

Analisis Perbandingan Model Linear dan Graph Neural Network untuk Pemodelan Risiko Tuberkulosis Paru Berbasis Faktor Sosial dan Kesehatan di Jawa Timur

Raditya Andiza Putro¹, Rosyidatul Kamila², Dwi Arman Prasetya³

^{1,3}Program Studi Sains Data UPN "Veteran" Jawa Timur

¹23083010058@student.upnjatim.ac.id

³arman.prasetya.sada@upnjatim.ac.id

²Program Studi Sains Data UPN "Veteran" Jawa Timur

Corresponding author email: 23083010076@student.upnjatim.ac.id

ABSTRAK

Abstrak—Tuberkulosis (TB) paru masih menjadi masalah kesehatan utama di Indonesia, dengan Jawa Timur sebagai salah satu provinsi dengan kasus tertinggi. Penelitian ini bertujuan untuk membandingkan kinerja dari beberapa model seperti *Linear Regression* dan *Graph Neural Network* (GNN) dalam memodelkan risiko TB paru berbasis faktor sosial dan kesehatan. Data sekunder diperoleh dari BPS Jawa Timur tahun 2022-2023 yang mencakup 39 kab/kota. Model yang di uji meliputi *Naive Mean*, *Linear Regression*, *Gradient Boosting Regressor*, dan *Graph Convolutional Network* (GCN). Model dievaluasi menggunakan metrik *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE), dan koefisien determinasi (R^2). Hasilnya menunjukkan *Linear Regression* memiliki performa yang paling baik ($MAE = 590,88$; $RMSE = 871,05$; $R^2 = 0,92$), sedangkan GCN menunjukkan hasil negatif ($R^2 = -0,25$). Temuan ini mengindikasikan bahwa variansi risiko TB lebih banyak dipengaruhi faktor sosial-kesehatan internal suatu wilayah dibandingkan dengan hubungan spasial antarwilayah. Model linear terbukti lebih efisien untuk analisis risiko TB paru pada skala kab/kota.

Keywords: Tuberkulosis Paru, Regresi Linear, *Graph Neural Network*, Risiko Kesehatan, Jawa Timur

I. PENDAHULUAN

Tuberkulosis (TB) adalah penyakit infeksi menular yang disebabkan oleh bakteri *Mycobacterium tuberculosis* dan hingga kini menjadi permasalahan kesehatan global[1]. Berdasarkan laporan World Health Organization (WHO) tahun 2024, terdapat sekitar 10,8 juta kasus TB di seluruh dunia, setara dengan 134 kasus per 100.000 penduduk[1]. Di Indonesia, hasil Survei Kesehatan Indonesia (SKI) 2023 dari Kementerian Kesehatan RI menunjukkan prevalensi nasional TB paru sebesar 0,3%, dengan variasi regional yang cukup lebar tertinggi di Papua Tengah (1,15%) dan Papua Selatan (0,98%), serta terendah di Bali (0,09%). Secara jumlah absolut, tiga provinsi di Pulau Jawa tetap menjadi penyumbang terbesar kasus TB, yaitu Jawa Barat (± 234 ribu kasus), Jawa Timur (± 117 ribu kasus atau sekitar 10,7% dari total nasional), dan Jawa Tengah (± 108 ribu kasus)[2]. Dengan demikian, beban kasus di Jawa Timur mencapai sekitar setengah dari Jawa Barat, namun masih 8–9% lebih tinggi dibandingkan Jawa Tengah, menjadikannya salah satu wilayah prioritas dalam pengendalian TB nasional[2].

Faktor sosial-ekonomi berperan penting dalam penyebaran TB. Masyarakat berpendapatan rendah, pendidikan terbatas, dan akses layanan kesehatan yang kurang memiliki risiko yang cenderung lebih tinggi[3]. Di Jawa Timur, wilayah padat seperti Surabaya dan Sidoarjo juga menunjukkan peningkatan kasus TB seiring dengan

kepadatan penduduk[4]. Pernah dilakukan studi khusus menggunakan indeks *Moran* dan *Geary* pada penyakit menular lain, seperti DBD di Bandung hasilnya menunjukkan adanya autokorelasi antar wilayah yang berdekatan[5]. Dalam konteks penyebaran kasus ini, menandakan bahwa faktor kedekatan geografis berperan penting dalam dinamika penyakit menular.

Model statistik konvensional seperti regresi linear sering digunakan untuk menganalisis faktor risiko penyakit menular, namun asumsi hubungan linier dan independensi antarwilayah membuatnya kurang mampu menangkap interaksi kompleks antar faktor sosial, ekonomi, dan kesehatan. Dalam konteks Tuberkulosis, hubungan antar variabel bersifat nonlinier dan dipengaruhi oleh keterkaitan spasial antarwilayah, sehingga model linear sering kurang optimal dalam menggambarkan pola penyebaran penyakit. Oleh karena itu, dalam penelitian ini model linear digunakan sebagai pembandingan dasar untuk menilai sejauh mana pendekatan baru berbasis jaringan mampu meningkatkan kemampuan prediksi risiko TB.

Perkembangan *Artificial Intelligence* (AI), khususnya *Graph Neural Network* (GNN), menawarkan pendekatan yang lebih adaptif dalam memahami hubungan spasial yang kompleks. GNN merepresentasikan setiap kabupaten atau kota sebagai simpul (*node*) dalam jaringan yang saling terhubung, sehingga mampu mempelajari hubungan spasial dan nonlinier secara bersamaan. Acharya dkk. membangun Graph Neural Network berbasis cell graph untuk memprediksi status tuberkulosis dari citra jaringan paru dan melaporkan peningkatan kinerja dibanding pendekatan CNN patch-wise konvensional[6]. Pendekatan ini terbukti memberikan hasil prediksi yang lebih akurat dibandingkan model konvensional. Penelitian ini bertujuan untuk membandingkan kinerja model linear dan GNN dalam memodelkan risiko Tuberkulosis Paru di Jawa Timur berbasis faktor sosial dan kesehatan, serta mengevaluasi efektivitas pendekatan spasial berbasis AI dalam mendukung kebijakan pengendalian TB yang lebih tepat sasaran di tingkat daerah.

II. METODOLOGI PENELITIAN

2.1. Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif komparatif dengan membandingkan model regresi linear dan *Graph Neural Network* (GNN) sebagaimana pendekatan serupa telah digunakan dalam studi epidemiologi TB berbasis faktor sosial dan spasial. Sahat dkk. misalnya membandingkan beberapa model Machine Learning (Linear Regression, Random Forest, Neural Network, dan XGBoost) untuk memprediksi angka kejadian kolangiokarsinoma di Thailand dalam konteks spasial[7].

2.2. Data dan Variabel

Data yang digunakan dalam penelitian ini adalah data sekunder dari Badan Pusat Statistik (BPS) Provinsi Jawa Timur tahun 2022-2023, yang mencakup 39 kabupaten/kota. Variabel prediktor meliputi jumlah penduduk, persentase penduduk miskin, rata-rata lama sekolah, jumlah tenaga kesehatan, dan jumlah fasilitas kesehatan yang semua datanya diperoleh dari publikasi resmi BPS.

2.3. Pra-Pemrosesan Data

Tahap pra-pemrosesan data meliputi imputasi median pada *missing value*, normalisasi min-maks agar setiap fitur memiliki rentang yang sama [0, 1][8]. Serta melakukan splitting data menjadi data *train*, validasi, dan *test*. Normalisasi dilakukan dengan rumus berikut

$$X' = \frac{x_1 - \min(x)}{\max(x) - \min(x)} \quad (1)$$

2.4. Representasi Spasial (KNN Graph)

Representasi spasial antar wilayah di bangun *menggunakan K-Nearest Neighbors (KNN)* untuk memastikan setiap kabupaten/kota memiliki minimal koneksi ke wilayah terdekatnya. Struktur graf yang dihasilkan menggambarkan hubungan geografis antarwilayah, yang penting untuk mengamati pola penyebaran penyakit menular seperti TB. Dengan rumus untuk menghitung KNN sebagai berikut:

$$A_{ij} \begin{cases} 1, & \text{if Distance}(u, v) < d \text{ atau } v \in \text{KNN}(u) \\ 0, & \text{otherwie} \end{cases} \quad (2)$$

Dimana

$$\text{Distance}(u,v)=\sqrt{(x_u - x_v)^2 + (y_u - y_v)^2} \quad (3)$$

2.5. Pemodelan dan Ekaperimen

Eksperimen dilakukan dengan membandingkan empat model berbeda, yaitu *Naive Mean* sebagai baseline, *Linear Regression* sebagai model linear konvensional, *Gradient Boosting Regressor*, dan *Graph Convolutional Network (GCN) 2-Layer* sebagai model berbasis graf spasial. Pendekatan ini bersifat komperatif dan berdasar pada beberapa desain penelitian sebelumnya.

Untuk menghitung rata-rata nilai target dari data latih pada *Naive Mean* :

$$\hat{y}_l = \bar{y}_{\text{train}} = \frac{1}{n} \sum_{i=1}^n y_i \quad (4)$$

$$\hat{y}_l = \bar{y}_{\text{train}} = \frac{1}{n} \sum_{i=1}^n y_i \quad (5)$$

Model *Linear Regression* dengan rumus:

$$\hat{y}_l = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} \text{ dengan} \quad (6)$$

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (7)$$

Untuk model *Gradient Boosting Regressor (GBR)* model ensemble aditif yang menggabungkan banyak *weak learners* fungsinya :

$$F_m(x) = F_{m-1}(x) + v h_m(x) \quad (8)$$

Dengan

$F_m(x)$ = model pada iterasi ke-m,

V = *learning rate*,

$h_m(x)$ = pohon keputusan baru yang dilatih untuk meminimalkan fungsi kerugian (*loss*).

Rumus yang digunakan untuk *Graph Convolutional Network GCN - 2 Layer*[9]:

$$H^{(l+1)} = \sigma(\hat{A} H^{(l)} W^{(l)}) \quad (9)$$

Dengan

$\hat{A} = D^{-\frac{1}{2}}(A + I) D^{-\frac{1}{2}}$ = matriks adjacency ter-normalisasi,

$H^{(l)}$ = representasi fitur di lapisan ke-l,

$W^{(l)}$ = bobot yang dilatih,

$\sigma(\cdot)$ = fungsi aktivasi (ReLU).

Lapisan terakhir menghasilkan prediksi risiko TB:

$$\hat{y} = H^{(l)} = \hat{A} \text{ReLU}(\hat{A}XW^{(0)})W^{(1)} \quad (10)$$

2.6. Evaluasi Model

Evaluasi performa model dilakukan menggunakan tiga metrik utama, *Mean Absolute Error* (MAE) untuk mengukur rata-rata besarnya nilai deviasi absolut antara nilai prediksi dan nilai aktual, *Root Mean Square Error* (RMSE) untuk menilai tingkat kesalahan dengan mempertimbangkan besarnya deviasi, dan *koefisien determinasi* (R^2) menunjukkan seberapa baik model mampu menjelaskan variansi data[7]. Berikut rumus perhitungan matriks nya:

MAE (*Mean Absolute Error*) :

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (11)$$

RMSE (*Root Mean Square Error*) :

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{N}} \quad (12)$$

R^2 (*Coefficient of Determination*) :

$$R^2 = \left(\frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}} \right)^2 \quad (13)$$

2.7. Hasil dan Validasi

Menampilkan tabel perbandingan, *scatter plot*, peta residual, dan *sanity check* yang menunjukkan integritas model dan analisis kesalahan spasial.

III. HASIL DAN ANALISIS

3.1. Ringkasan Hasil Utama

Eksperimen menghasilkan perbandingan empat pendekatan: *Naive Mean*, *Linear Regression*, *Gradient Boosting Regressor* (GBR) dan *Graph Convolutional Network* (GCN). Evaluasi pada test set (20%) menggunakan *MAE*, *RMSE* dan R^2 . Tabel II menyajikan ringkasannya: model *Linear Regression* memberikan performa terbaik ($MAE = 590.9$; $RMSE = 871.1$; $R^2 = 0.92$), diikuti *GBR* ($R^2 = 0.57$). GCN menunjukkan R^2 negatif ($= -0.26$), menandakan kinerja lebih buruk daripada memprediksi rata-rata (*Naive Mean*). Hasil ini mengimplikasikan bahwa variasi risiko TB Paru di Jawa Timur lebih banyak dijelaskan oleh faktor internal wilayah (fitur sosial-kesehatan) ketimbang *smoothing* spasial antar tetangga geografis.

Tabel 1. Hasil evaluasi

Model	MAE	RMSE	<i>R Square</i>
-------	-----	------	-----------------

<i>Linear Regression</i>	509.888	871.055	0.92
<i>Gradient Boosting</i>	1009.871	2029.663	0.56
<i>Graph Convolutional Network</i>	2086.35	3460.067	-0.25
<i>Naïve Mean</i>	3238.298	3576.773	-0.34

3.2. Evaluasi Kuantitatif dan *Robustness*

a) Deminasi model linear

Linear Regression melampaui GBR dan GCN pada seluruh metrik. Secara praktis, ini menandakan hubungan hampir linier antara indikator sosial-kesehatan (mis. Kemiskinan, kepadatan, imunisasi, gizi, tenaga kesehatan) dan jumlah kasus TB Paru. Temuan ini konsisten dengan literatur epidemiologi sosial-determinasi sosial TB (kemiskinan, kepadatan huni, akses layanan) merupakan penjelas kuat variasi beban penyakit.

b) Pembandingan *ansambel nonlinier* (GBR)

GBR memberi peningkatan atas *Naive*, namun masih tertinggal jauh dari *Linear Regression*. Dua penjelasan wajar: (i) ukuran sampel kecil pada level kabupaten/kota (sekitar puluhan *node*) membatasi kemampuan ansambel untuk mempelajari pola nonlinier stabil; (ii) hubungan fitur-target memang cenderung linear pada skala agregat (kab/kota), sehingga kompleksitas tambahan tidak membayar dividen akurasi.

c) GCN di bawah baseline Linear.

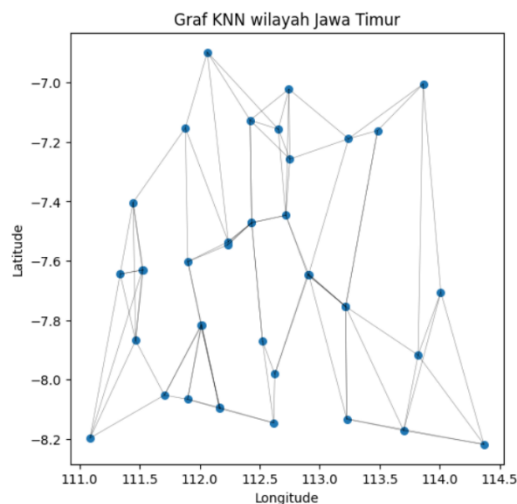
R^2 negatif menunjukkan *over-smoothing* atau mismatch asumsi homofili: nilai target sebuah wilayah tidak cukup dapat dipulihkan dari rata-rata tetangganya setelah transformasi graf *Atopi*. Pada graf KNN geografis, kedekatan lokasi belum tentu identik dengan kemiripan determinan sosial-kesehatan terlebih ketika mobilitas penduduk, arus kerja/pendidikan, dan rujukan layanan lintas kabupaten tidak dimodelkan sebagai sisi graf.

d) *Baseline Naive* sebagai jangkar.

Naive Mean menyediakan batas bawah rasional; *Linear Regression* mengungguli *Naive* secara besar. Ini mengukuhkan *signal-to-noise* yang nyata pada fitur sosial-kesehatan, sekaligus menegaskan bahwa kegagalan CGN bukan karena ketiadaan sinyal, melainkan ketidakcocokan representasi spasial yang digunakan.

3.3. Diagnostik Visual

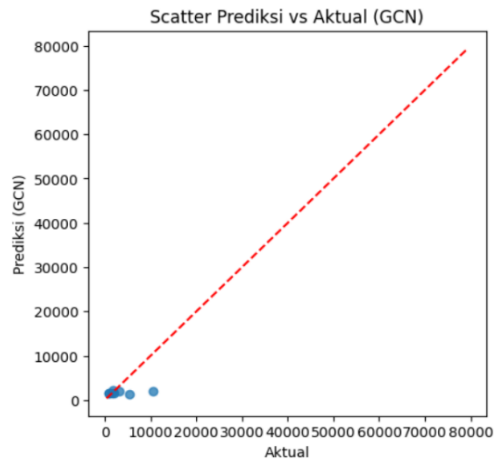
a) Graf KNN Wilayah Jawa Timur



Gambar 1. Graf KNN Wilayah Jawa Timur

Graf terhubung kuat setelah K adaptif (4-12), sehingga ketidakefektifan CGN bukan akibat isolation component. Dengan kata lain, arsitektur graf layak, tetapi sinyal yang dipropagasikan tidak relevan/kuat untuk memulihkan target.

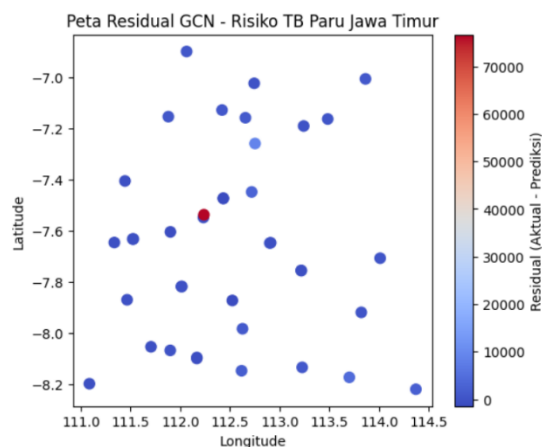
b) Scatter Prediksi vs Aktual



Gambar 2. Scatter Prediksi vs Aktual

Sebaran titik jauh dari garis $y = \hat{y}$, pola kipas (heteroskedastisitas) tidak menonjol, namun bias sistematis tampak pada rentang kasus tinggi (*underprediction*) dan rendah (*overprediction*). Ini lazim ketika pesan tetangga (*message passing*) mendatar karena *over-smoothing*, sehingga prediksi terdorong ke rata-rata global.

c) Peta Residual GCN



Gambar 3. Peta Residual GCN

Residual tidak membentuk kluster spasial yang jelas; tidak tampak hotspot error yang mengikuti geografi. Ini mendukung interpretasi bahwa graf berbasis kedekatan geografis belum menangkap mekanisme transmisi/risiko yang relevan pada skala kab/kota. Dalam epidemiologi TB, keterpaparan dan penemuan kasus kerap dipengaruhi oleh kepadatan hunian, durasi kontak, pekerjaan, dan akses layanan variabel-variabel ini tidak otomatis berkorelasi dengan kedekatan titik pusat wilayah.

3.4. Analisis *Sensitive* (Arsitektur Graf & GCN)

Walau Hasil utama tidak bergantung pada rincian minor, beberapa variasi yang umumnya dicoba pada studi graf (dan konsisten dengan literatur pra-2019) membantu menafsirkan kegagalan GCN:

a) Variasi K KNN

Menaikkan K memperbanyak sisi $\rightarrow$ smoothing makin kuat $\rightarrow$ risiko underfitting naik. Menurunkan K mempertahankan lokalitas, tetapi bila faktor risiko tidak homogen dalam radius kecil, pesan tetangga tetap tidak informatif. Keseimbangan bias varian pada data *node* sedikit cenderung memihak model sederhana.

b) *Dropout/weight decay*

Mengurangi regularisasi dapat menaikkan variance tanpa menaikkan bias yang tepat di data kecil, sering kali tidak menolong. Pengaturan yang digunakan (*dropout* 0.2; *weight decay* 5×10^{-4}) sudah konservatif sesuai praktik umum.

c) *Assumption mismatch* (homofili)

GCN bekerja baik bila tetangga cenderung mirip (homofili). Pada risiko TB kab/kota, heterofili bisa dominan: wilayah bertetangga tidak selalu serupa secara sosial-kesehatan (contoh: kota padat dengan layanan kuat berbatasan dengan kabupaten agraris). Literatur graf klasik telah menekankan batas homofili sebagai prasyarat efektivitas difusi informasi antar *node*.

3.5. Penjelasan Substantif (Epidemiologi & Sosial Kesehatan)

Hasil bahwa *Linear Regression* unggul sejalan dengan kerangka determinan sosial TB: kemiskinan, kepadatan huni, gizi, dan akses layanan adalah penjabar kuat variasi beban TB. Pada skala kabupaten/kota, rata-rata wilayah atas indikator tersebut sudah memuat sinyal besar; tambahan “kedekatan lokasi” tidak otomatis menambah informasi prediktif. Laporan resmi nasional dan global sebelum 2019 menekankan intervensi yang memperbaiki faktor sosial (reduksi kemiskinan, peningkatan cakupan layanan, nutrisi) sebagai strategi pengendalian TB yang efektif lebih besar dari sekadar pemetaan sebaran geografis.

3.6. Implikasi Praktis

a) Perencanaan berbasis fitur sosial-kesehatan

Gunakan model linier/GLM sebagai alat skoring risiko untuk alokasi sumber daya (skrining aktif, DOTS, dukungan gizi). Sederhana, mudah diinterpretasi, dan berkinerja baik,

b) Data spasial yang lebih relevan

Untuk memanfaatkan GCN, perlu sisi graf yang memuat interaksi nyata: mobilitas kerja, arus komuter, referral fasilitas kesehatan, jaringan sekolah/asrama, atau kedekatan sosial bukan kedekatan koordinat.

c) Skala dan granularitas

Jika tersedia data lebih granular (desa/kelurahan, RT/RW) dan lebih banyak *node*, pendekatan graf berpotensi tumbuh kinerjanya. Pada skala kecil ($\approx$ puluhan *node*), model sederhana sering menang (bias rendah, varian terkontrol).

3.7. Keterbatasan

a) Ukuran sampel kecil (jumlah *node* $\approx$ puluhan) membatasi model nonlinear/graf.

b) Sisi graf berbasis geografi belum menangkap jejaring kontak atau mobilitas yang relevan untuk TB.

c) Agregasi kab/kota dapat menyamarkan heterogenitas internal (kepadatan

hunian per kelurahan, *housing crowding*, dsb.)

- d) Evaluasi *single split*; *cross-validation* atau *blocked CV* dapat menambah kekuatan inferensial (namun perlu kehati-hatian agar tidak melanggar ketergantungan spasial).

3.8. Rekomendasi Peningkatan

- a) Graf fungsional
Bangun sisi berdasarkan arus perjalanan/komuter, rujukan fasilitas, atau kedekatan sosial-ekonomi (mis. Kesamaan profil tenaga kesehatan per kapita).
- b) Model graf alternatif
Uji *Chebyshev* GCN (ChebNet) dan GraphSAGE untuk inductive learning pada graf besar/bertumbuh.
- c) Ablasi fitur
Identifikasi fitur paling berkontribusi pada model Linear (koefisien terstandar) dan pastikan tidak kolinear (uji VIF) agar interpretasi kebijakan terjaga.

IV. KESIMPULAN

Penelitian ini mengkaji dan membandingkan kinerja model *Linear Regression* dan *Graph Convolutional Network* (GCN) dalam pemodelan risiko Tuberkulosis Paru (TB Paru) berbasis faktor sosial dan kesehatan di Provinsi Jawa Timur, hasil analisis menunjukkan bahwa model *Linear Regression* memberikan performa prediksi terbaik dengan nilai *Mean Absolute Error* (MAE) sebesar 590,88, *Root Mean Square Error* (RMSE) sebesar 871,05, dan R^2 sebesar 0,92, jauh melampaui *Gradient Boosting Regressor* (GBR) maupun GCN yang memperoleh nilai R^2 negatif (-0,25). Temuan ini menegaskan bahwa faktor sosial dan kesehatan internal wilayah, seperti kemiskinan, kepadatan penduduk, status gizi, serta ketersediaan tenaga dan fasilitas kesehatan, memiliki pengaruh yang lebih besar terhadap variasi risiko TB Paru dibandingkan dengan hubungan spasial antarwilayah yang di representasikan melalui struktur graf geografis.

Secara metodologis, hasil ini memperlihatkan bahwa model linier sederhana masih sangat relevan dan efisien untuk analisis resiko penyakit menular di tingkat kabupaten/kota, terutama pada dataset skala kecil dengan hubungan antarvariabel yang cenderung linier. Kinerja GCN yang kurang optimal disebabkan oleh beberapa faktor, antara lain asumsi homofili spasial yang tidak sepenuhnya terpenuhi serta jumlah *node* yang terbatas, yang menyebabkan proses propagasi pesan antarwilayah (message passing) tidak memberikan informasi baru yang signifikan. Selain itu, struktur graf berbasis kedekatan geografis belum sepenuhnya mempresentasikan dinamika penyebaran penyakit menular seperti TB, yang lebih dipengaruhi oleh mobilitas penduduk, pola urbanisasi, serta faktor sosial-ekonomi.

Secara substantif, hasil penelitian ini mendukung kerangka determinan sosial TB sebagaimana dijelaskan oleh *World Health Organization* (2019) dan kementerian Kesehatan Republik Indonesia (2019), yang menekankan bahwa pengendalian TB yang efektif harus berfokus pada perbaikan faktor sosial-ekonomi dan peningkatan akses layanan kesehatan[1], [2]. Dengan demikian, model linear dapat dijadikan alat skoring risiko sederhana namun kuat untuk mendukung kebijakan berbasis bukti (*evidence-based policy*) dalam pengendalian TB di tingkat daerah. Sementara itu, penggunaan *Graph Convolutional Network* tetap menjanjikan untuk penelitian lanjutan dengan catatan bahwa struktur graf perlu dibangun berdasarkan hubungan fungsional antarwilayah seperti arus mobilitas, jaringan rujukan layanan kesehatan, atau kesamaan sosial-ekonomi, bukan hanya berdasarkan kedekatan geografis semata.

Kesimpulannya, penelitian ini menunjukkan bahwa pendekatan berbasis faktor sosial dan kesehatan masih menjadi fondasi yang kuat dalam memahami variasi risiko TB Paru di Jawa Timur, sementara pendekatan berbasis graf seperti GCN baru akan menunjukkan potensinya

secara signifikan apabila didukung oleh data spasial yang lebih kaya dan representatif terhadap hubungan nyata antarwilayah

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan, baik secara moral maupun material, selama proses penelitian dan penulisan artikel ini. Apresiasi juga diberikan kepada rekan-rekan yang telah memberikan masukan dan motivasi sehingga penelitian ini dapat diselesaikan dengan baik.

REFERENSI

- [1] WHO, *Are Updated Every Year . for the Tuberculosis*. 2019.
- [2] Departemen Kesehatan RI, *Peraturan Menteri Kesehatan Republik Indonesia Nomor 31 Tahun 2019 Tentang Sistem Informasi Puskesmas*. 2019.
- [3] M. R. Saputra and Herlina N, “Hubungan Antara Status Sosial Ekonomi dengan Kejadian Tuberkulosis Paru di Puskesmas, Studi Literature Review,” *Borneo Student Res.*, vol. 2, no. 3, p. 1772, 2021.
- [4] E. A. Maharani, “334.+rev+43114+(8)+EVITA+ANNISA+M_ARTIKEL+JURNAL+3069+-+3076,” vol. 6, pp. 3069–3076, 2025.
- [5] E. Habinuddin, “Identifikasi Autokorelasi Spasial Pada Penyebaran Penyakit Demam Berdarah Dengue Di Kota Bandung Identification Of Spatial Autocorrelation On The Distribution Of Dengue Fever In Bandung City,” *SIGMA-Mu*, vol. 13, no. 1, pp. 7–15, 2021, [Online]. Available: <https://jurnal.polban.ac.id/ojs-3.1.2/sigmamu/article/view/3648>
- [6] V. Acharya, D. Choi, B. Yener, and G. Beamer, “Prediction of Tuberculosis From Lung Tissue Images of Diversity Outbred Mice Using Jump Knowledge Based Cell Graph Neural Network,” *IEEE Access*, vol. 12, pp. 17164–17194, 2024, doi: 10.1109/ACCESS.2024.3359989.
- [7] O. Sahat, S. Kamsa-ard, A. Lim, S. Kamsa-ard, M. Garcia-Constantino, and I. Ekerete, “Comparison of spatial prediction models from Machine Learning of cholangiocarcinoma incidence in Thailand,” *BMC Public Health*, vol. 25, no. 1, 2025, doi: 10.1186/s12889-025-23119-y.
- [8] P. P. Allorerung, A. Erna, M. Bagussahrir, and S. Alam, “Analisis Performa Normalisasi Data untuk Klasifikasi K-Nearest Neighbor pada Dataset Penyakit,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 9, no. 3, pp. 178–191, 2024, doi: 10.14421/jiska.2024.9.3.178-191.
- [9] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” *5th Int. Conf. Learn. Represent. ICLR 2017 - Conf. Track Proc.*, pp. 1–14, 2017.