

Makalah Seminar Hasil

PREDIKSI KANDIDAT SENYAWA HERBAL SEBAGAI OBAT OBESITAS MENGGUNAKAN PEMODELAN STACKED AUTOENCODER-DEEP NEURAL NETWORK

FADIL RISDIAN ANSORI (G641180111)^{1*}, WISNU ANANTA KUSUMA

ABSTRAK

Obesitas merupakan kondisi kronis akibat akumulasi lemak yang berlebihan dalam tubuh. Penyakit obesitas secara signifikan dapat meningkatkan resiko munculnya penyakit bawaan lainnya seperti diabetes, penyakit jantung, hipertensi, dan batu empedu. Penemuan obat obesitas dilakukan dengan menggunakan *machine learning* untuk memprediksi *Drug Target-Interaction* (DTI). Penelitian ini membangun model *Stacked Autoencoders-Deep Neural Network* (SAE-DNN) dengan klasifikasi multi label yang bertujuan untuk memprediksi kandidat senyawa herbal sebagai obat obesitas. Ekstraksi fitur dilakukan dengan menggunakan tiga *fingerprint*, yaitu PubChem *fingerprint*, MACCS *fingerprint*, dan Klekota-Roth *fingerprint*. Hasil penelitian menunjukkan bahwa model SAE-DNN mampu memprediksi DTI pada kasus obesitas dengan kinerja yang baik. Model SAE-DNN dengan menggunakan fitur Pubchem *fingerprint* menghasilkan nilai metrik rata-rata terbaik dengan akurasi 0,8161, *recall* 0,8556, presisi 0,8491, dan *F-measure* 0,8445. Hasil prediksi senyawa herbal menggunakan fitur Pubchem *fingerprint* SAE-DNN menghasilkan 208 interaksi senyawa-protein.

Kata Kunci: obesitas, *drug target-interaction*, klasifikasi multi label, *stack autoencoder-deep neural network*.

ABSTRACT

Obesity is a chronic condition due to excessive accumulation of fat in the body. Obesity can significantly increase the risk of other congenital diseases such as diabetes, heart disease, hypertension, and gallstones. Obesity drug discovery is done using machine learning to predict Drug Target-Interaction (DTI). This study builds a Stacked Autoencoders-Deep Neural Network (SAE-DNN) model which will later be used to predict candidate herbal compounds as obesity drugs. Feature extraction is carried out using three fingerprints, namely the PubChem fingerprint, MACCS fingerprint, and Klekota-Roth fingerprint. The results showed that the SAE-DNN model was able to predict DTI in obesity cases with good performance. The SAE-DNN model with a Pubchem fingerprint feature produces the best average metric value with an accuracy of 0.8161, recall of 0.8556, precision of 0.8491, and F-measure of 0.8445. The results of the prediction of herbal compounds using the Pubchem fingerprint SAE-DNN feature resulted in 208 compound-protein interactions.

Keywords: obesity, drug target-interaction, multilabel classification, stack autoencoder-deep neural network.

¹Departemen Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor, Bogor 16680

*Mahasiswa Program Studi S1 Ilmu Komputer, FMIPA-IPB; Surel: fadil_risdian@apps.ipb.ac.id

PENDAHULUAN

Latar Belakang

Obesitas merupakan salah satu masalah masyarakat global yang terus meningkat. The World Health Organization (WHO) mendefinisikan obesitas sebagai akumulasi lemak yang berlebihan sehingga dapat menyebabkan resiko bagi kesehatan. Penyakit obesitas secara signifikan dapat meningkatkan resiko munculnya penyakit bawaan lainnya seperti diabetes, penyakit jantung, hipertensi, dislipidemia, penyakit jantung koroner, batu empedu, sindrom apnea tidur obstruktif, asma, dan penyakit kejiwaan (De Lorenzo *et al.* 2019; Ye *et al.* 2020). Jumlah penderita obesitas meningkat hampir tiga kali lipat secara global sejak tahun 1975 (WHO 2021). Diperkirakan pada tahun 2030, hampir separuh penduduk dunia akan menderita obesitas (Ye *et al.* 2020). Pada tahun 2015, jumlah kematian di dunia akibat kelebihan berat badan mencapai 4 juta, lebih dari 60% kematian tersebut akibat dari obesitas. Obesitas juga memiliki dampak terhadap beban ekonomi di tingkat individu, keluarga, dan negara (Tremmel *et al.* 2017).

Penemuan obat secara eksperimen konvensional membutuhkan biaya besar dan waktu yang lama. Salah satu cara untuk mengatasi hal tersebut adalah menggunakan komputer untuk penemuan obat. Penemuan obat dengan menggunakan komputer dapat dilakukan dengan mengamati *drug target interaction* (DTI), yakni interaksi antara senyawa kimia atau obat dan target protein. Pencarian informasi mengenai DTI obesitas dapat dilakukan dengan melakukan eksplorasi pengobatan anti-obesitas. Salah satu cara eksplorasi obat anti-obesitas adalah dengan menargetkan fungsi jaringan adiposa yaitu *brown adipose tissue* (BAT) (Poher *et al.* 2015). BAT berperan besar terhadap proses *thermogenesis*. Proses tersebut akan mengubah energi yang disimpan dalam lemak menjadi energi panas (Cinti 2006). Protein-protein yang berkaitan dengan BAT akan menjadi target senyawa untuk diketahui interaksinya dalam penelitian DTI.

Telah terdapat penelitian terkait DTI dengan menggunakan *machine learning* sebelumnya. Sulistiawan *et al.* (2020) melakukan prediksi klasifikasi DTI menggunakan metode *deep semi-supervised learning* (DSSL) dengan model *stack autoencoder deep neural network* (SAE-DNN) untuk eksplorasi obat Covid-19. Namun, terdapat beberapa kekurangan dalam penggunaan klasifikasi biner dalam melakukan prediksi DTI yakni data kelas tidak seimbang, dan penyederhanaan permasalahan dengan memodelkan dimensi senyawa dan protein tanpa memperhatikan asosiasi antar senyawanya. Selain itu, terdapat konsep polifarmakologi yang menjelaskan bahwa kebanyakan obat bekerja tidak hanya menargetkan protein tunggal saja, tetapi banyak protein (Chen *et al.* 2016).

Klasifikasi multi label DTI dapat digunakan untuk mengatasi kekurangan klasifikasi biner DTI. Pada penelitian lain, Fadli *et al.* (2021) melakukan pencarian DTI multi label menggunakan model SAE-DNN pada data Covid-19. Ekstraksi fitur pada senyawa dilakukan dengan 4 fingerprint dengan 2 tipe: substructural fingerprint (PubChem fingerprint dan Molecular ACCess System (MACCS) fingerprint) dan topological fingerprint (daylight fingerprint and circular fingerprint (ECFPs)). Hasil akurasi, recall, precision, dan F-measure terbesar didapatkan dengan menggunakan ECFPs secara berurutan yaitu 83%, 91%, 88%, dan 89%. Pada klasifikasi multi label DTI, ruang masukan berisi kumpulan senyawa. Ruang keluaran terdiri dari interaksi antara senyawa dan protein targetnya, yang dimodelkan sebagai matriks. Model prediksi multi label DTI belajar dari beberapa output secara bersamaan. Hal tersebut lebih menguntungkan karena terdapat korelasi antara target sehingga dapat meningkatkan kinerja prediksi (Pliakos dan Vens 2020).

Penelitian DTI dengan menggunakan pendekatan *binary classification* menyederhanakan masalah yang dihadapi oleh DTI dan dapat menjadi bias pada model

yang digunakan. Oleh karena itu, diperlukan uji coba dengan menggunakan pendekatan klasifikasi multi label untuk memprediksi DTI. Kinerja model SAE-DNN yang baik dalam penelitian Fadli *et al.* (2021) juga memberikan inspirasi untuk melakukan penelitian prediksi DTI dengan menggunakan pendekatan klasifikasi multi label menggunakan algoritma SAE-DNN pada kasus obesitas dengan menargetkan protein yang berkaitan dengan BAT.

Perumusan Masalah

Perumusan masalah pada penelitian ini adalah:

1. Apakah klasifikasi multi label dapat digunakan untuk melakukan prediksi senyawa herbal antiobesitas?
2. Bagaimana mengembangkan pemodelan klasifikasi multi label SAE-DNN untuk prediksi kandidat senyawa herbal sebagai obat diabetes?
3. Bagaimana memprediksi interaksi yang tidak diketahui antara senyawa herbal dan protein target pada kasus obesitas?

Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Mengetahui keberhasilan klasifikasi multi label untuk melakukan prediksi senyawa herbal antiobesitas
2. Menerapkan model klasifikasi multi-label SAE-DNN untuk prediksi kandidat senyawa herbal sebagai obat diabetes
3. Memprediksi interaksi antara senyawa herbal dengan target protein baru untuk kasus obesitas

Manfaat Penelitian

Hasil penelitian ini diharapkan dapat membantu para peneliti untuk melakukan proses penemuan obat-obatan obesitas yang berasal dari senyawa herbal. Selain itu, masyarakat dapat mengetahui kandungan bahan alam yang mengandung senyawa herbal penyembuh obesitas sehingga dapat dikonsumsi langsung.

Ruang Lingkup Penelitian

Penelitian ini hanya berfokus kepada penemuan obat herbal untuk menyembuhkan obesitas. Oleh karena itu, protein yang digunakan hanya protein yang terkait dengan BAT. Penelitian ini tidak melakukan proses *molecular docking* untuk melihat efektivitas senyawa ketika berikatan dengan protein.

METODE

Tahapan Penelitian

Penelitian ini dilakukan dengan beberapa tahapan, dimulai dari pengambilan data dari basis data BindingDB, Pubchem, dan HerbalDB, praproses data dengan melakukan ekstraksi fitur dan pembentukan label, implementasi model menggunakan metode multi label learning dengan algoritma SAE-DNN, evaluasi model, dan pengujian model dengan data senyawa herbal.

Pengambilan Data

Penelitian ini menggunakan tiga dataset yaitu data protein, data senyawa, dan data senyawa herbal. Data protein yang berkaitan dengan BAT diperoleh melalui situs basis data Online Mendelian Inheritance in Man (OMIM) dengan menggunakan kata kunci "*brown adipose tissue*". Selanjutnya, diterapkan algoritma *skyline query* pada data tersebut

untuk mendapatkan protein signifikan sehingga data tersebut dapat digunakan untuk dilakukan pencarian data interaksi senyawa-protein.

Data interaksi senyawa-protein diperoleh dari situs BindingDB dan Pubchem. Pencarian senyawa dilakukan dengan menggunakan data protein pada situs BindingDB dan Pubchem. Situs BindingDB dapat diakses melalui alamat <https://www.bindingdb.org/bind/index.jsp>, sedangkan situs Pubchem dapat diakses melalui alamat <https://pubchem.ncbi.nlm.nih.gov/>. Data senyawa herbal diperoleh dari situs HerbalDB melalui situs <http://herbaldb.farmasi.ui.ac.id/v3/>.

Pra Proses Data

Tahapan pertama dalam praproses data adalah melakukan penyatuan dan penghapusan data berlebih pada data interaksi senyawa-protein yang diambil dari BindingDB dan Pubchem. Selanjutnya, data yang masih berupa fail teks diubah menjadi *Simplified Molecular-Input Line-Entry System* (SMILES) deskriptor. Data senyawa dengan format SMILES tersebut diubah ke bentuk sub struktur molekul (*fingerprint*) sebagai representasi numerik. *Fingerprint* merupakan pengkodean senyawa menjadi vektor *fingerprint* boolean, sehingga hasil perubahan data menjadi 1 dan 0. Ekstraksi fitur pada senyawa dilakukan dengan menggunakan deskriptor PubChem *fingerprint*, MACCS *fingerprint*, dan Klekota-Roth *fingerprint*. Data ekstraksi tersebut akan digunakan sebagai masukkan pada model SAE-DNN.

Langkah selanjutnya adalah mengubah atribut kelas dari label tunggal menjadi multi label. Data kelas berubah menjadi array P ($P = [p_1, p_2, p_3, \dots, p_n]$, dengan n jumlah protein). Pada setiap baris, p merupakan nilai interaksi antara senyawa dengan protein dengan label 1 adalah berinteraksi dan label 0 adalah tidak berinteraksi atau tidak diketahui interaksinya.

Pemodelan SAE-DNN

Dalam pengujian, digunakan salah satu jenis unsupervised learning yaitu Stacked Autoencoders (SAE) yang bertujuan untuk mendapatkan nilai bobot awal. Model SAE terdiri dari banyak *autoencoder*. SAE mempelajari data masukan dan berusaha mengonstruksi agar keluaran mirip dengan nilai masukan melalui perantara bottleneck. Hasil dari *bottleneck* akan diteruskan diteruskan *autoencoder* berikutnya. Hasil dari *autocoder* terakhir digunakan sebagai bobot untuk *deep neural network*.

Selanjutnya, dibuat model *deep neural network* dengan bobot yang didapatkan dari model SAE. Data latih digunakan sebagai data masukan. Untuk meningkatkan model DNN, dilakukan proses *batch normalization* (BN) dan *dropout* (DO). BN bekerja dengan mengurangi *internal covariate shift*. BN bekerja dengan memperbaiki *means* dan *variances* dari *inputs layer* (Ioffe dan Szegedy 2015). DO dilakukan untuk mencegah *overfitting* dan menyediakan cara untuk menggabungkan cara untuk menggabungkan banyak arsitektur jaringan syaraf tiruan secara efisien (Srivastava *et al.* 2014). Setelah hasil keluaran didapatkan, hasil model dan hasil asli dibandingkan untuk menghitung galat yang dihasilkan oleh model. *Backpropagation* dilakukan untuk menyesuaikan bobot pada model dengan harapan performa model akan meningkat.

Hyperparameter Tuning

Hyperparameter tuning membantu memilih nilai parameter yang terbaik dan menghasilkan performa klasifikasi yang optimal. *Hyperparameters* dapat diatur untuk mengontrol kinerja teknik yang diaplikasikan (Shankar *et al.* 2020). Metode *Bayesian optimization* digunakan pada pengujian *hyperparameter tuning* ini. Penerapan *bayesian optimization* bertujuan untuk menemukan pengaturan yang terbaik untuk hyperparameter

yang digunakan (Snoek *et al.* 2012). Daftar kombinasi *hyperparameter* yang akan digunakan dapat dilihat di Tabel 1.

Tabel 1 Daftar *hyperparameter* yang diujikan

<i>Hyperparameter</i>	Nilai
Jumlah <i>hidden neuron</i> di <i>layer</i> ke-0 (HL_0)	320-1800
Jumlah <i>hidden neuron</i> di <i>layer</i> ke-i (HL_i)	0.5 - 0.75
Jumlah <i>Hidden layer</i>	2-4
<i>Optimizer</i>	Adam,adgrad
<i>Learning Rate</i>	0.01- 0.1
<i>Dropout Rate</i>	0.5 - 0.8

Evaluasi Model

Evaluasi model dilakukan dengan menggunakan metode *iterative stratification*. Metode tersebut merupakan modifikasi *k-fold cross validation* yang bertujuan untuk menyeimbangkan jumlah kombinasi label dari data multi label (Sechidis *et al.* 2011). Pada pengujian ini, digunakan k sebanyak 5 kali.

Metrik yang digunakan untuk evaluasi model adalah akurasi, *precision*, *recall*, dan *F-measure*. Akurasi didefinisikan sebagai rasio antara data jumlah nilai yang benar dan jumlah total nilai (Cappiello *et al.* 2003). *Precision* mengacu seberapa dekat nilai yang dihitung secara individual sesuai satu dengan yang lain (Heidaryan 2019). *Recall* adalah persentase hasil prediksi yang benar-benar relevan dari jumlah total hasil prediksi yang relevan (Buckland dan Gey 1994). *F-measure* menghitung perbandingan rata-rata antara *precision* dan *recall* (Musicant *et al.* 2003).

Pengujian Model

Untuk mendapatkan kualitas prediksi yang baik, dilakukan perbandingan antara hasil dari tiga cara pengubahan fitur senyawa menjadi representasi numerik (PubChem, MACCS dan Klekota-Roth). Perbandingan dilakukan menggunakan *iterative stratification* sebanyak $k = 5$. Model yang terbaik akan diuji kepada data senyawa herbal yang diambil dari HerbalDB terhadap protein yang terkait dengan *brown adipose tissue*.

Lingkungan Pengembangan

Spesifikasi perangkat keras dan perangkat lunak yang digunakan untuk penelitian ini adalah sebagai berikut:

1. Perangkat keras berupa komputer personal dengan spesifikasi:
 - Intel Core i3-5005U
 - RAM 8 GB
2. Perangkat lunak yang digunakan, yaitu:
 - Bahasa pemrograman Python
 - Pandas, Numpy, Matplotlib, dan seaborn untuk manipulasi, analisa, dan visualisasi data
 - Sklearn untuk menghitung metrik dan membagi data
 - Tensorflow untuk pemodelan SAE-DNN
 - KerasTuner untuk melakukan *tuning* parameter
 - Google Colab untuk pengembangan model SAE-DNN dan pemanfaatan GPU
 - Pubchempy untuk menghasilkan fitur SMILES

- PyFingerprint untuk ekstraksi fitur

HASIL DAN PEMBAHASAN

Pengambilan Data

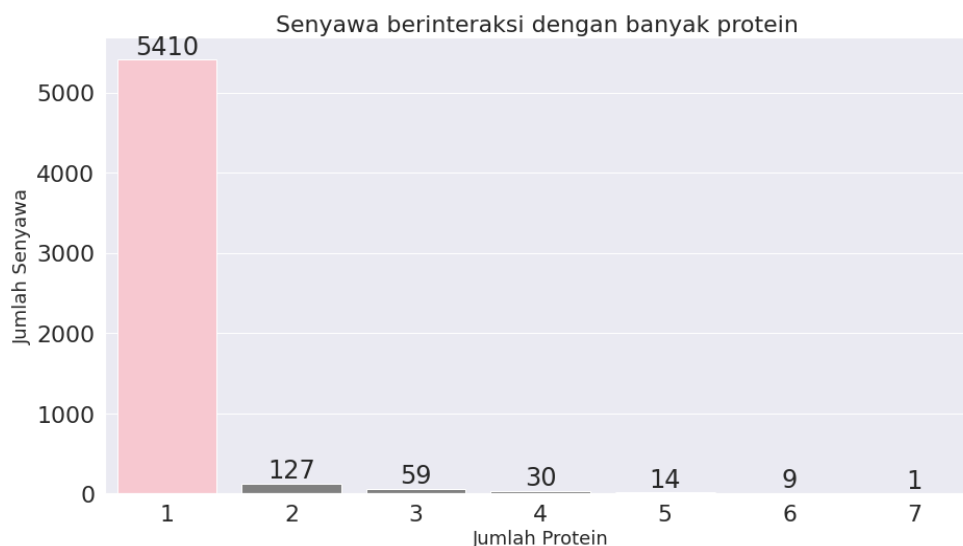
Data protein signifikan yang didapatkan berjumlah sebanyak tujuh yaitu PPARG, PPARGC1A, EP300, LPL, INS, LEP, dan B3GNTL1. Data interaksi senyawa protein diperoleh melalui situs BindingDB dan Pubchem dengan menggunakan data protein sebagai masukan pada kolom pencarian. Total data interaksi senyawa-protein yang didapatkan melalui situs BindingDB berjumlah 7514 interaksi, sedangkan data interaksi senyawa-protein yang diperoleh dari situs Pubchem berjumlah 5701 interaksi. Data senyawa herbal didapatkan dari situs HerbalDB berjumlah 403 senyawa herbal yang terdapat di Indonesia.

Pra Proses Data

Penghapusan data duplikat pada data interaksi dari situs BindingDB menghasilkan total interaksi sebanyak 4847, sedangkan penghapusan pada data interaksi dari situs Pubchem menghasilkan total 1259 interaksi. Selanjutnya dilakukan proses penggabungan data interaksi dari kedua situs dan proses penghapusan data duplikat. Total data interaksi senyawa-protein hasil penggabungan data dan penghapusan data yang sama adalah 6092 interaksi, dengan 4942 senyawa berinteraksi dengan PPARG, 428 senyawa berinteraksi dengan EP300, 244 senyawa berinteraksi dengan INS, 186 senyawa berinteraksi dengan PPARGC1A, 151 senyawa berinteraksi dengan LPL, 121 senyawa berinteraksi dengan LEP, dan 20 senyawa berinteraksi dengan B3GNTL1.

Proses ekstraksi fitur senyawa dilakukan dengan mengambil data yang hanya mengandung senyawa saja pada 6092 data interaksi senyawa-protein. Total data senyawa didapatkan sebanyak 5650. Proses ekstraksi fitur senyawa menghasilkan 881 atribut untuk PubChem fingerprint, 166 atribut untuk MACCS fingerprint, dan 4860 atribut untuk Klekota-Roth fingerprints.

Identifikasi protein unik pada data interaksi dilakukan terlebih dahulu untuk proses transformasi kelas data. Terdapat tujuh protein pada data interaksi. Selanjutnya, proses pembentukan data multi label menghasilkan 5650 baris data yang merupakan senyawa dan tujuh kelas yang merupakan himpunan protein yang berinteraksi dengan senyawa tersebut. Data didominasi oleh senyawa yang hanya memiliki satu protein yang berinteraksi dengan senyawa tersebut yakni sebanyak 5410 interaksi. Gambar 1 menunjukkan rincian jumlah senyawa yang berinteraksi dengan minimal satu protein.



Gambar 1. Jumlah senyawa yang berinteraksi dengan minimal satu protein.

Pemodelan SAE-DNN

Pemodelan SAE-DNN terlebih dahulu dibangun dengan menggunakan parameter standar. Selanjutnya hasil dibandingkan dengan DNN tanpa proses pre-training. Adapun default parameter yang digunakan yaitu: Node di $HL_0 = 1024$, Node di $HL_i = 0.5$, hidden layer = 2, Optimizer = Adam, activation function= ReLU, learning rate = 0.01 dan dropout rate = 0.5. Selanjutnya dicari hyperparameter optimal dari SAE-DNN untuk semua proses ekstraksi fitur. Adapun hasil hyperparameter tuning dari model SAE-DNN untuk semua proses ekstraksi fitur dapat dilihat pada Tabel 2.

Tabel 2 Daftar hyperparameter yang diujikan

Hyperparameter	Model		
	Pubchem	MACCS	Klekota-Roth
Node di HL_0	1800	1280	1700
Node di HL_i	0.5	0.66	0.5
Jumlah hidden layer	2	2	2
Optimizer	adam	adam	adam
Learning rate	0.07	0.1	0.1
Dropout rate	0.5	0.5	0.5

Evaluasi Model

Iterative stratification dilakukan dalam pemodelan masing-masing sebanyak 5 kali. Kemudian, nilai rata-rata tiap metrik didapatkan dan dibandingkan sebagai evaluasi. Tabel 3 menampilkan perbandingan kinerja model.

Tabel 3 Perbandingan kinerja model

Metrik	SAE-DNN			DNN		
	Pubchem	MACCS	Klekota-R	Pubchem	MACCS	Klekota-R

Akurasi	0.8161	0.7904	0.7790	0.7810	0.7699	0.7732
Recall	0.8556	0.8109	0.8536	0.8538	0.7930	0.8472
Precision	0.8491	0.8152	0.8260	0.8296	0.8012	0.8177
F-score	0.8445	0.8091	0.8259	0.8320	0.7922	0.8191
Akurasi_std	0.036	0.051	0.030	0.058	0.136	0.026
Recall_std	0.036	0.059	0.026	0.034	0.143	0.009
Precision_std	0.037	0.059	0.026	0.038	0.147	0.020
F-score_std	0.036	0.057	0.022	0.031	0.143	0.017

Model SAE-DNN menghasilkan kinerja yang lebih baik dibandingkan model DNN tanpa pre-training. Nilai masing-masing rata-rata akurasi, recall, precision dan f-measure pada setiap fingerprint memiliki nilai yang lebih tinggi dengan model SAE-DNN. Hal ini menunjukkan bahwa model SAE-DNN memiliki kemampuan yang baik dalam memprediksi DTI kelas multi label (akurasi), akurat dalam memprediksi kelas positif (recall tinggi), prediksi positif yang baik (Precision), dan mengenali kelas minoritas dengan baik (F1 tinggi).

Model SAE-DNN dengan fitur Pubchem menghasilkan nilai rata-rata metrik paling baik dibandingkan dengan proses ekstraksi fitur lainnya, dengan nilai akurasi 0.8161, recall 0.8556, precision 0.8491, dan f-measure 0.8445. Standar deviasi tiap metrik pada model MACCS memiliki nilai yang lebih tinggi dibandingkan dengan model yang lain. Hal ini menunjukkan bahwa kinerja model MACCS untuk tiap fold memiliki kestabilan yang rendah dibandingkan model lainnya.

Uji Senyawa Herbal

Uji senyawa herbal dilakukan menggunakan model SAE-DNN yang dilatih dengan ketiga fitur fingerprint. Data senyawa yang telah diambil dari situs HerbalDB, dilakukan pencarian SMILES dengan menggunakan ID PubChem. Pencarian ID PubChem senyawa herbal menghasilkan 305 senyawa herbal yang memiliki ID PubChem. Selanjutnya dilakukan ekstraksi fitur dengan masing-masing menggunakan ketiga fingerprint.

Hasil prediksi dengan model SAE-DNN fitur PubChem fingerprint menghasilkan 208 senyawa yang memiliki interaksi dengan protein *brown adipose tissue*. Dari 208 senyawa, 187 senyawa berinteraksi dengan satu protein, dua senyawa berinteraksi dengan dua protein, tujuh senyawa berinteraksi dengan tiga protein, tujuh senyawa berinteraksi dengan empat protein, dan lima senyawa berinteraksi dengan lima protein.

Hasil prediksi dengan model SAE-DNN fitur MACCS fingerprint menghasilkan 82 senyawa yang memiliki interaksi dengan protein *brown adipose tissue*. Dari 82 senyawa, 81 senyawa berinteraksi hanya dengan satu protein dan satu senyawa berinteraksi dengan empat protein.

Hasil prediksi dengan model SAE-DNN fitur Klekota-Roth fingerprint menghasilkan 298 senyawa yang memiliki interaksi dengan protein *brown adipose tissue*. Dari 298 senyawa, 215 senyawa berinteraksi dengan satu protein, 67 senyawa berinteraksi dengan dua protein, enam senyawa berinteraksi dengan tiga protein, sembilan senyawa berinteraksi dengan empat protein, tiga senyawa berinteraksi dengan lima protein, dan satu senyawa berinteraksi dengan enam protein.

KESIMPULAN DAN SARAN

Kesimpulan

Penerapan metode klasifikasi multi label untuk pencarian kandidat senyawa yang berinteraksi dengan protein obesitas (BAT) menggunakan model SAE-DNN telah berhasil dilakukan. Hasil penelitian menunjukkan bahwa model SAE-DNN mampu memprediksi interaksi antara senyawa-protein pada kasus obesitas (BAT) dengan kinerja yang cukup baik dan mampu mengungguli kinerja DNN. Hasil penelitian juga menunjukkan bahwa penggunaan fitur PubChem fingerprint sebagai variabel prediktor model menghasilkan nilai rata-rata metrik paling baik dengan akurasi sebesar 0,816, recall 0,856, precision 0,849, dan f-measure 0,844.

Hasil uji senyawa herbal menggunakan model SAE-DNN fitur PubChem menghasilkan 208 senyawa herbal yang berkaitan dengan protein BAT, sedangkan hasil klasifikasi senyawa herbal dengan menggunakan model SAE-DNN fitur MACCS menghasilkan 82 senyawa herbal yang berkaitan dengan protein BAT dan prediksi senyawa herbal dengan menggunakan model SAE-DNN fitur Klekota-Roth menghasilkan 298 senyawa herbal yang berkaitan dengan protein BAT.

Saran

Saran yang diberikan berdasarkan penelitian ini adalah:

1. Menggunakan fitur ekstraksi fingerprint yang lain untuk mengetahui hasil yang lebih baik
2. Menggunakan metode optimasi lain untuk melakukan tuning hyperparameter dengan model SAE-DNN

DAFTAR PUSTAKA

- Buckland M, Gey F. 1994. The relationship between Recall and Precision. *J. Am. Soc. Inf. Sci.* 45(1):12–19.doi:10.1002/(SICI)1097-4571(199401)45:1<12::AID-ASI2>3.0.CO;2-L.
- Cappiello C, Francalanci C, Pernici B. 2003. Time-Related Factors of Data Quality in Multichannel Information Systems. *J. Manag. Inf. Syst.* 20(3):71–92.doi:10.1080/07421222.2003.11045769.
- Chen X, Yan CC, Zhang Xiaotian, Zhang Xu, Dai F, Yin J, Zhang Y. 2016. Drug–target interaction prediction: databases, web servers and computational models. *Brief. Bioinform.* 17(4):696–712.doi:10.1093/bib/bbv066.
- Cinti S. 2006. The role of brown adipose tissue in human obesity. *Nutr. Metab. Cardiovasc. Dis.* 16(8):569–574.doi:10.1016/j.numecd.2006.07.009.
- De Lorenzo A, Gratteri S, Gualtieri P, Cammarano A, Bertucci P, Di Renzo L. 2019. Why primary obesity is a disease? *J. Transl. Med.* 17(1):169.doi:10.1186/s12967-019-1919-y.
- Fadli A, Kusuma WA, Annisa, Batubara I, Heryanto R. 2021. Screening of Potential Indonesia Herbal Compounds Based on Multi-Label Classification for 2019 Coronavirus Disease. *Big Data Cogn. Comput.* 5(4):75.doi:10.3390/bdcc5040075.
- Gilson MK, Liu T, Baitaluk M, Nicola G, Hwang L, Chong J. 2016. BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.* 44(D1):D1045–D1053.doi:10.1093/nar/gkv1072.
- Heidaryan E. 2019. A Note on Model Selection Based on the Percentage of Accuracy-Precision. *J. Energy Resour. Technol.* 141(4):045501.doi:10.1115/1.4041844.

- Ioffe S, Szegedy C. 2015 Mar 2. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *ArXiv150203167 Cs*. [diunduh 2021 Des 16]. Tersedia pada: <http://arxiv.org/abs/1502.03167>
- LeCun Y, Bengio Y, Hinton G. 2015. Deep learning. *Nature*. 521(7553):436–444.doi:10.1038/nature14539.
- Musicant DR, Kumar V, Ozgur A. 2003. Optimizing F-Measure with Support Vector Machines. Di dalam: *FLAIRS Conference 2003*. hlm. 5.
- Pliakos K, Vens C. 2020. Drug-target interaction prediction with tree-ensemble learning and output space reconstruction. *BMC Bioinformatics*. 21(1):49.doi:10.1186/s12859-020-3379-z.
- Poher A-L, Altirriba J, Veyrat-Durebex C, Rohner-Jeanrenaud F. 2015. Brown adipose tissue activity as a target for the treatment of obesity/insulin resistance. *Front. Physiol.* 6.doi:10.3389/fphys.2015.00004. [diunduh 2021 Nov 8]. Tersedia pada: <http://journal.frontiersin.org/article/10.3389/fphys.2015.00004/abstract>
- Sechidis K, Tsoumakas G, Vlahavas I. 2011. On the Stratification of Multi-label Data. Di dalam: Gunopulos D, Hofmann T, Malerba D, Vazirgiannis M, editor. Machine Learning and Knowledge Discovery in Databases. [internet] Vol. 6913. Berlin, Heidelberg. Berlin, Heidelberg: Springer Berlin Heidelberg. (Lecture Notes in Computer Science). hlm. 145–158. [diunduh 2022 Jul 11]. Tersedia pada: http://link.springer.com/10.1007/978-3-642-23808-6_10
- Shankar K, Zhang Y, Liu Y, Wu L, Chen C-H. 2020. Hyperparameter Tuning Deep Learning for Diabetic Retinopathy Fundus Image Classification. *IEEE Access*. 8:118164–118173.doi:10.1109/ACCESS.2020.3005152.
- Snoek J, Larochelle H, Adams RP. 2012 Jul. Practical Bayesian Optimization of Machine Learning Algorithms. :9.
- Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. 2014 Jun 14. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. :30.
- Sulistiani F, Kusuma WA, Ramadhanti NS, Tedjo A. 2020. Drug-Target Interaction Prediction in Coronavirus Disease 2019 Case Using Deep Semi-Supervised Learning Model. Di dalam: *2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*. [internet] *2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*. Depok, Indonesia. Depok, Indonesia: IEEE. hlm. 83–88. [diunduh 2021 Des 16]. Tersedia pada: <https://ieeexplore.ieee.org/document/9263241/>
- Tremmel M, Gerdtham U-G, Nilsson P, Saha S. 2017. Economic Burden of Obesity: A Systematic Literature Review. *Int. J. Environ. Res. Public. Health*. 14(4):435.doi:10.3390/ijerph14040435.
- World Health Organization. (2021). Obesity and overweight. [diakses 2021 Feb 8]. <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>.
- Ye P, Xi Y, Huang Z, Xu P. 2020. Linking Obesity with Colorectal Cancer: Epidemiology and Mechanistic Insights. *Cancers*. 12(6):1408.doi:10.3390/cancers12061408.