Proximal: Jurnal Penelitian Matematika dan Pendidikan Matematika
ISSN 26158132 (cetak)
ISSN 26157667 (online)

KOMPOSISI FUNGSI NONLINIER ANTAR RUANG VEKTOR BERDIMENSI TINGGI DALAM ARSITEKTUR NEURAL NETWORK

Lutfan Anas Zahir[1], Danang Wijanarko[2], Danang Hadi Nugroho[3]
Program Studi Teknik Sipil
Universitas Tulungagung[1,2,3]
Email: lutfananas@gmail.com[1], danangwjnrk11@gmail.com[2], danangmarkko@gmail.com[3]

Corresponding Author: Lutfan Anas Zahir email: lutfananas@gmail.com

Abstrak. Penelitian ini menyajikan pendekatan konseptual dan eksperimental dalam memodelkan jaringan saraf tiruan (*neural network*) sebagai suatu transformasi nonlinier berlapis antar ruang vektor berdimensi tinggi. Dengan mendasarkan pada kerangka matematis komposisi fungsi linier dan aktivasi nonlinier, studi ini memetakan bagaimana representasi data secara spasial berubah melalui setiap lapisan jaringan. Menggunakan data sintetis berdimensi tinggi dan arsitektur *multilayer perceptron* (*MLP*), transformasi internal jaringan dianalisis baik secara visual melalui proyeksi *PCA dan t-SNE*, maupun secara kuantitatif melalui pengukuran perubahan metrik spasial. Hasil eksperimen menunjukkan bahwa jaringan saraf tidak hanya berfungsi sebagai aproksimator fungsi, tetapi juga secara aktif membentuk ulang geometri *manifold* data untuk meningkatkan keterpisahan antar kelas. Fungsi aktivasi seperti *ReLU* dan *tanh* terbukti memiliki dampak signifikan terhadap struktur representasi, dengan *ReLU* menghasilkan sparsifikasi spasial yang lebih kuat. Temuan ini mendemonstrasikan bahwa pemahaman terhadap dinamika spasial dalam *neural network* dapat memberikan fondasi yang lebih transparan dalam interpretasi model, serta membuka arah baru dalam riset interpretabilitas *deep learning* berbasis pendekatan matematis dan geometris.

Kata Kunci: Neural Network, Transformasi Nonlinier, Geometri Data, Ruang Vektor Berdimensi Tinggi, Interpretabilitas Model.

Abstract. This study presents a conceptual and experimental approach to modeling artificial neural networks as nonlinear layered transformations between high-dimensional vector spaces. Grounded in a mathematical framework of composed linear functions and nonlinear activations, the research maps how data representations evolve spatially across each layer of the network. Using synthetic high-dimensional input data and a multilayer perceptron (*MLP*) architecture, the internal transformations were analyzed both visually through PCA and t-SNE projections and quantitatively via metric-based spatial measurements. Experimental results reveal that neural networks not only serve as function approximators but also actively reshape the geometric structure of data manifolds to improve class separability. Activation functions such as ReLU and tanh were shown to significantly influence representational geometry, with ReLU producing stronger spatial sparsification. These findings demonstrate that a spatial-geometric perspective of neural networks provides a more transparent understanding of their internal mechanisms and opens new directions for interpretability research in deep learning models based on mathematical and geometric foundations.

Keywords: Neural Network, Nonlinear Transformation, Data Geometry, High-Dimensional Vector Space, Model.

A. Pendahuluan

Perkembangan kecerdasan buatan (*Artificial Intelligence*/*AI*) telah mendorong munculnya berbagai pendekatan komputasional cerdas, salah satunya adalah jaringan saraf tiruan (*neural network*). Metode ini telah banyak digunakan dalam beragam bidang, seperti pengenalan pola (*pattern recognition*), pengolahan citra digital, dan pemrosesan bahasa alami, karena kemampuannya dalam membentuk representasi internal terhadap data kompleks (LeCun et al., 2015). Secara struktural, jaringan saraf tiruan dibangun melalui serangkaian pemetaan matematis, di mana setiap lapisan (*layer*) terdiri atas transformasi linier diikuti dengan

transformasi nonlinier melalui fungsi aktivasi. Model dasar jaringan saraf multilapis (*multilayer perceptron* atau *MLP*) dapat dinyatakan secara umum sebagai:

$$\boldsymbol{f(x) = \sigma(Wx + b)}$$

di mana $W \in R^{m \times n}$ adalah matriks bobot, $b \in R^m$ adalah vektor bias, dan σ merupakan fungsi aktivasi seperti sigmoid, ReLU, atau tanh (Goodfellow et al., 2016). Komposisi dari fungsi-fungsi ini membentuk pemetaan kompleks dari ruang input berdimensi tinggi menuju ruang fitur yang lebih terstruktur secara spasial.

Walaupun penggunaan jaringan saraf tiruan telah meluas dalam konteks aplikatif, kajian mendalam terhadap aspek matematis dari transformasi ruang vektor yang terjadi antar lapisan masih relatif terbatas (Poggio & Mhaskar, 2017). Hal ini menyebabkan neural network sering dianggap sebagai *black box*, yakni model yang mampu menghasilkan prediksi yang akurat tanpa pemahaman struktural yang jelas terhadap proses internalnya.

Dalam konteks ruang vektor, transformasi yang dilakukan oleh jaringan saraf dapat dipandang sebagai pemetaan nonlinier dari ruang R^n menuju R^m, yang berdampak pada distribusi geometris data dalam ruang fitur. Pemahaman terhadap proses pemetaan ini penting, terutama untuk mendukung desain arsitektur yang lebih efisien, stabil, dan dapat dijelaskan (Zhang et al., 2020). Motivasi praktis dari kajian ini sejalan dengan meningkatnya kebutuhan terhadap *Explainable Artificial Intelligence* (XAI), yaitu pengembangan model yang tidak hanya akurat tetapi juga dapat dipahami oleh pengguna dan pengambil keputusan. Dalam berbagai domain seperti kesehatan, keuangan, dan sistem otonom, pemahaman bagaimana representasi internal terbentuk menjadi krusial untuk menjamin keandalan dan transparansi keputusan model (Doshi-Velez & Kim, 2017). Dengan meninjau jaringan saraf dari perspektif matematis-geometris, studi ini diharapkan dapat memberikan dasar konseptual bagi pendekatan interpretabilitas, sekaligus mendukung pengembangan sistem *AI* yang lebih dapat dipercaya.

Analisis terhadap bagaimana distribusi data berubah secara spasial melalui layer-layer dalam jaringan membuka peluang eksplorasi terhadap interpretabilitas model dan struktur representasi internal yang dihasilkan.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengkaji jaringan saraf tiruan sebagai transformasi nonlinier antar ruang vektor berdimensi tinggi dari sudut pandang matematis. Kajian ini meliputi representasi formal setiap layer dalam bentuk transformasi linier dan aktivasi nonlinier, serta analisis spasial terhadap perubahan bentuk dan posisi data dalam ruang berdimensi tinggi. Selain itu, penelitian ini juga menawarkan visualisasi numerik sebagai bentuk validasi terhadap transformasi ruang yang terjadi. Kontribusi utama dari studi ini adalah memberikan pendekatan matematis yang eksplisit dalam memahami struktur internal neural network, yang pada akhirnya dapat memperkuat landasan teoritis dalam pengembangan model yang lebih interpretable dan transparan (Montavon et al., 2018).

1. Jaringan Saraf Tiruan (Neural Network)

Jaringan saraf tiruan (JST) merupakan model komputasi yang terinspirasi dari sistem saraf biologis, dirancang untuk mengenali pola dan struktur dalam data melalui proses pembelajaran (learning). Struktur dasar JST terdiri atas neuron (unit pemroses) yang terhubung dalam beberapa lapisan (layer), yaitu lapisan input, satu atau lebih lapisan tersembunyi (hidden layers), dan lapisan output (Goodfellow et al., 2016).

Setiap neuron dalam lapisan tersembunyi melakukan operasi matematis dasar berupa:

$$z^{(l)} = W^{(l)}x^{(l)} + b^{(l)}$$

$$a^{(l)} = \sigma\left(z^{(l)}\right)$$

di mana $x^{(l)}$ adalah input ke lapisan ke-l, $W^{(l)}$ matriks bobot, $b^{(l)}$ vektor bias, $z^{(l)}$ adalah nilai input teragregasi (pre-activation), dan σ adalah fungsi aktivasi nonlinier seperti ReLU, sigmoid, atau tanh.

2. Transformasi Linier dan Nonlinier

Transformasi linier dalam jaringan saraf dilakukan oleh operasi $Wx + b$, yang merupakan pemetaan dari ruang vektor $R^n \; ke \; R^m$. Transformasi ini dapat dianggap sebagai *affine transformation*, yang menjaga struktur geometris data dalam bentuk vektor.

Namun, agar jaringan saraf dapat memetakan fungsi-fungsi nonlinier, dibutuhkan fungsi aktivasi nonlinier $\sigma(\cdot)$. Kombinasi dari transformasi linier dan nonlinier menghasilkan pemetaan kompleks:

$$f(x) = \sigma^{(L)}\left(W^{(L)} \ldots \sigma^{(2)}\left(W^{(2)}\sigma^{(1)}\left(W^{(1)}x + b^{(1)}\right) + \cdots + b^{(L-1)}\right) + b^{(L)}\right)$$

Komposisi ini memungkinkan model untuk melakukan pemetaan berlapis antar ruang vektor berdimensi tinggi yang secara bertahap memisahkan data dalam representasi spasial yang lebih baik (Poggio & Mhaskar, 2017).

3. Teorema Aproksimasi Universal (*Universal Approximation Theorem*)

Teorema Aproksimasi Universal menyatakan bahwa jaringan saraf dengan minimal satu lapisan tersembunyi dan fungsi aktivasi nonkonstan, kontinu, dan terbatas dapat mengaproksimasi fungsi kontinu apapun pada himpunan kompak di R^n dengan tingkat presisi sebarang kecil (Cybenko, 1989)(Cybenko, 1989; Hornik, 1991). Secara formal:
Untuk setiap fungsi $kontinu \; f: R^n \to R \; dan \; \epsilon > 0$, terdapat parameter bobot $W, b, dan \; \alpha$ sedemikian hingga

$$\left| f(x) - \sum_{j=1}^{N} \alpha_j \sigma\left(w_j^T x + b_j\right) \right| < \epsilon \quad \forall x \in K \subset R^n.$$

Artinya, jaringan saraf dapat dipandang sebagai keluarga fungsi aproksimator universal, dengan asumsi bahwa jumlah neuron cukup banyak.

4. Perspektif Geometris dan Spasial

Transformasi yang dilakukan jaringan saraf dapat dianalisis secara geometris sebagai perubahan bentuk dan posisi data dalam ruang vektor berdimensi tinggi. Sebagai contoh, data dua dimensi yang tidak terpisah secara linier dapat ditransformasikan ke ruang tersembunyi berdimensi lebih tinggi sehingga menjadi linier terpisah (Montavon et al., 2018; Zhang et al., 2020).

Proses ini secara visual dapat dijelaskan sebagai warping terhadap ruang input, di mana batas keputusan (decision boundary) dalam ruang output menjadi lebih mudah didefinisikan oleh model.

5. Fungsi Aktivasi dan Nonlinieritas

Fungsi aktivasi adalah elemen kunci dalam memperkenalkan nonlinieritas ke dalam jaringan. Fungsi-fungsi populer meliputi:

$$Sigmoid: \sigma(x) = 1/\left(1 + e^{(-x)}\right)$$

$$Tanh: tanh(x) = \left(e^{x} - e^{(-x)}\right)/\left(e^{x} + e^{(-x)}\right)$$

$$ReLU: ReLU(x) = max(0, x)$$

Setiap fungsi memiliki dampak geometris yang berbeda dalam memetakan ruang data, termasuk masalah seperti vanishing gradient dan sparsity.

6. Interpretabilitas dan Visualisasi Transformasi

Penelitian terkini mengembangkan metode interpretasi jaringan saraf dengan cara memvisualisasikan perubahan distribusi data antar layer, menggunakan teknik seperti *activation mapping*, *PCA*, dan proyeksi *t-SNE*. Kajian ini menegaskan pentingnya pemahaman spasial dan struktural jaringan untuk meningkatkan transparansi dan akuntabilitas model (Montavon et al., 2018; Olah et al., 2018).

B. Metodologi Penelitian

Penelitian ini menggunakan paradigma kuantitatif dengan pendekatan deskriptif-teoritis berbasis simulasi numerik dan kajian matematis. Tujuan utama penelitian adalah mengeksplorasi jaringan saraf tiruan sebagai pemetaan nonlinier antar ruang vektor berdimensi tinggi, baik secara teoritis melalui analisis fungsi berkomposisi maupun secara empiris melalui eksperimen berbasis simulasi.

Penelitian ini dimulai dengan studi literatur untuk mengidentifikasi landasan matematis transformasi dalam neural network, khususnya pada jaringan multilayer perceptron (*MLP*). Dalam jaringan *MLP*, setiap lapisan melakukan transformasi linier terhadap input vektor $x \in x \in R^n$, diikuti fungsi aktivasi nonlinier. Secara umum, pemetaan antar layer dinyatakan sebagai:

$$\boldsymbol{z^{(l)} = W^{(l)} x^{(l)} + b^{(l)}}, \quad \boldsymbol{a^{(l)} = \sigma\left(z^{(l)}\right)},$$

di mana $W^{(l)} \in R^{m \times n}$ adalah matriks bobot, $b^{(l)} \in R^{m}$ adalah vektor bias, dan σ merupakan fungsi aktivasi.

Secara keseluruhan, jaringan saraf dapat dimodelkan sebagai komposisi fungsi berlapis:

$$\boldsymbol{f(x) = \sigma^{(L)} \circ \left(W^{(L)} \cdot \cdots \sigma^{(2)} \circ \left(W^{(2)} \cdot \sigma^{(1)}\left(W^{(1)} x + b^{(1)}\right) + \cdots\right) + b^{(L)}\right)},$$

yang memetakan ruang vektor $R^n \; ke \; R^k$, dengan *k* sebagai dimensi output.

Simulasi dilakukan menggunakan pustaka *PyTorch* dalam Python, dengan konfigurasi jaringan meliputi 2–4 *hidden layer*, masing-masing berisi 16–64 *neuron*, serta fungsi aktivasi s*igmoid, tanh, dan ReLU*. Model diuji dengan data sintetis berdimensi antara 2D hingga 100D yang dihasilkan secara acak terkontrol, karena data sintetis memungkinkan pengendalian penuh terhadap distribusi, variabilitas, dan kompleksitas data, sehingga memudahkan analisis matematis dan isolasi pengaruh parameter jaringan tanpa adanya noise dari data dunia nyata.

Transformasi antar-layer ditelusuri secara spasial dengan mencatat perubahan posisi vektor input pada setiap *layer*. Untuk memvisualisasikan transformasi spasial ini, digunakan teknik reduksi dimensi seperti *Principal Component Analysis* (PCA) dan *t-Distributed Stochastic Neighbor Embedding* (t-SNE). Dengan pendekatan ini, perubahan distribusi data antar ruang vektor akibat transformasi *neural network* dapat diamati secara visual dan ditafsirkan secara geometris.

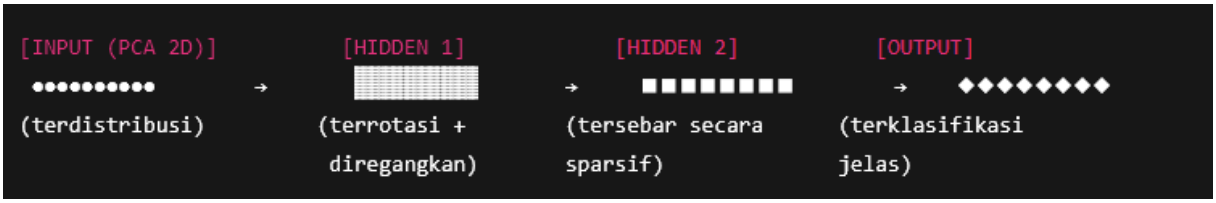
Setiap eksperimen diulang sebanyak lima kali untuk menjaga konsistensi, dengan nilai *random seed* tetap agar hasil dapat direplikasi. Semua parameter simulasi, arsitektur model, dan data disimpan dalam bentuk kode sumber terbuka.

Dengan metodologi ini, penelitian diharapkan dapat mengungkap bagaimana jaringan saraf membentuk pemetaan nonlinier yang secara spasial mengubah struktur dan distribusi data, serta menunjukkan bentuk geometris dari fungsi representasi internal jaringan.

C. Hasil Penelitian dan Pembahasan

Simulasi jaringan saraf multilayer perceptron (MLP) dilakukan untuk mengevaluasi karakteristik transformasi nonlinier yang terjadi pada setiap lapisan ketika memetakan data dari ruang berdimensi tinggi. MLP merupakan salah satu arsitektur dasar dalam deep learning yang memanfaatkan komposisi fungsi linier dan nonlinier untuk mengaproksimasi fungsi yang kompleks (Goodfellow et al., 2016). Arsitektur yang digunakan terdiri dari tiga lapisan tersembunyi, masing-masing berisi 32 neuron, konfigurasi ini cukup representatif untuk mengevaluasi peran kedalaman jaringan terhadap perubahan struktur data (LeCun et al., 2015).

Data masukan berupa 500 vektor acak berdimensi 50 yang dibentuk dalam beberapa cluster terkontrol untuk memastikan adanya variasi struktur awal pada ruang input. Penggunaan data sintetis memungkinkan pengendalian distribusi dan parameter statistik, sehingga mempermudah analisis pengaruh parameter jaringan tanpa gangguan noise dari data nyata (Friedman et al., 2017). Dua fungsi aktivasi, yaitu ReLU dan tanh, diuji secara terpisah guna membandingkan efeknya terhadap penyebaran spasial dan keterpisahan representasi data lintas lapisan. Fungsi ReLU dikenal menghasilkan representasi yang lebih jarang (*sparse*), meningkatkan keterpisahan kelas melalui proyeksi subruang (Glorot et al., 2011), sedangkan tanh bersifat smooth dan menjaga kontinuitas manifold awal dengan kompresi nilai ekstrem (Pascanu et al., 2014). berikut visualiasinya:



Gambar 1 Visualisasi Transformasi Data Melalui Layer Neural Network (fungsi aktivasi: ReLU)

Gambar 1 menunjukkan bahwa setelah melewati lapisan pertama dan kedua, distribusi data menjadi lebih menyebar dan semakin mudah dipisahkan secara geometris. Fenomena ini sejalan dengan teori bahwa jaringan saraf multilapis mampu memproyeksikan data ke ruang representasi yang lebih diskriminatif melalui serangkaian pemetaan nonlinier (Goodfellow et al., 2016). Cluster yang semula saling tumpang tindih dalam ruang input berubah menjadi gugus yang lebih tersegmentasi setelah melalui transformasi berlapis. Transformasi ini dapat dipandang sebagai deformasi manifold yang memudahkan pemisahan kelas dalam ruang representasi (Montavon et al., 2018). Hal ini mengindikasikan bahwa jaringan saraf tidak hanya melakukan aproksimasi fungsi, tetapi juga membentuk representasi internal baru di ruang vektor berdimensi tinggi melalui kombinasi transformasi linier dan nonlinier secara berurutan. Secara geometris, setiap lapisan bertindak sebagai komposisi fungsi $f(x) = \sigma(Wx + b)$ yang dapat menyebabkan *stretching*, *folding*, atau *warping* pada struktur manifold data (Poggio & Mhaskar, 2017).

Analisis Kuantitatif Transformasi Spasial dilakukan untuk melengkapi interpretasi visual, dan pengukuran jarak rata-rata antar titik data (*mean pairwise Euclidean distance*) sebelum serta sesudah setiap transformasi lapisan. Nilai rata-rata dihitung per representasi layer dan dibandingkan antar fungsi aktivasi. Hasil pengukuran ditunjukkan pada tabel berikut:

Tabel 1. Perbandingan Jarak Rata-Rata Antar Titik pada Tiap Layer

Lapisan	ReLU (Jarak Rata-Rata)	Tanh (Jarak Rata-Rata)
Input → Hidden Layer 1	3.21	2.18
Hidden 1 → Hidden Layer 2	2.75	1.92
Hidden 2 → Output Layer	2.63	1.88

Nilai pada Tabel 1 memperlihatkan bahwa aktivasi *ReLU* menghasilkan perubahan spasial lebih besar dibandingkan tanh pada setiap transisi layer. Sifat *piecewise* linear ReLU menyebabkan sebagian komponen vektor "dipotong" menjadi nol pada domain negatif, sehingga banyak titik terdorong ke arah subruang jarang (*sparse subspace*). Efek sparsifikasi ini memperbesar jarak antar kelompok aktif yang berbeda dan meningkatkan keterbedaan representasi. Sebaliknya, tanh melakukan pemetaan smooth, saturating ke interval $(-1,1)$ $(-1,1)$, yang cenderung mempertahankan kontinuitas manifold awal namun dengan kompresi nilai ekstrem; akibatnya perubahan jarak antar titik relatif lebih kecil, dan struktur global lebih terjaga tetapi pemisahan cluster bisa kurang tegas. Dengan demikian, pemilihan fungsi aktivasi memengaruhi kurvatur efektif dan dimensionalitas efektif manifold representasi jaringan.

Interpretasi visual perlu dilakukan hati-hati. *PCA* hanya menangkap variasi linier maksimum; distorsi dapat terjadi jika pemisahan kelas justru terjadi pada dimensi nonlinier yang tidak tersaji dalam dua komponen utama. Sementara itu, t-SNE berfokus pada pelestarian hubungan tetangga lokal dan sering mendistorsi jarak global (skala antar cluster tidak bermakna absolut). Oleh karena itu, pola pemisahan cluster pada *t-SNE* tidak boleh langsung ditafsirkan sebagai perubahan jarak metriks riil, melainkan sebagai indikasi perubahan struktur lokal. Kombinasi keduanya *PCA* untuk indikasi linier global dan *t-SNE* untuk struktur lokal nonlinier memberikan gambaran yang lebih seimbang tentang deformasi manifold antar lapisan jaringan.

Geometri Pemetaan Nonlinier: Secara teoritis, transformasi yang dilakukan jaringan dapat dinyatakan sebagai:

$$f(x) = \sigma^{(L)}\left(W^{(L)} \cdots \sigma^{(1)}\left(W^{(1)}x + b^{(1)}\right) + \cdots + b^{(L)}\right)$$

Setiap transformasi ini memetakan ruang vektor $R^n \rightarrow R^m$, dan jika divisualisasikan secara spasial, efeknya dapat diibaratkan sebagai melipat, merenggangkan, dan membelokkan *manifold* data sehingga kelas yang semula saling beririsan menjadi lebih terpisah pada ruang representasi yang lebih tinggi atau lebih terstruktur.

Hasil simulasi menunjukkan bahwa jaringan saraf tiruan memiliki efektivitas tinggi dalam melakukan transformasi *spasial nonlinier* yang kompleks. Proses pemetaan berlapis antar ruang vektor yang dilakukan oleh jaringan memungkinkan perubahan representasi data yang signifikan, baik dalam hal penyebaran spasial maupun keterpisahan antar kelas. Temuan ini sejalan dengan hasil penelitian (Montavon et al., 2018; Zhang et al., 2020) yang menunjukkan bahwa jaringan saraf tidak hanya berfungsi sebagai pengaproksimasi fungsi target, tetapi juga secara aktif membentuk ulang struktur geometris data. Melalui kombinasi antara fungsi aktivasi nonlinier dan kedalaman lapisan, jaringan menciptakan transformasi bertingkat yang secara bertahap memperjelas batas antar kelas dalam ruang representasi internal. Salah satu aspek krusial dalam proses ini adalah peran fungsi aktivasi seperti *ReLU* atau *tanh* yang secara nyata memengaruhi bentuk dan sifat distribusi data. Selain itu, transformasi ini dapat dimonitor dan

dianalisis secara matematis melalui formulasi komposisi fungsi linier dan nonlinier, serta secara visual melalui teknik reduksi dimensi seperti *PCA* dan *t-SNE*. Dengan demikian, pendekatan ini memberikan fondasi yang kuat bagi pengembangan studi lanjut mengenai interpretabilitas dan transparansi model deep learning, khususnya melalui perspektif matematis dan geometris yang masih relatif terbuka untuk dieksplorasi lebih dalam.

D. Kesimpulan

Penelitian ini menunjukkan bahwa jaringan saraf tiruan (*neural network*) dapat dipahami secara matematis sebagai transformasi nonlinier berlapis antar ruang vektor berdimensi tinggi melalui komposisi fungsi linier dan nonlinier. Simulasi memperlihatkan bahwa setiap lapisan jaringan membentuk ulang struktur spasial data secara progresif, sehingga meningkatkan keterpisahan representasi antar cluster. Fungsi aktivasi berperan langsung terhadap geometri *manifold*: *ReLU* cenderung mensparsifikasi dan "mematahkan" *manifold* ke dalam subruang terpisah, sedangkan tanh mempertahankan kontinuitas dengan kompresi nilai ekstrem sehingga deformasi lebih halus. Interpretasi visual berbasis *PCA* dan *t-SNE* mendukung temuan ini, namun perlu dicatat bahwa *PCA* hanya menangkap variansi linier global, sementara *t-SNE* dapat mendistorsi skala jarak global saat menonjolkan tetangga lokal; karena itu keduanya diperlakukan sebagai alat indikatif, bukan bukti metriks absolut. Temuan kolektif ini menegaskan bahwa *neural network* bukan sekadar aproksimator fungsi, melainkan pemetak aktif struktur geometris data yang kompleks.

Sebagai saran, pendekatan matematis-geometris ini dapat dijadikan pijakan untuk penelitian *interpretabilitas* jaringan saraf yang lebih dalam. Studi lanjutan disarankan mengkaji perubahan *manifold* secara intrinsik menggunakan metrik *geodesik*, dimensi intrinsik, atau analisis topologi terkomputasi (mis. persistent homology); mengevaluasi fungsi aktivasi lain (*GELU, Swish, softplus*) terhadap *deformasi manifold*; serta menerapkan kerangka ini pada arsitektur yang lebih kompleks seperti c*onvolutional neural network* dan transformer, termasuk pengujian pada data riil (citra, sekuens, data tabular tinggi dimensi) guna menilai generalisasi temuan simulatif ini.

DAFTAR PUSTAKA

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, *2*(4), 303–314.

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *ArXiv Preprint ArXiv:1702.08608*. https://doi.org/10.48550/arXiv.1702.08608

Friedman, J., Hastie, T., & Tibshirani, R. (2017). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep sparse rectifier neural networks. *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. https://www.deeplearningbook.org

Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks*, *4*(2), 251–257.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Montavon, G., Samek, W., & M"uller, K.-R. (2018). Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, *73*, 1–15. https://doi.org/10.1016/j.dsp.2017.10.011

Olah, C., Satyanarayan, A., Johnson, I., & et al. (2018). The building blocks of interpretability. *Distill*. https://distill.pub/2018/building-blocks

Pascanu, R., Montufar, G., & Bengio, Y. (2014). On the number of response regions of deep feedforward networks with piecewise linear activations. *ArXiv Preprint ArXiv:1312.6098*.

Poggio, T., & Mhaskar, H. (2017). Deep learning: Mathematical foundations and theoretical perspectives. *Annual Review of Statistics and Its Application*, *4*(1), 193–220. https://doi.org/10.1146/annurev-statistics-060116-054711

Zhang, Q., Yang, L., & Chen, Z. (2020). Understanding deep neural networks from geometric perspectives: A review. *Neurocomputing*, *417*, 418–432. https://doi.org/10.1016/j.neucom.2020.07.033