak digunakan adalah *over-representation analysis* (ORA) dan *gene set enrichment analysis* (GSEA); ORA menguji apakah gen dalam suatu daftar terpilih memiliki representasi berlebih dalam suatu anotasi biologis menggunakan uji Fisher atau uji chi-kuadrat, sedangkan GSEA menganalisis seluruh daftar gen yang diurutkan berdasarkan skor signifikansi tanpa memerlukan ambang batas dikotomis [32]. Dua

basis pengetahuan yang paling sering digunakan dalam enrichment analysis adalah *Gene Ontology* (GO) dan *Kyoto Encyclopedia of Genes and Genomes* (KEGG) [33].

Gene Ontology (GO) mengorganisasi anotasi fungsi gen dalam tiga domain independen yaitu *biological process* (BP) yang mendeskripsikan proses biologis yang dilakukan suatu gen, *molecular function* (MF) yang mendeskripsikan aktivitas molekuler suatu produk gen, dan *cellular component* (CC) yang mendeskripsikan lokasi seluler produk gen tersebut [33]. KEGG memetakan gen target ke dalam jaringan metabolik dan jalur sinyal yang terstruktur, memungkinkan identifikasi jalur biologis yang aktif atau terganggu pada kondisi penyakit tertentu [33]. Dalam konteks farmakologi jaringan antikanker, enrichment analysis digunakan untuk menginterpretasikan protein target yang diperoleh dari jaringan PPI, mengidentifikasi jalur biologis kritis yang dimodulasi oleh senyawa herbal, serta memberikan dasar mekanistik yang komprehensif bagi prediksi efek terapeutik [32, 34].

2.5 Network Pharmacology

Network pharmacology merupakan pendekatan dalam penemuan obat yang mengombinasikan konsep farmakologi, bioinformatika, dan biologi sistem untuk menganalisis hubungan antara senyawa bioaktif, target molekuler, serta mekanisme penyakit secara menyeluruh pada tingkat jaringan biologis [7]. Pendekatan ini dikembangkan sebagai alternatif terhadap paradigma konvensional *one drug-one target*, karena banyak penyakit kompleks tidak dapat dijelaskan hanya melalui interaksi dengan satu target molekuler saja [35, 36].

Berbeda dengan pendekatan tradisional, *network pharmacology* menerapkan konsep *multi-component*, *multi-target*, *multi-pathway*, yaitu suatu pendekatan yang mempertimbangkan bahwa beberapa senyawa dapat bekerja secara simultan pada banyak target dan berbagai jalur biologis yang saling berinteraksi [35, 36]. Konsep tersebut sangat sesuai untuk mengevaluasi mekanisme kerja tanaman herbal yang umumnya mengandung beragam senyawa bioaktif dengan aktivitas farmakologis yang berbeda-beda.

Pendekatan ini didasarkan pada pemahaman bahwa penyakit kompleks, termasuk kanker, muncul akibat gangguan pada jaringan biologis yang melibatkan interaksi berbagai gen, protein, dan jalur pensinyalan, bukan semata-mata disebabkan oleh perubahan pada satu gen atau satu protein tertentu [7]. Oleh sebab itu, *network pharmacology* mampu memberikan gambaran yang lebih

komprehensif mengenai mekanisme penyakit, mengidentifikasi protein target yang berpotensi menjadi sasaran terapi, serta memprediksi mekanisme kerja senyawa bioaktif melalui analisis hubungan antar komponen dalam jaringan biologis [35].

Alur kerja farmakologi jaringan secara umum meliputi: (1) identifikasi dan penyaringan senyawa bioaktif dari tanaman herbal berdasarkan kriteria *drug-likeness* seperti aturan Lipinski; (2) prediksi target protein untuk setiap senyawa menggunakan basis data dan alat komputasional; (3) identifikasi target penyakit dari basis data genomik dan proteomik; (4) konstruksi jaringan PPI dari target overlap; (5) analisis topologi jaringan menggunakan centrality untuk identifikasi protein *hub*; dan (6) validasi melalui *enrichment analysis* dan *molecular docking* [36, 35]. Penerapan farmakologi jaringan dalam penelitian antikanker herbal telah berkembang pesat sejak 2022, dengan jumlah publikasi yang meningkat tajam hingga lebih dari 30 persen pada 2024 dibandingkan tahun-tahun sebelumnya [35].

2.6 Molecular Docking

Molecular docking adalah metode komputasional yang digunakan untuk memprediksi orientasi dan konformasi yang disukai dari suatu ligan ketika berikatan dengan protein target (reseptor), membentuk kompleks yang stabil [37]. Metode ini memanfaatkan fungsi penilaian (*scoring function*) untuk memperkirakan kekuatan dan afinitas ikatan antara ligan dan protein, memungkinkan seleksi dan peringkat senyawa kandidat obat secara efisien sebelum pengujian eksperimental [37]. Selama proses docking, baik ligan maupun reseptor dapat menyesuaikan konformasinya untuk mencapai best-fit, dan penyesuaian konformasi ini dikenal sebagai model induced-fit yang lebih realistis dibandingkan model kunci-kunci (*lock-and-key*) yang kaku [38].

Dalam alur penelitian farmakologi jaringan, molecular docking berfungsi sebagai tahap validasi *in silico* terhadap prediksi interaksi senyawa bioaktif dengan protein target yang diidentifikasi dari analisis jaringan PPI [39]. Energi bebas ikatan yang dihasilkan dari simulasi docking memberikan estimasi kuantitatif terhadap afinitas pengikatan, di mana nilai negatif yang lebih besar mengindikasikan interaksi yang lebih kuat dan stabil [39]. Visualisasi ikatan hidrogen, interaksi hidrofobik, dan kontak van der Waals antara ligan dan residu asam amino pada situs aktif protein memberikan pemahaman mekanistik yang penting tentang bagaimana senyawa herbal memodulasi fungsi protein target terkait leukemia [37].