

INTISARI

PENDEKAKATAN MACHINE LEARNING DENGAN DEEP BELIEF NETWORK UNTUK KLASIFIKASI GENOTOKSISITAS PESTISIDA DI LINGKUNGAN AGRARIS

Shahnaz Izzati Frishila

21/472980/PA/20361

Indonesia sebagai negara agraris dengan penggunaan pestisida yang tinggi berisiko mengalami paparan senyawa genotoksik yang membahayakan kesehatan. Penelitian ini mengembangkan metode klasifikasi genotoksistas pestisida secara otomatis menggunakan *Deep Belief Network* (DBN) berbasis data struktur molekul dari EFSA. Data diolah menjadi Morgan *fingerprint* 2048-bit dan dilengkapi delapan deskriptor fisikokimia. Kemudian untuk mengatasi ketidakseimbangan kelas, digunakan teknik *Stable SMOTE*. Arsitektur DBN terdiri atas tiga RBM (512→128→128), dengan *pretraining* CD-1 dan *fine-tuning* menggunakan AdamW. *Hyperparameter* dioptimalkan dengan Optuna, termasuk *learning rate*, dropout, momentum, dan *batch size*.

Model terbaik menghasilkan akurasi 93,66 %, MCC 0,5716, dan AUC ROC 0,8982 pada data validasi, serta MCC 0,4093 dan AUC ROC 0,8535 pada data uji. Kinerja DBN dibandingkan dengan CNN dan Random Forest menunjukkan keunggulan DBN pada metrik MCC dan AUC, meskipun akurasi tertinggi dicapai oleh Random Forest. Hasil ini menunjukkan potensi DBN sebagai metode prediksi genotoksistas, dengan peluang peningkatan melalui pengayaan data dan tuning lanjutan.

Kata kunci: Klasifikasi genotoksistas, *Deep Belief Network*, *fingerprint molekul*, *Stable SMOTE*, prediksi toksikologi

ABSTRACT

MACHINE LEARNING APPROACH USING DEEP BELIEF NETWORK FOR CLASSIFYING PESTICIDE GENOTOXICITY IN AGRARIAN ENVIRONMENTS

Shahnaz Izzati Frishila

21/472980/PA/20361

Indonesia, as an agricultural country with high pesticide usage, faces significant risks of exposure to genotoxic chemical compounds that threaten public health. This study developed an automatic pesticide genotoxicity classification method using a Deep Belief Network (DBN) based on molecular structure data from the EFSA database. The data were converted into 2048-bit Morgan fingerprints and supplemented with eight physicochemical descriptors. Class imbalance was addressed using Stable SMOTE. The DBN architecture consisted of three RBM layers (512→128→128) pretrained with CD-1 and fine-tuned end-to-end using AdamW. Hyperparameters such as learning rate, dropout, momentum, and batch size were optimized using Optuna.

The best DBN model achieved 93.66% accuracy, an MCC of 0.5716, and an AUC-ROC of 0.8982 on the validation set, while the test set yielded an MCC of 0.4093 and an AUC-ROC of 0.8535. Compared to CNN and Random Forest models, DBN showed superior MCC and AUC scores, although Random Forest achieved slightly higher accuracy. These results demonstrate DBN's promise for predicting genotoxicity, with further improvement possible through data enrichment and advanced hyperparameter tuning.

Keyword: *Genotoxicity classification, Deep Belief Network, Morgan fingerprint, Stable SMOTE, Toxicology prediction*