

# *Comparative Evaluation of Support Vector Machine and Random Forest Algorithms for Heart Abnormality Detection Based on ECG Signals*

Evaluasi Komparatif Algoritma *Support Vector Machines* (SVM) Dan *Random Forest* Untuk Deteksi Kelainan Jantung Berdasarkan Sinyal EKG

Rizky Putri Intan Hafsa<sup>1\*</sup>, Basitha Febrinda Hidayatulail<sup>2</sup>, Rahman Arifuddin<sup>3</sup>

<sup>1,2,3</sup>Program Studi Teknik, Universitas Merdeka Malang, Indonesia

<sup>1</sup>intanrizky520@gmail.com

<sup>2</sup>basitha@unmer.ac.id

<sup>3</sup>rahman.arifuddin@unmer.ac.id

**Abstract** \_Electrocardiogram signals contain important information for identifying heart rhythm abnormalities and can be used to analyse heart activity and detect heart rhythm abnormalities. The aim of this study is to compare the Support Vector Machines and Random Forest methods for classifying signals in normal, apnoea and arrhythmia ECG recordings, which were processed through signal pre-processing, R-Peak detection, RR interval calculation, filter extraction, data partitioning for model training and performance evaluation. The apnoea data was obtained from the Apnoea-ECG Physionet database and the arrhythmia data from the MIT-BIH Arrhythmia database, both of which are available on Kaggle. The dataset was divided into two parts: 80% for training and 20% for testing to evaluate the performance of the model. Accuracy, precision, recall, F1-score and confusion matrix were used to classify the two parts. For apnoea classification, Random Forest showed an accuracy of 97.24%, while Support Vector Machine showed an accuracy of 91.41%. For arrhythmia classification, Random Forest showed an accuracy of 97.67%, while SVM showed an accuracy of 95.96%. These results show that Random Forest is superior and produces more effective classification performance than SVM in both tests. Therefore, Random Forest is considered more effective in classifying normal, apnoea, bradycardia and tachycardia conditions, and in classifying EKG signal patterns.

**Keywords:** Electrocardiogram; Apnoea; Arrhythmia; Support Vector Machines; Random Forest; Classification

**Abstrak** \_Sinyal Elektrokardiogram (EKG) memuat informasi penting untuk mengidentifikasi kelainan irama jantung dan dapat digunakan untuk menganalisis aktivitas listrik sinyal jantung dan menemukan kelainan irama pada sinyal jantung. Penelitian ini bertujuan untuk membandingkan metode *Support Vector Machines* (SVM) dan *Random Forest* untuk mengklasifikasikan sinyal pada kondisi normal, *apnea*, dan *arrhythmia*. Penelitian ini terdiri dari 20 rekaman EKG *apnea* dan 20 rekaman EKG *arrhythmia* yang diproses melalui rekaman tahap preprosesing sinyal, deteksi *R-Peak*, perhitungan *interval RR*, ekstraksi filter, pembagian data pelatihan model dan evaluasi performa. Data *apnea* diperoleh dari database *Apnea-ECG Physionet* dan data *Arrhythmia* diperoleh dari *MIT-BIH Arrhythmia database* yang tersedia di *Physionet*. Dataset dibagi menjadi dua bagian 80% data latihan dan 20% data uji untuk evaluasi performa model. *Accuracy*, *precision recall*, *F1-score* dan *confusion matrix* digunakan untuk membagi kedua bagian ini untuk klasifikasi kondisi, pada kondisi ini *apnea Random Forest* menunjukkan keakuratan 97.24% dan keakuratan *Support Vector Machine* 91.41% sementara klasifikasi *arrhythmia Random Forest* menunjukkan accurasi 97,67% sedangkan *Support Vector Machine* memperoleh accurasi 95.96% dari hasil tersebut menunjukkan bahwa *Random Forest* lebih unggul dan menghasilkan performa klasifikasi yang lebih efektif dibandingkan *Support Vector Machine* pada kedua pengujian. Dengan demikian *Random Forest* dinilai lebih efektif untuk klasifikasi kondisi normal, *apnea*, *bradycardia*, *tachycardia* dan normal dalam mengklasifikasikan pola sinyal EKG.

**Kata Kunci:** Elektrokardiogram; *Apnea*; *Arrhythmia*; *Support Vector Machines*; *Random Forest*; Klasifikasi

## I. INTRODUCTION

Elektrokardiogram (EKG) adalah salah satu jenis pemeriksaan non-invasif yang digunakan untuk merekam aktivitas permukaan jantung.[1] Sinyal EKG menunjukkan banyak hal tentang fungsi pada jantung terutama tentang irama jantung perubahan pola gelombang yang berbeda dan interval antara denyut jantung. Sinyal EKG biasanya terdiri dari beberapa komponen utama seperti gelombang P, kompleks QRS dan gelombang T yang menunjukkan aktivitas listrik jantung selama siklus jantung,[2] dengan menganalisis perubahan yang terjadi pada karakteristik sinyal EKG informasi ini dapat membantu dalam menentukan apakah sinyal tersebut normal atau abnormal. Untuk memahami sinyal EKG secara umumnya diperlukan keahlian khusus ketelitian tinggi dan pengalaman klinis selain itu, prosesnya dapat memakan waktu yang cukup lama terutama ketika data EKG yang digunakan berlangsung lama atau melibatkan jumlah sampel yang besar hasil interpretasi dapat dipengaruhi oleh perbedaan pengalaman pengamat akibatnya diperlukan metode komputasi yang dapat mengidentifikasi pola sinyal ECG. Dalam analisis sinyal biomedis *machine learning* adalah pendekatan yang banyak digunakan karena memiliki kemampuan untuk mempelajari pola dari data berdasarkan karakteristik,[3] yang diekstraksi dari sinyal EKG, fitur-fitur ini termasuk interval denyut jantung (RR), *rate* jantung (HR), dan karakteristik statistik sinyal, fitur ini dapat menggambarkan keteraturan denyut jantung, variasi antar denyut jantung dan perubahan pola sinyal, jika ada *tachycardia* denyut jantung lebih tinggi dari batas normal yang ditunjukkan dengan denyut jantung lebih dari 100 denyut per menit dan interval RR yang lebih pendek, sebaliknya *bradycardia* biasanya didefinisikan sebagai denyut jantung yang lebih rendah dari 60 denyut per menit,[4] namun gangguan pernapasan saat tidur dapat mempengaruhi denyut jantung dan sinyal EKG pada orang yang mengalami *apnea* gangguan pernafasan saat tidur dapat mempengaruhi variasi denyut jantung dan pola sinyal EKG.

Data *arrythmia* diperoleh dari MIT-BIH *Arrhythmia* dalam format CSV sederhana yang tersedia untuk database ini berisi 48 rekaman EKG *ambulatory Physionet*,[5] *Apnea-ECG Database* terdiri dari 70 rekaman EKG yang dibagi menjadi 35 set pembelajaran dan data tes dan durasi rekaman berkisar antara 7 jam dan hampir 10 jam, data *apnea* diperoleh dari dataset public yang bersumber dari *Physionet*[6]

*Support Vector Machines (SVM)* dan *Random Forest* adalah dua metode klasifikasi yang digunakan dalam penelitian ini. SVM merupakan pembelajaran yang diawasi yang bekerja dengan membentuk *hyperplane*

ideal untuk memisahkan data ke dalam kelas tertentu[7] sebaliknya *Random Forest* adalah metode pembelajaran kelompok yang terdiri dari banyak pohon keputusan dengan menghasilkan klasifikasi dari *Random Forest* diperoleh melalui mekanisme voting dari beberapa pohon keputusan[8] sehingga metode ini relative stabil dalam mengklasifikasikan variasi data dan hubungan *non linear* antar fitur.

Perbandingan antara SVM dan *Random Forest* dilakukan kedua metode memiliki karakteristik yang berbeda dalam proses klasifikasi jenis fitur, distribusi data dan karakteristik sinyal EKG, selain itu juga evaluasi performa tidak cukup hanya berdasarkan nilai *accuracy* terutama apabila data memiliki distribusi kelas yang tidak seimbang[9] karena itu penelitian ini menggunakan beberapa *matrix evaluasi*, yaitu *accuracy*, *precision recall*, *F1-score*, dan *confusion matrix* agar hasil perbandingan model dapat dianalisis secara lebih menyeluruh.[10]

Tujuan penelitian ini adalah untuk menganalisis perbandingan bagaimana kedua metode *Support Vector Machines dan Random Forest* bekerja dalam menentukan gangguan jantung berdasarkan sinyal EKG, rasio yang digunakan 80% data digunakan sebagai data latih dan 20% digunakan sebagai data uji,[11] agar proses perbandingan dapat dilakukan dengan seimbang kedua metode dilatih untuk menggunakan fitur yang sama kemudian hasil evaluasi digunakan untuk menentukan teknik yang memiliki kinerja klasifikasi yang lebih baik Ketika menentukan kondisi berdasarkan sinyal EKG.

## II. METHODS (FOR ORIGINAL RESEARCH ARTICLE ONLY)

Penelitian ini metode yang digunakan kuantitatif eksperimental digunakan untuk membandingkan kinerja dua metode pembelajaran mesin *Support Vector Machine (SVM)* dan *Random Forest*. SVM ini digunakan untuk mengklasifikasikan kondisi abnormal berdasarkan sinyal EKG yaitu dua skenario klasifikasi, klasifikasi *apnea* dan klasifikasi *Arrhythmia*, alur penelitian yang sama digunakan untuk setiap skenario ini mencakup pengumpulan data, *preprocessing* sinyal EKG, deteksi *R-Peak*, ekstraksi fitur pembagian data, pelatihan model.

Dataset yang digunakan dalam penelitian ini merupakan dataset public sinyal EKG. Database *Apnea-ECG physionet* sedangkan data *Arrhythmia* diperoleh dari database *Arrhythmia MIT-BIH*, Dataset *apnea* menunjukkan kondisi *normal breathing* dan *possible sleep apnea* dan dataset *Arrhythmia* digunakan untuk klasifikasi kondisi *heartbeat* dan *Arrhythmia*, dataset *apnea* dan *Arrhythmia* tidak diproses secara bersamaan karena mereka berasal dari

sumber yang berbeda hal ini dilakukan untuk mengurangi kemungkinan ketidaksesuaian dalam format data, sifat sinyal, dan struktur label yang dapat mempengaruhi hasil perbandingan model, oleh karena itu klasifikasi *apnea* dan *Arrhythmia* dianggap sebagai jenis tes yang berbeda.

*Preprocessing* sinyal EKG, sebelum proses ekstraksi fitur dan klasifikasi dimulai sinyal EKG diproses dengan preprosesing untuk meningkatkan kualitas sinyal, pada tahap ini data sinyal EKG dibaca dari file CSV, penyaringan sinyal dan persiapan sinyal untuk proses deteksi *R-Peak*, *bandpass* digunakan untuk mengurangi gangguan *baseline wander* dan *noise* pada frekuensi tinggi batas frekuensi bawah 0,5 hz dan batas frekuensi atas 40 hz

Sinyal ECG umumnya mengandung noise seperti *baseline wander*, noise otot, dan gangguan listrik. Salah satu bentuk umum filter digital adalah:

$$y(n) = \sum_{i=0}^M b_i x(n-i) - \sum_{j=1}^N a_j y(n-j)$$

Dimana, proses penyaringan sinyal digital dapat dinyatakan sebagai berikut :

$x(n)$  = sinyal input ECG  
 $y(n)$  = sinyal ECG setelah filtering  
 $a_j, b_i$  = koefisien filter

Selain menyaring sinyal, normalisasi juga dilakukan untuk membuat amplitudo sinyal EKG lebih sebanding sebelum diproses oleh SVM, ini penting karena perbedaan skala antar fitur dapat mempengaruhi proses pelatihan, ini terutama berlaku untuk metode SVM yang sensitive terhadap skala data.

Normalisasi digunakan agar amplitudo sinyal ECG berada pada skala yang sebanding sebelum diproses oleh model.

$$x_{norm}(n) = \frac{x(n) - \mu}{\sigma}$$

Dimana,  
 $x(n)$  = Sinyal EKG ke-n  
 $\mu$  = rata-rata sinyal EKG  
 $\sigma$  = standar deviasi sinyal EKG

digunakan dalam penelitian ini setelah proses *filtering* deteksi puncak R dilakukan untuk menemukan puncak utama pada sinyal EKG, puncak R yang berurutan yang terdeteksi kemudian digunakan untuk menghitung interval RR digunakan karena perubahan jarak antara denyut jantung dapat menunjukkan perubahan irama jantung atau beberapa kondisi fisiologis. Interval RR dapat dihitung dari posisi puncak R pada sinyal EKG

$$RR_i = \frac{R_{i+1} - R_i}{f_s}$$

Dimana,

$R_i$  = indeks puncak R ke-i

$f_s$  = frekuensi sampling EKG

$RR_i$  = interval antar puncak R

Nilai denyut jantung digunakan sebagai fitur karena dapat mempresentasikan pola denyut jantung pada sinyal EKG, denyut jantung juga dapat dihitung berdasarkan interval RR. Rumus heart rate adalah sebagai berikut:

$$HR = \frac{60}{RR}$$

Dimana,

$HR$  = heart rate dalam bpm

$RR$  = interval R-R dalam detik

Ekstraksi fitur dilakukan untuk mengubah sinyal EKG menjadi data numerik yang dapat digunakan sebagai masukan pada model *machine learning*, karakteristik sinyal EKG, posisi *R-peak*, interval RR, *heart rate* dan pola segmen sinyal, dalam klasifikasi data diperoleh dari ekstraksi fitur digunakan untuk membentuk matriks fitur yang memiliki fitur 401 dimensi, setiap baris data menunjukkan satu segmen sinyal EKG dan setiap kolom menunjukkan satu nilai fitur yang digunakan untuk klasifikasi, beberapa fitur statistik yang digunakan dapat dihitung menggunakan rumus berikut:

Rata-rata

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i$$

Standar Deviasi

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \mu)^2}$$

Energi Sinyal

$$E = \sum_{i=1}^N x_i^2$$

fitur-fitur tersebut kemudian digabungkan dengan fitur klinis EKG seperti:

*RR, HR, QRS duration, PR interval, QT interval*

input pada metode SVM dan *Random Forest* untuk membedakan antara kondisi normal dan abnormal

pada EKG.

Pembagian data dataset dibagi menjadi 80 persen data latih untuk membangun model klasifikasi, 80 persen data digunakan sebagai data latih untuk membangun model klasifikasi dan 20 persen digunakan sebagai data uji untuk mengevaluasi kemampuan model untuk mengenali data baru, agar proses perbandingan dapat dilakukan secara seimbang metode SVM dan Random Forest menggunakan rasio pembagian data yang sama.

Salah satu metode klasifikasi yang digunakan dalam penelitian ini adalah *Support Vector Machines, Hyperplane* optimal yang memiliki kemampuan untuk memisahkan data kedalam berbagai kelas adalah cara SVM bekerja, dalam penelitian ini SVM dilatih dengan sinyal EKG yang diekstraksi dari tahap *preprocessing* dan ekstraksi fitur, karena SVM *sensitive* terhadap perbedaan skala antara fitur data fitur dinormalisasikan atau diskalakan sebelum proses pelatihan, support vector machine digunakan untuk mencari hyperplane terbaik yang memisahkan data EKG normal dan abnormal

$$f(x) = w^T x + b$$

Dimana,

$w$  = vector bobot

$x$  = vector fitur EKG

$b$  = bias

selanjutnya model SVM dilatih dengan data latih dan data uji untuk mengetahui bagaimana kinerja klasifikasi model prediksi SVM. Jika data tidak dapat dipisahkan secara linear,

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i$$

Dengan kendala:

$$y_i(w^T x_i + b) \geq 1 - \xi_i$$

$$\xi_i \geq 0$$

Dimana,

$C$  = parameter penalty kesalahan klasifikasi

$\xi_i$  = slack variable

$y_i$  = label kelas, misalnya normal atau gangguan jantung

SVM dapat menggunakan fungsi kernel, radial basis function (RBF) adalah salah satu kernel yang paling umum digunakan karena mampu menangani pola data yang tidak linear, Rumus kernel RBF adalah sebagai berikut:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$$

Dimana,

$\gamma$  = parameter kernel

$x_i, x_j$  = dua vector fitur EKG

*Random Forest* digunakan sebagai metode pembandingan terhadap SVM. *Random Forest* adalah algoritma *ensemble learning* yang membangun sejumlah *decision tree* dan menggunakan voting mayoritas dari seluruh *tree* untuk menentukan hasil klasifikasi. Model random forest dilatih menggunakan data latih dan fitur yang sama dengan SVM tujuan penggunaan fitur dan pembagian data yang sama adalah untuk memastikan perbandingan performa kedua metode dilakukan pada kondisi eksperimen yang setara kemampuan *Random Forest* untuk menangani hubungan fitur yang kompleks yang cukup stabil terhadap variasi data. Hasil akhir klasifikasi random forest ditentukan menggunakan rumus voting berikut:

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), \dots, h_B(x)\}$$

Dimana,

$h_B(x)$  = prediksi dari pohon ke-  $B$

$\hat{y}$  = kelas akhir hasil klasifikasi

Evaluasi model dilakukan untuk mengetahui bagaimana SVM dan *Random Forest* mengklasifikasikan sinyal EKG, metrik evaluasi yang digunakan termasuk *accuracy, precision, recall, F1-score, error rate, confusion matrix dan person correlation*[12], proporsi prediksi yang benar terhadap seluruh data uji ukur dengan *accuracy*, ketepatan model dalam memprediksi suatu kelas diukur dengan *precision* kemampuan model. untuk menemukan data yang benar-benar termasuk kedalam kelas tertentu diukur dengan *recall, F1-score* digunakan untuk mengevaluasi ketepatan dan *recall, error rate* dihitung menggunakan rumus.

Accuracy

Accuracy digunakan untuk menghitung proporsi prediksi yang benar terhadap seluruh data uji

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Presisi

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall

$$\text{Recall} = \frac{TP}{TP + FN}$$

Specificity

$$Specificity = \frac{TN}{TN + FP}$$

F1-Score

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Error rate

$$Error Rate = 1 - Accuracy$$

Selain itu jumlah prediksi yang benar dan salah untuk masing-masing kelas dihitung dengan menggunakan *matrix confusion*, [13] analisis statistik tambahan dengan korelasi *pearson* digunakan untuk mengevaluasi hubungan pola performa antara SVM dan *Random Forest* berdasarkan metrik evaluasi yang diperoleh namun korelasi *pearson* hanya digunakan sebagai tambahan untuk analisis, *precision*, *accuracy*, *recall*, *F1-score*, *error rate* dan *confusion matrix* adalah factor utama yang menentukan perbandingan utama antara SVM dan *Random Forest*.

### III. RESULTS AND DISCUSSION (REVIEW ARTICLE USE DISCUSSION)

Data elektrokardiogram digunakan dalam skenario klasifikasi apnea untuk membedakan antara kondisi napas normal dan kemungkinan *possible sleep apnea*, karena data *apnea* dianalisis secara terpisah dari data *Arrythmia* karena berasal dari *notebook* yang berbeda, tahapan yang dilakukan meliputi *preprocessing* sinyal, deteksi *R-Peak*, perhitungan RR interval, ekstraksi fitur, pembagian data latih dan data uji, serta klasifikasi menggunakan SVM dan *Random Forest*, menurut hasil pengujian metode *Random Forest* menunjukkan kinerja yang lebih baik daripada SVM, metode *Random Forest* memperoleh nilai *accuracy* sebesar 97,23% sedangkan SVM memperoleh *accuracy* sebesar 91,44%.

Tabel 1. Hasil Klasifikasi Apnea

Tabel 1 menunjukkan bahwa *random forest* memiliki Tingkat error yang lebih rendah dibandingkan SVM dan selisih keakuratan 5.79% ini menunjukkan bahwa berdasarkan karakteristik sinyal EKG *random forest* lebih efektif dalam mengklasifikasikan kondisi napas normal dan kemungkinan *sleep apnea*, keunggulan *random forest* dapat dijelaskan melalui mekanisme *ensemble learning* yaitu pembentukan banyak pohon keputusan yang kemudian digabungkan untuk menghasilkan pohon keputusan akhir, disisi lain SVM masih menghasilkan hasil yang baik tetapi masih lebih rendah dibandingkan *random forest* karena distribusi data pemilihan kernel dan parameter model.

*Confusion matrix* dipakai untuk melihat distribusi hasil prediksi model terhadap kelas aktual, dalam klasifikasi *apnea* kelas 0 menunjukkan pernafasan normal sedangkan kelas 1 menunjukkan pernafasan yang merepresentasikan *possible sleep apnea*.

Figure 5 Confusion Matrix Random Forest Pada Klasifikasi Apnea

Figure 5 menunjukkan bahwa *random forest* berhasil mengklasifikasikan 18.976 data *possible sleep apnea* dengan benar sebagai kelas 1 dan 14.743 data normal *breathing* dengan benar sebagai kelas 0. Kesalahan klasifikasi terdiri dari 663 kesalahan positif yaitu data normal *breathing* yang diprediksi menunjukkan bahwa data *possible sleep apnea*, dan 296 kesalahan negatif yang menunjukkan bahwa data *possible sleep apnea* yang diprediksi sebagai normal *breathing*, jumlah kesalahan klasifikasi *random forest* sebesar 959 data dari 34.678 data uji mendukung nilai akurasi *random forest* sebesar 97,23.

Figure 6 Confusion Matriks SVM Pada Klasifikasi Apnea

Figure 6 SVM berhasil mengklasifikasikan 13.854 data normal *breathing* dengan benar sebagai kelas 0 dan 17.856 data *possible sleep apnea* dengan benar sebagai kelas 1. Tetapi SVM menghasilkan kesalahan yang lebih besar dibandingkan dengan *random forest* yaitu 1.552 false positive dan 1.416 negative, total kesalahan pada klasifikasi SVM 2.968 dari total 34.678 data uji, jumlah kesalahan ini menyebabkan *accuracy* SVM lebih rendah dibandingkan *random forest* yaitu 91,44%.

Dalam klasifikasi aritmia sinyal EKG digunakan untuk membedakan antara normal *heartbeat* dan *Arrythmia*, hasil diagnosis pasien menunjukkan bahwa ada 14 subjek normal, 3 subjek detak jantung lambat (*bradycardia*), dan 3 subjek detak jantung cepat (*tachycardia*), oleh karena itu kelas *Arrythmia* dalam penelitian ini terdiri dari *bradycardia*, *tachycardia* dan normal, hasil analisis interval RR dan BPM menunjukkan media RR minimum sebesar 536,11 ms dan median RR maksimum sebesar 1180,56 ms, nilai BPM minimum adalah 50,82 BPM dan nilai tertinggi BPM adalah 111,92 bpm subjek dengan diagnosis *bradycardia* memiliki nilai bpm terendah sebaliknya BPM tertinggi ditemukan pada subjek *tachycardia*, hasil ini menunjukkan bahwa fitur RR interval dan BPM dapat menunjukkan perbedaan pola denyut jantung antara *bradycardia*, *tachycardia* dan kondisi denyut jantung normal.

Tabel 2 Hasil Klasifikasi Aritmia

Tabel 2 menunjukkan hasil pengujian model pada skenario metode *random forest* kembali menunjukkan performa yang lebih baik dibandingkan SVM

keakuratan random forest adalah 97,67 % dengan Tingkat error 2,33% sedangkan keakuratan SVM adalah 95,96% dengan Tingkat error 4,04%, hasil menunjukkan bahwa kedua metode memiliki kemampuan untuk melakukan klasifikasi aritmia dengan sangat baik namun *random forest* tetap memberikan hasil yang lebih baik karena memiliki akurasi yang lebih tinggi dengan Tingkat kesalahan lebih rendah dibandingkan SVM perbedaan akurasi keduanya adalah 1,71%.

Analisis *confusion matrix* pada klasifikasi aritmia kelas 0 mempresentasikan normal *heartbeat* sedangkan kelas 1 mempresentasikan *Arrhythmia*.

Figure 7 Confusion Matrix Random Forest Pada Klasifikasi *Arrhythmia*

Figure 7 dari data uji 8.487 random forest menghasilkan 198 kesalahan klasifikasi terdiri dari 5.507 data aritmia normal dengan benar kelas 0 dan 2.782 data aritmia normal dengan benar kelas 1. Kesalahan klasifikasi terdiri dari 68 kesalahan positif yang menunjukkan data aritmia normal yang diprediksi dan 130 kesalahan negatif yang menunjukkan data aritmia normal yang diprediksi.

Figure 8 Confusion Matrix SVM Pada Klasifikasi *Arrhythmia*

Berdasarkan Figure 8 berhasil mengklasifikasikan 5.473 data normal *heartbeat* dengan benar sebagai kelas 0 dan 2.671 data *Arrhythmia* dengan benar kelas 1, dari 8.487 data uji, SVM menghasilkan 343 kesalahan klasifikasi data, terdiri dari 102 kesalahan positif, yang menunjukkan data *Arrhythmia* normal yang diprediksi, dan 241 kesalahan negatif yang menunjukkan data *Arrhythmia* normal yang diprediksi.

*Random forest* menghasilkan jumlah kesalahan klasifikasi yang lebih rendah dibandingkan SVM, menurut perbandingan *confusion matrix*. *Random forest* menghasilkan 198 kesalahan, sedangkan SVM menghasilkan 343 kesalahan. Selain itu, *Random Forest* memiliki nilai *false negative* yang lebih rendah daripada SVM, sehingga *Random Forest* lebih jarang gagal mendeteksi data *Arrhythmia*.

Perbandingan performa SVM dan Random Forest secara keseluruhan *Random Forest* lebih baik daripada SVM dalam kedua skenario klasifikasi (*apnea* dan *Arrhythmia*).

Tabel 3 Perbandingan performa SVM dan *Random Forest* pada klasifikasi *apnea* dan *Arrhythmia*

Tabel tersebut menunjukkan bahwa *Random Forest* memiliki akurasi yang lebih tinggi, Tingkat kesalahan yang lebih rendah, dan jumlah kesalahan klasifikasi yang lebih sedikit daripada SVM pada kedua skenario

klasifikasi. *Random Forest* menunjukkan performa yang lebih konsisten dibandingkan SVM pada dataset yang digunakan. Ini terbukti dengan selisih 5,79% pada klasifikasi *apnea* dan 1,71% pada klasifikasi *Arrhythmia*.

#### IV. CONCLUSION

Hasil penelitian menunjukkan bahwa metode *Random Forest* lebih baik dari pada *Support Vector Machines (SVM)* dalam klasifikasi *apnea* dan *Arrhythmia* berbasis sinyal elektrokardiogram (EKG). Metode *Random Forest* memperoleh akurasi sebesar 97,23% dengan *error rate* 2,77%, sedangkan SVM memperoleh akurasi sebesar 91,44% dengan *error rate* 8,56%. Selisih akurasi antara kedua metode pada klasifikasi *apnea* adalah 5,79% yang menunjukkan bahwa *Random Forest* lebih baik daripada SVM dalam klasifikasi.

*Random Forest* memiliki akurasi sebesar 97,67% dengan *error rate* 2,33% pada klasifikasi *Arrhythmia*, sedangkan SVM memiliki akurasi sebesar 95,96% dengan *error rate* 4,04%. Selisih akurasi antara dianggap sebagai metode adalah 1,71%. Oleh karena itu, *Random Forest* dapat dianggap sebagai metode yang lebih efektif dalam kedua skenario klasifikasi. Untuk pengembangan kedepan, penelitian selanjutnya bisa menambah jumlah subjek yang lebih banyak agar mendapatkan hasil yang lebih representatif, dan juga menambahkan metode klasifikasi lain seperti *K-Nearest Neighbor*, *Naive Bayes* atau metode *deep learning* lainnya agar memperoleh perbandingan performa yang lebih baik.

#### V. ACKNOWLEDGEMENTS

Penulis mengucapkan terima kasih kepada Program Studi Teknik Elektro, Ketua Program Studi, Bapak Ir. Subairi, ST., MT., IPM serta Dosen Pembimbing I Ibu Basitha Febrinda Hidayatulail, SST., MT. dan Dosen Pembimbing II Bapak Ir. Rahman Arifuddin, S.T., M.T. atas bimbingan, arahan, dan dukungan selama proses penyusunan penelitian ini. Penulis juga berterima kasih kepada semua pihak yang telah membantu, baik secara langsung maupun tidak langsung, dalam penyelesaian penelitian ini

#### REFERENCES

- [1] A. Darmawahyuni *et al.*, "An improved electrocardiogram arrhythmia classification performance with feature optimization," *BMC Med. Inform. Decis. Mak.*, vol. 24, no. 1, 2024, doi: 10.1186/s12911-024-02822-7.
- [2] A. K. Singh and S. Krishnan, *ECG signal feature extraction trends in methods and applications*, vol. 22, no. 1. BioMed Central, 2023. doi: 10.1186/s12938-023-01075-1.
- [3] P. N. Malleswari, V. Krishna Odugu, T. J. V. S.

- Rao, and T. V. N. L. Aswini, "Deep learning-assisted arrhythmia classification using 2-D ECG spectrograms," *EURASIP J. Adv. Signal Process.*, vol. 2024, no. 1, pp. 1–15, 2024, doi: 10.1186/s13634-024-01197-1.
- [4] D. A. S, F. Arisgraha, and R. Apsari, "Bradycardia and Tachycardia Detection System With Artificial Neural Network Method," *Indones. J. Trop. Infect. Dis.*, vol. 3, no. 2, p. 86, 2016, doi: 10.20473/ijtid.v3i2.206.
- [5] G. Moody and R. Mark, "MIT-BIH Arrhythmia Database," PhysioNet. Accessed: May 29, 2026. [Online]. Available: <https://physionet.org/content/mitdb/1.0.0/>
- [6] G. Moody and R. Mark, "Apnea-ECG Database," PhysioNet. Accessed: May 29, 2026. [Online]. Available: <https://physionet.org/content/apnea-ecg/1.0.0/>
- [7] Mariette Awad and Rahul Khanna, "Support Vector Machines for Classification SVM from a Geometric Perspective," *Efficint Learn. Mach.*, pp. 39–66, 2015.
- [8] Z. Jin, J. Shang, Q. Zhu, C. Ling, W. Xie, and B. Qiang, "RFRSF: Employee Turnover Prediction Based on Random Forests and Survival Analysis," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12343 LNCS, pp. 503–515, 2020, doi: 10.1007/978-3-030-62008-0\_35.
- [9] S. Riyanto, I. S. Sitanggang, T. Djatna, and T. D. Atikah, "Comparative Analysis using Various Performance Metrics in Imbalanced Data for Multi-class Text Classification," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 6, pp. 1082–1090, 2023, doi: 10.14569/IJACSA.2023.01406116.
- [10] Z. Maisat, E. Darmawan, and A. Fauzan, "Implementasi Optimasi Hyperparameter GridSearchCV Pada Sistem Prediksi Serangan Jantung Menggunakan SVM Implementation of GridSearchCV Hyperparameter Optimization in Heart Attack Prediction System Using SVM," *J. Ilm. Sist. Inf.*, vol. 13, no. 1, pp. 8–15, 2023.
- [11] A. Rácz, D. Bajusz, and K. Héberger, "Effect of Dataset Size and Train / Test Split Ratios in," *Eff. Dataset Size Train/Test Split Ratios QSAR/QSPR Multiclass Classif.*, vol. 26(4), no. 1111, pp. 1–16, 2021, [Online]. Available: <https://doi.org/10.3390/molecules26041111>
- [12] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, pp. 1–13, 2020, doi: 10.1186/s12864-019-6413-7.
- [13] Ž. Vujović, "Classification Model Evaluation Metrics," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 6, pp. 599–606, 2021, doi: 10.14569/IJACSA.2021.0120670.

\*Correspondent e-mail address  
[intanrizky520@gmail.com](mailto:intanrizky520@gmail.com) Peer reviewed under responsibility of Universitas Merdeka Malang, Indonesia

© 2026 Muhammadiyah University Sidoarjo, All right reserved, This is an open access article under the CC BY license(<http://creativecommons.org/licenses/by/4.0/>)

Received: 2026-06-03  
Accepted: 2026-06-03  
Published: 2026-07-02

**DAFTAR TABEL**

Tabel 1. Hasil Klasifikasi Apnea ..... 17  
Tabel 2. Hasil Klasifikasi Arrythmia ..... 17  
Tabel 3. Perbandingan performa SVM dan Random Forest pada klasifikasi apnea dan Arrythmia ..... 17

Tabel 1. Hasil Klasifikasi Apnea

Metode	Keakuratan	Tingkat Error	Interpretasi
Random Forest	97,23%	2,77%	Performa klasifikasi lebih baik
SVM	91,44%	8,56%	Performa klasifikasi lebih rendah
Selisih	5,79%	5,79%	Random Forest lebih unggul

Tabel 2. Hasil Klasifikasi Arrythmia

Metode	Keakuratan	Tingkat Error	Interpretasi
Random Forest	97,67%	2,33%	Performa klasifikasi lebih baik
SVM	95,96%	4,04%	Performa klasifikasi lebih rendah
Selisih	1,71%	1,71%	Random Forest lebih unggul

Tabel 3. Perbandingan performa SVM dan *Random Forest* pada klasifikasi *apnea* dan *Arrythmia*

Klasifikasi	Random Forest	SVM	Selisih	Metode Terbaik
Apnea	97,23%	91,44%	5,79%	Random Forest
	Error 2,77%	Error 8,56%		
Arrythmia	97,67%	95,96%	1,71	Random Forest
	Error 2,33%	4,04%		

**DAFTAR GAMBAR**

Figure 1. Grafik Sinyal Subjek Normal .....	19
Figure 2. Grafik Sinyal Subjek <i>Bradycardia</i> .....	19
Figure 3. Grafik Sinyal Subjek <i>Tachycardia</i> .....	19
Figure 4. Grafik Sinyal Subjek <i>Apnea</i> .....	19
Figure 5. <i>Confusion Matrix Random Forest</i> Pada Klasifikasi <i>Apnea</i> .....	20
Figure 6. <i>Confusion Matrix SVM</i> Pada Klasifikasi <i>Apnea</i> .....	20
Figure 7. <i>Confusion Matrix Random Forest</i> Pada Klasifikasi <i>Arrythmia</i> .....	20
Figure 8. <i>Confusion Matrix SVM</i> Pada Klasifikasi <i>Arrythmia</i> .....	20

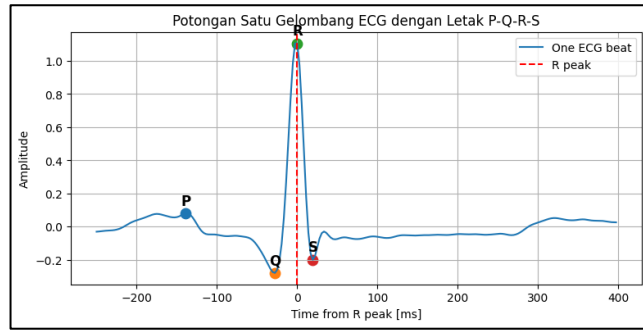


Figure 1. Grafik Sinyal Subjek Normal

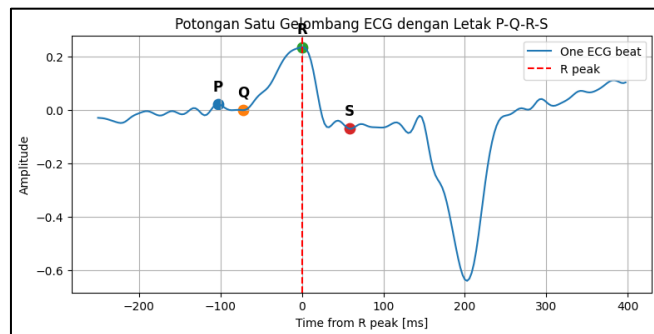


Figure 2. Grafik Sinyal Subjek Bradycardia

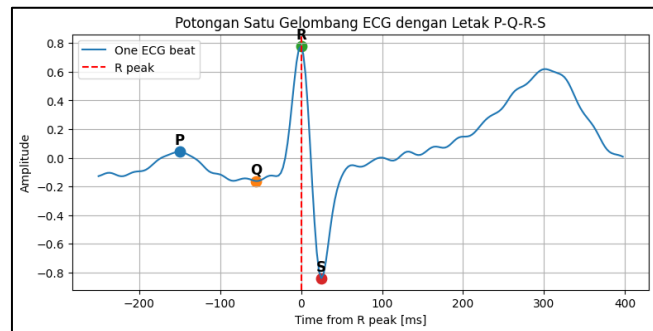


Figure 3. Grafik Sinyal Subjek Tachycardia

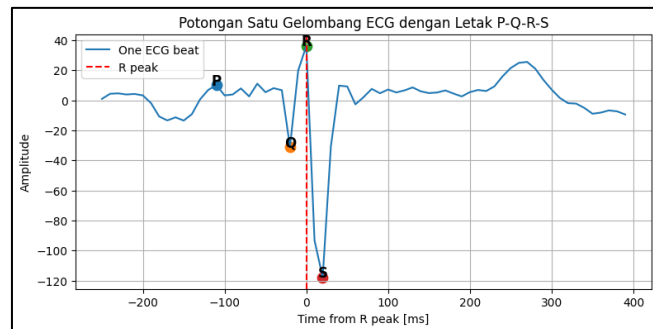


Figure 4. Grafik Sinyal Subjek Apnea

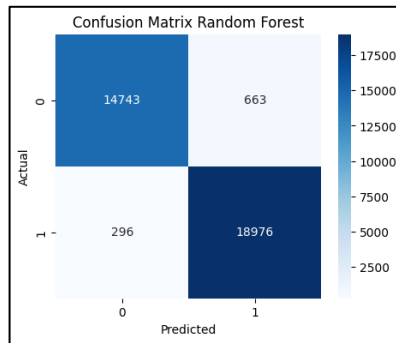


Figure 5. Confusion Matrix Random Forest Pada Klasifikasi Apnea

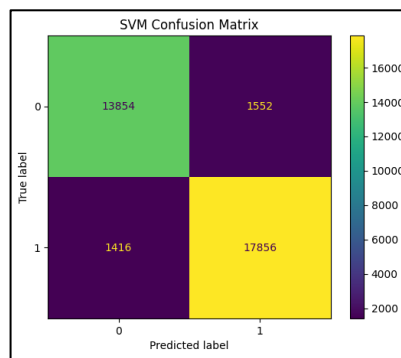


Figure 6. Confusion Matrix SVM Pada Klasifikasi Apnea

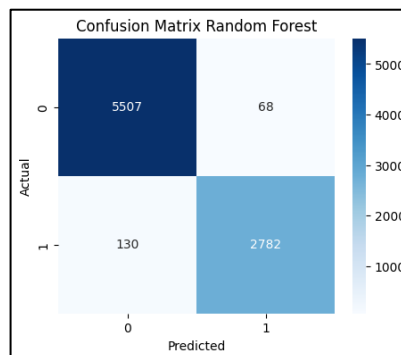


Figure 7. Confusion Matrix Random Forest Pada Klasifikasi Arrhythmia

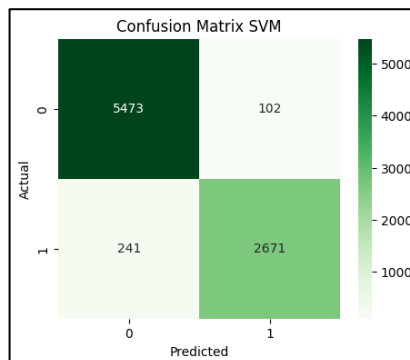


Figure 8. Confusion Matrix SVM Pada Klasifikasi Arrhythmia