PETIR: Jurnal Pengkajian dan Penerapan Teknik Informatika
Vol. 17, No. 2, September 2024, P-ISSN 1978-9262, E-ISSN 2655-5018
DOI: https://doi.org/10.33322/petir.v17i2.2556

Optimalisasi Prediksi Afinitas Interaksi Obat-Target dengan Graph Neural Network dan Attention Mechanism

Husni Fadhilah[1)]; Pawesi Siantika[1]; Dimmas Mulya[1]; Putri Saptawati[1]*

1. School of Electrical Engineering and Informatics Bandung Institute of Technology, Bandung, Jl. Ganesa No.10, Lb. Siliwangi, Coblong, Bandung, Jawa Barat 40132 Indonesia

[*)]Email: husnifadhilah62@gmail.com

Received: 30 Juli 2024 | Accepted: 18 Desember 2024 | Published: 10 Januari 2025

ABSTRACT

Virtual drug screening plays a crucial role in enhancing discovery throughput and minimizing R&D costs. Deep learning emerges as a promising solution, offering efficient outcomes without requiring extensive domain expertise or structural details. This study introduces iGanDTA, an improvement from a versatile model capable of accurately predicting drug-target binding affinities with high accuracy and classifying interactions with excellent performance. Using a residual graph neural network, iGanDTA processes fingerprint data from compounds to distinguish the level of binding within protein sequences. Evaluation on benchmark datasets demonstrates superior performance compared to existing methods, with iGanDTA achieving MSE, CIndex, and R2 scores of 0.238, 0.894, and 0.710 on the Davis dataset and 0.181, 0.864, and 0.746 on the KIBA dataset.

Keywords: *Drug–target interaction; graph neural networks; attention mechanism; drug SMILES; target protein; deep learning*

ABSTRAK

Virtual screening pada obat memainkan peran penting dalam meningkatkan throughput penemuan dan mengurangi biaya R&D. Deep learning muncul sebagai solusi menjanjikan, menawarkan hasil yang efisien tanpa memerlukan keahlian domain yang luas atau detail struktural. Studi ini memperkenalkan iGanDTA, sebuah perbaikan dari model multitask yang mampu memprediksi afinitas pengikatan obat-target dengan akurasi tinggi dan mengklasifikasikan interaksi dengan performa yang sangat baik. Menggunakan residual graph neural network, iGanDTA memproses data fingerprint dari senyawa untuk membedakan tingkat pengikatan dalam urutan protein. Evaluasi pada dataset benchmark menunjukkan performa yang superior dibandingkan dengan metode yang ada, dengan iGanDTA mencapai skor MSE, CIndex, dan R2 masing-masing 0.238, 0.894, dan 0.710 pada dataset Davis serta 0.181, 0.864, dan 0.746 pada dataset KIBA.

Kata Kunci: *Interaksi obat-target; graph neural network; mekanisme attention; SMILES obat; target protein; deep learning*

Vol. 17, No. 2, September 2024, P-ISSN 1978-9262, E-ISSN 2655-5018
DOI: https://doi.org/10.33322/petir.v17i2.2556

1. PENDAHULUAN

Proses penemuan obat baru adalah suatu proses yang rumit dan mahal, melibatkan penelitian dan pengembangan yang luas yang memakan waktu bertahun-tahun dan memerlukan investasi finansial yang signifikan. Prosedur standar untuk memperoleh persetujuan obat baru biasanya memerlukan biaya sekitar $2,8 miliar dan memakan waktu antara 10 hingga 15 tahun [1, 2]. Proses ini biasanya dimulai dengan identifikasi kandidat obat yang menjanjikan melalui berbagai metode penyaringan, diikuti dengan pengujian dan validasi yang ketat untuk menilai profil efektivitas dan keamanannya [3]. Namun, tingkat kegagalan yang signifikan dan waktu pengembangan yang panjang yang melekat dalam penemuan obat menekankan kebutuhan akan metode prediksi yang lebih efisien dan tepat [4].

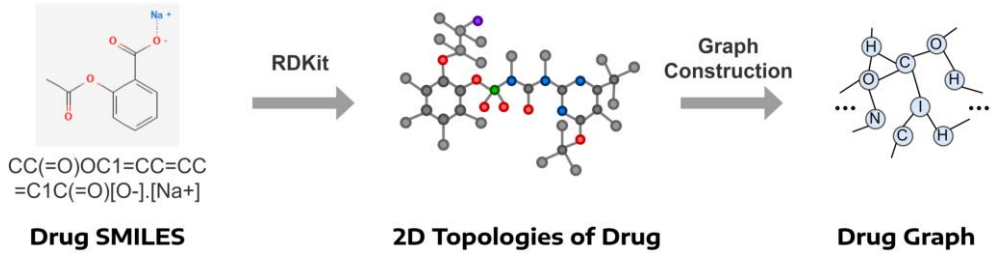
Memprediksi interaksi obat-target atau *drug-target interaction* (DTI) adalah aspek krusial dalam penemuan obat, penting untuk mengidentifikasi senyawa terapeutik potensial dan mengoptimalkan efektivitas pengobatan [5]. Pendekatan tradisional untuk prediksi DTI umumnya mengandalkan metode klasifikasi biner, sering kali mengabaikan aspek kekuatan interaksi [6]. Namun, kemajuan terbaru menekankan pentingnya mempertimbangkan afinitas pengikatan, yang memberikan wawasan kuantitatif mengenai kekuatan interaksi antara obat dan protein targetnya. Biasanya, obat yang diinginkan menunjukkan afinitas yang lebih kuat terhadap molekul targetnya dan afinitas yang lebih lemah terhadap molekul yang tidak relevan [3]. Akibatnya, tugas regresi untuk memprediksi afinitas obat-target atau *drug-target affinity* (DTA) telah menjadi fokus penting dalam penemuan obat dan upaya reposisi obat [2].

Proses penemuan obat melibatkan pemeriksaan menyeluruh terhadap interaksi antara molekul senyawa kecil dan protein untuk menyaring obat potensial. Proses ini memakan waktu dan intensif karena banyaknya kandidat obat yang perlu dipelajari [7]. Prediksi DTI, cukup penting untuk tugas klasifikasi biner dan regresi, secara tradisional mengandalkan simulasi dan pemodelan molekuler [8, 9]. Untuk mengatasi tantangan ini, metode komputasi untuk *virtual screening* menawarkan solusi yang menjanjikan [3]. Metode komputasi alternatif telah muncul, diklasifikasikan menjadi pendekatan berbasis ligan dan berbasis struktur [10]. Metode berbasis ligan mengandalkan kemiripan kimia di antara ligan [11], sementara metode berbasis struktur, seperti simulasi pemodelan, memanfaatkan struktur tiga dimensi untuk prediksi, meskipun memerlukan investasi waktu yang signifikan [12]. Metode berbasis *sequence*, yang menggunakan *deep learning*, telah mendapatkan perhatian karena kemampuannya untuk menghasilkan karakteristik dari urutan protein dan obat, sedangkan metode sebelumnya mengandalkan ekstraksi fitur [13]. Meskipun demikian, kemajuan terbaru dalam *deep learning*, khususnya *convolutional neural network* (CNN) [14, 15] dan *recurrent neural network* (RNN) [16], telah menunjukkan potensi yang cukup besar dalam menangkap representasi fitur yang disederhanakan dari urutan protein dan struktur molekuler. Meskipun efektif, pendekatan ini sering kali memerlukan dataset yang luas untuk mencapai kinerja puncak [3].

Dalam skenario ini, *graph neural network* (GNN) muncul sebagai kerangka kerja yang menjanjikan untuk memprediksi afinitas DTI, memanfaatkan struktur graf yang ada dalam senyawa molekuler dan protein target [9]. Tidak seperti metode tradisional yang memperlakukan molekul dan protein sebagai urutan linier atau representasi statis, GNN dapat menangkap informasi relasional kompleks dan karakteristik struktural yang melekat dalam interaksi biomolekuler [10, 17]. Dengan memanfaatkan informasi yang terkandung dalam node dan hubungan mereka, jaringan ini secara cermat mengekstrak detail fitur penting, memungkinkan identifikasi dan prediksi yang akurat dari node atau koneksi [18]. Selain itu, mempertimbangkan korelasi dan keragaman yang signifikan dalam data biologis, graf molekuler muncul sebagai representasi luar biasa dari informasi biologis,

Vol. 17, No. 2, September 2024, P-ISSN 1978-9262, E-ISSN 2655-5018
DOI: https://doi.org/10.33322/petir.v17i2.2556

terutama menunjukkan struktur molekuler dan asosiasi fungsional antara molekul. Model GNN menonjol sebagai pendekatan paling efektif untuk membedakan sifat molekuler, karena mereka mengasimilasi data lokal dan global dari jaringan dengan mengonsolidasi informasi tetangga [17].



Gambar 1. Proses ilustrasi pembuatan grafik molekul obat dari representasi SMILES

Berdasarkan kemajuan ini, model perbaikan dari GanDTI [3] yang disebut iGanDTA *(improved Graph attention network for Drug Target Affinity prediction)* dikembangkan dalam studi ini, yang menggabungkan GNN dengan modul *attention* untuk prediksi afinitas DTI. Dengan memanfaatkan struktur graf molekul dan protein serta mengintegrasikan mekanisme *attention*, tujuan model ini adalah untuk memprediksi dengan tepat afinitas pengikatan antara obat dan protein target, sehingga membantu proses penemuan obat. Melalui eksperimen dan evaluasi menyeluruh pada dataset KIBA dan Davis, efektivitas pendekatan yang diusulkan ditunjukkan dalam prediksi afinitas DTI yang tepat.

2. METODE PENELITIAN

2.1. Pengumpulan Data

Pengumpulan data memegang peranan penting dalam setiap penelitian, berfungsi sebagai dasar untuk analisis dan pemahaman. Dalam penelitian ini, data dikumpulkan secara cermat dari dua dataset yang terpercaya dan banyak digunakan dalam domain prediksi afinitas interaksi obat-target (DTI) yaitu dataset Davis [22] dan dataset KIBA *(Kinase Inhibitor BioActivity)* [23].

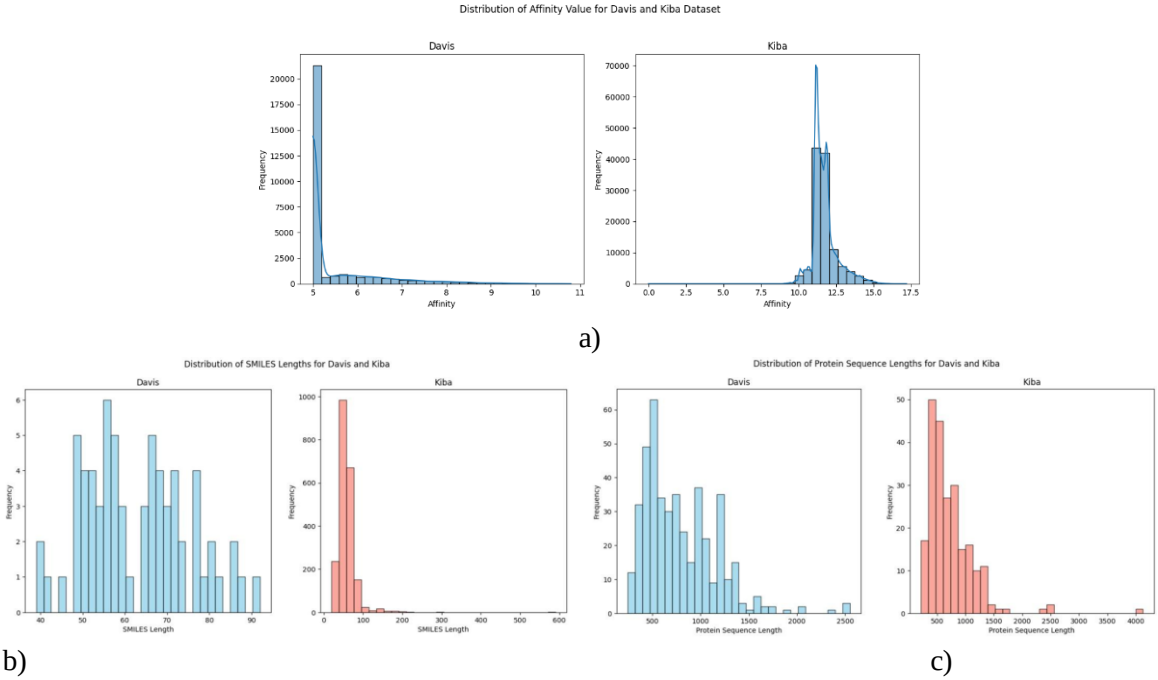
Dataset Davis, yang diperkenalkan oleh Davis dkk [22], terdiri dari uji selektivitas keluarga protein kinase dan penghambatnya masing-masing, bersama dengan nilai konstanta disosiasi (K_d). Dataset ini mencakup interaksi antara 442 protein dan 68 ligan, dengan total 30.056 nilai afinitas, memberikan perspektif yang komprehensif tentang interaksi kinase-penghambat. Sebaliknya, dataset KIBA, yang dibuat oleh Tang et al. [23], menggabungkan data bioaktivitas penghambat kinase dari berbagai sumber, termasuk nilai K_i (konstanta penghambat), K_d (konstanta disosiasi), dan IC_{50} (konsentrasi penghambat setengah maksimum). Database KIBA mencakup 52.498 obat dan 467 target, dengan total 246.088 skor KIBA. Namun, SimBoost [24] menerapkan prosedur filtrasi pada database ini. Akibatnya, database KIBA yang telah disaring kini mencakup 2.111 obat dan 229 protein target, dengan total 118.254 skor yang mewakili bioaktivitas obat-target.

Untuk memastikan kualitas dan konsistensi dataset, protokol yang ditetapkan untuk pra pemrosesan data diikuti. Untuk dataset Davis, nilai K_d diubah ke ruang logaritmik (pK_d) dalam persamaan 1, sesuai dengan praktik umum di bidang ini [14].

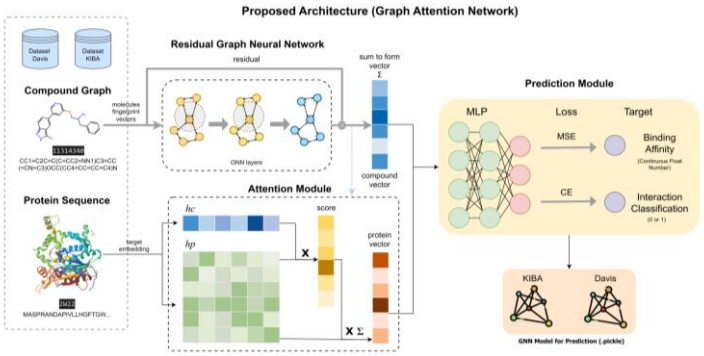
$$pK_d = -log_{10} \frac{K_d}{10^9} \quad (1)$$

Vol. 17, No. 2, September 2024, P-ISSN 1978-9262, E-ISSN 2655-5018
DOI: https://doi.org/10.33322/petir.v17i2.2556

Langkah-langkah pra-pemrosesan serupa diterapkan untuk menstandarisasi skor KIBA untuk analisis. Untuk kedua dataset, string dalam format SMILES diperoleh dari database PubChem, menawarkan gambaran standar tentang struktur kimia. Selain itu, urutan protein diperoleh dari database protein UniProt, memastikan representasi yang akurat dan andal dari protein target. Gambar 4 menunjukkan contoh data PDB dari Protein Data Bank dan representasi peta kontak untuk setiap dataset Davis dan KIBA. Peta kontak menggambarkan kedekatan spasial antara atom dalam kompleks protein-ligan, membantu dalam analisis interaksi mereka. Sepanjang proses pengumpulan data, transparansi dan reproduksibilitas diprioritaskan, dengan setiap langkah didokumentasikan secara cermat untuk memfasilitasi upaya penelitian di masa depan. Dengan memanfaatkan dataset yang telah disaring ini, analisis dan evaluasi yang kuat dari model yang diusulkan untuk prediksi afinitas DTI dilakukan.



Gambar 2. Ilustrasi dataset Davis (di sisi kiri) dan dataset KIBA (di sisi kanan), yang menampilkan: a) pola distribusi nilai afinitas pengikatan dalam setiap dataset, b) pola distribusi panjang string SMILES, dan c) distribusi panjang urutan protein.

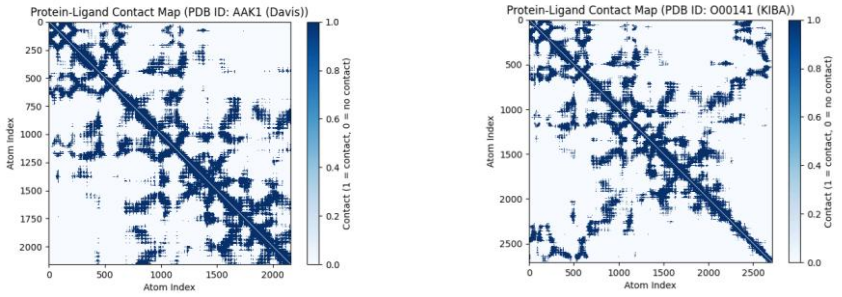


Gambar 3. Ilustrasi yang menunjukkan struktur jaringan saraf graf (GNN) untuk pemrosesan molekul senyawa, termasuk pembentukan vektor fingerprint dari molekul dan embedding target protein. Matriks dan operasi vektor dari modul attention menyoroti perannya dalam menilai pengaruh urutan protein. Pada akhirnya, modul prediksi menandakan tahap akhir untuk membuat prediksi afinitas berdasarkan keluaran mekanisme attention.

PETIR: Jurnal Pengkajian dan Penerapan Teknik Informatika
Vol. 17, No. 2, September 2024, P-ISSN 1978-9262, E-ISSN 2655-5018
DOI: https://doi.org/10.33322/petir.v17i2.2556

Tabel 1. Statistik Ringkas Dataset Davis dan KIBA

Dataset	Jumlah *Compounds*	Jumlah *Sequence* Protein	Jumlah *Interactions*
Davis	68	442	30056
KIBA	2111	229	118254



a) b)

Gambar 4. Contoh ilustrasi data PDB beserta representasi peta kontaknya dengan jarak *cutoff* = 17 untuk kedua dataset: a) Davis (ID PDB: AAK1) dan b) KIBA (ID PDB: 000141)

2.2. Arsitektur Pipeline Model

Dalam merancang model untuk memprediksi afinitas pengikatan antara senyawa dan protein, metode dari GanDTI [3] diterapkan, menggabungkan komponen utama seperti Modul *Load* Data, *Residual Graph Neural Network* (GNN), Modul *Attention*, dan Modul Prediksi. Seperti yang diilustrasikan dalam Gambar 3, struktur jaringan saraf graf disesuaikan untuk menganalisis molekul senyawa. Jaringan ini mencakup tahap pembentukan vektor *fingerprint* molekul dan *embedding* target dari protein. Vektor *fingerprint* molekul dihasilkan melalui serangkaian transformasi yang melibatkan atom dan ikatan senyawa, yang pada akhirnya menghasilkan representasi kompak yang menangkap fitur-fitur penting dari molekul tersebut. Demikian pula, *embedding* target dari protein melibatkan konversi urutan protein menjadi urutan vektor, memungkinkan eksplorasi asosiasi antara vektor senyawa dan urutan protein.

Selain itu, ilustrasi tersebut menggambarkan proses operasi matriks dan vektor dalam modul *attention*. Modul ini memfasilitasi penyelidikan bagian mana dari urutan protein yang memberikan pengaruh lebih besar pada interaksi dengan senyawa. Awalnya, baik vektor senyawa maupun protein ditransformasikan ke ruang lain menggunakan matriks bobot yang dapat dipelajari dan menggunakan fungsi aktivasi ReLU. Selanjutnya, *attention dot-product* diterapkan pada vektor yang telah ditransformasi ini, memungkinkan model untuk berkonsentrasi pada segmen-segmen yang relevan dari urutan protein terkait dengan senyawa.

Menuju akhir ilustrasi, modul prediksi digambarkan, di mana menggambarkan fase akhir dari desain model. Dalam modul ini, keluaran dari mekanisme *attention*, yang terdiri dari fitur senyawa dan protein, digunakan untuk membuat prediksi tentang afinitas interaksi obat-target. Selain itu, pembentukan model keseluruhan, yang mencakup integrasi jaringan saraf graf, mekanisme *attention*, dan modul prediksi, menggambarkan pendekatan komprehensif yang diterapkan untuk prediksi afinitas interaksi obat-target yang akurat.

2.2.1. Modul *Load* Data

Dalam pengembangan model prediksi afinitas obat-target, Modul *Load* Data memainkan peran penting dalam memuat data yang telah diproses untuk digunakan dalam pelatihan dan evaluasi model. Tahap pra-pemrosesan data dilakukan melalui modul pra-pemrosesan, yang

Vol. 17, No. 2, September 2024, P-ISSN 1978-9262, E-ISSN 2655-5018
DOI: https://doi.org/10.33322/petir.v17i2.2556

menerima urutan target dan representasi struktur molekul dalam format SMILES sebagai input. Langkah-langkah pra-pemrosesan ini melibatkan pembuatan *fingerprint* molekul, matriks ketetanggaan *(adjacency)*, dan pemisahan urutan target menjadi kata-kata menggunakan *n-gram* yang telah ditentukan. Hasil yang diproses kemudian dimuat dalam format yang sesuai dengan kebutuhan model, seperti tensor untuk memfasilitasi perhitungan model yang efisien.

Setelah data diproses, langkah berikutnya melibatkan pemuatan kamus yang berisi *fingerprint* molekul dan fitur kata yang digunakan dalam dataset. Proses ini akan membaca file *pickle* yang berisi kamus tersebut. Kamus-kamus ini diperlukan untuk menerapkan *fingerprint* molekul dan fitur kata dalam model. Selanjutnya, data yang telah diproses dan kamus fitur yang dimuat dikonversi menjadi tensor. Tensor-tensor ini digunakan dalam perhitungan model untuk memperkirakan afinitas antara obat dan target.

2.2.2. Modul *Residual* GNN

Residual GNN diimplementasikan untuk menangani struktur kompleks graf molekul dan protein. Metode ini memungkinkan kita untuk mengekstrak data dari berbagai tingkat representasi dalam graf dan memperbarui fitur node secara iteratif menggunakan fungsi pembaruan dan penyampaian pesan *(message passing)*. Dengan demikian, model ini dapat memahami hubungan antara berbagai entitas dalam graf dan menangkap informasi relevan untuk memprediksi afinitas pengikatan. Formula utama yang diterapkan pada modul *Residual* GNN digambarkan dalam persamaan 2.

$$c_i^{(l+1)} = U\left(c_i^{(l)},\ \sum m\left(c_j^{(l)}\right)\right) \tag{2}$$

Di sini $c_i^{(l)}$ mewakili fitur *fingerprint* molekul senyawa pada node i dan *edge* dalam lapisan tersembunyi (l). Struktur molekul digambarkan sebagai graf, dengan atom sebagai node dan ikatan sebagai *edge*. Setiap node mencakup atribut seperti tipe atom, jumlah atom hidrogen, *aromaticity*, dan valensi. Konektivitas antar node ditetapkan menggunakan matriks *adjacency* (A), yang menunjukkan atom mana yang terikat. Mekanisme penyampaian pesan (m) digunakan untuk memperbarui informasi pesan dari sebuah node ke node tetangganya. Bersama dengan fungsi pembaruan (u), informasi dalam graf akan digabungkan. Modul ini memperkenalkan konsep bobot *attention* dalam langkah penyampaian pesan.

$$\sum m = f_1\left(Wt_g A c_i^{(l)}\right) \tag{3}$$

Ini melibatkan penggunaan matriks berbobot Wt_g dan fungsi aktivasi non-linear f_1, dengan fungsi aktivasi leaky ReLU digunakan untuk meningkatkan kinerja model. Sub-lapisan terpisah t_g digunakan untuk menghitung bobot *attention*. Bobot-bobot ini dihitung berdasarkan fitur node yang telah ditransformasi. Fungsi sigmoid f_2 memastikan bobot berada di antara 0 dan 1.

$$Wt_g = f_2(W_g) \tag{4}$$

Perbedaan kuncinya terletak di sini. Bobot *attention* Wt_g dikalikan secara *element-wise* dengan representasi tersembunyi $c_i^{(l)}$ sebelum didistribusikan melalui matriks *adjacency* A. Ini memungkinkan jaringan untuk fokus pada pesan-pesan yang lebih relevan dari node tetangga berdasarkan skor *attention* yang telah dipelajari. Setelah tiga lapisan iterasi, fitur *fingerprint* awal atau *residual* digabungkan dengan cara *feed-forward* seperti yang dijelaskan di bawah ini:

$$c_i^{(l)} = c_i^{(l)} + c_i^{(1)} \tag{5}$$

Variabel $c_i^{(1)}$ di lapisan pertama melambangkan fitur *fingerprint* molekul senyawa. Atribut ini mencerminkan jaringan saraf *residual* yang sering digunakan dalam pengenalan gambar.

Vol. 17, No. 2, September 2024, P-ISSN 1978-9262, E-ISSN 2655-5018
DOI: https://doi.org/10.33322/petir.v17i2.2556

Akhirnya, fitur *fingerprint* yang diperbarui digabungkan melalui penjumlahan untuk membentuk fitur senyawa akhir.

$$c_e = \sum_{i \in compound} c_i^{(l)} \quad (6)$$

Di sini, c_e adalah vektor senyawa yang telah digabungkan secara keseluruhan. Bentuk penggabungan ini adalah teknik yang umum digunakan dalam Jaringan Saraf Graf (GNN).

2.2.3. Modul *Attention*

Selanjutnya, Modul *Attention* diintegrasikan ke dalam arsitektur untuk menyoroti bagian penting dari urutan protein yang berkontribusi pada interaksi dengan senyawa. Penggunaan mekanisme *attention* dalam model ini memungkinkan model untuk fokus pada fitur-fitur paling relevan dalam urutan protein, sehingga meningkatkan akurasi prediksi afinitas pengikatan.

Awalnya, informasi protein dikodekan menjadi representasi berurutan yang dilambangkan oleh $P = (p_{i,\dots}p_n)$, di mana setiap p_i mewakili vektor dalam ruang embedding berdimensi n. Untuk mengeksplorasi bagaimana vektor senyawa c berinteraksi dengan urutan protein, mekanisme *attention* digunakan untuk menentukan segmen mana dari urutan protein yang berkontribusi lebih menonjol pada interaksi. Dalam skenario ini, vektor senyawa berfungsi sebagai vektor kunci, sementara vektor protein berfungsi sebagai vektor nilai. Baik vektor senyawa maupun protein menjalani proses transformasi ke ruang yang berbeda, seperti dijelaskan di bawah ini:

$$h_c = f(W_a c_e) \quad (7)$$

$$h_p = f(W_a P) \quad (8)$$

Fitur senyawa yang ditransformasikan, dilambangkan sebagai h_c, dan fitur protein, dilambangkan sebagai h_p, mengalami transformasi menggunakan fungsi ReLU yang dilambangkan dengan f, dengan W_a mewakili matriks bobot yang dapat dilatih melalui proses *looping*. *Loop* ini diulang melalui beberapa lapisan linear, di mana di dalam *loop*, fitur senyawa ditransformasikan menggunakan sub-lapisan pertama. Sub-lapisan kedua dari setiap lapisan linear menghitung skor *attention*. Pertama, *attention* diterapkan dalam lapisan linear. Di setiap lapisan linear, transformasi dan perhitungan skor *attention* dilakukan sebagai berikut:

Setiap lapisan linear terdiri dari dua komponen transformasi: satu untuk menghasilkan representasi tersembunyi senyawa h_c, dan yang lainnya untuk menghasilkan skor *attention* $\alpha^{(l)}$. Representasi tersembunyi h_c dihitung dengan menerapkan fungsi aktivasi $tanh$ pada transformasi linear dari senyawa:

$$h_c^{(l)} = tanh\left(W_g^{(l)} c_e\right) \quad (9)$$

di mana $W_g^{(l)} c_e$ adalah matriks bobot untuk lapisan linear ke-l, dan c_e adalah vektor fitur senyawa awal. Skor *attention* $\alpha^{(l)}$ dihitung dengan menerapkan fungsi aktivasi sigmoid σ pada representasi tersembunyi $h_c^{(l)}$:

$$\alpha^{(l)} = \sigma\left(W_\alpha^{(l)} h_{(c)}^{(l)}\right) \quad (10)$$

di mana $W_\alpha^{(l)}$ adalah matriks bobot untuk menghitung skor *attention* di lapisan linear ke-l. Skor *attention* ini kemudian digunakan untuk memperbarui representasi senyawa dengan perkalian *element-wise* antara senyawa dan skor *attention*:

$$c_e^{(l+1)} = c_e^{(l)} \odot \alpha^{(l)} \quad (11)$$

di mana $\odot$ menunjukkan perkalian *element-wise*, dan $c_e^{(l)}$ adalah vektor fitur senyawa di lapisan ke- l. Setelah diproses melalui semua lapisan linear, skor *attention* dari semua lapisan digabungkan dengan mengambil rata-rata:

Vol. 17, No. 2, September 2024, P-ISSN 1978-9262, E-ISSN 2655-5018
DOI: https://doi.org/10.33322/petir.v17i2.2556

$$\alpha_{agg} = \frac{1}{L}\sum_{l=1}^{L} \alpha^{(l)} \quad (12)$$

di mana L adalah jumlah lapisan linear. Selanjutnya, skor *attention* yang digabungkan ini digunakan untuk menimbang urutan protein. Protein diperbarui dengan perkalian *element-wise* antara skor *attention* yang ditransposisikan dengan protein:

$$P_{\alpha^{(l)}} = \alpha_{agg} \odot h_p \quad (13)$$

di mana h_p adalah vektor fitur protein awal. Akhirnya, vektor fitur protein yang diberi bobot ini dirata-ratakan sepanjang dimensi urutan untuk membentuk vektor fitur protein akhir:

$$p_e = \frac{1}{N}\sum_{i=1}^{N} P_{\alpha^{(i)}} \quad (14)$$

di mana N adalah panjang urutan protein, dan $P_{\alpha^{(i)}}$ mewakili elemen ke-i dalam vektor fitur protein berbobot P_α. Kemudian menambahkan dimensi ke vektor fitur protein yang dirata-ratakan:

$$p_e = [p_e] \quad (15)$$

di mana p_e kini diperlakukan sebagai vektor kolom, secara efektif menambahkan dimensi ekstra untuk memastikan kompatibilitasnya dengan operasi berikutnya dalam model.

2.2.4. Modul Prediksi

Pada bagian akhir, Modul Prediksi diimplementasikan untuk menggabungkan fitur-fitur yang diekstrak dari *Residual* GNN dan Modul *Attention* dan menghasilkan prediksi akhir untuk afinitas pengikatan. Modul ini menggunakan jaringan saraf untuk memetakan fitur-fitur ini untuk memprediksi nilai yang menggambarkan kekuatan interaksi antara senyawa dan protein.

Vektor senyawa c_e dan vektor fitur protein p_e keduanya mengandung informasi yang cukup untuk memprediksi afinitas pengikatan. Selanjutnya, vektor-vektor ini digabungkan untuk menghasilkan vektor yang akan diproses melalui *Multilayer Perceptron* (MLP). Proses ini dapat digambarkan sebagai berikut:

$$o = MLP([c_e, p_e]) \quad (16)$$

Di sini, o mewakili vektor output, dan $[.;..]$ menunjukkan penggabungan. Dengan demikian, seluruh masalah dapat digambarkan secara singkat oleh fungsi berikut:

$$o = g(f(c),\ h(p, c) \quad (17)$$

Dalam persamaan 18, f melambangkan komponen GNN, h mewakili modul *attention*, dan g menunjukkan bagian MLP. Model ini menangani baik prediksi afinitas pengikatan maupun klasifikasi interaksi, oleh karena itu fungsi *loss*-nya dijelaskan secara terpisah. Tujuan dari tugas afinitas pengikatan adalah untuk meminimalkan kesalahan kuadrat rata-rata, yang dinyatakan sebagai:

$$L_{MSE} = \frac{1}{n}\sum_{i=1}^{n}(o_i - y_i)^2 + \frac{\lambda}{2}\,||\Theta||^2 \quad (18)$$

di mana y_i mewakili nilai yang diamati atau aktual, sementara o_i adalah nilai yang diprediksi. Θ mencakup penggabungan bobot dan bias dalam jaringan, dengan λ mewakili *hyperparameter* regularisasi $L2$. Dengan menggabungkan ketiga komponen ini, sebuah model yang kuat dan efektif telah dikembangkan untuk memprediksi afinitas pengikatan antara senyawa dan protein, memberikan dukungan berharga dalam penemuan obat dan pembuatan terapi yang lebih efisien.

2.3. Metrik Evaluasi

Metrik evaluasi seperti *mean squared error* (MSE), *concordance index* (CI), dan koefisien determinasi (${r^2}_m$) digunakan untuk menilai korelasi antara prediksi dan nilai kebenaran dasar, yang menunjukkan keunggulan dan efektivitas pendekatan yang diusulkan dalam prediksi afinitas obat-

Vol. 17, No. 2, September 2024, P-ISSN 1978-9262, E-ISSN 2655-5018
DOI: https://doi.org/10.33322/petir.v17i2.2556

target dibandingkan dengan metode dan tolok ukur yang ada.

MSE, sebagai fungsi kerugian dalam persamaan 19, digunakan untuk mengukur seberapa dekat prediksi model regresi dengan nilai aktual.

$$MSE(\mathrm{y},\mathrm{p}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - p_i)^2 \quad (19)$$

Di sini, nilai eksperimen y_i mewakili titik data ke-i, sedangkan nilai prediksi p_i sesuai dengan titik data yang sama. *Concordance Index* (CI) [24] dalam persamaan 20 berfungsi sebagai metrik evaluasi tambahan, yang mengukur kesepakatan dalam peringkat kekuatan afinitas pengikatan antara nilai yang diprediksi dan nilai aktual untuk pasangan obat dan target yang dipilih secara acak.

$$CI = \frac{1}{Z}\sum_{y_i>y_j} h(b_i - b_j) \quad (20)$$

Persamaan 21 melibatkan b_i dan b_j sebagai nilai prediksi untuk afinitas yang lebih besar dan lebih kecil. Z mewakili konstanta normalisasi, dan $h(x)$ menunjukkan fungsi langkah [25]:

$$h(x) = \begin{cases} 1, & x > 0 \\ 0.5, & x = 0 \\ 0, & x < 0 \end{cases} \quad (21)$$

$r^2{}_m$ digunakan sebagai metrik evaluasi untuk kinerja prediksi eksternal model.

$$r^2{}_m = r^2 * \left(1 - \sqrt{r^2 - \gamma_0{}^2}\right) \quad (22)$$

Kuadrat dari koefisien korelasi r^2 dan $\gamma_0{}^2$ menilai asosiasi antara nilai observasi dan prediksi, mempertimbangkan dengan dan tanpa istilah intercept, masing-masing. Model yang dapat diandalkan hanya ditunjukkan ketika $r^2{}_m$ lebih besar dari 0.5.

3. HASIL DAN PEMBAHASAN

Pada bagian ini, efektivitas model dalam memprediksi afinitas pengikatan dinilai. Model idealnya harus mendekati nilai yang diprediksi dengan nilai yang diukur dengan kesalahan minimal, sebaiknya membentuk garis linear dengan $R = 1$.

3.1. Pengaturan Eksperimen

Efektivitas model GNN dalam memprediksi afinitas pengikatan dievaluasi menggunakan GPU Nvidia Tesla P100 yang kuat yang tersedia di Kaggle *Kernels*. Sumber daya perangkat keras mutakhir ini memiliki memori HBM2 berkecepatan tinggi 16 GB dan memberikan daya komputasi yang luar biasa, secara signifikan mempercepat proses pelatihan dibandingkan dengan CPU. Efektivitas model dalam memprediksi afinitas pengikatan dinilai, dengan eksperimen dilakukan menggunakan berbagai *hyperparameter* untuk mengoptimalkan kinerja dalam prediksi afinitas interaksi obat-target.

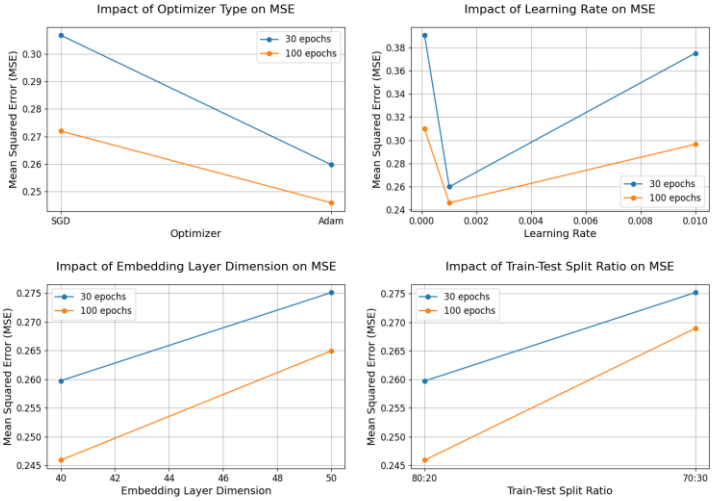
Dalam eksperimen ini, kami menggunakan radius (r) sebesar 2 untuk membangun grafik berbasis jarak dalam protein. Penelitian ini juga memanfaatkan urutan protein 3-gram, memungkinkan untuk mempertimbangkan konteks sekitar dalam urutan asam amino. Selain itu, ukuran *depth* pada *Graph Attention Network* (GAT) diatur menjadi 3, memungkinkan model untuk mengeksplorasi hubungan yang lebih dalam pada grafik. Untuk MLP yang digunakan sebagai bagian dari model ini, nilai *depth* diatur menjadi 2, yang telah terbukti optimal dalam penelitian sebelumnya. Tabel 2 menunjukkan pengaturan eksperimen untuk penyetelan *hyperparameter* dalam penelitian ini.

Vol. 17, No. 2, September 2024, P-ISSN 1978-9262, E-ISSN 2655-5018
DOI: https://doi.org/10.33322/petir.v17i2.2556

Tabel 2. Penyetelan *Hyperparameter* pada penelitian

No	*Hyperparameter*	Nilai
1	*Epochs*	[30,100]
2	*Optimizer*	[SGD, Adam]
3	Laju pembelajaran *(learning rate)*	[0.0001, 0.001, 0.01]
4	*Weight decay*	[10^6]
5	Dimensi lapisan *embedding* (fitur)	[30, 40, and 50]
6	Rasio pembagian data *train-test*	[80:20, 70:30]

3.2. Hasil Eksperimen dan Diskusi



Gambar 5. Contoh hasil penyetelan *hyperparameter* pada Dataset Davis

Hasil eksperimen yang ditunjukkan dalam Gambar 5 memberikan wawasan signifikan mengenai kinerja model di bawah berbagai konfigurasi *hyperparameter*. Pada berbagai jumlah *epoch*, jenis *optimizer*, dan laju pembelajaran, terlihat tren yang bervariasi dalam konvergensi model dan akurasi prediksi. Teramati bahwa menambah jumlah *epoch* umumnya meningkatkan kinerja model, meskipun dengan hasil yang semakin menurun setelah mencapai ambang batas tertentu. Misalnya, peningkatan *epoch* dari 30 ke 100 mengakibatkan penurunan MSE dari 0,25973 menjadi 0,24590 (perbaikan 5,34%) untuk *optimizer* Adam, dan dari 0,30678 menjadi 0,27199 (perbaikan 11,34%) untuk *optimizer* SGD.

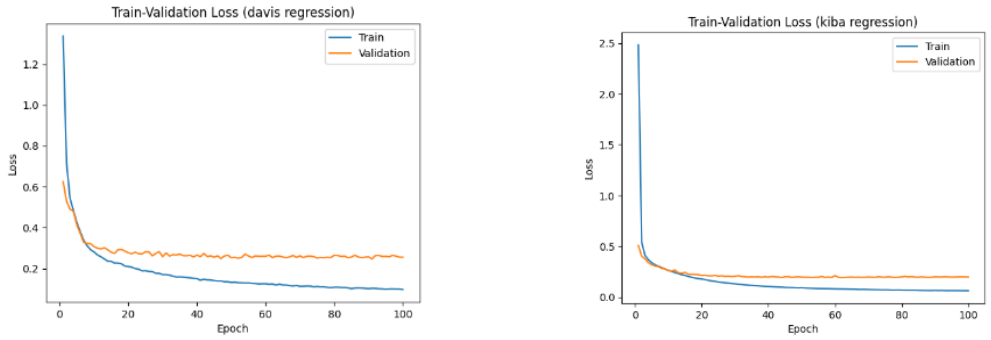
Mengenai jenis *optimizer*, baik Adam maupun SGD menunjukkan perilaku yang berbeda dalam mengoptimalkan parameter model. Sementara Adam umumnya lebih cepat konvergen, SGD menunjukkan stabilitas yang lebih baik selama periode pelatihan yang lebih lama. Pada 100 *epoch*, Adam mencapai MSE yang lebih rendah (0,24590) dibandingkan dengan SGD (0,27199), menunjukkan bahwa Adam lebih efisien dalam mengoptimalkan parameter model dalam rentang *epoch* yang diberikan. Pilihan laju pembelajaran mempengaruhi kecepatan konvergensi model dan akurasi prediksi akhir secara signifikan. Laju pembelajaran 0,001 secara konsisten menunjukkan kinerja terbaik, dengan MSE sebesar 0,25973 pada 30 *epoch* dan 0,24590 pada 100 *epoch*. Sebaliknya, laju pembelajaran 0,0001 menghasilkan MSE yang lebih tinggi (0,39064 pada 30 *epoch* dan 0,30979 pada 100 *epoch*), menyoroti konvergensi yang lebih lambat. Laju pembelajaran

Vol. 17, No. 2, September 2024, P-ISSN 1978-9262, E-ISSN 2655-5018
DOI: https://doi.org/10.33322/petir.v17i2.2556

yang lebih tinggi sebesar 0,01 mengakibatkan pelatihan yang tidak stabil, dengan MSE sebesar 0,37515 pada 30 *epoch* dan 0,29652 pada 100 *epoch*.

Selain itu, dimensi lapisan *embedding* memainkan peran penting dalam menangkap dan merepresentasikan fitur dasar dari senyawa dan protein. Dimensi *embedding* sebesar 40 menghasilkan MSE sebesar 0,25973 pada 30 *epoch* dan 0,24590 pada 100 *epoch*, sementara peningkatan dimensi ke 50 menghasilkan MSE masing-masing sebesar 0,27509 dan 0,26495. Ini menunjukkan bahwa *embedding* dengan dimensi lebih tinggi, meskipun sedikit meningkatkan kinerja prediktif, juga meningkatkan kompleksitas komputasi tanpa peningkatan yang proporsional. Rasio pembagian *train-test* mempengaruhi kapasitas model untuk menggeneralisasi data yang tidak dikenal. Pembagian 80:20 secara konsisten menghasilkan nilai MSE yang lebih baik (0,25973 pada 30 *epoch* dan 0,24590 pada 100 *epoch*) dibandingkan dengan pembagian 70:30 (0,27519 pada 30 *epoch* dan 0,26894 pada 100 *epoch*), menunjukkan bahwa proporsi set pelatihan yang lebih besar membantu dalam generalisasi yang lebih baik.

Berdasarkan temuan eksperimen, *hyperparameter* optimal untuk memprediksi afinitas interaksi obat-target ditentukan. Evaluasi yang ketat terhadap berbagai konfigurasi mengungkapkan bahwa penggunaan 100 *epoch*, *optimizer* Adam dengan laju pembelajaran 0,001, dimensi lapisan *embedding* 40, dan rasio pembagian *train-test* 80:20 memberikan hasil yang paling menguntungkan. Selain itu, penerapan penurunan bobot sebesar 10^6 secara signifikan meningkatkan kinerja model. Temuan ini menekankan pentingnya penyetelan *hyperparameter* dalam mengoptimalkan kemampuan prediktif model untuk prediksi afinitas interaksi obat-target. Dengan memilih kombinasi *hyperparameter* yang sesuai, akurasi dan kekuatan model prediktif dapat ditingkatkan secara efektif, sehingga memajukan bidang penemuan obat dan memfasilitasi identifikasi senyawa terapeutik potensial.



a) b)

Gambar 6. *Loss* pada tahap pelatihan dan validasi selama proses pelatihan model dari (a) dataset Davis dan (b) dataset KIBA

Secara keseluruhan, eksperimen ini menyoroti pentingnya penyetelan *hyperparameter* dalam mengoptimalkan kinerja model jaringan saraf grafis untuk prediksi afinitas interaksi obat-target. Eksplorasi sistematis berbagai konfigurasi *hyperparameter* memberikan wawasan berharga tentang faktor-faktor yang mempengaruhi kinerja model dan mengidentifikasi pengaturan optimal untuk mencapai akurasi prediktif mutakhir seperti yang ditunjukkan dalam Tabel 3 dan Tabel 4.

Tabel 3. Perbandingan Kinerja dengan Model lain pada Dataset Davis

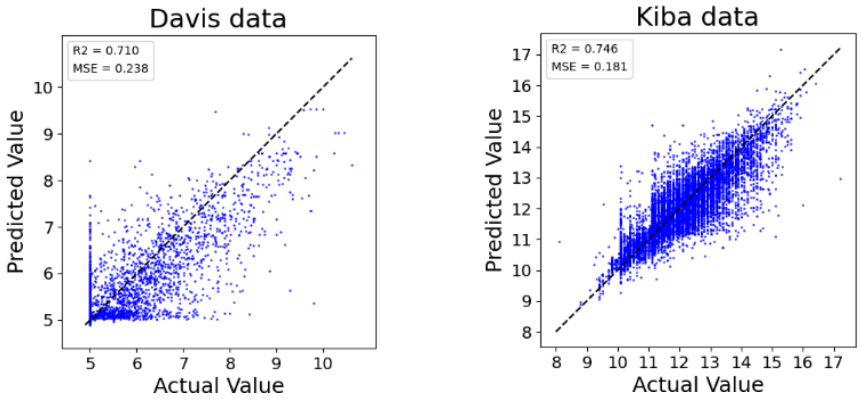
Model	presentasi Protein	Representasi Obat	MSE	CIndex	r2m

Vol. 17, No. 2, September 2024, P-ISSN 1978-9262, E-ISSN 2655-5018
DOI: https://doi.org/10.33322/petir.v17i2.2556

KronRLS	Smith-Waterman	PubChem-Sim	0.379	0.871	0.407
SimBoost	Smith-Waterman	PubChem-Sim	0.282	0.872	0.644
DeepDTA	1D-Subseq	1D	0.283	0.871	0.634
DeepGS	Amino acid seq.	SMILES seq. + Mol. graph	0.252	0.882	0.686
GLFA-Split avg	Protein graph	Mol graph	0.241	0.886	0.699
GEFA-Split avg	Protein graph	Mol graph	0.250	0.887	0.688
GraphDTA-GCNNet	1D-Subseq	Mol. graph	0.293	0.863	0.635
iGanDTA	1D-Subseq	Mol. graph	**0.238**	**0.894**	**0.710**

Tabel 4. Perbandingan Kinerja dengan Model lain pada Dataset KIBA

Model	presentasi Protein	Representasi Obat	MSE	CIndex	r2m
KronRLS	Smith-Waterman	PubChem-Sim	0.411	0.782	0.342
SimBoost	Smith-Waterman	PubChem-Sim	0.222	0.836	0.629
DeepDTA	1D-Subseq	1D	0.194	0.863	0.673
DeepGS	Amino acid seq.	SMILES seq. + Mol. graph	0.193	0.860	0.684
GLFA-Split avg	Protein graph	Mol graph	0.215	0.858	0.673
GEFA-Split avg	Protein graph	Mol graph	0.217	0.855	0.669
GraphDTA-GCNNet	1D-Subseq	Mol. graph	0.188	0.862	0.724
iGanDTA	1D-Subseq	Mol. graph	**0.181**	**0.864**	**0.746**



b)

Gambar 7. Perbandingan antara korelasi nilai afinitas yang diprediksi dan diukur untuk dataset a) Davis dan b. KIBA

iGanDTA dibandingkan dengan metode mutakhir yang ada termasuk KronRLS [25], SimBoost [24], DeepDTA [14], DeepGS [26], GEFA dan GLFA [27], serta GraphDTA-GCNNet [9]. Hasil dari Tabel 3 menunjukkan bahwa iGanDTA mengungguli semua model lainnya pada dataset Davis, mencapai *Mean Squared Error* (MSE) sebesar 0,238, *Concordance Index* (CIndex) sebesar 0,894, dan skor $r^2{}_m$ sebesar 0,710. Demikian pula, pada dataset KIBA (Tabel 4), iGanDTA melampaui model lain dengan MSE sebesar 0,181, CIndex sebesar 0,864, dan skor $r^2{}_m$ sebesar

Vol. 17, No. 2, September 2024, P-ISSN 1978-9262, E-ISSN 2655-5018
DOI: https://doi.org/10.33322/petir.v17i2.2556

0,746. Kinerja superior iGanDTA dapat dikaitkan dengan arsitekturnya, yang mengintegrasikan *Residual Graph Neural Network* (GNN) dan *Attention Module*.

GNN sangat efektif dalam menangkap informasi struktural dan hubungan antara node dalam representasi grafis, memberikan representasi fitur yang lebih kaya dan berarti dibandingkan dengan representasi berbasis urutan. iGanDTA dapat menggambarkan koneksi struktural, domain fungsional, dan atribut multimodal yang ada antara obat dan protein, menghasilkan prediksi yang lebih akurat dan dapat diinterpretasikan. Mekanisme *attention* dalam GNN memungkinkan model untuk menyoroti bagian-bagian penting dari urutan protein yang berinteraksi dengan molekul obat, sehingga meningkatkan akurasi prediksi. Mekanisme ini bekerja dengan memberikan bobot lebih besar pada interaksi yang lebih relevan, membantu dalam penemuan mekanisme di balik interaksi obat-protein dan menentukan sub-komponen fungsional yang relevan. Selain itu, penggunaan graf molekuler yang mencakup informasi struktural dan kimia tentang obat, serta peta kontak protein, memberikan pandangan yang lebih komprehensif tentang interaksi ini. Temuan ini menegaskan potensi iGanDTA dalam merevolusi proses penemuan obat dengan memberikan prediksi yang lebih akurat tentang interaksi obat-target.

4. KESIMPULAN

Skrining obat *virtual* yang efisien sangat penting untuk meningkatkan throughput penemuan dan mengurangi biaya R&D. *Deep learning* menunjukkan potensi yang signifikan dalam hal ini, karena dapat menghasilkan hasil yang menjanjikan tanpa memerlukan keahlian domain atau data struktural yang rumit dalam waktu yang singkat. Di sini, kami memperkenalkan model serbaguna dan sederhana yang disebut iGanDTA, yang mampu memprediksi afinitas pengikatan dan mengklasifikasikan interaksi dengan kinerja yang luar biasa. Model ini menggunakan *residual graph neural network* untuk menangani informasi *fingerprint* senyawa, menghasilkan vektor yang memanfaatkan *attention* berbasis *dot-product* pada urutan protein untuk mengevaluasi signifikansi pengikatan dalam urutan tersebut. Selanjutnya, kedua set data digabungkan untuk diproses menggunakan *Multilayer Perceptron* (MLP). Evaluasi pada dua dataset *benchmark* yang mapan menunjukkan bahwa model ini melampaui metode *deep learning* mutakhir di berbagai metrik.

Upaya masa depan dalam studi ini dapat berfokus pada peningkatan keterbacaan modelnya. Mengembangkan metode untuk memvisualisasikan kontribusi atom atau residu individu dalam GNN dan mekanisme *attention* dapat memberikan wawasan yang lebih dalam tentang interaksi obat-target. Transparansi yang ditingkatkan ini akan membantu peneliti untuk lebih memahami fitur molekuler spesifik yang mempengaruhi afinitas pengikatan, yang mengarah pada desain dan optimasi obat yang lebih terinformasi.

DAFTAR PUSTAKA

[1] J. A. DiMasi, “Research and Development Costs of New Drugs,” *JAMA*, vol. 324, no. 5, p. 517, Aug. 2020, doi: 10.1001/jama.2020.8648.

[2] H. Wu *et al.*, “AttentionMGT-DTA: A multi-modal drug-target affinity prediction using graph transformer and attention mechanism,” *Neural Networks*, vol. 169, no. September 2023, pp. 623–636, 2024, doi: 10.1016/j.neunet.2023.11.018.

[3] S. Wang, P. Shan, Y. Zhao, and L. Zuo, “GanDTI: A multi-task neural network for drug-target interaction prediction,” *Comput. Biol. Chem.*, vol. 92, no. December 2020, p. 107476, 2021, doi: 10.1016/j.compbiolchem.2021.107476.

[4] S. Bonner *et al.*, “A review of biomedical datasets relating to drug discovery: a knowledge graph perspective,” *Brief. Bioinform.*, vol. 23, no. 6, pp. 1–19, 2022, doi:

10.1093/bib/bbac404.

[5] B. C. Zhiqin Zhu, Zheng Yao, Guanqiu Qi, Neal Mazur, Pan Yang, "CAAI Trans on Intel Tech - 2023 - Zhu - Associative learning mechanism for drug-target interaction prediction.pdf," *CAAI Transactions on Intelligence Technology*. John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology and Chongqing University of Technology, pp. 1558–1577, 2023. doi: 10.1049/cit2.12194.

[6] N. R. C. Monteiro, J. L. Oliveira, and J. P. Arrais, "TAG-DTA: Binding-region-guided strategy to predict drug-target affinity using transformers," *Expert Syst. Appl.*, vol. 238, no. PE, p. 122334, 2024, doi: 10.1016/j.eswa.2023.122334.

[7] Q. Zhao, H. Zhao, K. Zheng, and J. Wang, "HyperAttentionDTI: improving drug-protein interaction prediction by sequence-based deep learning with attention mechanism," *Bioinformatics*, vol. 38, no. 3, pp. 655–662, 2022, doi: 10.1093/bioinformatics/btab715.

[8] M. Yazdani-Jahromi *et al.*, "AttentionSiteDTI: An interpretable graph-based model for drug-Target interaction prediction using NLP sentence-level relation classification," *Brief. Bioinform.*, vol. 23, no. 4, pp. 1–14, 2022, doi: 10.1093/bib/bbac272.

[9] T. Nguyen, H. Le, T. P. Quinn, T. Nguyen, T. D. Le, and S. Venkatesh, "GraphDTA: Predicting drug target binding affinity with graph neural networks," *Bioinformatics*, vol. 37, no. 8, pp. 1140–1147, 2021, doi: 10.1093/bioinformatics/btaa921.

[10] Y. Zhang, Y. Hu, N. Han, A. Yang, X. Liu, and H. Cai, "A survey of drug-target interaction and affinity prediction methods via graph neural networks," *Comput. Biol. Med.*, vol. 163, no. May, p. 107136, 2023, doi: 10.1016/j.compbiomed.2023.107136.

[11] H. Qi, T. Yu, W. Yu, and C. Liu, "Drug–target affinity prediction with extended graph learning-convolutional networks," *BMC Bioinformatics*, vol. 25, no. 1, pp. 1–21, 2024, doi: 10.1186/s12859-024-05698-6.

[12] A. C. Cheng *et al.*, "Structure-based maximal affinity model predicts small-molecule druggability," *Nat. Biotechnol.*, vol. 25, no. 1, pp. 71–75, 2007, doi: 10.1038/nbt1273.

[13] L. Chen *et al.*, "TransformerCPI: Improving compound-protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments," *Bioinformatics*, vol. 36, no. 16, pp. 4406–4414, 2020, doi: 10.1093/bioinformatics/btaa524.

[14] H. Öztürk, A. Özgür, and E. Ozkirimli, "DeepDTA: Deep drug-target binding affinity prediction," *Bioinformatics*, vol. 34, no. 17, pp. i821–i829, 2018, doi: 10.1093/bioinformatics/bty593.

[15] S. Hu, C. Zhang, P. Chen, P. Gu, J. Zhang, and B. Wang, "Predicting drug-target interactions from drug structure and protein sequence using novel convolutional neural networks," *BMC Bioinformatics*, vol. 20, no. Suppl 25, pp. 1–12, 2019, doi: 10.1186/s12859-019-3263-x.

[16] F. Wan *et al.*, "DeepCPI: A Deep Learning-based Framework for Large-scale in silico Drug Screening," *Genomics, Proteomics Bioinforma.*, vol. 17, no. 5, pp. 478–495, 2019, doi: 10.1016/j.gpb.2019.04.003.

[17] S. G. Paul, A. Saha, M. Z. Hasan, S. R. H. Noori, and A. Moustafa, "A Systematic Review of Graph Neural Network in Healthcare-Based Applications: Recent Advances, Trends, and Future Directions," *IEEE Access*, vol. 12, no. February, pp. 15145–15170, 2024, doi: 10.1109/ACCESS.2024.3354809.

[18] Z. Zhang, P. Cui, and W. Zhu, "Deep Learning on Graphs: A Survey," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 1, pp. 249–270, 2022, doi: 10.1109/TKDE.2020.2981333.

[19] Z. Yang, W. Zhong, L. Zhao, and C. Yu-Chian Chen, "MGraphDTA: Deep multiscale graph

Vol. 17, No. 2, September 2024, P-ISSN 1978-9262, E-ISSN 2655-5018
DOI: https://doi.org/10.33322/petir.v17i2.2556

neural network for explainable drug-target binding affinity prediction," *Chem. Sci.*, vol. 13, no. 3, pp. 816–833, 2022, doi: 10.1039/d1sc05180f.

[20] H. He, G. Chen, and C. Y. C. Chen, "NHGNN-DTA: a node-adaptive hybrid graph neural network for interpretable drug-target binding affinity prediction," *Bioinformatics*, vol. 39, no. 6, 2023, doi: 10.1093/bioinformatics/btad355.

[21] D. Hu, "An Introductory Survey on Attention Mechanisms in NLP Problems BT - Intelligent Systems and Applications," Y. Bi, R. Bhatia, and S. Kapoor, Eds., Cham: Springer International Publishing, 2020, pp. 432–448.

[22] M. I. Davis *et al.*, "Comprehensive analysis of kinase inhibitor selectivity," *Nat. Biotechnol.*, vol. 29, no. 11, pp. 1046–1051, 2011, doi: 10.1038/nbt.1990.

[23] J. Tang *et al.*, "Making Sense of Large-Scale Kinase Inhibitor Bioactivity Data Sets: A Comparative and Integrative Analysis," *J. Chem. Inf. Model.*, vol. 54, no. 3, pp. 735–743, Mar. 2014, doi: 10.1021/ci400709d.

[24] T. He, M. Heidemeyer, F. Ban, A. Cherkasov, and M. Ester, "SimBoost: a read-across approach for predicting drug-target binding affinities using gradient boosting machines," *J. Cheminform.*, vol. 9, no. 1, pp. 1–14, 2017, doi: 10.1186/s13321-017-0209-z.

[25] T. Pahikkala *et al.*, "Toward more realistic drug-target interaction predictions," *Brief. Bioinform.*, vol. 16, no. 2, pp. 325–337, 2015, doi: 10.1093/bib/bbu010.

[26] X. Lin, K. Zhao, T. Xiao, Z. Quan, Z. J. Wang, and P. S. Yu, "Deepgs: Deep representation learning of graphs and sequences for drug-target binding affinity prediction," *Front. Artif. Intell. Appl.*, vol. 325, no. i, pp. 1301–1308, 2020, doi: 10.3233/FAIA200232.

[27] T. M. Nguyen, T. Nguyen, T. M. Le, and T. Tran, "GEFA: Early Fusion Approach in Drug-Target Affinity Prediction," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 19, no. 2, pp. 718–728, 2022, doi: 10.1109/TCBB.2021.3094217.