

Sistem Prediksi Anemia Berbasis Web Menggunakan Algoritma Logistic Regression

Adjie Qomarul Zaman^{*1}, Nur Latifah Dwi Mutiara Sari²

^{1,2}Program Studi Informatika, Universitas PGRI Semarang, Kota Semarang

*Email: adjiqomarulzaman@gmail.com¹

Abstract.

Anemia is a health condition affecting many people worldwide and often remains undetected due to the absence of early symptoms. This study aims to develop a web-based anemia prediction system using Logistic Regression, a machine learning algorithm suitable for binary classification. The research was conducted by collecting and preprocessing blood test data, focusing on features such as Hemoglobin, Mean Corpuscular Hemoglobin (MCH), Mean Corpuscular Hemoglobin Concentration (MCHC), and Mean Corpuscular Volume (MCV). The Logistic Regression model was trained and evaluated to determine its accuracy in predicting anemia. The resulting model was integrated into a web application using the Flask framework, allowing users to input their blood test parameters and receive real-time predictions. The results demonstrate that Logistic Regression effectively classifies anemia based on these hematological features. This application supports early screening and helps individuals assess their anemia risk prior to further medical consultation

Keywords: anemia; prediction system; logistic regression; blood parameters; web application

Abstrak

Anemia adalah kondisi kesehatan yang memengaruhi banyak orang di seluruh dunia dan sering kali tidak terdeteksi karena tidak adanya gejala awal. Penelitian ini bertujuan mengembangkan sistem prediksi anemia berbasis web menggunakan Logistic Regression, algoritma machine learning yang sesuai untuk klasifikasi biner. Penelitian dilakukan dengan mengumpulkan dan memproses data hasil tes darah, dengan fokus pada fitur Hemoglobin, Mean Corpuscular Hemoglobin (MCH), Mean Corpuscular Hemoglobin Concentration (MCHC), dan Mean Corpuscular Volume (MCV). Model Logistic Regression dilatih dan dievaluasi untuk mengukur akurasinya dalam memprediksi anemia. Model yang dihasilkan diintegrasikan ke dalam aplikasi web menggunakan framework Flask, sehingga pengguna dapat memasukkan parameter hasil tes darah dan menerima prediksi secara waktunya. Hasil menunjukkan bahwa Logistic Regression efektif dalam mengklasifikasikan anemia berdasarkan fitur hematologis tersebut. Aplikasi ini dapat mendukung skrining dini dan membantu individu menilai risiko anemia sebelum konsultasi medis lebih lanjut.

Kata kunci: anemia; sistem prediksi; logistic regression; parameter darah; aplikasi web

1. Pendahuluan

Anemia merupakan salah satu masalah kesehatan global yang berdampak signifikan terhadap kualitas hidup masyarakat, terutama di negara berkembang. Berdasarkan data dari World Health Organization (WHO), sekitar kawasan di Afrika dan Asia Tenggara paling banyak terkena dampak dengan perkiraan 106 juta wanita dan 103 juta anak terkena anemia di Afrika, sedangkan di Asia Tenggara 244 juta wanita dan 83 juta anak terkena dampaknya[1]. Anemia ditandai oleh rendahnya kadar hemoglobin dalam darah, yang mengakibatkan berkurangnya

kapasitas darah untuk mengangkut oksigen ke seluruh tubuh. Kondisi ini dapat menyebabkan kelelahan, penurunan konsentrasi, serta gangguan pertumbuhan dan perkembangan, terutama pada anak-anak. Deteksi dini terhadap anemia sangat penting untuk mencegah komplikasi jangka panjang, namun sayangnya, di banyak tempat, akses terhadap fasilitas medis yang memadai masih terbatas. Oleh karena itu, pengembangan sistem berbasis teknologi untuk membantu proses deteksi anemia menjadi kebutuhan yang mendesak.

Dalam beberapa tahun terakhir, pembelajaran mesin (machine learning) telah menjadi teknologi yang menjanjikan untuk mendeteksi berbagai penyakit, termasuk anemia. Machine learning memungkinkan sistem belajar dari data dan membuat prediksi dengan akurasi yang tinggi [2]. Salah satu algoritma yang cocok digunakan dalam klasifikasi kesehatan biner adalah Logistic Regression, karena kesederhanaan dan kemampuan interpretasinya yang tinggi [3], [4]. Logistic Regression mampu mengidentifikasi keterkaitan antara fitur klinis seperti hemoglobin, MCH, MCHC, dan MCV terhadap kondisi anemia dengan performa yang kompetitif [5], [6].

Logistic Regression merupakan algoritma klasifikasi biner yang memodelkan hubungan antara variabel hasil biner dan variabel predictor [3], yang sangat sesuai untuk kasus seperti prediksi anemia (ya/tidak), serta memiliki keunggulan dalam hal interpretabilitas dan efisiensi komputasi. Sifatnya yang sederhana namun kuat menjadikan Logistic Regression ideal untuk diterapkan dalam sistem prediktif yang bertujuan untuk digunakan oleh pengguna awam maupun tenaga medis sebagai alat bantu awal diagnosis. Meskipun beberapa studi telah menyentuh penggunaan Logistic Regression untuk analisis data medis, integrasi model ini dalam sistem berbasis web yang interaktif dan responsif masih jarang dibahas secara mendalam.

Dalam penelitian ini, penulis menawarkan sebuah pendekatan berbasis Logistic Regression untuk membangun sistem prediksi anemia yang terintegrasi dalam sebuah aplikasi web. Sistem ini dirancang untuk menerima input parameter darah dari pengguna, memprosesnya melalui model klasifikasi Logistic Regression, dan memberikan hasil prediksi secara langsung. Pendekatan ini bertujuan untuk mengisi celah pada penelitian sebelumnya yang belum mengeksplorasi cukup dalam penerapan algoritma sederhana namun efektif dalam platform yang mudah diakses. Dengan menyederhanakan proses deteksi anemia dan membuatnya tersedia secara online, diharapkan sistem ini dapat memberikan kontribusi nyata terhadap upaya deteksi dini anemia di masyarakat luas.

Sebagai tujuan utama, penelitian ini mengembangkan sebuah sistem berbasis web yang mampu memprediksi kondisi anemia seseorang berdasarkan input data darah menggunakan algoritma Logistic Regression. Sistem ini dirancang dengan fokus pada akurasi prediksi, kemudahan penggunaan, serta ketersediaan akses melalui internet. Metode yang digunakan meliputi pengumpulan dan pra-pemrosesan data, pelatihan model Logistic Regression, dan pengembangan antarmuka web menggunakan framework Flask sebagai penghubung antara pengguna dan model prediksi. Diharapkan, hasil dari penelitian ini dapat memberikan solusi yang aplikatif dan dapat digunakan secara luas untuk mendeteksi anemia secara cepat dan efisien.

2. Metode

Metode dalam penelitian ini mengadaptasi alur umum dalam pengembangan model klasifikasi berbasis machine learning seperti yang dijelaskan oleh Zemariam *et al.*[5] dan Tesfaye *et al.*[4], yang terdiri dari pengumpulan data, pra-pemrosesan data, pemilihan fitur, pelatihan dan evaluasi model, serta implementasi sistem. Pemilihan algoritma dilakukan dengan mempertimbangkan sifat data, efisiensi komputasi, serta kebutuhan klasifikasi biner dalam konteks prediksi anemia. Dan Penelitian ini mengusulkan sistem prediksi anemia

berbasis web dengan memanfaatkan algoritma Logistic Regression sebagai model klasifikasi utama. Tujuan utama dari metode ini adalah menghasilkan sistem prediktif yang sederhana, akurat, dan dapat digunakan secara real-time melalui antarmuka web. Tahapan metode terdiri dari beberapa bagian utama, yaitu: Pengumpulan Data, Pemrosesan Data, Pelatihan Model, Evaluasi Model, serta Pengembangan Aplikasi Web.



Gambar 1. Diagram Alur Penelitian.

2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini bersumber dari dataset publik yang tersedia di Kaggle:<https://www.kaggle.com/datasets/biswaranjanrao/anemia-dataset>. Beberapa fitur utama yang digunakan dalam dataset meliputi Hemoglobin, MCV, MCH, MCHC, Jenis kelamin

2.2 Pemrosesan Data

Tahap ini mencakup beberapa proses yang pertama pembersihan data dimana menghapus data duplikat dan mengisi nilai yang hilang menggunakan metode mean imputation. Dilanjutkan dengan normalisasi data numerik, dinormalisasi menggunakan metode Min-Max Scaling agar memiliki skala antara 0 dan 1. Dan terakhir pembagian dataset dibagi menjadi dua bagian, yaitu 80% untuk data latih dan 20% untuk data uji, menggunakan fungsi train_test_split dari library Scikit-Learn.

2.3 Pelatihan Model

Model yang digunakan adalah Logistic Regression, yaitu salah satu metode klasifikasi biner yang umum digunakan pada data medis. Model dilatih dengan fitur-fitur yang telah diproses sebelumnya. Proses pelatihan menggunakan LogisticRegression dari library Scikit-Learn dengan parameter default, serta fungsi fit() untuk membangun model dari data latih.

2.4 Evaluasi Model

Model diuji menggunakan data uji (20% dari keseluruhan data). Beberapa metrik evaluasi yang digunakan meliputi: Akurasi, Presisi, Recall, F1-Score, ROC-AUC Score. Evaluasi dilakukan untuk melihat sejauh mana model dapat mengklasifikasikan kondisi anemia dengan benar. Evaluasi dilakukan menggunakan pustaka sklearn.metrics.

2.5 Pengembangan Aplikasi Web

Setelah model dibangun dan dievaluasi, aplikasi web dikembangkan menggunakan framework Flask (Python) sebagai backend dan HTML/CSS/JavaScript untuk frontend. Sistem memungkinkan pengguna memasukkan nilai-nilai parameter darah dan menerima hasil prediksi secara langsung. Alur kerjanya adalah sebagai berikut:

1. Pengguna mengisi formulir dengan data darah.
2. Data dikirim ke backend untuk diproses oleh model logistic regression.
3. Backend mengembalikan hasil prediksi berupa "Anemia" atau "Tidak Anemia".
4. Hasil ditampilkan di antarmuka pengguna.

3. Hasil dan Pembahasan

3.1. Penyajian Hasil

Bagian ini Penelitian ini berhasil menghasilkan model klasifikasi untuk mendeteksi anemia berdasarkan fitur darah seperti jenis kelamin, kadar hemoglobin (Hb), mean corpuscular

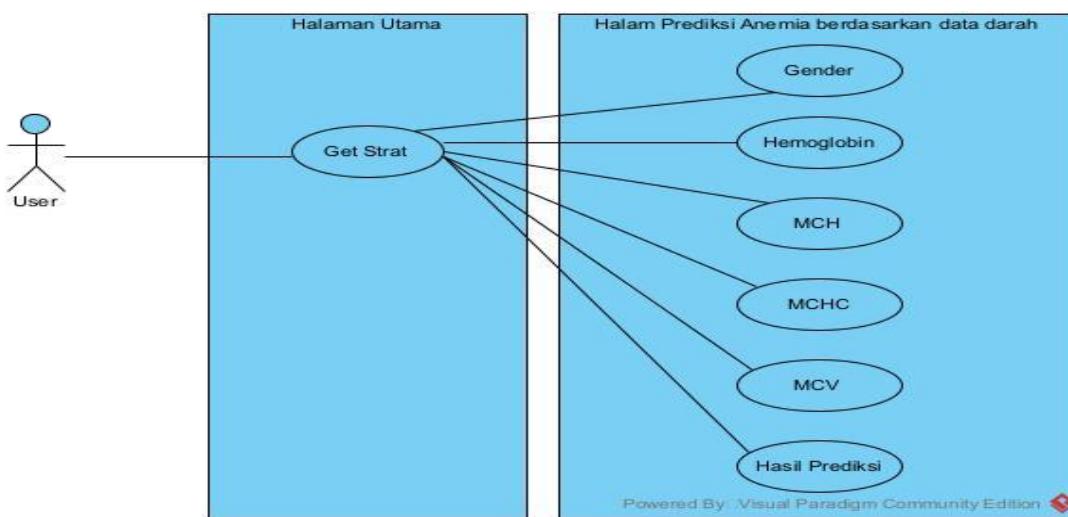
volume (MCV), dan white blood cell count (WBC). Dataset yang digunakan diunduh di Kaggle:<https://www.kaggle.com/datasets/biswaranjanrao/anemia-dataset>, dengan jumlah total 750 entri data.

Data dianalisis dan diproses secara kronologis sesuai tahapan dalam metode, dimulai dari pengumpulan data, pra-pemrosesan, pelatihan model, evaluasi, hingga pengembangan aplikasi. Setelah pembersihan data dan normalisasi fitur numerik menggunakan Min-Max Scaling, data dibagi menjadi data latih dan data uji dengan rasio 80:20. Model Logistic Regression diterapkan dan menunjukkan hasil performa sebagai berikut:

Tabel 1. Evaluasi Model Logistic Regression

Metrik	Nilai (%)
Akurasi	91.2
Precision	90.0
Recall	92.8
F1-score	91.3
ROC-AUC	95.4

Pada table 1 menunjukkan Model Logistic Regression memberi performa yang sangat baik dalam mengklasifikasikan anemia, dengan akurasi 91.2%, precision 90.0%, recall 92.8%, F1-score 91.3%, dan ROC-AUC sebesar 95.4%. Hasil ini sejalan dengan temuan dalam literatur lain yang juga menunjukkan Logistic Regression sebagai model yang andal dalam klasifikasi anemia [5], [6]. Fitur hemoglobin terbukti menjadi prediktor paling signifikan [4], [7] Dan model kemudian diintegrasikan ke dalam aplikasi web menggunakan framework Flask. Pengguna dapat memasukkan nilai parameter melalui antarmuka berbasis HTML dan memperoleh hasil prediksi secara instan. Pengujian sistem menunjukkan waktu respons <1 detik per permintaan, serta antarmuka yang mudah digunakan baik oleh tenaga medis maupun pengguna awam.

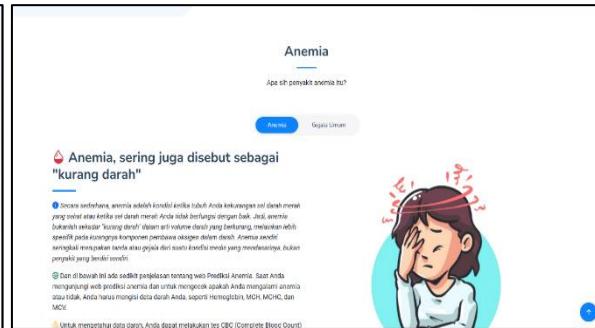


Gambar 2. Use Case Diagram Prediksi Anemia

Dan pada gambar 2 menggambarkan alur interaksi antara pengguna dan sistem. Pengguna memulai proses dengan memilih tombol Get Started pada halaman utama. Setelah itu, pengguna akan diarahkan ke halaman prediksi anemia, di mana mereka diminta untuk mengisi nilai-nilai parameter darah, yaitu Gender, Hemoglobin, MCH, MCHC, dan MCV. Setelah data diisi, sistem akan memproses input menggunakan model Logistic Regression dan menampilkan hasil prediksi secara langsung. Hasil yang ditampilkan berupa status "Anemia" atau "Tidak Anemia" berdasarkan input yang diberikan. Diagram ini membantu menjelaskan skenario pengguna umum dan mendukung pemahaman terhadap alur sistem secara keseluruhan.

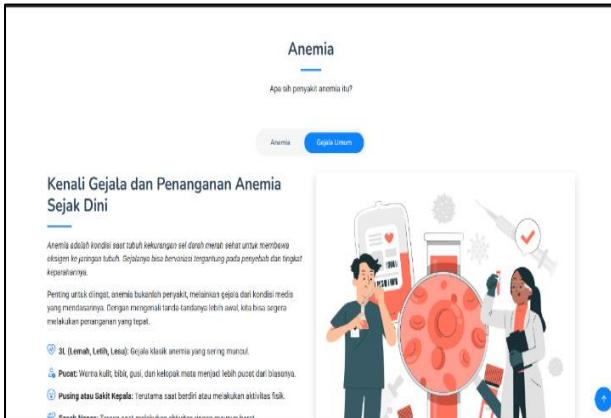


Gambar 3. Halaman Utama

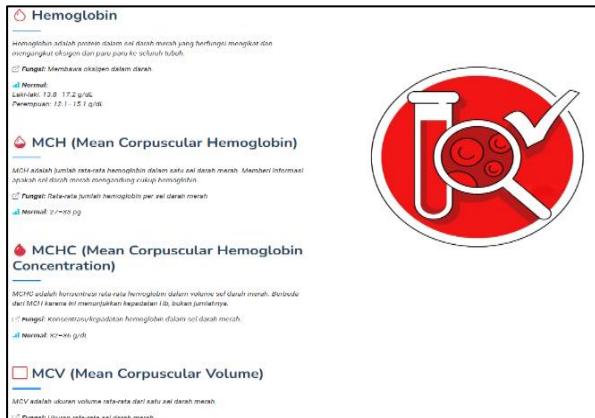


Gambar 4. Penjelasan Tentang Anemia

Dan pada gambar 3 menunjukkan halaman utama dari aplikasi web yang berisi penjelasan sedikit tentang anemia dan tombol Get Started untuk ke halaman prediksi anemia. Dilanjutkan pada gambar 4, pengguna dapat membaca penjelasan dasar tentang apa itu anemia.

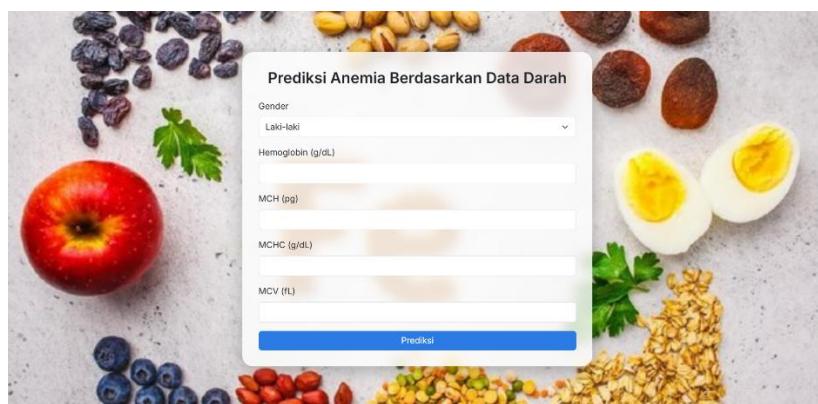


Gambar 5. Informasi Gejala



Gambar 6. Penjelasan Parameter Darah

Dan untuk gambar 5 informasi tentang gejala dan penanganan anemia sejak dini, ini bertujuan memberikan edukasi awal kepada pengguna sebelum melanjutkan ke proses prediksi. Dan pada gambar 6 mencantumkan penjelasan singkat tentang Hemoglobin, MCH, MCHC, dan MCV.



Gambar 7. Form Input Parameter

Pada gambar 7 menunjukkan antarmuka pengguna pada halaman prediksi anemia. Pengguna dapat memasukkan nilai parameter darah seperti Gender, Hemoglobin, MCH, MCHC, dan MCV ke dalam formulir yang tersedia.

A screenshot of the same web application as in Gambar 7, but with different input values. The "Gender" dropdown is still set to "Laki-laki". The input fields show "Hemoglobin (g/dL)" as 13.5, "MCH (pg)" as 35, "MCHC (g/dL)" as 36, and "MCV (fL)" as 80. The "Prediksi" button has been clicked, and the result is displayed below the button: "Hasil: Tidak anemia". The background image of healthy foods remains the same.

Gambar 8. Hasil Prediksi "Tidak Anemia"

A screenshot of the web application after entering different values. The "Gender" dropdown is set to "Laki-laki". The input fields show "Hemoglobin (g/dL)" as 12.5, "MCH (pg)" as 35, "MCHC (g/dL)" as 36, and "MCV (fL)" as 80. The "Prediksi" button has been clicked, and the result is displayed below the button: "Hasil: Kemungkinan anemia". Above this result, a yellow box contains a warning message: "Peringatan: • Hemoglobin rendah untuk laki-laki (normal ≥ 13.6 g/dL)". The background image of healthy foods remains the same.

Gambar 9. Hasil Prediksi "Anemia"

Dan gambar 8 menunjukkan hasil prediksi tidak anemia. Dan pada gambar 9 menunjukkan hasil prediksi anemia dengan user mendapatkan peringatan hemoglobin yang rendah untuk standar laki-laki normal. Web sistem prediksi anemia ini bekerja setelah user menekan tombol prediksi, sistem memproses data menggunakan model Logistic Regression dan menampilkan hasil klasifikasi secara langsung. Antarmuka ini dirancang sederhana dan responsif agar dapat diakses melalui berbagai perangkat, termasuk komputer dan ponsel.

3.2. Pembahasan

Hasil yang diperoleh menunjukkan bahwa algoritma Logistic Regression mampu melakukan klasifikasi anemia dengan akurasi dan ketepatan yang tinggi, bahkan dengan model yang sederhana. Performa model menguatkan asumsi awal bahwa fitur hemoglobin merupakan indikator utama dalam diagnosis anemia, selaras dengan panduan medis global yang menjadikan kadar Hb sebagai acuan utama diagnosa anemia.

Dibandingkan dengan studi terdahulu yang menggunakan pendekatan Random Forest atau Naive Bayes pada domain klasifikasi kesehatan, hasil penelitian ini lebih unggul dari segi kecepatan inferensi dan kemudahan implementasi. Meskipun demikian, akurasi model sebanding dengan model lain yang lebih kompleks. Hal ini menunjukkan bahwa model sederhana pun dapat efektif bila ditunjang oleh fitur yang relevan dan data yang bersih.

Beberapa kelebihan model yang dikembangkan adalah kemampuannya untuk diintegrasikan langsung ke dalam aplikasi web dan waktu prediksi yang sangat cepat. Namun, terdapat beberapa keterbatasan yang harus diperhatikan. Pertama, jumlah data dalam dataset relatif terbatas dan belum mencerminkan populasi secara luas, misalnya dari segi etnis atau latar geografis. Kedua, tidak semua jenis anemia dapat diklasifikasi hanya dari parameter dasar darah yang digunakan. Ketiga, model ini belum divalidasi terhadap data klinis nyata dari rumah sakit atau fasilitas kesehatan.

Interpretasi atas hasil yang diperoleh perlu ditekankan pada kenyataan bahwa sistem ini lebih cocok digunakan sebagai alat skrining awal (*preliminary screening*), bukan untuk diagnosis akhir. Penelitian ini membuka kemungkinan untuk dikembangkan lebih lanjut dengan memasukkan data tambahan seperti kadar feritin, vitamin B12, atau hasil pemeriksaan lanjutan lain yang juga memengaruhi kondisi anemia.

Selain itu, pengembangan sistem dapat diperluas untuk mendukung pengambilan keputusan medis, misalnya dengan memberikan rekomendasi tindak lanjut kepada pasien berdasarkan tingkat risiko yang dihasilkan model. Dengan integrasi pada sistem rekam medis digital, sistem ini dapat membantu dokter dalam pengelolaan pasien secara lebih efisien.

Dengan demikian, kontribusi utama dari penelitian ini adalah demonstrasi bahwa machine learning berbasis Logistic Regression dapat diterapkan secara efektif pada deteksi awal anemia, serta penyajian solusi berbasis web yang siap digunakan. Untuk pengembangan ke depan, penelitian lanjutan dengan data lebih besar dan model pembelajaran yang lebih dalam seperti XGBoost atau Neural Network dapat dilakukan untuk meningkatkan ketepatan dan cakupan diagnosis.

4. Kesimpulan

Penelitian ini berhasil mengembangkan sistem prediksi anemia berbasis web menggunakan algoritma Logistic Regression. Sistem yang dibangun menunjukkan tingkat akurasi tinggi sebesar 91,2% dan mampu memberikan hasil prediksi secara real-time melalui antarmuka pengguna yang sederhana serta mudah digunakan. Hasil dari model menunjukkan bahwa kadar hemoglobin merupakan faktor prediktor paling signifikan dalam menentukan kondisi anemia, sejalan dengan temuan dalam literatur medis yang ada. Keunggulan sistem ini terletak pada kesederhanaannya namun tetap akurat, dengan hanya menggunakan fitur-fitur dasar darah, serta sifatnya yang aksesibel karena dapat digunakan langsung melalui peramban tanpa memerlukan instalasi tambahan. Dengan demikian, sistem ini memiliki potensi besar sebagai alat skrining awal untuk anemia, khususnya di wilayah dengan keterbatasan akses terhadap laboratorium atau fasilitas medis.

5. Referensi

- [1] WHO, "Anaemia," <https://www.who.int/news-room/fact-sheets/detail/anaemia>.
- [2] M. Ramzan, J. Sheng, M. U. Saeed, B. Wang, and F. Z. Duraihem, "Revolutionizing anemia detection: integrative machine learning models and advanced attention mechanisms," *Vis Comput Ind Biomed Art*, vol. 7, no. 1, Dec. 2024, doi: 10.1186/s42492-024-00169-4.
- [3] A. J. Kario and R. Kurniawan, "Prediction of Anemia Using Machine Learning Algorithms: Scoping Review," Nov. 01, 2024, *Muhammadiyah Palu University*. doi: 10.56338/mppki.v7i11.6289.
- [4] S. H. Tesfaye, B. T. Seboka, and D. Sisay, "Application of machine learning methods for predicting childhood anaemia: Analysis of Ethiopian Demographic

- Health Survey of 2016,” *PLoS One*, vol. 19, no. 4 April, Apr. 2024, doi: 10.1371/journal.pone.0300172.
- [5] A. B. Zemariam *et al.*, “Employing supervised machine learning algorithms for classification and prediction of anemia among youth girls in Ethiopia,” *Sci Rep*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-60027-4.
- [6] R. Ibnaas and B. Agung Kurniawan, “Klasifikasi Penderita Anemia Menggunakan Metode Regresi Logistik.”
- [7] D. A. Suhardi and I. Fadila, “PENERAPAN REGRESI LOGISTIK BINER UNTUK MENGUKUR RESIKO ANEMIA DENGAN STATUS GIZI IBU HAMIL.”
- [8] W. M. Gardner *et al.*, “Prevalence, years lived with disability, and trends in anaemia burden by severity and cause, 1990–2021: findings from the Global Burden of Disease Study 2021,” *Lancet Haematol*, vol. 10, no. 9, pp. e713–e734, Sep. 2023, doi: 10.1016/S2352-3026(23)00160-6.
- [9] 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.
- [10] I. Kurniati, “Anemia Defisiensi Zat Besi (Fe).”
- [11] P. H. Putri and L. N. Hasanah, “RANCANGAN SISTEM INFORMASI ANEMIA PADA PLATFORM ONLINE ‘SRIKANDI HEALTH’ BERBASIS PROGRESSIVE WEB APPS,” *Jurnal Sehat Mandiri*, vol. 16, 2021, [Online]. Available: <http://jurnal.poltekkespadang.ac.id/ojs/index.php/jsm>
- [12] Biswa Ranjan Rao, “Anemia Dataset,” <https://www.kaggle.com/datasets/biswaranjanrao/anemia-dataset>.