

1. Pendahuluan

Latar Belakang

Obat adalah salah satu solusi utama untuk menangani penyakit pada makhluk hidup, akan tetapi mengonsumsi obat tersebut dapat mengarah pada gejala toksisitas [1]. Toksisitas obat memiliki efek yang sangat berbahaya, salah satu kasus yang disebabkan oleh toksisitas obat adalah karsinogenik, teratogenik, mutagenik, dan lain-lain [1], [2]. Oleh karena itu, toksisitas masih menjadi pendorong utama kegagalan kandidat obat dalam pengembangan obat, yang mengakibatkan tingginya biaya obat yang berhasil masuk ke pasar [3]. Sering kali, potensi masalah dan efek samping dari suatu obat baru ditemukan pada tahap akhir uji klinis, ketika banyak uang telah diinvestasikan ke dalam obat tersebut. Faktanya, sekitar 90% kandidat obat gagal hanya pada tahap uji klinis [4], dan juga dalam beberapa tahun terakhir, perkiraan biaya untuk membawa obat baru ke pasar telah mencapai sekitar \$ 1,8 miliar USD [5]. Akibatnya, terjadi peningkatan kebutuhan untuk dapat mengukur dengan akurat dan efisien mengenai toksisitas klinik dari kandidat obat baru terhadap manusia [3].

Pendekatan konvensional seperti pengujian laboratorium basah, validasi, dan prosedur sintesis tidak hanya mahal tetapi juga memakan waktu, yang menjadikannya kurang efisien dalam proses penemuan obat [6]. Oleh karena itu, muncul kebutuhan mendesak untuk mengembangkan metode alternatif yang lebih efisien dan akurat dalam mengevaluasi toksisitas klinik kandidat obat baru. Dalam konteks ini, salah satu pendekatan alternatif yang dapat digunakan adalah penggunaan teknik *In Silico* misalnya metode *Machine Learning* (ML), yang menawarkan cara yang lebih cepat dan lebih hemat biaya untuk simulasi dan pengukuran toksisitas klinik obat.

Telah dilakukan beberapa penelitian terkait implementasi *machine learning* pada prediksi toksisitas. Pada Tahun 2013 S. Gupta dan kolega model berbasis dipeptide untuk prediksi toksisitas peptida dan protein, dengan akurasi mencapai 0.9450 [7]. Pada tahun 2015, C. Gunavathi dan kolega melakukan penelitian pemilihan fitur dalam klasifikasi kanker menggunakan data ekspresi gen mikroarray dengan model klasifikasi kNN dan Cuckoo Search untuk seleksi fitur, mencapai akurasi rata-rata 100% pada beberapa dataset kanker, termasuk DLBCL Harvard, Lung Michigan, Ovarian Cancer, AML-ALL, dan Lung Harvard2 [8]. Pada Tahun 2017 E. Kim dan Kolega mengembangkan model prediksi hepatotoksitas yang diinduksi oleh obat dengan memanfaatkan sidik jari molekuler yang tertimbang menggunakan model *Random Forest* (RF) dan *Support Vector Machine* (SVM). Hasil validasi silang menunjukkan akurasi 0.7380 dan 0.7260 serta AUC 0.7910 dan 0.7680, sedangkan pada pengujian independen akurasi mencapai 0.6010 untuk RF dan 0.6110 untuk SVM [9]. Pada tahun 2019 L. Pu mengembangkan *eToxPred*, pendekatan berbasis *machine learning* untuk memperkirakan toksisitas kandidat obat dengan menggunakan *Deep Belief Networks* (DBN) dan *Extremely Randomized Trees* (ET), menghasilkan akurasi 0.7200 [10]. Pada Tahun 2023 R. R Rizwandy mengeksplorasi prediksi toksisitas obat menggunakan data masukan sidik jari molekuler yang akurat menggunakan metode *Artificial Neural Networks* (ANN) dan Teknik Optimasi *Cuckoo Search*, dengan *F1-Score* sebesar 0.6153 dan akurasi sebesar 0.9652 [11].

Lebih baru, pada tahun 2024, F. Siddique dan kolega mengulas kembali derivatif metotreksat dan fototreksat dengan pendekatan desain obat *in-silico* berbasis *deep learning* menggunakan model QSAR, yang mencatat *R squared* 0.8100 dan *Q squared* 0.7700 [12]. Pada tahun yang sama Y. Zhou dan kolega mengeksplorasi prediksi toksisitas okular senyawa menggunakan metode *deep learning* GCN dan *Random Forest*, dengan AUC 0.9150 untuk GCN dan 0.8690 untuk *Random Forest* [13]. Pada tahun 2024 Fahad Mostafa dan meneliti prediksi cedera hati akibat obat menggunakan *Random Forest* dan *MultiLayer Perceptron*, mencapai akurasi 0.9090 dan *Matthews Correlation Coefficient* 0.2450 untuk MLP [14]. A. Setiya dan kolega memperkenalkan *MolToxPred*, prediksi toksisitas molekul kecil dengan pendekatan *machine learning* menggunakan *Random Forest*, *Multi-Layer Perceptron*, *LightGBM* sebagai klasifikasi dasar dan *Logistic Regression* sebagai meta klasifikasi, mencapai AUROC 0.8776 pada set pengujian dan 0.8884 pada validasi eksternal [15].

Dari beberapa penelitian yang sudah dilakukan, metode seperti *Random Forest*, *Support Vector Machine*, dan *Deep Learning* telah banyak diaplikasikan dan masing-masing memiliki keunggulan serta keterbatasan terkait akurasi, sensitivitas, dan spesifisitas. Meskipun metode *ensemble* seperti *Random Forest* dan *Extremely Randomized Trees* telah meningkatkan performa model prediktif dengan menggabungkan prediksi dari beberapa model, pendekatan ini sering terbatas oleh *overfitting* dan kesulitan dalam menggeneralisasi hasil pada data baru. Hal ini menunjukkan adanya ruang signifikan untuk peningkatan dan mendesaknya pengembangan metode *ensemble* baru yang lebih efektif. Beberapa pendekatan yang bisa diterapkan untuk mengatasi masalah ini antara lain peningkatan fitur dan seleksi fitur yang lebih efektif, serta penggunaan teknik meta-heuristik seperti *Cuckoo Search* dalam proses seleksi fitur. Dengan solusi-solusi tersebut, diharapkan dapat mengembangkan model yang tidak hanya memiliki akurasi tinggi tetapi juga ketahanan yang kuat dalam menghadapi berbagai jenis data.

Penelitian ini bertujuan untuk mengembangkan model prediktif toksisitas uji klinis menggunakan metode *Cuckoo Search-Ensemble*. Metode *ensemble* adalah teknik yang menggabungkan prediksi dari beberapa model pembelajaran mesin untuk menghasilkan prediksi yang lebih akurat daripada model tunggal. Teknik ini efektif

dalam mengurangi varians dan bias yang mungkin muncul dari model prediktif tunggal. Sementara itu, *Cuckoo Search* adalah algoritma meta-heuristik yang terinspirasi dari perilaku parasitisme burung cuckoo. Algoritma ini digunakan untuk mencari solusi optimal dalam masalah optimasi, dan dalam konteks ini, digunakan untuk optimasi seleksi fitur dalam model *ensemble*. Dengan menggabungkan kedua metode ini, diharapkan model yang dikembangkan tidak hanya akurat tetapi juga robust dalam menghadapi berbagai variasi data dalam uji klinis.

Topik dan Batasannya

Rumusan Masalah

Berdasarkan masalah yang disebutkan dalam latar belakang, rumusan masalah yang akan dibahas adalah sebagai berikut:

1. Bagaimana efektivitas algoritma *Cuckoo Search* dalam seleksi fitur untuk memprediksi toksisitas klinis menggunakan model *ensemble*?
2. Bagaimana proses dan dampak dari penyetelan hyperparameter pada model *Cuckoo Search-Ensemble*?
3. Bagaimana performa model *Cuckoo Search-Ensemble* dalam praktiknya?

Batasan Masalah

Agar penelitian menjadi lebih terfokus dan tidak meluas dari pembahasan yang dimaksudkan, maka penelitian ini membatasi ruang lingkup penelitian kepada penggunaan dataset Clintox untuk memprediksi toksisitas klinis. Metode *ensemble* yang digunakan adalah *Random Forest* dan *AdaBoost*, yang dipadukan dengan *Cuckoo Search* untuk seleksi fitur. Penyetelan hyperparameter hanya difokuskan pada algoritma tersebut. Evaluasi performa model menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*. Penelitian ini hanya membahas prediksi toksisitas klinis dan tidak membahas aspek uji klinis lainnya.

Tujuan

Untuk membantu menjelaskan hubungan antara tujuan, pengujian, dan kesimpulan yang diharapkan, tabel berikut merangkum poin-poin tersebut:

Tabel 1. Tabel keterkaitan antara tujuan, pengujian dan kesimpulan

No	Tujuan	Pengujian	Kesimpulan
1	Untuk mengevaluasi efektivitas model <i>ensemble</i> dalam seleksi fitur untuk prediksi toksisitas klinis.	Pengujian efektivitas model <i>ensemble</i> untuk seleksi fitur	Kesimpulan efektivitas model terhadap toksisitas klinis
2	Untuk mengkaji proses dan dampak dari penyetelan hyperparameter pada model <i>Cuckoo Search-Ensemble</i> .	Pengujian proses dan dampak penyetelan hyperparameter pada model <i>Cuckoo Search-Ensemble</i>	Kesimpulan dampak penyetelan terhadap performa model
3	Untuk menilai performa model <i>Cuckoo Search-Ensemble</i> dalam kondisi praktik.	Pengujian performa model <i>Cuckoo Search-Ensemble</i> dalam praktik	Kesimpulan performa model dalam kondisi nyata

Organisasi Tulisan

Bagian-bagian selanjutnya dalam jurnal tugas akhir ini disusun sebagai berikut:

- Studi Terkait

Pada sub-bab Studi Terkait, dibahas berbagai penelitian sebelumnya yang relevan, termasuk metode yang telah digunakan dalam prediksi toksisitas klinis dan hasil yang diperoleh. Penjelasan ini bertujuan untuk memberikan gambaran tentang posisi penelitian ini dalam konteks kajian yang lebih luas.

- Sistem yang Dibangun

Pada sub-bab Sistem yang Dibangun, dijelaskan rancangan dan implementasi sistem yang digunakan dalam penelitian ini. Pembahasannya meliputi detail algoritma *ensemble* (*Random Forest* dan *AdaBoost*), penerapan algoritma *Cuckoo Search* untuk seleksi fitur, serta proses integrasi keduanya dalam sebuah framework prediksi.

- Evaluasi

Pada sub-bab Evaluasi, dipaparkan prosedur pengujian model yang meliputi pengolahan data, pengukuran performa menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*, serta analisis terhadap hasil yang diperoleh untuk menilai keunggulan dan keterbatasan model.

- Kesimpulan

Pada sub-bab Kesimpulan, dirangkum hasil-hasil utama yang diperoleh dari penelitian ini. Bagian ini juga mencakup saran untuk pengembangan lebih lanjut, terutama dalam memperluas cakupan atau meningkatkan pendekatan yang telah digunakan.