

Komparasi Klasifikasi antara ANN dan *Binary Gradient Boosted Decision Tree* pada Citra Sel Darah Merah yang Terinfeksi Malaria

Dimas Maulana Arbi
Program Studi Teknik Elektro, Fakultas Teknologi Industri
Universitas Islam Indonesia
Yogyakarta, Indonesia
18524078@students.uii.ac.id

Abstract—Sudah sangat umum pada zaman sekarang penggunaan *machine learning* digunakan pada praktis kehidupan sehari-hari seperti *Chat GPT*, *Autopilot* pada mobil, dll. Selain pada praktis kehidupan sehari-hari *machine learning* juga dapat digunakan dalam bidang medis, salah satu contohnya dalam pengolahan citra medis. Yang sangat jelas penggunaan *machine learning* yaitu pada pengolahan atau penggabungan dari citra-citra yang diambil oleh *ct-scan*. Selain hal tersebut *machine learning* dapat digunakan sebagai alat bantu mengklasifikasi sebuah penyakit ataupun kondisi medis yang lain, yang dimana klasifikasi tersebut didasarkan dari fitur-fitur yang diekstraksi dari *dataset*. Sedangkan metode *machine learning* yang dipilih juga akan mempengaruhi hasil akurasi dari model yang dibuat untuk melakukan klasifikasi. Secara umum metode algoritma *machine learning* tradisional sangat cocok dengan fitur data yang diekstraksi secara manual, sedangkan pada *machine learning* moderen tidak terlalu cocok untuk melakukan klasifikasi dengan fitur data yang diekstraksi secara manual

Keywords—*component, formatting, style, styling, insert (key words)*

I. INTRODUCTION

Klasifikasi pada citra digital di dunia sudah semakin banyak dan kompleks. Salah satu contohnya di bidang citra medis dapat digunakan sebagai alat bantu memperbanyak dan memperdalam data pengetahuan tentang citra medis tersebut. Citra medis sendiri merupakan sebuah citra atau gambar dari sebuah organisme baik kompleks maupun sederhana, ataupun hanya sebuah sel yang saling berkaitan [1]–[5]. Kaitan-kaitan tersebut baik berupa simbiosis mutualisme, komensalisme, maupun parasitisme pada tubuh hewan maupun manusia yang menguntungkan bagi manusia [6]. Contoh-contoh dari citra medis adalah citra mikroskopis, citra *X-ray*, citra *CT-Scan*, citra MRI, citra USG, dll [7]. Secara umum pada analisis citra-citra tersebut dilakukan oleh radiolog dan analisis sehingga dapat digunakan untuk diagnosis keadaan yang sedang dialami oleh hewan maupun manusia.

Data-data dari analisis citra medis tersebut serta data citra itu sendiri dapat dipelajari lebih lanjut keterkaitannya satu sama lain berdasarkan statistik. Analisis statistik tersebut menggunakan algoritma *rule-based learning*, *lazy learning*, *error back propagation*, *decision tree induction*, *nearest-neighbor*, dan algoritma baru dalam bidang analisis statistik [8]. Dari semua algoritma analisis statistik tersebut terhadap data citra medis dapat menghasilkan berbagai akurasi yang bervariasi salah satu contohnya analisis citra medis *X-ray* paru-paru sehat dan terjangkit penyakit, serta citra medis *dermoscopy melanoma* kanker kulit dengan menggunakan algoritma *Convolutional Neural Network* (CNN), *Artificial Neural Network* (ANN), *K Nearest Neighbor* (KNN),

Support Vector Machine (SVM), *Naive Bayes*, *Decision Tree*, *Random Forest*, *Random Subspace*, *Logistic Regression*, dan *Fuzzy Logic*. Dari *X-ray* paru-paru menghasilkan akurasi algoritma CNN sebesar 95%, ANN sebesar 91%, SVM sebesar 84%, KNN sebesar 86%, DT sebesar 74%, NB sebesar 81%, LR sebesar 92%, RF sebesar 93%, RS sebesar 84%, dan *Fuzzy* sebesar 81%. Sedangkan untuk *melanoma* kanker kulit menghasilkan akurasi CNN sebesar 93%, ANN sebesar 96%, SVM sebesar 84%, KNN sebesar 95%, DT sebesar 84%, NB sebesar 80%, LR sebesar 87%, RF sebesar 94%, RS sebesar 93%, dan *Fuzzy* sebesar 90% [9].

Pada jurnal ini saya menggunakan algoritma *Artificial Neural Network* (ANN) dan *Decision Tree* berdasarkan *Light Gradient Boost Machine* (LGBM) menggunakan fitur *Gray Level co-Occurrence Matrix* (GLCM), *gray image histogram*, *red channel image histogram*, *green channel image histogram*, *blue channel histogram*, *gradient edge image*. Untuk *dataset* citra medis yang digunakan adalah citra medis mikroskopis sel darah merah yang normal dan terjangkit malaria yang sudah diproses di laboratorium medis menggunakan *Peripheral Blood Smear* (PBS) yang diambil dari Kaggle. Malaria merupakan sebuah penyakit yang berasal dari mikro organisme bersel tunggal dari grup *Plasmodium* yang menyebar berasal dari nyamuk *Anopheles* dan menginfeksi sel darah merah [10].

II. LITERATURE REVIEW

A. Red and White Blood Cell Classification using Artificial Neural Network

Dalam jurnal penelitian klasifikasi sel darah merah dan putih menggunakan ANN dengan bantuan dari Monte-Carlo *sampling* menghasilkan akurasi sebesar 92% dengan rincian presisi dapat mengklasifikasikan sel *erythrocyte* sebesar 88%, *lymphocyte* sebesar 95%, *platelets* sebesar 91%, *neutrophil* sebesar 93%, *monocyte* sebesar 87%, dan *eosinophil* sebesar 91%. Pada *dataset* yang digunakan sudah melewati pemrosesan di laboratorium medis menggunakan *Peripheral Blood Smear* (PBS) [11].

B. A Comparative Analysis of Neural Network and Decision Tree Model for Detecting Result Anomalies

Selain itu dalam penelitian sebuah komparasi analisis model dari *Neural Network* dan *Decision Tree* untuk mendeteksi anomali jawaban pada ujian pelajar dan *Continuous Assessment* (CA) menghasilkan akurasi *decision tree* sebesar 91% dan *Neural Network* sebesar 96%, dari total *dataset* sebesar 1300 buah yang berisi 660 hasil jawaban anomali ujian pelajar, 629 jawaban CA, dan 11 tidak anomali [12].

C. DTBC: Decision Tree based Binary Classification using with Feature Selection and Optimizer for Malaria Infected Erythrocyte Detection

Dalam jurnal penelitian klasifikasi *erythrocyte* yang terinfeksi malaria menggunakan klasifikasi biner *Decision Tree* (DTBC) memberikan hasil perbandingan antara DTBC dan Skala *Heuristic* untuk Mengestimasi Parasit Malaria Prematur (SEMPs) dengan akurasi sebesar 92% untuk DTBC dan 86% untuk SEMPS [13].

III. METODE

Agar citra sel darah merah dapat diekstraksi fiturnya maka perlu dilakukan segmentasi, yang pertama adalah membuat *Area of Interest* (AoI) dengan cara membuat *mask* sehingga informasi pada area yang bukan AoI tidak terambil. Untuk membuat *mask* diperlukan untuk merubah citra berwarna menjadi citra keabuan dengan persamaan (Eq1) setelah itu melakukan *thresholding* berdasarkan *gray-level* tiap piksel dengan cara melihat histogram citra tersebut. setelah itu aplikasikan nilai *threshold* untuk membuat *mask* berupa filter bahwa *gray-level* yang tidak termasuk pada nilai *threshold* akan dijadikan bernilai 0.

$$\omega(x,y,z) = \frac{\sum_{i=0}^{|P|} (r_i x + g_i y + b_i z)}{|x|+|y|+|z|} \quad (\text{Eq1})$$

Pada persamaan (Eq 1) variabel x, y, z merupakan keluaran *channel* warna dari *channel* warna citra asli R, G, B (r_i, g_i, b_i) dari tiap piksel ke i^{th} pada citra.

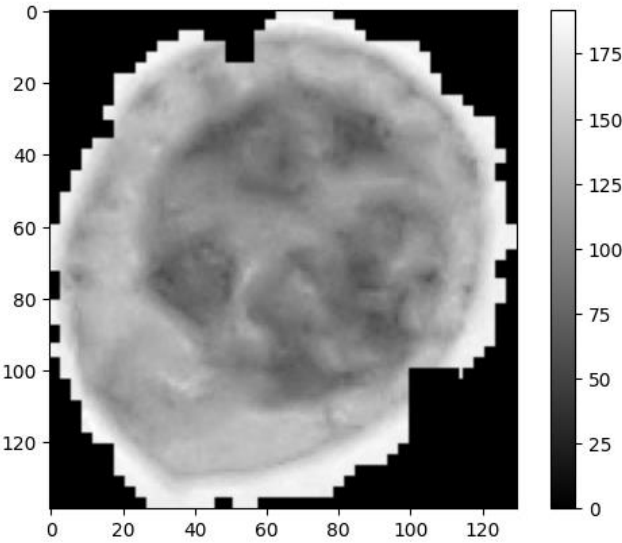


Fig. 1. Citra sel darah merah yang sudah menerapkan mask sehingga tercipta AoI

Lakukan perkalian *gray-level* ataupun tiap *red-level*, *green-level*, *blue-level* pada tiap piksel citra keabuan ataupun citra berwarna menggunakan persamaan (Eq2) citra sehingga hanya mendapatkan informasi pada AoI seperti pada (Fig 1.).

$$x_i = \sum_{i=0} (gl_i \times \alpha_i) \quad (\text{Eq 2})$$

Pada persamaan (Eq2) x_i merupakan hasil *gray-level* pada tiap piksel citra keabuan ataupun *red-level*, *green-level*, *blue-*

level pada tiap piksel citra warna. Setelah itu dapatkan histogram dari citra keabuan, citra berwarna pada masing-masing *channel* merah, hijau, dan biru.

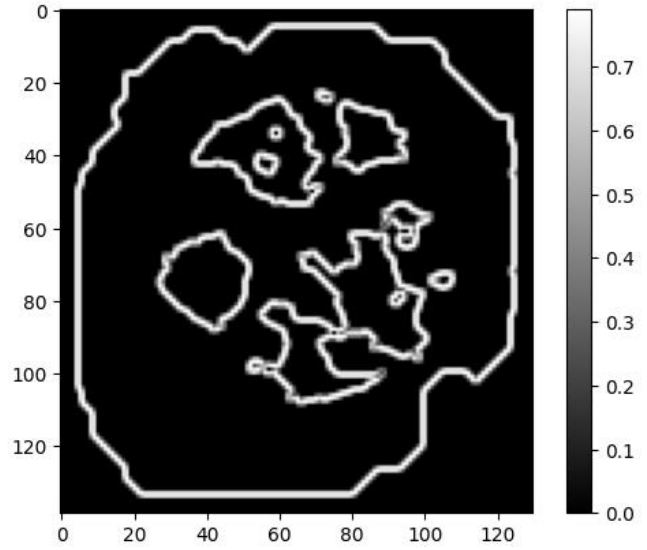


Fig. 2. Edge dari citra

Selain selain fitur histogram terdapat juga fitur gradien dari edge dengan cara melakukan *thresholding* kedua dengan cara melihat *gray-level* pada piksel di tengah sel darah merah pada citra. Setelah itu sama seperti pada membuat *mask* AoI dengan cara menambahkan antara *threshold* kedua dengan *threshold* pertama dan menerapkan *filter sobel* seperti pada (Fig 2.). Dari menerapkan *filter sobel* tersebut akan mendapatkan *edge* bentuk dari sel sehingga mendapatkan gradien dari *edge* citra dengan persamaan (Eq 3).

$$G = \sqrt{G_x^2 + G_y^2} \quad (\text{Eq 3})$$

Fitur yang digunakan selain di atas juga menggunakan fitur *Gradient Level co-Occurrence Matrix* (GLCM) dari citra keabuan dengan menerapkan persamaan (Eq 4).

$$C_{\Delta x, \Delta y}(i, j) = \sum_{z=1}^n \sum_{y=1}^m \begin{cases} 1, \text{if } I(x,y)=i \text{ and } I(x+\Delta x, y+\Delta y)=j \\ 0, \text{otherwise} \end{cases} \quad (\text{Eq 4})$$

Yang mana citra keabuan berupa (I), ($\Delta x, \Delta y$) merupakan variabel posisi piksel pada citra, (i, j)th merupakan nilai dari *co-occurrence* matriks dari variabel posisi piksel yang diberikan.

Dari semua metode-metode ekstraksi fitur di atas didapatkan deskriptor-deskriptor fitur sebanyak 127 buah, dan dapat dilihat 5 buah sampel fitur pada kolom dari 4 buah citra pada baris yang sudah diekstraksi pada tabel (Table I).

TABLE I. TANEL DESKRIPSI FITUR

skewness_gray	skewness_blue	skewness_green	skewness_red	entropy_gray
9,382016	8,422295	10,136413	8,221491	4,3275714
12,842174	12,346506	13,325415	11,420147	4,090796
9,0590725	8,779307	9,502989	8,3155985	4,120892
12,469995	11,945644	12,879091	11,7134	3,9860554

A. Gradient Boosting Decision Tree

Decision Tree merupakan sebuah hirarki pengambil keputusan berbentuk seperti pohon yang percabangannya memberikan hasil pemilihan keputusan serta konsekuensinya, yang umum digunakan pada penelitian yang memiliki perhitungan membuat strategi pengambilan keputusan untuk mencapai hasil yang diinginkan dengan cara menganalisis hasil dari keputusan sesuai dengan parameter yang digunakan [14]. *Decision Tree* merupakan sebuah sistem *supervised learning* pada ilmu *machine learning*, yang berisikan 3 buah level berupa *root node* di level pertama, *several sub-node* di level kedua dan seterusnya sebelum akhir, serta *leaves* di level terakhir proses pengambilan keputusan [15]. Di dalam *Decision Tree* merupakan gabungan dari beberapa fungsi matematika sehingga dapat melakukan klasifikasi [16].

Light Gradient Boosting Machine (LGBM) merupakan sebuah metode pembuatan *Decision Tree* berdasarkan menyusun model prediksi yang lemah dan biasanya mengungguli metode *Random Forest* [17], [18]. LGBM cara kerjanya bukan dengan cara mengembangkan *sub-nodes wise* tetapi menumbuhkan *leaves wise*. seperti pada gambar (Fig 3).

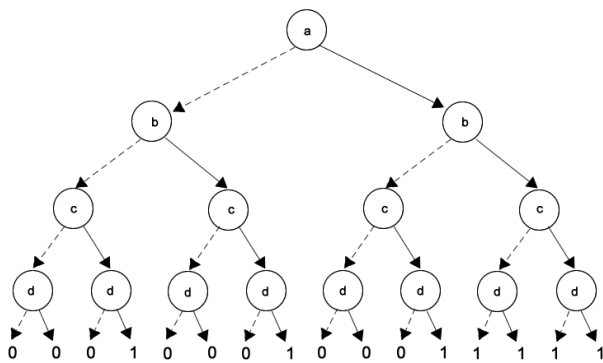


Fig. 3. Contoh *Decision Tree* Biner

B. Artificial Neural Network

Artificial Neural Network (ANN) merupakan cabang dari model *machine learning* yang dibangun menggunakan prinsip *neuronal organization* yang ditiru dari jaringan sistem saraf hewan [19], [20]. Sehingga pada setiap neuron terdapat pemrosesan informasi seperti pada otak secara biologis, dan di dalam tiap sebuah neuron melakukan pengelompokkan berdasarkan nilai *input* tersebut. Tiap neuron tersebut

terhubung melalui *edges*, pada *edges* tersebut pengolahan secara matematis dengan menggunakan persamaan fungsi *non-linear* dari nilai *input*. Hasil dari perhitungan nilai *input* akan dikelompokkan ke neuron selanjutnya berdasarkan pertimbangan dari penjumlahan nilai dari input serta umum disebut sebagai *activation*, yang dimana neuron tersebut dikelompokkan dan disebut *layer*. Umumnya pada sebuah sistem ANN terdapat minimal 2 buah *layer* berupa *input* dan *output*, sedangkan layer yang terdapat di tengah-tengah *layer input* dan *output* tersebut disebut *hidden layer*. Di dalam *edges* tersebut terdapat *hyperparameter* yang menentukan parameter-parameter perhitungan untuk mengelompokkan nilai input. Contoh dari model ANN dengan 2 *hidden layer* dapat dilihat pada gambar (Fig 4).

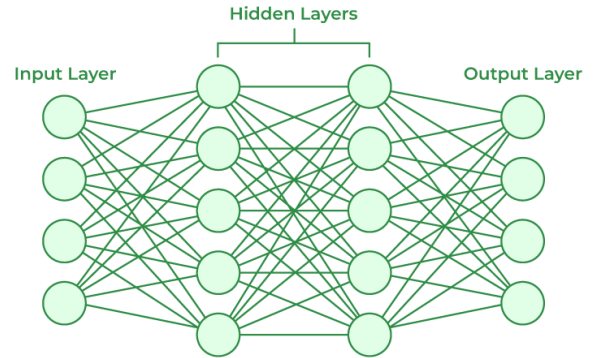


Fig. 4. Contoh Model *Artificial Neural Network*

IV. HASIL DAN ANALISI

Dari *dataset* 27.600 buah citra medis mikroskopis sel darah merah terbagi menjadi 13.800 buah citra sel darah merah yang terinfeksi malaria dan 13.800 buah citra sel darah merah yang tidak terinfeksi malaria. Dari fitur-fitur yang telah diekstrak dari citra yang digunakan sebagai variabel dan terinfeksi atau tidak terinfeksi adalah 1 dan tidak terinfeksi adalah 0. Label fitur tersebut dapat dilihat sampelnya pada tabel (Tabel II).

TABLE II. TABEL LABEL FITUR

Label
0
0
0
0

A. Gradient Boosted Decision Tree Biner

Dari jumlah total *dataset* dibagi menjadi 2 bagian, yaitu *train dataset* sebesar 60% dan *test dataset* sebesar 40% dengan *learning rate* sebesar 100 kali, *boosting type* menggunakan *dart*, *objective* menggunakan *multiclass*,

number leaves sebesar 100 buah, Dari hyperparameter tersebut menghasilkan model *decision tree* pada gambar (Fig 5).

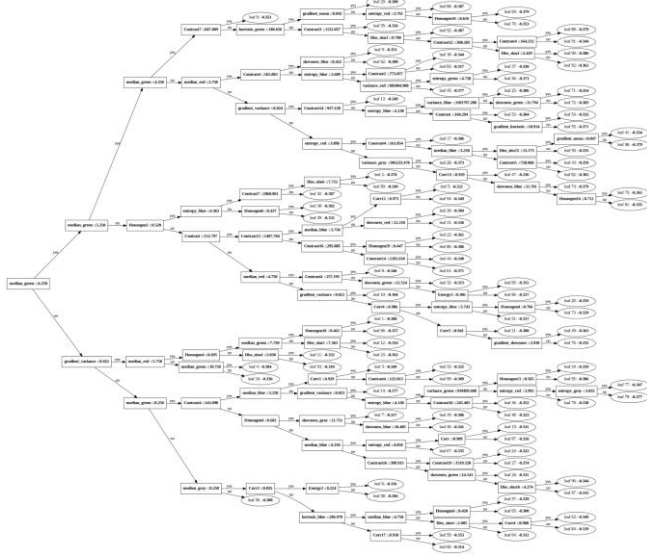


Fig. 5. Model *Decision Tree* yang dihasilkan dari 127 deskriptor fitur

Dari model tersebut dilakukan validasi menggunakan 40% data test menghasilkan *confusion matrix* pada gambar (Fig 6).

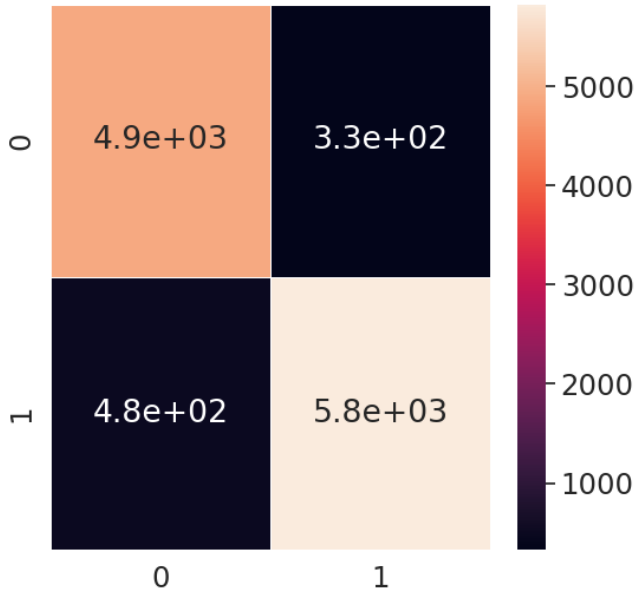


Fig. 6. \Confusion matrix dari model *Decision Tree* yang telah dibuat

Yang dimana terdapat nilai *True Positive* (TP) pada posisi matriks (0,0), *False Negative* (FN) pada posisi matriks (1,0), *False Positive* (FP) pada posisi matriks (0,1), dan *True Negative* (TN) pada posisi matriks (1,1). Dari *confusion matriks* tersebut dapat dihitung nilai presisi dan recall nya menggunakan persamaan (Eq 5).

$$Precision = \frac{TP}{(TP+FP)} \quad (Eq 5)$$

$$Precision = \frac{4900}{(4900 + 480)}$$

$$Precision = 0,91$$

Dari perhitungan presisi menggunakan persamaan di atas didapatkan bahwa nilai presisi model yang didapatkan sebesar 91%. Dan untuk dapat menghitung nilai *Recall* dengan menggunakan persamaan (Eq 6).

$$Recall = \frac{TP}{(TP+FN)} \quad (Eq 6)$$

$$Recall = \frac{4900}{(4900 + 330)}$$

$$Recall = 0.93$$

Dari perhitungan *Recall* menggunakan persamaan di atas didapatkan bahwa nilai *Recall* model yang didapatkan sebesar 93%. Sedangkan untuk mendapatkan nilai F1 dapat dihitung menggunakan persamaan (Eq 7).

$$F1l = 2 \times \frac{Precision \times Recall}{(Precision + Recall)} \quad (Eq 7)$$

$$F1 = 2 \times \frac{0,91 + 0,93}{(0,91 + 0,93)}$$

$$F1 = 0,91$$

Setelah melakukan perhitungan nilai F1 didapatkan nilai sebesar 0,91 atau dapat dikatakan bahwa model *decision tree* yang didapatkan memiliki akurasi sebesar 91%.

B. Artificial Neural Network

.Dari jumlah total *dataset* dibagi menjadi 2 bagian, yaitu *train dataset* sebesar 60% dan *test dataset* sebesar 40% dengan menggunakan *hidden layer* sebanyak 3 buah yang dimana pada tiap *hidden layer* secara berturut-turut memiliki neutron sebanyak 90, 50, dan 20 buah, sedangkan untuk *input layer* memiliki neutron sebanyak 127 buah, dan untuk *output layer* memiliki neutron sebanyak 2 buah. Sedangkan *hyperparameter* berupa *activation* dari *input layer* sampai *layer* terakhir sebelum *output layer* menggunakan *ReLU function* dan pada *layer output* menggunakan *Sigmoid function*. Serta terdapat *hyperparameter dropout* nilai *input* sebanyak 2 kali di antara *layer 1* dan *layer 2* sebesar 50% serta di antara *layer 2* dan *3* sebesar 25%. Selain itu terdapat *hyperparameter* berupa *optimizer adam* dan iterasi sebanyak 100 kali. Dari beberapa *hyperparameter* yang sudah diatur tadi menghasilkan model *Artificial Neural Network* pada gambar (Fig 7).

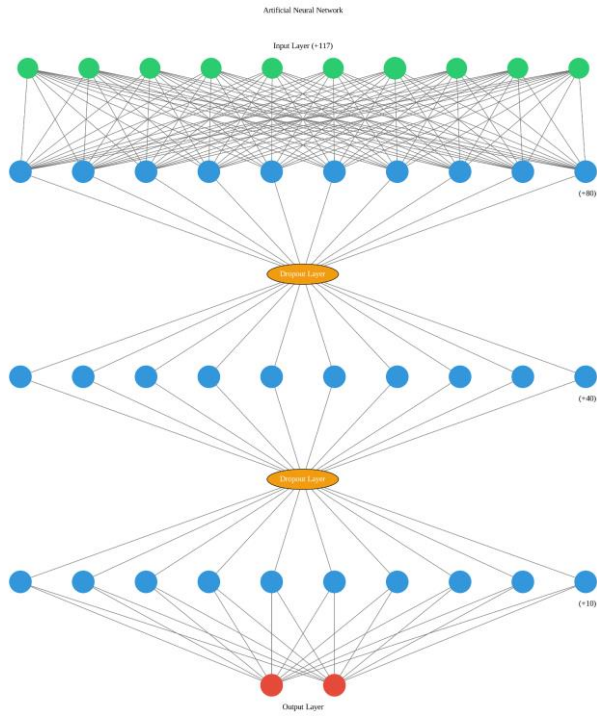


Fig. 7. Model ANN yang dihasilkan dari 127 deskriptor fitur sebagai *input* dan label 0 dan 1 sebagai *output* neuron

Dari model tersebut dilakukan validasi menggunakan 40% data test menghasilkan *confusion matrix* pada gambar (Fig 8).

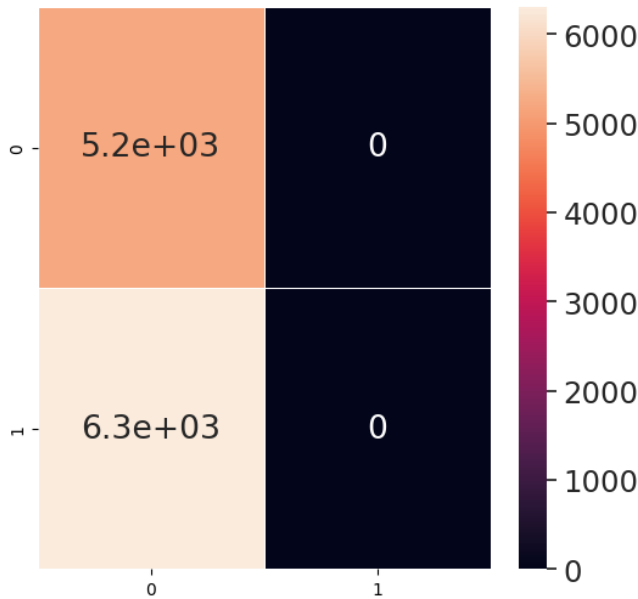


Fig. 8. *Confusion matrix* dari model ANN yang telah dibuat

Yang dimana terdapat nilai *True Positive* (TP) pada posisi matriks (0,0), *False Negative* (FN) pada posisi matriks (1,0), *False Positive* (FP) pada posisi matriks (0,1), dan *True Negative* (TN) pada posisi matriks (1,1). Dari *confusion matriks* tersebut dapat dihitung nilai presisi dan recall nya menggunakan persamaan (Eq 8).

$$Precision = \frac{TP}{(TP+FP)} \quad (Eq\ 8)$$

$$Precision = \frac{5200}{(5200 + 6300)}$$

$$Precision = 0,45$$

Dari perhitungan presisi menggunakan persamaan di atas didapatkan bahwa nilai presisi model yang didapatkan sebesar 91%. Dan untuk dapat menghitung nilai *Recall* dengan menggunakan persamaan (Eq 9).

$$Recall = \frac{TP}{(TP+FN)} \quad (Eq\ 9)$$

$$Recall = \frac{5200}{(5200 + 0)}$$

$$Recall = 1$$

Dari perhitungan *Recall* menggunakan persamaan di atas didapatkan bahwa nilai *Recall* model yang didapatkan sebesar 93%. Sedangkan untuk mendapatkan nilai F1 dapat dihitung menggunakan persamaan (Eq 10).

$$F1l = 2 \times \frac{Precision \times Recall}{(Precision + Recall)} \quad (Eq\ 10)$$

$$F1 = 2 \times \frac{0,45 + 1}{(0,45 + 1)}$$

$$F1 = 0,62$$

Setelah melakukan perhitungan nilai F1 didapatkan nilai sebesar 0,62 atau dapat dikatakan bahwa model *decision tree* yang didapatkan memiliki akurasi sebesar 62%.

Dari hasil analisis akurasi dari dua buah metode algoritma klasifikasi yang digunakan, menghasilkan akurasi pada model *Gradient Boosting Decision Tree* sebesar 91%, sedangkan pada model ANN dengan *hidden layer* berjumlah 3 menghasilkan akurasi sebesar 62%. Sehingga dapat dikatakan bahwa model *Artificial Neural Network* memiliki kelemahan dalam membuat model klasifikasi yang fiturnya diekstraksi secara manual, sedangkan jika pada metode *machine learning* tradisional sangat cocok digunakan dengan fitur data yang diekstraksi secara manual [21].

KESIMPULAN

Metode algoritma ANN tidak cocok digunakan untuk membuat model klasifikasi dari sebuah data set yang memiliki fitur diekstraksi secara manual, sehingga memiliki akurasi yang rendah sebesar 62%, selain itu sangat memungkinkan terdapat kesalahan dalam pemberian jumlah *hidden layer* serta *hyperparameter* yang diset. Sehingga dari penelitian ini dapat dikatakan bahwa metode algoritma yang cocok untuk mengklasifikasikan sebuah data yang memiliki banyak fitur data. Saya sangat membuka kritik dan saran terhadap metode penelitian yang saya buat ini .

REFERENCES

- [1] Mosby, Inc, Ed., *Mosby's dictionary of medicine, nursing & health professions*, Tenth edition. St.

Louis, Missouri: Elsevier, 2017.

- [2] R. Rosen, 'A relational theory of biological systems', *Bull. Math. Biophys.*, vol. 20, no. 3, pp. 245–260, Sep. 1958, doi: 10.1007/BF02478302.
- [3] B. Santelices, 'How many kinds of individual are there?', *Trends Ecol. Evol.*, vol. 14, no. 4, pp. 152–155, Apr. 1999, doi: 10.1016/S0169-5347(98)01519-5.
- [4] R. W. Piast, 'Shannon's information, Bernal's biopoiesis and Bernoulli distribution as pillars for building a definition of life', *J. Theor. Biol.*, vol. 470, pp. 101–107, Jun. 2019, doi: 10.1016/j.jtbi.2019.03.009.
- [5] Oxford University Press, Ed., *A dictionary of biology*, 6th ed. in Oxford paperback reference. Oxford ; New York: Oxford University Press, 2008.
- [6] E. G. Pringle, 'Orienting the Interaction Compass: Resource Availability as a Major Driver of Context Dependence', *PLoS Biol.*, vol. 14, no. 10, p. e2000891, Oct. 2016, doi: 10.1371/journal.pbio.2000891.
- [7] A. Swarnambiga, Ed., *Medical Image Processing for Improved Clinical Diagnosis*: in Advances in Medical Technologies and Clinical Practice. IGI Global, 2019. doi: 10.4018/978-1-5225-5876-7.
- [8] S. A. Lashari and R. Ibrahim, 'A Framework for Medical Images Classification Using Soft Set', *Procedia Technol.*, vol. 11, pp. 548–556, 2013, doi: 10.1016/j.protcy.2013.12.227.
- [9] B. M. Rashed and N. Popescu, 'Performance Investigation for Medical Image Evaluation and Diagnosis Using Machine-Learning and Deep-Learning Techniques', *Computation*, vol. 11, no. 3, p. 63, Mar. 2023, doi: 10.3390/computation11030063.
- [10] L. S. Garcia, 'Malaria', *Clin. Lab. Med.*, vol. 30, no. 1, pp. 93–129, Mar. 2010, doi: 10.1016/j.cll.2009.10.001.
- [11] S. Çelebi and M. Burak Çötel, 'Red and white blood cell classification using Artificial Neural Networks', *AIMS Bioeng.*, vol. 5, no. 3, pp. 179–191, 2018, doi: 10.3934/bioeng.2018.3.179.
- [12] S. Ziweritin, B. B. Baridam, and U. A. Okengwu, 'A Comparative Analysis of Neural Network and Decision Tree Model for Detecting Result Anomalies', *OALib*, vol. 09, no. 03, pp. 1–15, 2022, doi: 10.4236/oalib.1108549.
- [13] M. Sadiq and V. V. S. S. Balaram, 'DTBC: Decision tree based binary classification using with feature selection and optimization for malaria infected erythrocyte detection', *Int. J. Appl. Eng. Res.*, vol. 12, pp. 15923–15934, Jan. 2017.
- [14] D. Von Winterfeldt and W. Edwards, *Decision analysis and behavioral research*. Cambridge [Cambridgeshire] ; New York: Cambridge University Press, 1986.
- [15] L. Jiang, W. Wang, X. Yang, N. Xie, and Y. Cheng, 'Classification Methods of Remote Sensing Image Based on Decision Tree Technologies', in *Computer and Computing Technologies in Agriculture IV*, D. Li, Y. Liu, and Y. Chen, Eds., in IFIP Advances in Information and Communication Technology, vol. 344. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 353–358. doi: 10.1007/978-3-642-18333-1_41.
- [16] B. Hssina, A. Merbouha, H. Ezzikouri, and M. Erritali, 'A comparative study of decision tree ID3 and C4.5', *Int. J. Adv. Comput. Sci. Appl.*, vol. 4, no. 2, 2014, doi: 10.14569/SpecialIssue.2014.040203.
- [17] S. M. Pirayonesi and T. E. El-Diraby, 'Data Analytics in Asset Management: Cost-Effective Prediction of the Pavement Condition Index', *J. Infrastruct. Syst.*, vol. 26, no. 1, p. 04019036, Mar. 2020, doi: 10.1061/(ASCE)IS.1943-555X.0000512.
- [18] T. Hastie, R. Tibshirani, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, 2nd ed. in Springer series in statistics. New York, NY: Springer, 2009.
- [19] M. Puri, A. Solanki, T. Padawar, S. M. Tipparaju, W. A. Moreno, and Y. Pathak, 'Introduction to Artificial Neural Network (ANN) as a Predictive Tool for Drug Design, Discovery, Delivery, and Disposition', in *Artificial Neural Network for Drug Design, Delivery and Disposition*, Elsevier, 2016, pp. 3–13. doi: 10.1016/B978-0-12-801559-9.00001-6.
- [20] D. Nasution, T. H. F. Harumy, E. Haryanto, F. Fachrizal, Julham, and A. Turnip, 'A classification method for prediction of qualitative properties of multivariate EEG-P300 signals', in *2015 International Conference on Automation, Cognitive Science, Optics, Micro Electro-Mechanical System, and Information Technology (ICACOMIT)*, Bandung, Indonesia: IEEE, Oct. 2015, pp. 82–86. doi: 10.1109/ICACOMIT.2015.7440180.
- [21] C. N. Kamath, S. S. Bukhari, and A. Dengel, 'Comparative Study between Traditional Machine Learning and Deep Learning Approaches for Text Classification', in *Proceedings of the ACM Symposium on Document Engineering 2018*, Halifax NS Canada: ACM, Aug. 2018, pp. 1–11. doi: 10.1145/3209280.3209526.