

## Model Artificial Neural Network (ANN) dan Multi-Layer Perceptron (MLP) untuk Prakiraan Sebaran Kasus Tuberkulosis.

Muhammad Resha<sup>1</sup>, Ansar Suyuti<sup>2</sup>, Rhiza S Sadjad<sup>3</sup>, Muhammad Niswar<sup>4</sup>

<sup>1</sup>Computer Science, STMIK AKBA, Makassar, Indonesia. <sup>2,3,4</sup>Electrical Engineering, 90245 Hasanuddin University, Indonesia

<sup>1</sup>mresha@akba.ac.id, <sup>2</sup>asuyuti@unhas.ac.id, <sup>3</sup>rhiza@unhas.ac.id, <sup>4</sup>niswar@unhas.ac.id

### Abstract

**Background:** Tuberculosis di wilayah Kota Makassar merupakan salah satu kasus tertinggi pada tahun 2019. Penelitian ini bertujuan untuk memprediksi penyebaran kasus tuberculosis dengan menggunakan metode *Artificial Neural Network* (ANN) dan *Multi-Layer Perceptron* (MLP).

**Methods:** Desain penelitian menggunakan metode *Artificial Neural Network* (ANN) dan *Multi-Layer Perceptron* (MLP). Populasi penelitian ini sebagai kasus adalah pasien TB dengan variable terikat adalah pasien TB di kota Makassar. Sampel pada penelitian ini berjumlah 5148 responden, data training yang digunakan data tahun 2017 - 2018 dan tahun 2019 untuk data testing.

**Results:** Hasil penelitian didapatkan bahwa ANN adalah model terbaik untuk memprediksi penyebaran kasus TB di Kota Makassar dengan tingkat akurasi yang diperoleh 82.48% dibandingkan dengan model MLP dengan tingkat akurasi yang diperoleh 73.99%.

**Conclusion:** Model ANN memperoleh hasil bahwa penyebaran kasus TB dipengaruhi oleh jumlah penduduk dan luas suatu daerah, sehingga perlu tindakan kusus untuk daerah yang memiliki jumlah penduduk tertinggi di kota Makassar untuk mencegah peningkatan jumlah kasus TB di kota Makassar.

**Keywords:** Tuberculosis, ANN, MLP, training, testing

### Introduction

Infeksi bakteri *Mycobacterium tuberculosis* menjadi penyebab terjadinya Penyakit Tuberkulosis pada manusia dan menjadi penyebab kematian tertinggi penyakit menular di Indonesia[1]. Hal ini terlihat pada laporan WHO tahun 2019 indonesia menjadi urutan ketiga terbesar di dunia[2] dengan jumlah kasus 146.856 di Sulawesi Selatan[3] Jumlah tersebut cukup besar sehingga kasus ini merupakan masalah besar untuk Sulawesi Selatan . Oleh karena itu, dalam penelitian Tuberkulosis, jumlah kasus menjadi penting untuk di prediksi.

Muhammad Tahir Khan et al. telah melakukan penelitian tentang prediksi kasus tuberculosis di Pakistan tahun 2019. Prediksi kasus TB dilakukan dengan *Artificial Neural Network*. Penelitian ini menggunakan data dari laboratorium rujukan TB provinsi, Khyber Pakhtunkhwa, Pakistan dengan

akurasi 94,58% [4]. Pada tahun 2018, Abolfazl Mollalo dkk juga melakukan penelitian penyebaran tuberculosis secara spasial menggunakan model *Artificial Neural Network*. Pada peneltian ini data kasus pada tahun 2006 sampai 2010 dikumpulkan dan diperoleh dari makalah Scales et al.

Prediksi kasus Tuberkulosis juga dilakukan di China tahun 2020. Yanling Zheng et al. membangun model prediksi TB dengan metode Box-Jenkins dan menggunakan RMSE, MAE dan MAPE untuk menguji kinerja dan akurasi model. Jumlah data yang dikumpulkan berjumlah 587.344 kasus dalam rentang waktu 2012 - 2019. Hasil prediksi model AIC dan SC ditemukan angka 2,87 dan 2,98. Sedangkan indeks akurasi RMSE, MAE dan MAPE masing-masing adalah 0,98, 0,77 dan 5,8. [5]

Model algoritma Multi Layer Perceptron untuk memprediksi penyakit menular juga di lakukan oleh Halim[6]. Prediksi penyebaran penyakit dilakukan perhitungan data penyebaran di wilayah sekitarnya

kemudian data tersebut dihitung secara terpisah untuk setiap wilayah. Namun, model ini berlaku untuk penyakit DBD dengan jumlah 252 data training tahun 2012-2015 and 126 data tahun 2017-2018 data testing.

Pada penelitian ini model yang diusulkan berupa jaringan syaraf tiruan dengan arsitektur nearest neighbor multi layer perceptron untuk memprediksi jumlah kasus TB yang terinfeksi pada saat ini dan tahun depan.

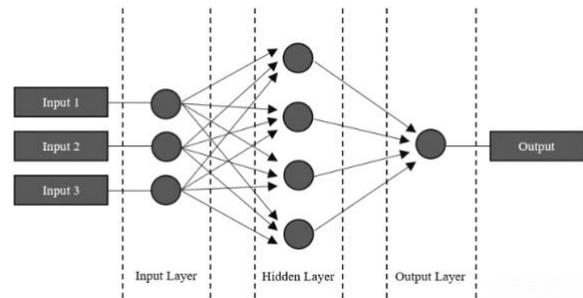
### Proposed Methods

Dalam penelitian ini, kami mengusulkan model Artificial Neural Network dan perceptron spasial multi-layer untuk memprediksi penyebaran kasus Tuberkulosis. Model yang diusulkan ini memperhatikan cara penyebaran penyakit TB yang ditularkan melalui udara[7], maka jika kasus TB menginfeksi suatu daerah tertentu maka daerah sekitarnya akan rentan terhadap penyebaran penyakit tersebut[8]. Prediksi penyakit di suatu daerah sangat dipengaruhi oleh penyakit TB di daerah sekitarnya. Sehingga untuk memprediksi kemungkinan penyebaran penyakit TB di suatu wilayah, perlu dilakukan perhitungan data penyebaran TB dari wilayah sekitarnya[9].

### Artificial Neural Network

Dalam penelitian ini kami mengusulkan model *Artificial Neural Network* (ANN) yaitu metode kecerdasan buatan yang menirukan cara kerja otak manusia dalam menyelesaikan suatu masalah. ANN mengidentifikasi kegiatan berdasarkan data masa lalu kemudian mempelajari data masa lalu untuk memberi keputusan terhadap data yang akan datang[10].

ANN bekerja dengan menggunakan algoritma untuk menginterpretasikan data non-linear dari pola sekuensial. Jaringan terdiri dari sejumlah unit yang lebih kecil yang disebut neuron, diatur antara input data dan output hasil menjadi banyak lapisan[11]. ANN merupakan bentuk arsitektur yang didistribusikan secara paralel dengan jumlah node yang banyak dan hubungan antar node atau di sebut *neuron*. Hubungan satu *neuron* dengan *neuron* lain memiliki nilai disebut sebagai bobot sehingga setiap *neuron* memiliki nilai yang terkait sebagai nilai aktivasi *neuron*[12].

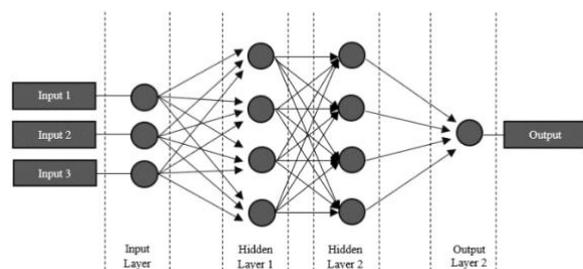


Gambar 1. ANN dengan 1 hidden layer.

ANN memiliki dasar *artificial neuron* yaitu model dengan aturan perkalian, penjumlahan, dan aktivasi. Pada masukan *artificial neuron*, input memiliki bobot setiap nilai input dikalikan dengan bobot individual. Kemudian penjumlahan bobot pada input dan bias. Selanjutnya penjumlahan bobot dan bias melewati fungsi aktivasi kemudian diteruskan ke output dari *neural network*. Meskipun prinsip kerja dan aturan ANN tampak seperti hal yang sederhana, Tetapi potensi dan kalkulasi dari model menjadi kuat jika ANN disambungkan antara satu dengan yang lain (Gambar 1). Kompleksitasnya dapat dikembangkan dari beberapa aturan dasar sederhana[11].

### Multi-Layer Perceptron

*Multi-layer perceptron* (MLP) metode yang paling sederhana yang digunakan untuk pengenalan pola[13], terdiri dari *neuron* yang diurutkan menjadi lapisan. Lapisan pertama disebut lapisan input, dan yang terakhir disebut sebagai lapisan output, lapisan diantaranya adalah lapisan tersembunyi[14].



Gambar 2. MLP dengan 2 hidden layer

Tujuan utama dari MLP adalah untuk memperkirakan beberapa fungsi  $f$ . Dalam model regresi,  $y = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n$ . Fungsi  $y = f(x)$  memetakan vektor input  $x$  ke dalam nilai  $y$ . Jaringan *feedforward* mendefinisikan pemetaan  $y = f(X, \beta)$  dan mempelajari nilai parameter  $\beta$  yang akan menghasilkan perkiraan fungsi terbaik.

## Results

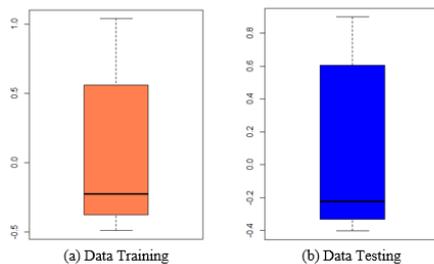
### Data Collection

Data yang kumpulkan dari kota Makassar provinsi Sulawesi Selatan yang terdiri dari data karakteristik jenis kelamin, umur, kecamatan pasien, kelurahan pasien, tipe diagnosis, lokasi anatomi, hasil akhir pengobatan, lama rawat, jumlah penduduk. dan luas daerah dalam tiga tahun terakhir.

Tabel 1. Statistik deskriptif perbandingan data *training* dengan data *testing*

Variabel	N		Max		Min		Mean		Var	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing
Y	3,432	1,715	44.000	44.000	0.000	0.000	2.714	2.721	13.936	14.089
X1	3,432	1,715	61039	61039	973	973	10405	10412	103302682	103440699
X2	3,432	1,715	10.940	10.940	0.050	0.050	1.284	1.29	4.187	4.190
X3	3,432	1,715	12	12	1	1	6.5	6.5	13	13
X4	3,432	1,715	2018	2019	2017	2019	2019	2018	0.250	0.250
<b>Total</b>	<b>5148</b>									

Variansi data *training* dengan data *testing* tidak jauh berbeda yang berarti pengelompokkan data *training* cukup mewakili dari dataset agar model yang diperoleh dapat di terapkan pada data *testing*. Box plot dari data *training* (Gambar 3) menunjukkan bahwa terdapat outlier pada dataset yang sebenarnya dan outlier tersebut tidak dapat ditangkap oleh model. Karena data *training* terdapat outlier maka digunakan median untuk memprediksi data real.



Gambar 3. Perbandingan data *training* dengan data *testing*

## Discussion

Hal utama yang dilakukan sebelum melakukan analisis yaitu menormalkan data terlebih dahulu, kemudian dari dataset diambil data tahun 2017-2018 sebagai data *training* yang nantinya dilakukan analisis sampai menemukan model dengan menggunakan dua metode yaitu ANN dan MLP. Model yang diperoleh akan diterapkan pada data *testing* untuk memvalidasi model tersebut.

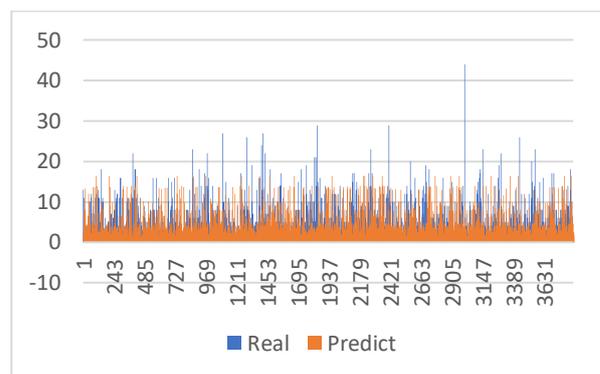
### Data Training and Testing

Data yang digunakan dari tahun 2017-2019 sebagai dataset terdiri dari data 5,148. Kemudian dari dataset diambil data tahun 2017-2018 sebagai data *training* untuk membuat model dan data tahun 2019 sebagai data *testing* yang digunakan untuk memvalidasi model yang diperoleh. Secara lengkap, perbandingan data *training* dengan data *testing* dapat dilihat pada tabel berikut:

Kemudian membandingkan kedua model dengan melihat tingkat akurasi.

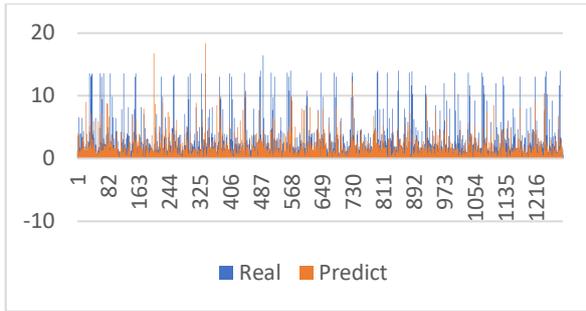
### Artificial Neural Network (ANN)

Memodelkan data *training* yang telah dibentuk menggunakan metode ANN kemudian membandingkan hasil prediksi yang diperoleh dengan data asli.



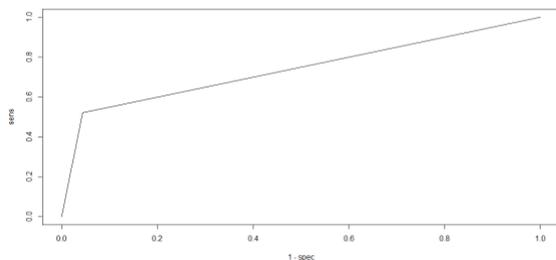
Gambar 4 data real vs predict data *training*

Gambar 4 menunjukkan bahwa hasil prediksi dapat mengikuti pola data asli selama tahun 2017-2018. Karena hasil prediksi dari data *training* cukup mewakili data asli, maka model yang diperoleh di terapkan pada data *testing*.



Gambar 5 data real vs predict data *testing*

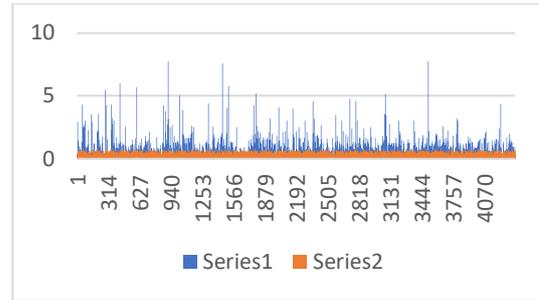
Gambar 5 menunjukkan bahwa model prediksi dapat mengikuti pola dari data yang sebenarnya. Namun, *mean square error* (MSE) yang diperoleh dari prediksi untuk tahun 2019 adalah 2.808. Jika dibandingkan dengan MSE data asli yaitu 2.724 sedikit lebih rendah dibandingkan MSE hasil prediksi. Artinya hasil uji-t menunjukkan bahwa tidak ada perbedaan yang signifikan antara rata-rata data asli dengan rata-rata hasil prediksi dengan tingkat akurasi yang diperoleh 82.48%. Dengan melihat tingkat akurasi yang diperoleh, maka dapat dikatakan jumlah penduduk dan luas suatu daerah sangat berpengaruh penyebaran TB yang menandakan bahwa semakin banyak jumlah penduduk dan luas daerah tidak bertambah maka semakin meningkat pula penyebaran kasus TB serta kurva ROC ANN dapat dilihat pada Gambar 6.



Gambar 6 plot ROC ANN

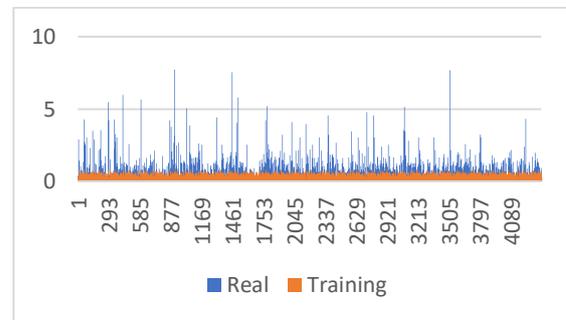
### *Multi-Layer Perceptron (MLP)*

Model MLP dengan menggunakan data *training* diperoleh tingkat akurasi 73.99%, sehingga dapat dikatakan bahwa model yang diperoleh cukup akurat untuk di terapkan pada data *testing*.



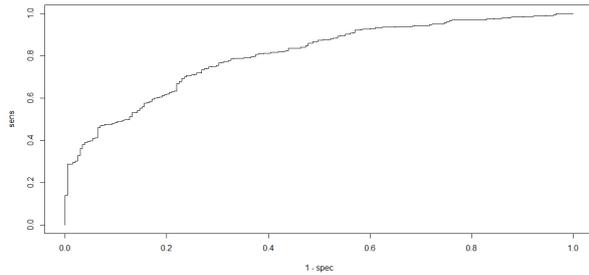
Gambar 7 data real vs predict data testing

Gambar 7 menunjukkan bahwa hasil prediksi dapat tidak mengikuti pola data asli selama tahun 2017-2018. Hal ini dikarena model MLP tidak memperhatikan jumlah kasus melainkan terinfeksi atau tidak, namun ketika dilihat dari tingkat akurasi yang diperoleh (74.94%), hal ini menjamin model cukup akurat untuk diterapkan pada data *testing*.



Gambar 8 data real vs predict data testing

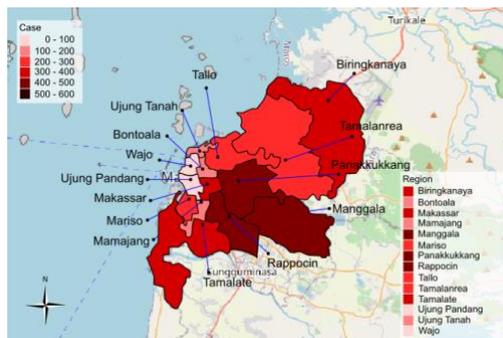
Gambar 8 menunjukkan bahwa model prediksi yang diperoleh lebih rendah dari data yang sebenarnya. Namun, *mean square error* (MSE) yang diperoleh dari prediksi untuk tahun 2019 adalah 0.378. Jika dibandingkan dengan MSE data asli yaitu 2.724 sedikit lebih rendah dibandingkan MSE hasil prediksi. Artinya hasil uji-t menunjukkan bahwa ada perbedaan antara rata-rata data asli dengan rata-rata hasil prediksi namun tingkat akurasi yang diperoleh 73.99%. Bila dibandingkan dengan tingkat akurasi yang diperoleh menggunakan model ANN (82.48%), maka terlihat jelas model ANN lebih baik memodelkan data TB tahun 2017-2018 dibandingkan model MLP. Hal ini dikarenakan model MLP memerlukan prediksi dengan melihat apakah disuatu daerah terinfeksi TB atau tidak dengan tidak memperhatikan jumlah kasusnya, namun model MLP jauh lebih cepat proses running programnya bila di bandingkan dengan model ANN. Tetapi model ANN melakukan prediksi dengan tingkat keakuratan yang tinggi serta kurva ROC MLP dapat dilihat pada Gambar 9.



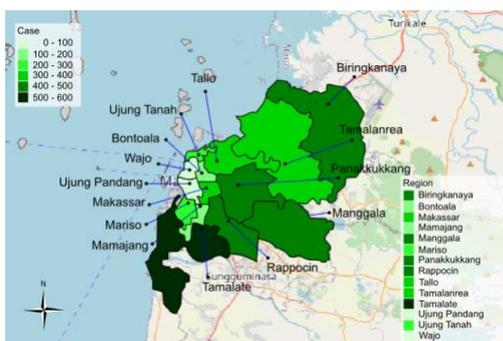
Gambar 9 Plot ROC MLP

### Map implementation

Pemodelan ini diimplementasikan dalam basis peta dalam memantau kasus TB [15] [16]. Dari peta ini dapat melihat data sebaran TB untuk setiap kecamatan di Kota Makassar pada tahun 2017-2018. Data jumlah kasus di setiap kecamatan akan ditampilkan dalam warna. Mulai warna gelap (jumlah korban besar) sampai warna yang paling terang (jumlah korban kasus terkecil).

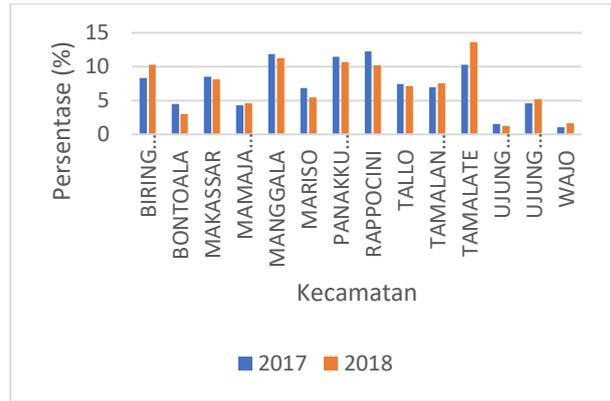


Gambar 10 peta jumlah kasus TB tahun 2017



Gambar 11 peta jumlah kasus TB tahun 2018

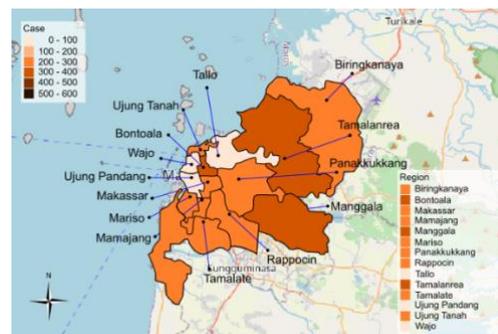
Secara keseluruhan, jumlah kasus TB di Kota Makassar dari tahun 2017-2018 dapat dilihat pada gambar berikut:



Gambar 13 plot jumlah kasus TB tahun 2017-2018

### Forecast TB in 2019

Hasil prakiraan penyebaran jumlah kasus TB pada tahun 2019 menggunakan metode ANN dapat dilihat pada peta berikut



Gambar 14 peta jumlah kasus TB tahun 2019

### Conclusion

Model ANN adalah model terbaik untuk memprediksi penyebaran kasus *tuberculosis* (TB) di kota Makassar dengan tingkat akurasi adalah 82.48% dibandingkan model MLP dengan tingkat akurasi yang diperoleh adalah 73.99%. Model ANN memperoleh hasil bahwa penyebaran kasus TB dipengaruhi oleh jumlah penduduk dan luas suatu daerah, sehingga perlu tindakan kusus untuk daerah yang memiliki jumlah penduduk tertinggi di kota Makassar untuk mencegah peningkatan jumlah kasus TB di kota Makassar.

Dalam penelitian ini, menggunakan empat variable bebas yaitu jumlah penduduk, luas daerah, bulan dan tahun serta variable terikatnya jumlah kasus TB. Untuk penelitian berikutnya prediksi jumlah kasus TB dengan memperhatikan krakteristik penduduk pada daerah tersebut.

---

### Acknowledgement

---

Penulis ingin mengucapkan terima kasih atas masukan dari reviewer yang tentunya meningkatkan kejelasan makalah ini. Kami juga ingin mengucapkan terima kasih kepada Dinas Kesehatan Makassar atas diskusi yang bermanfaat.

---

### References

---

- [1] D. Collins, F. Hafidz, and D. Mustikawati, "The economic burden of tuberculosis in Indonesia," vol. 21, no. December 2016, pp. 1041–1048, 2017.
- [2] WHO, *Guidelines on tuberculosis infection prevention and control 2019*, vol. 91, no. 5, 2019.
- [3] "Profil-Kesehatan-indonesia-2019."
- [4] M. T. Khan, A. C. Kaushik, L. Ji, S. I. Malik, S. Ali, and D. Wei, "Artificial Neural Networks for Prediction of Tuberculosis Disease," vol. 10, no. March, pp. 1–9, 2019, doi: 10.3389/fmicb.2019.00395.
- [5] Y. Zheng, L. Zhang, L. Wang, and R. Rifhat, "Statistical methods for predicting tuberculosis incidence based on data from Guangxi, China," *BMC Infect. Dis.*, vol. 20, no. 1, Apr. 2020, doi: 10.1186/s12879-020-05033-3.
- [6] S. Halim, A. Handojo, I. E. Widodo, Felecia, and T. Octavia, "Spatial multi-layer perceptron model for predicting dengue fever outbreaks in surabaya," *Adv. Sci. Technol. Eng. Syst.*, vol. 5, no. 5, pp. 103–108, 2020, doi: 10.25046/aj050514.
- [7] E. A. Nardell, "Transmission and Institutional Infection Control of Tuberculosis," 2015.
- [8] A. R. Abdul Rasam, N. M. Shariff, and J. F. Dony, "Identifying high-risk populations of tuberculosis using environmental factors and gis based multi-criteria decision making method," in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, Sep. 2016, vol. 42, no. 4W1, pp. 9–13, doi: 10.5194/isprs-archives-XLII-4-W1-9-2016.
- [9] M. Idrus, A. Mallongi, and J. Ibrahim, "Current Research in Tuberculosis Surveillance System Model for Pulmonary Tuberculosis Suspected in Pangkep Region , Indonesia," doi: 10.3923/crt.2017.1.7.
- [10] M. T. Khan, A. C. Kaushik, L. Ji, S. I. Malik, S. Ali, and D. Q. Wei, "Artificial neural networks for prediction of tuberculosis disease," *Front. Microbiol.*, vol. 10, no. MAR, 2019, doi: 10.3389/fmicb.2019.00395.
- [11] J. Zupan, "Introduction to Artificial Neural Network (ANN) Methods: What They Are and How to Use Them Introduction to Artificial Neural Network (ANN) Methods: What They Are and How to Use Them\*." [Online]. Available: <https://www.researchgate.net/publication/251626579>.
- [12] A. A. El-Solh, C. Bin Hsiao, S. Goodnough, J. Serghani, and B. J. B. Grant, "Predicting active pulmonary tuberculosis using an artificial neural network," *Chest*, vol. 116, no. 4, pp. 968–973, 1999, doi: 10.1378/chest.116.4.968.
- [13] L. Sarangi, M. N. Mohanty, and S. Pattanayak, "Design of MLP Based Model for Analysis of Patient Suffering from Influenza," in *Procedia Computer Science*, 2016, vol. 92, pp. 396–403, doi: 10.1016/j.procs.2016.07.396.
- [14] G. E. Hinton, "Learning multiple layers of representation," *Trends in Cognitive Sciences*, vol. 11, no. 10, pp. 428–434, Oct. 2007, doi: 10.1016/j.tics.2007.09.004.
- [15] K. A. Alene, Z. A. Wagaw, and A. C. A. Clements, "Mapping tuberculosis prevalence in Ethiopia: Protocol for a geospatial meta-analysis," *BMJ Open*, vol. 10, no. 5, May 2020, doi: 10.1136/bmjopen-2019-034704.
- [16] A. Mollalo, L. Mao, P. Rashidi, and G. E. Glass, "A gis-based artificial neural network model for spatial distribution of tuberculosis across the continental united states," *Int. J. Environ. Res. Public Health*, vol. 16, no. 1, Jan. 2019, doi: 10.3390/ijerph16010157.