



Pengembangan Model Prediktif Multilevel untuk Determinan Anemia pada Ibu Hamil Berbasis Big Data Kesehatan

Asyima^{1*}, Ruqaiyah², Masriadi³, Hukmiyah Aspar⁴, Ikrawanty Ayu Wulandari⁵, Nurasia Natsir⁶

¹⁻² Institut Ilmu Kesehatan Pelamonia, Makassar, Indonesia

³ Universitas Muslim Indonesia, Makassar, Indonesia

⁴⁻⁶ Institut Ilmu Kesehatan Pelamonia, Makassar, Indonesia

* Korespondensi: asyima@iikpelamonia.ac.id

Abstract. Anemia during pregnancy remains a major public health challenge in Indonesia due to its contribution to maternal morbidity, adverse birth outcomes, and neonatal mortality. This study aimed to develop and validate a multilevel predictive model to identify determinants of anemia among pregnant women using healthcare big data collected from primary healthcare centers in South Sulawesi Province. Secondary data were obtained from the national Maternal and Child Health information system, the e-Cohort application, and electronic medical records from 2020 to 2023. The dataset included 12,847 antenatal visits from 4,312 pregnant women across 87 primary healthcare centers. Data analysis employed multilevel logistic regression and machine learning algorithms, including Random Forest and Gradient Boosting. The findings revealed that non-adherence to iron supplementation, chronic energy deficiency, and high parity were the strongest individual determinants of anemia during pregnancy. In addition, higher quality antenatal care programs at the healthcare facility level demonstrated a protective effect against anemia. Intra-class correlation analysis confirmed significant clustering of anemia prevalence at the facility level. Among all predictive approaches, the Gradient Boosting model achieved the highest predictive performance. These findings highlight the importance of integrating individual and contextual healthcare system factors into predictive modeling to support more targeted and effective anemia prevention strategies for pregnant women in Indonesia.

Keywords: Healthcare Big Data; Machine Learning; Maternal Health; Multilevel Model; Pregnancy Anemia

Abstrak: Anemia pada kehamilan masih menjadi tantangan kesehatan masyarakat di Indonesia karena berkontribusi terhadap meningkatnya morbiditas maternal, luaran persalinan buruk, dan mortalitas neonatal. Penelitian ini bertujuan mengembangkan dan memvalidasi model prediktif multilevel untuk menentukan faktor-faktor yang memengaruhi anemia pada ibu hamil dengan memanfaatkan big data kesehatan dari berbagai Puskesmas di Provinsi Sulawesi Selatan. Penelitian menggunakan data sekunder dari sistem informasi Kesehatan Ibu dan Anak, aplikasi e-Kohort, dan rekam medis elektronik selama periode 2020–2023. Dataset mencakup 12.847 kunjungan antenatal dari 4.312 ibu hamil di 87 Puskesmas. Analisis dilakukan menggunakan regresi logistik multilevel serta algoritma *machine learning* berupa *Random Forest* dan *Gradient Boosting*. Hasil penelitian menunjukkan bahwa ketidakpatuhan konsumsi tablet besi, kekurangan energi kronis, dan paritas tinggi merupakan determinan individu utama anemia pada kehamilan. Selain itu, kualitas program antenatal care di tingkat fasilitas kesehatan bersifat protektif terhadap kejadian anemia. Analisis *intra-class correlation* menunjukkan adanya pengelompokan risiko anemia pada tingkat fasilitas kesehatan. *Model Gradient Boosting* menghasilkan performa prediksi terbaik dibandingkan model lainnya. Temuan ini menegaskan pentingnya integrasi faktor individu dan kontekstual sistem kesehatan dalam pengembangan strategi pencegahan anemia pada ibu hamil di Indonesia secara lebih terarah dan efektif.

Kata Kunci: Anemia Kehamilan; Big Data Kesehatan; Kesehatan Maternal; Machine Learning; Model Multilevel

1. LATAR BELAKANG

Anemia pada kehamilan, yang didefinisikan oleh *World Health Organization* (WHO) sebagai konsentrasi hemoglobin di bawah 11 g/dL pada trimester pertama dan ketiga atau di bawah 10,5 g/dL pada trimester kedua, merupakan salah satu gangguan gizi yang paling prevalensi dan secara klinis paling bermakna yang mempengaruhi wanita usia reproduksi secara global (WHO, 2023). Di seluruh dunia, diperkirakan 36,5% ibu hamil mengalami

anemia, dengan beban yang secara tidak proporsional terkonsentrasi di negara-negara berpenghasilan rendah dan menengah (Stevens et al., 2022). Di Indonesia, data nasional dari Survei Kesehatan Dasar (Riskesdas) menunjukkan bahwa prevalensi anemia pada ibu hamil mencapai 48,9% pada tahun 2018, menempatkan Indonesia di antara negara-negara dengan beban anemia tertinggi di Asia Tenggara beban yang terbukti sangat resisten terhadap intervensi kebijakan meskipun telah berlangsung selama beberapa dekade program suplementasi besi (Kementerian Kesehatan Republik Indonesia, 2018). Konsekuensi klinis dan kesehatan masyarakat dari anemia pada kehamilan sangat besar: anemia dikaitkan dengan peningkatan risiko perdarahan postpartum, sepsis nifas, dan kematian maternal yang signifikan, serta hambatan pertumbuhan intrauterin, kelahiran prematur, berat badan lahir rendah, dan gangguan perkembangan saraf pada masa kanak-kanak dini (Daru et al., 2018; Haider et al., 2013).

Meskipun terdapat banyak literatur tentang determinan anemia pada kehamilan, beberapa keterbatasan metodologis kritis telah membatasi utilitas transisi dari bukti yang ada. Pertama, penggunaan model regresi logistik konvensional yang dominan memperlakukan faktor risiko individual secara terpisah, mengabaikan sifat hierarkis dan berkelompok dari data kesehatan di mana wanita individual tersarang dalam fasilitas kesehatan, yang pada gilirannya tersarang dalam konteks geografis dan administratif yang lebih luas (Goldstein, 2011). Keterbatasan analitis ini mengaburkan determinan kontekstual yang beroperasi di tingkat fasilitas dan kabupaten yang mungkin secara substansial memodifikasi risiko individu. Kedua, proliferasi sistem kesehatan digital di Indonesia—termasuk aplikasi kesehatan ibu e-Kohort, Sistem Informasi Manajemen Puskesmas (SIMPUS), dan platform rekam medis elektronik nasional—telah menghasilkan volume data kesehatan yang dikumpulkan secara rutin yang sebagian besar belum dimanfaatkan untuk pemodelan epidemiologis dan analitik prediktif (Kementerian Kesehatan RI, 2022). Ketiga, konteks Sulawesi Selatan sangat kurang terwakili dalam literatur yang diterbitkan meskipun memiliki geografi yang heterogen dan beban anemia yang terdokumentasi tinggi, yang membatasi spesifisitas kontekstual dari model prediktif yang ada.

Kemunculan analitik *big data* dan *machine learning* dalam epidemiologi menawarkan peluang transformatif untuk melampaui keterbatasan tersebut. Kerangka pemodelan multilevel, bila dikombinasikan dengan volume dan granularitas data kesehatan yang dikumpulkan secara rutin, memungkinkan estimasi simultan dari faktor risiko individual dan kontekstual sembari memperhitungkan non-independensi observasi dalam fasilitas yang sama (Merlo et al., 2016). Selain itu, algoritma *machine learning* seperti *Random Forest* dan *Gradient Boosting* telah

menunjukkan kinerja prediktif yang superior dibandingkan regresi konvensional dalam dataset kesehatan yang kompleks dan berdimensi tinggi dengan menangkap hubungan non-linear dan efek interaksi (Obermeyer & Emanuel, 2016). Berdasarkan latar belakang tersebut, penelitian ini mengejar tiga tujuan: (1) mengidentifikasi determinan anemia di tingkat individu dan fasilitas menggunakan regresi logistik *multilevel*; (2) mengkuantifikasi proporsi varians prevalensi anemia yang dapat diatribusikan kepada faktor kontekstual tingkat fasilitas melalui analisis korelasi *intraclass*; dan (3) mengevaluasi kinerja komparatif algoritma *machine learning versus model regresi multilevel* untuk prediksi anemia.

2. TINJAUAN PUSTAKA

Epidemiologi dan Etiologi Anemia pada Kehamilan

Anemia pada kehamilan secara etiologis bersifat heterogen, timbul dari interaksi kompleks faktor gizi, infeksi, genetik, dan sosio-demografis. Defisiensi besi sejauh ini merupakan penyebab paling umum, menyumbang sekitar 50% dari semua kasus secara global, diikuti oleh defisiensi folat, defisiensi vitamin B12, malaria, hemoglobinopati, dan infeksi kronis (WHO, 2023). Dalam konteks Indonesia, anemia defisiensi besi mendominasi tetapi seringkali hidup berdampingan dengan kekurangan energi kronis, keragaman pangan yang rendah, dan infeksi cacing di lingkungan pedesaan, menciptakan pola anemia gizi multifaktorial yang membuat strategi suplementasi nutrisi tunggal menjadi tidak memadai (Sumarmi, 2017). Determinan tingkat individu yang konsisten teridentifikasi dalam literatur meliputi usia ibu yang muda, paritas tinggi, interval kehamilan yang pendek, status sosial-ekonomi rendah, pendidikan formal terbatas, asupan besi makanan yang tidak adekuat, ketidakpatuhan suplementasi besi-asam folat (IFA), dan pemanfaatan antenatal care yang tidak mencukupi (Fatimah et al., 2019; Gebre & Mulugeta, 2015). Kumpulan bukti yang terus berkembang dari studi multilevel yang dilakukan di Afrika Sub-Sahara dan Asia Selatan menunjukkan bahwa faktor tingkat fasilitas dan komunitas termasuk kualitas layanan antenatal, kompetensi tenaga kesehatan, dan aksesibilitas geografis memberikan efek independen terhadap prevalensi anemia yang tidak dapat sepenuhnya dijelaskan oleh kovariat tingkat individu saja (Aheto, 2019; Akombi et al., 2017).

Analisis Multilevel dalam Penelitian Kesehatan Maternal

Pemodelan multilevel menyediakan kerangka statistik yang ketat untuk menganalisis data dengan struktur hierarkis yang melekat di mana unit-unit tingkat bawah (wanita individual) tersarang dalam unit-unit tingkat lebih tinggi (fasilitas kesehatan atau distrik geografis) (Raudenbush & Bryk, 2002). Dalam konteks penelitian kesehatan maternal,

pendekatan ini secara metodologis imperatif karena wanita yang menerima perawatan antenatal dalam Puskesmas yang sama berbagi paparan umum terkait kualitas, cakupan, dan konsistensi pemberian layanan kesehatan, yang melanggar asumsi independensi regresi konvensional. Koefisien korelasi *intraclass* (ICC) mengkuantifikasi derajat kesamaan di antara observasi dalam kelompok yang sama; ICC yang melebihi 0,05 umumnya dianggap cukup untuk membenarkan pemodelan *multilevel* (Merlo et al., 2005). Studi yang menerapkan regresi logistik *multilevel* untuk anemia pada kehamilan di negara-negara berpenghasilan rendah dan menengah melaporkan nilai ICC berkisar dari 0,10 hingga 0,25, mengkonfirmasi pengelompokan substansial di tingkat fasilitas dan komunitas (Aheto, 2019; Nguyen et al., 2018).

Big Data Kesehatan dan Machine Learning dalam Prediksi Anemia

Big data kesehatan mencakup dataset besar, kompleks, dan heterogen yang dihasilkan dari rekam medis elektronik, sistem informasi kesehatan, basis data laboratorium, aplikasi kesehatan mobile, dan registri kesehatan populasi (Raghupathi & Raghupathi, 2014). Di Indonesia, peluncuran progresif aplikasi *e-Kohort* telah mengkatalisasi digitalisasi data kesehatan maternal, menghasilkan catatan berbasis fasilitas yang bersifat longitudinal untuk wanita individual dalam skala besar (Kementerian Kesehatan RI, 2022). Algoritma machine learning telah menunjukkan potensi yang sangat menjanjikan untuk tugas prediksi anemia. Metode ensemble berbasis pohon Keputusan termasuk *Random Forest* (Breiman, 2001) dan *Gradient Boosting* (Friedman, 2001) sangat cocok untuk dataset kesehatan yang dicirikan oleh non-linearitas, interaksi kompleks, dan pola data yang hilang. Xu et al. (2021) menunjukkan bahwa *Random Forest* mengungguli regresi logistik konvensional dalam memprediksi anemia defisiensi besi dalam dataset rekam medis elektronik yang besar, sementara Hasan et al. (2020) memperlihatkan sensitivitas *superior Gradient Boosting* dalam mengidentifikasi ibu hamil berisiko tinggi.

Kesenjangan Penelitian dan Justifikasi Studi

Tinjauan kritis terhadap literatur yang ada mengungkapkan tiga kesenjangan persisten yang ingin diatasi oleh penelitian ini. Pertama, tidak ada studi yang dipublikasikan yang mengintegrasikan regresi *multilevel* dengan *machine learning* dalam kerangka analitis terpadu untuk memodelkan determinan anemia di kalangan ibu hamil Indonesia. Kedua, data yang dikumpulkan secara rutin dari *e-Kohort* dan SIMPUS belum dimanfaatkan secara sistematis untuk pengembangan model prediktif yang menargetkan pencegahan anemia pada skala besar. Ketiga, Sulawesi Selatan dengan geografi heterogen yang mencakup daerah pantai urban, semi-urban, dan dataran tinggi terpencil yang menciptakan variasi signifikan dalam akses layanan

kesehatan dan kualitas pelayanan secara substansial kurang terwakili dalam literatur epidemiologis. Kesenjangan-kesenjangan ini secara kolektif membenarkan kebaruan dan signifikansi kesehatan masyarakat dari penelitian ini.

3. METODE PENELITIAN

Desain dan Setting Penelitian

Penelitian ini mengadopsi desain potong lintang retrospektif yang memanfaatkan data kesehatan sekunder yang dikumpulkan secara rutin dari 87 Puskesmas di 24 kabupaten dan kota dalam Provinsi Sulawesi Selatan, Indonesia. Fasilitas dipilih untuk mewakili spektrum penuh klasifikasi geografis (urban, semi-urban, pedesaan, dan terpencil) serta kategori kinerja fasilitas. Sulawesi Selatan dipilih berdasarkan prevalensi anemia yang tinggi yang terdokumentasi pada ibu hamil (52,3% berdasarkan Riskesdas 2018), infrastruktur kesehatan digital yang relatif maju yang memungkinkan ekstraksi data e-Kohort yang sistematis, dan heterogenitas geografis serta sosiodemografis yang memaksimalkan kekuatan analitis dari pendekatan pemodelan multilevel.

Sumber Data dan Integrasi

Big data kesehatan diintegrasikan dari tiga sumber kesehatan digital yang saling melengkapi. Sumber utama adalah aplikasi pengawasan kesehatan maternal e-Kohort, yang menyediakan catatan kunjungan antenatal longitudinal termasuk pengukuran hemoglobin, data antropometri, usia kehamilan, paritas, catatan suplementasi besi-asam folat, dan dokumentasi komorbiditas. Sumber ini dilengkapi dengan SIMPUS, yang menyediakan data administratif tingkat fasilitas termasuk indikator cakupan program, skor penilaian kualitas ANC, dan rasio tenaga kesehatan per penduduk. Sumber ketiga adalah Sistem Informasi Kesehatan Daerah (SIKDA) yang dikelola oleh Dinas Kesehatan Provinsi Sulawesi Selatan, yang menyediakan indikator kontekstual tingkat kabupaten. Integrasi data dilakukan melalui *probabilistic record linkage* menggunakan nomor identifikasi maternal yang unik. Nilai yang hilang diimputasi menggunakan imputasi berganda dengan persamaan berantai (MICE) untuk variabel dengan *missingness* di bawah 20%. Dataset analitik *final* terdiri dari 12.847 catatan kunjungan *antenatal* dari 4.312 ibu hamil yang unik, mencakup periode Januari 2020 hingga Desember 2023.

Definisi Variabel

Variabel Hasil

Anemia pada kehamilan didefinisikan sebagai hemoglobin di bawah 11 g/dL, sesuai kriteria diagnostik WHO (WHO, 2023). Klasifikasi *biner* (anemia vs. tidak anemia) diterapkan pada model regresi primer; klasifikasi keparahan (ringan: Hb 10,0–10,9 g/dL; sedang: 7,0–9,9 g/dL; berat: <7,0 g/dL) digunakan secara deskriptif.

Variabel Individu Level-1

Prediktor tingkat individu meliputi: usia ibu (≤ 19 , 20–34, ≥ 35 tahun); paritas (0, 1–3, ≥ 4); trimester kehamilan pada pengukuran Hb pertama; status gizi berdasarkan lingkaran lengan atas (LILA; <23,5 cm = kekurangan energi kronis); kepatuhan suplementasi besi-asam folat (patuh: ≥ 90 tablet; tidak patuh: <90 tablet); frekuensi kunjungan antenatal (adekuat: ≥ 4 ; tidak adekuat: <4); dan komorbiditas penyakit kronis (hipertensi, diabetes mellitus, atau penyakit ginjal kronis).

Variabel Fasilitas Level-2

Prediktor tingkat fasilitas meliputi: klasifikasi geografis Puskesmas (urban, semi-urban, pedesaan, terpencil); rasio tenaga kesehatan per penduduk; skor implementasi kualitas program ANC (indeks komposit 0–100); dan tingkat cakupan program suplementasi IFA tahunan.

Analisis Statistik

Analisis dilaksanakan dalam tiga tahap berurutan. Pertama, statistik deskriptif dihitung untuk semua variabel, distratifikasi berdasarkan status anemia. Kedua, model regresi logistik multilevel dua level dibangun menggunakan estimasi maximum likelihood dengan kuadratur *Gauss-Hermite* adaptif. Tiga model bertingkat diestimasi: Model 0 (null model) untuk mengestimasi ICC dan mengkonfirmasi perlunya pemodelan multilevel; Model 1 (prediktor tingkat individu saja); dan Model 2 (model penuh dengan kedua level dan istilah interaksi lintas level). Kesesuaian model dinilai menggunakan AIC, BIC, dan uji rasio kemungkinan. ICC dikalkulasi dari komponen varians null model sebagai $ICC = \tau^2 / (\tau^2 + \pi^2/3)$, di mana τ^2 adalah varians antar-fasilitas. Ketiga, tiga algoritma machine learning dilatih dan dievaluasi melalui validasi silang berlapis lima terstratifikasi: Regresi Logistik Konvensional (baseline), *Random Forest*, dan *Gradient Boosting*. Kepentingan fitur dinilai menggunakan nilai SHAP untuk model berkinerja terbaik. Diskriminasi dievaluasi melalui AUC-ROC; kalibrasi melalui uji *Hosmer-Lemeshow*. Semua analisis dilakukan dalam R versi 4.3.1. Signifikansi statistik ditetapkan pada $\alpha = 0,05$.

4. HASIL DAN PEMBAHASAN

Statistik Deskriptif dan Prevalensi Anemia

Dari 4.312 ibu hamil yang dimasukkan dalam analisis, 2.203 (51,1%) diklasifikasikan sebagai anemia (Hb <11 g/dL). Di antara wanita anemia, 38,4% mengalami anemia ringan, 55,7% anemia sedang, dan 5,9% anemia berat. Rerata hemoglobin adalah 9,84 g/dL (SD = 1,12) di antara wanita anemia *versus* 11,87 g/dL (SD = 0,63) di antara wanita tidak anemia. Prevalensi anemia tingkat fasilitas berkisar dari 31,2% hingga 74,8% di 87 Puskesmas, mencerminkan heterogenitas geografis yang substansial. Ketidakpatuhan IFA secara nyata lebih prevalen di antara wanita anemia (79,0%) dibandingkan wanita tidak anemia (29,7%), demikian pula kekurangan energi kronis (40,6% vs. 13,4%), paritas tinggi (30,5% vs. 8,7%), dan kunjungan ANC tidak adekuat (51,4% vs. 23,6%), semuanya signifikan secara statistik pada $p < 0,001$.

Tabel 1. Karakteristik Populasi Studi yang Distratifikasi berdasarkan Status Anemia (N = 4.312)

Karakteristik	Anemia n=2.203 (%)	Tidak Anemia n=2.109 (%)	p-value
Ketidakpatuhan IFA	1.741 (79,0)	627 (29,7)	<0,001
Kekurangan Energi Kronis (LILA <23,5 cm)	894 (40,6)	283 (13,4)	<0,001
Paritas Tinggi (≥ 4 kelahiran)	672 (30,5)	184 (8,7)	<0,001
Komorbidity Penyakit Kronis	437 (19,8)	148 (7,0)	<0,001
Kunjungan ANC Tidak Adekuat (<4)	1.132 (51,4)	498 (23,6)	<0,001

ANC = Antenatal Care; IFA = Iron-Folic Acid (Besi-Asam Folat); LILA = Lingkar Lengan Atas. Uji chi-square diterapkan untuk semua perbandingan kategoris.

Korelasi Intraclass dan Partisi Varians

Null model multilevel menghasilkan varians antar-fasilitas sebesar $\tau^2 = 0,641$ (SE = 0,093), yang berkorespondensi dengan ICC sebesar 0,163 (16,3%). ICC ini jelas melampaui ambang batas konvensional 0,05, memberikan justifikasi statistik dan substantif yang kuat untuk kerangka analitis multilevel. Setelah dimasukkannya kovariat tingkat individu dalam Model 1, varians antar-fasilitas sebagian teratenuasi ($\tau^2 = 0,487$; ICC = 0,129), mengindikasikan bahwa sekitar 20,9% varians antar-fasilitas dapat diatribusikan kepada perbedaan komposisional dalam faktor risiko individu di seluruh Puskesmas. Pada Model 2 penuh dengan prediktor tingkat fasilitas yang ditambahkan, varians antar-fasilitas berkurang lebih lanjut ($\tau^2 = 0,312$; ICC = 0,087), mengkonfirmasi bahwa variabel tingkat fasilitas secara kolektif menjelaskan 35,9% varians kontekstual tambahan. *ICC residual* sebesar 8,7% persisten, yang dapat diatribusikan kepada faktor kontekstual yang tidak terukur. Temuan ini

menunjukkan bahwa konteks Puskesmas dalam pemberian layanan secara independen dan substansial terkait dengan risiko anemia.

Hasil Model Prediktif Multilevel

Tabel 2 menyajikan *adjusted odds ratio* dari model regresi logistik multilevel penuh. Di tingkat individu, ketidakpatuhan suplementasi IFA merupakan prediktor anemia terkuat (aOR = 3,21; 95% CI: 2,67–3,86; p <0,001), konsisten dengan defisiensi besi sebagai etiologi primer anemia dalam populasi ini. Kekurangan energi kronis (aOR = 2,84; 95% CI: 2,31–3,49; p <0,001) dan paritas tinggi (aOR = 2,17; 95% CI: 1,73–2,72; p <0,001) menempati peringkat kedua dan ketiga, masing-masing. Temuan paritas tinggi mencerminkan deplesi progresif cadangan besi di seluruh kehamilan berturut-turut, terutama dalam konteks interval kehamilan pendek yang umum di populasi pedesaan Sulawesi Selatan (Gebre & Mulugeta, 2015). Komorbiditas penyakit kronis (aOR = 1,89; 95% CI: 1,41–2,54), kunjungan ANC tidak adekuat (aOR = 1,78; 95% CI: 1,43–2,21), dan usia maternal remaja (aOR = 1,64; 95% CI: 1,28–2,09) maupun lanjut (aOR = 1,52; 95% CI: 1,21–1,91) juga secara independen signifikan. Di tingkat fasilitas, skor implementasi kualitas program ANC yang lebih tinggi secara signifikan bersifat protektif (aOR = 0,61 per kenaikan 10 poin; 95% CI: 0,48–0,78; p <0,001), menggarisbawahi peran kritis kualitas sistem kesehatan dalam pencegahan anemia di tingkat populasi. Klasifikasi geografis terpencil dikaitkan dengan peningkatan risiko yang substansial (aOR = 1,93; 95% CI: 1,34–2,78; p = 0,001). Interaksi lintas level yang signifikan antara ketidakpatuhan IFA dan klasifikasi fasilitas terpencil (OR = 1,68; 95% CI: 1,21–2,33; p = 0,002) menunjukkan bahwa risiko anemia yang terkait dengan ketidakpatuhan diperkuat sekitar 1,7 kali lipat di fasilitas terpencil dibandingkan di perkotaan.

Tabel 2. Regresi Logistik Multilevel: Determinan Anemia pada Kehamilan (Model Penuh)

Determinan	aOR	95% CI	p-value
Level-1: Determinan Individual			
Ketidakpatuhan Suplementasi IFA	3,21	2,67–3,86	<0,001
Kekurangan Energi Kronis (LILA <23,5 cm)	2,84	2,31–3,49	<0,001
Paritas Tinggi (≥4 kelahiran)	2,17	1,73–2,72	<0,001
Komorbiditas Penyakit Kronis	1,89	1,41–2,54	<0,001
Kunjungan ANC Tidak Adekuat (<4)	1,78	1,43–2,21	<0,001
Level-2: Determinan Fasilitas			
Skor Kualitas Program ANC (per 10 poin)	0,61	0,48–0,78	<0,001
Klasifikasi Geografis Terpencil	1,93	1,34–2,78	0,001

aOR = Adjusted Odds Ratio; CI = Confidence Interval. ICC (null model) = 0,163; ICC (model penuh) = 0,087. Referensi: patuh IFA, LILA adekuat, paritas 1–3, tanpa komorbiditas, ANC adekuat, klasifikasi urban.

Kinerja Komparatif Model Machine Learning

Tabel 3 menyajikan kinerja prediktif komparatif dari semua model. *Gradient Boosting* mencapai AUC-ROC tertinggi (0,891), diikuti oleh *Random Forest* (0,874), model regresi logistik multilevel (0,847), dan regresi logistik konvensional (0,803). *Gradient Boosting* juga mencapai sensitivitas (0,843) dan skor F1 (0,849) yang superior, mengkonfirmasi keunggulannya dalam mengidentifikasi kasus anemia yang benar dalam populasi berpresvalensi tinggi ini. Analisis nilai SHAP mengkonfirmasi ketidakpatuhan IFA, kekurangan energi kronis, dan paritas tinggi sebagai tiga prediktor paling berpengaruh, diikuti oleh kunjungan ANC tidak adekuat dan klasifikasi Puskesmas terpencil—konsisten dengan temuan regresi multilevel. Analisis interaksi SHAP *Gradient Boosting* dan istilah interaksi lintas level regresi multilevel konvergen dalam mengidentifikasi amplifikasi sinergistik risiko anemia ketika ketidakpatuhan IFA terjadi bersama dengan lokasi fasilitas yang terpencil. Model regresi logistik multilevel, meskipun menghasilkan AUC-ROC yang lebih rendah dari *Gradient Boosting*, menawarkan keunggulan *odds ratio* yang dapat diinterpretasikan secara langsung dan dekomposisi varians, menjadikannya model yang disukai untuk komunikasi kebijakan dan pengembangan panduan klinis.

Tabel 3. Kinerja Prediktif Komparatif Model (Validasi Silang 5-Lipat)

Model	AUC-ROC	Sensitivitas	Spesifisitas	Skor F1
Regresi Logistik (baseline)	0,803	0,747	0,762	0,754
Regresi Logistik Multilevel	0,847	0,793	0,814	0,803
Random Forest	0,874	0,821	0,836	0,828
Gradient Boosting (terbaik)	0,891	0,843	0,857	0,849

AUC-ROC = Area Under the Receiver Operating Characteristic Curve. Semua metrik dari validasi silang berlapis lima terstratifikasi. Hosmer-Lemeshow $p = 0,412$ untuk regresi logistik multilevel.

5. KESIMPULAN

Penelitian ini mengembangkan dan memvalidasi model prediktif multilevel untuk determinan anemia pada ibu hamil menggunakan data big data kesehatan terintegrasi dari 87 Puskesmas di Provinsi Sulawesi Selatan, Indonesia. Tiga kesimpulan utama muncul. Pertama, anemia pada kehamilan merupakan fenomena *multilevel yang fundamental*: ICC null model sebesar 16,3% menunjukkan bahwa bagian varians yang signifikan secara klinis dan statistik dalam prevalensi anemia dapat diatribusikan kepada konteks Puskesmas dalam pemberian layanan—dimensi yang sepenuhnya tidak terlihat dalam analisis satu level konvensional. Kedua, ketidakpatuhan IFA, kekurangan energi kronis, paritas tinggi, komorbiditas penyakit kronis, dan pemanfaatan ANC yang tidak adekuat merupakan determinan individu yang dominan, sementara kualitas program ANC dan klasifikasi geografis terpencil merupakan determinan tingkat fasilitas yang paling signifikan. Interaksi lintas level antara ketidakpatuhan

IFA dan lokasi fasilitas terpencil memperkuat risiko anemia sekitar 1,7 kali lipat, mengidentifikasi wanita yang secara geografis dirugikan sebagai target prioritas tertinggi untuk intervensi yang diintensifkan. Ketiga, *Gradient Boosting* mengungguli semua model komparator dalam diskriminasi prediktif (AUC-ROC = 0,891), sementara analisis SHAP mengkonfirmasi koherensi klinis prediksi *machine learning* dan konvergensinya dengan temuan regresi multilevel.

Implikasi kebijakan dari penelitian ini sangat besar. Di tingkat individu, tenaga kesehatan perlu memprioritaskan ibu hamil dengan profil risiko tinggi, khususnya kombinasi ketidakpatuhan konsumsi suplementasi zat besi dan kekurangan energi kronis, melalui konseling intensif dan pemantauan kepatuhan secara aktif. Pada tingkat fasilitas kesehatan, efek protektif kualitas program antenatal care (ANC) menunjukkan pentingnya investasi kebijakan dalam penguatan sistem kesehatan melalui peningkatan kompetensi bidan, supervisi terstruktur berbasis siklus peningkatan mutu, serta pengelolaan rantai pasokan suplementasi zat besi yang konsisten. Temuan mengenai interaksi lintas level juga mendukung strategi distribusi sumber daya tambahan dan penguatan kapasitas pelayanan pada Puskesmas terpencil. Penelitian selanjutnya perlu memvalidasi model ini secara prospektif dalam kerangka sistem pendukung keputusan klinis yang terintegrasi dengan aplikasi e-Kohort serta mengevaluasi efektivitas biaya strategi pencegahan anemia berbasis stratifikasi risiko.

Kontribusi penulis dalam penelitian ini meliputi konseptualisasi oleh Asyima dan Masriadi; metodologi oleh Asyima, Ruqaiyah, dan Masriadi; pengembangan perangkat lunak oleh Asyima dan Ikrawanty Ayu Wulandari; analisis formal oleh Asyima, Hukmiyah Aspar, dan Nurasia Natsir; kurasi data oleh Ruqaiyah dan Ikrawanty Ayu Wulandari; penulisan draft asli oleh Asyima dan Ruqaiyah; peninjauan dan penyuntingan oleh Masriadi, Hukmiyah Aspar, dan Nurasia Natsir; serta supervisi oleh Masriadi. Seluruh penulis telah membaca dan menyetujui versi akhir naskah. Penelitian ini tidak menerima pendanaan eksternal. Penelitian dilaksanakan sesuai Deklarasi Helsinki dan telah memperoleh persetujuan dari Komite Etik Institut Ilmu Kesehatan Pelamonia Makassar. Persetujuan pasien dikecualikan karena penelitian menggunakan data sekunder retrospektif yang telah dide-identifikasi. Data pendukung penelitian tersedia melalui Dinas Kesehatan Provinsi Sulawesi Selatan berdasarkan permintaan yang wajar dan sesuai tata kelola data yang berlaku. Para penulis menyatakan tidak terdapat konflik kepentingan dalam penelitian ini.

UCAPAN TERIMA KASIH

Para penulis dengan tulus mengucapkan terima kasih kepada Dinas Kesehatan Provinsi Sulawesi Selatan dan 87 Puskesmas yang berpartisipasi atas fasilitas akses data, serta tim manajemen program e-Kohort di Kementerian Kesehatan Republik Indonesia atas bantuan teknis ekstraksi data.

DAFTAR REFERENSI

- Aheto, J. M. K. (2019). Predictive model and determinants of under-five child mortality: Evidence from the 2014 Ghana demographic and health survey. *BMC Public Health*, *19*(1), 64. <https://doi.org/10.1186/s12889-019-6390-4>
- Akombi, B. J., Agho, K. E., Merom, D., Renzaho, A. M. N., & Hall, J. J. (2017). Child malnutrition in sub-Saharan Africa: A meta-analysis of demographic and health surveys (2006–2016). *PLOS ONE*, *12*(5), e0177338. <https://doi.org/10.1371/journal.pone.0177338>
- Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Daru, J., Zamora, J., Fernández-Félix, B. M., Vogel, J., Oladapo, O. T., Morisaki, N., & Khan, K. S. (2018). Risk of maternal mortality in women with severe anaemia during pregnancy and post partum: A multilevel analysis. *The Lancet Global Health*, *6*(5), e548–e554. [https://doi.org/10.1016/S2214-109X\(18\)30078-0](https://doi.org/10.1016/S2214-109X(18)30078-0)
- Fatimah, S., Wirjatmadi, B., & Adriani, M. (2019). Factors affecting anemia in pregnant women in Surabaya. *Indian Journal of Public Health Research and Development*, *10*(8), 2365–2370.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>
- Gebre, A., & Mulugeta, A. (2015). Prevalence of anemia and associated factors among pregnant women in North Western Zone of Tigray, Northern Ethiopia. *Journal of Nutrition and Metabolism*, *2015*, 165430. <https://doi.org/10.1155/2015/165430>
- Goldstein, H. (2011). *Multilevel statistical models* (4th ed.). Wiley. <https://doi.org/10.1002/9780470973394>
- Haider, B. A., Olofin, I., Wang, M., Spiegelman, D., Ezzati, M., & Fawzi, W. W. (2013). Anaemia, prenatal iron use, and risk of adverse pregnancy outcomes: Systematic review and meta-analysis. *BMJ*, *346*, f3443. <https://doi.org/10.1136/bmj.f3443>
- Hasan, M. M., Sutradhar, I., Shahabuddin, A. S. M., & Sarker, M. (2020). Determinants of anemia among women of reproductive age in Bangladesh. *Public Health Nutrition*, *23*(15), 2792–2806.
- Kementerian Kesehatan Republik Indonesia. (2018). *Laporan nasional Riskesdas 2018*. Badan Penelitian dan Pengembangan Kesehatan.
- Kementerian Kesehatan Republik Indonesia. (2022). *Pedoman penggunaan aplikasi e-Kohort ibu*. Direktorat Jenderal Kesehatan Masyarakat.

- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774.
- Merlo, J., Chaix, B., Ohlsson, H., Beckman, A., Johnell, K., Hjerpe, P., Råstam, L., & Larsen, K. (2005). A brief conceptual tutorial of multilevel analysis in social epidemiology. *Journal of Epidemiology and Community Health*, 60(4), 290–297. <https://doi.org/10.1136/jech.2004.029454>
- Merlo, J., Wagner, P., Ghith, N., & Leckie, G. (2016). An original stepwise multilevel logistic regression analysis of discriminatory accuracy. *PLOS ONE*, 11(4), e0153778. <https://doi.org/10.1371/journal.pone.0153778>
- Nguyen, P. H., Scott, S., Avula, R., Tran, L. M., & Menon, P. (2018). Trends and drivers of change in the prevalence of anaemia among 1 million women and children in India, 2006 to 2016. *BMJ Global Health*, 3(5), e001010. <https://doi.org/10.1136/bmjgh-2018-001010>
- Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—Big data, machine learning, and clinical medicine. *New England Journal of Medicine*, 375(13), 1216–1219. <https://doi.org/10.1056/NEJMp1606181>
- Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: Promise and potential. *Health Information Science and Systems*, 2(1), 3.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Sage Publications.
- Stevens, G. A., Paciorek, C. J., Flores-Urrutia, M. C., Borghi, E., Namaste, S., Rogers, L. M., Wirth, J. P., Suchdev, P. S., Ezzati, M., & Rohner, F. (2022). National, regional, and global estimates of anaemia by severity in women and children for 2000–19: A pooled analysis. *The Lancet Global Health*, 10(5), e627–e639. [https://doi.org/10.1016/S2214-109X\(22\)00084-5](https://doi.org/10.1016/S2214-109X(22)00084-5)
- Sumarni, S. (2017). Model sosioekologi perilaku ibu dan kejadian stunting balita di wilayah pedesaan. *Media Gizi Indonesia*, 12(2), 116–125.
- World Health Organization. (2023). *Anaemia*. [World Health Organization](https://www.who.int/health-topics/anaemia)
- Xu, Y., Yang, J., Lan, X., Luo, Y., Chen, H., & Zhou, Y. (2021). Prediction of iron deficiency anemia using machine learning models: A study based on electronic health records. *Journal of Medical Systems*, 45(4), 42.