

Klasifikasi Aritmia pada sinyal ECG menggunakan Ensemble Machine Learning dengan kerangka kerja Bootstrap Aggregating

1st Jennifer Celine Winatra

Penulis-Teknik Biomedis

Telkom University

Bandung, Indonesia

jenniferceline@student.telkomuniversity.ac.id

2nd Tito Waluyo Purboyo

Pembimbing-Teknik Biomedis

Telkom University

Bandung, Indonesia

titowaluyo@telkomuniversity.ac.id

3rd Dziban Naufal

Pembimbing-Teknik Biomedis

Telkom University

Bandung, Indonesia

dziban@telkomuniversity.ac.id

Abstrak — Penyakit jantung masih menjadi penyebab utama kematian global, terutama di negara-negara berkembang. Aritmia, salah satu jenis gangguan irama jantung, berperan besar dalam angka kematian ini. Dengan meningkatnya jumlah kasus fibrilasi atrium (FA), deteksi dini menjadi sangat penting. Penelitian ini mengeksplorasi penggunaan ensemble machine learning dengan kerangka kerja bootstrap aggregating (bagging) untuk klasifikasi sinyal elektrokardiogram (EKG). Dataset MIT-BIH digunakan sebagai data latih dengan berbagai pendekatan bagging seperti Traditional Bagging, Bayesian Bagging, Feature Bagging, dan Random Subspace. Hasil menunjukkan bahwa Bayesian Bagging memberikan performa terbaik (akurasi 93.40%, F1-score 93.50%, dan AUC 99.30%) dengan waktu inferensi hanya 0.10 detik. Keunggulan ini dipengaruhi oleh mekanisme pembobotan Dirichlet yang mampu mempertahankan kontribusi tiap sampel secara efektif. Studi ini menunjukkan bahwa metode bagging menawarkan keseimbangan optimal antara akurasi, stabilitas metrik, efisiensi komputasi, serta potensi penerapan klinis untuk membantu diagnosis aritmia secara lebih cepat, akurat, dan andal di berbagai kondisi.

Kata kunci — aritmia, EKG, bootstrap aggregating, ensemble machine learning, Bayesian Bagging, MIT-BIH

I. PENDAHULUAN

Penyakit kardiovaskular (cardiovascular disease/CVD) merupakan salah satu penyebab utama kematian di seluruh dunia. Kondisi ini mencakup berbagai gangguan pada jantung dan pembuluh darah, termasuk penyakit jantung koroner, hipertensi, dan stroke [1]. Menurut laporan Kementerian Kesehatan Indonesia, pada tahun 2023 terdapat 887.531 kasus penyakit jantung, meningkat sebesar 0,85% dibandingkan tahun sebelumnya [2]. Penyakit ini khususnya berdampak besar di negara-negara berpenghasilan rendah dan menengah.

Salah satu jenis penyakit kardiovaskular yang cukup umum adalah aritmia, yaitu gangguan irama jantung yang dapat menimbulkan risiko serius seperti stroke atau gagal jantung. Aritmia yang paling sering dijumpai adalah fibrilasi

atrium (FA), dengan jumlah penderita secara global diperkirakan mencapai 46,3 juta orang [3]. Di Indonesia, kasus FA diproyeksikan meningkat signifikan seiring pertumbuhan jumlah penderita penyakit jantung secara umum.

Deteksi aritmia secara akurat dapat dilakukan menggunakan elektrokardiogram (EKG), yakni alat non-invasif yang merekam aktivitas listrik jantung. Sinyal EKG memiliki tiga komponen utama yaitu gelombang P, kompleks QRS, dan gelombang T, yang dapat digunakan sebagai indikator pola detak jantung. Dalam penelitian ini, klasifikasi sinyal EKG dilakukan berdasarkan standar Association for the Advancement of Medical Instrumentation (AAMI) dengan memanfaatkan data dari MIT-BIH Arrhythmia Database [4].

Namun, analisis sinyal EKG tidaklah sederhana karena dipengaruhi oleh variasi fisiologis dan gangguan (noise) dari alat perekam. Oleh sebab itu, pendekatan machine learning, khususnya teknik ensemble seperti bagging (bootstrap aggregating), dipertimbangkan sebagai solusi untuk mengatasi tantangan ini. Bagging merupakan metode yang melatih beberapa model dasar (base learner) pada subset data yang berbeda, lalu menggabungkan hasilnya untuk meningkatkan akurasi dan stabilitas [5].

Penelitian ini bertujuan untuk mengevaluasi performa metode bagging dalam mengklasifikasikan aritmia berbasis sinyal EKG. Berbagai arsitektur bagging, termasuk Traditional Bagging, Bayesian Bagging, Feature Bagging, dan Random Subspace diuji untuk mengetahui kontribusinya terhadap peningkatan akurasi deteksi aritmia.

II. KAJIAN TEORI

A. Dataset

Dataset yang digunakan dalam penelitian ini bersumber dari MIT-BIH Arrhythmia Database yang dapat diakses melalui situs Physionet. Dataset ini telah banyak digunakan

dalam penelitian aritmia dan menjadi tolok ukur standar karena keandalannya yang telah terbukti. Dataset ini terdiri dari data rekaman EKG dari 47 pasien dan mencakup 97.360 denyut jantung normal, 3.026 denyut yang dikategorikan sebagai Supraventricular, 7.235 denyut sebagai Ventricular, 1.784 sebagai Fusion Beats, serta 505 denyut yang tidak terklasifikasi. Meskipun data ini menyediakan dua jenis sinyal rekaman yaitu MLII dan V1, penelitian ini hanya menggunakan sinyal dari lead MLII. [6]

B. Metode

Setelah data diperoleh, dilakukan analisis awal untuk mendapatkan gambaran umum melalui Exploratory Data Analysis (EDA) tanpa menarik kesimpulan langsung. Analisis ini mencakup perhitungan nilai mean, median, modus, standar deviasi, dan variansi. Langkah awal pre-processing dimulai dengan teknik Random Undersampling yang bertujuan menyeimbangkan jumlah data antar kelas dan mempercepat proses komputasi. Setiap kelas diseimbangkan hingga 500 data (total 2.500 data), dengan pemilihan data secara acak namun merata dari berbagai pasien, mencakup 48 pasien untuk kelas N, 32 pasien untuk kelas S, 37 pasien untuk kelas V, 20 pasien untuk kelas F, dan 7 pasien untuk kelas Q.

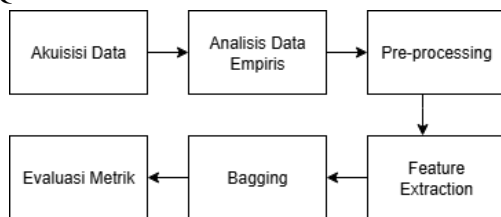


Fig. 1 Framework sistem penelitian

Selanjutnya dilakukan proses pembersihan sinyal, dimulai dari detrending menggunakan polinomial orde-6 untuk mengoreksi deviasi baseline. Pengurangan noise dilakukan menggunakan metode Discrete Wavelet Transform (DWT) dengan wavelet Daubechies 6 (db6) pada level dekomposisi ke-4 dan ambang batas 35%. Koefisien wavelet yang berada di bawah ambang tersebut dihilangkan sebagai noise. Pemilihan db6 didasarkan pada kemiripan morfologi dengan kompleks PQRST pada sinyal EKG.

Untuk sinyal yang memiliki rentang frekuensi luas, dilakukan pemulusan tambahan menggunakan filter Savitzky-Golay dengan jendela 150 milidetik dan polinomial orde-3. Pemilihan parameter ini bertujuan untuk menjaga bentuk morfologi sinyal tanpa menyebabkan distorsi.

Setelah noise dihilangkan, dilakukan segmentasi sinyal berdasarkan puncak R. Deteksi puncak R dilakukan dua tahap: pertama, dengan fungsi `find_peaks` dari pustaka SciPy yang menetapkan jarak antar puncak minimal 200 ms; kedua, puncak kandidat disempurnakan dengan mencari nilai maksimum lokal di sekitar posisi awal. Segmentasi sinyal kemudian dilakukan dengan mengekstraksi jendela waktu mengelilingi puncak R untuk mencakup satu siklus jantung

lengkap (kompleks PQRST). Rentang segmentasi meliputi 100 ms sebelum puncak gelombang P hingga 200 ms setelah puncak gelombang T.

Setelah segmentasi, fitur diekstraksi dari setiap segmen menggunakan pustaka Time Series Feature Extraction Library (TSFEL). TSFEL membagi fitur ke dalam tiga domain utama: statistik, temporal, dan spektral. Lebih dari 60 jenis fitur tersedia dan sebanyak 314 fitur diekstraksi dari tiap segmen EKG menggunakan parameter default. Namun, hanya 156 fitur yang berhasil digunakan dalam proses selanjutnya.

C. Bagging

Bootstrap aggregating, atau yang biasa disebut bagging merupakan salah satu penggabungan algoritma machine learning. Algoritma ini menggunakan beberapa base learner yang kemudian melakukan klasifikasi pada setiap subset secara mandiri. Bagging dirancang untuk meningkatkan akurasi dan stabilitas algoritma pada saat klasifikasi. Algoritma ini baik digunakan untuk mengurangi varians dan menghindari overfitting [7]. Fungsi dari metode bagging adalah (1):

$$F_n = \sum_{i=1}^n \omega_i \delta_{Z_i} \quad (1)$$

δ_{Z_i} adalah ukuran probabilitas yang merosot pada Z_i , ω_i adalah bobot pada Z_i , $\omega_i \geq 0$, $\sum \omega_i = 1$. Dalam bootstrap, distribusi $T(F)$ diperoleh dengan membuat replikasi bootstrap secara berulang-ulang, di mana satu replikasi bootstrap adalah sampel yang diambil dengan penggantian ukuran n dari Z_1 hingga Z_n . Distribusi bootstrap dari $T(F)$ didasarkan pada pertimbangan semua replikasi bootstrap yang mungkin terjadi $T(Fn^{(r)})$ di mana $\omega_i^{(r)}$ di $Fn^{(r)}$ sesuai dengan proporsi waktu Z_i muncul di replikasi bootstrap ke- r , dengan $\omega_i^{(r)}$ memiliki nilai di $\{0, 1/n, \dots, n/n\}$. [9]

1) *Traditional Bagging*: Traditional Bagging adalah teknik ensemble yang bekerja dengan mengambil subset data secara acak dengan bagging untuk melatih setiap model dasar. Setelah semua model dasar dilatih, hasil prediksi dari model-model tersebut digabungkan menggunakan voting untuk klasifikasi atau rata-rata untuk regresi. Pendekatan ini bertujuan untuk mengurangi variansi model dan meningkatkan stabilitas serta generalisasi, terutama pada dataset yang rawan overfitting.

2) *Bayesian Bagging*: Bayesian Bagging adalah pendekatan yang menggabungkan prinsip-prinsip Bayesian dengan teknik bagging tradisional. Dalam metode ini, beberapa model klasifikasi dibangun berdasarkan sampel bootstrap dari data pelatihan, dan setiap model memberikan kontribusi terhadap keputusan akhir berdasarkan probabilitas yang dihasilkan. Pendekatan ini memungkinkan penggabungan ketidakpastian dari setiap model, sehingga dapat meningkatkan stabilitas dan akurasi prediksi. Bayesian

Bagging sering digunakan ketika data memiliki karakteristik yang tidak pasti atau ketika model dasar memiliki bias yang tinggi.

3) *Feature Bagging*: Feature Bagging adalah pendekatan yang mirip dengan bagging tradisional, namun yang dibedakan adalah pengambilan subset acak fitur dari data untuk setiap model dasar. Alih-alih memilih subset data secara acak seperti pada teknik bagging biasa, dalam Feature Bagging, hanya subset fitur yang diambil secara acak dan diterapkan pada model. Feature Bagging berfokus pada pemilihan subset fitur secara acak untuk setiap model dalam ensemble. Dalam metode ini, setiap model dilatih menggunakan hanya sebagian dari fitur yang tersedia, bukan seluruh dataset. Ini membantu mengurangi overfitting dan meningkatkan generalisasi model. Feature Bagging sangat efektif dalam situasi di mana data memiliki banyak fitur redundan atau tidak relevan, karena dapat mengekstrak informasi yang lebih relevan dari subset fitur yang dipilih.

4) *Random Subspace*: Random Subspace adalah teknik yang digunakan untuk meningkatkan keberagaman model dalam ensemble dengan memilih subset acak dari fitur untuk setiap model dalam ensemble. Dalam Random Subspace, untuk setiap model dasar, hanya sejumlah fitur acak yang digunakan untuk melatih model tersebut, sementara contoh data tetap sama. Dalam Random Subspace, setiap model dilatih pada subset acak dari fitur tanpa melakukan sampling pada baris data (data points). Metode ini memungkinkan penggunaan semua data pelatihan tetapi hanya memanfaatkan sebagian dari fitur untuk membangun setiap model. Hal ini dapat meningkatkan kinerja model dengan mengurangi variabilitas dan meningkatkan robustnes terhadap noise dalam data. Random Subspace sering digunakan dalam kombinasi dengan algoritma pembelajaran seperti pohon keputusan untuk menciptakan ensemble yang kuat

Tabel 1 Karakteristik Algoritma Bagging

Model	Sampel diambil	Keberagaman	Output
Traditional Bagging	√	Mengurangi variansi	Meningkatkan stabilitas
Bayesian Bagging	–	Memberi bobot	Mengurangi overfitting
Feature Bagging	√	Menambah keberagaman	Subset fitur berbeda
Random Subspace	–	Mengurangi korelasi	Meningkatkan variasi fitur

Evaluasi digunakan untuk menentukan seberapa efektif sebuah model. Terdapat beberapa hal dalam evaluasi, yaitu:

A. Model klasifikasi menjalani evaluasi menggunakan serangkaian metrik yang kuat. Efektivitas dan efisiensi keseluruhannya dinilai dengan memanfaatkan akurasi (2), area di bawah kurva Karakteristik Operasional Penerima (ROC) (AUC), dan waktu pengujian rata-rata. Untuk menganalisis secara spesifik kemampuan model dalam

mengidentifikasi setiap kelas dengan benar, recall (3), precision (4), dan F1-score (5) diterapkan.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \quad (2)$$

$$Precision = \frac{TP}{TP + FN} \times 100\% \quad (3)$$

$$Recall = \frac{TP}{TP + FP} \times 100\% \quad (4)$$

$$F1\ Score = 2 \times \frac{precision \times recall}{precision + recall} \times 100\% \quad (5)$$

III. HASIL PENELITIAN

A. Perangkat Penelitian

Kumpulan data sebanyak 2500 data, yang berisi 500 data per kelas, dipartisi menjadi 80% set pelatihan dan 20% set validasi menggunakan pustaka scikit-learn. Pengujian algoritma juga dilakukan menggunakan personal computer (PC) yang dilengkapi prosesor AMD Ryzen 5 5600G dengan grafis Radeon, memiliki kecepatan 3,9 GHz, 12 CPU, serta RAM 16 GB, yang berfungsi sebagai basis komputasi tambahan untuk pemrosesan data.

B. Hasil metrik

Berdasarkan hasil metrik yang didapatkan dari masing-masing model berdasarkan standar AAMI, yaitu N (Normal), S (Supraventricular), V (Ventricular), F (Fusion), dan Q (Unclassified) seperti pada tabel 3.

Tabel 2 Hasil Metrik

Model	Kelas	Precision	Recall	F1 Score
Traditional Bagging	N	0.94	0.92	0.93
	S	0.93	0.95	0.94
	V	0.88	0.88	0.88
	F	0.94	0.90	0.92
	Q	0.93	0.98	0.95
Bayesian Bagging	N	0.94	0.93	0.93
	S	0.94	0.95	0.95
	V	0.88	0.90	0.89
	F	0.97	0.92	0.94
	Q	0.95	0.99	0.97
Feature Bagging	N	0.93	0.93	0.93
	S	0.93	0.94	0.94
	V	0.88	0.90	0.89
	F	0.96	0.90	0.93
	Q	0.94	0.98	0.96
Random Subspace	N	0.94	0.93	0.93
	S	0.93	0.95	0.94
	V	0.90	0.90	0.90
	F	0.94	0.91	0.93
	Q	0.95	0.98	0.96

Hasil evaluasi menunjukkan bahwa model Bayesian Bagging secara konsisten menghasilkan nilai metrik yang

paling tinggi dan stabil di seluruh kelas. Model ini memperoleh F1 score tertinggi sebesar 0,97 pada kelas F dan 0,96 pada kelas Q, serta precision dan recall yang seimbang dan tinggi di semua kelas, menunjukkan kapabilitas yang sangat baik dalam mengidentifikasi pola aritmia.

Model Traditional Bagging juga menunjukkan performa yang baik namun sedikit menurun pada kelas V, dengan F1 score hanya sebesar 0,88. Hal ini mengindikasikan bahwa model ini masih kesulitan dalam membedakan denyut jantung jenis Ventricular dari kelas lainnya. Feature Bagging memberikan hasil yang mirip dengan Traditional Bagging, namun sedikit lebih baik pada kelas V, meskipun tidak setinggi performa Bayesian Bagging. Sementara itu, model Random Subspace menampilkan performa yang kompetitif, terutama pada kelas Q dengan F1 score sebesar 0,96 dan kelas F sebesar 0,93, tetapi memiliki penurunan pada kelas S dan V.

Secara keseluruhan, F1 score digunakan sebagai metrik utama karena mencerminkan keseimbangan antara precision dan recall, yang penting dalam konteks klasifikasi medis. Hasil evaluasi menunjukkan bahwa Bayesian Bagging tidak hanya unggul dalam akurasi keseluruhan, tetapi juga memiliki performa per kelas yang paling merata dan stabil. Dengan demikian, model ini dapat dianggap sebagai pendekatan yang paling efektif untuk diterapkan dalam sistem klasifikasi aritmia berbasis sinyal EKG.

IV. PEMBAHASAN

Hasil yang diperoleh dari penelitian ini memperlihatkan keunggulan penggunaan pendekatan ensemble machine learning, khususnya metode bagging, dalam menangani tantangan utama pada klasifikasi sinyal EKG, seperti ketidakseimbangan data dan variabilitas sinyal antarindividu. Bayesian Bagging menjadi metode yang paling menonjol karena mampu menghasilkan metrik evaluasi yang tinggi secara konsisten pada seluruh kelas. Hal ini menunjukkan bahwa mekanisme pembobotan probabilistik seperti distribusi Dirichlet dalam proses voting dapat meningkatkan ketahanan model terhadap noise dan outlier pada data medis.

Selain itu, model-model lain seperti Traditional Bagging dan Random Subspace tetap menunjukkan performa yang dapat diterima, meskipun tidak sebaik Bayesian Bagging. Hal ini menegaskan pentingnya pemilihan metode ensemble yang sesuai dengan karakteristik data. Perbedaan performa antar kelas juga mencerminkan adanya variasi kompleksitas sinyal dari masing-masing kategori aritmia, di mana kelas V (Ventricular) dan kelas S (Supraventricular) cenderung memberikan tantangan lebih besar dibanding kelas N dan Q.

Penting juga untuk dicatat bahwa pemrosesan awal yang teliti, seperti detrending, denoising, dan segmentasi berdasarkan puncak R, memberikan kontribusi signifikan

terhadap keberhasilan klasifikasi. Proses ini memungkinkan ekstraksi fitur yang lebih representatif melalui TSFEL. Dengan jumlah fitur yang cukup besar (156 fitur), sistem memperoleh pemahaman kontekstual yang baik terhadap pola gelombang PQRS.

Ke depan, pengembangan sistem klasifikasi aritmia dapat difokuskan pada integrasi model dengan perangkat EKG portable secara real-time, serta optimalisasi efisiensi komputasi agar tetap responsif pada lingkungan klinis. Evaluasi tambahan menggunakan dataset yang lebih bervariasi dan pengujian pada data uji eksternal juga penting untuk memastikan generalisasi model.

V. KESIMPULAN

Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa metode ensemble machine learning, khususnya Bayesian Bagging, merupakan pendekatan yang sangat efektif dalam klasifikasi aritmia berdasarkan sinyal elektrokardiogram. Metode ini mampu memberikan akurasi tinggi, metrik evaluasi yang stabil di seluruh kelas, serta waktu inferensi yang efisien. Proses pre-processing dan ekstraksi fitur yang dilakukan secara sistematis juga memberikan kontribusi besar terhadap performa model. Dengan demikian, pendekatan ini sangat menjanjikan untuk diterapkan dalam sistem deteksi dini aritmia berbasis EKG, terutama dalam konteks klinis yang membutuhkan keandalan dan efisiensi secara bersamaan.

REFERENSI

- [1] World Health Organization. (2021). Cardiovascular diseases (CVDs). <https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-cvds>
- [2] Kementerian Kesehatan Republik Indonesia, B. K. P. K. (2023). Survei Kesehatan Indonesia (SKI) 2023. [https://www.badankebijakan.kemkes.go.id/hasil-ski-2023/\[3\]](https://www.badankebijakan.kemkes.go.id/hasil-ski-2023/[3]) Indonesia Heart Rhythm Society. (2023, October 13). Satu Dekade InaHRS: An Overview and Outlook. <https://inahrs.or.id/news/satu-dekade-inahrs-an-overview-and-outlook>
- [4] Zakaria, H., Nurdiniyah, E. S. H., Kurniawati, A. M., Naufal, D., & Sutisna, N. (2024). Morphological Arrhythmia Classification Based on Inter-patient and Two Leads ECG using Machine learning. IEEE Access. <https://doi.org/10.1109/ACCESS.2024.3469640>.
- [5] Ahmet Mert, N. K. & A. A. (2012). Evaluation of bagging ensemble method with time-domain feature extraction for diagnosing of arrhythmia beats. 24, 317–326. <https://link.springer.com/article/10.1007/s00521-012-1232-7>
- [6] George B. Moody. (1997). MIT-BIH Arrhythmia Database Directory. Harvard-MIT Division of Health Sciences and Technology Biomedical Engineering Center. <https://archive.physionet.org/physiobank/database/html/mitdbdir/mitdbdir.htm>
- [7] Wikipedia. (2022). Bootstrap aggregating. https://id.wikipedia.org/wiki/Bootstrap_aggregating
- [8] Evangelista, E. (2023). An Optimized Bagging Ensemble Learning Approach Using BESTrees for Predicting Students' Performance. International Journal of Emerging Technologies in Learning, 18(10), 150–165. <https://doi.org/10.3991/ijet.v18i10.38115>