

Model machine learning dan molecular docking untuk prediksi senyawa fitokimia dengan aktivitas epidermal growth factor receptor tyrosine kinase inhibitor terhadap mutasi reseptor T790M, C797S dan insersi ekson 20 = Machine learning and molecular docking model to predict phytochemical with EGFR TKI activity against EGFR mutation T790M, C797S and exon 20 insertion

Nathanael Andry Mianto, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920538375&lokasi=lokal>

Abstrak

Latar belakang: Terapi target dengan EGFR TKI merupakan terapi utama untuk kanker paru, khususnya KPKBSK. Selain permasalahan efektivitas harga, permasalahan lain terkait penggunaan EGFR TKI adalah resistansi primer dan sekunder seperti T790M, C797S dan insersi ekson 20. Penelitian *in silico* QSAR berbasis machine learning dan molecular docking dapat mempercepat penemuan senyawa EGFR TKI baru berbasis senyawa fitokimia.

Metode: Penelitian ini adalah penelitian *in silico* untuk mengembangkan model QSAR berbasis machine learning yang dilanjutkan dengan molecular docking untuk penapisan senyawa fitokimia yang memiliki efek EGFR TKI terhadap mutasi T790M, C797S dan insersi ekson 20. Model machine learning dibuat untuk memprediksi nilai IC₅₀ berdasarkan struktur kimia senyawa. Data pembelajaran berasal dari pangkalan data ChEMBL, data senyawa fitokimia dari KEGG, dan data makromolekul dari RCSB PDB. Simulasi molecular dynamics dilakukan untuk memvalidasi lebih lanjut hasil molecular docking.

Hasil: Didapatkan 8627 senyawa EGFR TKI dari ChEMBL, yang dibagi menjadi set pembelajaran dan set uji dengan rasio 9:1. Model machine learning dengan algoritma LGBM berhasil dikembangkan dengan nilai R² sebesar 73%. Lima senyawa fitokimia dengan nilai IC₅₀ prediksi terbaik adalah orobol, norswertianin, isokaempferide, isoathyriol, dan norathyriol. Kelima senyawa memiliki profil farmakokinetik dan toksikologi yang cukup baik. Hasil molecular docking menunjukkan norswertianin memiliki kemampuan ikatan terbaik terhadap EGFR mutasi T790M dan T790M/C797S/L858R, sedangkan orobol terhadap mutasi T790M/C797S dan insersi ekson 20. Kedua senyawa dapat membentuk ikatan yang stabil pada simulasi molecular dynamics.

Kesimpulan: Model QSAR berbasis machine learning dengan algoritma LGBM dapat digunakan untuk memprediksi nilai IC₅₀ EGFR TKI senyawa berdasarkan struktur kimianya. Senyawa fitokimia dapat digunakan sebagai dasar pengembangan EGFR TKI baru. Senyawa norswertianin dan orobol memiliki potensi terbesar sebagai EGFR TKI yang efektif untuk mutasi T790M, C797S dan insersi ekson 20.

.....Background: Targeted therapy with EGFR TKI has been the mainstay of lung cancer treatment, especially NSCLC. Besides the poor cost-effectiveness, primary and secondary resistances, such as T790M, C797S, and exon 20 insertion mutation, have been problematic in EGFR TKI clinical utilization. *In silico* research, such as machine learning-based QSAR dan molecular docking, has the potential to hasten the discovery of novel EGFR TKI based on phytochemicals.

Method: This study is an *in silico* research to develop a machine learning-based QSAR model, followed by molecular docking experiments for virtual screening of phytochemicals that have bioactivity as EGFR TKI against T790M, C797S, and exon 20 insertion mutation. A machine learning model will be developed to

predict IC₅₀ based on chemical structure. The learning set is sourced from the ChEMBL database, phytochemical data from KEGG, and macromolecule data from RCSB PDB. Molecular dynamic simulation is carried out to validate the molecular docking result further.

Results: A total of 8627 compounds was procured from ChEMBL database, which was split into training and test sets with a ratio of 9:1. LGBM based machine learning model with considerable accuracy can be developed, with R² of 73%. The five compounds with the best predicted IC₅₀ value was orobol, norswertianin, isokaempferide, isoathyriol, and norathyriol. All compounds possess a good pharmacokinetics and toxicology profile. The molecular docking result showed that norswertianin has the best binding affinity against EGFR with T790M and T790M/C797S/L8568R. Orobol has the best binding affinity against EGFR with T790M/C797S and exon 20 insertion. Both compounds can form a stable binding in molecular dynamics simulation.

Conclusions: Machine learning-based QSAR model utilizing the LGBM algorithm can predict IC₅₀ value as EGFR TKI based on the compound's chemical structure. Phytochemicals can be used as the basis for novel EGFR TKI. Norswertianin and orobol have the best EGFR TKI potential against T790M, C797S, and exon 20 insertion mutations.