



Arrhythmia Classification Using Electrocardiogram Signal Features and Heart Rate Variability with the XGBoost Algorithm

Klasifikasi Aritmia Menggunakan Fitur Sinyal Elektrokardiogram dan Heart Rate Variability dengan Algoritma XGBoost

**Siti Salsabi¹, Bella Antika², Zahara Martiza³,
Theo Helena Rikie Lam⁴, Agung Prabowo^{5*}**

^{1,2,3,4,5}Program Studi Teknik Informatika, Fakultas Sains dan Teknologi,
Universitas Prima Indonesia, Indonesia

E-Mail: ¹sitisalsabi17@gmail.com, ²bellaantika2004@gmail.com, ³zaharamartiza2004@gmail.com,
⁴lamhelenarikie@gmail.com, ⁵agungprabowo@unprimdn.ac.id

Received Apr 25th 2026; Revised Jun 15th 2026; Accepted Jul 07th 2026; Available Online Jul 29th 2026

Corresponding Author: Agung Prabowo

Copyright © 2026 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

Abstract

Cardiac arrhythmia is one of the most common cardiovascular disorders and may lead to serious complications, including stroke, heart failure, and sudden cardiac death if not detected early. Electrocardiogram (ECG) signals are widely used for arrhythmia detection; however, accurate classification remains challenging due to the complexity of ECG waveforms, similarities among arrhythmia classes, and class imbalance in available datasets. Previous studies have generally utilized either ECG morphological features or Heart Rate Variability (HRV) features separately, limiting the comprehensive representation of cardiac electrical activity and rhythm dynamics. This study aims to classify cardiac arrhythmias by combining heartbeat-based ECG features and time-domain HRV features using the Extreme Gradient Boosting (XGBoost) algorithm. The dataset was obtained from the MIT-BIH Arrhythmia Database, which contains single-lead ECG signals and heartbeat annotations. Preprocessing included exploratory data analysis, handling missing values, zero-padding, linear interpolation, and normalization using StandardScaler. The HRV features consisted of MeanRR, SDNN, RMSSD, and pNN50. The model was evaluated using accuracy, precision, recall, F1-score, and a confusion matrix. The proposed model achieved an accuracy of 99.16%, demonstrating excellent performance in distinguishing normal and arrhythmic heartbeats. The integration of HRV and ECG features improved data representation and classification performance, although class imbalance remained a challenge for several categories.

Keywords: Arrhythmia Classification, Electrocardiogram, EDA, Time-Domain HRV, XGBoost.

Abstrak

Aritmia jantung merupakan gangguan kardiovaskular yang dapat menyebabkan stroke, gagal jantung, dan kematian mendadak akibat gangguan irama jantung apabila tidak terdeteksi sejak dini. Sinyal Elektrokardiogram (ECG) digunakan untuk mendeteksi aritmia, namun klasifikasi masih menjadi tantangan akibat kompleksitas pola gelombang, kemiripan antarkelas, dan ketidakseimbangan data. Penelitian sebelumnya umumnya memanfaatkan fitur morfologi ECG atau Heart Rate Variability (HRV) secara terpisah, sehingga belum merepresentasikan karakteristik sinyal secara menyeluruh. Penelitian ini bertujuan untuk mengklasifikasikan aritmia menggunakan kombinasi fitur ECG berbasis denyut jantung dan fitur HRV pada domain waktu dengan algoritma Extreme Gradient Boosting (XGBoost). Data berasal dari MIT-BIH Arrhythmia Database yang berisi sinyal ECG satu kanal beserta anotasi denyut jantung. Tahap prapemrosesan meliputi exploratory data analysis, penanganan nilai hilang, zero padding, interpolasi linier, dan normalisasi menggunakan StandardScaler. Fitur HRV yang digunakan meliputi MeanRR, SDNN, RMSSD, dan pNN50. Model dievaluasi menggunakan accuracy, precision, recall, F1-score, dan confusion matrix. Hasil menunjukkan akurasi sebesar 99,16%, yang menunjukkan kemampuan sangat baik dalam membedakan denyut jantung normal dan aritmia. Integrasi fitur HRV dan ECG meningkatkan representasi data serta performa klasifikasi, meskipun ketidakseimbangan kelas masih menjadi tantangan pada beberapa kategori.

Kata kunci: EDA, Elektrokardiogram, HRV Domain Waktu, Klasifikasi Aritmia, XGBoost.

1. PENDAHULUAN

Penyakit kardiovaskular hingga saat ini masih menjadi salah satu penyebab kematian tertinggi di dunia, baik di negara maju maupun berkembang, serta menjadi masalah kesehatan global yang terus mendapat perhatian serius karena dampaknya terhadap morbiditas dan mortalitas [1], [2]. Salah satu gangguan kardiovaskular yang banyak ditemukan adalah aritmia, yaitu kondisi gangguan irama jantung yang menyimpang dari irama sinus normal akibat kelainan dalam pembentukan maupun penghantaran impuls listrik jantung [3]. Salah satu bentuk aritmia yang sering terjadi ialah takiaritmia, yakni kondisi ketika denyut jantung berdetak lebih cepat daripada rentang normal. Kejadian aritmia dipengaruhi oleh berbagai faktor risiko, seperti gaya hidup tidak sehat, gangguan metabolik, serta riwayat penyakit jantung dalam keluarga yang meningkatkan kemungkinan terjadinya gangguan kardiovaskular [4]. Kondisi ini dapat menimbulkan berbagai gejala klinis, mulai dari palpitasi, kelelahan, nyeri dada, sinkop, hingga komplikasi serius yang berpotensi mengancam jiwa apabila tidak dideteksi secara dini [5]. Oleh karena itu, diperlukan metode deteksi yang akurat dan efisien untuk mendukung identifikasi dini aritmia.

Salah satu metode diagnostik yang banyak digunakan dalam mendeteksi gangguan irama jantung adalah Elektrokardiogram (EKG), yaitu alat diagnostik noninvasif yang merekam aktivitas listrik jantung dan mampu memberikan informasi penting terkait kondisi jantung pasien. Sinyal EKG normal terdiri atas gelombang P, kompleks QRS, dan gelombang T yang merepresentasikan proses depolarisasi dan repolarisasi jantung [6]. Penggunaan EKG memungkinkan proses identifikasi kelainan jantung dilakukan secara cepat dan akurat, termasuk dalam mendeteksi berbagai jenis aritmia [7]. Namun demikian, pemanfaatan sinyal EKG untuk klasifikasi aritmia di Indonesia masih relatif terbatas, khususnya penelitian yang secara spesifik mengintegrasikan fitur *Heart Rate Variability* (HRV) sebagai parameter tambahan dalam sistem klasifikasi berbasis pembelajaran mesin [8]. Beberapa penelitian terdahulu telah menunjukkan performa yang baik dalam klasifikasi aritmia menggunakan berbagai pendekatan machine learning maupun deep learning. Penelitian sebelumnya memperoleh akurasi hingga 100% menggunakan transformasi spektrogram dan Convolutional Neural Network (CNN) pada dataset Massachusetts Institute of Technology-Beth Israel Hospital (MIT-BIH) Arrhythmia Database [1]. Penelitian Ramadhan dan Mandala memanfaatkan fitur RR Interval, PR Interval, dan QRS Complex dengan algoritma CNN dan menghasilkan akurasi sebesar 98,21% [9]. Meskipun menunjukkan performa yang tinggi, sebagian besar penelitian tersebut masih berfokus pada fitur morfologi sinyal EKG atau ekstraksi fitur tertentu tanpa mengombinasikannya dengan informasi variabilitas denyut jantung yang direpresentasikan oleh HRV.

Heart Rate Variability (HRV) merupakan variasi interval antar detak jantung yang dihitung berdasarkan interval R-R pada sinyal EKG dan mencerminkan aktivitas sistem saraf otonom dalam mengatur fungsi jantung [8]. Analisis HRV berkembang melalui berbagai pendekatan, mulai dari *time domain*, *frequency domain*, hingga metode non-linear seperti *Poincaré plot*, yang digunakan untuk mengevaluasi karakteristik variabilitas denyut jantung secara lebih komprehensif [10]. Di antara berbagai pendekatan tersebut, analisis HRV *time domain* banyak digunakan karena memiliki proses perhitungan yang relatif sederhana, kebutuhan komputasi yang rendah, serta mampu memberikan informasi fisiologis yang penting melalui parameter seperti Mean of R-R intervals (MeanRR), Standard Deviation of NN intervals (SDNN), Root Mean Square of Successive Differences (RMSSD), dan Percentage of successive NN intervals that differ by more than 50 ms (pNN50). Penelitian Rumpa dan Panggalo menunjukkan bahwa parameter pNN50 mampu merepresentasikan perubahan kondisi fisiologis yang berkaitan dengan aktivitas sistem saraf otonom sehingga berpotensi digunakan sebagai fitur penting dalam sistem klasifikasi berbasis machine learning. Selain itu, beberapa penelitian sebelumnya masih memiliki keterbatasan dari sisi cakupan klasifikasi maupun representasi fitur. Penelitian Novianto dkk. hanya berfokus pada identifikasi kondisi normal dan takikardia berdasarkan nilai BPM sehingga belum mampu membedakan berbagai jenis aritmia secara rinci [11]. Penelitian Siahaan dkk. juga masih menerapkan klasifikasi biner antara kondisi normal dan berpotensi aritmia sehingga belum mengakomodasi skenario klasifikasi multikelas yang lebih kompleks [12]. Sementara itu, penelitian Rumpa dan Panggalo hanya memanfaatkan satu parameter HRV yaitu pNN50 tanpa mengintegrasikannya dengan fitur EKG maupun algoritma machine learning untuk klasifikasi aritmia [13]. Keterbatasan tersebut menunjukkan bahwa informasi fisiologis yang terkandung dalam HRV masih belum dimanfaatkan secara optimal dalam banyak penelitian terdahulu.

Dalam perkembangan analisis data medis, penerapan *machine learning* menjadi pendekatan yang semakin banyak digunakan untuk meningkatkan akurasi klasifikasi penyakit, termasuk pada data sinyal biomedis. Salah satu algoritma yang menunjukkan performa tinggi adalah *Extreme Gradient Boosting* (XGBoost), yang dikenal memiliki kemampuan baik dalam menangani data kompleks, hubungan non-linear antar variabel, serta efisiensi komputasi yang tinggi [14]. Penelitian sebelumnya menunjukkan XGBoost mampu menghasilkan akurasi klasifikasi penyakit jantung yang lebih tinggi dibandingkan algoritma konvensional, bahkan mencapai akurasi sebesar 93% dibandingkan *decision tree* sebesar 90% [15]. Temuan tersebut diperkuat oleh penelitian lain yang menunjukkan bahwa XGBoost memiliki stabilitas kinerja yang baik, prediksi yang seimbang antar kelas, serta efektif digunakan pada klasifikasi data medis [16], [17]. Selain itu, peningkatan performa model juga dapat didukung melalui optimasi parameter adaptif dan tahapan

preprocessing yang tepat sehingga mampu meningkatkan akurasi serta efisiensi model, khususnya pada pengolahan data berskala besar [18], [19].

Meskipun berbagai penelitian terkait klasifikasi aritmia telah dilakukan, sebagian besar masih berfokus pada penggunaan sinyal EKG secara langsung atau penerapan algoritma klasifikasi konvensional tanpa mengoptimalkan fitur HRV sebagai komponen utama dalam model klasifikasi. Penelitian oleh Terapan et al. menunjukkan bahwa fitur HRV *time domain* berkontribusi terhadap performa klasifikasi, namun masih terbatas pada seleksi fitur dan penggunaan pendekatan klasifikasi tertentu tanpa mengintegrasikan fitur morfologi sinyal EKG berbasis denyut secara langsung [20]. Penelitian lain telah menggunakan fitur HRV, namun masih terbatas pada jumlah parameter tertentu atau belum menerapkan klasifikasi aritmia multikelas secara komprehensif [13]. Selain itu, pendekatan deep learning yang banyak digunakan cenderung membutuhkan data dalam jumlah besar, sumber daya komputasi tinggi, serta memiliki keterbatasan interpretabilitas terhadap fitur yang berkontribusi pada keputusan model [1], [12].

Algoritma XGBoost dipilih karena memiliki kemampuan yang baik dalam menangani hubungan non-linear antar fitur, mendukung mekanisme regularisasi untuk mengurangi risiko overfitting, serta mampu menghasilkan performa tinggi pada data tabular yang berasal dari proses ekstraksi fitur sinyal biomedis. Dibandingkan pendekatan deep learning yang membutuhkan proses ekstraksi fitur otomatis dan sumber daya komputasi besar, XGBoost lebih efisien untuk dataset berbasis fitur hasil ekstraksi EKG dan HRV serta memberikan interpretasi model yang lebih baik melalui analisis kontribusi fitur. Berdasarkan kondisi tersebut, penelitian ini bertujuan mengembangkan model klasifikasi aritmia berbasis sinyal EKG dengan mengintegrasikan fitur Heart Rate Variability (HRV) domain waktu menggunakan algoritma Extreme Gradient Boosting (XGBoost) guna meningkatkan akurasi dan kestabilan klasifikasi pada kasus multikelas. Research gap penelitian ini terletak pada masih terbatasnya penelitian yang mengombinasikan fitur morfologi sinyal EKG berbasis denyut dengan fitur HRV domain waktu dalam satu model klasifikasi aritmia multikelas berbasis XGBoost. Adapun kebaruan (*novelty*) penelitian ini terletak pada integrasi fitur morfologi sinyal EKG dan parameter HRV domain waktu (MeanRR, SDNN, RMSSD, dan pNN50) yang dihitung menggunakan pendekatan sliding window sebagai masukan model XGBoost untuk klasifikasi aritmia berdasarkan standar AAMI ke dalam empat kelas, yaitu Normal (N), Supraventricular (S), Ventricular (V), dan Fusion (F). Pendekatan ini diharapkan mampu menghasilkan representasi data yang lebih komprehensif dibandingkan dengan penggunaan fitur EKG atau HRV secara terpisah, disertai tahapan *preprocessing* yang komprehensif untuk menghasilkan sistem deteksi aritmia yang lebih efektif dan efisien. Penelitian ini memanfaatkan MIT-BIH Arrhythmia Database dengan klasifikasi berdasarkan standar AAMI ke dalam empat kelas, yaitu *Normal (N)*, *Supraventricular (S)*, *Ventricular (V)*, dan *Fusion (F)*. Hasil penelitian ini diharapkan dapat berkontribusi dalam pengembangan metode deteksi dini gangguan irama jantung berbasis *machine learning* serta menjadi referensi bagi pengembangan sistem diagnosis cerdas di bidang kesehatan.

2. METODE PENELITIAN

Penelitian ini diawali dengan pengumpulan data Elektrokardiogram (EKG) dari dataset Massachusetts Institute of Technology-Beth Israel Hospital (MIT-BIH) Arrhythmia Database yang merupakan dataset standar untuk penelitian aritmia. Setelah data diperoleh, dilakukan *preprocessing* untuk membersihkan dan menyiapkan data agar kualitasnya lebih baik sebelum diproses lebih lanjut. Tahap berikutnya adalah ekstraksi fitur, yaitu mengambil informasi penting dari sinyal EKG dan menghitung fitur Heart Rate Variability (HRV) yang dapat menggambarkan variasi denyut jantung. Kemudian dataset dibagi menjadi data training dan testing dengan rasio 80:20. Data training digunakan untuk melatih model XGBoost, sedangkan data testing digunakan untuk mengevaluasi kemampuan model dalam mengklasifikasikan aritmia.

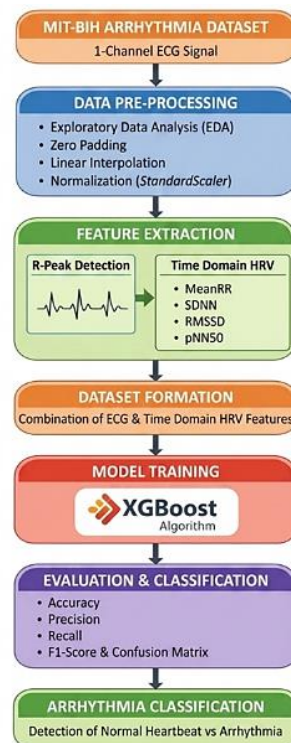
Penelitian [12] mengembangkan sistem deteksi aritmia menggunakan arsitektur Deep CNN berbasis Conv1D pada dataset MIT-BIH Arrhythmia. Model yang dibangun melalui tahapan *data cleaning*, normalisasi Z-score, dan *weighting* berhasil mencapai akurasi 98,88%. Namun, pendekatan tersebut masih bergantung pada satu dataset dan belum mengintegrasikan informasi fisiologis dari fitur HRV yang dapat merepresentasikan dinamika ritme jantung. Oleh karena itu, penelitian ini mengusulkan kombinasi fitur sinyal EKG dan HRV domain waktu dengan algoritma XGBoost untuk memperoleh informasi yang lebih lengkap dalam proses klasifikasi aritmia. Penelitian ini [21] mengembangkan sensor wearable berbasis Heart Rate Variability (HRV) untuk memantau tingkat stres secara real-time. Penelitian tersebut melibatkan 100 responden dan menunjukkan bahwa sistem yang dikembangkan mampu mencapai akurasi rata-rata sebesar 86%, dengan akurasi tertinggi mencapai 95% pada kategori stres rendah.

Penelitian [13] menganalisis fitur Heart Rate Variability (HRV) berupa pNN50 yang diperoleh dari sinyal EKG untuk mengidentifikasi perubahan kondisi fisiologis saat seseorang mengalami emosi marah. Hasil penelitian menunjukkan bahwa nilai pNN50 mengalami perubahan pada sebagian besar partisipan, sehingga parameter tersebut dinilai mampu merepresentasikan aktivitas sistem saraf otonom. Penelitian [11] mengembangkan sistem elektrokardiogram (EKG) berbasis Internet of Things (IoT) menggunakan sensor AD8232 dan NodeMCU ESP8266 untuk pemantauan kondisi jantung secara real-time. Hasil pengujian

menunjukkan bahwa sistem memperoleh akurasi pengukuran BPM sebesar 97,96% dengan rata-rata error 2,04%, serta mampu membedakan kondisi jantung normal dan takikardia berdasarkan nilai BPM.

Penelitian [22] mengusulkan metode deteksi aritmia berbasis Deep Convolutional Neural Network (Deep CNN) dengan arsitektur Conv1D menggunakan dataset MIT-BIH Arrhythmia Database. Hasil penelitian menunjukkan performa yang sangat baik dengan nilai akurasi, precision, recall, dan F1-score sebesar 98,88%. Pada Penelitian [23] mengembangkan sistem klasifikasi kondisi jantung pada mahasiswa yang mengalami gangguan tidur menggunakan citra EKG dan algoritma K-Nearest Neighbor (KNN). Hasil penelitian menunjukkan bahwa penggunaan 20 komponen PCA dengan nilai K=6 menghasilkan performa terbaik dengan akurasi sebesar 96,23%, precision 98,46%, recall 92,75%, dan F1-score 95,50%.

Flowchart penelitian dapat ditunjukkan pada Gambar 1.



Gambar 1. Flowchart Penelitian

Penelitian ini menggunakan pendekatan kuantitatif berbasis eksperimen komputasional untuk menganalisis sinyal Elektrokardiogram (EKG) dalam proses klasifikasi aritmia. Penelitian ini berfokus pada pengolahan sinyal EKG dan fitur Heart Rate Variability (HRV) pada domain waktu menggunakan algoritma Extreme Gradient Boosting (XGBoost), dengan memanfaatkan dataset *MIT-BIH Arrhythmia Database* dari PhysioNet. Penelitian ini dilaksanakan pada periode Agustus 2025 hingga Juni 2026 dengan menggunakan dataset sekunder *MIT-BIH Arrhythmia Database* yang diperoleh melalui platform PhysioNet. Seluruh tahapan pengolahan data, pembangunan model, serta evaluasi dilakukan secara komputasional menggunakan perangkat milik peneliti. Peralatan yang digunakan dalam penelitian ini terdiri atas perangkat keras dan perangkat lunak. Perangkat keras yang digunakan meliputi laptop dan mouse. Perangkat lunak yang digunakan berupa aplikasi Visual Studio Code dengan bahasa pemrograman Python.

Metode penelitian yang digunakan pada penelitian ini terdiri atas beberapa tahapan, yaitu pra-pemrosesan data, ekstraksi fitur, pembagian data, pelatihan model menggunakan algoritma XGBoost, dan evaluasi kinerja model. Alur penelitian secara keseluruhan ditunjukkan pada Gambar 1. Tahapan-tahapan tersebut dirancang untuk menghasilkan model klasifikasi aritmia yang mampu memanfaatkan informasi morfologi sinyal EKG dan fitur Heart Rate Variability (HRV) secara optimal.

2.1. Pra-Pemrosesan Data (*Data Preprocessing*)

Tahap pra-pemrosesan dilakukan untuk meningkatkan kualitas data sebelum digunakan dalam proses pelatihan model. Tahapan ini bertujuan untuk mengurangi gangguan pada data serta memastikan seluruh sampel memiliki format yang seragam.

1. **Penanganan Nilai Hilang**
Penanganan nilai hilang dilakukan untuk memastikan seluruh data dapat diproses secara optimal. Nilai yang tidak tersedia diperbaiki menggunakan metode interpolasi sehingga kontinuitas data tetap terjaga.
2. **Zero Padding pada Sinyal Elektrokardiogram**
Zero padding digunakan untuk menyeragamkan panjang seluruh segmen sinyal Elektrokardiogram (EKG). Proses ini dilakukan dengan menambahkan nilai nol pada bagian akhir sinyal yang panjangnya lebih pendek dari panjang target.
3. **Interpolasi Data**
Interpolasi digunakan untuk memperkirakan nilai yang hilang pada data sehingga karakteristik sinyal tetap dipertahankan dan dapat digunakan dalam proses ekstraksi fitur.
4. **Normalisasi Data**
Normalisasi dilakukan menggunakan *StandardScaler* untuk menyamakan rentang nilai antarfitur sehingga proses pelatihan model menjadi lebih stabil dan efektif.

2.2. Ekstraksi Fitur

Tahap ekstraksi fitur dilakukan untuk mengubah sinyal Elektrokardiogram (EKG) menjadi representasi numerik yang dapat diproses oleh algoritma machine learning. Pada penelitian ini digunakan kombinasi fitur sinyal EKG berbasis denyut jantung dan fitur Heart Rate Variability (HRV) domain waktu yang terdiri atas MeanRR, SDNN, RMSSD, dan pNN50. Perhitungan fitur HRV dilakukan berdasarkan interval RR menggunakan pendekatan sliding window

2.3. Pembagian Data Train-Test

Data yang telah melalui tahap pra-pemrosesan dan ekstraksi fitur kemudian dibagi menjadi data pelatihan (training set) dan data pengujian (testing set) dengan rasio 80:20. Data pelatihan digunakan untuk membangun model klasifikasi, sedangkan data pengujian digunakan untuk mengevaluasi kemampuan model pada data yang tidak terlibat dalam proses pelatihan.

2.4. Klasifikasi Menggunakan Algoritma XGBoost

Algoritma Extreme Gradient Boosting (XGBoost) digunakan sebagai model klasifikasi karena mampu menangani data nonlinier. Proses pelatihan model dilakukan menggunakan data latih dan diuji pada data uji. Persamaan 1 merupakan persamaan dari algoritma XGBoost [25].

$$L_t = \sum_{i=1}^n l\left(y_i, y_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \quad (1)$$

2.5. Evaluasi Model

Evaluasi model dilakukan menggunakan confusion matrix serta metrik accuracy, precision, recall, dan F1-score untuk mengukur kemampuan model dalam mengklasifikasikan setiap kelas aritmia

3. HASIL DAN PEMBAHASAN

3.1 Data Collecting

Dataset yang digunakan dalam penelitian ini berupa sinyal elektrokardiogram (EKG) yang diperoleh dari Massachusetts Institute of Technology-Beth Israel Hospital (MIT-BIH) *Arrhythmia Database* melalui platform PhysioNet. Dataset tersebut merupakan salah satu dataset standar yang banyak dimanfaatkan dalam penelitian klasifikasi aritmia serta telah dilengkapi dengan anotasi denyut jantung yang telah divalidasi secara klinis. Contoh data mentah disajikan pada Tabel 1, sedangkan karakteristik lengkap dataset dirangkum pada Tabel 2.

Dalam penelitian ini, data dikelompokkan ke dalam 4 kelas berdasarkan standar AAMI, yaitu Normal (N), Supraventricular Ectopic Beat (S), Ventricular Ectopic Beat (V), dan Fusion Beat (F). Distribusi jumlah data pada masing-masing kelas ditunjukkan pada Tabel 3. Distribusi tersebut menunjukkan bahwa dataset memiliki karakteristik tidak seimbang (*imbalanced*), di mana kelas Normal (N) memiliki jumlah sampel yang jauh lebih banyak dibandingkan kelas lainnya.

Tabel 1. Dataset

Record	Beat Number	Sample Index	Time_second	Amplitude	Label
100	1	18	0.050000	-0.170	+
100	2	77	0.213889	0.840	N
100	3	370	1.027778	0.940	N

Record	Beat_Number	Sample_Index	Time_second	Amplitude	Label
100	4	662	1.838889	0.885	N
100	5	946	2.627778	0.810	N
...	...	...	...	...	...
234	2760	648797	1802.213889	1.725	N
234	2761	649040	1802.888889	1.680	N
234	2762	649292	1803.588889	1.540	N
234	2763	649536	1804.266667	1.660	N
234	2764	649772	1804.922222	1.635	N

Setiap data denyut jantung yang digunakan dalam penelitian ini disajikan pada Tabel 2 sedangkan distribusi kelas ditunjukkan pada Tabel 3.

Tabel 2. Karakteristik Dataset MIT-BIH *Arrhythmia Database*

Nama Dataset	Jumlah Sampel	Deskripsi
Record 100–234 (48 record)	Rekaman ± 30 menit per record	Sinyal EKG mentah (raw signal) single-lead dari MIT-BIH Arrhythmia Database dengan laju sampling 360 Hz, digunakan sebagai data awal untuk segmentasi denyut jantung dan ekstraksi fitur.
Dataset Denyut Jantung	± 100.000 denyut jantung	Data denyut jantung hasil segmentasi sinyal EKG berdasarkan puncak R, masing-masing berdurasi 600 ms (216 sample data), serta telah dipetakan ke dalam kelas AAMI (N, S, V, F).
Subset Data Latih	$\pm 80\%$ dari total denyut	Data denyut jantung hasil fusi fitur sinyal EKG dan fitur HRV time-domain yang digunakan untuk pelatihan model XGBoost.
Subset Data Uji	$\pm 20\%$ dari total denyut	Data denyut jantung yang digunakan untuk pengujian dan evaluasi performa model klasifikasi, terpisah sepenuhnya dari data pelatihan.

Tabel 3. Distribusi Kelas

Kelas	Deskripsi	Jumlah Data
N	Normal Beat	17.465
S	Supraventricular Ectopic Beat	249
V	Ventricular Ectopic Beat	219
Total		18.090

Dihitung dari interval R-R. Fitur Heart Rate Variability (HRV) yang digunakan meliputi *MeanRR*, *SDNN*, *RMSSD*, dan *pNN50*. Perhitungan HRV dilakukan dengan pendekatan *sliding window* merepresentasikan dinamika ritme jantung secara lokal di sekitar periode pengamatan. Pada penelitian ini, digunakan ukuran jendela (*window size*) selama 30 detik dengan pergeseran (*step size*) 15 detik, sehingga menghasilkan overlap sebesar 50% antar-jendela. Setiap jendela dibentuk berdasarkan deret interval RR normal yang diperoleh dari hasil deteksi puncak R dan hanya diproses apabila mengandung minimal 10 interval RR untuk memastikan kestabilan dalam perhitungan parameter HRV. Dari setiap jendela dihitung fitur-fitur HRV domain waktu, yaitu *MeanRR*, *SDNN*, *RMSSD*, dan *pNN50*. Pemilihan jendela 30 detik dilakukan untuk menangkap variasi ritme jantung jangka pendek yang relevan dengan deteksi aritmia, sementara overlap 50% digunakan untuk menjaga kontinuitas informasi temporal antar-jendela serta meningkatkan representativitas karakteristik HRV. Fitur-fitur HRV yang dihasilkan kemudian digabungkan dengan fitur morfologi sinyal EKG sebagai masukan bagi model XGBoost dalam proses klasifikasi aritmia.

3.2 Data Preprocessing dan EDA

Pada tahap preprocessing data dilakukan untuk memastikan kualitas data yang digunakan dalam proses pelatihan model klasifikasi aritmia. Pada tahap awal, dilakukan ekstraksi sinyal Elektrokardiogram (EKG) berdasarkan posisi puncak R yang diperoleh dari anotasi dataset. Selanjutnya, interval RR dihitung dari selisih waktu antar puncak R yang berurutan dan dilakukan penyaringan untuk menghilangkan nilai yang tidak fisiologis, yaitu interval RR di luar rentang 0,3 hingga 2,0 detik. Berdasarkan interval RR yang telah difilter, fitur Heart Rate Variability (HRV) dihitung menggunakan pendekatan *sliding window* dengan ukuran jendela 30 detik dan *step size* 15 detik sehingga menghasilkan overlap sebesar 50%. Setiap jendela harus mengandung minimal 10 interval RR agar perhitungan parameter HRV tetap stabil. Dari setiap jendela kemudian diekstraksi fitur-fitur HRV domain waktu, yaitu *MeanRR*, *SDNN*, *RMSSD*, dan *pNN50*. Setelah proses *preprocessing* selesai, dilakukan *Exploratory Data Analysis* (EDA) untuk memahami karakteristik dataset sebelum proses pelatihan model. Analisis distribusi kelas menunjukkan bahwa jumlah sampel pada masing-masing kelas aritmia tidak seimbang, dengan kelas Normal (N) memiliki jumlah sampel yang jauh lebih besar dibandingkan dengan kelas Supraventricular Ectopic Beat (S), Ventricular Ectopic Beat (V), Fusion Beat (F), dan Unknown

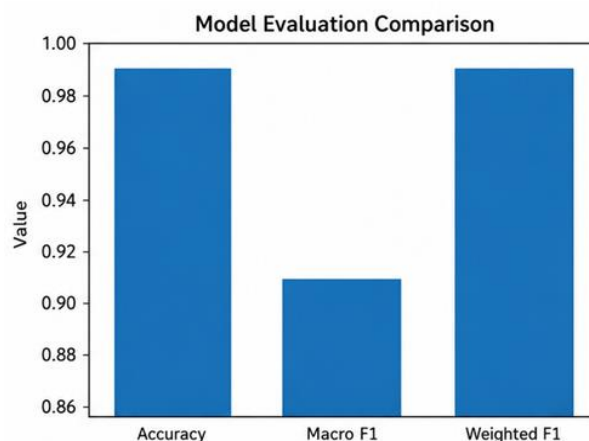
Beat (Q). Selain itu, analisis terhadap fitur HRV menunjukkan adanya perbedaan pola nilai MeanRR, SDNN, RMSSD, dan pNN50 pada setiap kelas aritmia. Perbedaan ini menunjukkan bahwa fitur HRV memiliki kemampuan untuk membantu membedakan karakteristik ritme jantung antarkelas. Melalui EDA, peneliti juga dapat memastikan bahwa data yang digunakan sudah sesuai dan tidak mengandung nilai yang tidak wajar yang dapat memengaruhi hasil klasifikasi.

3.3 Parameter Model

Parameter model XGBoost dalam penelitian ini ditetapkan berdasarkan karakteristik dataset Massachusetts Institute of Technology-Beth Israel Hospital (MIT-BIH) Arrhythmia Database dan kebutuhan klasifikasi multikelas. Fungsi objektif multi:softprob dipilih karena menghasilkan nilai probabilitas untuk setiap kelas sehingga prediksi ditentukan berdasarkan kelas dengan probabilitas tertinggi. Model dibangun melalui 300 iterasi boosting dengan kedalaman pohon maksimum sebesar 6, serta laju pembelajaran 0,1 agar konvergensi berlangsung bertahap dan stabil. Untuk mencegah *overfitting*, diterapkan subsampling data sebesar 80% per iterasi dan subsampling fitur sebesar 80% per pohon. Mengingat distribusi kelas yang tidak seimbang pada dataset, pembobotan sampel berbasis keseimbangan kelas diterapkan selama pelatihan agar model tidak condong pada kelas Normal yang mendominasi dan tetap mampu mengenali kelas minoritas secara memadai.

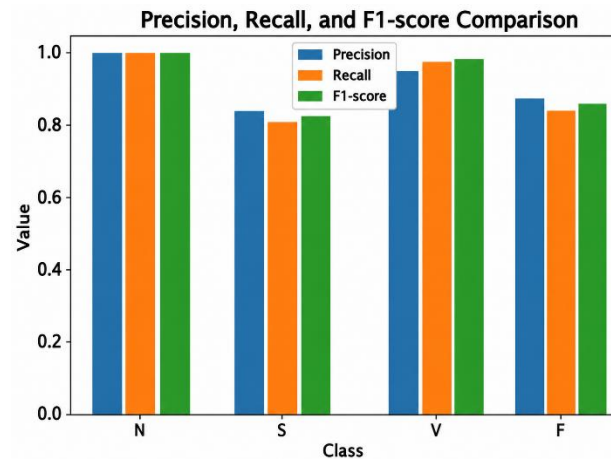
3.4 Analisis Hasil

Analisis dilakukan setelah model XGBoost dilatih menggunakan fitur Elektrokardiogram (EKG) berbasis denyut dan *Heart Rate Variability (HRV) time domain*. Evaluasi dilakukan untuk mengukur kemampuan model dalam mengklasifikasikan denyut jantung pada data uji. Hasil akurasi sebesar 99,15% menunjukkan bahwa model mampu mengenali pola aritmia dengan sangat baik pada data yang digunakan. Nilai ini tergolong sangat tinggi untuk klasifikasi multiclass, sehingga perlu dilihat lebih lanjut faktor-faktor yang mungkin memengaruhinya. Salah satu penyebab tingginya akurasi adalah penggunaan kombinasi fitur morfologi sinyal ECG dan fitur Heart Rate Variability (HRV) yang memberikan informasi lebih lengkap mengenai kondisi jantung. Selain itu, algoritma XGBoost memang dikenal mampu menangani data yang kompleks dan menemukan pola-pola penting dengan baik sehingga performanya sering lebih unggul dibandingkan metode klasifikasi konvensional. *macro F1-score* sebesar 91,07%. Dataset menunjukkan ketidakseimbangan distribusi kelas (*class imbalance*), yang ditandai dengan dominasi kelas *Normal (N)* sehingga nilai akurasi cenderung dipengaruhi oleh kelas mayoritas. Oleh karena itu, nilai *macro F1-score* sebesar 91,07% digunakan sebagai evaluasi tambahan untuk mengukur performa model secara lebih representatif pada seluruh kelas, termasuk kelas minoritas. Perbandingan Nilai Evaluasi Model XGBoost ditunjukkan pada Gambar 2.

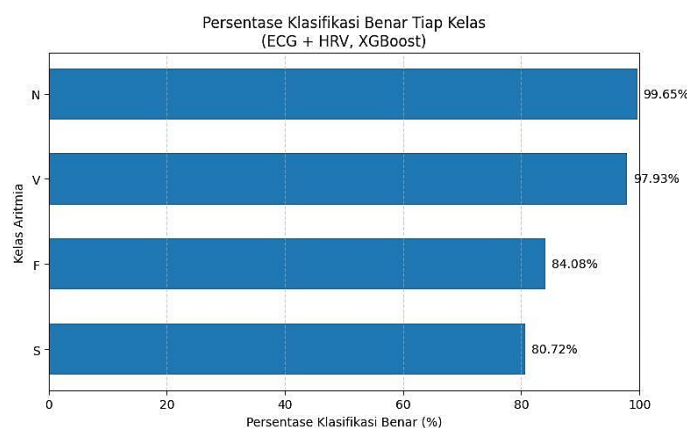


Gambar 2. Perbandingan Nilai Evaluasi Model XGBoost

Gambar 4 dan 5 menunjukkan bahwa model XGBoost memiliki performa yang bervariasi pada setiap kelas. Kelas *Normal (N)* memiliki nilai *F1-score* tertinggi sebesar 99,59%, diikuti kelas *Ventricular (V)* sebesar 98,24%. Sementara itu, kelas *Supraventricular (S)* dan *Fusion (F)* memiliki nilai *F1-score* yang lebih rendah, masing-masing sebesar 82,30% dan 84,14%. Hal ini dipengaruhi oleh distribusi data yang tidak seimbang. Kondisi tersebut menyebabkan model cenderung lebih optimal dalam mengenali kelas mayoritas, sehingga meningkatkan nilai akurasi secara keseluruhan, namun berdampak pada penurunan performa pada kelas minoritas. Meskipun demikian, penggunaan fitur HRV *time domain* seperti *MeanRR*, *SDNN*, *RMSSD*, dan *pNN50* membantu model menangkap variasi interval denyut jantung yang tidak terlihat secara langsung pada sinyal EKG mentah.



Gambar 4. Perbandingan Precision, Recall, dan F1-score per Kelas



Gambar 5. Performa Algoritma XGBoost setiap kelas

Confusion matrix digunakan untuk menganalisis kesalahan prediksi secara lebih rinci. Hasil analisis menunjukkan bahwa kesalahan prediksi lebih banyak terjadi pada kelas minoritas, yang mengindikasikan bahwa ketidakseimbangan distribusi data masih menjadi tantangan utama dalam model. Faktor lain yang dapat berkontribusi terhadap tingginya hasil adalah karakteristik dataset MIT-BIH Arrhythmia Database yang memiliki kualitas sinyal cukup baik dan telah banyak digunakan dalam penelitian klasifikasi aritmia. Jika data pelatihan dan data pengujian memiliki pola yang sangat mirip, model akan lebih mudah melakukan prediksi dengan benar. Oleh karena itu, untuk memastikan bahwa model benar-benar memiliki kemampuan generalisasi yang baik, pengujian pada dataset lain atau data pasien yang berbeda tetap diperlukan. Dengan demikian, nilai akurasi 99,15% dapat dianggap sangat menjanjikan, tetapi masih perlu dievaluasi lebih lanjut untuk memastikan performanya tetap stabil dalam kondisi dunia nyata.

3.5 Pembahasan

Penelitian menunjukkan bahwa model XGBoost yang dibangun menggunakan kombinasi fitur sinyal Elektrokardiogram (EKG) dan fitur Heart Rate Variability (HRV) domain waktu mampu mencapai akurasi sebesar 99,16%, weighted F1-score sebesar 99,15%, serta macro F1-score sebesar 91,07%. Tingginya nilai akurasi dan weighted F1-score menunjukkan bahwa model mampu mengenali sebagian besar pola denyut jantung dengan tepat. Kinerja yang tinggi ini diduga dipengaruhi oleh kemampuan algoritma XGBoost dalam membangun kumpulan pohon keputusan secara bertahap sehingga mampu menangkap hubungan kompleks antara karakteristik morfologi sinyal EKG dan pola variasi interval R-R yang direpresentasikan oleh fitur HRV. Penambahan fitur HRV memberikan informasi tambahan mengenai dinamika ritme jantung yang tidak sepenuhnya tercermin dalam bentuk gelombang EKG. Fitur seperti MeanRR, SDNN, RMSSD, dan pNN50 memungkinkan model mengenali perubahan variabilitas denyut jantung yang sering terjadi pada kondisi aritmia. Dengan demikian, kombinasi fitur EKG dan HRV menghasilkan representasi data yang lebih komprehensif dibandingkan dengan penggunaan fitur EKG saja.

Meskipun demikian, terdapat perbedaan yang cukup besar antara nilai weighted F1-score dan macro F1-score. Kondisi ini mengindikasikan bahwa performa model belum merata di seluruh kelas. Nilai weighted F1-score yang tinggi dipengaruhi oleh dominasi kelas dengan jumlah data yang besar, sedangkan nilai macro

F1-score memberikan gambaran bahwa beberapa kelas minoritas masih memiliki tingkat kesalahan klasifikasi yang relatif lebih tinggi. Fenomena ini menunjukkan bahwa ketidakseimbangan distribusi data (class imbalance) masih menjadi tantangan dalam proses klasifikasi aritmia.

Hasil penelitian ini sejalan dengan penelitian sebelumnya yang menunjukkan bahwa algoritma XGBoost memiliki kemampuan yang baik dalam tugas klasifikasi data medis. Namun, penelitian ini menawarkan kontribusi tambahan melalui integrasi fitur HRV domain waktu dengan fitur sinyal EKG sehingga model dapat menangkap informasi morfologi dan dinamika ritme jantung secara bersamaan. Pendekatan tersebut terbukti mampu menghasilkan performa klasifikasi yang sangat tinggi pada dataset MIT-BIH Arrhythmia Database.

Sebagai evaluasi tambahan, model diuji menggunakan data pengujian yang terpisah dari data pelatihan untuk mengurangi risiko overfitting. Meskipun hasil yang diperoleh menunjukkan performa yang sangat baik, penelitian selanjutnya dapat menerapkan teknik validasi yang lebih komprehensif, seperti k-fold cross-validation atau metode penyeimbangan data, guna memastikan stabilitas model dan meningkatkan kemampuan klasifikasi pada kelas minoritas. Dengan demikian, model yang dikembangkan berpotensi diterapkan sebagai sistem pendukung deteksi dini aritmia yang lebih akurat dan andal.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan model klasifikasi aritmia menggunakan kombinasi fitur sinyal Elektrokardiogram (EKG) berbasis denyut jantung dan fitur Heart Rate Variability (HRV) domain waktu dengan algoritma Extreme Gradient Boosting (XGBoost). Tujuan penelitian untuk meningkatkan representasi data dalam proses klasifikasi aritmia melalui integrasi informasi morfologi sinyal EKG dan dinamika ritme jantung telah tercapai. Hasil pengujian menunjukkan bahwa kombinasi fitur EKG dan HRV mampu menghasilkan performa klasifikasi yang sangat baik dengan akurasi sebesar 99,16%, sehingga efektif dalam membedakan denyut jantung normal dan aritmia pada dataset Massachusetts Institute of Technology-Beth Israel Hospital (MIT-BIH) Arrhythmia Database. Temuan ini menunjukkan bahwa informasi variabilitas denyut jantung yang direpresentasikan oleh fitur HRV dapat melengkapi informasi morfologi sinyal EKG dan meningkatkan kemampuan model dalam mengenali karakteristik aritmia.

Kontribusi utama (novelty) penelitian ini terletak pada integrasi fitur sinyal EKG berbasis denyut jantung dengan fitur HRV domain waktu yang dihitung menggunakan pendekatan *sliding window* dalam kerangka klasifikasi berbasis XGBoost. Pendekatan tersebut memungkinkan model memanfaatkan informasi bentuk gelombang dan pola ritme jantung secara bersamaan sehingga menghasilkan representasi data yang lebih komprehensif dibandingkan dengan penggunaan salah satu jenis fitur secara terpisah. Meskipun memperoleh performa yang tinggi, penelitian ini masih menghadapi tantangan berupa ketidakseimbangan distribusi data pada beberapa kelas aritmia. Oleh karena itu, penelitian selanjutnya dapat mengkaji penerapan metode penyeimbangan data, validasi yang lebih komprehensif, serta eksplorasi fitur HRV dan algoritma klasifikasi lainnya untuk meningkatkan stabilitas dan kemampuan generalisasi model.

REFERENSI

- [1] K. Setiyanto and A. Marwanto, "Analisis Deteksi Aritmia Jantung Berbasis Spektrogram Dan Convolutional Neural Network (CNN) Untuk Asuhan Keperawatan Analysis of Cardiac Arrhythmia Detection Using Spectrograms and Convolutional Neural Networks (CNN) for Nursing Care," vol. 7, no. 1, pp. 12–18, 2026.
- [2] L. F. Tampubolon, A. Ginting, and F. E. Saragi Turnip, "Gambaran Faktor yang Mempengaruhi Kejadian Penyakit Jantung Koroner (PJK) di Pusat Jantung Terpadu (PJT)," *J. Ilm. Permas J. Ilm. STIKES Kendal*, vol. 13, no. 3, pp. 1043–1052, 2023.
- [3] A. Hanifa, "Sistem Pakar Untuk Mendiagnosa Penyakit Aritmia Menggunakan Certainty Factor," *J. SANTI - Sist. Inf. dan Tek. Inf.*, vol. 2, no. 1, pp. 41–48, 2022.
- [4] H. Hidayat, A. Sunyoto, and H. Al Fatta, "Klasifikasi Penyakit Jantung Menggunakan Random Forest Clasifier," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 7, no. 1, pp. 31–40, 2023.
- [5] Q. Ayuni, R. C. Wihandika, and N. Yudistira, "Klasifikasi Aritmia Dari Hasil Elektrokardiogram Menggunakan Metode Support Vector Machine," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 6, pp. 2163–2170, 2021.
- [6] B. W. Putra, R. F. I. Isnanto, P. Sari, A. Saputra, M. R. Sanjaya, and A. N. S. Mukti, "Klasifikasi arritmia pada sinyal EKG menggunakan deep neural network," *Jupiter*, vol. 13, pp. 29–38, 2021.
- [7] J. Given Hamonangan, A. S. Rahmi, S. J. Lase, N. Suhendi Syafei, and A. Turnip, "JIIF (Jurnal Ilmu Dan Inovasi Fisika) Deteksi Dini Aritmia Menggunakan K-Nearest Neighbour," vol. 08, no. 01, pp. 86–95, 2024.
- [8] D. Hasbi, P. Studi, T. Elektro, F. T. Industri, and U. I. Indonesia, "Analisis Heart Rate Variability (Hrv) Antara Elektrokardiogram (Ekg) Dengan Stetoskop Elektronik Littmann 3200," *Skripsi*, vol. 1, p. 35, 2020.

- [9] Y. Ramadhan and S. Mandala, "Analysis of Electrocardiogram Dynamic Features for Arrhythmia Classification," vol. 8, no. 2, pp. 204–212, 2023.
- [10] D. Pang *et al.*, "Rancang Bangun Oksimeter Pulsa yang Diintegrasikan dengan EKG sebagai Alat Ukur Multimodal untuk Analisis HRV pada L .)," vol. 11, no. 1, pp. 43–46, 2022.
- [11] A. I. Novianto, S. I. Purnama, and I. Hikmah, "Analisis dan Implementasi Perangkat Elektrokardiogram (EKG) Berbasis Internet of Things Analysis and Implementation Electrocardiogram (ECG) Device Based on Internet of Things," vol. 8275, no. July, pp. 150–163, 2025.
- [12] A. Melalui, S. Ekg, M. Arsitektur, and Y. J. Nurriski, "Optimasi Deep Convolutional Neural Network (Deep CNN) untuk Deteksi," vol. 46, no. 1, pp. 10–20, 2023.
- [13] P. Guru, S. Dasar, U. Kristen, and I. Toraja, "Analisis Fitur HRV pNN50 pada Sinyal Psikofisiologis Marah Manusia," vol. 6, no. 2, pp. 177–179, 2022.
- [14] F. I. Adani and H. Amalia, "Penerapan Metode Algoritma XGBoost untuk Prediksi Risiko Penyakit Jantung," vol. 5, no. 2, pp. 117–125, 2025.
- [15] D. M. Pratiwi and L. Mufidah, "Perbandingan Metode Decision Tree Classifier dan XGBoost Classifier Dalam Memprediksi Penyakit Jantung," pp. 991–1000, 2024.
- [16] R. Soelistijadi *et al.*, "Pemodelan Prediktif Menggunakan Metode Ensemble Learning XGBoost dalam Peningkatan Akurasi Klasifikasi Penyakit Ginjal," vol. 5, no. 4, pp. 1866–1875, 2024.
- [17] A. Brahmandjati, A. M. A. Rahim, and F. Asharudin, "Optimasi Prediksi Diabetes Dengan Algoritma XGBoost Dan Teknik Preprocessing Data," vol. 3, no. 1, pp. 116–125, 2024.
- [18] B. S. C. Putra, I. Tahyudin, B. A. Kusuma, and K. N. Isnaini, "Efektivitas Algoritma Random Forest, XGBoost, dan Logistic Regression dalam Prediksi Penyakit Paru-paru," *Techno.Com*, vol. 23, no. 4, pp. 909–922, 2024.
- [19] M. Hasanudin, S. Dwiasnati, and W. Gunawan, "Pelatihan Datascience pada Pra-Pemrosesan Data untuk Siswa SMK Media Informatika - Jakarta," *J. Pengabd. Pada Masy.*, vol. 9, no. 4, pp. 882–888, 2024.
- [20] I. Terapan, U. Telkom, T. Elektro, and U. Telkom, "Analisis Fitur Domain Waktu ECG Heart Rate Variability Berdasarkan Gain Informasi," vol. 10, no. 2, 2022.
- [21] C. Widowati and K. Purwantini, "Inovasi Sensor Wearable untuk Monitoring Kesehatan Mental melalui Variabilitas Denyut Jantung kesehatan mental secara berkelanjutan . Perangkat seperti smartwatch , gelang pintar , dan memberikan rekomendasi yang lebih personal (Mitra & Ul Rehman , 2024).," 2024.
- [22] M. Cahya and A. M. Simarmata, "I a l a k n d e k," vol. 31, pp. 232–236, 2026.
- [23] P. Studi, T. Informatika, F. Sains, and U. P. Indonesia, "Analisis Sinyal EKG Untuk Deteksi Gangguan Jantung Pada," vol. 30, pp. 361–368, 2025.
- [24] R. S. Hidayana *et al.*, "Sistem Prediksi Risiko Penyakit Jantung Berbasis Machine Learning dan Framework Streamlit," vol. 8, no. 6, pp. 3672–3680, 2025.
- [25] M. Syukron, R. Santoso, and T. Widiarihi, "Perbandingan Metode Smote Random Forest Dan Smote Xgboost Untuk Klasifikasi Tingkat Penyakit Hepatitis C Pada Imbalance Class Data," *J. Gaussian*, vol. 9, no. 3, pp. 227–236, 2020.