

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Salah satu penyakit yang memiliki tingkat kematian tinggi adalah kanker, dan jumlah penderitanya terus bertambah setiap tahun. Selama tiga dekade terakhir, klasifikasi kanker telah berkembang, tetapi pendekatan umum belum mengidentifikasi jenis kanker (Putri et al., 2020). Setiap tahun, 1,2 hingga 1,56 juta orang meninggal karena kanker paru-paru. Jumlah ini diproyeksikan akan meningkat. Diperkirakan bahwa pada tahun 2030, kanker paru-paru akan mencapai angka puncaknya dan menyebabkan kematian sekitar 17 juta orang (Jawas & Sentana, 2018). Berdasarkan informasi dari *Cancer Information & Support Center (CISC)*, persentase kematian yang disebabkan oleh kanker paru-paru di Indonesia mencapai 88% (Luo et al., 2019).

Kanker adalah istilah yang dipakai untuk menggambarkan kelompok penyakit di mana sel-sel tumbuh secara tidak terkendali dan merusak jaringan normal di dalam tubuh. Proses ini dikenal sebagai transformasi seluler, yaitu ketika sel-sel yang seharusnya berkembang dengan teratur menjadi tidak normal dan terus berkembang tanpa batas. Sel-sel kanker ini bisa menyerang dan merusak jaringan serta organ di sekitarnya, dan juga dapat menyebar ke bagian tubuh lainnya, yang dikenal dengan istilah metastasis (Prasetio & Susanti, 2019). Perkembangan sel kanker dapat menyebar dengan cepat ke seluruh bagian tubuh, mulai dari aliran darah atau sistem limfatik hingga kematian. Ini bisa terjadi karena cacat genetik (Rusamsi et al., 2018). Saat ini, pembedahan, terapi radiasi, dan kemoterapi adalah pilihan pengobatan utama untuk kanker, tetapi obat herbal masih dipertimbangkan sebagai pengobatan tambahan (Golbeck et al., 2018).

Pendekatan eksperimental untuk mempelajari interaksi protein-protein

membutuhkan banyak sumber daya seperti waktu, tenaga, dan uang. Oleh karena itu, pendekatan komputasi, termasuk metode komputasi, dapat menjadi alternatif untuk memprediksi interaksi protein-protein. Pendekatan ini menggunakan komputer dan tidak membutuhkan banyak sumber daya. *Deep learning* merupakan salah satu komponen yang terdapat dalam teknik machine learning yang melibatkan banyak lapisan (*layer*) untuk memproses informasi *nonlinier* yang bertujuan untuk melakukan ekstraksi fitur dan transformasi, baik terpadu (*supervised*) maupun tidak terpadu (*unsupervised*), serta analisis dan klasifikasi pola (Deng & Yu, 2014). Metode ini baik diterapkan pada beberapa domain *big data*, seperti *computer vision*, pengenalan suara, dan penentuan keputusan karena dapat menangani data dengan karakteristik yang kompleks (Krizhevsky et al., 2012). Metode *deep learning* juga telah diterapkan pada penelitian bioinformatika, terutama untuk memprediksi interaksi protein-protein. (Sun et al., 2017) menjadikan fitur-fitur dari pengolahan sekuens protein dengan metode *autocovariance* dan *conjoint triad* sebagai masukan untuk *stacked autoencoder*. (Hashemifar et al., 2018) menggunakan convolutional neural network (CNN) yang memberikan keluaran nilai biner yang mengindikasikan ada atau tidaknya interaksi pada sepasang protein berdasarkan masukan vektor profil pasangan protein tersebut. (Xu et al., 2020) menggabungkan penduga *autocovariance*, lokal, serta kontinu dan takkontinu multiskala sebagai masukan untuk model *deep learning* yang dinamai *ensemble deep neural network* (EnsDNN).

Obat herbal adalah obat paling populer di dunia selama tiga dekade terakhir, dan bahkan lebih dari 80% populasi global menggantungkan diri pada penggunaan obat-obatan herbal. Secara umum obat herbal terdiri dari tanaman obat seperti kunyit, daun jarak, tanaman obat akar wangi, dan lain-lain. Banyak orang telah memanfaatkan obat herbal karena memberikan potensi khasiat yang lebih baik daripada obat konvensional. Selain itu, penggunaan obat herbal dalam jangka waktu yang panjang menimbulkan dampak negatif yang lebih sedikit jika dibandingkan

dengan penggunaan obat konvensional (Ekor, 2014). Dalam proses penemuan obat herbal, menganalisis interaksi obat-target adalah salah satu proses penting. Penentuan interaksi antara target dan obat harus dilakukan secara eksperimental di laboratorium. Sebagai konsekuensinya, diperlukan usaha yang sangat besar seperti biaya yang mahal, waktu proses yang lama, dan tugas yang relatif sulit. Oleh karena itu, menganalisis interaksi obat-target sangat menantang (Chen et al., 2016). Interaksi target obat menggambarkan hubungan antara senyawa dan protein. Masing-masing mewakili calon obat herbal dan penyakit sasaran masing-masing. Faktanya, informasi interaksi obat-target yang terkait dengan obat herbal masih sedikit. Salah satu solusinya adalah dengan menghasilkan model yang dapat memprediksi interaksi antara obat dan protein sebagai target. Penelitian ini mengusulkan model prediksi interaksi obat-target menggunakan pendekatan *machine learning*.

Protein memiliki peranan yang krusial dalam melakukan fungsi seluler pada organisme hidup. Untuk melakukan fungsi tersebut, protein sering berinteraksi dengan protein lain (Athanasios et al., 2017). Interaksi ini dikenal sebagai interaksi protein-protein (*protein-protein interaction*) dan dapat diwakili oleh jaringan *node* (*vertex*) yang mewakili protein dan tepi (*edge*) yang mewakili interaksi protein-protein (Pyrogova I, 2017). Dengan mengetahui lokasinya dalam jaringan PPI, PPI dapat dimanfaatkan untuk menggambarkan peran protein dalam fungsi-fungsinya. Posisi protein dalam jaringan interaksi protein (PPI) mempengaruhi signifikansi perannya dalam jaringan tersebut. Mendeteksi protein-protein penting ini memiliki manfaat yang beragam, termasuk dalam upaya penemuan obat-obatan baru (Athanasios et al., 2017).

Model *deep learning* yang digunakan untuk memprediksi interaksi protein-protein pada penelitian ini adalah *Recurrent Neural Network* (RNN) dengan banyak lapisan (*layer*). RNN merupakan sebuah tipe arsitektur jaringan saraf tiruan yang melakukan pemrosesan berulang pada data input sekuensial. RNN masuk ke dalam

kelompok deep learning karena pengolahan data dilakukan melalui beberapa tingkatan yang berbeda. RNN telah merevolusi bidang seperti Natural language processing (NLP) dengan *deep learning* (Liu et al., 2016). *Long Short Term Memory* (LSTM) adalah sebuah komponen dalam struktur RNN yang merupakan jenis jaringan berulang. LSTM memiliki kemampuan yang lebih baik dalam mengenali pola yang diberikan dan digunakan untuk mengatur dan mengoptimalkan bobot dalam jaringan, sehingga memudahkan proses klasifikasi (Zhou et al., 2015).

Berdasarkan penjelasan dari latar belakang yang telah diuraikan, maka dilakukan penelitian dengan data interaksi protein-protein kanker yang akan didapatkan dari pangkalan data OMIM dan data interaksi protein-protein yang berasal dari pangkalan data STRING kemudian akan dimodelkan menggunakan metode *Recurrent Neural Network* dan dievaluasi dengan metode *Mean Square Error*. Interaksi protein-protein selanjutnya akan dianalisis untuk mendapatkan protein yang signifikan dan dapat memberikan rujukan tanaman obat potensial untuk penyakit kanker.

## 1.2. Rumusan Masalah

Berdasarkan uraian latar belakang diatas, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

- (a) Bagaimana memodelkan interaksi protein-protein yang berpotensi mempengaruhi perkembangan kanker dengan model *Recurrent Neural Network* (RNN)?
- (b) Bagaimana mengevaluasi model *Recurrent Neural Network* (RNN) berdasarkan data interaksi protein-protein dengan metode *Mean Square Error* (MSE)?

## 1.3. Tujuan Penelitian

Berdasarkan rumusan masalah diatas, maka tujuan dari penelitian ini adalah:

- (a) Memodelkan interaksi protein-protein yang berpotensi mempengaruhi perkembangan kanker dengan model *Recurrent Neural Network* (RNN).
- (b) Mengevaluasi model *Recurrent Neural Network* (RNN) berdasarkan data interaksi protein-protein dengan metode Mean Square Error (MSE).

#### 1.4. Ruang Lingkup Penelitian

Agar penelitian ini tidak keluar dari rumusan masalah, maka peneliti membuat ruang lingkup penelitian sebagai berikut:

- (a) Penelitian ini menggunakan data penyakit kanker paru-paru yang tersedia pada pangkalan data STRING (<https://string-db.org/>) dan tanaman obat yang tersedia pada IJAH Analytics (<https://ijah.apps.cs.ipb.ac.id/>) yang diakses pada Mei 2023.
- (b) Penelitian ini hanya mencakup proses *screening*, yaitu identifikasi senyawa potensial yang ingin dijadikan obat.
- (c) Penelitian ini tidak melakukan proses *molecular docking* untuk melihat efektivitas senyawa ketika berkaitan dengan protein.
- (d) Hasil pemodelan interaksi protein-protein dari model *Recurrent Neural Network* (RNN) yang diperoleh pada penelitian ini tidak diuji di laboratorium basah.

#### 1.5. Manfaat Penelitian

Berdasarkan permasalahan dan juga tujuan penelitian, maka pada penelitian ini ada beberapa manfaat yang diharapkan oleh penulis, yaitu:

- (1) Manfaat bagi penulis
  - (a) Memberikan pengetahuan yang lebih luas lagi kepada penulis dalam melakukan pemodelan interaksi protein-protein dari tanaman obat pada penyakit kanker menggunakan metode *Recurrent Neural Network* (RNN).
  - (b) Sebagai salah satu syarat agar penulis menyelesaikan gelar Strata satu (S1).

(c) Hasil penelitian ini juga dapat menjadi referensi kepada penulis ketika ingin melanjutkan penelitian di masa yang akan datang.

(2) Manfaat bagi Universitas

(a) Mengetahui kemampuan mahasiswa dalam penerapan dan penguasaan ilmu yang diperoleh selama berada di bangku perkuliahan dan dapat dijadikan sebagai bahan evaluasi.

(b) Menjadi tolak ukur bagi universitas untuk menentukan keberhasilan dan juga kemampuan peneliti dalam mengimplementasikan ilmu yang telah diperoleh pada saat perkuliahan.

(3) Manfaat bagi masyarakat

(a) Memberikan informasi kepada masyarakat dalam melakukan pemodelan interaksi protein-protein dari tanaman obat pada penyakit kanker menggunakan metode *Recurrent Neural Network* (RNN).

(b) Memberikan informasi yang dapat digunakan sebagai referensi penelitian tahap selanjutnya.