BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 5, August 2026 | Hal 2049-2058
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i5.1215

Klasifikasi Anemia pada Data Tidak Seimbang Menggunakan Random Forest, SMOTETomek, dan GridSearchCV

M Alfathan Haris*, Solikhun

[1] Sistem Informasi, Sekolah Tinggi Ilmu Komputer Tunas Bangsa, Pematangsiantar, Indonesia
[2] Teknik Informatika, Sekolah Tinggi Ilmu Komputer Tunas Bangsa, Pematangsiantar, Indonesia
Email: [1,*]alfathanharis07@gmail.com, [2]solikhun@amiktunasbangsa.ac.id
Email Penulis Korespondensi: alfathanharis07@gmail.com

Abstrak−Penelitian ini memanfaatkan Machine Learning sebagai dukungan awal untuk mengenali anemia berdasarkan informasi warna citra dan kadar hemoglobin. Permasalahan utama pada data yang digunakan adalah ketimpangan jumlah kelas, karena data normal lebih banyak dibandingkan data anemia sehingga model berpeluang mempelajari kelas mayoritas secara dominan. Untuk mengatasi kondisi tersebut, penelitian ini menerapkan Random Forest yang dipadukan dengan SMOTETomek dan GridSearchCV. Dataset Anemia Diagnosis dari Kaggle berisi 104 data dengan empat fitur numerik, yaitu %Red Pixel, %Green Pixel, %Blue Pixel, dan Hb, serta satu target klasifikasi anemia. Komposisi awal data terdiri atas 78 data normal dan 26 data anemia. SMOTETomek digunakan pada data latih untuk menyeimbangkan kelas melalui pembentukan sampel sintetis dan pembersihan Tomek Links, sedangkan GridSearchCV digunakan untuk memperoleh hyperparameter Random Forest yang paling sesuai. Evaluasi dilakukan menggunakan validasi silang 10-Fold dan Wilcoxon Signed-Rank Test. Model usulan memperoleh accuracy 97.09%, precision 97.50%, recall 91.67%, dan F1-score 93.24%. Nilai p-value Wilcoxon sebesar 0.03125 pada accuracy dan F1-score menunjukkan perbedaan signifikan pada taraf 0.05. Hasil ini memperlihatkan bahwa integrasi Random Forest, SMOTETomek, dan GridSearchCV mampu meningkatkan klasifikasi anemia pada dataset tidak seimbang.

Kata Kunci: Anemia Diagnosis; Class Imbalance; GridSearchCV; Random Forest; SMOTETomek

Abstract−This study applies a Machine Learning approach to support preliminary anemia identification using image-color information and hemoglobin levels. The main issue in the dataset is an imbalanced class composition, where normal records are more numerous than anemia records and may cause the classifier to learn the majority-class pattern more strongly. To address this issue, Random Forest is integrated with SMOTETomek and GridSearchCV. The Kaggle Anemia Diagnosis dataset contains 104 records, four numerical attributes (%Red Pixel, %Green Pixel, %Blue Pixel, and Hb), and one anemia classification target. The original class distribution consists of 78 normal records and 26 anemia records. SMOTETomek is applied to the training data to balance the classes by generating synthetic minority samples and removing Tomek Links, while GridSearchCV searches for the most appropriate Random Forest hyperparameters. Model performance is assessed using 10-Fold Cross Validation and the Wilcoxon Signed-Rank Test. The proposed model obtains 97.09% accuracy, 97.50% precision, 91.67% recall, and 93.24% F1-score. The Wilcoxon p-value of 0.03125 for accuracy and F1-score indicates a significant difference at the 0.05 level. These findings indicate that the integrated Random Forest, SMOTETomek, and GridSearchCV approach improves anemia classification on imbalanced data.

Keywords: Anemia Diagnosis; Class Imbalance; GridSearchCV; Random Forest; SMOTETomek

1. PENDAHULUAN

Anemia merupakan gangguan hematologi yang masih sering dijumpai, terutama pada kelompok rentan di negara berkembang. Secara umum, kondisi ini muncul ketika kadar hemoglobin atau jumlah sel darah merah tidak memenuhi batas normal, sehingga distribusi oksigen dari darah menuju jaringan tubuh menjadi kurang optimal[2], [3]. Dampaknya dapat berupa rasa lemah, mudah lelah, penurunan konsentrasi, berkurangnya produktivitas, serta meningkatnya risiko gangguan kesehatan lain. Karena itu, identifikasi anemia sejak awal diperlukan agar penanganan dapat diberikan lebih cepat dan sesuai dengan kondisi pasien[4], [5].

Perkembangan teknologi komputasi membuka peluang bagi bidang kesehatan untuk memanfaatkan Machine Learning sebagai alat bantu pengambilan keputusan[6], [7]. Melalui pendekatan tersebut, sistem dapat mengenali pola dari data terdahulu dan membuat prediksi berdasarkan fitur yang tersedia[8]. Pada kasus anemia, informasi citra dari konjungtiva mata, kuku, telapak tangan, dan mukosa bibir sering dimanfaatkan karena perubahan warna pada bagian tubuh tersebut dapat berkaitan dengan kadar hemoglobin[9]-[10]. Komponen warna RGB dari citra dapat menjadi masukan bagi model klasifikasi[11], sedangkan penggabungan informasi visual dengan kadar hemoglobin dapat memperkaya karakteristik data yang dipelajari model[12].

Dalam kajian diagnosis anemia, beragam metode klasifikasi telah digunakan, misalnya Naive Bayes, K-Nearest Neighbor, dan Random Forest[13], [14]. Pemilihan metode tersebut didasarkan pada penerapannya yang sederhana serta kemampuannya mengolah data medis dengan jumlah sampel terbatas[15]. Studi terdahulu juga menunjukkan bahwa pembelajaran mesin berpotensi membantu proses skrining awal anemia[16]. Akan tetapi, data kesehatan sering tidak memiliki jumlah sampel yang seimbang pada setiap kelas[17], [18]. Kondisi tersebut dapat membuat model lebih dominan mengenali kelas mayoritas, sementara kelas minoritas berisiko kurang terdeteksi[19].

Ketidakseimbangan kelas biasanya ditangani melalui resampling, termasuk oversampling seperti SMOTE dan ADASYN[20], [21]. SMOTE membentuk data buatan dari kelas minoritas melalui interpolasi dengan tetangga terdekat, sedangkan ADASYN menambah sampel sintetis secara adaptif pada area yang lebih sulit dipelajari. Pada data kesehatan, pendekatan ini membantu model memperoleh informasi yang lebih cukup dari kelas minoritas sehingga evaluasi menjadi lebih berimbang[22].

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 5, August 2026 | Hal 2049-2058
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i5.1215

Penelitian Zuhanda et al. dijadikan acuan karena membahas penggunaan ADASYN pada Naive Bayes untuk klasifikasi anemia dengan distribusi kelas tidak seimbang[1]. Pada dataset berjumlah 104 data dengan rasio kelas 3:1, ADASYN meningkatkan accuracy dari 0,57 menjadi 0,78, precision dari 0,17 menjadi 0,74, recall dari 0,20 menjadi 0,88, dan F1-score dari 0,18 menjadi 0,80. Temuan tersebut menunjukkan bahwa pengolahan data tidak seimbang berperan dalam membantu model mengenali anemia sebagai kelas minoritas. Namun, masih terdapat ruang perbaikan melalui pemilihan algoritma yang lebih fleksibel, pembersihan data pada area batas kelas, serta penentuan parameter model yang lebih sistematis.

Berangkat dari kondisi tersebut, penelitian ini menyoroti tiga aspek pengembangan. Pertama, asumsi independensi pada Naive Bayes kurang sesuai apabila fitur warna RGB dan kadar hemoglobin saling berkaitan. Kedua, ADASYN menambah data sintetis, tetapi tidak dirancang untuk menghapus sampel ambigu di sekitar batas keputusan. Ketiga, strategi pencarian hyperparameter perlu dijelaskan secara terstruktur karena konfigurasi parameter dapat memengaruhi kestabilan dan capaian model.

Berdasarkan uraian tersebut, terdapat kesenjangan penelitian yang belum terjawab pada studi-studi sebelumnya. Penelitian terdahulu mengenai klasifikasi anemia umumnya berhenti pada penerapan oversampling tunggal, seperti SMOTE atau ADASYN, yang hanya menambah sampel sintetis tanpa membersihkan sampel ambigu di sekitar batas keputusan. Di sisi lain, penelitian yang memanfaatkan Random Forest pada data medis tidak seimbang jarang menguraikan proses penentuan hyperparameter secara sistematis, sehingga hasilnya sulit direproduksi . Kombinasi antara metode hibrida SMOTETomek, algoritma ensemble Random Forest, optimasi GridSearchCV, dan pengujian signifikansi statistik pada dataset anemia berbasis fitur warna citra serta kadar hemoglobin belum ditemukan pada penelitian sebelumnya. Kesenjangan inilah yang menjadi fokus penelitian ini.

Tujuan penelitian ini adalah membangun model klasifikasi anemia yang akurat dan stabil pada kondisi data tidak seimbang. Kontribusi penelitian ini dapat diuraikan sebagai berikut. Pertama, penelitian ini mengintegrasikan SMOTETomek sebagai metode hibrida yang menggabungkan oversampling dengan pembersihan Tomek Links pada algoritma Random Forest, sehingga mengurangi keterbatasan asumsi independensi fitur yang terdapat pada penelitian rujukan. Kedua, penelitian ini menerapkan GridSearchCV untuk menentukan kombinasi hyperparameter secara sistematis dan terdokumentasi. Ketiga, evaluasi dilakukan menggunakan validasi silang 10-Fold yang diperkuat Wilcoxon Signed-Rank Test, sehingga peningkatan performa yang diperoleh dapat dibuktikan signifikan secara statistik dan bukan sekadar selisih angka. Keempat, penelitian ini menyajikan pembanding kuantitatif terhadap penelitian rujukan pada dataset yang sama.

Sebagai solusi, penelitian ini menerapkan Random Forest yang dipadukan dengan SMOTETomek dan GridSearchCV[24], [25]. Random Forest digunakan karena membangun sejumlah pohon keputusan dan menggabungkan keluarannya melalui voting, sehingga mampu menangkap hubungan antarfitur secara lebih adaptif pada data medis[26]. SMOTETomek dipakai untuk menata kembali komposisi data latih sekaligus mengurangi sampel yang berada di area batas keputusan, sedangkan GridSearchCV digunakan untuk memilih kombinasi hyperparameter secara terarah.

Pengujian kinerja dilakukan dengan validasi silang 10-Fold agar hasil tidak bergantung pada satu komposisi data latih dan data uji. Wilcoxon Signed-Rank Test diterapkan untuk menilai apakah perbedaan antarskenario bersifat signifikan secara statistik. Dengan rancangan tersebut, penelitian ini diarahkan untuk menghasilkan model klasifikasi anemia yang lebih stabil pada data tidak seimbang serta memperlihatkan kontribusi integrasi ensemble learning, penyeimbangan kelas, dan optimasi parameter.

2. METODOLOGI PENELITIAN

2.1 Dataset

Data penelitian berasal dari Anemia Diagnosis di Kaggle yang disusun oleh Shahzad Aslam dan juga digunakan pada penelitian rujukan. Dataset ini memuat 104 instance. Variabel prediktif terdiri atas Red Pixel, Green Pixel, Blue Pixel, dan Hb, sedangkan atribut Anaemic menjadi label klasifikasi dengan dua kategori, yaitu Yes untuk anemia dan No untuk normal.

Pada pemeriksaan awal, ditemukan 78 data normal dan 26 data anemia sehingga perbandingan kelasnya 3:1. Fitur utama tidak memiliki missing value. Kolom Number dan Name dikeluarkan dari pemodelan karena hanya berisi identitas dan tidak merepresentasikan kondisi medis yang dibutuhkan model.

Deskripsi lengkap setiap fitur yang digunakan dalam penelitian ini ditunjukkan pada Tabel 1.

Tabel 1. Deskripsi Fitur Dataset

No	Fitur	Tipe Data	Deskripsi
1	Red Pixel	Numerik	Persentase piksel merah pada data citra
2	Green Pixel	Numerik	Persentase piksel hijau pada data citra
3	Blue Pixel	Numerik	Persentase piksel biru pada data citra
4	Hb	Numerik	Kadar hemoglobin darah
5	Anaemic	Kategorikal	Target klasifikasi anemia: Ya atau Tidak

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 5, August 2026 | Hal 2049-2058
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i5.1215

Tabel 1 memperlihatkan empat fitur prediktif, yaitu %Red Pixel, %Green Pixel, %Blue Pixel, dan Hb, serta satu atribut target berupa Anaemic yang menjadi label klasifikasi.

Pada tahap pemodelan, fitur yang digunakan hanya persentase Red Pixel, Green Pixel, Blue Pixel, dan kadar Hb. Atribut Anaemic dijadikan target dengan kategori Yes dan No. Kolom identitas tidak dimasukkan agar proses pelatihan berfokus pada variabel yang berkaitan dengan indikasi anemia.

Contoh data pada Tabel 2 ditampilkan untuk memperlihatkan bentuk input yang dipakai sebelum proses analisis.

Tabel 2. Contoh Data Anemia

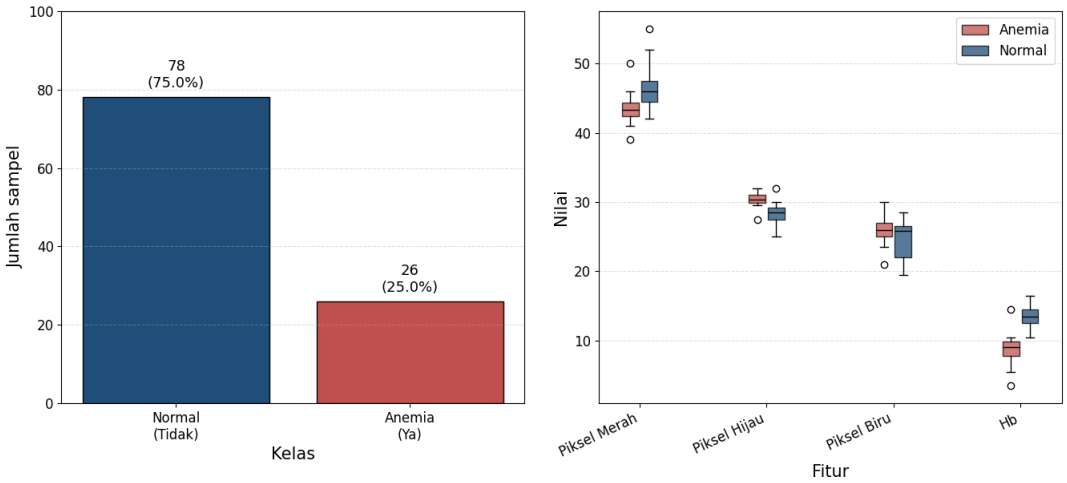
No	Red Pixel	Green Pixel	Blue Pixel	Hb	Status Anemia
1	43.2555	30.8421	25.9025	6.3	Ya
2	45.6033	28.1900	26.2067	13.5	Tidak
3	45.0107	28.9677	26.0215	11.7	Tidak
4	44.5398	28.9899	26.4703	13.5	Tidak
5	43.2870	30.6972	26.0158	12.4	Tidak
6	45.0994	27.9645	26.9361	16.2	Tidak
7	43.1457	30.1628	26.6915	8.6	Ya
8	43.6103	29.1099	27.2798	10.3	Tidak
9	45.0423	29.1660	25.7918	13.0	Tidak
10	46.5143	27.4282	26.0575	9.7	Ya
...	...	...	...	...	...
104	43.5706	29.8094	26.6199	12.2	Tidak

Tabel 2 menyajikan beberapa baris contoh data anemia. Setiap record berisi nilai piksel merah, hijau, biru, kadar hemoglobin, serta status anemia, sehingga struktur masukan model dapat dipahami sebelum tahap preprocessing. Selanjutnya, distribusi kelas pada data awal sebelum proses penyeimbangan disajikan pada Tabel 3.

Tabel 3. Distribusi Kelas Awal

Kelas	Jumlah Sampel	Persentase	Keterangan
Normal (Tidak)	78	75.00%	Kelas mayoritas
Anemia (Ya)	26	25.00%	Kelas minoritas
Total	104	100.00%	-

Tabel 3 menunjukkan bahwa data awal belum seimbang. Kelas Normal (No) terdiri atas 78 data atau 75.00%, sedangkan kelas Anemia (Yes) terdiri atas 26 data atau 25.00%. Perbedaan jumlah ini berpotensi membuat model lebih dominan mempelajari pola kelas normal sehingga penyeimbangan data diperlukan sebelum pelatihan. Kondisi awal dataset tersebut divisualisasikan secara grafis pada Gambar 1.



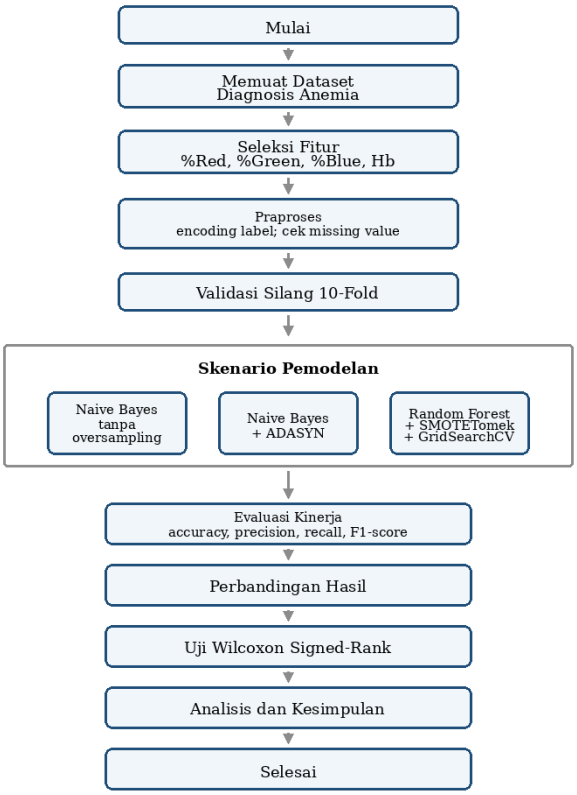
Gambar 1. Hasil Eksplorasi Awal Data

Gambar 1 memperlihatkan kondisi awal dataset. Diagram batang menunjukkan jumlah data normal lebih besar daripada data anemia, yaitu 78 dibanding 26 data. Boxplot fitur memperlihatkan variasi Red Pixel, Green Pixel, Blue Pixel, dan Hb pada tiap kelas. Informasi awal ini menjadi dasar pemilihan strategi penanganan data tidak seimbang.

2.2 Tahapan Penelitian

Alur penelitian dibuat untuk menggambarkan proses mulai dari pemilihan data, preprocessing, pelatihan model, hingga evaluasi. Rangkaian tersebut disajikan pada Gambar 2.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 5, August 2026 | Hal 2049-2058
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i5.1215



Gambar 2. Tahapan Penelitian Klasifikasi Anemia

Gambar 2 menjelaskan tahapan klasifikasi anemia. Proses diawali dengan pemuatan dataset Anemia Diagnosis, dilanjutkan pemilihan Red Pixel, Green Pixel, Blue Pixel, dan Hb sebagai variabel prediktif. Kolom Anaemic digunakan sebagai target, sedangkan Number dan Name tidak dimasukkan karena hanya berfungsi sebagai identitas. Tahap preprocessing meliputi pengecekan missing value dan perubahan label target ke bentuk numerik agar dapat diproses oleh model Machine Learning.

Pengujian dilakukan menggunakan validasi silang 10-Fold melalui tiga skenario. Skenario pertama menggunakan Naive Bayes tanpa oversampling sebagai baseline. Skenario kedua menggunakan Naive Bayes dengan ADASYN sebagai pembanding penelitian rujukan. Skenario ketiga menggunakan Random Forest yang dipadukan dengan SMOTETomek dan GridSearchCV sebagai metode usulan. Performa model dinilai melalui accuracy, precision, recall, dan F1-score, kemudian dibandingkan menggunakan Wilcoxon Signed-Rank Test.

2.3 Skenario Evaluasi

Evaluasi dirancang dalam tiga skenario agar perbandingan model dapat dilakukan secara konsisten. Naive Bayes tanpa oversampling digunakan sebagai baseline, Naive Bayes dengan ADASYN menjadi pembanding, sedangkan Random Forest dengan SMOTETomek dan GridSearchCV menjadi skenario utama. Seluruh skenario diuji melalui 10-Fold Cross Validation. Pada setiap fold, resampling hanya dilakukan pada data latih agar data uji tetap menjadi data yang belum pernah dilihat model.

Kinerja model dinilai menggunakan empat ukuran evaluasi, yaitu accuracy, precision, recall, dan F1-score. Accuracy digunakan untuk melihat jumlah prediksi benar dibandingkan seluruh data yang diuji. Precision menunjukkan tingkat ketepatan model ketika memberikan prediksi pada kelas anemia, sedangkan recall menggambarkan kemampuan model dalam menemukan sampel anemia yang benar-benar termasuk kelas tersebut. F1-score digunakan sebagai ukuran gabungan antara precision dan recall, sehingga lebih sesuai untuk membaca performa model pada data dengan komposisi kelas yang tidak seimbang. Selanjutnya, Wilcoxon Signed-Rank Test digunakan untuk menguji apakah perbedaan performa antar metode memiliki signifikansi statistik.

Perhitungan evaluasi mengacu pada empat komponen confusion matrix, yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN), sebagaimana ditunjukkan pada Persamaan (1)–(4). True Positive menunjukkan data anemia yang berhasil dikenali sebagai anemia, sedangkan True Negative menunjukkan data normal yang benar diklasifikasikan sebagai normal. False Positive terjadi ketika data normal diprediksi sebagai anemia, sementara False Negative terjadi ketika data anemia tidak berhasil terdeteksi oleh model.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 5, August 2026 | Hal 2049-2058
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i5.1215

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1\text{-}score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Persamaan (1) digunakan untuk menghitung proporsi seluruh prediksi yang benar. Pada kasus data tidak seimbang, accuracy perlu dilihat bersama metrik lain karena model dapat terlihat baik hanya karena sering memprediksi kelas mayoritas. Oleh sebab itu, Persamaan (2)-(4) digunakan untuk mengukur precision, recall, dan F1-score. Precision menilai ketepatan prediksi anemia, recall menilai kemampuan mendeteksi seluruh sampel anemia, sedangkan F1-score memberikan nilai gabungan yang lebih seimbang.

2.4 Metode yang Diusulkan

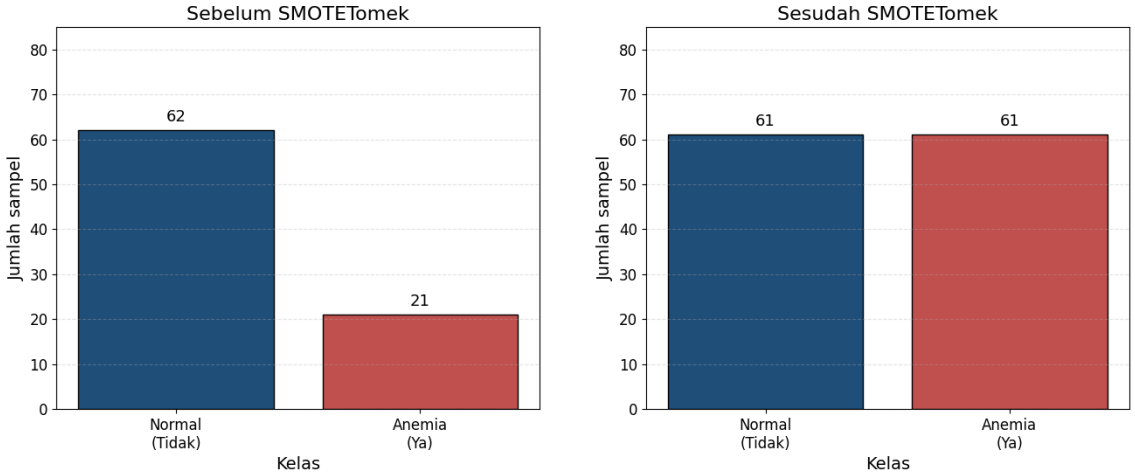
Metode usulan mengintegrasikan Random Forest, SMOTETomek, dan GridSearchCV untuk memperbaiki klasifikasi anemia pada data tidak seimbang. Random Forest dipakai sebagai classifier utama karena membentuk banyak pohon keputusan dan menentukan kelas akhir melalui voting mayoritas. Mekanisme ensemble tersebut membantu model mempelajari hubungan antarfitur secara lebih fleksibel, khususnya pada kombinasi fitur warna dan kadar hemoglobin.

SMOTETomek digunakan pada data latih untuk memperbaiki komposisi kelas. Teknik ini menggabungkan SMOTE sebagai pembentuk sampel sintetis kelas minoritas dan Tomek Links sebagai mekanisme pembersihan sampel yang ambigu di dekat batas keputusan. Dengan cara tersebut, data anemia tidak hanya ditambah jumlahnya, tetapi juga dibuat lebih bersih untuk proses pelatihan.

GridSearchCV berfungsi mencari konfigurasi hyperparameter Random Forest yang paling sesuai. Parameter yang dievaluasi meliputi jumlah pohon, kedalaman pohon, jumlah minimum sampel pada percabangan, jumlah minimum sampel pada daun, jumlah fitur yang dipakai saat pemisahan node, dan kriteria pemisahan. Kombinasi terbaik dari proses pencarian tersebut dipakai untuk membangun model akhir.

Untuk mencegah data leakage, SMOTETomek dan GridSearchCV hanya dijalankan pada data latih di setiap fold. Data uji tidak digunakan dalam resampling, pembuatan sampel sintetis, maupun pencarian hyperparameter. Strategi ini diterapkan agar hasil evaluasi lebih mencerminkan kemampuan generalisasi model.

Perbandingan distribusi kelas pada data latih sebelum dan sesudah penerapan SMOTETomek ditampilkan pada Gambar 3.



Gambar 3. Perbandingan Distribusi Kelas Latih Sebelum dan Sesudah Penerapan SMOTETomek

Gambar 3 membandingkan jumlah data latih sebelum dan sesudah SMOTETomek. Sebelum penyeimbangan, data latih terdiri atas 62 sampel Normal dan 21 sampel Anemia. Setelah SMOTETomek diterapkan, jumlah kedua kelas menjadi sama, yaitu masing-masing 61 sampel. Hasil tersebut menunjukkan bahwa data latih berhasil dibuat lebih proporsional sebelum proses pembelajaran model.

3. HASIL DAN PEMBAHASAN

Bagian ini memaparkan hasil pengujian pada tiga skenario klasifikasi anemia, yaitu Naive Bayes tanpa oversampling, Naive Bayes dengan ADASYN, dan Random Forest dengan SMOTETomek serta GridSearchCV. Analisis disusun berdasarkan accuracy, precision, recall, F1-score, uji signifikansi statistik, serta feature importance.

3.1 Hasil Perbandingan Model

Hasil evaluasi memperlihatkan bahwa skenario usulan memberikan performa paling baik. Naive Bayes tanpa oversampling memperoleh accuracy 94.09%, precision 92.50%, recall 86.67%, dan F1-score 86.57%. Naive Bayes dengan ADASYN menghasilkan accuracy 87.36%, precision 74.33%, recall 91.67%, dan F1-score 79.31%. Sementara

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 5, August 2026 | Hal 2049-2058
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i5.1215

itu, Random Forest dengan SMOTETomek dan GridSearchCV memperoleh accuracy 97.09%, precision 97.50%, recall 91.67%, dan F1-score 93.24%.

Capaian Random Forest yang lebih tinggi menunjukkan bahwa pendekatan ensemble lebih mampu membaca pola pada fitur warna dan kadar hemoglobin. Hubungan antara %Red Pixel, %Green Pixel, dan %Blue Pixel dapat saling memengaruhi sehingga asumsi independensi pada Naive Bayes menjadi kurang menguntungkan. Random Forest tidak bergantung pada asumsi tersebut karena keputusan akhir diperoleh dari kombinasi banyak pohon keputusan. Perbandingan performa ketiga skenario model berdasarkan validasi silang 10-Fold disajikan pada Tabel 4.

Tabel 4. Perbandingan Performa Model Menggunakan Validasi Silang 10-Fold

Metode	Akurasi	Presisi	Recall	F1-score
Naive Bayes tanpa oversampling	94.09%	92.50%	86.67%	86.57%
Naive Bayes + ADASYN	87.36%	74.33%	91.67%	79.31%
Random Forest + SMOTETomek + GridSearchCV	97.09%	97.50%	91.67%	93.24%

Berdasarkan Tabel 4, skenario Random Forest dengan SMOTETomek dan GridSearchCV memiliki nilai accuracy, precision, dan F1-score paling tinggi. Precision yang tinggi menunjukkan bahwa prediksi kelas anemia lebih tepat, sedangkan recall yang setara dengan Naive Bayes + ADASYN menunjukkan bahwa kemampuan menemukan data anemia tetap terjaga. Kombinasi tersebut menandakan bahwa metode usulan lebih seimbang dalam menangani data yang tidak merata.

Hasil perbandingan juga memperlihatkan bahwa peningkatan kinerja tidak hanya disebabkan oleh algoritma klasifikasi. SMOTETomek membantu menyusun ulang distribusi data latih, sementara GridSearchCV memastikan parameter Random Forest dipilih melalui proses pencarian yang terarah. Kedua komponen tersebut membuat pelatihan lebih sesuai dengan karakteristik dataset anemia.

3.2 Ablation Study

Studi ablasi dilakukan untuk menilai kontribusi masing-masing komponen dalam metode usulan. Empat konfigurasi Random Forest dibandingkan, yaitu model bawaan, model dengan SMOTETomek, model dengan GridSearchCV, serta model yang menggabungkan SMOTETomek dan GridSearchCV. Semua konfigurasi diuji menggunakan 10-Fold Cross Validation dengan metrik accuracy, precision, recall, dan F1-score.

Hasil studi ablasi dari keempat konfigurasi tersebut ditunjukkan pada Tabel 5.

Tabel 5. Hasil Studi Ablasi

Model	Akurasi	Std. Akurasi	Presisi	Recall	F1-score
Random Forest bawaan	96.09%	4.79%	97.50%	88.33%	91.24%
Random Forest + SMOTETomek	97.09%	4.45%	97.50%	91.67%	93.24%
Random Forest + GridSearchCV	97.09%	4.45%	97.50%	91.67%	93.24%
Random Forest + SMOTETomek + GridSearchCV	97.09%	4.45%	97.50%	91.67%	93.24%

Berdasarkan Tabel 5, Random Forest bawaan telah memberikan hasil yang tinggi dengan accuracy 96.09%, precision 97.50%, recall 88.33%, dan F1-score 91.24%. Setelah SMOTETomek digunakan, accuracy meningkat menjadi 97.09%, recall menjadi 91.67%, dan F1-score menjadi 93.24%. Kenaikan recall menunjukkan bahwa penyeimbangan data membantu model mengenali kelas anemia yang jumlahnya lebih sedikit.

Random Forest dengan GridSearchCV menghasilkan nilai yang sama dengan kombinasi SMOTETomek dan GridSearchCV pada dataset ini, yaitu accuracy 97.09%, precision 97.50%, recall 91.67%, dan F1-score 93.24%. Kondisi tersebut menunjukkan bahwa Random Forest cukup stabil pada dataset kecil, sedangkan SMOTETomek tetap memberikan manfaat pada deteksi kelas minoritas. Dengan demikian, kontribusi utama metode terletak pada integrasi algoritma, penyeimbangan kelas, dan optimasi parameter.

3.3 Detail Hasil 10-Fold Cross Validation

Validasi silang 10-Fold digunakan untuk melihat kestabilan performa pada setiap fold. Dataset dibagi menjadi sepuluh bagian, kemudian proses pelatihan dan pengujian dilakukan bergantian. Melalui skema ini, setiap data memiliki kesempatan menjadi data uji sehingga hasil evaluasi tidak bergantung pada satu pembagian data saja. Pendekatan ini sesuai untuk dataset dengan jumlah sampel terbatas. Rincian akurasi yang diperoleh setiap model pada masing-masing fold disajikan pada Tabel 6.

Tabel 6. Akurasi Setiap Fold pada Validasi Silang 10-Fold

Fold	NB	NB + ADASYN	RF + SMOTETomek + GridCV
1	90.91%	90.91%	100.00%
2	100.00%	90.91%	100.00%
3	100.00%	81.82%	90.91%
4	100.00%	100.00%	100.00%
5	100.00%	100.00%	100.00%

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 5, August 2026 | Hal 2049-2058
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i5.1215

6	90.00%	90.00%	90.00%
7	90.00%	80.00%	90.00%
8	100.00%	100.00%	100.00%
9	90.00%	80.00%	100.00%
10	80.00%	60.00%	100.00%
Rata-rata	94.09%	87.36%	97.09%

Tabel 6 memperlihatkan bahwa Random Forest dengan SMOTETomek dan GridSearchCV lebih stabil daripada dua skenario pembanding. Model ini mencapai accuracy 100.00% pada fold ke-1, ke-2, ke-4, ke-5, ke-8, ke-9, dan ke-10, serta nilai terendah 90.00%. Sebaliknya, Naive Bayes + ADASYN turun sampai 60.00% pada fold ke-10. Hasil tersebut menunjukkan bahwa metode usulan tidak hanya unggul pada rata-rata, tetapi juga lebih konsisten pada validasi silang.

Konsistensi antar fold penting karena data yang digunakan relatif sedikit. Model dengan rata-rata accuracy tinggi belum tentu stabil apabila performanya turun tajam pada fold tertentu. Pada penelitian ini, gabungan SMOTETomek dan GridSearchCV pada Random Forest menghasilkan variasi performa yang lebih terkendali sehingga model lebih layak digunakan untuk klasifikasi anemia pada data tidak seimbang.

3.4 Uji Signifikansi Statistik

Uji signifikansi dilakukan untuk memastikan bahwa selisih performa antara metode usulan dan pembanding tidak hanya terlihat secara deskriptif. Wilcoxon Signed-Rank Test dipilih karena perbandingan dilakukan berdasarkan hasil tiap fold. Pengujian difokuskan pada accuracy dan F1-score karena keduanya mewakili performa umum serta keseimbangan antara precision dan recall. Hasil pengujian Wilcoxon Signed-Rank Test terhadap skenario yang dibandingkan ditunjukkan pada Tabel 7.

Tabel 7. Hasil Uji Wilcoxon Signed-Rank

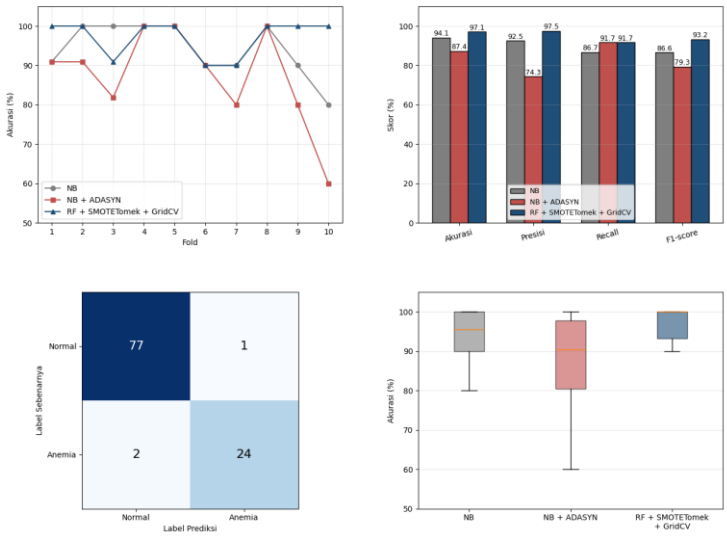
Metrik	Nilai-p	Keputusan
Akurasi	0.03125	Signifikan karena $p < 0.05$
F1-score	0.03125	Signifikan karena $p < 0.05$

Berdasarkan Tabel 7, nilai p-value untuk accuracy dan F1-score adalah 0.03125. Nilai ini berada di bawah taraf signifikansi 0.05, sehingga perbedaan performa antara metode usulan dan metode pembanding dapat dinyatakan signifikan. Temuan tersebut memperkuat bahwa peningkatan Random Forest dengan SMOTETomek dan GridSearchCV tidak hanya tampak pada nilai rata-rata, tetapi juga didukung oleh pengujian statistik.

Penggunaan uji statistik diperlukan karena dataset berukuran kecil dan hasil setiap fold dapat dipengaruhi oleh variasi pembagian data. Dengan Wilcoxon Signed-Rank Test, kesimpulan mengenai peningkatan performa model memiliki dasar yang lebih kuat daripada sekadar membandingkan nilai rata-rata.

3.5 Visualisasi Hasil dan Feature Importance

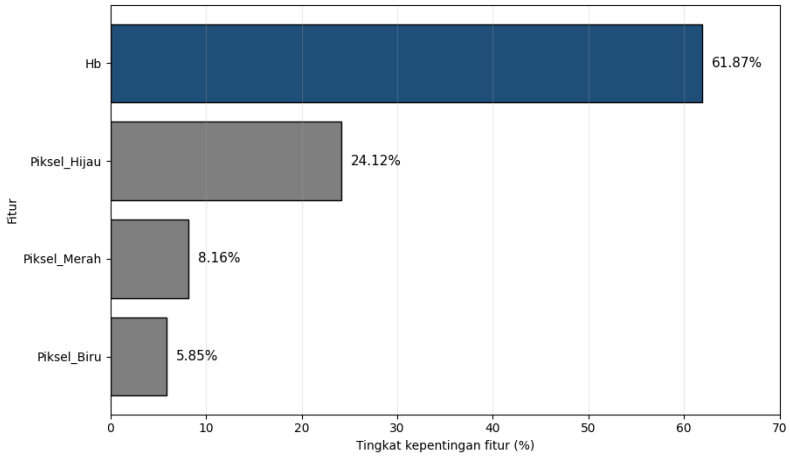
Visualisasi evaluasi digunakan untuk memperjelas perbandingan performa antar model pada validasi silang 10-Fold. Selain menyajikan metrik numerik, visualisasi membantu melihat pola accuracy di setiap fold, variasi performa, dan hasil klasifikasi melalui confusion matrix. Dengan demikian, pembahasan didukung oleh tabel dan representasi grafis. Ringkasan hasil evaluasi model pada validasi silang 10-Fold ditampilkan pada Gambar 4.



Gambar 4. Hasil Evaluasi Model Menggunakan Validasi Silang 10-Fold

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 5, August 2026 | Hal 2049-2058
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i5.1215

Gambar 4 menyajikan ringkasan evaluasi model secara visual. Grafik accuracy per fold menunjukkan bahwa Random Forest dengan SMOTETomek dan GridSearchCV lebih stabil dibandingkan skenario pembanding. Diagram metrik memperlihatkan keunggulan metode usulan pada accuracy, precision, dan F1-score. Confusion matrix agregat memperlihatkan kemampuan model membedakan kelas Normal dan Anemia, sedangkan boxplot accuracy menunjukkan variasi performa yang lebih terkendali. Selanjutnya, tingkat kepentingan setiap fitur pada model Random Forest ditampilkan pada Gambar 5.



Gambar 5. Tingkat Kepentingan Fitur Model Random Forest

Gambar 5 menunjukkan tingkat kepentingan fitur pada model Random Forest. Fitur Hb menjadi atribut paling berpengaruh karena kadar hemoglobin merupakan indikator utama dalam diagnosis anemia. Fitur Red Pixel, Green Pixel, dan Blue Pixel juga berkontribusi karena informasi warna dapat menggambarkan karakteristik visual yang berhubungan dengan anemia. Artinya, model memanfaatkan informasi hemoglobin dan fitur warna secara bersamaan.

3.6 Pembahasan Penelitian Rujukan

Jika dibandingkan dengan penelitian rujukan yang menggunakan Naive Bayes dan ADASYN, penelitian ini memperluas pengembangan pada sisi algoritma, penanganan ketidakseimbangan kelas, dan optimasi parameter. Naive Bayes memiliki keterbatasan karena mengasumsikan fitur saling independen, sedangkan fitur warna dan hemoglobin dapat berkaitan. Oleh karena itu, Random Forest digunakan karena lebih fleksibel dalam mempelajari pola data. SMOTETomek dipilih karena penyeimbangan dilakukan sekaligus dengan pembersihan sampel ambigu di sekitar batas keputusan.

GridSearchCV digunakan untuk menentukan konfigurasi hyperparameter Random Forest secara sistematis. Pemilihan parameter yang tepat dapat memengaruhi kemampuan model dalam membaca pola data. Berbeda dari parameter default, GridSearchCV menilai beberapa kombinasi parameter sehingga model akhir lebih menyesuaikan karakteristik dataset. Dengan demikian, peningkatan kinerja berasal dari perpaduan algoritma, penyeimbangan data, dan optimasi parameter.

Metode usulan menghasilkan accuracy 97.09%, lebih tinggi daripada Naive Bayes dengan ADASYN pada skenario evaluasi yang sama. Namun, pembandingan dengan penelitian rujukan perlu dilakukan secara cermat karena rancangan evaluasi dapat memengaruhi hasil akhir. Oleh karena itu, penelitian ini memakai 10-Fold Cross Validation sebagai evaluasi utama agar hasil pengujian lebih stabil. Ringkasan perbandingan kontribusi antara penelitian rujukan dan penelitian ini disajikan pada Tabel 8.

Tabel 8. Perbandingan Kontribusi Penelitian

Aspek	Zuhanda et al.	Penelitian Ini	Kontribusi
Algoritma	Naive Bayes	Random Forest	Mengurangi keterbatasan asumsi independensi fitur
Penyeimbangan	ADASYN	SMOTETomek	Oversampling dikombinasikan dengan pembersihan Tomek Links
Tuning	Belum dijelaskan secara sistematis	GridSearchCV	Hyperparameter dipilih melalui pencarian sistematis
Evaluasi	Hold-out atau evaluasi terbatas	Validasi Silang 10-Fold	Evaluasi lebih stabil pada dataset kecil
Uji signifikansi	Tidak dilakukan	Wilcoxon Signed-Rank Test	Peningkatan performa dibuktikan signifikan (p = 0,03125)
Capaian (accuracy)	78%	97,09%	Peningkatan 19,09 poin persentase

Tabel 8 memperlihatkan perbedaan antara penelitian rujukan dan penelitian ini. Perbedaan tersebut mencakup algoritma yang digunakan, strategi penyeimbangan data, mekanisme optimasi hyperparameter, serta rancangan evaluasi yang lebih stabil.

4. KESIMPULAN

Penelitian ini mengimplementasikan Random Forest yang dipadukan dengan SMOTETomek dan GridSearchCV untuk klasifikasi diagnosis anemia pada data tidak seimbang, menggunakan dataset berjumlah 104 data yang terdiri atas 78 data normal dan 26 data anemia, dengan SMOTETomek diterapkan untuk menata distribusi kelas pada data latih dan GridSearchCV digunakan untuk memperoleh hyperparameter terbaik pada Random Forest. Hasil validasi silang 10-Fold menunjukkan bahwa metode usulan memperoleh accuracy 97.09%, precision 97.50%, recall 91.67%, dan F1-score 93.24%, yang lebih baik daripada Naive Bayes tanpa oversampling maupun Naive Bayes dengan ADASYN pada skenario evaluasi yang sama, sedangkan Wilcoxon Signed-Rank Test menghasilkan p-value 0.03125 pada accuracy dan F1-score sehingga peningkatan performa tersebut dinyatakan signifikan pada taraf 0.05. Meskipun demikian, penelitian ini masih memiliki keterbatasan pada ukuran data yang relatif kecil serta fitur yang hanya mencakup persentase piksel warna dan kadar hemoglobin, sehingga penelitian berikutnya disarankan menggunakan dataset klinis yang lebih besar, menambahkan fitur hematologi seperti MCH, MCHC, MCV, RBC, dan WBC, membandingkan model dengan XGBoost, LightGBM, Support Vector Machine, atau pendekatan deep learning, serta melakukan validasi menggunakan data klinis dari fasilitas kesehatan agar model dapat dikaji lebih lanjut sebagai sistem pendukung keputusan.

REFERENCES

[1] A. Z. Alem *et al.*, “Prevalence and factors associated with anemia in women of reproductive age across low- and middle-income countries based on national data,” *Sci. Rep.*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-46739-z.
[2] S. Safiri *et al.*, “Burden of anemia and its underlying causes in 204 countries and territories, 1990–2019: results from the Global Burden of Disease Study 2019,” *J. Hematol. Oncol.*, vol. 14, no. 1, Dec. 2021, doi: 10.1186/s13045-021-01202-2.
[3] Fachira Kasmarini and Ratih Kurniasari, “The Indonesian Journal of Health Promotion,” *Media Publikasi Promosi Kesehatan Indonesia*, 2022, doi: 10.31934/mppki.v2i3.
[4] K. Aprilianti Cia *et al.*, “Asupan Zat Besi dan Prevalensi Anemia pada Remaja Usia 16-18 Tahun,” 2021. Accessed: May 26, 2026. [Online]. Available: https://doi.org/10.33096/woh.vi.248
[5] Hirmayant and E. Utami, “Enhanced Heart Disease Diagnosis Using Machine Learning Algorithms: A Comparison of Feature Selection,” *Jurnal RESTI*, vol. 9, no. 2, pp. 385–392, Apr. 2025, doi: 10.29207/resti.v9i2.6175.
[6] R. Faurina, M. J. Gazali, and I. D. A. Herani, “Optimization Of Disease Prediction Accuracy Through Artificial Neural Network (Ann) Algorithms In Diagnese Application,” *Jurnal Teknik Informatika (Jutif)*, vol. 5, no. 2, pp. 339–347, Apr. 2024, doi: 10.52436/1.jutif.2024.5.2.1182.
[7] M. M. Ahsan, S. A. Luna, and Z. Siddique, “Machine-Learning-Based Disease Diagnosis: A Comprehensive Review,” Mar. 01, 2022, *MDPI*. doi: 10.3390/healthcare10030541.
[8] J. W. Asare, P. Appiahene, E. J. Arthur, S. Korankye, S. Afrifa, and E. T. Donkoh, “Detection of anemia using conjunctiva images: A smartphone application approach,” *Med. Nov. Technol. Devices*, vol. 18, Jun. 2023, doi: 10.1016/j.medntd.2023.100237.
[9] M. Mansour, T. B. Donmez, M. Kutlu, and S. Mahmud, “Non-invasive detection of anemia using lip mucosa images transfer learning convolutional neural networks,” *Front. Big Data*, vol. 6, 2023, doi: 10.3389/fdata.2023.1291329.
[10] J. R. Navarro-Cabrera, M. A. Valles-Coral, M. E. Farro-Roque, N. Reátegui-Lozano, and L. Arévalo-Fasanando, “Machine vision model using nail images for non-invasive detection of iron deficiency anemia in university students,” *Front. Big Data*, vol. 8, 2025, doi: 10.3389/fdata.2025.1557600.
[11] J. W. Asare, W. L. Brown-Acquaye, M. M. Ujakpa, E. Freeman, and P. Appiahene, “Application of machine learning approach for iron deficiency anaemia detection in children using conjunctiva images,” *Inform. Med. Unlocked*, vol. 45, Jan. 2024, doi: 10.1016/j.imu.2024.101451.
[12] R. Magdalena *et al.*, “Convolutional Neural Network For Anemia Detection Based On Conjunctiva Palpebral Images,” *Jurnal Teknik Informatika (JUTIF)*, 2022, doi: 10.20884/1.jutif.2022.3.2.197.
[13] J. G. Gómez, C. Parra Urueta, D. S. Álvarez, V. Hernández Riaño, and G. Ramirez-Gonzalez, “Anemia Classification System Using Machine Learning,” *Informatics*, vol. 12, no. 1, Mar. 2025, doi: 10.3390/informatics12010019.
[14] P. Appiahene, J. W. Asare, E. T. Donkoh, G. Dimauro, and R. Maglietta, “Detection of iron deficiency anemia by medical images: a comparative study of machine learning algorithms,” *BioData Min.*, vol. 16, no. 1, Dec. 2023, doi: 10.1186/s13040-023-00319-z.
[15] M. M. Ahsan, S. A. Luna, and Z. Siddique, “Machine-Learning-Based Disease Diagnosis: A Comprehensive Review,” Mar. 01, 2022, *MDPI*. doi: 10.3390/healthcare10030541.
[16] S. Ayu Wulandari, H. Al Azies, M. Naufal, W. Adi Prasetyanto, and F. Az Zahra, “Breaking Boundaries in Diagnosis: Non-Invasive Anemia Detection Empowered by AI,” *IEEE Access*, vol. 12, pp. 9292–9307, 2024, doi: 10.1109/ACCESS.2017.Doi.
[17] M. K. Zuhanda, L. Permata, Hartono, E. Ongko, and Desniarti, “Impact of Adaptive Synthetic on Naïve Bayes Accuracy in Imbalanced Anemia Detection Datasets,” *Jurnal RESTI*, vol. 9, no. 1, pp. 85–93, Feb. 2025, doi: 10.29207/resti.v9i1.6031.
[18] M. Salmi, D. Atif, D. Oliva, A. Abraham, and S. Ventura, “Handling imbalanced medical datasets: review of a decade of research,” *Artif. Intell. Rev.*, vol. 57, no. 10, Oct. 2024, doi: 10.1007/s10462-024-10884-2.
[19] A. X. Wang, V. T. Le, H. N. Trung, and B. P. Nguyen, “Addressing imbalance in health data: Synthetic minority oversampling using deep learning,” *Comput. Biol. Med.*, vol. 188, Apr. 2025, doi: 10.1016/j.compbiomed.2025.109830.

[20] A. G. Coimbra *et al.*, "Approaches for handling imbalanced data used in machine learning in the healthcare field: A case study on Chagas disease database prediction," *PLoS One*, vol. 20, no. 5 May, May 2025, doi: 10.1371/journal.pone.0320966.
[21] Y. Yang, H. A. Khorshidi, and U. Aickelin, "A review on over-sampling techniques in classification of multi-class imbalanced datasets: insights for medical problems," 2024, *Frontiers Media SA*. doi: 10.3389/fdgth.2024.1430245.
[22] J. Zhu *et al.*, "Processing imbalanced medical data at the data level with assisted-reproduction data as an example," *BioData Min.*, vol. 17, no. 1, Dec. 2024, doi: 10.1186/s13040-024-00384-y.
[23] H. Hairani, T. Widiyaningtyas, and D. Dwi Prasetya, "International Journal On Informatics Visualization journal homepage : www.joiv.org/index.php/joiv International Journal On Informatics Visualization Addressing Class Imbalance of Health Data: a Systematic Literature Review on Modified Synthetic Minority Oversampling Technique (SMOTE) Strategies," 2024. [Online]. Available: www.joiv.org/index.php/joiv
[24] M. K. Zuhanda, L. Permata, Hartono, E. Ongko, and Desniarti, "Impact of Adaptive Synthetic on Naïve Bayes Accuracy in Imbalanced Anemia Detection Datasets," *Jurnal RESTI*, vol. 9, no. 1, pp. 85–93, Feb. 2025, doi: 10.29207/resti.v9i1.6031.
[25] M. I. Elim and E. Utami, "Performance Comparison of Child Stunting Prediction Support Vector Machine vs Random Forest with Grid Search Optimization," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 5, pp. 5305–5319, Oct. 2025, doi: 10.52436/1.jutif.2025.6.5.5285.
[26] H. Hairani, A. Anggrawan, and D. Priyanto, "International Journal On Informatics Visualization journal homepage : www.joiv.org/index.php/joiv INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION Improvement Performance of the Random Forest Method on Unbalanced Diabetes Data Classification Using Smote-Tomek Link," 2024. [Online]. Available: www.joiv.org/index.php/joiv
[27] Y. Y. Yuruk, "Uncover This Tech Term: Random Forest," Sep. 01, 2025, *Korean Radiological Society*. doi: 10.3348/kjr.2025.0800.