

Makalah Kolokium

PREDIKSI KANDIDAT SENYAWA HERBAL SEBAGAI OBAT OBESITAS MENGGUNAKAN PEMODELAN SEMI-SUPERVISED LEARNING

FADIL RISDIAN ANSORI (G641180111)*, 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) yang nantinya digunakan untuk memprediksi kandidat senyawa herbal sebagai obat obesitas. Alur penelitian dimulai dari pengambilan data, praproses data, pemodelan, evaluasi model dan diakhiri dengan pengujian model pada data uji untuk memperoleh kandidat senyawa herbal.

Kata Kunci: obesitas, *drug target-interaction*, *semi-supervised learning*.

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. The research flow starts from data collection, data preprocessing, modelling, model evaluation and ends with model testing on test data to obtain candidate herbal compounds.

Keywords: obesity, drug target-interaction, semi-supervised learning.

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.

Departemen 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

Obesitas juga memiliki dampak terhadap beban ekonomi di tingkat individu, keluarga, dan negara (Tremmel *et al.* 2017)

Penemuan obat obesitas telah banyak dilakukan sebelumnya. Penemuan obat secara eksperimen konvensional membutuhkan biaya besar dan waktu yang lama. Untuk itu, diperlukan suatu cara untuk mengatasi kekurangan penemuan obat secara konvensional. Salah satu cara yang dilakukan 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. Menargetkan komponen spesifik dari regulasi neuroendokrin asupan energi, seperti leptin atau neuropeptida hipotalamus, telah banyak menunjukkan kegagalan. Alternatif baru dilakukan 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. DTI dimodelkan sebagai matriks $Y \in R^{M \times N}$, dengan M adalah jumlah senyawa dan N adalah jumlah target protein. 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 algoritme SAE-DNN pada kasus obesitas dengan menargetkan protein yang berkaitan dengan BAT. Ekstraksi fitur pada senyawa dilakukan dengan menggunakan PubChem *fingerprint* dan ECFPs.

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 *brown adipose tissue*. Penelitian ini tidak melakukan proses *molecular docking* untuk melihat efektivitas senyawa ketika berikatan dengan protein.

TINJAUAN PUSTAKA

Obesitas

Obesitas adalah gangguan metabolisme kronis yang ditandai dengan peningkatan lemak di dalam tubuh. Obesitas merupakan faktor risiko untuk beberapa penyebab utama kematian, termasuk penyakit jantung, diabetes, dan banyak jenis kanker. Dalam praktik klinis, kegemukan tubuh biasanya diukur dengan *body mass index* (BMI). BMI dihitung sebagai berat badan (*kg*) dibagi dengan tinggi badan (m^2). Seseorang dengan BMI 25 kg/m^2 atau lebih diklasifikasikan sebagai kelebihan berat badan. Pra obesitas dan obesitas kelas I, II dan III didefinisikan sebagai BMI 25 kg/m^2 sampai 29,9 kg/m^2 , 30 kg/m^2 sampai 34,9 kg/m^2 , 35 kg/m^2 sampai 39,9 kg/m^2 , dan 40 kg/m^2 atau lebih besar (Kaila dan Raman 2008)

Machine Learning untuk penemuan obat

Dalam penemuan obat, kecerdasan buatan memberikan banyak pilihan teknik untuk menganalisis, mempelajari, dan menjelaskan bagaimana obat-obat dapat diidentifikasi (Dara *et al.* 2021). Salah satu teknik kecerdasan buatan yang dapat digunakan untuk menemukan obat adalah *machine learning*. Penerapan *machine learning* dilakukan dengan mengamati *Drug-Target Interaction* (DTI). DTI mengacu pada interaksi antara senyawa kimia dan target protein dalam tubuh manusia (Sachdev dan Gupta 2019). Fungsi biologis protein target diaktifkan atau dilemahkan oleh senyawa dengan mengikat langsung ke protein target. Penelitian telah menemukan bahwa obat dapat mengikat banyak protein target dan sebaliknya. Oleh karena itu, diperlukan untuk melakukan identifikasi DTI dalam jumlah yang besar. Untuk mengatasi tersebut, banyak pihak yang menyediakan database biologi

dengan skala yang besar. DTI dapat bernilai positif atau negatif. DTI yang bernilai positif menunjukkan adanya interaksi sedangkan DTI yang bernilai negatif menunjukkan tidak ada atau tidak diketahui interaksi antara pasangan obat-target yang diberikan (Bahi dan Batouche 2018)

Semi-Supervised Learning

Semi-supervised learning (SSL) adalah cabang dari machine learning yang menggunakan data berlabel dan tidak berlabel untuk melakukan tugas tertentu. Secara konsep, SSL memungkinkan untuk memanfaatkan sejumlah data besar yang tidak berlabel yang tersedia dalam banyak kasus penggunaan dalam kombinasi dengan kumpulan data berlabel yang biasanya lebih kecil. Tujuan dari penggunaan SSL adalah memanfaatkan data yang tidak berlabel untuk meningkat hasil akurasi model klasifikasi.

Terdapat dua pendekatan untuk memanfaatkan metode *semi-supervised learning*, *transductive learning* dan *inductive learning*. *Transductive learning* bertujuan untuk menerapkan pengklasifikasi terlatih pada contoh tidak berlabel yang diamati pada waktu pelatihan. *Inductive learning* bertujuan untuk mempelajari pengklasifikasi yang mampu menggeneralisasi ke kejadian yang tidak teramati pada waktu pengujian (Ouali *et al.* 2020).

Stacked Autoencoders

Autoencoders (AE) merupakan salah satu jenis dari jaringan syaraf tiruan yang didesain untuk mengkodekan masukan menjadi representasi yang memiliki makna, dan kemudian mengkodekannya kembali sehingga masukan yang direkonstruksi mirip dengan yang asli. *Autoencoders* termasuk salah jenis *unsupervised learning*, namun dapat digunakan untuk meningkatkan hasil performa klasifikasi *supervised learning* (Bank *et al.* 2021). *Stacked Autoencoder* (SAE) terdiri dari tumpukan AE. Lapisan AE pertama memetakan variabel masukan ke vektor lapisan tersembunyi pertama. Setelah pelatihan lapisan AE pertama selesai dilakukan, rekonstruksi lapisan AE pertama dihapus dan lapisan tersembunyi pertama dijadikan sebagai masukan di lapisan AE selanjutnya. Secara umum, lapisan masukan AE selanjutnya, merupakan lapisan tersembunyi dari AE sebelumnya. Setiap lapisan dilatih menggunakan algoritme *gradient descent* yang sama sebagai lapisan AE tunggal (Bao *et al.* 2017).

Deep Neural Network

Deep Neural Network (DNN) merupakan salah satu bentuk jaringan saraf tiruan. DNN memiliki nama lain yaitu *multilayer perceptron*. DNN memiliki beberapa lapisan dari sekumpulan neuron yang saling terhubung (Schmidhuber 2015). Input bergerak di jaringan melalui layer-layer yang menghitung kemungkinan output-nya. Tiap *hidden layer* memetakan input dari *layer* sebelumnya (x_j) ke output (y_j) yang akan dikirim ke *layer* selanjutnya yang terlihat di Rumus 1.

$$y_j = f(x_j) \quad x_j = b_j + \sum_i y_i w_{ij} \quad (1)$$

$f(x_j)$ adalah fungsi aktivasi, b_j adalah bias dari layer j , i adalah *layer* sebelumnya, j adalah *layer* selanjutnya, dan w_{ij} adalah bobot yang menghubungkan *layer* i dan *layer* j . Dalam pembentukan awal model, bobot dan bias di setiap layer biasanya diinisialisasi secara acak (LeCun *et al.* 2015).

Ketika model telah menghasilkan nilai keluaran, nilai keluaran tersebut akan dibandingkan dengan nilai aslinya dan dilakukan perhitungan nilai galat. *Cross-entropy* digunakan untuk perhitungan galat pada kasus klasifikasi sebanyak 2 kelas. *Cross-entropy*

membandingkan antara nilai asli dengan nilai prediksi yang dihasilkan (Sharma dan Rani 2018)

Setelah didapatkan nilai galat, proses *backpropagation* dilakukan untuk mengurangi nilai galat dalam pelatihan. Nilai galat diperoleh dari selisih output target dan output aktual. *Backpropagation* menggunakan turunan fungsi cost ($f'(C)$) untuk menghitung galat. *Backpropagation* dilakukan dengan menghitung galat pada sebagian data. Pembaruan bobot dilakukan dengan menggunakan *stochastic gradient descent*, yaitu metode iteratif yang digunakan dalam mengoptimasi fungsi objektif (LeCun *et al.* 2015).

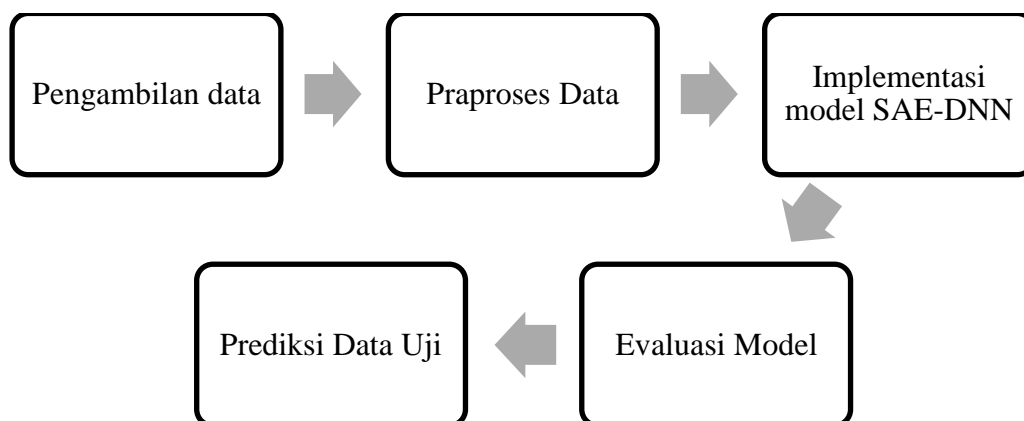
METODE

Data Penelitian

Penelitian ini menggunakan tiga dataset yaitu data protein, data interaksi senyawa-protein, dan data senyawa herbal. Data protein yang digunakan dalam penelitian ini adalah ADRB3, PPAR α , PPAR γ , dan SIRT1. Protein-protein tersebut memiliki peran yang besar terhadap proses *thermogenesis* yang ada di BAT (Qiao *et al.* 2014; Festuccia *et al.* 2011; Andrade *et al.* 2014). Data interaksi senyawa-protein diperoleh dengan mencari interaksi antara protein ADRB3, PPAR α , PPAR γ , dan SIRT1 dengan senyawa yang melalui situs BindingDB (Gilson *et al.* 2016). Data senyawa herbal yang akan digunakan untuk pengujian model diperoleh dari *database* HerbalDB (Erlina *et al.* 2020).

Tahapan Penelitian

Penelitian dilakukan melalui beberapa tahapan, dimulai dari pengambilan data, praproses data, implementasi model, evaluasi dan pengujian model pada data senyawa herbal. Gambar 1 menampilkan alur dari tahapan penelitian yang dilakukan.



Gambar 1 Tahapan Penelitian

Pengambilan Data

Data interaksi senyawa-protein diperoleh dari situs dan BindingDB dengan melakukan pencarian sebanyak empat kali dengan menggunakan kata kunci “ADRB3”, “PPAR α ”, “PPAR γ ”, dan “SIRT1”. Situs BindingDB dapat diakses melalui alamat <https://www.bindingdb.org/bind/index.jsp> sedangkan situs. Hasil dari dari pencarian data interaksi senyawa-protein akan menghasilkan data senyawa yang berkaitan.

Praproses Data

Tahapan pertama dalam praproses data adalah melakukan penyatuan dan penghapusan data berlebih pada data interaksi senyawa-protein yang diambil dari BindingDB. Selanjutnya melakukan ekstraksi fitur dari senyawa. Data senyawa dengan format SDF

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* dan ECFPs. Data ekstraksi tersebut akan digunakan sebagai masukan 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

Pembuatan model didasarkan pada pendekatan *deep semi-supervised*, yakni menggabungkan *unsupervised learning* dan *supervised learning*. Dalam pengujian, digunakan salah satu jenis *unsupervised learning* yaitu *Stacked Autoencoders* (SAE) yang bertujuan untuk mendapatkan nilai bobot awal. Bobot tersebut digunakan untuk model prediksi. Untuk *Supervised learning*, digunakan *Deep Neural Network* (DNN) untuk melakukan proses klasifikasi.

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, membuat 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)	[300; 500; 750; 1000]
Jumlah <i>hidden neuron</i> di <i>layer</i> ke-0 (HL_i)	[0,5; 0,66; 0,75]
<i>Activation function</i>	[ReLU, Sigmoid, tanh]
<i>Dropout rate</i>	[0,1;0,3;0,5]

<i>Hidden layer</i>	[2;3;4]
<i>Learning rate</i>	[0,01;0,001;0,0001;0,00001]

Evaluasi Model

Evaluasi model dilakukan dengan menggunakan metode *Cross-validation*. Metode tersebut bekerja dengan membagi data menjadi dua bagian dengan tujuan yang berbeda, yaitu untuk melatih model dan untuk memvalidasi model. Pada pengujian ini digunakan salah satu metode *Cross-validation* yaitu *k-fold cross validation*. Pada *k-fold cross validation*, data dibagi menjadi bagian yang berukuran sama sebanyak k. Selanjutnya, tiap bagian k akan dilakukan iterasi pelatihan dan validasi (Refaeilzadeh *et al.* 2016). Pada pengujian ini, digunakan k sebanyak 10 kali.

Metrik yang digunakan untuk evaluasi model adalah *accuracy*, *precision*, *recall*, *F-measure* dan *Area Under Receiver-Operating Characteristic* (AUROC). *Accuracy* 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). *Precision* 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). ROC berfungsi mengevaluasi keakuratan model statistik yang mengklasifikasikan subjek ke dalam 1 dari 2 kategori. AUROC merupakan ukuran akurasi tes yang memberikan bobot yang sama untuk rentang penuh nilai ambang (Zou *et al.* 2007).

Pengujian Model

Untuk mendapatkan kualitas prediksi yang baik, dilakukan perbandingan antara hasil dari 2 cara pengubahan fitur senyawa menjadi representasi numerik (PubChem *fingerprint* dan ECFPs). Perbandingan dilakukan menggunakan *k-fold cross-validation* dan AUROC. 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 dan R
 - Pandas, Numpy, dan Matplotlib untuk manipulasi, analisa, dan visualisasi data
 - Tensorflow untuk pemodelan SAE-DNN
 - Google Colab untuk pengembangan model SAE-DNN dan pemanfaatan GPU
 - Library ChemmineR untuk ekstraksi fitur senyawa

JADWAL PENELITIAN

Penelitian ini akan dilaksanakan mulai bulan Februari hingga Juli tahun 2022. Jadwal penelitian dapat dilihat pada Tabel 2.

Tabel 2 Jadwal Penelitian

No	Kegiatan	Tahun 2022
----	----------	------------

		Februari	Maret	April	Mei	Juni	Juli
1	Kolokium	■					
2	Pengumpulan Data		■				
3	Praproses Data		■				
4	Pengembangan Model		■	■			
5	Evaluasi Model		■	■			
6	Pengujian Model			■			
7	Penulisan Laporan Tugas Akhir				■		
8	Seminar				■		
9	Revisi					■	■
10	Sidang					■	■

DAFTAR PUSTAKA

- Andrade JMO, Frade ACM, Guimarães JB, Freitas KM, Lopes MTP, Guimarães ALS, de Paula AMB, Coimbra CC, Santos SHS. 2014. Resveratrol increases brown adipose tissue thermogenesis markers by increasing SIRT1 and energy expenditure and decreasing fat accumulation in adipose tissue of mice fed a standard diet. *Eur. J. Nutr.* 53(7):1503–1510.doi:10.1007/s00394-014-0655-6.
- Bahi M, Batouche M. 2018. Drug-Target Interaction Prediction in Drug Repositioning Based on Deep Semi-Supervised Learning. Di dalam: Amine A, Mouhoub M, Ait Mohamed O, Djebbar B, editor. *Computational Intelligence and Its Applications*. [internet] Vol. 522. Cham. Cham: Springer International Publishing. (IFIP Advances in Information and Communication Technology). hlm. 302–313. [diunduh 2021 Nov 8]. Tersedia pada: http://link.springer.com/10.1007/978-3-319-89743-1_27
- Bank D, Koenigstein N, Giryes R. 2021 Apr 3. Autoencoders. *ArXiv200305991 Cs Stat.* [diunduh 2021 Des 17]. Tersedia pada: <http://arxiv.org/abs/2003.05991>
- Bao W, Yue J, Rao Y. 2017. A deep learning framework for financial time series using stacked autoencoders and long-short term memory. Podobnik B, editor.[editorial]. *PLOS ONE*. 12(7):e0180944.doi:10.1371/journal.pone.0180944.
- 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.
- Dara S, Dhamercherla S, Jadav SS, Babu CM, Ahsan MJ. 2021 Agu 11. Machine Learning in Drug Discovery: A Review. *Artif. Intell. Rev.*.doi:10.1007/s10462-021-10058-4. [diunduh 2022 Feb 16]. Tersedia pada: <https://link.springer.com/10.1007/s10462-021-10058-4>

- De Lorenzo A, Grattereri 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.
- Erlina L, Paramita RI, Kusuma WA, Fadilah F, Tedjo A, Pratomo IP, Ramadhanti NS, Nasution AK, Surado FK, Fitriawan A, *et al.* 2020. Virtual Screening on Indonesian Herbal Compounds as COVID-19 Supportive Therapy: Machine Learning and Pharmacophore Modelling Approaches. In Review. [diunduh 2022 Feb 17]. Tersedia pada: <https://www.researchsquare.com/article/rs-29119/v2>
- 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.
- Festuccia WT, Blanchard P-G, Deshaies Y. 2011. Control of Brown Adipose Tissue Glucose and Lipid Metabolism by PPAR γ . *Front. Endocrinol.* 2.doi:10.3389/fendo.2011.00084. [diunduh 2022 Feb 16]. Tersedia pada: <http://journal.frontiersin.org/article/10.3389/fendo.2011.00084/abstract>
- 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>
- Kaila B, Raman M. 2008. Obesity: A Review of Pathogenesis and Management Strategies. *Can. J. Gastroenterol.* 22(1):61–68.doi:10.1155/2008/609039.
- 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.
- Ouali Y, Hudelot C, Tami M. 2020 Jul 6. An Overview of Deep Semi-Supervised Learning. *ArXiv200605278 Cs Stat.* [diunduh 2022 Feb 8]. Tersedia pada: <http://arxiv.org/abs/2006.05278>
- 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>
- Qiao L, Yoo H sun, Bosco C, Lee B, Feng G-S, Schaack J, Chi N-W, Shao J. 2014. Adiponectin reduces thermogenesis by inhibiting brown adipose tissue activation in mice. *Diabetologia.* 57(5):1027–1036.doi:10.1007/s00125-014-3180-5.
- Refaeilzadeh P, Tang L, Liu H. 2016. Cross-Validation. Di dalam: Liu L, Özsu MT, editor. *Encyclopedia of Database Systems*. [internet] New York, NY. New York, NY: Springer New York. hlm. 1–7. [diunduh 2021 Des 17]. Tersedia pada: http://link.springer.com/10.1007/978-1-4899-7993-3_565-2
- Sachdev K, Gupta MK. 2019. A comprehensive review of feature based methods for drug target interaction prediction. *J. Biomed. Inform.* 93:103159.doi:10.1016/j.jbi.2019.103159.

- Schmidhuber J. 2015. Deep learning in neural networks: An overview. *Neural Netw.* 61:85–117.doi:10.1016/j.neunet.2014.09.003.
- 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.
- Sharma A, Rani R. 2018. BE-DTI': Ensemble framework for drug target interaction prediction using dimensionality reduction and active learning. *Comput. Methods Programs Biomed.* 165:151–162.doi:10.1016/j.cmpb.2018.08.011.
- 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.
- Sulistiawan 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.
- Zou KH, O'Malley AJ, Mauri L. 2007. Receiver-Operating Characteristic Analysis for Evaluating Diagnostic Tests and Predictive Models. *Circulation.* 115(5):654–657.doi:10.1161/CIRCULATIONAHA.105.594929.