

SKRINING VIRTUAL BERBASIS LIGAN SENYAWA PRODUK NATURAL TERHADAP SARS-COV-2 MAIN PROTEASE (M^{pro}) MENGGUNAKAN PENDEKATAN SIMILARITAS SUBSTRUKTUR DAN PEMBELAJARAN MESIN

GP. Wahyunanda Crista Yuda^{1a}

¹Fakultas Farmasi Universitas Gadjah Mada

^aCorresponding author's email: gpwahunanda@mail.ugm.ac.id

ABSTRAK

Pendahuluan : COVID-19 merupakan penyakit infeksi yang disebabkan oleh SARS-CoV-2 dan merupakan kerabat dekat SARS-CoV. Dengan afinitasnya 10-20 kali terhadap reseptor ACE2 melebihi SARS-CoV, COVID-19 mencapai prevalensi hingga 515,764,048 kasus. Beberapa penelitian telah dilakukan untuk mencari terapi COVID-19, mulai penemuan vaksin hingga repurposing obat sehingga menemukan SARS-CoV-2 Main Protease (M^{pro}) adalah target terapi yang potensial dikarenakan perannya yang esensial dalam replikasi dan sekuensnya yang konservatif. Beberapa studi terapi antiviral tertarget M^{pro} telah dilakukan, namun beberapa tidak efektif dan berpotensi membahayakan. Alternatif yang dapat diajukan adalah dengan memanfaatkan senyawa produk natural (PN).

Metode : melakukan skrining sebanyak 407.270 senyawa PN dengan menggunakan pendekatan berbasis ligan: similaritas substruktur dan pembelajaran mesin. Sebanyak 16.473 ligand yang berinteraksi dengan M^{pro} diperoleh dan diklasifikasi menjadi aktif dan inaktif berdasarkan *cut-off* dan senyawa PN disaring dengan menggunakan parameter *Lipinski's Rule of Five*. Kedua data ini distandardisasi menggunakan protocol Galadeta dkk. Skrining dilakukan pada KNIME dengan menggunakan *syntax* RDkit.

Hasil : penelitian menunjukkan sebanyak 25 *scaffold* berhasil menghasilkan 59.406 senyawa PN poten dengan 7-bromo-4-(chloromethyl)-2H-chromen-2-one dan isomernya memiliki skor similaritas 0,93, mengalahkan kontrol positif GC-376 (0,85), nirmatrelvir (0,64), ritonavir (0,57), dan lopinavir (0,53).

Kesimpulan : Potensi inhibisi M^{pro} oleh 7-bromo-4-(chloromethyl)-2H-chromen-2-one perlu dikaji lebih lanjut untuk pengembangan senyawa terapi COVID-19 berbasis bahan alam.

Kata Kunci: COVID-19, produk natural, perancah, similaritas sidik jari, KNIME

ABSTRACT

Introduction : COVID-19 is an infectious disease caused by SARS-CoV-2, close relative of SARS-CoV. With 10-20 times affinity for the ACE2 receptor over SARS-CoV, COVID-19 has reached 515,764,048 cases. Several studies have searched for COVID-19 therapy, from vaccine discovery to drug repurposing, finding SARS-CoV-2 Main Protease (M^{pro}) a potential therapeutic target due to its essential role in replication with conservative sequence. Several studies of M^{pro} targeted therapy have been conducted, but some are ineffective and potentially harmful. An alternative proposed is to use natural product (NP) compounds.

Method : This study aims to screen 407,270 NP compounds using a ligand-based approach: substructure similarity and machine learning. 16,473 ligands that interact with M^{pro} were obtained and classified into active and inactive based on the *cut-off* and NP compounds were filtered using *Lipinski's Rule of Five* parameters. Both of these data were



standardized using the Gadaleta protocol. Screening is conducted on KNIME using the RDKit syntax.

Results : showed that 25 scaffolds produced 59,406 potent NP compounds with 7-bromo-4-(chloromethyl)-2H-chromen-2-one had a similarity score of 0.93, beating the positive control GC-376 (0.85), nirmatrelvir (0.64), ritonavir (0.57), and lopinavir (0.53).

Conclusion : The potential for inhibition of M^{pro} by 7-bromo-4-(chloromethyl)-2H-chromen-2-one needs to be studied further to develop NP-based COVID-19 therapeutic.

Keywords: COVID-19, natural products, scaffold, fingerprint similarity, KNIME

PENDAHULUAN

COVID-19 merupakan penyakit yang disebabkan oleh infeksi virus SARS-CoV-2 dan mencapai prevalensi 515,764,048 kasus dengan mortalitas 6,246,241 jiwa per 5 Mei 2022 (1). SARS-CoV-2 merupakan kerabat dekat dari SARS-CoV dan MERS-CoV yang sempat menyebabkan epidemi pada tahun 2002 dan 2013 lalu (2). Namun, salah satu keunikan yang menyebabkan SARS-CoV-2 menjadi pandemi global adalah afinitas proteinnya terhadap reseptor ACE2, yang mana 10-20 kali lebih tinggi dibandingkan SARS-CoV (3), sehingga tingkat replikasi dan penyebarannya dapat tinggi melebihi kedua pendahulunya. Berbagai pihak telah mengembangkan vaksin untuk mengatasi SARS-CoV-2 (4–7). Selain itu juga, beberapa studi juga menyarankan untuk menggunakan obat yang ada sebagai terapi alternatif untuk COVID-19 menggunakan pendekatan drug repurposing (8–11).

Dari semua studi yang menarget beberapa protein autentik SARS-CoV-2, M^{pro} atau 3CL^{pro} merupakan yang paling potensial dikarenakan fungsi fisiologisnya yang esensial dalam proses replikasi virus dan sekuensnya yang terbilang masih konservatif

(99,02% identik dengan SARS-CoV) (12). Beberapa inhibitor protein ini telah dikembangkan, seperti GC-373 dan 376 (13–17). Bahkan, banyak studi yang melakukan skrining virtual drug repurposing dengan menggunakan metode penambatan molekular sehingga menghasilkan hidroksiklorokuin, remdesivir, lopinavir, dan beberapa obat lainnya sebagai alternatif terapi (18–21). Namun, beberapa obat ini ternyata tidak menunjukkan efikasi yang memadai, bahkan dapat memberikan efek samping yang tidak diketahui dikarenakan regimen dosis yang digunakan juga berbeda (22). Oleh karena itu, alternatif pengobatan tertarget M^{pro} yang potensial dan aman perlu dilakukan. Salah satunya adalah menggunakan produk natural (PN) dari tumbuhan obat.

Indonesia memiliki lebih dari 30.000 spesies tumbuhan obat (80%) dari total tumbuhan obat di dunia dan sekitar sekitar 9.600 tumbuhan tersebut memiliki berbagai aktivitas biologis (23). Beberapa senyawa fitokimia yang dihasilkan oleh tumbuhan ini tentunya memiliki aktivitas antivirus. Sebuah *literature review* merangkum kumpulan tanaman obat dan senyawa aktifnya beserta kemungkinan potensi targetnya dalam menyerang SARS-CoV-2 (24).



Namun perlu diingat, satu tumbuhan obat dapat menghasilkan berbagai senyawa bioaktif sebagai metabolit sekundernya sehingga pendekatan penambatan molekular sangat tidak praktis dilakukan untuk skrining ratusan ribu fitokimia. Maka dari itu, pendekatan berbasis ligan menggunakan gabungan metode similaritas substruktur dan pembelajaran mesin dapat menjadi alternatif (25) untuk skrining cepat dengan hasil yang lebih akurat dibandingkan penambatan molekular (26,27). Hal ini memungkinkan karena didukung oleh banyaknya ligan yang sudah tercatat berinteraksi dengan M^{pro} pada database keminformatik sehingga pendekatan berbasis ligan dapat dilakukan.

Tahap awal dari skrining virtual berbasis ligan dilakukan mulai dari persiapan ligan dengan melakukan standarisasi menggunakan protokol Gadaleta dkk. (28) dan membagi ligan menjadi aktif dan inaktif berdasarkan *cut-off*. Ligan yang aktif nantinya akan dijadikan input untuk *scaffold mining* dan *hierarchical clustering* sebagai preparasi pembuatan *Maximum Common Substructure* (MCS) dalam bentuk *SMiles ARbitrary Target Specification* (SMARTS) (29) yang berfungsi sebagai referensi skrining dengan menyamakan similaritas MCS dengan substruktur fitokimia. Sedangkan, ligan yang aktif dan inaktif dijadikan input untuk membentuk model pembelajaran mesin berbasis similaritas sidik jari. Algoritma sidik jari yang digunakan adalah *Molecular Access*

System (MACCS) dan ECFP4 atau sering dikenal dengan *Morgan* (30–32), lalu algoritma pembelajaran mesin yang digunakan adalah *classification supervised: Random Forest* (RF) (33), *Artificial Neural Network* (ANN) (34), dan *Support Vector Machine* (SVM) (33). Model yang terbaik nantinya digunakan untuk menentukan aktivitas fitokimia yang terkskrining dengan mengurutkan skor similaritasnya dengan ligan-ligan yang berinteraksi dengan M^{pro}, yang mana semakin tinggi skornya maka semakin poten senyawa fitokimia tersebut karena semakin mirip dengan kumpulan ligan yang aktif.

METODE

Semua Langkah dimulai dari tahap satu hingga terakhir mengadaptasi alur kerja pada penelitian Tuerkova dan Sydow (35,36). Alur kerja ini dilakukan pada aplikasi sumber terbuka KNIME menggunakan nodus dengan sedikit koding (37) dan nodus tersebut diadaptasi dari perintah (*syntax*) yang berasal dari komunitas RDKit (38)

Preparasi Ligan dan Senyawa Fitokimia

Sebanyak 16.473 ligan yang berinteraksi dengan M^{pro} didapatkan dari database PDB, ChEMBL, dan PubChem beserta format *Simplified Molecular-Input Line-Entry System* (SMILES)nya. Sebanyak 407.270 senyawa fitokimia juga didapatkan dari database Collection of Open Natural Products (COCONUT) beserta SMILESnya. Kedua data tersebut distandardisasi menggunakan



protocol Gadaleta dkk. dengan menghilangkan stereoisomer pada SMILES, menghilangkan ion dan garam, dan menyaring senyawa yang hanya memiliki atom H,C,N,O,F,Br,I,Cl,P,S. Untuk seleksi ligan, sebelum distandardisasi perlu dilakukan klasifikasi potensi senyawa aktif/inaktif dengan melabeli biner 0 untuk inaktif dan 1 untuk aktif, di mana senyawa aktif memiliki aktivitas $IC_{50}/K_i < 1 \mu M$ (39) dan %Inhibition $> 80\%$ (40). Khusus untuk preparasi ligan *scaffold mining*, interaksi protein-ligan yang valid dari PDB disaring menggunakan resolusi $< 2.05 \text{ \AA}$ (41). Untuk database fitokimia, sebelum distandardisasi perlu dilakukan penyaringan dengan menggunakan *Lipinski's Rule of Five* (42) dengan kriteria senyawa yang lolos adalah memenuhi ≥ 4 syarat dan dianggap aktif di dalam penelitian ini.

1. *Scaffold Mining, Hierarchical Clustering*, dan Pembuatan *Maximum Common Substructure*

Ligan aktif yang telah dipreparasi dan distandardisasi diteruskan ke proses pembuatan *scaffold* dengan algoritma *unsupervised* menggunakan nodus RDKit *Find Murcko Scaffold* untuk menghasilkan *Murcko framework* (43). Setelah kumpulan *scaffold* terbentuk, dilanjutkan dengan melakukan klustering metode *Average Linkage* (44) dengan *distance threshold* 0.5 yang diatur pada nodus *cluster assigner*. Terakhir, *scaffold* dengan nomor klusternya diteruskan ke proses pembuatan dan kalkulasi MCS

menggunakan nodus RDKit MCS sehingga menghasilkan *scaffold* yang teroptimasi ke dalam format SMARTS.

2. Skrining Substruktur Produk Natural Berdasarkan MCS

Scaffold teroptimasi ini kemudian dijadikan sebagai standar referensi untuk skrining substruktur seluruh PN yang telah disaring dan terstandardisasi dengan menggunakan nodus RDKit *Substructure Filter* dan RDKit *Molecular Highlighter*. Proses ini terus diulang setiap putaran dengan satu putaran mewakili *scaffold* yang menyocokkan diri dengan struktur PN yang ada sehingga dihasilkan kumpulan senyawa yang telah diskining yang mungkin satu senyawa memiliki lebih dari satu *scaffold* berbeda.

3. Penentuan Skor Similaritas dengan Metode Sidik Jari dan Pembelajaran Mesin

Ligan yang telah dipreparasi sebelumnya terbagi menjadi aktif dan inaktif berdasarkan parameter *cut-off* yang telah ditentukan. Semua ligan aktif dan inaktif ini kemudian diteruskan ke proses pembuatan model similaritas sidik jari dan pembelajaran mesin. Data ligan berupa SMILES dan label biner 0 dan 1 masuk ke dalam pemrosesan pemecahan sidik jari menggunakan nodus RDKit *Fingerprint* dengan format sidik jari yaitu *bitstring*. Dua metode sidik jari yang digunakan adalah MACCS dan *Morgan* (31). Setelah SMILES terbagi menjadi beberapa *bitstring* dan membentuk sebuah *array*, *array* ini kemudian diteruskan ke dalam tiga



algoritma pembelajaran mesin terawasi yang sering digunakan untuk klasifikasi, *Random Forest*, *Artificial Neural Network*, dan *Support Vector Machine*.

Ketiga algoritma ini dikonfigurasi dengan *k-cross value* sebesar 10, yang berarti 10% data untuk uji percobaan dan 90% data untuk pelatihan mesin. Metode sampling yang digunakan masing-masing algoritma adalah *random stratified sampling*, mengingat rasio jumlah ligan aktif dengan inaktif adalah 1:40 agar proporsi ligan untuk pelatihan dan tes seimbang. Selain itu, nodus *Synthetic Minority Over-sampling*

Technique (SMOTE) juga membantu untuk membuat dataset ligan buatan untuk menyeimbangkan jumlah ligan yang dijadikan pelatihan. Selanjutnya, semua konfigurasi pembelajaran mesin dari ketiga algoritma dibuat *default* dari aplikasi. Setelah pembuatan model pembelajaran mesin selesai, model dari ketiga algoritma untuk masing-masing dari metode MACCS dan *Morgan* dievaluasi menggunakan analisis statistik *confusion matrix* dan *ROC Curve* (45). Model dengan hasil statistik terbaik akan dijadikan prediktor untuk menentukan skor similaritas PN yang telah terskrining sebelumnya.

HASIL DAN PEMBAHASAN

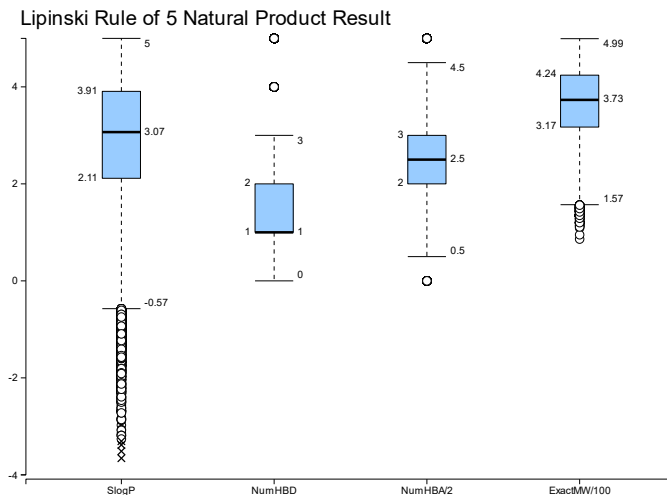
1. Preparasi Ligan dan Senyawa Produk Natural

Tabel 1. Hasil Pengambilan Data Ligan dan Senyawa Produk Natural.

Proses dilakukan dengan mengajukan permintaan daring ke masing-masing database dan dilanjut dengan pelabelan dan Standardisasi

Data		Jumlah Ligan M ^{pro}	Jumlah Senyawa Produk Natural
Hasil Pengambilan Database	PDB	317	
	ChEMBL	8.819	
	PubChem	7.337	
	COCONUT		407.270
Belum Terstandardisasi	Ligan/PN aktif	5.259	301.277
	Ligan/PN inaktif	11.214	
Telah Terstandardisasi	Ligan/PN aktif	269	292.384
	Ligan/PN inaktif	3.873	





Gambar 1. Distribusi Parameter *Lipinski's Rule of Five* dari Senyawa Produk Natural yang Terskrining.

Terlihat rerata logP yang diperoleh yaitu 3,07. rerata jumlah ikatan hidrogen donor dan akseptor yaitu 1 dan 5. Terakhir rerata berat molekul senyawa adalah 373 Da

Hasil preparasi ligan yang berinteraksi atau memiliki bioaktivitas dengan M^{Pro} di Tabel 1 menghasilkan 269 ligan aktif dari 16.473 yang dapat dijadikan sebagai bahan untuk membuat *scaffold*. Selain itu, total 3.969 ligan aktif dan inaktif yang hanya berasal dari data bioaktivitas digunakan sebagai bahan untuk membuat model pembelajaran mesin dan similaritas sidik jari. Hasil preparasi senyawa PN menghasilkan 292.384 senyawa yang telah lolos penyaringan *Lipinski's Rules of Five* dan standarisasi dari 407.270 senyawa yang ada di database COCONUT. Berdasarkan Gambar 1, rerata senyawa ini memiliki 0-5 ikatan hidrogen donor (<5), 0-10 ikatan hidrogen akseptor (<10), berat molekul 86,08- 499,31 Da (<500), dan LogP - 3.657-5 (<5) sehingga memenuhi kriteria.

Pembagian label berdasarkan *cut-off* ini penting mengingat acuan skrining virtual berbasis ligan terletak pada pengolahan ligannya. Alasan penulis menentukan *cut-off* resolusi <2.05Å dan bioaktivitas IC₅₀/K_i <1µM dan >80% Inhibisi adalah untuk memastikan bahwa senyawa PN yang terskrining memiliki beberapa karakteristik yang mirip dengan ligan-ligan ini, baik dari potensi bioaktivitas dan posisi substruktur yang memiliki sifat farmakofor, walau pendekatan yang digunakan adalah similaritas sidik jari.

Preparasi lainnya adalah dengan melakukan standarisasi menggunakan protocol Gadaleta dkk. yang bertujuan untuk menyamaratakan ligan dan senyawa PN ke dalam bentuk 2D dan normal (tanpa ion, garam, dan atom selain H,C,N,O,F,Br,I,Cl,P,S). Penyamataan ini dilakukan agar memudahkan pembuatan *scaffold* dalam pemecahan substruktur sehingga benar-

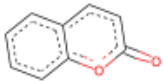
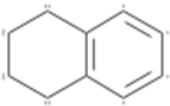
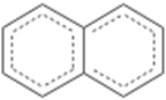

benar dihasilkan *Murcko framework* yang sedikit gangguan dan mengurangi bias untuk pembelajaran mesin ketika membentuk sebuah model. Preparasi lainnya, khusus untuk senyawa PN, penulis menggunakan *Lipinski's Rule of Five* sebagai kriteria seleksi senyawa yang memiliki sifat *drug-likeness* dan


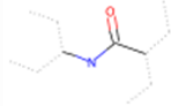

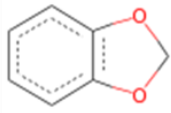
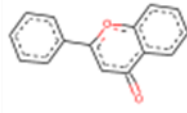
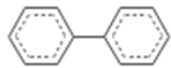
menentukan senyawa PN layak dikonsumsi secara oral (42). Parameter ini sangat kritis dalam pengembangan obat, terutama dalam menghilangkan masalah farmakokinetik dan bioavailabilitas obat ketika masuk ke fase absorpsi.

2. Pembentukan dan Klustering Hasil *Scaffold* yang Terbentuk dan Jumlah Senyawa yang Terskrining

Tabel 2. Daftar 10 *Scaffold* Teroptimasi yang Menghasilkan Senyawa PN Terbanyak beserta Klusternya.

Dapat dilihat bahwa urutan *scaffold* dimulai dari kumarin, lalu dilanjut dengan alil karbon 9, sikloheksan mengandung N, benzodioksol, flavonoid, naftalena, alil karbon 7, dan bifenil

Kelompok Kluster	<i>Scaffold</i> (MCS)	Format SMARTS (String)	Jumlah Senyawa PN yang Terskrining
1		<chem>[*]1:[*]:[*]:[*]2:[*]([*]:1):[*8] :[*]([*]:[*]:2)=[*8]</chem>	14.231
		<chem>[*]1-[*]-[*]-[*]2:[*](-[*]-1):[*]:[*]:[*]:[*]:2</chem>	3.222
		<chem>[*]1:[*]:[*]:[*]:[*]2:[*]:[*]:[*]6:[*]:[*]:1:2</chem>	3.098
2		<chem>[*7](-:[*]:[*]:[*]:[*]:[*]:[*]:[*])-.:[*]-.:[*]:[*]</chem>	10.340

Kelompok Kluster	Scaffold (MCS)	Format SMARTS (String)	Jumlah Senyawa PN yang Terskrining
		<chem>[#7]-,:[#6]:[#7]:[#6]:[#6]:[#6]:[#7]</chem>	3.089
		<chem>[#6]-,:[#6]-,:[#6](-[#6](=[#8])-[#7]-[#6](,:[#6]:,-[#6]):,-[#6]:,-[#6])-,:[#6]-,:[#6]</chem>	1.749
3		<chem>[#7]1-[#6]-[#6]-[#7]-[#6]-[#6]-1</chem>	8.368
4		<chem>[#6]1:[#6]:[#6]:[#6]2:[#6](:[#6]:1)-[#8]-[#6]-[#8]-2</chem>	6.751
5		<chem>[#6]1:[#6]:[#6]:[#6]2:[#6](=[#8]):[#6]:[#6](-[#6]3:[#6]:[#6]:[#6]:[#6]:[#6]:3):[#8]:[#6]:2:[#6]:1</chem>	3.248
6		<chem>[#6]1:[#6]:[#6]:[#6](:[#6]:[#6]:1)-[#6]1:[#6]:[#6]:[#6]:[#6]:1</chem>	2.103

Hasil pembuatan *scaffold*, klustering, dan optimasi MCS menghasilkan 25 *scaffold* yang berhasil menskrining senyawa PN yang dapat dilihat pada Tabel 2. *Scaffold* ini terbagi menjadi empat kelompok: (1) naftalena dan kumarin, (2) grup alil, (3)

sikloheksana mengandung N, (4) benzodioksol dan turunannya, (5) flavonoid, dan (6) bifenil. Total senyawa yang terskrining pun juga semakin berkurang, 59.406 dari 292.384 (20,3%) yang merupakan hasil dari pencocokan 25 *scaffold* di atas. *Scaffold* dengan

jumlah senyawa PN terbanyak diperoleh oleh *scaffold* kumarin dengan total senyawa 14.231, lalu disusul oleh alil karbon 9 (10.340), sikloheksana mengandung N (8.368), dan seterusnya. Dapat disimpulkan bahwa senyawa PN yang memiliki substruktur yang kita kenal seperti kumarin, flavonoid, dan benzodioksol berpotensi besar untuk menjadi terapi inhibitor M^{pro}.

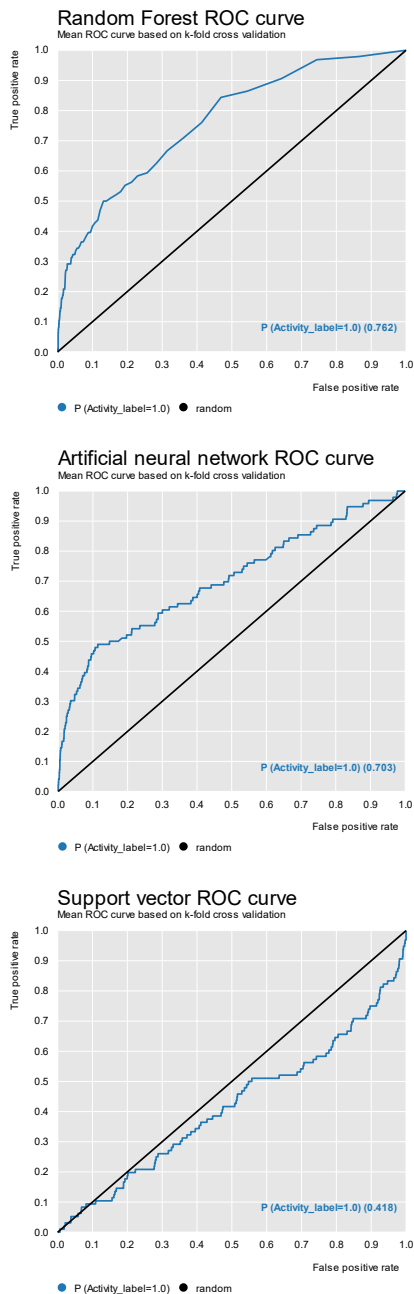
Pembuatan Murcko framework, pengelompokan kluster, dan optimasi *scaffold* menggunakan pendekatan MCS secara berurutan masuk ke dalam salah satu metode skrining berbasis ligan menggunakan deskriptor 2D. Maksud dari descriptor di sini adalah, skrining dilakukan berlandaskan properti dari ligan yang sudah ada yang kemudian digunakan untuk mencari senyawa yang memiliki kesamaan dengan properti ligan yang dijadikan variabel, berbeda dengan skrining berbasis struktur yang

prosesnya mengacu pada struktur 3D dan posisi atau koordinat ligan. Pendekatan menggunakan deskriptor pada penelitian ini ada dua: pembentukan substruktur umum dengan membentuk *scaffold* seperti pada hasil di atas dan pemecahan struktur ligan menjadi sidik jari yang mewakili substruktur-substruktur ligan aktif dan inaktif. Tentu hasil yang diberikan pendekatan pertama memiliki banyak kekurangan. Tiga kekurangan utamanya adalah hasil skrining yang tidak terorganisir, biasanya hasil skrining dengan mengandalkan similaritas *scaffold*, dan tidak adanya kuantifikasi senyawa mana yang poten dalam menghambat protein target. Oleh karena itu, disinilah peran metode similaritas sidik jari dan pembelajaran mesin membantu memprediksi semua senyawa PN.

3. Hasil Evaluasi Model Pendekatan Similaritas Sidik Jari dan Pembelajaran Mesin

Tabel 3. Rangkuman Parameter Pembelajaran Mesin dan Metode Similaritas Sidik Jari yang Akan Digunakan untuk Memprediksi Senyawa Produk Natural. Dari semua model algoritma dan sidik jari, dapat disimpulkan bahwa model *Random Forest* Sidik Jari MACCS digunakan sebagai model predictor dikarenakan hasilnya mengungguli model lainnya

Parameter Evaluasi Pembelajaran Mesin	MACCS			Morgan		
	RF	ANN	SVM	RF	ANN	SVM
Akurasi Keseluruhan (%)	95.94	93.32	N/A	89.29	95.99	N/A
Akurasi Seimbang (%)	0.880	0.924	N/A	0.940	0.98	N/A
<i>Cohen's Kappa</i> (κ)	0.223	0.161	0.000	0.107	0.129	0.000
<i>ROC AUC</i>	0.762	0.703	0.418	0.748	0.584	0.433

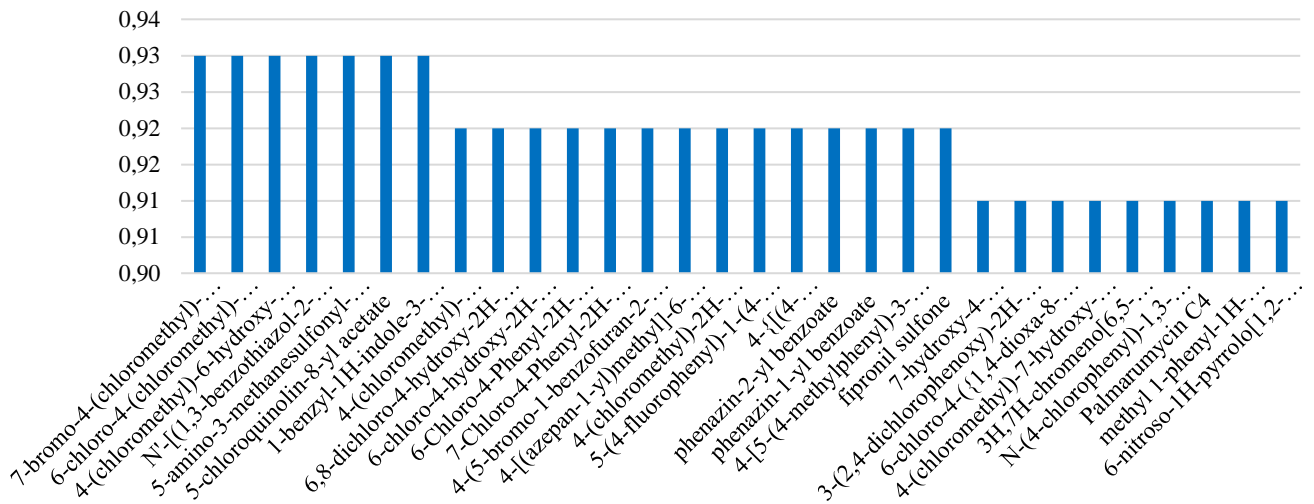


Gambar 2. Hasil Kurva ROC dari Masing-masing Algoritma Pembelajaran Mesin Metode Sidik Jari MACCS. Dapat dilihat bahwa model *Random Forest* mengungguli kedua model lainnya, sedangkan model Support Vector Machine tidak layak digunakan untuk memprediksi skor senyawa PN

Pendekatan skrining berbasis ligan kedua yang digunakan dalam penelitian ini adalah gabungan similaritas sidik jari

dengan pembelajaran mesin. Dengan memasukkan 3.969 ligan aktif dan inaktif, melakukan pre-proses, dan melakukan pelatihan dan tes model, didapatkan hasil statistik yang tertera pada Tabel 3. Evaluasi pembelajaran mesin dari masing-masing tiga algoritma dan dua metode sidik jari menyimpulkan model *Random Forest* MACCS adalah yang terbaik dikarenakan keseluruhan aspek memiliki nilai tertinggi dibandingkan dengan model-model lainnya. Selain itu, nilai masing-masing evaluasi juga memenuhi kriteria minimal, di mana akurasi keseluruhan $\geq 90\%$ (46) dan ROC AUC berada pada rentang cukup diterima (0,7-0,8) (47). Namun, berbeda dengan *Cohen's Kappa* dan Akurasi Seimbang di mana nilai Kappa berada pada rentang persetujuan rendah (0.00-0.40) (48) dan Akurasi Seimbang $< 1\%$ dikarenakan ketidakseimbangan ligan aktif dan inaktif sedari awal. Disamping evaluasi model RF, model SVM, baik MACCS dan Morgan menunjukkan ketidakmampuan untuk mengklasifikasi ligan inaktif dan aktif. Hal ini terlihat pada hasil Akurasi Keseluruhan dan *Cohen's Kappa* yang tidak memunculkan hasil dan kurva ROC AUC mereka berada di bawah 0,5 yang dapat dilihat pada gambar 2. Kedua nilai ini menandakan bahwa model SVM gagal (46–48). Setelah menentukan bahwa model yang digunakan adalah RF MACCS, selanjutnya model diekspor ke tahap terakhir untuk memprediksi similaritas senyawa PN.

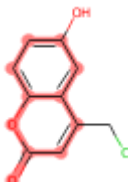
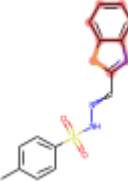
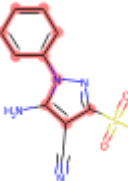
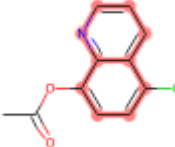
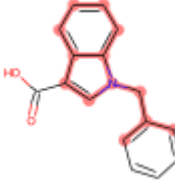
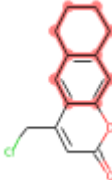
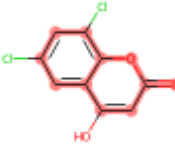
Daftar Senyawa PN Hasil Skrining Model RF MACCS (Kepercayaan >90)

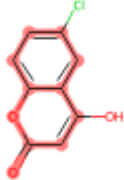


Gambar 3. Daftar Senyawa PN Kandidat dengan Skor Similaritas >90%. Daftar ini diurutkan dari yang tertinggi hingga terendah. Terdapat 30 senyawa PN yang melebihi kepercayaan >90% dari 59.406 senyawa PN yang lolos sebelumnya

Tabel 4. Top 10 Senyawa Hasil Prediksi Model Random Forest MACCS. Terlihat bahwa scaffold kumarin dan alil mendominasi posisi tertinggi prediksi similaritas dengan ligan yang aktif berinteraksi dengan M^{pro}

COCONUT ID	Nama Senyawa	SMILES	Struktur Senyawa (+MCS berwarna)	Skor Similaritas
CNP0044068	7-bromo-4-(chloromethyl)-2H-chromen-2-one	<chem>O=c1cc(CCl)c2ccc(Br)cc2o1</chem>		0,93
CNP0050175	6-chloro-4-(chloromethyl)-2H-chromen-2-one	<chem>O=c1cc(CCl)c2cc(Cl)ccc2o1</chem>		0,93

COCONUT ID	Nama Senyawa	SMILES	Struktur Senyawa (+MCS berwarna)	Skor Similaritas
CNP0061561	4-(chloromethyl)-6-hydroxy-2H-chromen-2-one	<chem>O=c1cc(CCl)c2cc(O)ccc2o1</chem>		0,93
CNP0161048	N'-[(1,3-benzothiazol-2-yl)methylidene]-4-methylbenzene-1-sulfonohydrazide	<chem>Cc1ccc(S(=O)(=O)NN=Cc2nc3ccccc3s2)cc1</chem>		0,93
CNP0054393	5-amino-3-methanesulfonyl-1-phenyl-1H-pyrazole-4-carbonitrile	<chem>CS(=O)(=O)c1nn(-c2ccccc2)c(N)c1C#N</chem>		0,93
CNP0316537	5-chloroquinolin-8-yl acetate	<chem>CC(=O)Oc1ccc(Cl)c2cccnc12</chem>		0,93
CNP0389503	1-benzyl-1H-indole-3-carboxylic acid	<chem>O=C(O)c1cn(Cc2ccccc2)c2ccc1cc2</chem>		0,93
CNP0061285	4-(chloromethyl)-2H,6H,7H,8H,9H-cyclohexa[g]chromen-2-one	<chem>O=c1cc(CCl)c2cc3c(cc2o1)CC3</chem>		0,92
CNP0062084	6,8-dichloro-4-hydroxy-2H-chromen-2-one	<chem>O=c1cc(O)c2cc(Cl)cc(Cl)c2o1</chem>		0,92

COCONUT ID	Nama Senyawa	SMILES	Struktur Senyawa (+MCS berwarna)	Skor Similaritas
CNP0107403	6-chloro-4-hydroxy- 2H-chromen-2-one	O=c1cc(O)c2cc(Cl)ccc2o1		0,92

Hasil prediksi model RF MACCS menunjukkan bahwa terdapat 30 senyawa yang memiliki skor similaritas dengan kepercayaan >90%. Skor kepercayaan tertinggi diperoleh senyawa 7-bromo-4-(chloromethyl)-2H-chromen-2-one dan isomernya dengan skor 0,93 sedangkan skor terendah diperoleh senyawa N-(3-methylphenyl)-4-[[6-oxo-4-propyl-1,6-dihydropyrimidin-2-yl)sulfanyl]methyl}benzamide dengan skor 0,04. Apabila kita sedikit membedah lebih dalam klasifikasi senyawa top 10 pada Tabel 4 berdasarkan kelas senyawanya, dapat disimpulkan kelas senyawa kumarin dan turunannya, pirazol dan turunannya, naftalena, dan quinoline dan turunannya berpotensi dalam menghambat M^{Pro}. Hasil ini juga didukung oleh beberapa penelitian tentang kelas senyawa ini dapat menghambat M^{Pro} (49–52). Untuk memperkuat skor kepercayaan senyawa-senyawa di atas. Penulis membandingkannya dengan menggunakan model yang sama terhadap keempat kontrol positif: ritonavir, lopinavir, nirmatrelvir, GC 376 protease inhibitor.

Dari hasil prediksi skor similaritas, didapatkan hasil dengan skor tertinggi diperoleh sebagai berikut: GC 376 (0,85) > nirmatrelvir (0,64) > ritonavir (0,57) > lopinavir (0,53). Hal ini memberikan hasil bahwa senyawa-senyawa yang berada minimal di atas skor nirmatrelvir berpotensi sebagai senyawa inhibitor M^{Pro}, dengan tetap mempertimbangkan metabolit sekunder yang mudah didapatkan di tanaman. Perlu diingat, hasil ini juga dapat dikatakan sedikit bias dikarenakan skor yang diperoleh hanya merepresentasikan kesamaan substruktur senyawa NP dengan ligan yang dianggap positif berdasarkan *cut-off*, tidak merepresentasikan potensi aktivitas yang sebenarnya sehingga penelitian terkait penambatan molekular terhadap 30 senyawa ini disarankan untuk dilakukan sekaligus mencari nilai IC50/Ki mereka.

KESIMPULAN

COVID-19 merupakan penyakit yang disebabkan oleh SARS-CoV-2 dan pertama kali menjadi *coronavirus* yang mencetak prestasi pandemi global. Berbagai upaya yang dilakukan untuk mengatasi penyakit ini, namun belum

ada satupun terapi yang spesifik, efektif, dan aman. Oleh karena itu, senyawa produk natural yang memiliki aktivitas antivirus dapat menjadi solusi alternatif. Hasil penelitian ini menunjukkan bahwa terdapat 30 senyawa yang dapat menjadi obat yang potensial, di mana 7-bromo-4-(chloromethyl)-2H-chromen-2-one dan isomernya memiliki skor kepercayaan 0,93 terhadap M^{pro}. Selain itu, hasil dari top 10 senyawa PN menunjukkan kelas senyawa kumarin dan turunannya, pirazol dan turunannya, naftalena, dan quinoline dan turunannya berpotensi untuk diteliti lebih jauh.

SARAN

Penelitian ini merupakan langkah yang sangat awal dalam mencari terapi antiviral COVID-19 tertarget M^{pro}. Diperlukan penelitian tambahan terkait penambatan molekuler dan dinamika molekuler untuk mencari tahu nilai IC₅₀/K_i sebelum masuk ke tahap *in vitro* dan *in vivo*. Selain itu, diperlukan juga penelitian mengenai senyawa yang termasuk ke dalam empat kelas teratas, di mana senyawa tersebut mudah diperoleh dan mudah diekstraksi dibandingkan menggunakan senyawa dengan skor kepercayaan tertinggi.

UCAPAN TERIMA KASIH

Terima kasih penulis sampaikan kepada Bapak Dr. rer.nat. apt. Adam Hermawan, M.Sc., Dr. B.S. Ari Sudarmanto, M.Si., Fakultas Farmasi Universitas Gadjah Mada, dan pihak-pihak lainnya yang telah membantu kelancaran penelitian

ini. Keseluruhan alur kerja dan hasil penelitian dapat diunduh pada pranala berikut:

<https://github.com/gpwahyunanda/Natural-Product-LBVS>

DAFTAR PUSTAKA

1. JHU C. COVID-19 Map [Internet]. Johns Hopkins Coronavirus Resource Center. 2022 [cited 2022 Feb 11]. Available from: <https://coronavirus.jhu.edu/map.html>
2. NHS. SARS (severe acute respiratory syndrome) [Internet]. nhs.uk. 2017 [cited 2022 Feb 11]. Available from: <https://www.nhs.uk/conditions/sars/>
3. Abd El-Aziz TM, Al-Sabi A, Stockand JD. Human recombinant soluble ACE2 (hrsACE2) shows promise for treating severe COVID-19. *Signal Transduct Target Ther*. 2020 Nov 3;5(1):1–2.
4. Clinical Trials. COVID-19 Clinical Trials Drug [Internet]. 2022 [cited 2022 Jan 23]. Available from: https://clinicaltrials.gov/ct2/results?cond=2019nCoV&Search=Clear&age_v=&gndr=&type=&rslt=
5. Clinical Trials Arena. Coronavirus treatment: Vaccines/drugs in the pipeline for COVID-19 [Internet]. Clinical Trials Arena. 2020 [cited 2022 Jan 21]. Available from: <https://www.clinicaltrialsarena.com/analysis/coronavirus-MERS-CoV-drugs/>
6. Gautret P, Lagier JC, Parola P, Meddeb L, Mailhe M, Doudier B, et al. Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. *Int J Antimicrob Agents*. 2020;56(1):105949.
7. RAPS. COVID-19 vaccine tracker [Internet]. 2022 [cited 2022 Jan 23]. Available from: <https://www.raps.org/news-and-articles/news-articles/2020/3/covid-19-vaccine-tracker>



8. Akinlalu AO, Chamundi A, Yakumbur DT, Afolayan FID, Duru IA, Arowosegbe MA, et al. Repurposing FDA-approved drugs against multiple proteins of SARS-CoV-2: An in silico study. *Sci Afr*. 2021 Sep;13:e00845.
9. Singh AK, Singh A, Dubey AK. Repurposed Therapeutic Strategies towards COVID-19 Potential Targets Based on Genomics and Protein Structure Remodeling [Internet]. IntechOpen; 2021 [cited 2022 Feb 16]. Available from: <https://www.intechopen.com/online-first/75622>
10. Hoertel N, Sánchez-Rico M, Cougoule C, Gulbins E, Kornhuber J, Carpinteiro A, et al. Repurposing antidepressants inhibiting the sphingomyelinase acid/ceramide system against COVID-19: current evidence and potential mechanisms. *Mol Psychiatry*. 2021 Dec;26(12):7098–9.
11. Kandeel M, Abdelrahman AHM, Oh-Hashi K, Ibrahim A, Venugopala KN, Morsy MA, et al. Repurposing of FDA-approved antivirals, antibiotics, anthelmintics, antioxidants, and cell protectives against SARS-CoV-2 papain-like protease. *J Biomol Struct Dyn*. 2021 Sep;39(14):5129–36.
12. Qamar MT ul, Alqahtani SM, Alamri MA, Chen LL. Structural basis of SARS-CoV-2 3CLpro and anti-COVID-19 drug discovery from medicinal plants. *J Pharm Anal*. 2020 Aug 1;10(4):313–9.
13. Lu J, Chen SA, Khan MB, Brassard R, Arutyunova E, Lamer T, et al. Crystallization of Feline Coronavirus Mpro With GC376 Reveals Mechanism of Inhibition. *Front Chem* [Internet]. 2022 [cited 2022 Apr 24];10. Available from: <https://www.frontiersin.org/article/10.3389/fchem.2022.852210>
14. Sharun K, Tiwari R, Dhama K. Protease inhibitor GC376 for COVID-19: Lessons learned from feline infectious peritonitis. *Ann Med Surg*. 2020 Dec 28;61:122–5.
15. Mahase E. Covid-19: Pfizer's paxlovid is 89% effective in patients at risk of serious illness, company reports. *BMJ*. 2021 Nov 8;n2713.
16. Mody V, Ho J, Wills S, Mawri A, Lawson L, Ebert MCCJC, et al. Identification of 3-chymotrypsin like protease (3CLPro) inhibitors as potential anti-SARS-CoV-2 agents. *Commun Biol*. 2021 Jan 20;4(1):1–10.
17. Mengist HM, Mekonnen D, Mohammed A, Shi R, Jin T. Potency, Safety, and Pharmacokinetic Profiles of Potential Inhibitors Targeting SARS-CoV-2 Main Protease. *Front Pharmacol* [Internet]. 2021 [cited 2022 Apr 24];11. Available from: <https://www.frontiersin.org/article/10.3389/fphar.2020.630500>
18. Andreani J, Le Bideau M, Duflo I, Jardot P, Rolland C, Boxberger M, et al. In vitro testing of combined hydroxychloroquine and azithromycin on SARS-CoV-2 shows synergistic effect. *Microb Pathog*. 2020;145:104228.
19. Choudhary R, Sharma AK. Potential use of hydroxychloroquine, ivermectin and azithromycin drugs in fighting COVID-19: trends, scope and relevance. *New Microbes New Infect*. 2020;35:100684.
20. Chu CM, Cheng VCC, Hung IFN, Wong MML, Chan KH, Chan KS, et al. Role of lopinavir/ritonavir in the treatment of SARS: initial virological and clinical findings. *Thorax*. 2004 Mar;59(3):252–6.
21. Wang M, Cao R, Zhang L, Yang X, Liu J, Xu M, et al. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Res*. 2020 Mar;30(3):269–71.
22. Javelot H, El-Hage W, Meyer G, Becker G, Michel B, Hingray C. COVID-19 and (hydroxy) chloroquine–azithromycin combination: Should we take the risk for our patients? *Br J Clin Pharmacol*. 2020;86(6):1176.
23. BPOM. Potensi Obat Herbal Indonesia [Internet]. 2020 [cited



- 2022 May 6]. Available from: <https://www.pom.go.id/new/view/more/pers/531/Potensi-Obat-Herbal-Indonesia.html>
24. Benarba B, Pandiella A. Medicinal Plants as Sources of Active Molecules Against COVID-19. *Front Pharmacol* [Internet]. 2020 [cited 2022 May 6];11. Available from: <https://www.frontiersin.org/article/10.3389/fphar.2020.01189>
 25. Tresadern G, Bemporad D, Howe T. A comparison of ligand based virtual screening methods and application to corticotropin releasing factor 1 receptor. *J Mol Graph Model*. 2009 Jul;27(8):860–70.
 26. McGaughey GB, Sheridan RP, Bayly CI, Culberson JC, Kreatsoulas C, Lindsley S, et al. Comparison of Topological, Shape, and Docking Methods in Virtual Screening. *J Chem Inf Model*. 2007 Jul 1;47(4):1504–19.
 27. Hamza A, Wei NN, Zhan CG. Ligand-Based Virtual Screening Approach Using a New Scoring Function. *J Chem Inf Model*. 2012 Apr 23;52(4):963–74.
 28. Gadaleta D, Lombardo A, Toma C, Benfenati E. A new semi-automated workflow for chemical data retrieval and quality checking for modeling applications. *J Cheminformatics*. 2018 Dec 10;10(1):60.
 29. DCIS. Daylight Theory: SMARTS - A Language for Describing Molecular Patterns [Internet]. 2007 [cited 2022 Apr 18]. Available from: <https://www.daylight.com/dayhtml/doc/theory/theory.smarts.html>
 30. Maggiora G, Vogt M, Stumpfe D, Bajorath J. Molecular Similarity in Medicinal Chemistry. *J Med Chem*. 2014 Apr 24;57(8):3186–204.
 31. Massagué AC, Ojeda MJ, Valls C, Mulero M, Garcia-Vallvé S, Pujadas G. Molecular fingerprint similarity search in virtual screening. *Methods*. 2015 Jan;71:58–63.
 32. Bajorath J. Machine learning and similarity-based virtual screening techniques. In: *In Silico Drug Discovery and Design* [Internet]. Future Science Ltd; 2013 [cited 2022 Apr 18]. p. 134–46. (Future Science Book Series). Available from: <https://www.futuremedicine.com/doi/full/10.4155/ebo.12.419>
 33. Mitchell JBO. Machine learning methods in chemoinformatics. *Wiley Interdiscip Rev Comput Mol Sci*. 2014 Sep;4(5):468–81.
 34. Li H, Ung CY, Yap CW, Xue Y, Li ZR, Chen YZ. Prediction of estrogen receptor agonists and characterization of associated molecular descriptors by statistical learning methods. *J Mol Graph Model*. 2006 Nov;25(3):313–23.
 35. Tuerkova A, Zdražil B. A ligand-based computational drug repurposing pipeline using KNIME and Programmatic Data Access: case studies for rare diseases and COVID-19. *J Cheminformatics*. 2020 Nov 25;12(1):71.
 36. Sydow D, Morger A, Driller M, Volkamer A. TeachOpenCADD: a teaching platform for computer-aided drug design using open source packages and data. *J Cheminformatics*. 2019 Apr 8;11(1):29.
 37. Berthold MR, Cebron N, Dill F, Gabriel TR, Kötter T, Meinl T, et al. KNIME - the Konstanz information miner: version 2.0 and beyond. *ACM SIGKDD Explor Newsl*. 2009 Nov 16;11(1):26–31.
 38. Bento AP, Hersey A, Félix E, Landrum G, Gaulton A, Atkinson F, et al. An open source chemical structure curation pipeline using RDKit. *J Cheminformatics*. 2020 Sep 1;12(1):51.
 39. Mysinger MM, Carchia M, Irwin John J, Shoichet BK. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J Med Chem*. 2012 Jul 26;55(14):6582–94.
 40. Cichonska A, Ravikumar B, Allaway RJ, Park S, Wan F, Isayev O, et al. Crowdsourced mapping extends the target space of kinase inhibitors [Internet]. *bioRxiv*; 2020 [cited 2022 Apr 20]. p.



- 2019.12.31.891812. Available from: <https://www.biorxiv.org/content/10.1101/2019.12.31.891812v3>
41. PDB. Resolution - Proteopedia, life in 3D [Internet]. 2019 [cited 2022 Apr 20]. Available from: <https://proteopedia.org/wiki/index.php/Resolution>
 42. Lipinski CA. Lead-and drug-like compounds: the rule-of-five revolution. *Drug Discov Today Technol.* 2004;1(4):337–41.
 43. Hu Y, Stumpfe D, Bajorath J. Computational Exploration of Molecular Scaffolds in Medicinal Chemistry. *J Med Chem.* 2016 May 1;59(9):4062–76.
 44. Yang Y. Chapter 3 - Temporal Data Clustering. In: Yang Y, editor. *Temporal Data Mining Via Unsupervised Ensemble Learning* [Internet]. Elsevier; 2017 [cited 2022 Apr 19]. p. 19–34. Available from: <https://www.sciencedirect.com/science/article/pii/B9780128116548000038>
 45. Nighania K. Various ways to evaluate a machine learning models performance [Internet]. Medium. 2019 [cited 2022 Apr 20]. Available from: <https://towardsdatascience.com/various-ways-to-evaluate-a-machine-learning-models-performance-230449055f15>
 46. Brownlee J. How To Know if Your Machine Learning Model Has Good Performance [Internet]. Machine Learning Mastery. 2018 [cited 2022 Mar 18]. Available from: <https://machinelearningmastery.com/how-to-know-if-your-machine-learning-model-has-good-performance/>
 47. El Khouli RH, Macura KJ, Barker PB, Phil D, Habba MR, Jacobs MA, et al. The Relationship of Temporal Resolution to Diagnostic Performance for Dynamic Contrast Enhanced (DCE) MRI of the Breast. *J Magn Reson Imaging JMRI.* 2009 Nov;30(5):999–1004.
 48. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Medica.* 2012 Oct 15;22(3):276–82.
 49. Alexpandi R, De Mesquita JF, Pandian SK, Ravi AV. Quinolines-Based SARS-CoV-2 3CLpro and RdRp Inhibitors and Spike-RBD-ACE2 Inhibitor for Drug-Repurposing Against COVID-19: An in silico Analysis. *Front Microbiol.* 2020 Jul 23;11:1796.
 50. Lv Z, Cano KE, Jia L, Drag M, Huang TT, Olsen SK. Targeting SARS-CoV-2 Proteases for COVID-19 Antiviral Development. *Front Chem* [Internet]. 2022 [cited 2022 May 6];9. Available from: <https://www.frontiersin.org/article/10.3389/fchem.2021.819165>
 51. Negi M, Chawla PA, Faruk A, Chawla V. Role of heterocyclic compounds in SARS and SARS CoV-2 pandemic. *Bioorganic Chem.* 2020 Nov;104:104315.
 52. Özdemir M, Köksoy B, Ceyhan D, Sayın K, Erçağ E, Bulut M, et al. Design and in silico study of the novel coumarin derivatives against SARS-CoV-2 main enzymes. *J Biomol Struct Dyn.* :1–16.