

---

## KLASIFIKASI INHIBITOR SARS-COV-2 MENGGUNAKAN GEOMETRIC DEEP LEARNING

Salamat Nur Himawan, Robieth Sohiburoyyan dan Iryanto

Politeknik Negeri Indramayu

Jl. Lohbener Lama No.08, Legok, Kec. Lohbener, Indramayu, Jawa Barat 45252

snhimawan@polindra.ac.id, robieth.s@polindra.ac.id, iryanto@polindra.ac.id

### ABSTRAK

*SARS-CoV-2 merupakan virus penyebab wabah Covid-19. Saat ini virus tersebut masih berkembang dan memiliki jenis-jenis baru. Pada proses penyebaran dan pengendalian wabah Covid-19 dibuat antivirus yang dapat mencegah perkembangan virus. Penemuan antivirus tidak terlepas dari penemuan senyawa yang dapat menghambat perkembangan virus atau inhibitor. Geometric Deep Learning digunakan untuk klasifikasi senyawa kimia yang berpotensi menjadi inhibitor. Data latih merupakan data senyawa kimia yang berpotensi sebagai inhibitor berdasarkan FDA-Approve Drug Library. Hasil latih model Geometric Deep Learning ditunjukkan dengan loss dan akurasi. Loss terbaik mencapai 0.781 dan akurasi terbaik mencapai 0.553. Hasil menunjukkan model berpotensi dalam klasifikasi inhibitor Covid-19. Model dapat dioptimalisasi dengan menambah data latih yang ada.*

**Kata Kunci:** *Inhibitor, SARS-CoV-2, Geometric Deep Learning*

### PENDAHULUAN

Pandemi COVID-19 telah memiliki dampak yang signifikan terhadap kesehatan global, ekonomi, dan masyarakat. Sejak munculnya pada akhir tahun 2019, penyakit ini menyebar dengan cepat di seluruh dunia dan menyebabkan jutaan kematian. Menurut World Health Organization sampai bulan Februari 2023 ada kasus baru sebanyak 1 juta dan kematian baru sebanyak 7000 [1]. Saat ini jenis virus SARS-CoV-2 masih terus berkembang dan memiliki jenis-jenis baru. Antivirus COVID-19 sangat penting untuk mengendalikan pandemi COVID-19. Antivirus dapat ditemukan melalui eksperimen inhibitor COVID-19.

Penghambat (inhibitor) adalah zat yang dapat mengikat dan mengurangi aktivitas enzim, reseptor, atau biomolekul lainnya [2]. Inhibitor secara potensial dapat mengurangi beban virus pada individu yang terinfeksi, memperpendek durasi penyakit, dan mencegah penularan virus ke orang lain. Beberapa penelitian yang telah berkembang dalam identifikasi inhibitor, seperti yang telah dilakukan oleh Touret, dkk dan Gawriljuk, dkk. Dalam penelitiannya Touret menunjukkan hasil eksperimen beberapa obat seperti Azithromycine, Opipramol dan Quinidine menunjukkan potensi untuk menjadi inhibitors SARS-CoV-2 [3]. Sementara Gawriljuk telah membangun model machine learning yang dapat

mengidentifikasi senyawa yang merupakan inhibitor dari SARS-CoV-2 [4].

Salah satu bagian dari Machine Learning adalah Deep Learning. Deep Learning merupakan metode yang banyak digunakan dalam mengenali pola kompleks pada gambar, teks dan suara [5]. Geometric Deep Learning merupakan teknik dalam membuat model Deep Learning yang dapat belajar dari data non-Euclidean Inhibitor [6]. Geometric Deep Learning mampu memodelkan jenis data yang lebih kompleks dan objek yang saling terhubung, sehingga sangat cocok dalam memodelkan inhibitor SARS-CoV-2.

Oleh karena itu, mempelajari inhibitor COVID-19 merupakan area penelitian yang penting dan dapat memiliki implikasi signifikan bagi pengembangan antivirus. Dengan mengidentifikasi dan mengevaluasi inhibitor yang efektif terhadap SARS-CoV-2, peneliti dapat memberikan kontribusi pada upaya global untuk mengendalikan dan pada akhirnya mengeliminasi pandemi COVID-19.

Hasil latih model Geometric Deep Learning ditunjukkan dengan loss dan akurasi. Loss terbaik mencapai 0.781 dan akurasi terbaik mencapai 0.553. Hasil menunjukkan model berpotensi dalam klasifikasi inhibitor Covid-19. Hasil pengujian model terlihat dari confusion matrix. Terlihat dari confusion matrix

menunjukkan ketidakseimbangan kelas pada data mengakibatkan hasil latih model belum maksimal. Model dapat dioptimalisasi dengan menambah data latih yang ada.

## METODE PENELITIAN

### Pengumpulan Data

Pada penelitian ini data yang digunakan merupakan data senyawa kimia inhibitors Sars-CoV-2. Data merupakan senyawa kimia dari FDA-Approve Drug Library yang telah diuji secara in vitro dapat berpotensi menjadi inhibitor SARS-CoV-2 [3]. Data berisikan SMILE kode senyawa kimia dan potensi untuk menjadi inhibitor. Data terdiri dari 1484 senyawa kimia, 88 diantaranya merupakan senyawa yang berpotensi menjadi inhibitor SARS-CoV-2. Berikut merupakan beberapa bagian dari data yang digunakan.

**Tabel 1. Sample Data**

No	SMILES	Class
0	<chem>CC(C)C1=C(C(=CC=C1)C(C)C)O</chem>	0
1	<chem>C1=NC2=C(N1CCC(CO)CO)N</chem>	0
2	<chem>CCCNC(C)C(=O)NC1=CC=C(C=C1)C1</chem>	0
3	<chem>C1CC(=O)N1CCN(CC1)S(=O)(=O)c2ccc(Cl)s2</chem>	1
4	<chem>C[N+]1=CC=CC=C1/C=N/O.[Cl-]</chem>	0
5	<chem>C1=CC2=C(C=C1O)C(=CN2)CCN.C1</chem>	0

### Pre-processing Data

#### Fiturisasi Data

Fiturisasi data merupakan tahap ekstraksi fitur dari data graf. Graf terdiri dari node (verteks) dan edge (koneksi atau tautan) yang merepresentasikan hubungan antara node. Setiap node dalam graf dapat memiliki fitur atau atribut terkait, dan edge menggambarkan hubungan atau interaksi antara node. Dalam hal ini node merupakan atom dan edge merupakan ikatan antar atom.

Representasi node dan edge yang digunakan merupakan WeaveNet yang dikembangkan oleh Kearnes dkk [7]. Graf direpresentasikan dalam bentuk matriks,

seperti matriks adjacency yang menunjukkan hubungan antar node. Setiap node dalam graf diberikan fitur awal berupa representasi numerik tau vector yang menggambarkan atribut node.

Tujuan dari fiturisasi adalah untuk mempelajari makna dari node dan struktur graf secara keseluruhan untuk melakukan berbagai tugas, seperti klasifikasi node, prediksi koneksi, atau klasifikasi graf. Fiturisasi graf menjadi penting dalam proses ini dengan mengkodekan struktur graf dan atribut node menjadi representasi numerik yang dapat dimasukkan ke dalam jaringan saraf.

#### Oversampling

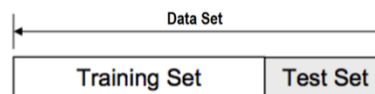
Oversampling adalah metode yang digunakan dalam pemrosesan data yang tidak seimbang. Dalam konteks klasifikasi, oversampling mengacu pada meningkatkan jumlah sampel pada kelas minoritas agar seimbang dengan kelas mayoritas. Tujuan dari oversampling adalah untuk meningkatkan kemampuan model untuk mempelajari pola dari kelas minoritas kurang diwakili dalam dataset.

Dataset memiliki jumlah kelas positif dan negatif yang tidak seimbang. Kelas positif memiliki jumlah yang sangat sedikit dibandingkan dengan kelas negatif. Penambahan jumlah kelas positif mengikuti proporsi antara kelas positif dan negatif, seperti pada persamaan berikut.

$$multiplier = \text{int} \left( \frac{\text{negatif class}}{\text{positif class}} \right) - 1 \quad (1)$$

#### Data Split

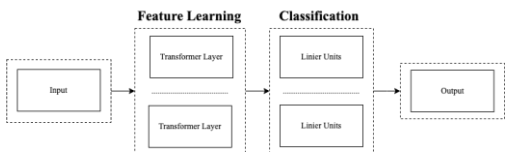
Data split merupakan proses pembagian data menjadi data training dan testing. Tujuan dari data split adalah memastikan model dapat melakukan generalisasi yang pada data baru yang belum pernah dipelajari sebelumnya. Data testing diambil dengan proporsi tertentu. Pembagian data training dan data testing yang digunakan adalah 70:30.



**Gambar 1. Data Split**

### Membangun Model

Model yang akan dibangun untuk penelitian ini kurang lebih akan berbentuk seperti berikut.



Gambar 2. Arsitektur Model

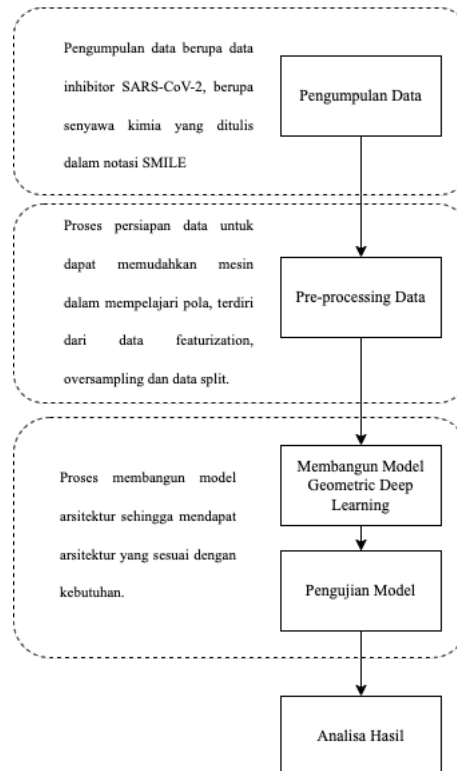
Secara umum, hanya ada 2 bagian utama dalam arsitektur yang akan digunakan, yakni feature extractor berupa beberapa lapisan Transformer layer, dan predictor berupa tumpukan FCN. Graph Transformer merupakan model yang menggunakan self-attention untuk mengambil informasi dari node-node dalam graf. Mekanisme self-attention memungkinkan model untuk memberikan bobot berbeda pada setiap node dan memperhatikan koneksi antar node. Transformer layer yang digunakan merupakan pengembangan model oleh Shi Yungseng, dkk [8].

Bagian pertama akan mengolah bundel data yang masuk untuk membaca korelasi informasi antar tiap titik data dan menghasilkan suatu kumpulan fitur abstrak yang akan dimanfaatkan predictor untuk memprediksi output.

Model bertugas untuk prediksi dalam level graf. Prediksi level graf melibatkan penggunaan representasi graf yang telah diperbarui dan informasi yang disebarkan melalui model untuk membuat prediksi pada tingkat keseluruhan graf. Dalam hal ini adalah dalam kasus penentuan kelas antara inhibitor dan non-inhibitor SARS-CoV-2.

### Alur Penelitian

Tahapan-tahapan dalam penelitian dari proses awal hingga selesai, dimana pada penelitian ini akan menghasilkan model Geometric Deep Learning yang mampu mengklasifikasi inhibitors Sars-CoV-2. Berikut diagram alur yang dilakukan pada penelitian ini.



Gambar 3. Alur Penelitian

### HASIL DAN PEMBAHASAN

Setelah pre-processing data berupa representasi graf dan fiturisasi data. Model Geometric Deep Learning dapat diterapkan untuk memproses graf tersebut. Model akan menyebarkan informasi dan memperbarui representasi fitur simpul melalui iterasi beberapa lapisan. Ini memungkinkan model untuk menangkap pola dan struktur yang kompleks dalam jaringan senyawa kimia.

Eksperimen model dilakukan dengan arsitektur yang ditampilkan pada Tabel 1. Model terdiri dari beberapa layer utama yaitu Transformer layer, Linear Layer, Batch Normalization dan Top-K Pooling.

Tabel 1. Detail Arsitektur Model

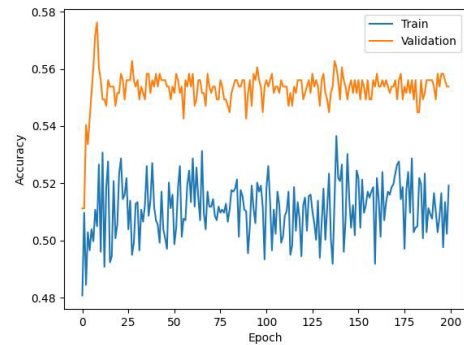
Proses	Parameter
Transformer Convolution	Input size = 192 Output size = 64 Heads = 3 Dropout = 0.2
Linear Transformation	Input size = 192 Output size = 64
Batch Normalization	Input size = 4

Top-K Pooling	Input size = 64 Ratio = 0.5
Transformer Convolution	Input size = 64 Output size = 64 Heads = 3 Dropout = 0.2
Linear Transformation	Input size = 192 Output size = 64
Batch Normalization	Input size = 64
Top-K Pooling	Input size = 64 Ratio = 0.5
Linear Layer	Input size = 192 Output size = 128
Linear Layer	Input size = 128 Output size = 64
Linear Layer	Input size = 64 Output size = 1

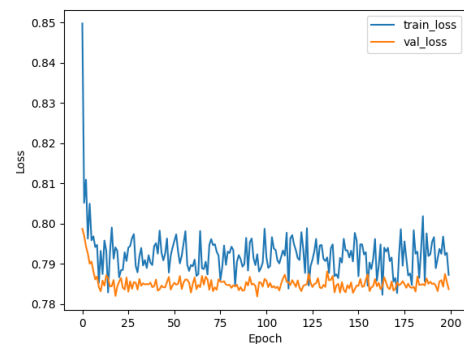
**Tabel 2.** Data Hyperparameter

Hyperparameter	Nilai
Epoch	200
Optimizer	Stochastic Gradient Descent
Learning Rate	0.01
Weight Decay	0.0001
Loss Function	Binary Cross Entropy

Hasil latih model menggunakan dataset dengan 200 epoch. Evaluasi dari model ditunjukkan dengan loss dan akurasi. Grafik loss dan akurasi pada data latih dan validasi ditunjukkan pada Gambar 4 dan Gambar 5. Loss terbaik mencapai 0.781 dan akurasi terbaik mencapai 0.553. Grafik loss dan akurasi menunjukkan potensi model untuk dapat mengklasifikasikan inhibitor SARS-CoV-2.



**Gambar 4.** Grafik Akurasi vs Epoch Train dan Validasi



**Gambar 5.** Grafik Loss vs Epoch Train dan Validasi

**Tabel 3.** Confusion Matrix Pengujian Model

	Prediksi (True)	Prediksi (False)
Aktual (True)	383	597
Aktual (False)	317	604

Keberhasilan model dalam mengklasifikasikan inhibitor dilihat dari pengujian menggunakan data yang belum dikenali. Pada Tabel 3 confusion matrix hasil pengujian menunjukkan model dapat dengan baik memprediksi inhibitor pada kelas negatif tetapi kesulitan dalam memprediksi kelas positif. Hal tersebut terjadi karena data kelas negatif jauh lebih banyak dibandingkan dengan kelas positif. Teknik oversampling untuk memperbanyak kelas negatif tidak begitu optimal karena perbedaan kelas yang sangat jauh. Permasalahan ini dapat diatasi dengan memperkaya data latih yang digunakan.

## PENUTUP

Data inhibitor digolongkan ke dalam data non-euclidian. Data non-euclidian memiliki karakteristik struktur yang kompleks. Pada penelitian ini menunjukan Model Geometric Deep Learning memiliki potensi dalam mempelajari pola pada data non-euclidian, yaitu sebagai klasifikasi SARS-CoV-2. Pada penelitian selanjutnya disarankan untuk memperbanyak data latih dengan perbandingan data kelas yang lebih proporsional.

## DAFTAR PUSTAKA

- [1] Berg J. M, Tymoczko, J.L., Gatto Jr., G.J. dan Stryer, L. Biochemistry 8th Edition. W.H Freeman dan Company. 2015. IEEE Signal Processing Magazine. 2017.
- [2] World Health Organization. Current COVID-19 Situation: Overview Of SARS-CoV-2 Circulating Variants. 2023.
- [3] Franck T., Magali G., Karine B., Antoine N., Etienne D., Xavier de L., dan Bruno C. In Vitro Screening Of A FDA Approved Chemical Library Reveals Potential Inhibitors Of SARS-CoV-2 Replication. Sci Rep 10, 13093. 2020.
- [4] Gawriljuk V. O., Zin P. P. K., Puhl A. C., Zorn K. M., Foil D. H., Lane T. R., Hurst B., Tavella T. A., Costa F. T. M., Lakshmanane P., Bernatchez J., Godoy A. S, Oliva G., Siqueira-Neto J. L., Madrid P. B., dan Ekins S. Machine Learning Identify Inhibitors Of SARS-CoV-2. Journal of Chemical Information and Modelling. 2021.
- [5] LeCun Y, Bengio Y dan Hinton G. Deep Learning. Nature. 2015
- [6] Michael M. B., Joan B., Yann L., Arthur S., dan Pierre V. Geometric Deep Learning: Going Beyond Euclidian Data. 2016.
- [7] Kearnes S., McCloskey K., Berndl M., Pande V., dan Riley P. "Molecular graph convolutions: moving beyond fingerprints." Journal of computer-aided molecular design 30.8 (2016):595-608.
- [8] Yunsheng S., Zhengjie H., Shikun F., Hui Z., Wenjing W., dan Yu S. Mask Label Prediction: Unified Message Passing Model for Semi-Supervised Classification. Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence Main Track. 2021.