

# Analisis Fitur Domain Waktu ECG *Heart Rate Variability* Berdasarkan *Gain Informasi*

UNANG SUNARYA<sup>1</sup>, LYRA VEGA UGI<sup>2</sup>

<sup>1</sup>Fakultas Ilmu Terapan, Universitas Telkom, Indonesia

<sup>2</sup>Fakultas Teknik Elektro, Universitas Telkom, Indonesia

Email: unangsunarya@telkomuniversity.ac.id

*Received* 6 November 2021 | *Revised* 8 Desember 2021 | *Accepted* 25 Januari 2022

## ABSTRAK

*Salah satu analisis yang sering digunakan untuk mendeteksi berbagai penyakit yang terkait fungsi jantung adalah dengan menghitung rasio tidur dan terjaganya seseorang saat tidur pada malam hari. Namun dalam praktiknya sering ditemukan kendala baik pada saat akuisisi data ataupun pada saat pemrosesan data untuk menyimpulkan sebuah hasil analisis. Salah satu penyebabnya adalah kurang tepatnya dalam pemilihan fitur ECG untuk proses implementasi. Oleh karena itu, pada penelitian ini dilakukan seleksi fitur ECG pada domain waktu dan klasifikasi kondisi tidur dan terjaga dari 10 subjek berdasarkan fitur Heart Rate Variability (HRV) dengan menggunakan algoritma random forest. Wavelet digunakan untuk mendapatkan komponen sinyal ECG yang tepat sedangkan gain informasi digunakan untuk memilih fitur ECG yang dominan. Hasil akhir dari implementasi menunjukkan nilai akurasi rata-rata sebesar 80,26 % dengan fitur terbaik adalah median nearest neighbor index.*

**Kata kunci:** ECG wavelet, HRV, domain waktu, gain informasi, akurasi

## ABSTRACT

*One of the most frequent methods to detect heart-related diseases is by calculating the ratio of total sleep and awake of someone during night sleep. However, it is often encountered problems either in data acquisition or data processing to output the results of the analysis. One of the reasons is selecting ECG features improperly during implementation. Therefore, this study has been conducted ECG features selection in the time domain and classification of sleep and awake states across 10 subjects based on Heart Rate Variability (HRV) features obtained using a random forest algorithm. Wavelet was used to get the proper ECG signal components while information gain was used to select the dominant ECG features. The implementation results showed an average accuracy of 80.26 % with the median nearest neighbor index as the best feature.*

**Keywords:** ECG, wavelet, HRV, time-domain, information gain, accuracy

## 1. PENDAHULUAN

Jantung merupakan organ vital pada manusia yang berfungsi memompakan darah ke seluruh tubuh. Berbagai penelitian tentang penyakit yang berkaitan dengan organ jantung telah banyak dilakukan. Pada tahun 2020, Islam (**Islam, dkk, 2020**) melakukan penelitian prediksi dini penyakit jantung dengan menggunakan metode PCA. Prediksi penyakit jantung dapat juga dilakukan dengan menggunakan metode *hidden naïve Bayes* seperti yang dilakukan oleh Jabbar (**Jabbar & Samreen, 2016**). Selain untuk memprediksi penyakit jantung, sinyal ECG juga dapat digunakan untuk mendeteksi kelelahan (**Szypulska & Piotrowski, 2012**), gangguan tidur seseorang di malam hari (**Zarei & Asl, 2019**) (**Jung, dkk, 2017**), yang nantinya dapat digunakan untuk mengukur kualitas tidur seseorang (**Mendonca, dkk, 2020**).

Berbagai macam algoritma dan *feature engineering* telah banyak dilakukan untuk mendapat fitur yang paling menonjol dalam mendeteksi ataupun mengklasifikasi karakteristik sinyal ECG, mulai dari fitur statistik level amplitudo ataupun pola sinyal ECG (**Kansal, dkk, 2015**). Zheng (**Zheng, dkk, 2013**) menggunakan *Support Vector Machine* (SVM) untuk mendeteksi *anomali* pada sinyal ECG. Algoritma *Artificial Neural Network* (ANN) telah digunakan untuk mendiagnosis penyakit jantung dengan metode auto analisis oleh Raj (**Raj, dkk, 2014**). Pada tahun 2016, Deepak (**Gautam & Giri, 2016**) melakukan penelitian sinyal ECG untuk mendiagnosis penyakit menggunakan analisis *Heart Rate Variability* (HRV). Pada penelitiannya mengekstrak fitur yang dihasilkan oleh perbedaan jarak di antara puncak sinyal ECG (*R-peaks*) yang berurutan. Namun dalam implementasinya belum mendapatkan fitur yang optimal. Sulitnya mendapatkan fitur yang menonjol sedikit lebih banyak dipengaruhi karena keberadaan *noise* atau pun sinyal lain selama proses akuisisi, seperti *noise* yang berasal dari pernafasan (**Vityazeva, dkk, 2018**). Oleh karena itu diperlukan sebuah teknik untuk menghilangkan atau mengurangi *noise* pada sinyal sebelum diekstrak cirinya. Fitur-fitur yang telah diekstrak dari sinyal ECG dapat mempengaruhi kinerja algoritma, baik dalam hal akurasi ataupun waktu komputasi dalam mengambil keputusan (**Vityazeva, dkk, 2018**). Maka dari itu diperlukan sebuah metode pemilihan fitur yang tepat sebelum digunakan.

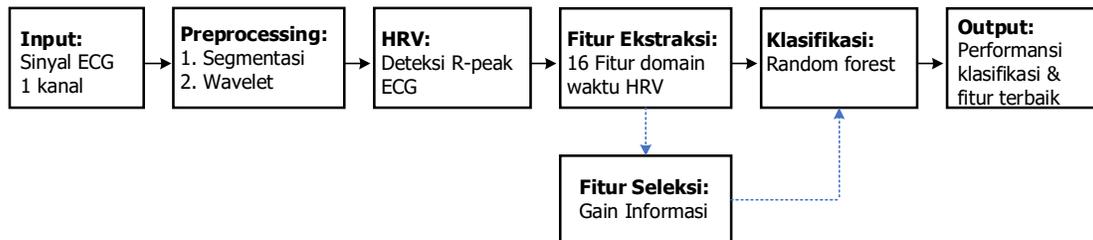
Pada penelitian ini, sinyal ECG dari subjek pada saat tidur dianalisis menggunakan HRV pada domain waktu, kemudian diklasifikasikan ke dalam dua kondisi yaitu kondisi tidur dan terjaga menggunakan metode *random forest*. Untuk menghilangkan *noise* frekuensi tinggi pada sinyal dan untuk mendapatkan fokus pada bagian QRS sinyal ECG, *wavelet* model *symlet4* digunakan dan kemudian dipilih komponen yang paling bersesuaian dengan pola QRS pada sinyal ECG sebelum dikonstruksi kembali. Kemudian fitur seleksi *information gain* digunakan untuk mencari fitur ECG yang paling dominan dalam mencapai kinerja akurasi. Validasi pengujian dilakukan dengan metode *One-Subject-Leave-Out cross-validation* (OSLO), dimana setiap satu subjek uji, sembilan subjek lainnya bertindak sebagai data latih.

Selanjutnya pada bab 2, dibahas tentang metodologi penelitian yang digunakan selama implementasi. Bab 3 membahas hasil yang dicapai selama implementasi. Pada bagian akhir ditarik kesimpulan tentang hasil yang dicapai sesuai metode yang digunakan.

## 2. METODOLOGI PENELITIAN

### 2.1 Diagram Blok Implementasi

Gambar 1 menunjukkan diagram lengkap implementasi yang diawali dari proses akuisisi data ECG, pemrosesan awal sinyal ECG, proses ekstraksi fitur ECG, dan klasifikasi. Dalam skema implementasi ini, digunakan seleksi fitur untuk mencari fitur terbaik dari kumpulan fitur yang telah diekstraksi, yang memiliki peran penting dalam mencapai kinerja klasifikasi.



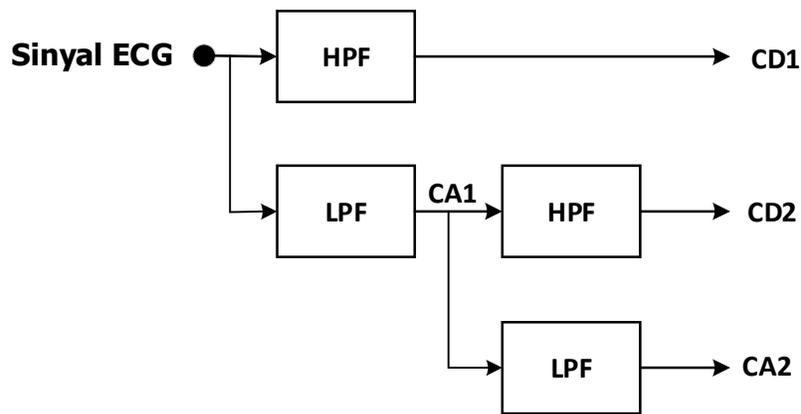
**Gambar 1. Diagram Blok implementasi sistem klasifikasi keseluruhan**

### 2.2 Dataset

Pada penelitian ini digunakan sepuluh subjek sebagai sampel penelitian yang diambil dari *dataset 'St. Vincent's University Hospital / University College Dublin Sleep Apnea Database' (Goldberger, dkk, 2000)*. Sinyal ECG diambil dari subjek yang sedang tidur di malam hari dan semua subjek diduga memiliki gangguan tidur. Rekaman sinyal ECG dari subjek terdiri dari kondisi tidur dan kondisi terbangun dari tidur. Setiap sampel subjek diambil durasi empat jam pada satu kanal *Hotler V5* menggunakan *Reynolds Lifecard CF system* dengan frekuensi *sampling* 100 Hz. Pada penelitian ini, sebelum sinyal ECG didekomposisi dengan *wavelet*, setiap sinyal ECG dari masing-masing subjek disegmentasi dengan ukuran 3000 sampel untuk memudahkan dalam proses dekomposisi.

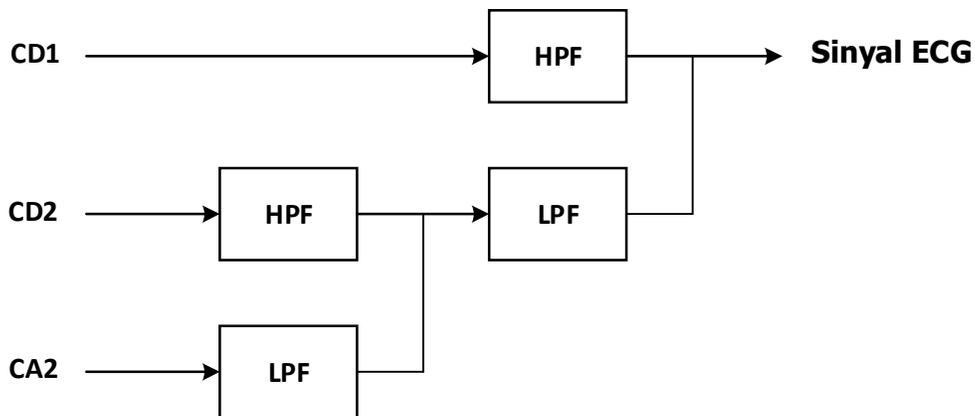
### 2.3 Wavelet Dekomposisi

*Wavelet* merupakan salah satu metode yang dapat digunakan untuk kompresi data ECG (Hilton, 2004) (Ku, dkk, 2006). Selain data yang berupa sinyal, *wavelet* juga sering digunakan untuk *watermarking* pada data berupa gambar, seperti yang dilakukan oleh Hafizhana (Hafizhana, dkk, 2020). *Wavelet* bekerja dengan cara mendekomposisi sinyal ke dalam beberapa komponen sinyal penyusunnya. Seperti halnya transformasi *fourier* yang dapat mendekomposisi sinyal menjadi beberapa sinyal penyusunnya, *wavelet* juga dapat melakukannya bahkan dengan tambahan informasi waktu kejadian. *Wavelet* mendekomposisi sinyal ke dalam komponen aproksimasi yang merupakan bagian sinyal dengan frekuensi rendah dan komponen *detail* yang mengandung sinyal dengan frekuensi yang lebih tinggi (Weeks, 2007). Adapun gambaran umum untuk dekomposisi sinyal ECG dua level dapat dilihat pada Gambar 2. Pada setiap level sinyal ECG dilewatkan pada *filter bank* yang terdiri dari pasangan *High Pass Filter* (HPF) dan *Low Pass Filter* (LPF), yang selanjutnya masing-masing diikuti dua kali penurunan level *sampling* (*down sampling*). *Component Detail* (CD1) dan *Component Average* (CA1) masing-masing merupakan keluaran dari pasangan filter HPF dan LPF dekomposisi tingkat pertama, yang mengalami tingkat penurunan *sampling* dua kali, artinya panjang sinyal pada CD1 dan CA1 menjadi setengah panjang sinyal asalnya. Sedangkan CD2 dan CA2 merupakan keluaran dari pasangan filter HPF dan LPF tingkat kedua dimana masing-masing sinyal mengalami penurunan sampel dua kali dari sinyal sebelumnya.



**Gambar 2. Dekomposisi Sinyal ECG dengan *Wavelet* pada Level 2**

Gambar 3 menunjukkan gambaran secara umum proses dasar rekonstruksi sinyal-sinyal hasil dekomposisi *wavelet*. Dari hasil dekomposisi dapat dipilih bagian sinyal mana saja yang paling dominan untuk dikonstruksi.



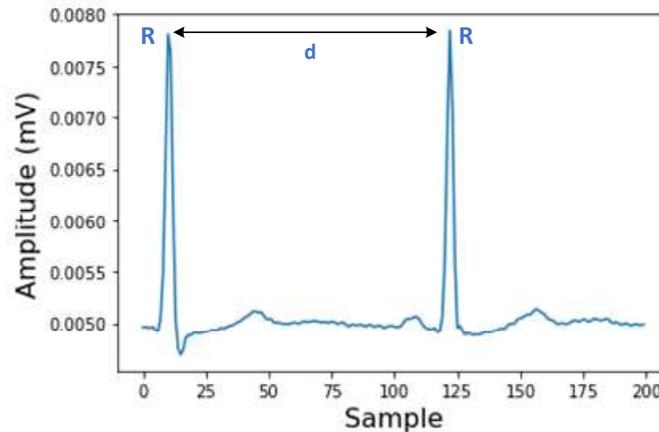
**Gambar 3. Rekonstruksi Komponen Sinyal ECG Menjadi Sinyal ECG Penuh**

### 2.4 Heart Rate Variability

*Heart Rate Variability* (HRV) merupakan perbedaan rata-rata detak jantung (HR) yang dapat terjadi karena usia atau pengaruh lingkungan. HRV memungkinkan digunakan untuk mendeteksi kondisi normal atau tidak normalnya jantung seseorang (Jose, dkk, 2015). Detak jantung seseorang dapat berubah dipengaruhi oleh aktivitasnya. Dapat dirasakan dengan jelas perbedaan detak jantung seseorang ketika sedang berlari atau sedang dalam kondisi istirahat. Saat seseorang sedang berlari detak jantung menjadi lebih cepat dibandingkan ketika sedang beristirahat. HRV dapat dihitung dari jarak puncak ke puncak (*R-R peak*) dari dua siklus sinyal ECG yang berdekatan, seperti dapat dilihat pada Gambar 4. Jarak antara dua puncak yang berdekatan (Jose, dkk, 2015) dapat dihitung dengan menggunakan Persamaan (1) berikut:

$$d(ms) = \frac{R_s - R_{s-1}}{f_s} \times 1000 \quad (1)$$

Dimana  $d$  merupakan jarak dalam milidetik,  $R$  adalah posisi puncak dari satu siklus sinyal, dan  $f_s$  adalah frekuensi *sampling* yang digunakan, dalam hal ini adalah sebesar 100 Hz.



**Gambar 4. Jarak Dua Puncak (R-R Peak) dari Dua Siklus ECG yang Berdekatan**

## 2.5 Ekstraksi Fitur ECG

Ekstraksi fitur domain waktu diambil dari setiap sinyal ECG yang sudah dihitung nilai HRV-nya. Perbedaan nilai HRV akan menjadi parameter untuk 16 *ekstraktor* fitur yang digunakan. Dimana pada penelitian ini fokus pada fitur domain waktu (**Shaffer & Ginsberg, 2017**) seperti yang terlihat pada Tabel 1.

**Tabel 1. Fitur Domain Waktu HRV**

No.	Fitur	Keterangan
1	mean_nni	<i>Mean nearest neighbor index</i>
2	sdnn	<i>Standard deviation nearest neighbor</i>
3	sdsd	<i>Standard deviation differences of nearest neighbor interval</i>
4	nni_50	<i>Nearest neighbor index 50</i>
5	pnni_50	<i>Percentage of success rate nearest neighbor index 50 ms</i>
6	nni_20	<i>Nearest neighbor index 20</i>
7	pnni_20	<i>Percentage of success rate nearest neighbor index 20 ms</i>
8	rmssd	<i>Rms of successive R-R interval differences</i>
9	median_nni	<i>Median nearest neighbor index</i>
10	range_nni	<i>Range nearest neighbor index</i>
11	cvsd	<i>The variation coefficient of successive differences</i>
12	cvnni	<i>The variation coefficient nearest neighbor</i>
13	mean_hr	<i>Mean heart rate</i>
14	max_hr	<i>Maximum heart rate</i>
15	min_hr	<i>Minimum heart rate</i>
16	std_hr	<i>Standard deviation heart rate</i>

## 2.6 Fitur Seleksi

Ada 16 fitur domain waktu yang diekstrak dari HRV pada setiap segmen sinyal ECG. Banyaknya fitur yang diekstrak ini mengakibatkan semakin kompleksnya model yang berpotensi menyebabkan *overfitting*, dimana pada saat proses pelatihan dapat mencapai akurasi

maksimum tetapi pada saat pengujian menghasilkan nilai akurasi yang relatif rendah jauh dari nilai akurasi pelatihan (**Peker, dkk, 2015**) (**Liu & Tang, 2014**). Dari total 16 fitur dapat ditentukan fitur mana saja yang memberikan kontribusi besar terhadap kinerja sistem dan fitur mana saja yang berkontribusi kecil. Besar kecilnya kontribusi setiap fitur dapat dilihat dari besar kecilnya nilai *information gain* (*gain* informasi). Semakin besar nilai *gain* informasi, semakin besar pula kontribusi fitur tersebut terhadap kinerja algoritma (**Lei, 2012**).

### 2.7 Random Forest

*Random Forest* (RF) merupakan salah satu algoritma klasifikasi dan regresi dalam *Machine Learning* (ML) yang paling banyak digunakan (**Shi & Horvath, 2006**). Algoritma RF merupakan kumpulan dari algoritma *decision tree* (pohon keputusan) dengan memperbaiki kelemahan yang ada pada algoritma pohon keputusan. Dalam implementasinya, RF menerapkan metode *bagging* untuk mencegah terjadinya *overfitting* karena kompleksnya model (**Raschka, 2015**).

### 2.8 Kinerja Akurasi

Adapun kinerja akurasi tiap subjek pada penelitian ini dihitung berdasarkan jumlah sampel data yang berhasil diprediksi dengan benar terhadap jumlah total seluruh data yang digunakan dan dirumuskan pada Persamaan (2) sebagai berikut:

$$Akurasi_i (\%) = \frac{JDB_i}{TSD_i} \times 100\% \quad (2)$$

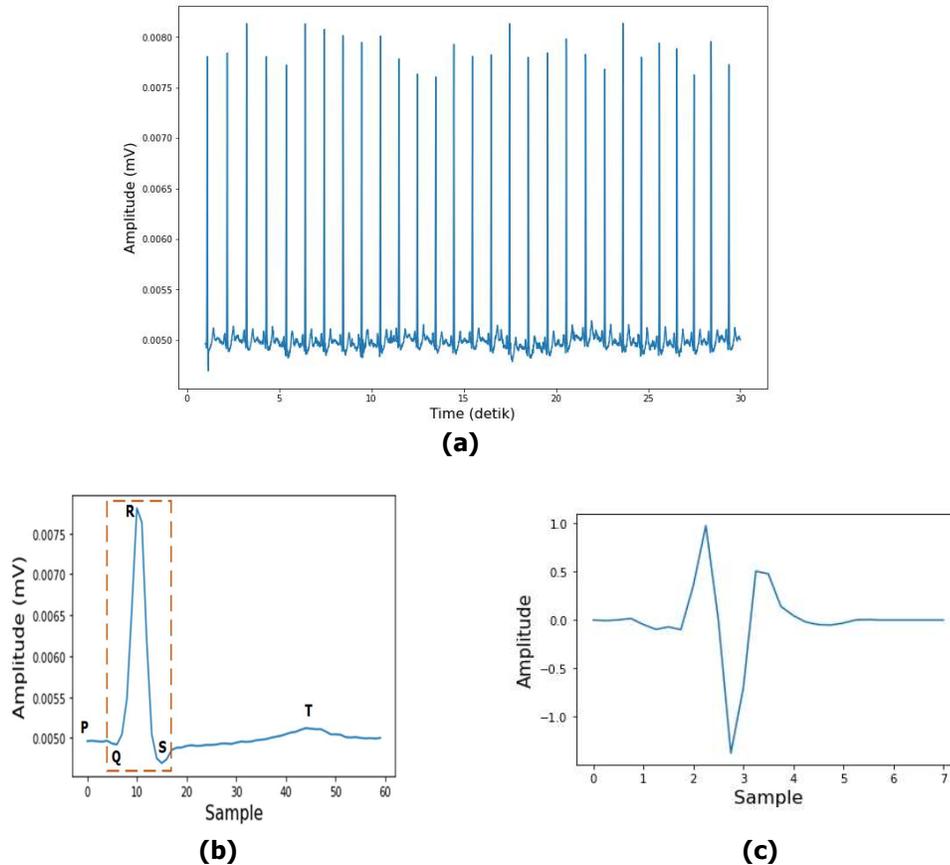
Dimana *i* menunjukkan tiap subjek data, *JDB* merupakan jumlah data yang diprediksi benar, sedangkan *TSD* menunjukkan jumlah total seluruh data yang digunakan dalam implementasi. Sedangkan untuk menghitung akurasi rata-rata dari sepuluh subjek dapat digunakan Persamaan (3) sebagai berikut:

$$Akurasi (\%) = \frac{\sum_{i=1}^{10} Akurasi_i}{total\ subjek} \times 100\% \quad (3)$$

## 3. HASIL DAN PEMBAHASAN

### 3.1 Dekomposisi dan Rekonstruksi Sinyal ECG menggunakan *Wavelet*

Gambar 5 menunjukkan satu segmen sinyal ECG dan satu bagian sinyal QRS pada satu siklus sinyal ECG. Sinyal ECG yang sudah disegmentasi pada frekuensi *sampling* 100 Hz, seperti terlihat pada Gambar 5(A), selanjutnya didekomposisi menggunakan *wavelet* tipe *symlet* 4, Gambar 5(C), untuk mendapatkan komponen sinyal yang diperlukan. Pemilihan tipe *wavelet* ini didasarkan pada bentuk sinyal ECG yang mirip dengan tipe kernel *wavelet* ini. Pada Gambar 5(B) menunjukkan bagian dari sinyal ECG (QRS) yang mirip dengan tipe kernel *symlet* 4 pada *wavelet*. Dari hasil dekomposisi sinyal ECG diambil bagian sinyal yang dianggap paling merepresentasikan bentuk sinyal yang akan diproses pada bagaian *Heart Rate Variability* (HRV).

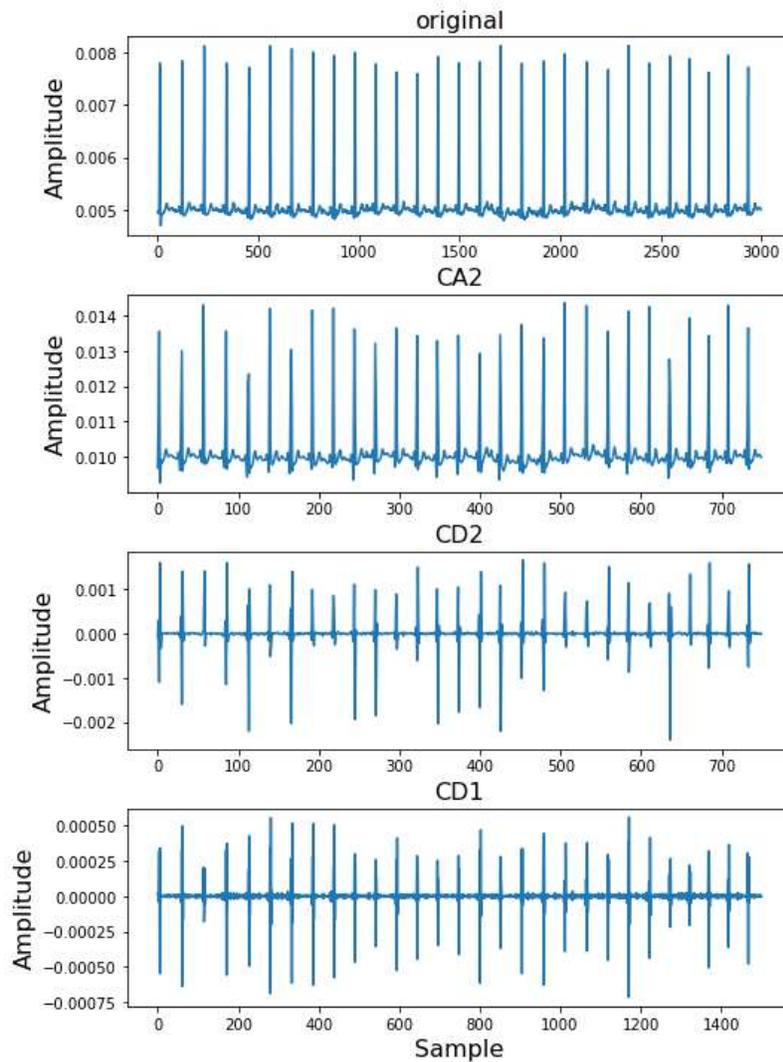


**Gambar 5. Potongan Sinyal ECG dari Salah Satu Subjek Implementasi. (A) Satu Segmen ECG dapat Terdiri dari Banyak Siklus ECG, (B). Satu Siklus ECG Terdiri dari Bagian PQRST, (C). Tipe *wavelet Symlet 4***

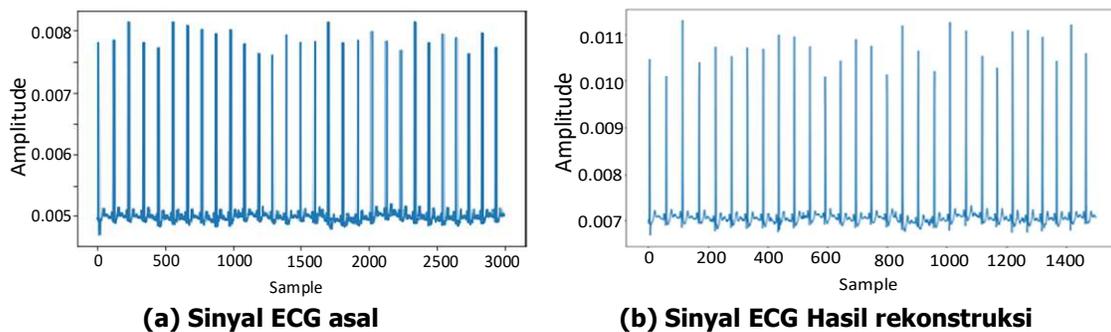
Gambar 6 menunjukkan sinyal ECG sebelum didekomposisi (*original*) dan sinyal ECG yang telah didekomposisi pada level dekomposisi dua, menghasilkan komponen aproksimasi (CA2) dan komponen detail (CD1 dan CD2).

Dari hasil dekomposisi sinyal ECG, kami hanya menggunakan dua komponen sinyal CA2 dan CD2 untuk direkonstruksi. Pemilihan dua komponen ini dilandaskan pada kebutuhan untuk perhitungan HRV dimana yang paling dominan adalah bagian *R-peak* sedangkan bagian lain dapat diabaikan.

Gambar 7 menunjukkan hasil rekonstruksi dari dua komponen *wavelet* CA2 dan CD2. Hasil rekonstruksi menunjukkan *R-peaks* yang sangat jelas antar setiap satu sinyal ECG dan sangat memudahkan proses perhitungan jarak antara *R-peaks* yang berdekatan. Hasil rekonstruksi juga menunjukkan jumlah sampel yang lebih sedikit dibandingkan sinyal asal, sehingga proses deteksi puncak relatif lebih cepat.



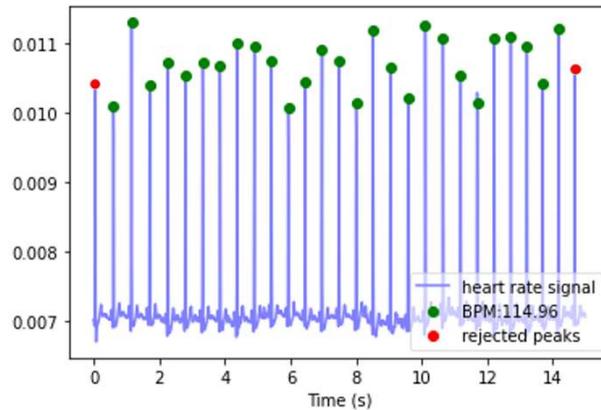
**Gambar 6. Dekomposisi Sinyal ECG. CA2: Komponen Aproksimasi Level 2, CD2: Komponen Detail Level 2, dan CD1: Komponen Detail Level 1**



**Gambar 7. Perbandingan Sinyal ECG Sebelum dan Setelah Rekonstruksi**

### 3.2 Heart Rate Variability Deteksi Puncak

Gambar 8 menunjukkan hasil deteksi puncak dari sinyal hasil rekonstruksi *wavelet*. Untuk menghindari kesalahan dalam perhitungan jarak antar puncak (*R-peak*) akibat proses segmentasi, Kami menghilangkan bagian awal dan akhir deteksi puncak.

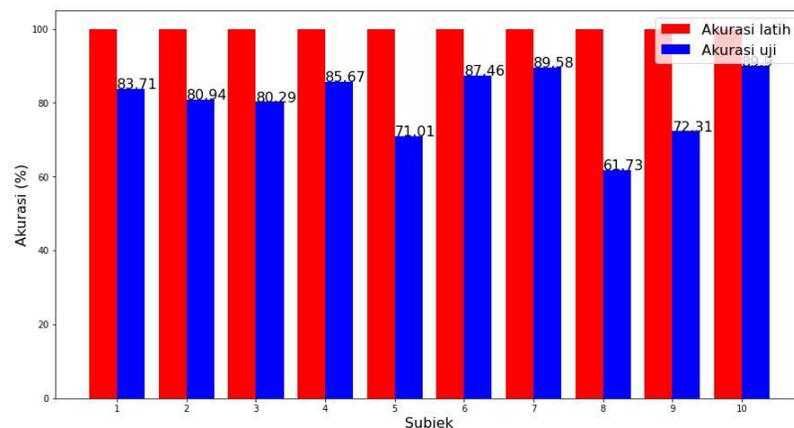


**Gambar 8. Deteksi Puncak Sinyal ECG dalam Domain Waktu**

Sekalipun amplitudo tiap puncak berbeda nilainya, namun algoritma deteksi puncak ini mampu mencari titik puncak (*R-peak*) setiap siklus sinyal ECG. HRV dihitung dari jarak antara dua titik puncak berdekatan yang selanjutnya jarak antar puncak ini diekstrak menggunakan fitur domain waktu.

### 3.3 Kinerja Klasifikasi Sinyal ECG

Gambar 9 menunjukkan kinerja klasifikasi dari data latih dan data uji yang berasal dari sepuluh subjek. Dalam penelitian ini digunakan *10-fold cross validation (one-subject-leave-out)*, dimana setiap satu subjek uji sembilan subjek lainnya bertindak sebagai data latih. Setiap subjek terdiri dari segmen data selama empat jam yang setiap segmennya memiliki kondisi tidur atau terjaga. Kemudian akurasi rata-rata dari setiap subjek dihitung dan terakhir dihitung akurasi rata-rata dari sepuluh subjek.

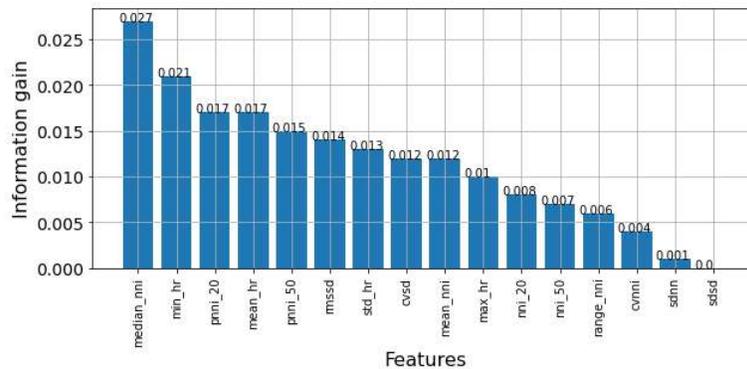


**Gambar 9. Kinerja Akurasi Rata-Rata Setiap Subjek**

Setiap segmen ECG setelah diekstrak fiturnya, selanjutnya diklasifikasikan ke dalam dua kondisi yaitu kondisi tidur dan terjaga dengan menggunakan algoritma *Random Forest*. Gambar 9 menunjukkan hasil akurasi dari sepuluh subjek yang digunakan dalam penelitian dengan akurasi data latih rata-rata  $100 \pm 0,0$  % dan akurasi data uji  $80,26 \pm 8,7$  %. Adapun akurasi rata-rata tertinggi terjadi pada subjek kesepuluh (89,9%) dan akurasi terendah pada subjek kedelapan (61,73%). Rendahnya akurasi pada subjek delapan dapat diakibatkan karena adanya gangguan pada sebagian data subjek saat akuisisi, baik karena *noise* yang dihasilkan organ lain atau pun gangguan tidur yang dapat menghasilkan fluktuasi sinyal ECG yang sedang diambil. Sekalipun nilai akurasi pada subjek delapan relatif lebih rendah, namun secara keseluruhan model sistem menunjukkan hasil yang baik.

**3.4 Fitur Seleksi Berdasarkan Gain Informasi**

Gambar 10 menunjukkan tingkat signifikansi setiap fitur domain waktu yang digunakan terhadap kinerja klasifikasi algoritma. Semakin tinggi nilai *gain* informasi, maka semakin tinggi pula pengaruhnya terhadap kinerja klasifikasi.



**Gambar 10. Distribusi Fitur Berdasarkan Tingkat Signifikansinya Terhadap Kinerja Klasifikasi Kondisi Tidur dan Terjaga**

Dalam penelitian ini ditunjukkan bahwa *'mean\_nni'* merupakan fitur yang sangat berkontribusi dibandingkan dengan fitur lainnya. Sebaliknya fitur *'sdsd'* merupakan fitur yang sangat tidak berkontribusi dalam proses implementasi, sehingga ketika fitur ini dibuang tidak akan terlalu berpengaruh terhadap hasil klasifikasi sebaliknya dapat mengurangi waktu komputasi karena berkurang satu parameter untuk diproses.

**3.5 Kinerja Klasifikasi Berdasarkan Jumlah Fitur**

Tabel 2 menunjukkan tingkat akurasi rata-rata berdasarkan jumlah fitur yang digunakan.

**Tabel 2. Akurasi Berdasarkan Jumlah Fitur**

No.	Fitur yang digunakan	Akurasi (%)	Waktu (menit)
1	<i>Median_nni</i>	73,40 ±13,09	15,03
2	<i>Median_nni, min_hr</i>	77,80 ±13,07	17,1
3	Semua fitur domain waktu (16 fitur)	80,26 ± 8,70	34,7
4	15 fitur domain waktu (tanpa fitur <i>sdsd</i> )	80,24 ± 8,72	30,2

Dapat dilihat bahwa pada kondisi awal sebelum menggunakan fitur seleksi *gain* informasi, akurasi rata-rata sebesar 80,26 % dengan standar deviasi  $\pm 8,70$  %. Setelah diterapkan fitur seleksi didapatkan bahwa fitur '*median\_nni*' merupakan fitur yang berkontribusi paling besar di antara fitur lainnya. Hanya dengan menggunakan fitur '*median\_nni*' dapat mencapai akurasi 73,40 %. Ketika ditambahkan satu fitur terbaik kedua, akurasi meningkat menjadi 77,80%. Sedangkan ketika satu fitur '*sdsd*' dihilangkan (15 fitur), hasil akurasi menunjukkan nilai yang relatif sama dengan nilai akurasi menggunakan 16 fitur domain waktu. Hal tersebut dapat diakibatkan karena pada analisis fitur seleksi dengan *gain* informasi, fitur '*sdsd*' menunjukkan kontribusi yang sangat kecil, ditunjukkan dengan kecilnya nilai *gain* informasinya. Selain itu, banyaknya penggunaan fitur sangat berpengaruh pada waktu komputasi algoritma dalam mengambil keputusan. Hasil menunjukkan bahwa penggunaan seluruh fitur (16 fitur) dengan kinerja akurasi 80,26 % memerlukan waktu komputasi selama 34,7 menit sedangkan dengan kinerja akurasi yang hampir sama (80,24%) dengan menggunakan 15 fitur membutuhkan waktu komputasi 30,2 menit. Lebih jauh lagi untuk kinerja akurasi 77,80 % dengan 2 fitur hanya dibutuhkan waktu komputasi 17,1 menit. Waktu komputasi ini hanya setengah dari waktu komputasi dengan penggunaan seluruh fitur, padahal selisih akurasinya hanya terpaut 2,46 % tapi dapat menekan waktu komputasi 50 % lebih cepat.

#### 4. KESIMPULAN

Fitur seleksi sangat diperlukan dalam memilih fitur yang tepat untuk digunakan dalam model *machine learning*. Tanpa fitur seleksi pemilihan fitur menjadi acak dan banyak yang dapat membuat model menjadi sangat kompleks dan pada akhirnya mengakibatkan *overfitting*. Selain dapat mengakibatkan *overfitting*, model yang kompleks juga memakan banyak waktu saat proses komputasi. Hasil akhir menunjukkan bahawa nilai akurasi dengan total 16 fitur domain waktu (80,26 %) tidak terpaut terlalu jauh dengan akurasi yang berasal dari dua fitur terbaik '*median\_nni*' dan '*min\_hr*' sebesar 77,80 %. Selain itu, penggunaan fitur seleksi menunjukkan bahwa pengurangan fitur dengan nilai *gain* informasi kecil (*sdsd*) tidak menurunkan nilai akurasi secara signifikan (80,24%). Dengan menggunakan fitur seleksi dapat melihat seberapa signifikan peran fitur-fitur yang digunakan terhadap kinerja algoritma sehingga pada akhirnya dapat diputuskan nilai akurasi yang optimal dengan *trade-off* terhadap waktu komputasi yang harus dicapai.

#### DAFTAR RUJUKAN

- Gautam, D. D., & Giri, V. K. (2016). Analysis of HRV signal for disease diagnosis. *2016 11th International Conference on Industrial and Information Systems (ICIIS), 2018-Janua*, (pp. 639–643). <https://doi.org/10.1109/ICIINFS.2016.8263017>
- Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-K., & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation, 101*(23), 1–2. <https://doi.org/10.1161/01.CIR.101.23.e215>
- Hafizhana, Y., Safitri, I., Novamizanti, L., & Ibrahim, N. (2020). Image Watermarking pada Citra Medis menggunakan Compressive Sensing berbasis Stationary Wavelet Transform.

- ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, 8(1), 43. <https://doi.org/10.26760/elkomika.v8i1.43>
- Hilton, M. L. (2004). Wavelet and wavelet packet compression of phonocardiograms. *Electronics Letters*, 40(17), 1040. <https://doi.org/10.1049/el:20045476>
- Islam, M. T., Rifa, S. R., & Kibria, M. G. (2020). Early Prediction of Heart Disease Using PCA and Hybrid Genetic Algorithm with k-Means. *2020 23rd International Conference on Computer and Information Technology (ICCIT)*, (pp. 1–6). <https://doi.org/10.1109/ICCIT51783.2020.9392655>
- Jabbar, M. A., & Samreen, S. (2016). Heart disease prediction system based on hidden naïve bayes classifier. *2016 International Conference on Circuits, Controls, Communications and Computing (I4C)*, (pp. 1–5). <https://doi.org/10.1109/CIMCA.2016.8053261>
- Jose, S. K., Shambharkar, C. M., & Chunkath, J. (2015). HRV analysis using ballistocardiogram with LabVIEW. *2015 International Conference on Computing and Communications Technologies (ICCT)*, (pp. 128–132). <https://doi.org/10.1109/ICCT2.2015.7292732>
- Jung, D. W., Hwang, S. H., Lee, Y. J., Jeong, D.-U., & Park, K. S. (2017). Apnea–Hypopnea Index Prediction Using Electrocardiogram Acquired During the Sleep-Onset Period. *IEEE Transactions on Biomedical Engineering*, 64(2), 295–301. <https://doi.org/10.1109/TBME.2016.2554138>
- Kansal, S., Bansod, P. P., & Kumar, A. (2015). Statistical Approach for Determination of ECG Markers. *2015 International Conference on Computational Intelligence and Communication Networks (CICN)*, (pp. 446–451). <https://doi.org/10.1109/CICN.2015.93>
- Ku, C.T., Wang, H.S., Hung, K.C. & Hung, Y.S. (2006). A Novel ECG Data Compression Method Based on Nonrecursive Discrete Periodized Wavelet Transform. *IEEE Transactions on Biomedical Engineering*, 53(12), 2577–2583. <https://doi.org/10.1109/TBME.2006.881772>
- Lei, S. (2012). A Feature Selection Method Based on Information Gain and Genetic Algorithm. *2012 International Conference on Computer Science and Electronics Engineering*, (pp. 355–358). <https://doi.org/10.1109/ICCSEE.2012.97>
- Liu, X., & Tang, J. (2014). Mass Classification in Mammograms Using Selected Geometry and Texture Features, and a New SVM-Based Feature Selection Method. *IEEE Systems Journal*, 8(3), 910–920. <https://doi.org/10.1109/JSYST.2013.2286539>
- Mendonca, F., Mostafa, S. S., Morgado-Dias, F., Julia-Serda, G., & Ravelo-Garcia, A. G. (2020). A Method for Sleep Quality Analysis Based on CNN Ensemble With Implementation in a Portable Wireless Device. *IEEE Access*, 8, 158523–158537. <https://doi.org/10.1109/ACCESS.2020.3019734>

- Peker, M., Arslan, A., Sen, B., Celebi, F. V., & But, A. (2015). A novel hybrid method for determining the depth of anesthesia level: Combining ReliefF feature selection and random forest algorithm (ReliefF&#x002B;RF). *2015 International Symposium on Innovations in Intelligent Systems and Applications (INISTA)*, (pp. 1–8). <https://doi.org/10.1109/INISTA.2015.7276737>
- Raj, A. A. S., Dheetsith, N., Nair, S. S., & Ghosh, D. (2014). Auto analysis of ECG signals using artificial neural network. *2014 International Conference on Science Engineering and Management Research (ICSEMR)*, (pp. 1–4). <https://doi.org/10.1109/ICSEMR.2014.7043597>
- Raschka, S. (2015). *Python Machine Learning*. Packt Publishing.
- Shaffer, F., & Ginsberg, J. P. (2017). An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health*, *5*, 1–17. <https://doi.org/10.3389/fpubh.2017.00258>
- Shi, T., & Horvath, S. (2006). Unsupervised Learning With Random Forest Predictors. *Journal of Computational and Graphical Statistics*, *15*(1), 118–138. <https://doi.org/10.1198/106186006X94072>
- Szypulska, M., & Piotrowski, Z. (2012). Prediction of fatigue and sleep onset using HRV analysis. *Proceedings of the 19th International Conference - Mixed Design of Integrated Circuits and Systems, MIXDES 2012*, (pp. 543–546).
- Vityazeva, T., Vityazev, S., & Mikheev, A. (2018). Synchronization of heart rate and respiratory signals for HRV analysis. *2018 7th Mediterranean Conference on Embedded Computing (MECO), June*, (pp. 1–4). <https://doi.org/10.1109/MECO.2018.8405989>
- Weeks, M. (2007). Digital Signal Processing Using matlab and Wavelets. In *Infinity Science Press LLC* (1st ed., Vol. 1).
- Zarei, A., & Asl, B. M. (2019). Automatic Detection of Obstructive Sleep Apnea Using Wavelet Transform and Entropy-Based Features From Single-Lead ECG Signal. *IEEE Journal of Biomedical and Health Informatics*, *23*(3), 1011–1021. <https://doi.org/10.1109/JBHI.2018.2842919>
- Zheng, Q., Chen, C., Li, Z., Huang, A., Jiao, B., Duan, X., & Xie, L. (2013). A novel multi-resolution SVM (MR-SVM) algorithm to detect ECG signal anomaly in WE-CARE project. *2013 ISSNIP Biosignals and Biorobotics Conference: Biosignals and Robotics for Better and Safer Living (BRC)*, (pp. 1–6). <https://doi.org/10.1109/BRC.2013.6487453>