

# Evaluasi Klasifikasi Hubungan Kuantitatif Struktur Aktivitas Molekul dengan Model Hybrid Deep Learning dan Pemilihan Fitur Recursive Feature Elimination pada Inhibitor Dipeptidyl Peptidase-4 = Evaluation of the Classification in Quantitative Structures Activity Relationships of Molecular with Hybrid Deep Learning Models and Selection Features of Recursive Feature Elimination in Dipeptidyl Peptidase-4 Inhibitors

Adawiyah Ulfa, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20516871&lokasi=lokal>

---

## Abstrak

Pengembangan inhibitor Dipeptidyl Peptidase-4 (DPP-4) sangat diperlukan dalam pengobatan Diabetes Mellitus tipe 2 dengan efek samping yang rendah. Pemodelan hubungan kuantitatif struktur aktivitas (QSAR) merupakan pendekatan analisis hubungan struktur kimia dengan aktivitasnya yang banyak digunakan dalam desain obat penyakit Diabetes. Pada tesis ini, model QSAR klasifikasi dibangun untuk memprediksi struktur aktivitas senyawa pada inhibitor DPP-4 yang dapat memblokir kerja enzim DPP-4. Dalam representasi molekul digunakan circular fingerprint ECFP dan FCFP yang menyajikan notasi SMILES dalam format vektor biner. Fingerprint ECFP dan FCFP yang berdiameter 4 dan 6 sebagai input data dalam membangun model QSAR klasifikasi. Pada QSAR klasifikasi dengan pendekatan deep learning memberikan waktu yang cepat dalam proses virtual screening senyawa aktif atau tidak aktif dalam inhibitor DPP-4. Penelitian ini menggunakan model Hybrid Deep Learning 1D CNN-LSTM untuk memprediksi aktivitas senyawa inhibitor dalam kelas aktif atau tidak aktif berdasarkan nilai aktivitas biologis dengan proporsi data latih dan data uji yang berbeda. Dalam arsitektur 1D CNN-LSTM terdiri dari model 1D CNN sebagai tahap ekstraksi fitur dan output dari lapisan konvolusi 1D CNN digunakan dalam lapisan LSTM. Selain itu, pemilihan fitur dengan metode Random Forest-Recursive Feature Elimination (RF-RFE) digunakan untuk memperoleh fitur yang optimal dari dataset ECFP dan FCFP. Selanjutnya, penelitian ini membandingkan performa model dengan menerapkan pemilihan fitur RF-RFE dan tanpa pemilihan fitur RF-RFE. Hasil penelitian ini menunjukkan bahwa model QSAR klasifikasi menggunakan Hybrid Deep Learning yaitu 1D CNN-LSTM dengan pemilihan fitur RF-RFE memperoleh performa model yang lebih baik dibandingkan model tanpa pemilihan fitur optimal. Performa model 1D CNN-LSTM dengan pemilihan fitur RF-RFE menggunakan data ECFP\_4 dengan proporsi data latih 80% memiliki akurasi sebesar 0.9075, sensitivitas 0.9008, spesifisitas 0.9142, dan nilai MCC 0.8151.

.....The development of Dipeptidyl Peptidase-4 (DPP-4) inhibitors is urgently needed in the treatment of Type 2 Diabetes Mellitus with low side effects. Activity structure quantitative relationship modeling (QSAR) is an analytical approach to the relationship between chemical structure and activity which is widely used in diabetes drug design. In this thesis, a classification QSAR model was built to predict the structure of the activity of the DPP-4 inhibitor compound that can block the action of the DPP-4 enzyme. In molecular representation, ECFP and FCFP circular fingerprints are used which present SMILES notation in binary vector format. ECFP and FCFP fingerprints with diameters of 4 and 6 as input data in building a classification QSAR model. The QSAR classification with a deep learning approach provides fast time in the virtual screening process for active or inactive compounds in DPP-4 inhibitors. This study uses the Hybrid Deep Learning 1D CNN-LSTM model to predict the activity of inhibitor compounds inactive or

inactive classes based on the value of biological activity with different proportions of training data and test data. The 1D CNN-LSTM architecture consists of a 1D CNN model as the feature extraction stage and output of 1D CNN convolution layer is used in the LSTM layer. In addition, feature selection using the Random Forest-Recursive Feature Elimination (RF-RFE) method was used to obtain optimal features from the ECFP and FCFP datasets. Furthermore, this study compares the performance of the model by applying the RF-RFE feature selection and without the RF-RFE feature selection. The results of this study indicate that the classification QSAR model using Hybrid Deep Learning, namely 1D CNN-LSTM with RF-RFE feature selection, obtains better model performance than the model without optimal feature selection. The performance of the CNN-LSTM 1D model with RF-RFE feature selection using ECFP\_4 data with a proportion of 80% training data has an accuracy of 0.9075, sensitivity of 0.9008, specificity of 0.9142, and an MCC value of 0.8151.