Journal of Computer Science and Information Technology (JCSIT)
Volume 3, No 3 – Juni 2026
e-ISSN : 3031-8467

KLASIFIKASI RISIKO PENYAKIT JANTUNG BERBASIS DATA REKAM MEDIS MENGGUNAKAN CATBOOST

Hanna Maryam*, Fahrim Irhamna Rachman, Chyquitha Danuputri
[1,2,3]Universitas Muhammadiyah Makassar
105841110922@student.unismuh.ac.id,
Fachrim141020@unismuh.ac.id,
chyquithadanuputri@unismuh.ac.id
*Corresponding Author

Received: 03-06- 2026	Revised: 15-06-2026	Approved: 25-06-2026

ABSTRAK

Penyakit jantung memerlukan deteksi dini agar pasien berisiko dapat diidentifikasi lebih cepat berdasarkan data klinis yang tersedia. Penelitian ini bertujuan menerapkan algoritma CatBoost untuk klasifikasi risiko penyakit jantung menggunakan data rekam medis pasien RSUD Haji Makassar. Dataset terdiri atas 640 data pasien periode 2021–2025 dengan fitur jenis kelamin, usia, glukosa, ureum, kreatinin, SGOT, SGPT, tekanan darah sistolik, dan tekanan darah diastolik. Target klasifikasi dibentuk berdasarkan diagnosis ICD-10 menjadi dua kelas, yaitu tidak berisiko dan berisiko penyakit jantung. Tahapan penelitian meliputi pembersihan data, imputasi nilai hilang menggunakan median, pembagian data latih dan data uji, pelatihan model CatBoost, serta evaluasi menggunakan *accuracy*, *precision*, *recall*, *F1-score*, *confusion matrix*, *AUC-ROC*, dan 5-*fold cross-validation*. Hasil pengujian menunjukkan *accuracy* 96,09%, *weighted F1-score* 94,18%, *macro F1-score* 49,00%, dan *AUC-ROC* 0,8943. Meskipun metrik agregat terlihat tinggi, *confusion matrix* menunjukkan model memprediksi seluruh data uji sebagai kelas berisiko dan gagal mengenali kelas tidak berisiko. Oleh karena itu, model belum layak diarahkan pada aplikasi klinis sebelum dilakukan perbaikan melalui pembobotan kelas, penyesuaian ambang keputusan, evaluasi imputasi yang lebih adaptif, dan validasi eksternal.
Kata Kunci: CatBoost, klasifikasi, *machine learning*, penyakit jantung, rekam medis

PENDAHULUAN

Penyakit kardiovaskular masih menjadi masalah kesehatan utama karena berhubungan dengan angka kematian yang tinggi dan kebutuhan deteksi dini. World Health Organization menyatakan bahwa penyakit kardiovaskular merupakan penyebab kematian utama secara global, dengan estimasi 19,8 juta kematian pada tahun 2022 dan sebagian besar kematian tersebut berkaitan dengan serangan jantung serta stroke [1]. Kondisi ini mendorong perlunya pemanfaatan data klinis yang telah tersedia di fasilitas kesehatan untuk mendukung identifikasi risiko pasien secara lebih cepat dan konsisten.

Dalam praktik pelayanan kesehatan, data rekam medis menyimpan informasi yang dapat merepresentasikan kondisi fisiologis pasien, seperti usia, jenis kelamin, kadar glukosa, fungsi ginjal, enzim hati, serta tekanan darah. Variabel tersebut tidak menggantikan penilaian klinis dokter, tetapi dapat digunakan sebagai bahan pemodelan risiko awal. Pelaporan model prediksi medis juga perlu dilakukan secara transparan agar tujuan, data, metode, metrik, dan keterbatasannya dapat dinilai dengan baik, sebagaimana ditekankan dalam pedoman TRIPOD+AI untuk model prediksi berbasis regresi maupun *machine learning* [2].

Machine learning dapat digunakan untuk menemukan pola pada data tabular medis dan melakukan klasifikasi terhadap label tertentu. Di antara berbagai

algoritma, metode ensemble berbasis boosting relevan karena mampu memodelkan hubungan nonlinier dan interaksi antarvariabel. CatBoost menerapkan *ordered boosting* untuk menghitung gradien berdasarkan urutan permutasi data sehingga mengurangi *prediction shift* dan *target leakage* yang dapat muncul ketika observasi yang sama digunakan sekaligus untuk membentuk statistik target dan memperbarui model [3]. Mekanisme ini penting pada dataset penelitian yang hanya berjumlah 640 pasien, karena ukuran sampel yang terbatas meningkatkan risiko model mempelajari pola yang terlalu spesifik terhadap data pelatihan.

Penanganan fitur kategorikal secara *native* juga krusial karena satu dari sembilan prediktor, yaitu jenis kelamin, bersifat kategorikal, sedangkan delapan fitur lainnya berupa pengukuran numerik klinis. CatBoost membentuk ordered target statistics tanpa memerlukan *one-hot encoding* manual yang dapat menambah dimensi, menghasilkan representasi jarang, dan mengabaikan keterkaitan kategori dengan target. Pada saat yang sama, pohon keputusan tetap dapat menangkap interaksi antara jenis kelamin, usia, tekanan darah, glukosa, indikator fungsi ginjal, dan enzim hati. Dengan demikian, pemilihan CatBoost didasarkan bukan hanya pada performa komputasi, tetapi juga pada kesesuaiannya terhadap struktur campuran dan ukuran data rekam medis penelitian ini [3].

Beberapa penelitian terbaru mendukung pemanfaatan *machine learning* untuk prediksi penyakit kardiovaskular. Qiu et al. mengembangkan kerangka CVD-OCSCatBoost untuk prediksi risiko kardiovaskular dan melaporkan bahwa optimasi CatBoost dapat meningkatkan kinerja klasifikasi [5]. Tompra et al. juga menekankan bahwa ketidakseimbangan kelas pada dataset penyakit jantung perlu diperhatikan karena dapat membuat nilai akurasi tampak tinggi, sementara kemampuan model mengenali kelas minoritas tetap rendah [6]. Dengan demikian, evaluasi model perlu melihat *confusion matrix*, *recall*, *F1-score*, dan *AUC-ROC*, bukan hanya akurasi.

Meskipun penelitian terdahulu melaporkan performa yang tinggi, komparasi antarpublikasi memiliki keterbatasan metodologis. Dataset yang digunakan berbeda dalam ukuran sampel, prevalensi kelas, definisi luaran, jumlah fitur, dan sumber data. Sebagian studi menggunakan dataset publik yang relatif seimbang dan telah melalui kurasi, sedangkan data rekam medis rutin cenderung mengandung nilai hilang, variasi pengukuran, dan ketidakseimbangan kelas. Akibatnya, perbedaan *accuracy* atau AUC antarpublikasi tidak dapat langsung dimaknai sebagai keunggulan algoritma tanpa mempertimbangkan karakteristik data dan rancangan validasinya [5], [6].

Keterbatasan lain adalah dominannya penggunaan *accuracy* sebagai dasar perbandingan. Pada data tidak seimbang, model yang selalu memilih kelas mayoritas dapat memperoleh *accuracy* tinggi meskipun gagal mengenali kelas minoritas. Selain itu, penggunaan resampling, optimasi hyperparameter, pemilihan threshold, dan skema *cross-validation* yang tidak seragam dapat menghasilkan estimasi performa yang berbeda. Oleh sebab itu, penelitian ini menilai weighted dan *macro F1-score*, *AUC-ROC*, serta *confusion matrix* agar performa model dibaca secara lebih kritis dan tidak hanya bergantung pada metrik agregat [6].

Penelitian ini berfokus pada klasifikasi risiko penyakit jantung menggunakan algoritma CatBoost berdasarkan data rekam medis pasien RSUD Haji Makassar. Target klasifikasi dibentuk dari diagnosis ICD-10, terutama kelompok penyakit sistem sirkulasi pada Chapter IX ICD-10 [4]. Tujuan penelitian adalah menerapkan CatBoost pada data medis pasien dan menilai performanya melalui metrik evaluasi klasifikasi yang mencakup *accuracy*, *precision*, *recall*, *F1-score*, *confusion matrix*, *AUC-ROC*, dan *cross validation*.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan desain pemodelan klasifikasi *supervised learning*. Objek penelitian adalah data rekam medis pasien RSUD Haji Makassar periode 1 Januari 2021 sampai 31 Desember 2025. Data yang dianalisis berjumlah 640 data pasien dengan rentang usia 2 sampai 93 tahun. Penelitian dilakukan selama Februari sampai April 2026, meliputi pengumpulan data, *preprocessing*, pembentukan label klasifikasi, pelatihan model, evaluasi, dan interpretasi hasil.

Fitur prediktor yang digunakan terdiri atas sembilan variabel, yaitu jenis kelamin, usia, glukosa, ureum, kreatinin, SGOT, SGPT, tekanan darah sistolik, dan tekanan darah diastolik. Variabel jenis kelamin digunakan sebagai fitur kategorikal, sedangkan variabel lain digunakan sebagai fitur numerik. Target klasifikasi dibentuk menjadi dua kelas, yaitu 0 untuk tidak berisiko dan 1 untuk berisiko penyakit jantung. Kelas berisiko ditentukan berdasarkan diagnosis pasien yang termasuk kelompok ICD-10 penyakit sistem sirkulasi, sedangkan diagnosis di luar kelompok tersebut diberi label tidak berisiko.

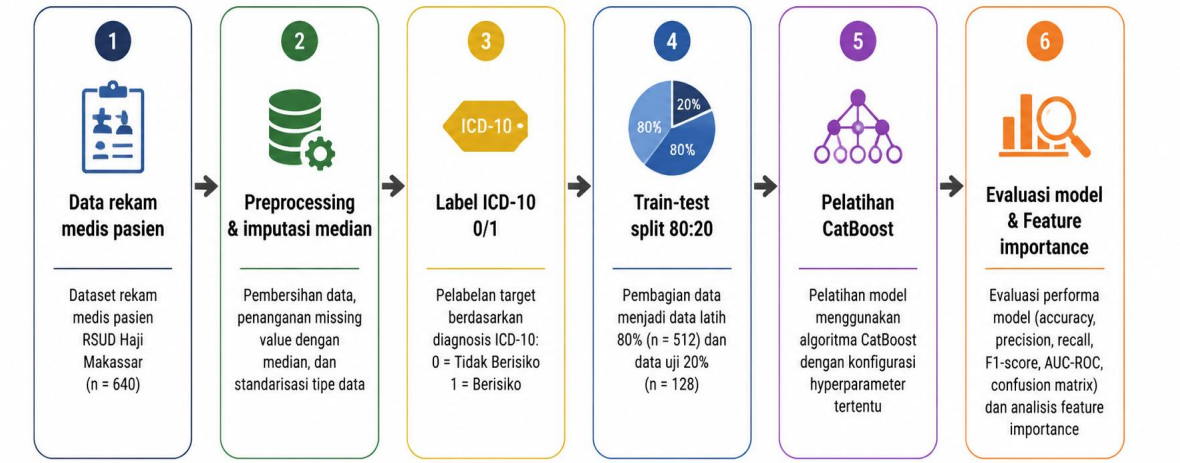
Tabel 1. Definisi operasional variabel penelitian

Variabel	Jenis	Keterangan
Jenis kelamin	Kategorikal	Kode jenis kelamin pasien sebagai fitur demografis.
Usia	Numerik	Usia pasien dalam tahun pada saat data direkam.
Glukosa	Numerik	Kadar glukosa darah pasien sebagai indikator metabolik.
Ureum	Numerik	Nilai ureum sebagai salah satu indikator fungsi ginjal.
Kreatinin	Numerik	Nilai kreatinin sebagai indikator fungsi ginjal.
SGOT dan SGPT	Numerik	Enzim transaminase yang mencerminkan kondisi fungsi hati.
Sistolik dan diastolik	Numerik	Nilai tekanan darah pasien yang digunakan sebagai faktor klinis.
Target risiko	Biner	0 = tidak berisiko, 1 = berisiko berdasarkan diagnosis ICD-10.

Tahap *preprocessing* dilakukan dengan memeriksa kelengkapan data, memisahkan tekanan darah menjadi sistolik dan diastolik, mengonversi format data numerik, serta menangani nilai hilang. Pada model utama, nilai hilang diimputasi menggunakan median yang dihitung hanya dari data latih agar informasi data uji tidak masuk ke proses pembentukan model. Median digunakan sebagai *baseline* karena lebih tahan terhadap nilai ekstrem daripada rata-rata pada variabel klinis. Namun, karena nilai klinis dapat berbeda menurut jenis kelamin dan kelompok usia, penggunaan median global belum sepenuhnya

mempertahankan heterogenitas pasien. Oleh sebab itu, hasil model utama perlu dibaca sebagai hasil *baseline* dan belum sebagai konfigurasi imputasi terbaik.

Sebagai tindak lanjut metodologis, evaluasi berikutnya harus membandingkan tiga skenario imputasi pada partisi data yang sama, yaitu median global, median terstratifikasi berdasarkan jenis kelamin dan kelompok usia, serta *K-Nearest Neighbors* (KNN) Imputer. Seluruh parameter imputasi wajib dipelajari hanya dari data latih pada setiap *fold* untuk mencegah *data leakage*. Pemilihan metode terbaik tidak cukup didasarkan pada *accuracy*, tetapi perlu mempertimbangkan *macro F1-score*, *balanced accuracy*, *recall* setiap kelas, dan stabilitas antar*fold*. Setelah *preprocessing*, dataset dibagi menjadi data latih dan data uji dengan komposisi 80:20 secara *stratified*. Model yang digunakan adalah *CatBoostClassifier* dengan *loss function Logloss* untuk klasifikasi biner, sedangkan evaluasi dilakukan menggunakan *accuracy*, *weighted precision*, *weighted recall*, *weighted F1-score*, *macro F1-score*, *AUC-ROC*, *confusion matrix*, *classification report*, dan 5-*fold Stratified* K-*Fold cross-validation* [7], [8].



Gambar 1. Alur penelitian klasifikasi risiko penyakit jantung menggunakan CatBoost

Tabel 2. Keonfigurasi hyperparameter model CatBoost

Hyperparameter	Nilai
iterations	200
learning_rate	0,05
depth	6
loss_function	Logloss
eval_metric	Accuracy
random_seed	42

HASIL DAN PEMBAHASAN

Preprocessing dan distribusi target

Hasil pemeriksaan data menunjukkan bahwa *missing value* ditemukan pada usia, kreatinin, tekanan darah sistolik, dan tekanan darah diastolik. Pada analisis *baseline*, nilai hilang diisi menggunakan median data latih masing-masing fitur, yaitu usia 58,0; kreatinin 0,88; sistolik 145,0; dan diastolik 81,0. Variabel glukosa, ureum, SGOT, dan SGPT tidak memiliki *missing value* sehingga dapat digunakan setelah penyesuaian format numerik. Karena penelitian ini belum membandingkan

median global dengan median terstratifikasi dan *KNN Imputer*, pengaruh pilihan imputasi terhadap batas keputusan serta kemampuan mengenali kelas minoritas belum dapat disimpulkan. Keterbatasan ini perlu dipertimbangkan saat menafsirkan performa model.

Distribusi target menunjukkan ketidakseimbangan kelas yang sangat kuat. Dari 640 data, sebanyak 613 data atau 95,78% termasuk kelas berisiko, sedangkan 27 data atau 4,22% termasuk kelas tidak berisiko. Setelah pembagian data, data uji terdiri atas 128 data, yaitu 5 data tidak berisiko dan 123 data berisiko. Komposisi ini penting diperhatikan karena akurasi model dapat terlihat tinggi hanya karena dominasi kelas mayoritas.

Tabel 3. Distribusi kelas target dan pembagian data

Kelas	Total Data	Persentase	Data Latih	Data Uji
Tidak Berisiko (0)	27	4,22%	22	5
Berisiko (1)	613	95,78%	490	123
Total	640	100,00%	512	128

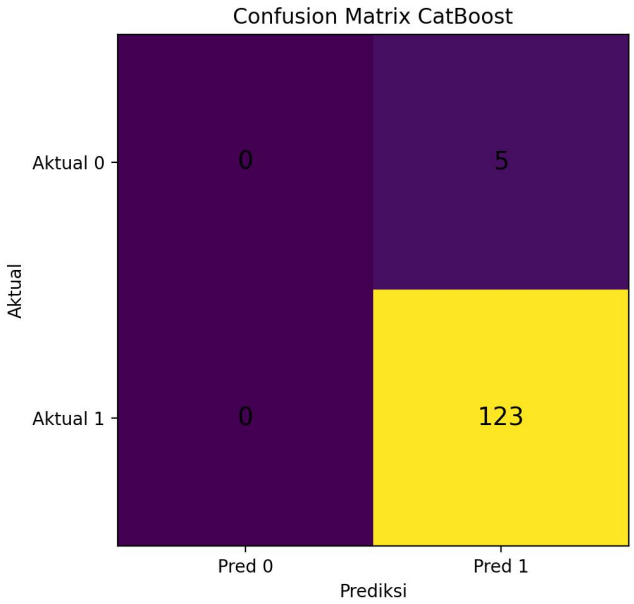
Hasil evaluasi model

Pengujian model CatBoost pada 128 data uji menghasilkan *accuracy* sebesar 96,09%, *weighted precision* sebesar 92,34%, *weighted recall* sebesar 96,09%, *weighted F1-score* sebesar 94,18%, *macro F1-score* sebesar 49,00%, dan *AUC-ROC* sebesar 0,8943. Nilai weighted metrics menunjukkan performa agregat yang tinggi karena jumlah data kelas berisiko sangat dominan. Namun, *macro F1-score* yang hanya 49,00% memberi sinyal bahwa kinerja antar kelas tidak seimbang.

Tabel 4. Hasil evaluasi model CatBoost

Metrik	Nilai	Keterangan
Accuracy	96,09%	Jumlah prediksi benar dibanding seluruh data uji.
Weighted Precision	92,34%	Ketepatan prediksi berbobot berdasarkan support kelas.
Weighted Recall	96,09%	Kemampuan mengenali kelas aktual secara berbobot.
Weighted F1-Score	94,18%	Rata-rata harmonik precision dan recall berbobot.
Macro F1-Score	49,00%	Rata-rata F1-score tanpa memperhatikan dominasi kelas.
AUC-ROC	0,8943	Kemampuan diskriminasi probabilitas model.

Confusion matrix memperlihatkan bahwa model memprediksi seluruh data uji sebagai kelas berisiko. Sebanyak 123 data berisiko berhasil diklasifikasikan dengan benar, tetapi 5 data tidak berisiko ikut diprediksi sebagai berisiko. Dengan demikian, tidak ada data yang diprediksi sebagai kelas tidak berisiko. Kondisi ini menjelaskan mengapa *accuracy* tinggi tidak sepenuhnya mencerminkan kemampuan model dalam membedakan dua kelas secara seimbang.



Gambar 2. *Confusion matrix* hasil pengujian model CatBoost

Tabel 5. *Classification report* per kelas

Kelas	Precision	Recall	F1-Score	Support
Tidak Berisiko (0)	0,00	0,00	0,00	5
Berisiko (1)	0,96	1,00	0,98	123
Macro Average	0,48	0,50	0,49	128
Weighted Average	0,92	0,96	0,94	128

Validasi silang dan *feature importance*

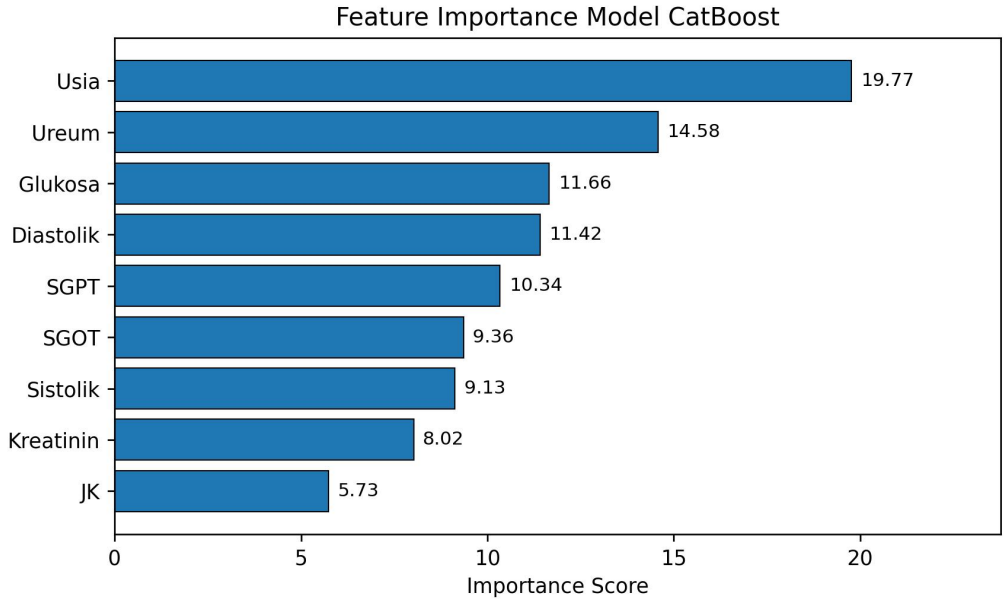
Pengujian dengan 5-*fold cross validation* menghasilkan *accuracy* pada *fold* pertama sebesar 96,09%, sedangkan *fold* kedua sampai kelima masing-masing sebesar 95,31%. Rata-rata *accuracy* sebesar 95,47% dengan standar deviasi 0,35%. Nilai ini menunjukkan performa yang relatif stabil antar*fold*, tetapi kestabilan tersebut tetap perlu ditafsirkan bersama distribusi kelas yang sangat tidak seimbang.

Tabel 6. Hasil 5-*fold cross validation*

Fold	Accuracy
Fold 1	96,09%
Fold 2	95,31%
Fold 3	95,31%
Fold 4	95,31%
Fold 5	95,31%
Rata-rata	95,47%
Standar Deviasi	0,35%

Analisis *feature importance* menunjukkan bahwa usia menjadi fitur paling berpengaruh dengan skor 19,77. Fitur berikutnya adalah ureum 14,58, glukosa 11,66, tekanan darah diastolik 11,42, SGPT 10,34, SGOT 9,36, tekanan darah sistolik 9,13, kreatinin 8,02, dan jenis kelamin 5,73. Temuan ini sejalan dengan

pemahaman klinis bahwa faktor demografis, metabolik, tekanan darah, serta fungsi ginjal dapat berasosiasi dengan risiko kardiovaskular. Namun, *feature importance* pada model ini tidak dapat langsung diartikan sebagai hubungan kausal karena model hanya mempelajari pola dari data yang tersedia.



Gambar 3. *Feature importance* model CatBoost

Tabel 7. Ringkasan *feature importance* model CatBoost

Ranking	Fitur	Importance Score
1	Usia	19,77
2	Ureum	14,58
3	Glukosa	11,66
4	Diastolik	11,42
5	SGPT	10,34
6	SGOT	9,36
7	Sistolik	9,13
8	Kreatinin	8,02
9	Jenis kelamin	5,73

PEMBAHASAN

Secara agregat, model menghasilkan *accuracy* 96,09% dan *AUC-ROC* 0,8943. Namun, nilai tersebut tidak menunjukkan klasifikasi dua kelas yang seimbang karena seluruh data uji diprediksi sebagai kelas berisiko. Oleh sebab itu, performa penelitian ini perlu dibandingkan dengan studi terdahulu bukan hanya berdasarkan *accuracy*, tetapi juga berdasarkan komposisi data, strategi penanganan ketidakseimbangan, dan kemampuan mengenali kelas minoritas.

Tabel 8. Komparasi performa penelitian ini dengan penelitian terdahulu

Penelitian	Data/strategi	Accuracy	Recall	F1-score	AUC
Penelitian ini	640 rekam medis lokal; rasio kelas 95,78:4,22;	96,09%	Kelas 0: 0,00; kelas 1: 1,00	Macro: 0,49	0,8943

	tanpa class weighting				
Qiu et al. [5]	Dataset CVD; optimasi OCS-CatBoost	73,67%	72,17%	Tidak dilaporkan pada ringkasan	0,8024
Tompra et al. [6]	Data besar tidak seimbang; CatBoost awal	74%	75%	Macro: 0,58; weighted: 0,80	0,82
Tompra et al. [6]	CatBoost + optimasi SMOTE	70%	79%	0,30	0,81
Tompra et al. [6]	CatBoost + optimasi SMOTE-ENN	63%	88%	0,27	0,82

Catatan: Nilai performa antarpublikasi tidak sepenuhnya dapat dibandingkan secara langsung karena terdapat perbedaan definisi kelas positif, distribusi kelas, jumlah dan jenis fitur, skema validasi, strategi resampling, serta metode perhitungan rata-rata metrik. Angka pada tabel digunakan untuk menunjukkan kecenderungan performa, bukan untuk menetapkan superioritas algoritma secara absolut.

Tabel 8 memperlihatkan bahwa *accuracy* penelitian ini lebih tinggi daripada Qiu et al. dan beberapa konfigurasi CatBoost pada Tompra et al., tetapi keunggulan tersebut bersifat semu karena 95,78% observasi berasal dari kelas berisiko. Prediksi konstan terhadap kelas mayoritas saja sudah menghasilkan *accuracy* sekitar 95,78%, sangat dekat dengan *accuracy* model sebesar 96,09%. Sebaliknya, Qiu et al. memperoleh *accuracy* lebih rendah tetapi *recall* 72,17%, sedangkan optimasi resampling pada Tompra et al. menurunkan *accuracy* CatBoost menjadi 63–70% namun meningkatkan *recall* menjadi 79–88%. Disparitas ini menunjukkan trade-off antara ketepatan agregat dan sensitivitas terhadap kelas yang kurang terwakili [5], [6].

Fenomena *model collapse* pada penelitian ini dapat dijelaskan melalui fungsi *Logloss* standar yang memberikan kontribusi total kesalahan lebih besar kepada kelas mayoritas karena jumlah observasinya jauh lebih banyak. Parameter *class_weights* pada CatBoost dapat mengalikan loss setiap observasi sesuai bobot kelas, sehingga kesalahan pada kelas yang kurang terwakili memperoleh penalti lebih besar. Bobot awal dapat ditentukan secara berbanding terbalik dengan frekuensi kelas, tetapi nilainya tetap harus dituning karena rasio ekstrem tidak selalu menghasilkan batas keputusan terbaik. Parameter *scale_pos_weight* juga dapat digunakan pada klasifikasi biner, tetapi penerapannya bergantung pada penetapan kelas positif. Dalam penelitian ini, kelas yang gagal dikenali adalah kelas 0, sehingga penggunaan *scale_pos_weight* tanpa meninjau kembali orientasi label berisiko justru memperkuat kelas berisiko sebagai kelas dominan.

Pemberian bobot kelas diperkirakan menggeser batas keputusan agar model tidak selalu memilih kelas mayoritas. Konsekuensinya, *recall* kelas tidak berisiko dan *macro F1-score* dapat meningkat, sementara *accuracy* agregat dan *precision* kelas

dominan dapat menurun karena bertambahnya prediksi terhadap kelas minoritas. Oleh karena itu, *class_weights* tidak boleh dipilih hanya berdasarkan rasio kelas, tetapi perlu dituning melalui *Stratified* K-*Fold* dengan *objective macro F1-score*, *balanced accuracy*, atau *area under the precision-recall curve*. Eksperimen pembobotan juga harus dibandingkan dengan *threshold tuning* dan kalibrasi probabilitas pada partisi yang sama. Karena konfigurasi berbobot belum diuji pada model utama, pembahasan ini merupakan dasar teoritis dan agenda validasi, bukan klaim bahwa model telah berhasil mengatasi collapse.

Hasil *confusion matrix* menegaskan bahwa *weighted precision*, *weighted recall*, dan *weighted F1-score* tidak boleh dibaca secara terpisah dari metrik per kelas. Nilai nol pada *precision*, *recall*, dan *F1-score* kelas tidak berisiko menunjukkan bahwa model belum mempelajari batas keputusan yang efektif bagi kelas minoritas. Dengan demikian, *AUC-ROC* yang relatif baik merefleksikan kemampuan pengurutan probabilitas, tetapi belum diterjemahkan menjadi keputusan kelas yang memadai pada threshold default.

Dari sisi penerapan, model ini masih merupakan prototipe eksperimental dan belum layak digunakan sebagai alat diagnosis maupun skrining klinis. Pengembangan berikutnya harus menguji *class weighting*, *threshold tuning*, median terstratifikasi, *KNN Imputer*, SMOTE, ADASYN, atau *hybrid resampling* secara terkontrol; melaporkan metrik per kelas, *balanced accuracy*, AUC-PR, dan kalibrasi; serta melakukan validasi eksternal pada rumah sakit lain. Seluruh perbandingan harus menggunakan partisi dan skema validasi yang sama agar perubahan performa dapat diatribusikan pada intervensi metodologis, bukan pada perbedaan data uji.

Keunikan penelitian ini dibandingkan beberapa penelitian berbasis dataset publik adalah penggunaan data rekam medis lokal RSUD Haji Makassar dan pembentukan target berdasarkan diagnosis ICD-10. Pendekatan ini membuat model lebih kontekstual terhadap data layanan kesehatan setempat. Namun, penelitian ini masih memiliki keterbatasan karena jumlah kelas tidak berisiko sangat kecil, variabel prediktor terbatas pada data yang tersedia di rekam medis, serta belum mencakup faktor lain seperti riwayat merokok, kolesterol, riwayat keluarga, indeks massa tubuh, terapi obat, dan hasil pemeriksaan penunjang lain.

Keterbatasan metodologis

Keterbatasan utama penelitian ini adalah belum dilakukannya eksperimen langsung terhadap metode imputasi adaptif dan parameter pembobotan kelas. Karena itu, median terstratifikasi, *KNN Imputer*, *class_weights*, *scale_pos_weight*, dan *threshold tuning* belum dapat dinyatakan meningkatkan performa sebelum diuji pada data dan skema validasi yang sama. Keterbatasan tersebut dibedakan secara tegas dari hasil empiris agar rekomendasi metodologis tidak ditafsirkan sebagai hasil yang telah terbukti.

Implikasi pengembangan sistem

Apabila dikembangkan lebih lanjut, model dapat diintegrasikan ke dalam sistem informasi rumah sakit sebagai modul skrining awal yang menampilkan probabilitas risiko dan faktor yang berkontribusi pada prediksi. Luaran sistem sebaiknya ditampilkan sebagai informasi pendukung bagi tenaga kesehatan, bukan sebagai keputusan diagnosis otomatis. Dengan cara tersebut, model dapat

membantu prioritisasi pemeriksaan tanpa menghilangkan peran verifikasi klinis dokter.

Pengembangan sistem juga perlu memperhatikan keamanan data, audit penggunaan model, pembaruan dataset secara berkala, serta monitoring performa setelah implementasi. Selain itu, perlu dibuat prosedur validasi internal dan eksternal agar model tidak hanya bekerja baik pada data penelitian, tetapi juga tetap stabil ketika digunakan pada data pasien baru dengan karakteristik yang mungkin berbeda.

KESIMPULAN

Penelitian ini menerapkan CatBoost pada sembilan fitur rekam medis untuk mengklasifikasikan risiko penyakit jantung berdasarkan diagnosis ICD-10. Meskipun model menghasilkan *accuracy* dan *AUC-ROC* yang tinggi secara agregat, *confusion matrix* menunjukkan *model collapse* karena seluruh data uji diprediksi sebagai kelas berisiko. Dengan demikian, performa saat ini belum memadai untuk penggunaan klinis. Model perlu dievaluasi ulang melalui perbandingan median global, median terstratifikasi, dan *KNN Imputer*; pembobotan kelas dan penyesuaian threshold; penambahan fitur klinis; serta validasi eksternal. Secara teoretis, penelitian ini menegaskan pentingnya membaca metrik per kelas pada data klinis yang tidak seimbang. Secara praktis, temuan ini memberikan dasar evaluatif bagi pengembangan *Clinical Decision Support System* di Indonesia, yaitu bahwa kelayakan model harus ditentukan oleh kemampuan membedakan seluruh kelompok pasien secara seimbang, bukan oleh *accuracy* agregat semata.

DAFTAR PUSTAKA

[1] World Health Organization, “Cardiovascular diseases (CVDs),” Fact sheet, Jul. 31, 2025. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[2] G. S. Collins et al., “TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods,” BMJ, vol. 385, Art. no. e078378, 2024, doi: 10.1136/bmj-2023-078378.

[3] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, “CatBoost: unbiased boosting with categorical features,” in Advances in Neural Information Processing Systems, vol. 31, pp. 6638–6648, 2018. [Online]. Available: https://proceedings.neurips.cc/paper/2018/hash/14491b756b3a51daac41c24863285549-Abstract.html

[4] World Health Organization, “ICD-10 Version: 2019: International Statistical Classification of Diseases and Related Health Problems, 10th Revision,” 2019. [Online]. Available: https://icd.who.int/browse10/2019/en

[5] Z. Qiu, Y. Qiao, W. Shi, and X. Liu, “A robust framework for enhancing cardiovascular disease risk prediction using an optimized category boosting model,” Mathematical Biosciences and Engineering, vol. 21, no. 2, pp. 2943–2969, 2024, doi: 10.3934/mbe.2024131.

[6] K.-V. Tompra, G. Papageorgiou, and C. Tjortjis, “Strategic machine learning optimization for cardiovascular disease prediction and high-risk patient identification,” Algorithms, vol. 17, no. 5, Art. no. 178, 2024, doi: 10.3390/a17050178.

[7] Scikit-learn Developers, "Cross-validation: evaluating estimator performance," scikit-learn documentation, 2026. [Online]. Available: https://scikit-learn.org/stable/modules/cross_validation.html
[8] Scikit-learn Developers, "Metrics and scoring: quantifying the quality of predictions," scikit-learn documentation, 2026. [Online]. Available: https://scikit-learn.org/stable/modules/model_evaluation.html
[9] M. M. Ali, B. K. Paul, K. Ahmed, F. M. Bui, J. M. W. Quinn, and M. A. Moni, "Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison," Computers in Biology and Medicine, vol. 136, Art. no. 104672, 2021, doi: 10.1016/j.compbiomed.2021.104672.
[10] D. Asif, M. Bibi, M. S. Arif, and A. Mukheimer, "Enhancing heart disease prediction through ensemble learning techniques with hyperparameter optimization," Algorithms, vol. 16, no. 6, Art. no. 308, 2023, doi: 10.3390/a16060308.
[11] T. R. Mahesh et al., "AdaBoost ensemble methods using K-fold cross validation for survivability with the early detection of heart disease," Computational Intelligence and Neuroscience, vol. 2022, Art. no. 9005278, 2022, doi: 10.1155/2022/9005278.