Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK) DOI: 10.25126/jtiik.2023106603
Vol. 10, No. 3, Juni 2023, hlm. 531-542 p-ISSN: 2355-7699
Akreditasi KEMENRISTEKDIKTI, No. 36/E/KPT/2019 e-ISSN: 2528-6579

PREDIKSI INTERAKSI *DRUG TARGET* PADA GEN KANKER MENGGUNAKAN METODE LASSO-XGBOOST

Muh. Fadhil Al-Haaq Ginoga[1], Wisnu Ananta Kusuma*[2], Mushthofa[3]

[1,2,3]Institut Pertanian Bogor, Bogor
Email: [1]dilloginoga@apps.ipb.ac.id, [2]ananta@apps.ipb.ac.id, [3]M@apps.ipb.ac.id
*Penulis Korespondensi

(Naskah masuk: 06 Oktober 2022, diterima untuk diterbitkan: 20 Juni 2023)

Abstrak

Pengobatan kanker saat ini sering dilakukan dengan kemoterapi menggunakan obat kimia dan dapat menyebabkan efek samping. Alternatif pengobatan dapat menggunakan senyawa herbal yang diketahui memiliki efek samping lebih sedikit. Analisis *Drug Target Interaction* (DTI) dapat dilakukan untuk mengetahui interaksi senyawa herbal terhadap protein kanker. Pada penelitian ini dilakukan perancangan model prediksi DTI dengan melakukan seleksi fitur pada *dataset* menggunakan *Least Absolute Shrinkage and Selection Operator* (LASSO) lalu dilakukan penyeimbangan data dengan *Synthetic Minority Oversampling Technique* (SMOTE) dan diprediksi menggunakan *Extreme Gradient Boosting* (XGBoost). Data protein terkait kanker didapatkan dari daftar *Cancer Gene Census*, dari daftar tersebut dilakukan penelusuran pada database GDSC, DrugCentral, dan DrugBank untuk menghasilkan daftar senyawa obat yang berinteraksi dengan protein tersebut. Selain itu, senyawa herbal dihasilkan dari *database* HerbalDB dan Knapsack. Pengujian dilakukan pada beberapa jenis ekstraksi fitur seperti CTD, DC, PseAAC, dan PSSM. Hasil prediksi menunjukkan beberapa senyawa herbal seperti *andrographolide*, *ursolic acid* dan *oleanolic acid* memiliki interaksi pada protein terkait kanker. Selain itu, LASSO-XGBoost dapat memprediksi DTI pada kanker dengan skor F1 0,861; AUROC 0,927; *recall* 0,85; *precision* 0,866; dan *accuracy* 0,897.

Kata kunci: *interaksi obat target, LASSO, XGBoost, SMOTE, kanker, senyawa herbal, protein, senyawa*

DRUG TARGET INTERACTION PREDICITION ON CANCER GENE USING LASSO-XGBOOST METHOD

Abstract

Currently, cancer treatment is usually done with chemotherapy using chemical drugs that can cause side effects. An alternative treatment can use herbal compounds that known have fewer side effects. Drug Target Interaction analysis (DTI) can be performed to determine the interaction of herbal compounds with cancer proteins. In this study, a DTI prediction model is built by selecting features on the data set using Least Absolute Shrinkage and Selection Operator (LASSO) then data balancing performed with Synthetic Minority Oversampling Technique (SMOTE) and Extreme Gradient Boosting (XGBoost) performed to predict the interaction. The cancer-associated protein data were obtained from the Cancer Gene Census list, then the list used to search on the GDSC, DrugCentral and DrugBank databases to generate a list of drug compounds that interact with these proteins. In addition, plant compounds to be generated from the HerbalDB and Knapsack databases. Tests were performed on several types of feature extraction such as CTD, DC, PseAAC and PSSM. Predictive results suggest that several herbal compounds such as andrographolide, ursolic acid and oleanolic acid interact with cancer-associated proteins. In addition, LASSO-XGBoost was able to predict DTI in cancer with score of F1 0,861; AUROC 0,927; recall 0,857, precision 0,866; and accuracy 0,897.

Keywords: *drug target interaction, LASSO, XGBoost, SMOTE, cancer, herbal compounds, protein, compound*

1. PENDAHULUAN

Kanker merupakan suatu penyakit mematikan yang disebabkan oleh pertumbuhan sel yang tidak normal (Tim CancerHelps, 2019). Penyakit kanker dapat disebabkan oleh faktor eksternal (kimia, radiasi, dan infeksi) dan faktor internal (mutasi, hormon, kondisi imun, dan mutasi acak) (Mathur et al., 2015). Terapi pengobatan penyakit kanker atau kemoterapi saat ini umumnya masih menggunakan obat kimia. Namun, penggunaan obat kimia dapat memberikan berbagai efek samping pada penderita kanker seperti resistensi obat (Hussain et al., 2021).

Oleh karena itu, diperlukan alternatif pengobatan lain untuk mengurangi efek samping tersebut, salah satunya menggunakan tanaman obat.

Tanaman obat sering digunakan untuk menjaga kesehatan dan khasiatnya diketahui secara turun-temurun. Tanaman obat mengandung banyak zat aktif atau senyawa herbal yang berkhasiat dan diyakini tidak memiliki efek samping seperti obat kimia (Hernani, 2011). Terdapat beberapa senyawa herbal yang baik digunakan untuk pengobatan penyakit kanker, salah satunya *flavonoid*. *Flavonoid* secara alami terdapat dalam tumbuhan dan dapat digunakan sebagai kemoterapi atau agen kemoprefentif untuk mengurangi risiko kanker dengan efek samping yang lebih sedikit (Gürler et al., 2020). Oleh sebab itu, terdapat potensi besar bagi senyawa herbal untuk digunakan dalam mengobati penyakit kanker. Namun, permasalahan yang dihadapi adalah pencarian senyawa herbal yang dapat digunakan dalam pengobatan protein kanker cukup sulit diidentifikasi mengingat kombinasi interaksinya yang cukup banyak. Oleh karena itu, diperlukan suatu metode yang efisien, yaitu penyaringan *in silico* dengan membangun model prediksi interaksi senyawa obat dan protein target (*drug target*) atau *Drug Target Interaction* (DTI) (Wijaya et al., 2021).

Prediksi DTI dengan data senyawa obat dan protein dapat dilakukan dengan teknik *machine learning* (Xu et al., 2021). Penerapan teknik *machine learning* membutuhkan fitur yang berisi informasi senyawa dan protein sebagai inputnya. Fitur senyawa dan protein berperan penting untuk mendapatkan kualitas model *machine learning* yang baik. Namun, umumnya fitur yang terbentuk dari ekstraksi fitur tersebut sangat besar dan akan menjadi suatu masalah baru karena dapat menyebabkan model tidak efisien (Shi et al., 2019). Berbagai metode seleksi fitur telah dikembangkan, salah satunya metode *Least Absolute Shrinkage and Selection Operator* (LASSO) yang diusulkan oleh Tibshirani (2011). Metode LASSO dapat mengurangi dimensi dengan menghapus fitur yang *noisy* dan *redundant* (Patil dan Kim, 2020).

Penelitian sebelumnya terkait optimasi pemodelan DTI menggunakan LASSO sebagai seleksi fitur telah dilakukan oleh You et al. (2019), yang melakukan prediksi DTI menggunakan *Deep Neural Network* untuk mendapatkan obat potensial pada penyakit kanker payudara. Penelitian ini menggunakan *chemical properties, topological properties,* dan *geometrical properties* dari senyawa sedangkan *amino acid composition* (AAC), *dipeptide composition* (DC), dan *tripeptide composition* sebagai fitur protein. Penelitian oleh Shi et al. (2019), yang mengusulkan metode LASSO *Random Forest*-DTI (LRF-DTI) menggunakan *molecular fingerprint* FP2 dan *pseudo-position specific scoring matrix* (PsePSSM) sebagai *feature vector*, yang selanjutnya dimodelkan menggunakan *Random Forest* sebagai *classifier* yang dioptimasi dengan teknik *oversampling Synthetic Minority Oversampling Technique* (SMOTE). Penelitian oleh Mahmud et al. (2020) yang mengusulkan metode DeepACTION menggunakan struktur senyawa dan informasi sekuens protein sebagai *feature vector* dengan pemodelan menggunakan *Convolutional Neural Network*. Penelitian tersebut menggunakan *topological* dan *geometrical feature* dari senyawa sedangkan AAC, *pseudo*-AAC (PseAAC), *dipeptide composition*, *autocorrealation*, *quasi-sequence-order* dan *composition*, *transition, and distribution* (CTD) sebagai fitur protein.

Penelitian DTI lainnya dilakukan oleh Thafar et al. (2021) yang mengusulkan metode DTi2Vec. Metode tersebut mengidentifikasi DTI dengan pembelajaran jaringan dan *ensemble*. Metode tersebut membangun jaringan untuk menghasilkan *feature vector* yang selanjutnya diprediksi dengan metode *ensemble Extreme Gradient Boosting* (XGBoost). Hasilnya, XGBoost dapat memberikan prediksi yang sangat baik dan dianggap *robust* karena memiliki *regularization parameter* yang dapat mengurangi *variance* pada model.

Tujuan dari penelitian ini yaitu mengusulkan pendekatan baru dengan mengkombinasikan metode LASSO dan XGBoost dalam melakukan prediksi DTI dan mengevaluasi performanya dalam memprediksi DTI pada kanker. Selain itu, penggunaan beberapa jenis ekstraksi fitur juga diperlukan untuk mengetahui jenis ekstraksi fitur yang paling sesuai dengan model yang digunakan. Prediksi DTI yang dilakukan berbasis *feature based* dengan membentuk *feature vector* dari *descriptor* senyawa dan protein. Selanjutnya, dilakukan seleksi fitur menggunakan metode LASSO dan menangani *data imbalance* dengan SMOTE. Setelah itu, pemodelan dilakukan menggunakan metode XGBoost dan dilakukan prediksi pada senyawa herbal untuk mengetahui kandidat alternatif obat penyakit kanker.

Penelitian ini diharapkan dapat membantu peneliti dibidang kesehatan khususnya penerapan senyawa herbal sebagai obat dan jamu yang bertujuan memperkecil ruang pencarian dalam mengidentifikasi kandidat senyawa herbal yang berpotensi sebagai obat dalam pengobatan kanker.

2. METODE PENELITIAN

2.1. Data Penelitian

Data penelitian yang digunakan dikhususkan pada protein gen terkait kanker. Protein yang digunakan diambil dari daftar *Cancer Gene Census* (CGC). CGC merupakan daftar gen-gen yang dapat bermutasi dan berperan dalam penyakit kanker. Daftar CGC diperoleh dari *database International Cancer Genome Consortium* (ICGC). Penelitian ini juga menggunakan data *screening* dari *database Genomics of Drug Sensitivity in Cancer* (GDSC) sebagai sumber data. GDSC merupakan suatu *website* yang menyediakan informasi hasil *screening* interaksi senyawa obat dan gen kanker. GDSC dapat di akses pada tautan https://www.cancerrxgene.org/.

Informasi *sequence* protein diperoleh dari Uniprot (Bateman et al., 2022). Selanjutnya, senyawa yang berinteraksi dengan protein diperoleh dari DrugBank (Wishart et al., 2006) dan DrugCentral (Avram et al., 2021). Sedangkan, informasi senyawa diperoleh dari PubChem (Kim et al., 2021). Jumlah data yang dihasilkan, diantaranya 180 protein, 1072 senyawa obat, 2193 interaksi. Selain itu, diperoleh 403 senyawa herbal dari pengumpulan data pada *database* Knapsack (Afendi et al., 2012) dan herbalDB (Syahdi et al., 2019).

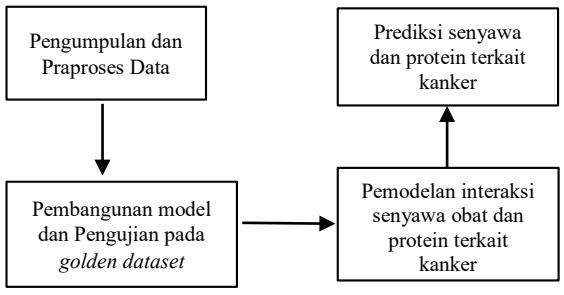
Penelitian ini juga menggunakan *golden dataset* dari penelitian Yamanishi et al., (2008) yang sudah sering digunakan sebagai data pembanding dalam penelitian terkait DTI. *Golden dataset* terdiri dari beberapa jenis protein, yaitu *ion channel* (IC), *G-protein coupled receptor* (GPCR), *enzyme*, dan *nuclear receptor* (NR). Jumlah data pada *golden dataset* dapat dilihat pada Tabel 1

Tabel 1. Jumlah data *golden dataset*

Dataset	*Drug*	*Protein*	*Interaction*
IC	210	204	1476
GPCR	223	95	635
NR	54	26	90
Enzyme	445	664	2926

2.2. Tahapan Penelitian

Penelitian yang dilakukan terdiri atas beberapa tahapan (Gambar 1). Tahapan dimulai dengan mengumpulkan data yang akan digunakan lalu di praproses untuk mendapatkan data yang bersih. Setelah itu, dilakukan pembangunan model dan pengujian pada setiap fitur yang digunakan untuk dibandingkan pada *golden dataset*. Fitur terbaik akan digunakan pada pemodelan interaksi senyawa obat dan kanker. Model dengan *hyperparameter* terbaik selanjutnya digunakan untuk memprediksi interaksi senyawa herbal dan protein terkait kanker.



Gambar 1. Tahapan penelitian

Setiap pemodelan dilakukan beberapa tahapan (Gambar 2) dimulai dari praproses dan ekstraksi data, seleksi fitur menggunakan metode LASSO, *minority oversampling* dengan SMOTE, pemodelan menggunakan metode XGBoost, dan evaluasi model serta prediksi senyawa herbal pada protein terkait kanker.

2.3. Praproses Data

Data yang digunakan terdiri dari interaksi senyawa obat dan protein (DTI). Data protein terlebih dahulu di ekstraksi menjadi beberapa jenis fitur protein, diantaranya PseAAC, CTD, PSSM, dan DC.

PseAAC dapat digunakan untuk mendapatkan informasi sekuens protein dengan model diskrit tanpa kehilangan informasi urutan sekuens. PseAAC memiliki 20 fitur komponen asam amino dan beberapa informasi urutan asam amino pada *sequence*.

CTD merupakan *descriptor* protein yang diusulkan oleh Dubchak et al., pada tahun 1995. CTD bisa menjelaskan posisi dari asam amino dan informasi *physicochemical* dari *sequence*. Selain itu, metode ini memberikan informasi komposisi dari asam amino dan frekuensi kemunculan *dipeptide*.

PSSM merupakan fitur yang memberikan nilai peluang untuk setiap asam amino dari sekuens. Nilai peluang merepresentasikan peluang residu protein bermutasi menjadi suatu asam amino selama proses evolusi pada *multiple sequence alignment* (Wang et al., 2018).

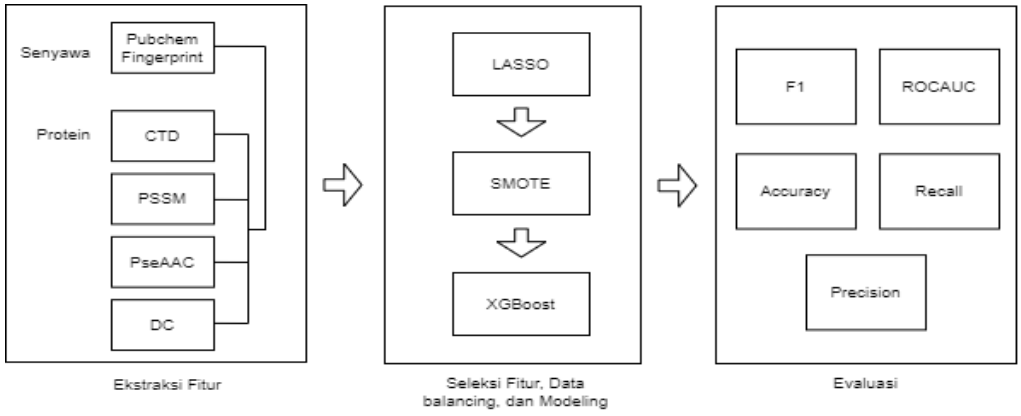
DC merupakan rasio kemunculan *dipeptide*. DC memberikan informasi ketetanggaan dari asam amino pada sekuens protein. Pada sekuens protein terdapat 400 *dipeptide.* Gambar 3 merepresentasikan proses ekstraksi fitur protein pada seluruh jenis fitur yang digunakan.

Fusion merupakan gabungan dari setiap fitur protein yang digunakan. PseAAC dapat memberikan informasi komposisi asam amino dan urutan sekuennya. CTD dapat memberikan informasi *physicochemical* dari sekuens, dan PSSM dapat memberikan informasi peluang evolusi dari asam amino. Fitur ini diharapkan memiliki informasi gabungan yang lebih banyak untuk dimodelkan. Setiap teknik ekstraksi fitur dilakukan dengan parameter *default*.

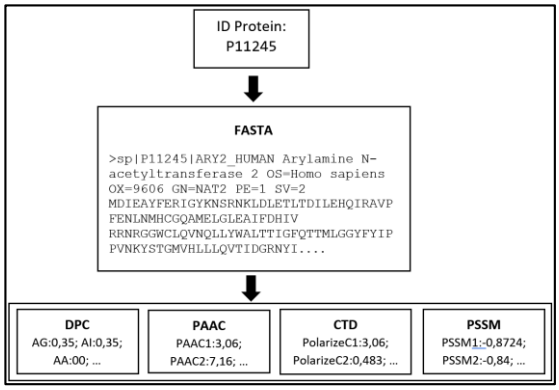
Jumlah fitur yang dihasilkan dari ekstraksi fitur dapat dilihat pada Tabel 2. Ekstraksi fitur pada senyawa menggunakan Pubchem *fingerprint.* Pubchem *fingerprint* menggambarkan keberadaan struktur pada senyawa yang di representasikan dengan nilai biner. Gambar 4 merupakan proses ekstraksi fitur pada senyawa.

Tabel 2. Jumlah fitur setiap jenis representasi fitur

Fitur	**Jumlah Fitur**
PseAAC	30
DC	400
PSSM	220
CTD	146
Pubchem	881



Gambar 2. Tahapan pemodelan

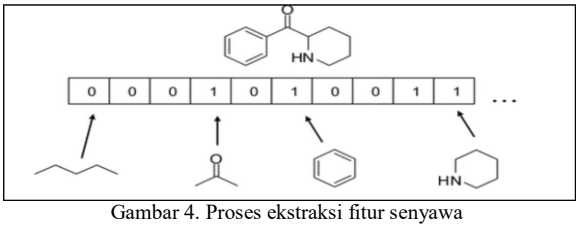


Gambar 3. Proses ekstraksi fitur protein

2.4. Pembangkitan Sampel Berlabel 0 dan 1

Interaksi DTI yang dimiliki hanya berlabel 1 (positif), sehingga dilakukan pembangkitan label 0 (belum diketahui interaksinya). Pembangkitan label 0 dilakukan dengan menyilangkan setiap kombinasi interaksi senyawa dan protein dalam bentuk *complete bipartite graph*.

Jumlah label 0 yang dibentuk sangat banyak dibandingkan label 1, maka dilakukan *random undersampling* dengan mengambil sampel berlabel 0 sejumlah tiga kali jumlah label 1. Setelah itu, untuk menyeimbangkan label 1 dan 0 dilakukan *oversampling* pada data berlabel minoritas menggunakan teknik SMOTE.



Gambar 4. Proses ekstraksi fitur senyawa

2.5. Metode Seleksi Fitur LASSO

Metode LASSO merupakan pengembangan dari metode *Ridge Regression*. Perbedaannya terletak pada koefisien regresi. *Ridge Regression* hanya mampu menyusutkan koefisien regresi mendekati nol, sedangkan LASSO dapat menyusutkan koefisien regresi hingga tepat nol. Selain itu, LASSO dapat menyelesaikan model regresi yang memiliki multikolinearitas. Berdasarkan hal tersebut LASSO memiliki kelebihan sebagai seleksi fitur. Metode LASSO mengecilkan koefisien taksiran pada model untuk mempertahankan fitur yang penting dalam seleksi fitur. Koefisien taksiran dapat dilihat pada rumus 1. LASSO meminimalkan koefisien taksiran untuk mengurangi jumlah *predictor* pada model. Selanjutnya, koefisien taksiran yang minimal dijumlahkan dengan nilai lambda. Semakin tinggi nilai lambda, koefisien semakin mendekati nol hingga tepat nol (Datta et al., 2017).

$$argmin\ \sum_{i=1}^{N}\left(y_i - \sum b_j x_{ij}\right)^2 + \lambda\sum|b_j| \qquad (1)$$

2.6. Teknik Oversampling SMOTE

SMOTE merupakan metode *oversampling* data yang dapat menangani data yang tidak seimbang dengan melakukan *resampling* pada data. Metode SMOTE dapat melakukan *oversampling* pada data minoritas dengan membuat replikasi atau data sintetis. Metode SMOTE melakukan *k-Nearest Neighbors* untuk membangkitkan data kelas minoritas sesuai dengan jumlah duplikasi yang dibutuhkan. Metode SMOTE telah digunakan di beberapa penelitian terkait DTI untuk menyeimbangkan data. Metode ini memiliki performa yang baik dalam kasus data yang tidak seimbang (Shi et al., 2019; Peng et al., 2022)

2.7. Model XGBoost

Gradient Boosting (GB) merupakan algoritme yang dapat digunakan sebagai *regressor* maupun *classifier*. GB menerapkan *ensemble learning* menggunakan *Decision Tree* dan dioptimasi dengan GB untuk meminimalkan nilai *loss function*. GB mengoptimasi dengan melakukan iterasi sebanyak M, $1 \leq m \leq M$ dengan membentuk F_m sebanyak M *tree*. Algoritme GB menghitung nilai *loss* dalam bentuk *sum of squared residual* secara *diferensiable* sehingga membentuk *loss* dalam bentuk *gradient*. Selanjutnya, *gradient descent* digunakan untuk

mendapatkan *loss* yang paling minimum. Setiap iterasi GB akan mengurangi nilai rata-rata *loss function* seminimal mungkin dari fungsi awal $F_0(x)$ menggunakan rumus 2.

$$argmin \sum L(y_i, f(m-1)(x_i) + \gamma_m h_m(x_i)) \quad (2)$$

XGBoost yang dikembangkan oleh Chen dan Guestrin (2016) merupakan suatu algoritme yang menerapkan konsep *Gradient Boosting*. XGBoost memiliki parameter tambahan yaitu pinalti pada tiap *tree*. Bobot dari *tree* yang dihasilkan akan dijumlahkan agar mendapatkan akurasi yang baik.

2.8. *Hyperparameter Tuning*

Hyperparameter tuning dilakukan untuk mendapatkan kombinasi *hyperparameter* yang optimal pada model. Setiap kombinasi *hyperparameter* dicobakan secara bergantian pada model dan dipilih satu kombinasi dengan hasil evaluasi terbaik. Nilai *hyperparameter* yang dicobakan pada model dapat dilihat pada Tabel 3.

Tabel 3. Nilai *hyperparameter* model pada percobaan

Hyperparameter	Nilai
n_estimators	100, 500, 1000
max_depth	4, 6, 8
learning_rate	0,01; 0,1; 0,2
min_child_weight	1, 4, 8
colsample_bytree	0,5; 0,25
subsample	0,5

Deskripsi *hyperparameter* yang digunakan adalah sebagai berikut.

1. *n_estimators* : Jumlah *gradient boosted trees*
2. *max_depth* : kedalaman maksimal *tree*
3. *learning_rate* : ukuran langkah untuk mendapatkan *loss function* minimum
4. *min_child_weight* : Jumlah minimum berat *instance* yang dibutuhkan pada *child tree*
5. *colsample_bytree* : rasio *subsampling* kolom setiap konstruksi *tree*
6. *subsample* : rasio pengambilan sampel pelatihan pada data

2.9. Evaluasi Performa

Tahap ini dilakukan bersamaan dengan pemodelan menggunakan *Cross Validation* (CV) untuk menguji kemampuan model dalam memprediksi data. Penelitian ini menggunakan *Stratified k-Fold* CV untuk menguji model. Pada penelitian ini hanya melakukan *hyperparameter* menggunakan CV, serta menggunakan rata-rata CV untuk setiap *fold*. Selanjutnya, model dievaluasi dengan meperhatikan label yang diprediksi dengan label aktualnya menggunakan *confussion matrix*. Tabel 4 merupakan representasi *confussion matrix*. Selain itu, evaluasi dilakukan dengan memperhatikan beberapa metrik seperti *accuracy*, *recall*, *precision*, dan *F-measure* (F1).

Tabel 4. *Confussion matrix*

	Predicted Positive	*Predicted Negative*
Actual Positive	*True Positive* (TP)	*False Negative* (FN)
Actual Negative	*False Positive* (FP)	*True Negative* (TN)

Accuracy merupakan persentase data uji yang benar diprediksi oleh model. Perhitungan *Accuracy* dapat dilihat pada rumus 2.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

Recall atau *True Positive Rate* (TPR) merupakan ketepatan model dalam memprediksi kelas positif dan menyatakan rasio kelas positif yang diprediksi benar dari kelas positif. Perhitungan *Recall* dapat dilihat pada rumus 4.

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

Precision merupakan ketepatan model dalam melakukan prediksi positif dan menyatakan rasio kelas positif yang diprediksi benar dari semua prediksi positif. Perhitungan *Precision* dapat dilihat pada rumus 5.

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

F1 biasa disebut sebagai *harmonic mean* dari *precision* dan *recall*. Metrik ini mampu memberikan evaluasi model dengan baik walaupun menggunakan data yang tidak seimbang. F1 dapat menyatakan performa kelas minoritas secara menyeluruh dan mengatasi perbandingan kualitas saat TP dan FP meningkat secara bersamaan (Lin et al., 2014). Perhitungan F1 dapat dilihat pada rumus 6.

$$F1 = \frac{2 \times Precision \times Recall}{Precision \times Recall} \quad (6)$$

Kurva *Area Under Receiver Operating Characteristic* (AUROC) didapatkan dengan cara membuat grafik antara *True Positive Rate* (TPR) dan *False Positive Rate* (FPR) dari beberapa nilai *threshold* di rentang [0,1] dan menghitung luas dibawah garis *Receiver Operating Characteristic* (ROC). Kurva *Precision* dan *Recall* atau *Area Under Curve Precision Recall* (AUCPR) juga digunakan dalam evaluasi hasil analisis.

3. HASIL DAN PEMBAHASAN

3.1. Praproses Data

Praproses dilakukan pada setiap data untuk mendapatkan representasi setiap jenis ekstraksi fitur.

Fitur senyawa Pubchem *fingerprint* diperoleh menggunakan *library* python PubChemPy, PubChemPy dapat diakses pada https://pypi.org/project/PubChemPy/. Fitur Protein CTD, DC dan PseAAC diperoleh menggunakan *library* python propy3, propy3 dapat diakses pada tautan https://pypi.org/project/propy3/. Fitur Protein PSSM diperoleh dengan melakukan *alignment* menggunkana psiblast terlebih dahulu dan dihasilkan matrik PSSM yang selanjutnya digunakan untuk membangkitkan *feature vector* PSSM menggunakan *library* R PSSMCOOL yang dikembangkan oleh Mohammadi et al. (2022). Fitur *fusion* dihasilkan dengan menggabungkan setiap jenis ekstraksi fitur untuk mendapatkan representasi fitur yang berbeda dan lebih banyak.

3.2. Pembangunan Model LASSO-XGBoost

Prediksi DTI dilakukan dengan metode XGBoost. Pemilihan fitur terlebih dahulu dilakukan pada *dataset* menggunakan metode LASSO. Seleksi fitur bertujuan untuk mengurangi jumlah dimensi pada *dataset* sehingga dapat mempercepat proses komputasi saat pemodelan. Setelah itu, dilakukan *minority oversampling* pada dataset menggunakan SMOTE untuk menyelesaikan permasalahan *data imbalance*. Pelatihan dengan model XGBoost dilakukan pada data yang sudah seimbang. Model yang dilatih dilakukan *hyparameter tuning* untuk mencari parameter yang optimal menggunakan CV dan dievaluasi untuk mendapatkan hasil terbaik.

3.3. Pengujian Pada *Golden Dataset*

Perbandingan dilakukan pada setiap kombinasi fitur DTI menggunakan *golden dataset*. Setiap kombinasi fitur protein dan senyawa dicoba pada empat jenis *dataset* tersebut. Pengujian dilakukan dengan 5-*fold* CV pada model LASSO-XGBoost. Hasil percobaan pada *golden dataset* menggunakan *grid search* selengkapnya dapat dilihat pada lampiran tambahan dengan rujukan *Folder* 1.

Tabel 5. Hasil pengujian pada model dengan *dataset enzyme*

Fitur	F1	AUROC	*Accuracy*	*Precision*	*Recall*
PseAAC	0,899	0,958	0,925	0,901	0,897
CTD	0,879	0,942	0,910	0,884	0,874
PSSM	**0,912**	**0,966**	**0,934**	**0,915**	**0,908**
DPC	0,880	0,942	0,912	0,890	0,872
Fusion	0,890	0,956	0,919	0,897	0,885

Tabel 6. Hasil pengujian pada model dengan *dataset* NR

Fitur	F1	AUROC	*Accuracy*	*Precision*	*Recall*
PseAAC	0,789	0,821	0,844	0,797	0,785
CTD	0,793	**0,828**	**0,847**	0,799	0,790
PSSM	**0,794**	0,825	**0,847**	0,798	0,794
DPC	0,791	0,814	**0,847**	**0,802**	0,787
Fusion	0,792	0,822	0,841	0,790	**0,798**

Tabel 7. Hasil pengujian pada model dengan *dataset* GPCR

Fitur	F1	AUROC	*Accuracy*	*Precision*	*Recall*
PseAAC	**0,844**	**0,908**	**0,885**	**0,853**	0,837
CTD	0,832	0,906	0,875	0,837	0,829
PSSM	0,843	0,907	0,883	0,845	**0,842**
DPC	0,837	0,900	0,881	0,848	0,828
Fusion	0,837	0,906	0,879	0,842	0,834

Tabel 8. Hasil pengujian pada model dengan *dataset* IC

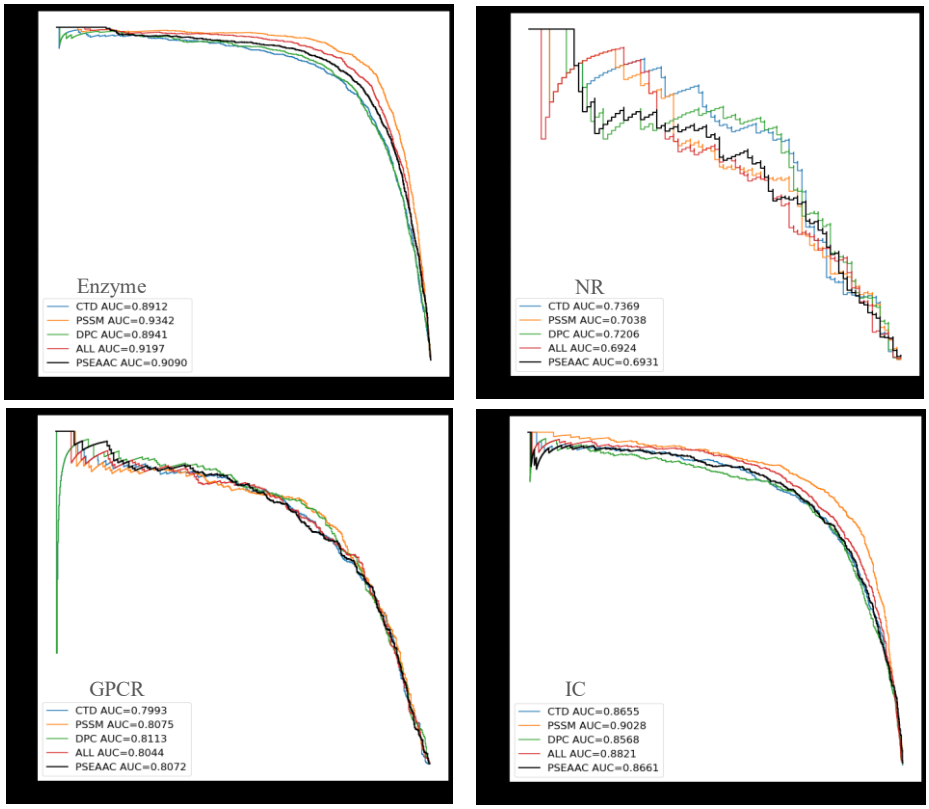
Fitur	F1	AUROC	*Accuracy*	*Precision*	*Recall*
PseAAC	0,866	0,942	0,902	0,875	0,859
CTD	0,847	0,930	0,887	0,853	0,842
PSSM	**0,885**	**0,954**	**0,914**	**0,888**	**0,881**
DPC	0,848	0,929	0,889	0,860	0,838
Fusion	0,856	0,936	0,892	0,857	0,854

Hasilnya, pada *dataset enzyme* (Tabel 5) fitur PSSM memberikan hasil yang paling baik dari setiap skor pengujian dengan skor F1 0,912; AUROC 0,966; *accuracy* 0,934; *precision* 0,915; dan *recall* 0,908. Pada *dataset* NR (Tabel 6) fitur PSSM memberikan hasil F1 paling baik dengan skor 0,794; fitur CTD paling baik pada metrik AUROC dengan skor 0,828; fitur DPC paling baik pada *precision* dengan skor 0,802; dan metrik *fusion* paling baik pada *recall* dengan skor 0,798. Pada *dataset* GPCR (Tabel 7) fitur pseAAC memberikan hasil paling baik, namun PSSM memberikan *recall* paling baik dengan skor 0,842. Pada *dataset* IC (Tabel 8), PSSM memberikan hasil prediksi paling baik dari keseluruhan metrik evaluasi dengan skor F1 0,885; AUROC 0,954; *accuracy* 0,914; *precision* 0,888; dan *recall* 0,881.

Dari hasil ini model LASSO-XGBoost menggunakan PSSM memberikan performa yang paling baik dalam memprediksi DTI (*accuracy* tinggi), memiliki prediksi kelas positif yang baik (*recall* tinggi), memiliki prediksi positif yang baik (*precision* tinggi), memiliki performa paling baik pada kelas minoritas (F1 tinggi) dan mampu membedakan kelas positif dan tidak diketahui dengan baik (AUROC tinggi).

Jika dilihat pada hasil evaluasi, skor pada *dataset enzyme* paling baik, lalu IC, GPCR dan NR. Skor hasil percobaan memiliki nilai yang cukup berbeda berdasarkan pada *dataset* yang digunakan. Dapat disimpulkan bahwa perbedaan jumlah data pada setiap jenis *dataset* berpengaruh pada hasil evaluasi.

Hasil *Precision Recall Curve* (PRC) menunjukkan *trade-off* antara *precision* dan *recall*. PRC pada *dataset enzyme* terlihat fitur PSSM memiliki AUCPR paling luas dengan AUCPR 0,9432. Pada *dataset* NR setiap fitur menunjukkan penurunan AUCPR dan fitur CTD memberikan hasil paling baik dengan AUCPR 0,7369. PRC *dataset* GPCR menunjukkan fitur DPC memberikan hasil paling baik dengan AUCPR 0,8113 diikuti PSSM dengan AUCPR 0,8075. PRC pada *dataset* IC menunjukkan fitur PSSM memberikan skor paling baik dengan nilai AUCPR 0,9028.



Gambar 5. PRC *enzyme* dan NR (atas*), PRC GPCR dan IC (bawah)*

Berdasarkan hasil pada Gambar 5 dapat disimpulkan bahwa model LASSO-XGBoost dengan fitur PSSM memiliki kurva yang lebih dekat ke sudut kanan atas yang menandakan performa model menggunakan fitur tersebut cukup baik.

Berdasarkan hasil tersebut, maka disimpulkan bahwa kombinasi fitur Pubchem-PSSM paling baik digunakan pada model yang dibangun. Sehingga kombinasi fitur tersebut digunakan dalam pemodelan dengan *dataset* DTI pada protein terkait kanker.

3.4. Pemodelan Dengan Interaksi Senyawa Obat dan Protein Pada Kanker

Pemodelan dilakukan menggunakan *dataset* DTI pada protein kanker. Pemodelan dilakukan menggunakan kombinasi fitur Pubchem-PSSM berdasarkan hasil sebelumnya. *Hyperparameter* dilakukan dengan 5-*fold* CV. *Hyperparameter* yang optimal dapat dilihat pada Tabel 9. Berdasarkan model yang optimal dihasilkan nilai evaluasi, yaitu F1 0,861; AUROC 0,927; *accuracy* 0,897; *recall* 0,857; *precision* 0,866. Hasil *hyperparameter* selengkapnya dapat dilihat pada tambahan *spreadsheet* 2.

Tabel 9. *Hyperparameter* optimal pada LASSO-XGBoost

Hyperparameter	Nilai
n_estimators	1000
max_depth	8
learning_rate	0,1
min_child_weight	1
colsample_bytree	0,5
subsample	0,5

Senyawa herbal selanjutnya di masukkan ke dalam model prediksi. Prediksi interaksi senyawa herbal dan protein terkait kanker dilakukan dengan metode LASSO-XGBoost menggunakan fitur Pubchem-PSSM.

Hasil prediksi pada model menghasilkan 396 senyawa herbal yang berinteraksi dengan 62 protein kanker dan membentuk 5152 interaksi. Terdapat sebanyak 379 senyawa herbal yang memiliki interaksi dengan beberapa protein berdasarkan pada pemodelan.

Penelitian ini tidak melakukan uji molekuler pada hasil prediksi terkait interaksi senyawa herbal dan protein. Untuk mengetahui hasil prediksi yang relevan maka dilakukan studi literatur untuk melihat hasil pengujian terkait dengan interaksi yang telah diprediksi. Tabel 10 merupakan beberapa hasil prediksi interaksi senyawa dan protein beserta peluangnya. Tabel 11 menunjukkan hasil studi literaturnya. Hasil prediksi selengkapnya dapat dilihat pada tambahan *Spreadsheet* 3.

Tabel 10. Hasil prediksi interaksi senyawa herbal-protein

Senyawa	Protein	Peluang
Andrographolide	IL2, AR	0,841; 0,882
Ascorbic Acid	IL2	0,925
Betulinic Acid	BCL2	0,740
Cucurbitacin E	IL2	0,757
Cucurbitacin R	IL2	0,961
Tetradecanol	IL2	0,925
Thymoquinone	IL2	0,525

Senyawa	Protein	Peluang
Ursolic Acid	IL2, EGFR, BCL2, MAPK1, SMO	0,975; 0,741; 0,780; 0,643; 0,557
Oleanolic Acid	EGFR, P53, BCL2	0,643; 0,615; 0,780
Borneol	BCL2	0,582
Cucurbitacin I	BCL2	0,634

Tabel 11. Studi literatur interaksi hasil prediksi

Protein	Senyawa Herbal
IL2	*Andrographolide* (Rajagopal et al., 2003), *Ascorbic Acid* (Schwager dan Schulze, 1998), *Cucurbitacin E* (Wang et al., 2015), *Cucurbitacin R* (Escandell et al., 2010), *Tetradecanol* (Park et al., 2017), *Thymoquinone* (Miliani et al., 2018), *Ursolic Acid* (Kanjoormana dan kuttan, 2010)
AR	*Andrographolide* (Liu et al., 2011)
EGFR	*Oleanolic Acid* (Tang et al., 2022), *Ursolic Acid* (Shan et al., 2009)
P53	*Oleanolic Acid* (Pratheeshkumar dan Kuttan, 2011)
BCL2	*Borneol* (Nasr et al., 2020), *Cucurbitacin I* (Yuan et al., 2014), *Oleanolic Acid* (Pratheeshkumar dan Kuttan, 2011), *Ursolic Acid* (Kassi et al., 2009), *Betulinic Acid* (Bhatia et al., 2015)
MAPK1	*Oleanolic Acid* (Tang et al., 2022), *Ursolic Acid* (Achiwa et al., 2013)
SMO	*Ursolic Acid* (Lyu et al., 2022)

Berdasarkan hasil studi literatur yang telah dilakukan ditemukan beberapa interaksi yang memberikan efek peningkatan dan penurunan aktivitas pada protein target. Beberapa diantaranya, yaitu interaksi senyawa *andrographolide* dengan IL2 dan AR, interaksi pada IL2 mengaktivasi imunostimulator yang dibuktikan dengan peningkatan produksi IL2 yang dapat mengaktivasi sel T (Rajagopal et al., 2003). *Andrographolide* dapat menghambat AR menyebabkan induksi apoptosis dan menekan pertumbuhan kanker prostat (Liu et al., 2011). Senyawa *andrographolide* dapat ditemukan pada tanaman *Andrographis paniculata* (Sambiloto) dan *Nicotiana tabacum* (Tembakau).

Senyawa *oleanolic acid* berinteraksi dengan protein P53 dan BCL2, ditemukan *upregulation* penekan tumor pada P53 dan *downregulation* BCL2 pada sel melanoma B16F-10 (Pratheeshkumar dan Kuttan, 2011). Senyawa *oleanolic acid* dapat ditemukan pada tanaman *Morinda citrifolia L.* (Mengkudu), *Plantago major* (Daun Sendok), *Allium cepa L.* (Bawang Merah), *Johar, Prunus avium* (Ceri Manis), *Foeniculum vulgare* (Adas), *Panax ginseng* (Ginseng Asia).

Senyawa *ursolic acid* dapat menghambat proliferasi pada sel HT-29 dengan menekan EGFR (Shan et al., 2009) dan menghambat ekspresi SMO (Lyu et al., 2022). Senyawa *ursolic acid* dapat ditemukan pada tanaman *Morinda citrifolia L.* (Mengkudu), *Olea europaea* (Zaitun), *Malus spp.* (Apel), *Pyrus spp.* (Pir).

Hasil prediksi dari model yang sudah dibangun masih perlu dilakukan pengujian lebih lanjut untuk memastikan bahwa interaski senyawa herbal dan protein memiliki potensi untuk digunakan.

4. KESIMPULAN DAN SARAN

Pada penelitian ini, prediksi DTI menggunakan LASSO-XGBoost memberikan performa yang cukup baik dalam memprediksi DTI. Percobaan pada *golden dataset* menunjukkan bahwa kombinasi fitur Pubchem-PSSM memberikan hasil paling baik pada model.

Hasilnya, pada *dataset* interaksi senyawa obat dan protein kanker dihasilkan nilai evaluasi berdasarkan model yang optimal, yaitu F1 0,861; AUROC 0,927; *recall* 0,857; *precision* 0,866; *accuracy* 0,897. Hasil prediksi interaksi senyawa herbal dan protein terkait kanker menggunakan model yang sudah dibangun menghasilkan 5152 interaksi dari 396 senyawa herbal dan 62 protein terkait kanker. Beberapa senyawa herbal yang berhasil diprediksi seperti *andrographolide*, *ursolic acid*, dan *oleanolic acid* memiliki interaksi pada protein terkait kanker berdasarkan model LASSO-XGBoost.

Pengembangan lebih lanjut dapat dilakukan dengan mencoba beberapa kombinasi fitur maupun kombinasi metode yang berbeda agar dapat dilakukan perbandingan dalam mencari model yang optimal.

BAHAN TAMBAHAN

Bahan Tambahan: Tersedia *online* pada tautan berikut https://github.com/TropBRC-BioinfoLab/lasso-xgb-dti, *Folder* 1: Hasil percobaan pada *golden dataset* menggunakan *grid search*, *Spreadsheet* 2: Hasil *hyperparameter* senyawa obat dan protein pada LASSO-XGBoost dengan PSSM, *Spreadsheet* 3: Hasil prediksi senyawa herbal dan protein LASSO-XGBoost dengan PSSM.

UCAPAN TERIMA KASIH

Terima kasih penulis ucapkan atas kesempatan yang diberikan dari Kementrian Riset, Teknologi, dan Pendidikan melalui program hibah Penelitian Tesis Magister (PTM) dengan nomor 3871/IT3.L1/PT.01.03/P/B/2022 sehingga hasil penelitian ini dapat dipublikasikan. Semoga tulisan ini dapat bermanfaat untuk penelitian lainnya

DAFTAR PUSTAKA

ACHIWA, Y., HASEGAWA, K. dan UDAGAWA, Y., 2013. Effect of ursolic acid on MAPK in cyclin D1 signaling and RING-type E3 ligase (SCF E3s) in two endometrial cancer cell lines. *Nutrition and cancer*, 65(7), pp.1026-1033.

AFENDI, F.M., OKADA, T., YAMAZAKI, M., HIRAI-MORITA, A., NAKAMURA, Y., NAKAMURA, K., IKEDA, S.,

TAKAHASHI, H., ALTAF-UL-AMIN, M., DARUSMAN, L.K. dan SAITO, K., 2012. KNApSAcK family databases: integrated metabolite–plant species databases for multifaceted plant research. *Plant and Cell Physiology*, 53(2), pp.e1-e1.

AVRAM, S., BOLOGA, C.G., HOLMES, J., BOCCI, G., WILSON, T.B., NGUYEN, D.T., CURPAN, R., HALIP, L., BORA, A., YANG, J.J. dan KNOCKEL, J., 2021. DrugCentral 2021 supports drug discovery and repositioning. *Nucleic acids research*, 49(D1), pp.D1160-D1169.

BATEMAN, A., MARTIN, M.J., ORCHARD, S., MAGRANE, M., AHMAD, S., ALPI, E., BOWLER-BARNETT, E.H., BRITTO, R., BYE-A-JEE, H., CUKURA, A. dan DENNY, P., 2022. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*.

BHATIA, A., KAUR, G. dan SEKHON, H.K., 2015. Anticancerous efficacy of betulinic acid: An immunomodulatory phytochemical. *J. PharmaSciTech,* 4, pp.39-46.

CHEN, T. dan GUESTRIN, C., 2016. Xgboost: a scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; 2016: 785-794. ACM, New York, NY.

DATTA, S., DEV, V.A. dan EDEN, M.R., 2017. Developing QSPR for predicting DNA drug binding affinity of 9-Anilinoacridine derivatives using correlation-based adaptive LASSO algorithm. *Computer Aided Chemical Engineering,* Vol. 40, pp. 2767-2772. Elsevier

ESCANDELL, J.M., RECIO, M.C., GINER, R.M., MANEZ, S., CERDA-NICOLAS, M., MERFORT, I. dan RÍOS, J.L., 2010. Inhibition of delayed-type hypersensitivity by cucurbitacin R through the curbing of lymphocyte proliferation and cytokine expression by means of nuclear factor AT translocation to the nucleus. *Journal of Pharmacology and Experimental Therapeutics*, 332(2), pp.352-363.

GÜRLER, S.B., KIRAZ, Y. dan BARAN, Y., 2020. Flavonoids in cancer therapy: Current and future trends. *Biodiversity and Biomedicine*, pp.403-440.

HERNANI, P., 2011. Pengembangan Biofarmaka Sebagai Obat Herbal Untuk Kesehatan. *Buletin Teknologi Pascapanen Pertanian*. 7(1), pp.20-9.

HUSSAIN, Y., ISLAM, L., KHAN, H., FILOSA, R., ASCHNER, M. dan JAVED, S., 2021. Curcumin–cisplatin chemotherapy: A novel strategy in promoting chemotherapy efficacy and reducing side effects. *Phytotherapy Research*, 35(12), pp.6514-6529.

KANJOORMANA, M. dan KUTTAN, G., 2010. Antiangiogenic activity of ursolic acid. *Integrative Cancer Therapies*, 9(2), pp.224-235.

KASSI, E., SOURLINGAS, T.G., SPILIOTAKI, M., PAPOUTSI, Z., PRATSINIS, H., ALIGIANNIS, N. dan MOUTSATSOU, P., 2009. Ursolic acid triggers apoptosis and Bcl-2 downregulation in MCF-7 breast cancer cells. *Cancer investigation*, 27(7), pp.723-733.

KIM, S., CHEN, J., CHENG, T., GINDULYTE, A., HE, J., HE, S., LI, Q., SHOEMAKER, B.A., THIESSEN, P.A., YU, B. dan ZASLAVSKY, L., 2021. PubChem in 2021: new data content and improved web interfaces. *Nucleic acids research*, 49(D1), pp.D1388-D1395.

LIN, K.B., WENG, W., LAI, R.K. dan LU, P., 2014. Imbalance data classification algorithm based on SVM and clustering function. *In 2014 9th International Conference on Computer Science & Education*, pp. 544-548. IEEE.

LIU, C., NADIMINTY, N., TUMMALA, R., CHUN, J.Y., LOU, W., ZHU, Y., SUN, M., EVANS, C.P., ZHOU, Q. dan GAO, A.C., 2011. Andrographolide targets androgen receptor pathway in castration-resistant prostate cancer. *Genes & cancer*, 2(2), pp.151-159.

LYU, X., ZHANG, X., SUN, L., WANG, J. dan WANG, D., 2022. Inhibitory Effect of Ursolic Acid on Proliferation and Migration of Renal Carcinoma Cells and Its Mechanism. *Computational Intelligence and Neuroscience*, 2022.

MAHMUD, S.H., CHEN, W., JAHAN, H., DAI, B., DIN, S.U. dan DZISOO, A.M., 2020. DeepACTION: A deep learning-based method for predicting novel drug-target interactions. *Analytical biochemistry*, 610, p.113978.

MATHUR, G., NAIN, S. dan SHARMA, P.K., 2015. Cancer: an overview. *Acad J Cancer Res*. 8(1), pp.01-09.

MILIANI, M., NOUAR, M., PARIS, O., LEFRANC, G., MENNECHET, F. dan ARIBI, M., 2018. Thymoquinone potently enhances the activities of classically activated macrophages pulsed with necrotic jurkat cell lysates and the production of antitumor Th1-/M1-related cytokines. *Journal of Interferon & Cytokine Research*, 38(12), pp.539-551.

MOHAMMADI, A., ZAHIRI, J., MOHAMMADI, S., KHODARAHMI, M. dan ARAB, S.S., 2022. PSSMCOOL: a comprehensive R package for generating evolutionary-based descriptors of protein sequences from PSSM profiles. *Biology Methods and Protocols*, 7(1), p.bpac008.

NASR, F.A., NOMAN, O.M., ALQAHTANI, A.S., QAMAR, W., AHAMAD, S.R., AL-MISHARI, A.A., ALYHYA, N. dan FAROOQ, M., 2020. Phytochemical constituents and anticancer activities of Tarchonanthus camphoratus essential oils grown in Saudi Arabia. S*audi Pharmaceutical Journal,* 28(11), pp.1474-1480.

PARK, J.U., KANG, B.Y., LEE, H.J., KIM, S., BAE, D., PARK, J.H. dan KIM, Y.R., 2017. Tetradecanol reduces EL-4 T cell growth by the down regulation of NF-κB mediated IL-2 secretion. *European Journal of Pharmacology*, 799, pp.135-142.

PATIL, A.R. dan KIM, S., 2020. Combination of ensembles of regularized regression models with resampling-based lasso feature selection in high dimensional data. *Mathematics*, 8(1), p.110.

PENG, Y., WANG, J., WU, Z., ZHENG, L., WANG, B., LIU, G., LI, W. dan TANG, Y., 2022. MPSM-DTI: prediction of drug–target interaction via machine learning based on the chemical structure and protein sequence. *Digital Discovery*, 1(2), pp.115-126.

PRATHEESHKUMAR, P. dan KUTTAN, G., 2011. Oleanolic acid induces apoptosis by modulating p53, Bax, Bcl-2 and caspase-3 gene expression and regulates the activation of transcription factors and cytokine profile in B16F. Journal of Environmental Pathology, *Toxicology and Oncology*, 30(1).

RAJAGOPAL, S., KUMAR, R.A., DEEVI, D.S., SATYANARAYANA, C. dan RAJAGOPALAN, R., 2003. Andrographolide, a potential cancer therapeutic agent isolated from Andrographis paniculata. *Journal of Experimental therapeutics and Oncology*, 3(3), pp.147-158.

SCHWAGER, J. dan SCHULZE, J., 1998. Modulation of interleukin production by ascorbic acid. *Veterinary immunology and immunopathology*, 64(1), pp.45-57.

SHAN, J.Z., XUAN, Y.Y., ZHENG, S., DONG, Q. dan ZHANG, S.Z., 2009. Ursolic acid inhibits proliferation and induces apoptosis of HT-29 colon cancer cells by inhibiting the EGFR/MAPK pathway. *Journal of Zhejiang University SCIENCE B*, 10(9), pp.668-674.

SHI, H., LIU, S., CHEN, J., LI, X., MA, Q. dan YU, B., 2019. Predicting drug-target interactions using Lasso with random forest based on evolutionary information and chemical structure. *Genomics*, 111(6), pp.1839-1852.

SYAHDI, R.R., IQBAL, J.T., MUNIM, A. dan YANUAR, A., 2019. HerbalDB 2.0: Optimization of construction of three-dimensional chemical compound structures to update Indonesian medicinal plant database. *Pharmacognosy Journal*, 11(6).

TANG, Z.Y., LI, Y., TANG, Y.T., MA, X.D. dan TANG, Z.Y., 2022. Anticancer activity of oleanolic acid and its derivatives: Recent advances in evidence, target profiling and mechanisms of action. *Biomedicine & Pharmacotherapy*, 145, p.112397.

THAFAR, M.A., OLAYAN, R.S., ALBARADEI, S., BAJIC, V.B., GOJOBORI, T., ESSACK, M. dan GAO, X., 2021. DTi2Vec: Drug–target interaction prediction using network embedding and ensemble learning. *Journal of cheminformatics*, 13(1), pp.1-18.

TIBSHIRANI, R., 2011. Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(3), pp.273-282.

Tim CancerHelps, 2019. Stop Kanker. AgroMedia.

WANG, L., LI, C., LIN, Q., ZHANG, X., PAN, H., XU, L., SHI, Z., OUYANG, D. dan HE, X., 2015. Cucurbitacin E suppresses cytokine expression in human Jurkat T cells through down-regulating the NF-κB signaling. *Acta biochimica et biophysica Sinica*, 47(6), pp.459-465.

WANG, L., YOU, Z.H., CHEN, X., YAN, X., LIU, G., ZHANG, W., 2018. Rfdt: A rotation forest-based predictor for predicting drug-target interactions using drug structure and protein sequence information. *Current Protein and Peptide Science*, 1;19(5), pp.445-54.

WIJAYA, S.H., AFENDI, F.M., BATUBARA, I., HUANG, M., ONO, N., KANAYA, S. dan ALTAF-UL-AMIN, M., 2021. Identification of Targeted Proteins by Jamu Formulas for Different Efficacies Using Machine Learning Approach. *Life*, 11(8), p.866.

WISHART, D.S., KNOX, C., GUO, A.C., SHRIVASTAVA, S., HASSANALI, M., STOTHARD, P., CHANG, Z. dan WOOLSEY, J., 2006. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic acids research*, 34(suppl_1), pp.D668-D672.

XU, L., RU, X. dan SONG, R., 2021. Application of machine learning for drug–target interaction prediction. *Frontiers in Genetics*, p.1077.

YAMANISHI, Y., ARAKI, M., GUTTERIDGE, A., HONDA, W. dan KANEHISA, M., 2008. Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics,* 24(13), pp.i232-i240.

YOU, J., MCLEOD, R.D. dan HU, P., 2019. Predicting drug-target interaction network using deep learning model. *Computational biology and chemistry,* 80, pp.90-101.

YUAN, G., YAN, S.F., XUE, H., ZHANG, P., SUN, J.T. dan LI, G., 2014. Cucurbitacin I induces protective autophagy in glioblastoma in vitro and in vivo. *Journal of Biological Chemistry*, 289(15), pp.10607-10619.

Halaman ini sengaja dikosongkan