Identifikasi Tipe Penyakit Anemia dengan Menggunakan Metode *Random Forest*

Laili Nur Farida[1], Dian Candra Rini Novitasari[1*], Wika Dianita Utami[1]

[1]Department of Mathematics, UIN Sunan Ampel Surabaya, Surabaya, Indonesia.
Email:diancrini@uinsa.ac.id*)

***Abstrak*—Anemia merupakan kelainan darah yang disebabkan karena kurangnya sel darah merah dan kadar hemoglobin. Berdasarkan survei yang dilakukan oleh *World Health Organization* (WHO), anemia memengaruhi 1.62 miliar orang di seluruh dunia. Sebanyak 93% pasien penyakit kronis mengalami anemia. Anemia dapat didiagnosis menggunakan *Complete Blood Count* (CBC) yang bertujuan untuk mengevaluasi jumlah total dan karakteristik komponen sel dalam darah. Tujuan dari penelitian adalah untuk melakukan klasifikasi *Random Forest* (RF) terhadap anemia. Penelitian ini menggunakan data numerik berupa karakteristik sel darah, yang terdiri dari 269 pasien *Normocytic Normochromic An*emia, 189 pasien *Iron Deficiency Anemia*, *mia*dan 336 pasien *Healthy*. Pada penelitian ini proses klasifikasi menggunakan RF yang kemudian dilakukan evaluasi hasil menggunakan *Confusion Matrix*, sehingga diperoleh hasil evaluasi klasifikasi berupa akurasi, sensitivitas, dan spesifisitas. Penelitian ini memperoleh hasil terbaik pada k sejumlah 5 dengan parameter *n estimators*, *max depth*, dan *min samples leaf* berturut-turut sebanyak 10, 90, dan 4. Nilai akurasi, sensitivitas, dan spesifisitas yang dihasilkan masing-masing sebesar 100%.**

***Keywords*— Anemia, *Iron Deficiency Anemia,* Klasifikasi, *Normocytic Normochromic An*emia, Random Forest**

***Abstract*— Anemia is a blood disorder caused by a lack of red blood cells and hemoglobin levels. Based on a survey conducted by the World Health Organization (WHO), anemia affects 1.62 billion people worldwide. As many as 93% of chronic disease patients experience anemia. Anemia can be diagnosed using Complete Blood Count (CBC) which aims to evaluate the total number and characteristics of cell components in the blood. The purpose of this study was to perform Random Forest (RF) classification of anemia. This study used numerical data in the form of blood cell characteristics, consisting of 269 Normocytic Normochromic Anemia patients, 189 Iron Deficiency Anemia patients, and 336 Healthy patients. In this study, the classification process used RF which was then evaluated using the Confusion Matrix, so that the classification evaluation results were obtained in the form of accuracy, sensitivity, and specificity. This study obtained the best results at k of 5 with parameters n estimators, max depth, and min samples leaf of 10, 90, and 4, respectively. The accuracy, sensitivity, and specificity values produced were each 100%.**

***Kata Kunci*— Anemia, *Iron Deficiency Anemia,* Classification, *Normocytic Normochromic An*emia, Random Forest**

I. INTRODUCTION

Anemia merupakan kelainan darah yang disebabkan oleh kurangnya sel darah merah dan hemoglobin. Berdasarkan survei yang dilakukan oleh *World Health Organization* (WHO), anemia diderita oleh 1.62 miliar orang di seluruh dunia [1]. Terdapat beberapa jenis anemia, yaitu anemia yang disebabkan oleh gangguan metabolisme zat besi dan penyakit kronis. Berdasarkan hal tersebut, anemia yang sering diderita adalah *Normocytic Normochromic Anemia* (NNA) dan *Iron Deficiency Anemia* (IDA) [2]. Sebanyak 93% pasien penyakit kronis mengalami anemia, dengan IDA sebagai penyebab yang paling umum yaitu 81.5% [3]. Sebanyak 37.37% (296 dari 792) pasien anemia mengalami NNA [4]. Berdasarkan banyaknya jumlah penderita IDA dan NNA, maka dibutuhkan suatu sistem untuk mendiagnosis lebih awal, serta mengurangi jumlah kasus dari anemia.

IDA dan NNA dapat didiagnosis menggunakan *Complete Blood Count* (CBC). CBC merupakan salah satu tes darah yang paling umum yang memberikan banyak informasi mengenai kesehatan seseorang. Pemeriksaan lebih lanjut di laboratorium atau klinik diperlukan untuk memahami hasil tes ini dengan tepat [5]. Tes ini bertujuan untuk mengevaluasi jumlah total dan karakteristik komponen sel dalam darah. Komponen tersebut, diantaranya *White Blood Cells* (WBC), *Red Blood Cells* (RBC), *Platelets* (PLT), *Hemoglobin* (HGB), *Hematocrit* (HCT), *Mean Corpuscular Volume* (MCV), *Mean Corpuscular Hemoglobin* (MCH), *Mean Corpuscular Hemoglobin Concentration* (MCHC), *Plateletcrit* (PCT), *Platelet Distribution Width* (PDW), *Lymphocytes Percentage* (LP), *Neutrophils Percentag*e (NP), *Lymphocytes Absolute Count* (LAC), dan *Neutrophils Absolute Count* (NAC) [6].

Variabel-variabel yang digunakan dalam diagnosis anemia menjadi faktor penentu dalam hasil diagnosa. Oleh karena itu, diupayakan mengurangi adanya kesalahan diagnosis pada anemia. Salah satu cara untuk meminimalisir adanya kekeliruan diagnosis yaitu menggunakan bantuan *Computer Aided Diagnosis* (CAD). Salah satu bagian dari CAD adalah klasifikasi anemia menggunakan *Machine Learning,* yaitu metode *Random Forest* (RF). RF merupakan suatu metode yang terdiri atas beberapa

Decision Tree (DT), dimana dalam membangun DT tersebut menggunakan *bootstrap*. *Bootstrap* merupakan suatu teknik yang digunakan untuk memperkirakan atau melakukan prediksi kuantitas dari sampel data [7]. Metode RF menggunakan beberapa parameter untuk melakukan prediksi, diantaranya yaitu jumlah pohon (*n estimators*), kedalaman maksimal pohon (*max depth*), dan jumlah minimal sampel daun (*min samples leaf*). Kedalaman maksimal pohon (*max depth*) adalah seberapa banyak lapisan simpul yang dapat dibentuk oleh pohon dari simpul akar hingga simpul daun, sedangkan jumlah minimal sampel daun (*min samples leaf*) adalah jumlah minimal data yang harus ada di setiap simpul daun. Metode ini memiliki kelebihan dapat mengurangi *overfitting* dan menangani data yang memiliki dimensi tinggi [8].

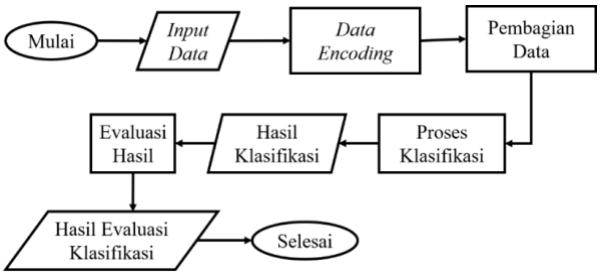
Berdasarkan kelebihan metode RF, terdapat beberapa penelitian yang menggunakan metode ini untuk melakukan klasifikasi. Penelitian terkait klasifikasi *Autism Spectrum Disorder* (ASD) menggunakan metode RF dengan 21 variabel menghasilkan akurasi sebesar 97.20% [9]. Penelitian yang melakukan klasifikasi *normocytic*, *microcytic*, dan *macrocytic anemia* berdasarkan tingkat keparahan menggunakan DT, *Logistic Regression* (LR), *Multi-Layer Perceptron* (MLP), NB, RF, dan *Support Vector Machines* (SVM). Hasil akurasi tertiggi diperoleh pada metode MLP, yaitu sebesar 99.35% [10]. Penelitian lain yang membandingkan metode RF dan MLP untuk klasifikasi penyakit alzheimer menunjukkan bahwa metode RF memiliki kinerja lebih baik daripada MLP dengan hasil akurasi sebesar 90.2% [11]. Penelitian yang melakukan klasifikasi anemia 2 kelas dengan membandingkan metode J48, RF, dan *Hoeffding Tree*. Penelitian ini memperoleh hasil terbaik pada metode RF dengan nilai akurasi sebesar 97.57% [12].

Penelitian terkait klasifikasi anemia menggunakan metode DT, SVM, dan *K-Nearest Neighbors* (KNN). Metode yang menghasilkan akurasi tertinggi adalah metode DT, yaitu sebesar 82.61% [13]. Penelitian lain melakukan klasifikasi IDA dan 11 jenis anemia lainnya menggunakan metode *Ensemble* DT, *Artificial Neural Networks*, SVM, dan *Naive Bayes* (NB). Penelitian ini menghasilkan akurasi tertinggi pada *Bagged* DT, yaitu sebesar 85.6% [14]. Namun, metode DT ini memiliki kelemahan, yaitu rentan terhadap *overfitting* [15]. *Overfitting* merupakan suatu kondisi ketika sebuah model *Machine Learning* (ML) terlalu baik dalam memahami data pelatihan, sehingga mempelajari *noise* yang tidak penting. *Overfitting* menyebabkan sebuah model tidak dapat melakukan prediksi yang akurat pada data baru [16]. Penanganan *overfitting* dapat dilakukan menggunakan metode RF [8]. Cara RF mengurangi *overfitting* dan menangani data yang memiliki dimensi tinggi dengan teknik *bootstrap* dan *majority voting* [17].

Berdasarkan beberapa kelebihan yang dimiliki, metode RF mampu memperoleh hasil akurasi yang lebih tinggi dibandingkan dengan metode lain, seperti pada penelitian yang membandingkan metode RF dengan SVM, KNN, *Backpopopagation*, dan *Deep Learning* menggunakan 21 variabel pada klasifikasi ASD menghasilkan sensitivitas masing-masing sebesar 100%, 99.33%, 82.11%, 99.33%, dan 94.03% [7]. Penelitian yang melakukan analisis sentimen terhadap penggunaan ahli anestesi Brasil dengan membandingkan metode RF dan MLP. Hasil penelitian ini menunjukkan bahwa metode RF memiliki kinerja lebih baik daripada MLP dengan hasil akurasi masing-masing sebesar 96.42% dan 94.44% [18]. Pada penelitian terkait klasifikasi bintil paru, akurasi yang dicapai RF merupakan yang tertinggi dibandingkan metode klasifikasi ML lainnya seperti KNN, SVM, dan DT dengan masing-masing sebesar 90.73%, 84.51%, 87.10%, dan 87.07% [19]. Berdasarkan pemaparan sebelumnya, bahwa anemia merupakan kelainan darah yang disebabkan karena kurangnya sel darah merah dan kadar hemoglobin. Penyakit ini dapat mengakibatkan berbagai komplikasi kesehatan yang serius apabila terlambat untuk didiagnosis. Oleh karena itu, perlu dilakukan diagnosis secara dini terkait anemia agar melakukan penanganan terhadap penyakit anemia.

II. METODOLOGI

Penelitian ini menggunakan data sekunder berupa karakteristik anemia yang diperoleh dari website Kaggle [20]. Penelitian ini menggunakan 794 data yang terdiri dari 269 pasien NNA, 189 pasien IDA, dan 336 pasien Healthy (HE). Penelitian ini melalui beberapa prosedur pengolahan data. Langkah langkah yang dilakukan di tunjukkan pada Gbr. 1.



Gbr. 1. Diagram Alir Penelitian

Langkah awal dalam klasifikasi anemia adalah menentukan data *input* berupa karakteristik sel-sel darah. Data karakteristik sel-sel darah ditunjukkan pada TABEL 1.

TABEL 1. DATA KARAKTERISTIK KLINIS PPOK

Variabel	Keterangan	*Range* Data
ID	Letak baris suatu data	$1 \leq x \leq 794$
A	Sel darah putih yang berperan dalam melawan infeksi (WBC)	$0.8 \leq x \leq 45.7$
B	Persentase limfosit dari total sel darah putih (LP)	$6.9 \leq x \leq 91.4$
C	Persentase neutrofil dari total sel darah putih (NP)	$7.6 \leq x \leq 671$
D	Jumlah absolut limfosit per unit volume darah (LAC)	$0.4 \leq x \leq 41.8$
E	Jumlah absolut neutrofil per unit volume darah (NAC)	$1.2 \leq x \leq 44$
F	Sel darah merah yang mengangkut oksigen dari paru-paru ke seluruh tubuh dan membawa karbon dioksida kembali ke paru-paru (RBC)	$1.42 \leq x \leq 6.9$

Variabel	Keterangan	*Range* Data
G	Protein dalam sel darah merah yang membawa oksigen dari paru-paru ke seluruh tubuh (HGB)	$-10 \leq x \leq 16.2$
H	Ukuran persentase volume darah yang terdiri dari sel darah merah (HCT)	$10.1\ x \leq 741$
I	Rata-rata volume sel darah merah (MCV)	$-79.3\ x \leq 111.2$
J	Rata-rata jumlah hemoglobin per sel darah merah (MCH)	$11.4\ x\ 3117$
K	Rata-rata konsentrasi hemoglobin dalam sel darah merah (MCHC)	$20.3 \leq x \leq 92.8$
L	Sel darah yang membantu dalam pembekuan darah (PLT)	$24 \leq x \leq 534$
M	Ukuran variasi dalam ukuran trombosit (PDW)	$8.4 \leq x\ 29.2$
N	Persentase volume darah yang terdiri dari trombosit (PCT)	$0.03\ x \leq 0.44$
O	Jenis anemia	NNA, IDA, dan HE

Data pada TABEL 1 menunjukkan adanya data kategorik pada variabel O, sehingga harus dilakukan *preprocessing* yaitu berupa *data encoding*. Setelah melakukan *preprocessing*, data dibagi menjadi data pelatihan dan data pengujian menggunakan *K-Fold Cross Validation* dengan jumlah fold sebanyak 5 fold. Setelah itu, data melakukan proses klasifikasi berupa pelatihan dan pengujian menggunakan metode RF terlebih dahulu untuk *tuning parameter* berupa *n estimators* sebanyak 10, 60, dan 110, *max depth* sebanyak 30, 60, dan 90, serta *min samples leaf* sebanyak 4, 8, dan 12. Proses klasifikasi menggunakan RF berdasarkan Persamaan. Langkah terakhir yaitu melakukan evaluasi hasil klasifikasi menggunakan *Confusion Matrix multi class* untuk mengetahui jumlah TP, TN, FP, dan FN. *Confusion Matrix multi class* tersebut digunakan untuk memperoleh akurasi, sensitivitas, dan spesifisitas yang terbaik dari hasil klasifikasi RF.

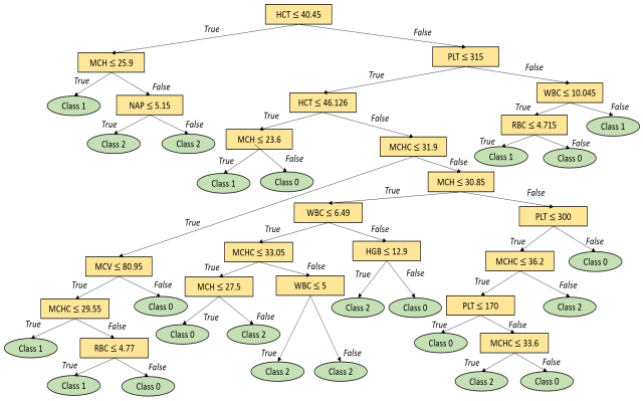
III. HASIL DAN PEMBAHASAN

Penelitian ini memperoleh hasil dan pembahasan berupa hasil uji coba parameter RF, pohon RF, hasil evaluasi klasifikasi menggunakan *Confusion Matrix*, dan perbandingan hasil penelitian pada klasifikasi anemia. Pada tahap awal penelitian ini adalah melakukan input data. TABEL 2 menunjukkan contoh data karakteristik sel-sel darah. Pada TABEL 2 menunjukkan adanya tipe data kategorik, sehingga proses *data encoding* perlu dilakukan. Hasil *data encoding* disajikan pada TABEL 2.

Langkah selanjutnya yaitu proses pelatihan dan pengujian menggunakan metode klasifikasi RF. Namun, sebelum memulai tahap klasifikasi, data terlebih dahulu dilakukan pembagian data pelatihan dan data pengujian melalui penerapan metode *K-Fold Cross Validation* dengan jumlah $k = 5$. Setelah melakukan pembagian data, proses selanjutnya yaitu melakukan klasifikasi menggunakan RF. Klasifikasi RF memperoleh keputusan kelas dari DT. Visualisasi DT pertama ditunjukkan pada Gbr. 2.

TABEL 2. HASIL DATA ENCODING

ID	A	B	C	D	E	F	G	H	I	J	⋯	O
1	7.2	30.7	60.7	2.2	4.4	4	9	30.5	77	22.6	⋯	1
2	6	30.2	63.5	1.8	3.8	4.2	3.8	32.8	77.9	23.2	⋯	1
3	5.2	19.7	72.4	1	3.8	4.9	13.2	41	84.7	27.2	⋯	0
4	5.1	31.3	60.6	1.6	3.1	5.6	14.6	45.6	81	25.9	⋯	0
5	4.5	43.4	51.2	2	2.3	5.4	13.5	44	82.4	25.7	⋯	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	…	⋮
794	8.8	25.8	77.5	1.9	5.1	4.9	15.2	46.2	89.7	30.6	⋯	0



Gbr. 2. DT ke-1

Berdasarkan Gbr. 2, simpul akar pada DT pertama adalah variabel HCT dengan nilai threshold cabang kiri HCT ≤ 40.45 dan nilai *threshold* cabang kanan HCT > 40.45. Selanjutnya, pada simpul cabang MCH > 25.9, dan NAP > 5.15 terklasifikasi kelas 2. Sedangkan pada simpul cabang MCH > 25.9, dan NAP > 5.15 terklasifikasi kelas 0.

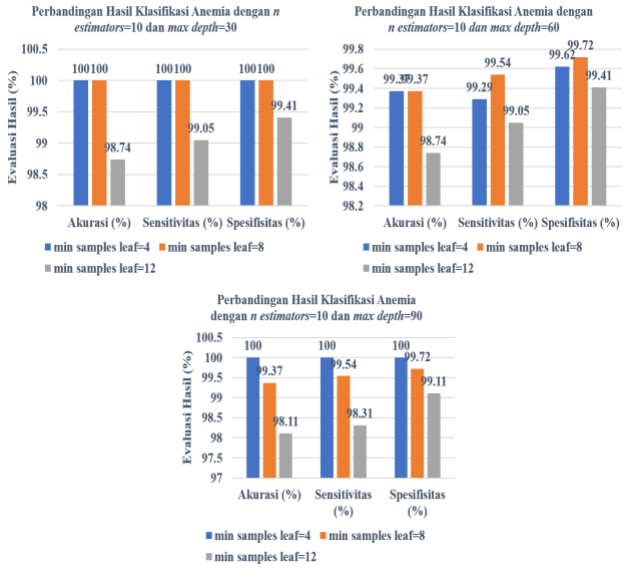
Setelah pembentukan DT, klasifikasi RF memperoleh hasil uji coba untuk tu*ning parameter* terbaik dengan parameter *n estimators*, *max depth*, dan *min samples leaf*. Hasil dari uji coba ditunjukkan pada TABEL 3.

TABEL 3. HASIL UJI COBA

k-fold	*n estimators*	*max depth*	*min samples leaf*	Accuracy (%)	Sensitivity (%)	Specificity (%)	Time (s)
5	10	30	4	100	100	100	0.06
			8	100	100	100	0.03
			12	98.74	99.05	99.41	0.02
		60	4	99.37	99.29	99.62	0.02
			8	99.37	99.54	99.72	0.02
			12	98.74	99.05	99.41	0.02
		90	4	100	100	100	0.02
			8	99.37	99.54	99.72	0.02
			12	98.11	98.31	99.11	0.03
	60	30	4	100	100	100	0.10
			8	100	100	100	0.10
			12	100	100	100	0.10
		60	4	100	100	100	0.10
			8	100	100	100	0.10
			12	99.37	99.44	99.67	0.10
		90	4	100	100	100	0.10
			8	99.37	99.54	99.72	0.10
			12	100	100	100	0.10
	110	30	4	100	100	100	0.19
			8	100	100	100	0.19
			12	99.37	99.44	99.67	0.18
		60	4	100	100	100	0.18
			8	100	100	100	0.18
			12	99.37	99.54	99.72	0.19
		90	4	100	100	100	0.18
			8	100	100	100	0.18
			12	99.37	99.54	99.72	0.18

Berdasarkan data pada TABEL 3, pada *n estimators*=10 akurasi tertinggi yang dicapai sebesar 100%

dalam waktu 0.02 s dengan *max depth* dan *min samples leaf* masing-masing senilai 90 dan 4. Waktu pelatihan yang diperoleh pada *n estimators* = 10 menyatakan bahwa semakin besar *max depth* dan semakin kecil *min samples leaf*, maka akurasi semakin tinggi serta waktu pelatihan menjadi lebih cepat. Pada *n estimators* = 60 akurasi tertinggi yang dicapai sebesar 100% dalam 59 waktu 0.10 s dengan *max depth*=30 dan *min samples leaf* = 4. Waktu pelatihan yang diperoleh pada *n estimators* = 60 menyatakan bahwa semakin kecil nilai *max depth* dan *min samples leaf* = 4, maka akurasi juga semakin tinggi. Sebaliknya, nilai *min samples leaf* pada *n estimators* = 60 tidak mempengaruhi waktu pelatihan. Pada *n estimators* = 110 akurasi tertinggi yang dicapai sebesar 100% dalam waktu 0.18 s dengan *max depth* senilai 60 dan 90, serta *min samples leaf* senilai 4 dan 8. Waktu pelatihan yang diperoleh pada *n estimators* = 110 menyatakan bahwa semakin besar *max depth* dan semakin kecil *min samples leaf*, maka waktu pelatihan lebih cepat serta akurasi juga semakin tinggi. Oleh karena itu, parameter terbaik dari hasil uji coba diperoleh pada *n estimators*, *max depth*, dan *min samples leaf* masing-masing senilai 10, 90, dan 4. Parameter terbaik menghasilkan akurasi, sensitivitas, dan spesifisitas sebesar 100% serta waktu pelatihan selama 0.02 s. Berdasarkan hasil uji coba yang ditunjukkan pada TABEL 3, grafik parameter uji coba ditampilkan pada Gbr. 3.



Gbr. 3. Hasil Perbandingan Akurasi, Sensitivitas, dan Spesifisitas Berdasarkan N Estimators 10 dengan Max Depth 30, 60, dan 90

Gbr. 3 menunjukkan diagram batang dari hasil uji coba dengan *max depth* senilai 30, 60, dan 90, sedangkan *min samples leaf* sejumlah 4, 8, serta 12. Pada *max depth* senilai 30 menunjukkan hasil akurasi, sensitivitas, dan spesifisitas tertinggi berupa 100%, yang diperoleh pada *min samples leaf* senilai 4 dan 8. Pada *max depth* senilai 60 menunjukkan hasil akurasi, sensitivitas, dan spesifisitas tertinggi masing-masing berupa 99.37%, 99.54%, dan 99.72% yang diperoleh pada *min samples leaf* senilai 8. Pada *max depth* senilai 90 menunjukkan hasil akurasi, sensitivitas, dan spesifisitas tertinggi berupa 100%, yang diperoleh pada *min samples leaf* senilai 4. Berdasarkan Gbr. 3, hasil akurasi, sensitivitas, dan spesifisitas meningkat apabila nilai *min samples leaf* semakin kecil dan nilai *max depth* semakin besar. Berdasarkan hasil terbaik dari proses pengujian, *Confusion Matrix* ditampilkan pada TABEL 4.

TABEL 4. HASIL CONFUSION MATRIX MULTI CLASS

Aktual \ Prediksi	0	1	2
0	68	0	0
1	0	34	0
2	0	0	55

(a)

Aktual \ Prediksi	0	1	2
0	68	0	0
1	0	34	0
2	0	0	55

(b)

Aktual \ Prediksi	0	1	2
0	68	0	0
1	0	34	0
2	0	0	55

(c)

TABEL 4 menunjukkan nilai TP, FN, FP, dan TN. Berdasarkan TABEL 4 (a), pada kelas 0 jumlah TP, TN, FP, dan FN masing-masing sebanyak 68, 89, 0, dan 0. Berdasarkan TABEL 4 (b), pada kelas 1 jumlah TP, TN, FP, dan FN masing-masing sebanyak 34, 123, 0, dan 0. Berdasarkan TABEL 4 (c), pada kelas 2 jumlah TP, TN, FP, dan FN masing-masing sebanyak 55, 102, 0, dan 0.

Hasil klasifikasi anemia pada penelitian ini menghasilkan akurasi, sensitivitas, dan spesifisitas yang sangat baik. Penelitian ini membuktikan bahwa metode RF sangat efektif dalam melakukan klasifikasi anemia apabila dibandingkan penelitian lain. Perbandingan hasil penelitian ini dengan penelitian lainnya terkait klasifikasi anemia fitunjukkan pada TABEL 5.

TABEL 5. PERBANDINGAN HASIL PENELITIAN TERKAIT KLASIFIKASI ANEMIA

Methods	Number of Variables	Number of Classes	Accuracy (%)
MLP [10]	11	3 (*mild*, *moderate* and *severe anemia*)	99.35
DT [13]	3	2 (anemia and nonanemia)	82.61
Bagged DT [14]	25	12 (anemia, nonanemia, IDA, IFDA, IVBDA, FDA, FVBDA, HA, ACD, Thal, Thal *Trait*, and VBDA	85.60
RF [12]	9	2 (anemia and nonanemia)	97.57
RF (Penelitian ini)	14	3 (NNA, IDA, and HE)	100

Berdasarkan data pada TABEL 5, penelitian lain menggunakan metode MLP dan DT untuk klasifikasi anemia, di mana hasil akurasi yang diperoleh lebih rendah dibandingkan dengan penelitian ini. Akurasi yang tinggi pada penelitian ini disebabkan oleh metode RF yang memiliki kelebihan dapat menangani *overfitting*, yang tidak dimiliki oleh MLP dan DT. Sebaliknya, kedua metode tersebut justru rentan terhadap *overfitting*. Oleh karena itu, penelitian ini menghasilkan akurasi lebih tinggi dibandingkan dengan penelitian lain pada klasifikasi anemia.

IV. KESIMPULAN

Berdasarkan hasil dari penerapan metode *Random Forest*, diperoleh kesimpulan bahwa skurasi tertinggi yang dicapai pada uji coba adalah 100% pada k=5 dengan

parameter *n estimators*=10, *max depth*=90, dan *min samples leaf* =4 dalam waktu 0.02s. Nilai TP, TN, FP, dan FN yang diperoleh pada kelas HE berturut-turut yaitu 62, 96, 0, dan 0. Pada kelas IDA berturut-turut yaitu 36, 122 0, dan 0. Sedangkan pada kelas NNA berturut-turut yaitu 60, 98, 0, dan 0. Oleh karena itu, dapat disimpulkan bahwa semakin rendah nilai *n estimators*, *max depth*, dan *min samples leaf*, maka waktu pelatihan RF juga lebih cepat.

DAFTAR PUSTAKA

[1] N. K.T, K. Prasad, and B. M. K. Singh, "Analysis of Red Blood Cells from Peripheral Blood Smear Images for Anemia Detection: A Methodological Review," *Med. Biol. Eng. Comput.*, vol. 60, no. 9, pp. 2445–2462, 2022, doi: 10.1007/s11517-022-02614-z.

[2] A. S. Satniyazovna, "Anemia is A Chronic Disease," *Am. J. Med. Sci. Pharm. Res.*, vol. 04, no. 09, pp. 12–19, 2022, doi: https://doi.org/10.37547/TAJMSPR/Volume04Issue09-03.

[3] V. Talarico, L. Giancotti, G. A. Mazza, R. Miniero, and M. Bertini, "Iron Deficiency Anemia in Celiac Disease," *Nutrients*, vol. 13, no. 5, pp. 1–11, 2021, doi: 10.3390/nu13051695.

[4] V. Pandey, "Study of Spectrum of Anemia in Various Age Groups," *IP Arch. Cytol. Histopathol. Res.*, vol. 6, no. 3, pp. 197–202, 2021, doi: 10.18231/j.achr.2021.044.

[5] L. Agnello *et al.*, "The Value of a Complete Blood Count (CBC) for Sepsis Diagnosis and Prognosis," *Diagnostics*, vol. 11, no. 10, pp. 1–19, 2021, doi: 10.3390/diagnostics11101881.

[6] I. H. Seo and Y. J. Lee, "Usefulness of Complete Blood Count (CBC) to Assess Cardiovascular and Metabolic Diseases in Clinical Settings: A Comprehensive Literature Review," *Biomedicines*, vol. 10, no. 11, 2022, doi: 10.3390/biomedicines10112697.

[7] E. S. Dewi and E. M. Imah, "Comparison of Machine Learning Algorithms for Autism Spectrum Disorder Classification," *Int. Jt. Conf. Sci. Eng. (IJCSE 2020) Comp.*, vol. 196, no. Ijcse, pp. 152–159, 2020, doi: 10.2991/aer.k.201124.028.

[8] J. Quist, L. Taylor, J. Staaf, and A. Grigoriadis, "Random Forest Modelling of High-Dimensional Mixed-Type Data for Breast Cancer Classification," *Cancers (Basel).*, vol. 13, no. 5, pp. 1–15, 2021, doi: 10.3390/cancers13050991.

[9] R. Sujatha, S. L. Aarthy, J. M. Chatterjee, A. Alaboudi, and N. Z. Jhanjhi, "A Machine Learning Way to Classify Autism Spectrum Disorder," *Int. J. Emerg. Technol. Learn.*, vol. 16, no. 6, pp. 182–200, 2021, doi: 10.3991/ijet.v16i06.19559.

[10] R. Vohra, A. Hussain, A. K. Dudyala, J. Pahareeya, and W. Khan, "Multi-Class Classification Algorithms for the Diagnosis of Anemia in an Outpatient Clinical Setting," *PLoS One*, vol. 17, no. 7 July, pp. 1–18, 2022, doi: 10.1371/journal.pone.0269685.

[11] M. Song, H. Jung, S. Lee, D. Kim, and M. Ahn, "Diagnostic Classification and Biomarker Identification of Alzheimer's Disease with Random Forest Algorithm," *Brain Sci.*, vol. 11, no. 4, pp. 1–14, 2021, doi: 10.3390/brainsci11040453.

[12] N. Sasikala, G. R. Banu, T. Babiker, and P. Rajpoot, "A Role of Data Mining Techniques to Predict Anemia Disease," *Int. J. Comput. Appl.*, vol. 174, no. 20, pp. 16–20, 2021, doi: 10.5120/ijca2021921090.

[13] N. Bin Noor, M. S. Anwar, and M. Dey, "Comparative Study between Decision Tree, SVM and KNN to Predict Anaemic Condition," *BECITHCON 2019 - 2019 IEEE Int. Conf. Biomed. Eng. Comput. Inf. Technol. Heal.*, pp. 24–28, 2019, doi: 10.1109/BECITHCON48839.2019.9063188.

[14] T. Karagül Yıldız, N. Yurtay, and B. Öne, "Classifying Anemia Types using Artificial Learning Methods," *Eng. Sci. Technol. an Int. J.*, vol. 24, no. 1, pp. 50–70, 2021, doi: 10.1016/j.jestch.2020.12.003.

[15] M. Al-akhras, K. El Hindi, M. Habib, and B. Abu, "Instance Reduction for Avoiding Overfitting in Decision Trees," *De Gruyter*, vol. 30, no. 1, pp. 438–459, 2021.

[16] J. G. Greener, S. M. Kandathil, L. Moffat, and D. T. Jones, "A Guide to Machine Learning for Biologists," *Nat. Rev. Mol. Cell Biol.*, vol. 23, pp. 40–55, 2022, doi: 10.1007/978-3-031-22206-1_11.

[17] X. Zhou, P. Lu, Z. Zheng, D. Tolliver, and A. Keramati, "Accident Prediction Accuracy Assessment for Highway-Rail Grade Crossings Using Random Forest Algorithm Compared with Decision Tree," *Reliab. Eng. Syst. Saf.*, p. 106931, 2020, doi: 10.1016/j.ress.2020.106931.

[18] J. Asian, M. Dholah Rosita, and T. Mantoro, "Sentiment Analysis for the Brazilian Anesthesiologist Using Multi-Layer Perceptron Classifier and Random Forest Methods," *J. Online Inform.*, vol. 7, no. 1, p. 132, 2022, doi: 10.15575/join.v7i1.900.

[19] N. S. El-Askary, M. A. M. Salem, and M. I. Roushdy, "Feature Extraction and Analysis for Lung Nodule Classification using Random Forest," *ACM Int. Conf. Proceeding Ser.*, pp. 248–252, 2019, doi: 10.1145/3328833.3328872.

[20] E. Aboelnaga, "Anemia Types Classification," *Kaggle*, 2024. https://www.kaggle.com/datasets/ehababoelnaga/anemia-types-classification