



Seminar Nasional LPPM UNSOED 2025

"Pengembangan Sumberdaya Perdesaan dan Kearifan Lokal Berkelanjutan"

8 Oktober 2025

Purwokerto

ANALISIS KOMPARATIF ALGORITMA K-NEAREST NEIGHBORS DAN LIGHTGBM UNTUK KLASIFIKASI KONTRAKSI JANTUNG PREMATUR

**Elsa Sari Hayunah Nurdiniyah^{1*}, A'isya Nur Aulia Yusuf¹, Norma Amalia¹, Widhiatmoko
Herry Purnomo¹, dan Azizah Najda Hafizha¹**

¹Program Studi Teknik Elektro, Fakultas Teknik, Universitas Jenderal Soedirman

*Corresponding Author: elsa.nurdiniyah@unsoed.ac.id

ABSTRAK

Kontraksi prematur merupakan salah satu kelainan jantung dimana detak jantung berdetak lebih dulu dibanding siklus normalnya. Secara umum kontraksi prematur dibagi menjadi kontraksi prematur yang berasal dari ventrikel (*Premature Ventricle Contraction*, PVC) dan kontraksi prematur yang berasal dari atrial (*Premature Atrial Contraction*, PAC). Meski keduanya dianggap bukan sesuatu yang berbahaya namun beberapa studi menunjukkan bahwa kontraksi prematur yang terjadi dengan frekuensi cukup tinggi diasosiasikan dengan meningkatnya resiko penyakit kardiovaskular baik untuk individu dengan dan tanpa penyakit kardiovaskular. Elektrokardiogram (EKG) menjadi instrumen utama dalam deteksi aritmia, dan penerapan machine learning berpotensi meningkatkan akurasi identifikasi pola kontraksi prematur. Penelitian ini difokuskan pada perbandingan performa dua algoritma machine learning yang berbeda karakteristik, yaitu *K-Nearest Neighbors* (KNN) sebagai algoritma berbasis instance dengan pendekatan sederhana namun efektif dalam mengenali pola lokal, serta *Light Gradient Boosting Machine* (LightGBM) sebagai algoritma berbasis gradient-boosted trees yang unggul dalam menangani data kompleks dengan efisiensi komputasi tinggi. Kedua algoritma diuji menggunakan kombinasi fitur temporal, statistik, spektral, dan interval RR dari sinyal EKG untuk mengklasifikasi tiga kelas: normal, PVC, dan PAC. Hasil evaluasi menunjukkan bahwa KNN mencapai akurasi keseluruhan 94,12% dengan kinerja relatif stabil pada PAC (F1-score 84,09%) namun menurun dibandingkan performanya pada PVC (F1-score 90,78%). Sebaliknya, LightGBM memperoleh akurasi keseluruhan lebih tinggi sebesar 94,80% dan unggul dalam mendeteksi PVC dengan recall 95,38%, meskipun pencapaiannya lebih rendah pada PAC (F1-score 81,65%). Perbedaan pola ini menegaskan bahwa meskipun KNN relatif stabil dalam klasifikasi PAC, LightGBM unggul dalam mendeteksi PVC dengan recall yang lebih tinggi. Temuan ini diharapkan dapat menjadi referensi dalam pemilihan algoritma klasifikasi yang sesuai untuk mendukung pengembangan metode yang lebih adaptif untuk implementasi klinis pada klasifikasi kontraksi jantung prematur.

Kata kunci: kontraksi jantung prematur; *machine learning*, KNN, LightGBM



Seminar Nasional LPPM UNSOED 2025

"Pengembangan Sumberdaya Perdesaan dan Kearifan Lokal Berkelanjutan"

8 Oktober 2025

Purwokerto

ABSTRACT

Premature contractions are a heart condition in which the heart beats earlier than its normal cycle. Premature contractions are generally divided into premature ventricular contractions (PVC) and premature atrial contractions (PAC). Although both are considered harmless, several studies have shown that premature contractions that occur with a high frequency are associated with an increased risk of cardiovascular disease in both individuals with and without cardiovascular disease. The electrocardiogram (ECG) is the primary instrument in arrhythmia detection, and the application of machine learning has the potential to improve the accuracy of premature contraction pattern identification. This study focused on comparing the performance of two machine learning algorithms with different characteristics: K-Nearest Neighbors (KNN), an instance-based algorithm with a simple yet effective approach in recognizing local patterns, and Light Gradient Boosting Machine (LightGBM), a gradient-boosted tree-based algorithm that excels in handling complex data with high computational efficiency. Both algorithms were tested using a combination of temporal, statistical, spectral, and RR interval features from the ECG signal to classify three classes: normal, PVC, and PAC. The evaluation results showed that KNN achieved an overall accuracy of 94.12% with relatively stable performance on PAC (F1-score: 84.09%), but its performance decreased compared to PVC (F1-score: 90.78%). In contrast, LightGBM achieved a higher overall accuracy of 94.80% and excelled in detecting PVC with a recall of 95.38%, although its achievement was lower on PAC (F1-score 81.65%). This difference in patterns confirms that although KNN is relatively stable in PAC classification, LightGBM excels in detecting PVC with a higher recall. These findings are expected to be a reference in selecting an appropriate classification algorithm to support the development of more adaptive methods for clinical implementation in premature cardiac contraction classification.

Keywords: premature cardiac contractions, machine learning, KNN, LightGBM

PENDAHULUAN

Kontraksi jantung prematur, seperti Premature Ventricular Contraction (PVC) dan Premature Atrial Contraction (PAC), adalah tipe kelainan jantung dimana detak jantung berdetak lebih dulu dibanding siklus normalnya. Pada umumnya kontraksi jantung prematur dianggap tidak terlalu berbahaya. Namun jika kontraksi jantung prematur sering terjadi (misalnya >1000/hari) maka seringkali dikaitkan dengan peningkatan risiko kejadian kardiovaskular yang merugikan termasuk fibrilasi atrium (FA), gagal jantung (HF), stroke, dan mortalitas pada individu dengan dan tanpa penyakit kardiovaskular (CVD) (Orini *et al.*, 2023). Elektrokardiogram (EKG) tetap menjadi standar diagnostik dalam mendeteksi aritmia termasuk PVC dan PAC karena sifatnya yang non-invasif dan kemampuan menangkap aktivitas listrik jantung dengan resolusi temporal tinggi. Namun, interpretasi EKG secara manual memerlukan keahlian tinggi, memakan waktu, dan bisa dipengaruhi oleh artefak sinyal, variasi anatomi individu, serta kondisi perekaman yang berbeda.

Seiring perkembangan teknologi data dan komputasi, metode machine learning (ML) telah diadopsi secara luas dalam analisis EKG untuk meningkatkan akurasi, kecepatan, serta otomatisasi diagnosis aritmia. Algoritma-algoritma seperti Support Vector Machine, Random



Seminar Nasional LPPM UNSOED 2025

"Pengembangan Sumberdaya Perdesaan dan Kearifan Lokal Berkelanjutan"

8 Oktober 2025

Purwokerto

Forest, dan Light Gradient Boosting Machine (LightGBM) telah banyak diuji di studi-terkait. Salah satu algoritma machine learning yang sederhana namun memberikan hasil yang cukup baik adalah K-Nearest Neighbours. KNN adalah algoritma berdasarkan instance (instance-based) yang mana tidak memerlukan asumsi bentuk fungsi target, efektif, cukup mudah dipahami, dan cukup baik dalam mengenali pola yang sangat lokal (misalnya, kemiripan antar sampel dalam ruang fitur). KNN juga serta relatif tahan terhadap outlier jika jarak atau metriknya dipilih dengan baik. Beberapa penelitian yang menggunakan KNN sebagai algoritma klasifikasi penyakit jantung diantaranya adalah Vera Artanti *et al.* (2024) yang memperoleh akurasi sebesar 91 %, dengan presisi sekitar 90 %, recall 93 %, dan F1-score 92 % untuk klasifikasi penyakit jantung (Artanti *et al.*, 2024). Selain itu, dalam penelitian Performance Evaluation of Machine Learning Techniques (MLT) for Heart Disease Prediction (2023), KNN menampilkan performa unggul dibanding beberapa model ML lainnya dalam prediksi penyakit jantung, dengan akurasi mendekati 99,04 % (Ansari *et al.*, 2023). Pada penelitian yang dilakukan Adi & Winarti (2022) terbukti bahwa untuk memprediksi penyakit gagal jantung KNN mampu mencapai akurasi yang cukup baik dengan running time yang singkat yakni dibawah 1 detik. Bahkan penelitian yang dilakukan Aprilliandhika dan Abdulloh (2024) juga menunjukkan bahwa KNN, setelah dioptimasi menggunakan *grid search*, berhasil mengungguli Support Vector Machine (SVM) dalam prediksi stroke, dengan nilai recall sebesar 98 %. Validitas klinis KNN juga tercermin dalam penelitian lain yang berfokus pada klasifikasi penyakit jantung, terutama regurgitasi katup mitral, di mana KNN menunjukkan performa yang kuat setelah penyetelan hiperparameter intensif, memperlihatkan bahwa algoritma ini mampu menyesuaikan diri dengan variasi data dan menghasilkan deteksi yang sensitif terhadap kondisi penyakit (Sudha *et al.*, 2025).

Selain KNN, LightGBM (Light Gradient Boosting Machine) adalah algoritma machine learning yang juga digunakan untuk klasifikasi penyakit jantung. LightGBM adalah algoritma boosting berbasis pada gradient-boosted decision trees. Ia menggunakan pendekatan leaf-wise tree growth dan optimasi histogram untuk mempercepat pelatihan, serta mendukung fitur seperti penanganan missing value, regularisasi, dan teknik penyeimbang kelas. Kelebihan LightGBM meliputi efisiensi komputasi, kemampuan menangani dataset besar, dan kemampuan menangkap interaksi antar fitur serta non-linearitas. Beberapa penelitian yang menggunakan LightGBM untuk klasifikasi diantaranya adalah studi Heart Rate Classification in ECG Signals Using Machine Learning and Deep Learning, LightGBM mencapai 99% akurasi dan F1 score 94% untuk klasifikasi heartbeat menggunakan fitur-handcrafted (temporal/statistik/interval) (Vo, 2025). Begitu juga penelitian Machine Learning-Based Approach for ECG Analysis and Cardiac Anomaly Detection menemukan bahwa model boosting seperti LightGBM lebih unggul dalam performa keseluruhan, sedangkan KNN membantu dalam kestabilan terhadap variasi data yang belum pernah dilihat sebelumnya (Monali Gulhane, 2024). Dalam artikel Analysis of Arrhythmia Features based on LightGBM and Kmeans Feature Extraction (Li *et al.*, 2023), LightGBM digunakan untuk mengklasifikasi risiko aritmia berdasarkan fitur-fitur EKG seperti gelombang P, interval P-R, kompleks QRS, ST dan gelombang T. Model mencapai akurasi 93,5 % dan hasilnya stabil di atas 90 % setelah validasi. Artikel Predictive Analytics in Heart Disease: Leveraging LightGBM for Improved Diagnostic Accuracy (Han, Meng, & Zhu, 2024) juga menggunakan LightGBM pada dataset penyakit jantung (14 indikator kesehatan) dan memperoleh akurasi 98,54 % dan AUC 98,5%, lebih baik dibanding SVM dan Logistic Regression (Han *et al.*, 2024).



Seminar Nasional LPPM UNSOED 2025

"Pengembangan Sumberdaya Perdesaan dan Kearifan Lokal Berkelanjutan"

8 Oktober 2025

Purwokerto

Dari literatur diatas KNN dan LightGBM menunjukkan performa yang kompetitif dalam klasifikasi penyakit jantung. LightGBM unggul dalam menangani dataset besar dan fitur kompleks, sedangkan KNN menonjol dalam mendeteksi pola lokal dan kelas minor. Oleh karena itu, pada penelitian dilakukan perbandingan kedua algoritma tersebut dengan menerapkan konsep inter-patient untuk melihat algoritma mana yang mampu memberikan performansi paling baik dalam mengklasifikasikan kontraksi jantung prematur terutama kelas PVC dan PAC. Penelitian ini diharapkan dapat memberi kontribusi terhadap pengembangan metode klasifikasi kontraksi jantung prematur yang lebih adaptif dan aplikatif dalam praktik medis atau sistem pemantauan pasien.

METODE

Data

Pada penelitian ini digunakan data sinyal EKG 1D MIT-BIH Arrhythmia Database yang didiskritisasi dengan frekuensi sampling 360 Hz. Dari total 48 rekaman EKG penelitian ini hanya menggunakan kanal MLII karena arah konduksi impuls jantung searah dengan rekaman pada kanal tersebut sehingga kompleks QRS terlihat paling jelas (Shi *et al.*, 2019). Temuan dari studi wearable ECG mendukung pemilihan ini dengan menunjukkan bahwa deteksi premature atrial contraction (PAC) dan premature ventricular contraction (PVC) dapat dilakukan dengan format lead tunggal, misalnya rekaman 15 detik pada perangkat wearable, yang memberikan sinyal memadai untuk mendeteksi premature beats (Orini *et al.*, 2023). Rekaman yang digunakan dalam penelitian ini dipilih secara selektif dengan kriteria pemisahan data latih dan data uji menggunakan pendekatan inter-patient, sehingga tidak ada rekaman yang digunakan secara bersamaan untuk pelatihan dan pengujian, dan proses pembagiannya mengikuti prosedur yang dijelaskan pada penelitian Shi *et al.* (2019).

Preprocessing

Sinyal EKG yang diperoleh dari database dibersihkan menggunakan median filter dengan lebar 200 ms dan 600 ms untuk menghilangkan baseline wander (Zakaria *et al.*, 2024). Sedangkan noise high frekuensi power line dihilangkan menggunakan low pass filter (LPF) butterworth dengan frekuensi cut-off 40 Hz (Zakaria *et al.*, 2024). Selanjutnya dilakukan segmentasi sinyal PQRST dengan mengambil 130 sampel sebelum dan 170 sampel sesudah posisi R (Jiang *et al.*, 2019).

Ekstraksi Fitur

Penelitian ini menggunakan fitur dari beberapa domain, mencakup temporal, spektral, statistik, dan morfologi interval R-R. Fitur temporal, spektral, dan statistik diekstraksi dengan Time Series Feature Extraction Library (TSFEL) berbasis Python, yang menyediakan berbagai parameter analisis sinyal waktu (Barandas *et al.*, 2020; Tsai & Morshed, 2025), dengan rincian ditunjukkan pada Tabel 1. Dari domain morfologi R-R, digunakan empat fitur utama, yaitu Posisi R, Pre_R, Post_R, dan Local_R (Srinivasulu & Sriraam, 2023; Zakaria *et al.*, 2024).

Seleksi Fitur



Seminar Nasional LPPM UNSOED 2025

"Pengembangan Sumberdaya Perdesaan dan Kearifan Lokal Berkelanjutan"

8 Oktober 2025

Purwokerto

Tahapan seleksi meliputi penskalaan fitur menggunakan z-score standardization untuk menormalkan nilai, penghapusan fitur dengan variansi rendah menggunakan *VarianceThreshold* dengan ambang batas 0, serta eliminasi fitur dengan korelasi tinggi menggunakan korelasi Pearson ($r > 0,7$) agar menghindari redundansi (Mukaka, 2012). Fitur yang tersisa kemudian dipilih menggunakan Recursive Feature Elimination with Cross-Validation (RFECV) dengan *Random Forest* sebagai estimator dan validasi silang 5-fold untuk memperoleh subset fitur yang paling relevan.

Klasifikasi

Untuk mengatasi ketidakseimbangan jumlah data antar kelas pada data latih, penelitian ini menerapkan Synthetic Minority Over-sampling Technique (SMOTE), yang menghasilkan sampel sintesis pada kelas minoritas melalui interpolasi antar tetangga pada ruang fitur, sehingga distribusi kelas menjadi lebih seimbang. Pendekatan ini memungkinkan model belajar pola dari kelas minoritas dengan lebih baik dan meningkatkan kinerja klasifikasi, khususnya pada metrik seperti recall dan F1-score. Setelah data diseimbangkan, dilakukan pelatihan dan pengujian menggunakan algoritma machine learning yang telah ditentukan.

K-Nearest Neighbors (KNN) merupakan salah satu algoritma *supervised learning* yang sederhana namun efektif dalam mengenali pola lokal. Sebagai algoritma berbasis *instance*, KNN tidak membangun model eksplisit, melainkan melakukan klasifikasi dengan membandingkan data uji terhadap sampel pelatihan terdekat. Algoritma ini bersifat non-parametrik, sehingga tidak memerlukan asumsi distribusi data. Prinsip kerjanya adalah menghitung jarak antara data uji dengan seluruh data latih, kemudian memilih K tetangga terdekat untuk menentukan kelas berdasarkan mayoritas label di sekitarnya. Dua parameter penting pada KNN adalah jumlah tetangga ($n_neighbors$) dan bobot (*weights*) dari setiap tetangga yang digunakan dalam voting. Dalam penelitian ini, dilakukan *hyperparameter tuning* dengan variasi nilai $n_neighbors$ antara 4 hingga 10, serta pengaturan bobot untuk meningkatkan kinerja klasifikasi.

Tabel 1. Fitur TSFEL

Domain	Fitur
Temporal	<i>absolute energy, area under the curve, autocorrelation, centroid, entropy, mean absolute difference, mean difference, median absolute difference, median difference, negative turning points, neighbourhood peaks, peak to peak distance, positive turning points, signal distance, slope, sum absolute difference, total energy, dan zero crossing rate</i>
Statistik	<i>empirical cumulative distribution function (ECDF), interquartile range, kurtosis, max, min, median, mean, mean absolute deviation, median absolute deviation, root mean square, skewness, standard deviation, dan variance</i>
Spektral	<i>FFT mean coefficient, fundamental frequency, human range energy, LPCC, MFCC, wavelet entropy, wavelet energy, wavelet standard deviation, wavelet variance, wavelet absolute mean, maximum power spectrum, maximum frequency, median frequency, power bandwidth, spectral centroid, spectral decrease, spectral distance, spectral entropy, spectral kurtosis, spectral positive turning points, spectral roll-off, spectral roll-on, spectral skewness, spectral slope, spectral spread, dan spectral variation</i>



Seminar Nasional LPPM UNSOED 2025

"Pengembangan Sumberdaya Perdesaan dan Kearifan Lokal Berkelanjutan"

8 Oktober 2025

Purwokerto

LightGBM adalah algoritma gradient-boosted trees yang dirancang untuk menangani data kompleks dengan efisiensi komputasi tinggi. Seperti decision tree, model ini membangun pohon keputusan secara bertahap, namun setiap pohon baru difokuskan untuk mengoreksi kesalahan pohon sebelumnya melalui minimisasi loss function dengan gradient descent. Keunggulan utama LightGBM terletak pada strategi leaf-wise growth, yaitu memperluas daun yang memberi potensi penurunan loss terbesar. Pendekatan ini memungkinkan peningkatan akurasi dengan iterasi lebih sedikit, khususnya pada dataset berskala besar. Prediksi akhir diperoleh dengan menjumlahkan kontribusi seluruh pohon yang terbentuk. Untuk meningkatkan kinerja, dilakukan hyperparameter tuning pada jumlah estimator (100, 200), kedalaman maksimum (3, 5, 7), learning rate (0.001, 0.01, 0.1), jumlah daun maksimum (31, 63), dan subsample (0.6, 0.8, 1.0) menggunakan GridSearchCV dengan stratified 5-fold cross-validation.

Untuk menganalisa hasil klasifikasi dari berbagai algoritma digunakan beberapa parameter yakni presisi, recall, dan F1-score. Ketiganya merupakan parameter yang dianggap mampu merepresentasikan performansi algoritma klasifikasi untuk data yang tidak seimbang. Presisi, recall, dan F1-score dapat diperoleh dari confusion matrix yakni True Positive (TP), False Positive (FP), True Negative (TN), dan False Negative (FN).

$$\text{Presisi} = \frac{TP}{TP+FP} \times 100 \quad (1)$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100 \quad (2)$$

$$\text{F1 score} = \frac{2 \times (\text{Presisi} \times \text{Recall})}{\text{Presisi} + \text{Recall}} \times 100 \quad (3)$$

HASIL DAN PEMBAHASAN

Hasil klasifikasi K-Nearest Neighbors

Parameter K-NN yang dipilih setelah proses hypertuning adalah $k = 4$ dengan skema pemberian bobot berdasarkan jarak (*distance weighting*). Pemilihan ini didasarkan pada trade-off optimal antara sensitivitas lokal dan kestabilan prediksi: dengan $k = 4$, model mempertimbangkan cukup tetangga (data sekitar) untuk mengurangi noise, sementara bobot jarak memastikan tetangga terdekat memiliki pengaruh lebih besar, yang membantu dalam situasi di mana data berada di perbatasan antar kelas. Skema ini juga mengurangi kemungkinan tied vote daripada bobot uniform. Pengaturan ini diharapkan untuk membentuk model yang sensitif terhadap struktur lokal data namun tidak terlalu rentan terhadap outlier. Evaluasi silang (misalnya stratified k-fold) digunakan dalam proses tuning untuk memastikan bahwa parameter yang terpilih memberikan performa yang konsisten antara subset data, tidak hanya pada data training.

Gambar 1a menunjukkan confusion matrix dari KNN. Dari hasil klasifikasi diperoleh bahwa model berhasil mengklasifikasikan sebagian besar sampel kelas Normal (N) dengan benar (2180 kasus benar), dan sebagian kecil diklasifikasikan sebagai PVC atau PAC. Untuk kelas



Seminar Nasional LPPM UNSOED 2025

"Pengembangan Sumberdaya Perdesaan dan Kearifan Lokal Berkelanjutan"

8 Oktober 2025

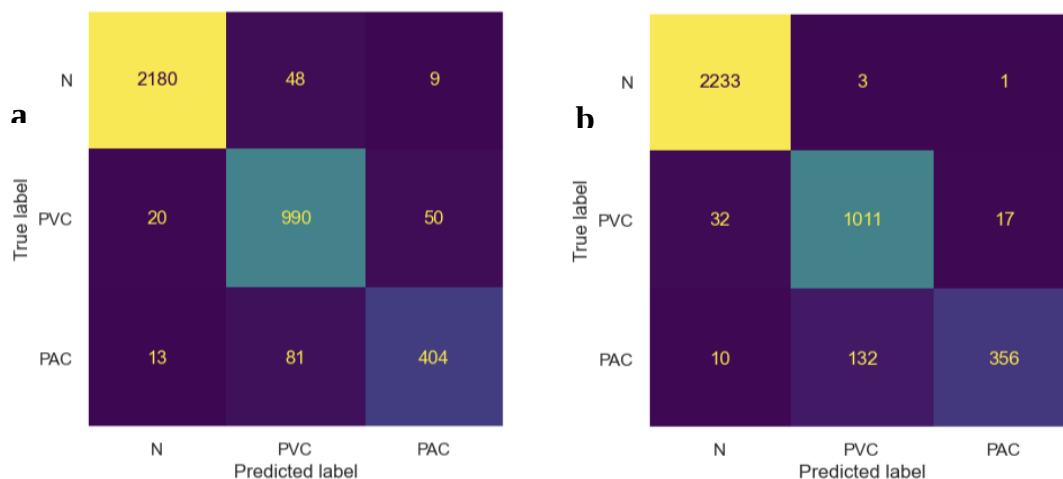
Purwokerto

PVC juga performanya tinggi dimana sebagian besar sampel PVC (990 dari 1060) diklasifikasikan dengan benar, meskipun ada sejumlah error terutama dari kelas PAC yang diklasifikasikan sebagai PVC. Sementara itu, kelas PAC memperlihatkan performa relatif paling rendah pada recall: meskipun presisi PAC tinggi, dimana 404 sampel diklasifikasikan benar sebagai PAC, tetapi ada banyak sampel PAC yang gagal terdeteksi sebagai PVC atau kelas lain. Akibatnya, meskipun akurasi keseluruhan cukup tinggi (94,2 %), model kurang sensitif dalam mendeteksi kelas PAC. Kesalahan jenis ini menunjukkan bahwa fitur antar kelas PAC dan PVC memiliki overlap, dan bahwa $k = 4$ - distance weighting menghasilkan bias ke kelas mayoritas ketika tetangga terdekat lebih banyak dari kelas lain, sehingga recall untuk kelas minoritas perlu diperbaiki.

Hasil klasifikasi LightGBM

Hasil dari *hypertuning* parameter algoritma LightGBM dilakukan untuk menemukan kombinasi yang optimal, dengan mempertimbangkan keseimbangan antara kompleksitas model dan generalisasi. Parameter yang dipilih adalah $n_estimators = 200$ (banyaknya pohon boosting), $learning_rate = 0.1$ (tingkat kontribusi tiap pohon relatif moderat), $max_depth = 5$ (membatasi kedalaman pohon agar tidak terlalu dalam sehingga mengurangi risiko overfitting), $num_leaves = 31$ (jumlah daun tiap pohon yang memungkinkan fleksibilitas menangkap variasi data tanpa membuat pohon terlalu kompleks), serta $subsample = 0.8$ (menggunakan 80 % barisan data tiap iterasi untuk memperkenalkan variabilitas/randomisasi agar model lebih robust). Penetapan ini juga memperhitungkan dokumentasi dan praktik-baik LightGBM, yang menyebutkan bahwa pembatasan num_leaves relatif terhadap max_depth dan penggunaan $subsample$ adalah strategi yang efektif untuk mencegah overfitting sambil menjaga daya prediktif yang tinggi.

Berdasarkan *confussion matrix* pada Gambar 1b, model menunjukkan kinerja sangat baik pada kelas Normal (N) dan PVC: untuk kelas Normal, hampir seluruh sampel diklasifikasikan dengan benar, hanya 3 sampel diklasifikasikan sebagai PVC dan 1 sebagai PAC; untuk kelas PVC, 1011 sampel tepat diklasifikasikan, dengan jumlah kesalahan ke kelas lain (N dan PAC) relatif kecil (32 ke N, 17 ke PAC). Kelas PAC juga memiliki hasil yang baik dalam presisi, namun recall-nya lebih rendah: terdapat 10 sampel PAC yang salah diklasifikasikan sebagai





Seminar Nasional LPPM UNSOED 2025

"Pengembangan Sumberdaya Perdesaan dan Kearifan Lokal Berkelanjutan"

8 Oktober 2025

Purwokerto

N dan 132 sebagai PVC, dibanding 356 sampel PAC yang berhasil diklasifikasikan dengan benar. Implikasi dari hasil ini menunjukkan bahwa model sudah sangat kuat untuk kelas mayoritas, sedangkan untuk kelas minoritas seperti PAC LightGBM belum bisa mengklasifikasikannya dengan baik.

Pembahasan

Tabel 2 menunjukkan perbandingan klasifikasi antara 2 algoritma. Perbandingan metrik antara kelas PVC dan PAC memperlihatkan bahwa LightGBM unggul dalam presisi saat memprediksi PAC (95,19 %) dan memiliki recall sangat tinggi untuk PVC (95,38 %), sementara KNN memiliki recall PAC yang lebih baik (81,12 %) dibandingkan LightGBM, tetapi dengan presisi PAC yang lebih rendah (87,21 %). Untuk PVC, LightGBM dan KNN keduanya menunjukkan performa yang cukup kuat, tapi LightGBM masih lebih baik dalam menjaga keseimbangan antara recall dan precision untuk PVC yang mana menghasilkan F1-score PVC di LightGBM sekitar 91,67 %, dibandingkan dengan F1-score PVC KNN yang sedikit lebih rendah. Ini menunjukkan bahwa LightGBM lebih konservatif dalam memprediksi kelas PAC sehingga ketika membuat prediksi PAC, kemungkinan besar prediksi tersebut benar (presisi tinggi), tetapi ada banyak sampel PAC yang gagal terdeteksi (recall rendah). Sebaliknya, KNN lebih agresif mendeteksi PAC yang mengakibatkan recall lebih baik untuk PAC, tetapi toleransi terhadap kesalahan prediksi PAC juga lebih besar (presisi sedikit turun).

KNN sebagai algoritma instance-based menawarkan keunggulan dalam mengenali pola lokal di ruang fitur. Karena keputusan dibuat berdasarkan tetangga terdekat (dengan bobot jarak), KNN sangat responsif terhadap variasi lokal fitur PAC dan PVC, yang mana jika sampel PAC memiliki tetangga PAC yang cukup dekat, KNN cenderung memprediksinya dengan benar sebagai PAC. Hal ini menjelaskan kenapa recall PAC KNN lebih tinggi: lebih sedikit kasus PAC asli yang "terlewat" menjadi kelas lain. Namun karakteristik ini juga membawa risiko ketika terjadi overlap fitur antara PVC dan PAC yang mana membuat beberapa sampel PAC menjadi ambigu, yang menyebabkan false positives PAC atau kesalahan prediksi antar kelas. KNN juga cenderung sangat dipengaruhi oleh pemilihan jumlah tetangga, skala fitur, dan distribusi tetangga. Jika banyak tetangga PAC kurang representatif, PAC bisa salah diklasifikasikan.

LightGBM, di sisi lain, menggunakan pendekatan gradient boosted trees yang membangun model secara global dengan memperbaiki kesalahan iterasi demi iterasi. Algoritma ini unggul dalam menangani interaksi antar fitur kompleks dan pola global, serta efisiensinya lewat

Tabel 2. Hasil klasifikasi berbagai algoritma

Algoritma	Kelas	Presisi	Recall	F1-Score	Akurasi
KNN	N	98,51%	97,45%	97,98%	94,12%
	PVC	88,40%	93,30%	90,78%	
	PAC	87,21%	81,12%	84,09%	
LightGBM	N	98,15%	99,82%	98,98%	94,80%
	PVC	88,23%	95,38%	91,67%	
	PAC	95,19%	71,48%	81,65%	



Seminar Nasional LPPM UNSOED 2025

"Pengembangan Sumberdaya Perdesaan dan Kearifan Lokal Berkelanjutan"

8 Oktober 2025

Purwokerto

teknik-teknik seperti leaf-wise growth dan histogram binning. Inilah yang memungkinkan LightGBM menjaga performa tinggi untuk PVC (both recall dan precision) dan memberikan presisi PAC yang sangat tinggi. Ketika LightGBM mengatakan bahwa suatu data atau sampel dikategorikan sebagai kelas PAC, besar kemungkinan klasifikasi tersebut benar, menunjukkan false positives PAC relatif rendah. Namun karena optimisasi keseluruhan dan kemampuan model untuk kelas mayoritas (yaitu PVC dan Normal) lebih mudah dikenali, banyak PAC yang berada di sekitar batas keputusan masih gagal terdeteksi sebagai PAC sehingga mengakibatkan recall PAC yang lebih rendah. Efisiensi komputasi LightGBM juga memungkinkan pelatihan dan inferensi yang relatif cepat dibanding KNN ketika data bertambah besar.

Berdasarkan perbandingan ini, jika prioritas klinis adalah mendeteksi sebanyak mungkin PAC (yaitu recall kelas PAC sangat penting), maka KNN saat ini lebih sesuai karena kemampuan deteksi PAC lebih baik meskipun ada beberapa false positives. Namun, jika prioritas adalah menjaga prediksi PAC yang dilakukan sangat akurat (presisi tinggi), dan meminimalkan kesalahan prediksi, maka LightGBM lebih cocok karena presisi PAC-nya sangat tinggi dan performa PVC dan Normal juga sangat baik. Tetapi untuk aplikasi umum di mana false negative baik PAC maupun PVC dapat berdampak besar (misalnya risiko medis jika PAC tidak terdeteksi) sehingga nilai false negatif harus dibuat seminimum mungkin untuk kelas PAC dan PVC maka KNN lebih cocok digunakan.

KESIMPULAN

Penelitian ini berhasil membandingkan kinerja algoritma K-Nearest Neighbors (KNN) dan Light Gradient Boosting Machine (LightGBM) dalam klasifikasi detak jantung ke tiga kelas: Normal, Premature Ventricular Contraction (PVC), dan Premature Atrial Contraction (PAC). LightGBM menunjukkan presisi tinggi untuk PAC (95,19 %) dan recall yang sangat baik untuk PVC (95,38 %), meski recall untuk PAC masih relatif rendah (71,48 %), menunjukkan bahwa banyak sample PAC yang tidak berhasil terdeteksi. Di sisi lain, KNN lebih baik dalam menangkap sampel PAC asli dengan recall PAC sekitar 81,12 %, walaupun presisinya lebih rendah dibanding LightGBM. Perbandingan ini menunjukkan bahwa LightGBM unggul jika prioritasnya adalah menghasilkan prediksi PAC yang akurat dan menghindari false positives, sedangkan KNN cocok ketika tujuan utamanya adalah memaksimalkan deteksi PAC, meski dengan beberapa kompromi terhadap keakuratan prediksi. Berdasarkan hasil ini, jika aplikasi klinis memerlukan sensitivitas/recall tinggi terhadap PAC (misalnya untuk deteksi dini), maka K-Nearest Neighbors (KNN) adalah algoritma yang direkomendasikan; tetapi jika aplikasi lebih menekankan keandalan prediksi PAC dan meminimalkan kesalahan prediksi (false positives), maka Light Gradient Boosting Machine (LightGBM) menjadi pilihan yang lebih baik. Temuan ini diharapkan dapat menjadi referensi dalam pemilihan algoritma klasifikasi yang sesuai untuk mendukung pengembangan metode yang lebih adaptif untuk implementasi klinis pada klasifikasi kontraksi jantung prematur. Untuk kedepannya mungkin penggunaan algoritma seperti hybrid atau ensemble dengan menggunakan LightGBM sebagai basis dan menggunakan metode lokal (seperti KNN) sebagai penguat dapat mengoptimalkan recall PAC tanpa kehilangan terlalu banyak presisi serta meningkatkan performansi sistem secara keseluruhan.



Seminar Nasional LPPM UNSOED 2025

"Pengembangan Sumberdaya Perdesaan dan Kearifan Lokal Berkelanjutan"

8 Oktober 2025

Purwokerto

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada LPPM UNSOED atas pembiayaan penelitian ini melalui Hibah Penelitian BLU UNSOED Skema Riset Peningkatan Kompetensi Tahun 2025 dengan Nomor Kontrak 14.453/UN23.34/PT.01/V/2025.

DAFTAR PUSTAKA

- Adi, S., & Wintarti, A. 2022. KOMPARASI METODE SUPPORT VECTOR MACHINE (SVM), K-NEAREST NEIGHBORS (KNN), DAN RANDOM FOREST (RF) UNTUK PREDIKSI PENYAKIT GAGAL JANTUNG. *MATHunesa: Jurnal Ilmiah Matematika*, 10(2), 258–268. <https://doi.org/10.26740/mathunesa.v10n2.p258-268>.
- Ansari, G. A., Bhat, S. S., Ansari, M. D., Ahmad, S., Nazeer, J., & Eljialy, A. E. M. 2023. Performance Evaluation of Machine Learning Techniques (MLT) for Heart Disease Prediction. *Computational and Mathematical Methods in Medicine*, 2023(1), 8191261. <https://doi.org/10.1155/2023/8191261>.
- Aprilliandhika, W., & Abdulloh, F. F. 2024. COMPARISON OF K-NEAREST NEIGHBOR AND SUPPORT VECTOR MACHINE ALGORITHM OPTIMIZATION WITH GRID SEARCH CV ON STROKE PREDICTION. *Jurnal Teknik Informatika (Jutif)*, 5(4), 991–1000. <https://doi.org/10.52436/1.jutif.2024.5.4.1951>.
- Artanti, V., Faisal, M., & Kurniawan, F. 2024. Klasifikasi Cardiovascular Diseases Menggunakan Algoritma K-Nearest Neighbors (KNN). *Techno.Com*, 23(2), 469–481. <https://doi.org/10.62411/tc.v23i2.10061>.
- Barandas, M., Folgado, D., Fernandes, L., Santos, S., Abreu, M., Bota, P., Liu, H., Schultz, T., & Gamboa, H. 2020. TSFEL: Time Series Feature Extraction Library. *SoftwareX*, 11, 100456. <https://doi.org/10.1016/j.softx.2020.100456>.
- Han, R., Meng, R., & Zhu, Q. 2024. *Predictive Analytics in Heart Disease: Leveraging LightGBM for Improved Diagnostic Accuracy*.
- Jiang, J., Zhang, H., Pi, D., & Dai, C. 2019. A novel multi-module neural network system for imbalanced heartbeats classification. *Expert Systems with Applications: X*, 1, 100003. <https://doi.org/10.1016/j.eswax.2019.100003>.
- Monali Gulhane. 2024. Machine Learning-Based Approach for ECG Analysis and Cardiac Anomaly Detection to Enhance Early Diagnosis and Treatment. *Advances in Nonlinear Variational Inequalities*, 28(2s), 555–572. <https://doi.org/10.52783/anvi.v28.2762>.
- Mukaka, M. M. 2012. Statistics Corner: A guide to appropriate use of Correlation coefficient in medical research. *Malawi Medical Journal : The Journal of Medical Association of Malawi*, 24(3), 369–371.
- Orini, M., van Duijvenboden, S., Young, W. J., Ramírez, J., Jones, A. R., Tinker, A., Munroe, P. B., & Lambiase, P. D. 2023. Premature atrial and ventricular contractions detected on wearable-format electrocardiograms and prediction of cardiovascular events.



Seminar Nasional LPPM UNSOED 2025

"Pengembangan Sumberdaya Perdesaan dan Kearifan Lokal Berkelanjutan"

8 Oktober 2025

Purwokerto

- European Heart Journal - Digital Health, ztad007.
<https://doi.org/10.1093/ehjdh/ztad007>.
- Shi, H., Wang, H., Zhang, F., Huang, Y., Zhao, L., & Liu, C. 2019. Inter-patient heartbeat classification based on region feature extraction and ensemble classifier. *Biomedical Signal Processing and Control*, 51, 97–105. <https://doi.org/10.1016/j.bspc.2019.02.012>
- Srinivasulu, A., & Sriraam, N. 2023. Signal Processing Framework for the Detection of Ventricular Ectopic Beat Episodes. *Journal of Medical Signals & Sensors*, 13(3), 239–251. https://doi.org/10.4103/jmss.jmss_12_22.
- Sudha, P., Aravinth, T. U., & Selvi, R. T. 2025. Prediction and Treatment of Mitral Valve Failure by Hyperparameter Tuning Using ML Algorithms with an Ayurvedic Approach. Dalam A. Swaroop, B. Virdee, S. D. Correia, & Z. Polkowski (Ed.), *Proceedings of Data Analytics and Management* (hlm. 455–471). Springer Nature Singapore.
- Tsai, I. H., & Morshed, B. I. 2025. Signal-Specific and Signal-Independent Features for Real-Time Beat-by-Beat ECG Classification with AI for Cardiac Abnormality Detection. *Electronics*, 14(13), 2509. <https://doi.org/10.3390/electronics14132509>.
- Vo, T. N. 2025. *Heart Rate Classification in ECG Signals Using Machine Learning and Deep Learning* (No. arXiv:2506.06349). arXiv. <https://doi.org/10.48550/arXiv.2506.06349>.
- 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*, 12, 147372–147386. <https://doi.org/10.1109/ACCESS.2024.3469640>.