
Prediksi Bioaktivitas Penghambat Protease NS3 sebagai Anti Virus Dengue Berbasis In-Silico Menggunakan LSTM - Algoritma Firefly

GirasMatahari¹, IsmanKurniawan²

^{1,2}Fakultas Informatika, Universitas Telkom, Bandung

¹girasaja@student.telkomuniversity.ac.id, ²ismankrn@telkomuniversity.ac.id

Abstrak

Dengue adalah penyakit yang disebabkan oleh virus yang ditularkan melalui nyamuk *Aedes* dan menjadi masalah kesehatan global dengan jutaan kasus setiap tahun. Salah satu target utama dalam pengembangan obat dengue adalah NS3 protease, protein esensial yang berperan dalam replikasi virus, namun pendekatan konvensional masih menghadapi keterbatasan biaya dan efektivitas. Teknologi kecerdasan buatan, khususnya metode *Machine Learning* telah menunjukkan potensi besar dalam mempercepat penemuan obat dengan meningkatkan efisiensi dan akurasi prediksi. Penelitian ini bertujuan untuk memprediksi bioaktivitas senyawa dengan menggunakan pendekatan berbasis kecerdasan buatan. Metode yang digunakan mencakup pendekatan in-silico untuk bioaktivitas, di mana model *Long Short-Term Memory* (LSTM) dioptimasi menggunakan *Algoritma Firefly*. Hasil terbaik dicapai pada skenario optimasi dengan nilai beta 0,8, yang menghasilkan model 3-layer dengan konfigurasi [125, 37, 73] dan fungsi aktivasi sigmoid. Model ini mencapai akurasi sebesar 0,7647 dan *F1score* sebesar 0,7705 pada data uji, yang menunjukkan performanya baik dan klasifikasi yang seimbang serta konsisten antar kelas. Hasil ini membuktikan bahwa pendekatan LSTM yang dioptimasi dengan *Algoritma Firefly* dapat menyaring senyawa secara efisien dan berkontribusi pada pengembangan obat anti-dengue yang lebih efektif.

Kata kunci: Dengue, NS3 protease, Kecerdasan buatan, Machine Learning, Long Short-Term Memory (LSTM), Algoritma Firefly

Abstract

Dengue is a disease caused by a virus transmitted through *Aedes* mosquitoes and has become a global health problem with millions of cases every year. One of the main targets in the development of dengue drugs is the NS3 protease, an essential protein that plays a role in viral replication, yet conventional approaches still face limitations in cost and effectiveness. Artificial intelligence technology, particularly Machine Learning methods, has shown great potential in accelerating drug discovery by enhancing the efficiency and accuracy of predictions. This research aims to predict the bioactivity of compounds using an artificial intelligence-based approach. The methods used include in-silico approaches for bioactivity, where the Long Short-Term Memory (LSTM) model is optimized using the Firefly Algorithm. The best results were achieved in the optimization scenario with a beta value of 0,8, resulting in a 3-layer model with the configuration [125, 37, 73] and a sigmoid activation function. This model achieved an accuracy of 0,7647 and an *F1 score* of 0,7705 on the Uji data, indicating that its performance is good and the classification is balanced and consistent across classes. These results prove that the LSTM approach optimized with the Firefly Algorithm can efficiently filter compounds and contribute to the development of more effective anti-dengue drugs.

Keywords: Dengue, NS3 protease, Artificial Intelligence, Machine Learning, Long Short-Term Memory (LSTM), Firefly Algorithm

1. Pendahuluan

1.1. Latar Belakang

Dengue merupakan penyakit yang disebabkan oleh infeksi virus yang ditularkan melalui gigitan nyamuk *Aedes aegypti* atau *Aedes albopictus*. Gejalanya bervariasi, mulai dari demam dengue (DF) yang ditandai dengan demam, ruam, dan nyeri, hingga *Dengue Shock Syndrom* (DSS) yang dapat berakibat fatal jika tidak segera ditangani [1]. Infeksi yang terjadi berulang kali disebabkan oleh perilaku manusia, adaptasi mikroorganisme, serta faktor lingkungan seperti globalisasi, yang berkaitan dengan kepadatan penduduk, sanitasi yang buruk, dan meningkatnya interaksi manusia dengan vektor penyakit [2]. Pada awal tahun 2023, tercatat adanya lonjakan kasus demam berdarah dengan lebih dari 6,5 juta kasus dan 7.300 kematian, yang tersebar di lebih dari 100 negara. Wilayah yang paling terdampak meliputi Amerika, Asia Tenggara, dan Pasifik Barat, dengan Asia menyumbang sekitar 70% dari total kasus global [3].

NS3 adalah protein penting dalam virus dengue yang memiliki peran signifikan dalam proses replikasi. Bersama dengan kofaktor NS2B, NS3 membentuk enzim protease yang memecah protein virus menjadi komponen yang diperlukan untuk reproduksi. Mengingat fungsinya yang vital, NS3 menjadi sasaran utama dalam pengembangan obat antivirus untuk menghambat penyebaran dengue [4]–[6]. Pendekatan in-silico telah

diterapkan untuk memahami fungsi protein NS3 dalam replikasi virus dengue. Melalui prediksi bioinformatika dan simulasi molekuler, kita dapat mengidentifikasi pengaruh fosforilasi terhadap interaksi NS3-NS5, yang memiliki peran krusial dalam proses replikasi. Temuan ini mendukung pengembangan metode prediktif yang berbasis kecerdasan buatan untuk menilai potensi bioaktivitas senyawa antivirus [7].

Bioaktivitas senyawa terhadap virus dengue dapat dilihat dari kemampuannya dalam menghambat protease NS2B/NS3 yang memiliki peran krusial dalam replikasi virus. Uji in-vitro menunjukkan bahwa beberapa senyawa peptidomimetik memiliki potensi tinggi dengan nilai IC_{50} μ M, yang menandakan adanya peluang besar untuk pengembangannya sebagai kandidat terapi dengue [8].

Pengembangan obat konvensional untuk virus dengue mengalami berbagai tantangan, seperti biaya yang tinggi, waktu yang lama, dan keterbatasan efektivitas akibat evolusi virus yang cepat [5], [9]. Selain itu, waktu terapi yang terbatas memerlukan diagnosis dan penanganan yang cepat [5]. Oleh karena itu, pendekatan in vitro dan in-silico, terutama metode in-silico, dianggap efektif untuk mempercepat pengujian antivirus sekaligus mengurangi biaya pengembangan [4].

Kecerdasan Buatan (AI) memberikan solusi alternatif dengan meniru kemampuan analisis serta pengambilan keputusan yang dimiliki manusia [10]. Pendekatan *Machine Learning* saat ini banyak digunakan dalam pengembangan obat anti dengue karena kemampuannya dalam menganalisis data medis yang besar dan kompleks. Teknologi ini memiliki potensi untuk menurunkan biaya, mempercepat penemuan obat, dan meningkatkan efektivitas pengobatan [11].

Berbagai penelitian menunjukkan bahwa protease NS3 adalah target terapi yang menjanjikan untuk dengue. Ho et al., (2020) menerapkan metode pembelajaran mesin seperti DNN, DT, dan LR untuk mendeteksi dengue dengan AUC 85,87%, yang efektif sebagai alat deteksi dini di daerah dengan sumber daya terbatas [12]. Natsrita et al., (2024) menggabungkan *Machine Learning* dan simulasi molekuler untuk mengidentifikasi antibodi netralisasi terhadap empat serotipe dengue dengan AUC 0,885 [13]. Altayb dan Alatawi (2024) menggunakan pendekatan QSAR berbasis *Machine Learning* untuk menemukan inhibitor NS3 virus Zika, yang menghasilkan akurasi 92,5% dan F1 score 0,91 [14]. Norshidah et al., (2023) mengintegrasikan metode in-silico dan in vitro untuk mengidentifikasi inhibitor NS2B/NS3 yang menghambat replikasi virus dengue [4]. Elumalai dan Muthuvel (2022) menggunakan k-NN dan *Random Forest* untuk mengklasifikasikan peptida penghambat dengue, dengan akurasi 90,47% dan F1 score 0,89 [15]. Selain itu, Mukhtar et., (2023) mengevaluasi potensi *Nigella sativa* sebagai inhibitor dengue melalui *molecular docking* [16], dan alomair et al., (2021) menemukan bahwa quercetin dapat menghambat fosforilasi residu S137 pada NS3 yang penting dalam replikasi virus [7]. Kurniawan et al., (2020) membangun model QSAR dengan menggunakan metode *Extremely Randomized Trees* (ERT) yang memiliki akurasi 73%, AUC 0,82, dan validasi *y-scrambling* untuk memastikan keandalan model [17].

Penelitian sebelumnya menunjukkan bahwa protease NS3 memiliki potensi yang signifikan sebagai target terapeutik untuk dengue, namun pendekatan *Machine Learning* seperti *molecular docking*, QSAR, k-NN, dan *Random Forest* masih memiliki keterbatasan dengan akurasi 70-90%, sehingga belum optimal dalam mengidentifikasi inhibitor dengan afinitas tinggi. Untuk mengatasi masalah ini, penelitian ini bertujuan untuk memprediksi bioaktivitas berbasis in-silico dalam menghambat protease NS3 sebagai antivirus dengue, dengan menggunakan model *Long Short-Term Memory* (LSTM) yang dioptimasi melalui Algoritma Firefly. LSTM dipilih karena kemampuannya dalam mengelola data sekuensial dan pola yang kompleks, sedangkan FA digunakan untuk meningkatkan efisiensi pencarian solusi optimal, mempercepat proses optimal, serta kelemahan metode konvensional seperti grid search. Kombinasi antara LSTM dan FA diharapkan dapat meningkatkan akurasi serta kinerja model secara keseluruhan.

1.2. Topik dan Batasannya

Berdasarkan latar belakang yang telah dijabarkan, maka rumusan masalah yang dapat diangkat menjadi topik dan batasan adalah sebagai berikut:

1. Bagaimana performa metode baseline seperti *molecular docking* dan QSAR dalam memprediksi bioaktivitas inhibitor protease NS3 dibandingkan dengan pendekatan in-silico berbasis LSTM yang dioptimasi dengan Algoritma Firefly?
2. Bagaimana efektivitas Algoritma Firefly dalam mengoptimasi arsitektur LSTM untuk meningkatkan akurasi prediksi bioaktivitas inhibitor NS3 sebagai terapi anti-virus dengue?
3. Bagaimana performa model LSTM yang dioptimasi dengan Algoritma Firefly dalam memprediksi bioaktivitas inhibitor NS3 dibandingkan dengan model yang menggunakan metode optimasi konvensional?

1.3. Tujuan

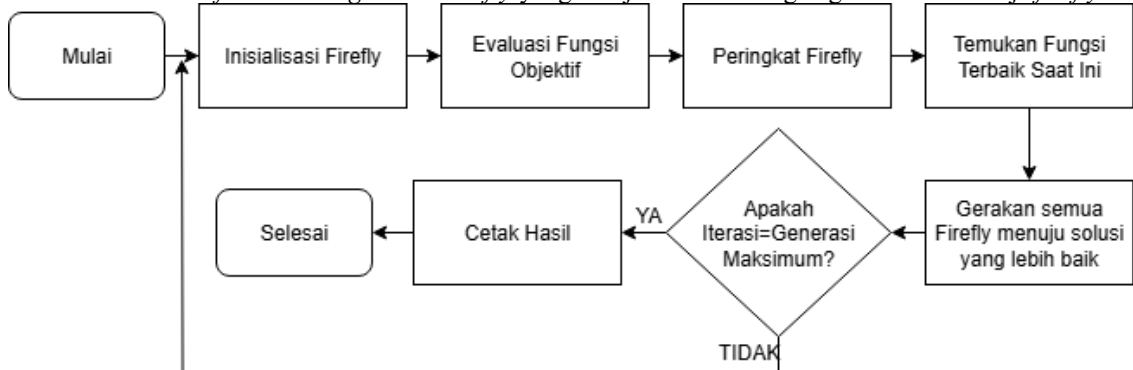
Berdasarkan topik dan batasannya yang telah dijabarkan, penelitian ini bertujuan untuk mengevaluasi performa metode baseline seperti *molecular docking* dan QSAR dalam memprediksi bioaktivitas inhibitor protease NS3, serta membandingkannya dengan pendekatan in-silico berbasis LSTM yang dioptimasi menggunakan Algoritma Firefly. Selain itu, penelitian ini juga bertujuan untuk menguji efektivitas Algoritma Firefly dalam mengoptimasi arsitektur LSTM guna meningkatkan akurasi prediksi bioaktivitas inhibitor NS3 sebagai terapi antivirus dengue. Terakhir, penelitian ini bertujuan untuk mengevaluasi performa model LSTM yang telah

dioptimasi dengan *Algoritma Firefly*, dan membandingkannya dengan model yang menggunakan metode optimasi konvensional.

2. Studi Terkait

2.1. Algoritma Firefly

Gambar 2.1 adalah *flowchart Algoritma Firefly* yang menjelaskan tentang bagaimana cara kerja *firefly*.



Gambar 2. 1 Flowchart Algoritma Firefly

Algoritma Firefly adalah algoritma optimasi yang terinspirasi dari perilaku kunang-kunang dalam mencari pasangan melalui cahaya yang mereka hasilkan. Diperkenalkan oleh Xin-She pada tahun 2007, algoritma ini termasuk dalam kategori *Nature Inspired Optimization* (NIO) dan *Swarm Intelligence* (SI). *Algoritma Firefly* telah terbukti efektif dalam menyelesaikan berbagai masalah kompleks di berbagai bidang, termasuk rekayasa biomedis dan kesehatan [18]. Algoritma ini pertama kali dikembangkan oleh Xhin-She digunakan untuk menyelesaikan berbagai masalah optimasi. Cahaya yang dipancarkan kunang-kunang bertujuan untuk menarik pasangan, dan pola cahayanya berbeda untuk setiap spesies. Dalam *Algoritma Firefly*, mekanisme ini dimodelkan sebagai cara kunang-kunang mencari solusi optimal dalam ruang pencarian [19]. Keunggulan *Algoritma Firefly* terletak pada fleksibilitasnya untuk menghasilkan solusi optimal serta kemampuannya dalam menangani masalah non-linear dan multi-model [18].

Daya Tarik (attractiveness) kunang-kunang dihitung sebagai fungsi eksponensial dari jarak antar kunang-kunang, dengan intensitas cahaya yang memudar seiring bertambahnya jarak. Dalam implementasinya, algoritma ini memiliki tiga aturan ideal, yaitu semua kunang-kunang bersifat uniseksual sehingga dapat saling tertarik, intensitas cahaya menentukan daya tarik, dan jika tidak ada kunang-kunang yang lebih terang maka kunang-kunang bergerak secara acak [19].

Algoritma ini terdiri dari beberapa komponen utama, yaitu jarak (r_{ij}) untuk mengukur kedekatan antara dua kunang-kunang, koefisien penyerap cahaya (γ) yang mengontrol reduksi cahaya seiring jarak, dan faktor randomisasi (α) untuk memperkenankan kerandoman dalam pergerakan [19]. Dalam *Algoritma Firefly* terdapat beberapa persamaan, yaitu pertama daya tarik (β) (1).

$$\beta = \beta_0 e^{-\gamma r_{ij}^2} \quad (1)$$

Di mana β_0 adalah nilai maksimum daya tarik (saat jarak = 0), γ adalah koefisien penyerap cahaya, dan r_{ij} adalah jarak antara kunang-kunang i dan j . Untuk persamaan kedua adalah pergerakan kunang-kunang (x_i) (2).

$$x_i = x_i + \beta(x_j - x_i) + \alpha(rand - 0,5) \quad (2)$$

Di mana term pertama adalah posisi kunang-kunang i . Term kedua adalah tarikan oleh kunang-kunang j berdasarkan daya tarik β . Term ketiga adalah faktor randomisasi dengan $rand$ adalah nilai acak dalam $[0, 1]$.

Dalam penerapannya di bidang *Machine Learning*, *Algoritma Firefly* dan *Long Short-Term Memory* terbukti efektif, khususnya dalam konteks pemilihan fitur. *Algoritma Firefly* dapat digunakan untuk mengidentifikasi subset fitur yang paling relevan, sehingga meningkatkan akurasi model *Machine Learning* dan mengurangi kompleksitas komputasi. Dengan demikian, *Algoritma Firefly* dapat diharapkan meningkatkan akurasi model dan mengurangi kompleksitas komputasi [19].

Berdasarkan persamaan-persamaan tersebut, penulis dapat menuliskan algoritma dari *Algoritma Firefly*, berikut.

ALGORITMA 1: ALGORITMA FIREFLY [19]

Fungsi objektif $f(x)$, di mana $x = (x_1, x_2, \dots, x_d)$

Hasilkan populasi awal firefly x_i ($i = 1, 2, \dots, n$)

Tentukan intensitas cahaya I_i di x_i melalui $f(x_i)$
