

Pengembangan Sistem Deteksi Kantuk Menggunakan Pengklasifikasi *Random Forest* pada Sinyal Elektrokardiogram

(*Development of Drowsiness Detection System Using Random Forest Classifier on Electrocardiogram Signals*)

Nuryani Nuryani¹, Khoirun Nisak¹, Artono Dwijo Sutomo¹

Abstract—Drowsiness is one of the main causes of accidents in traffic, manufacturing industry, and other fields. A system that can detect sleepiness early is important to reduce the number of accidents due to drowsiness. Changes in autonomous nervous activity associated with drowsiness were analyzed using heart rate variability (HRV) obtained from ECG. The Random Forest classifier is known to perform well and is strong against overfitting. In this study, a drowsiness detection system was developed based on electrocardiogram (ECG) signals and Random Forest. The drowsiness detection system was trained using ECG recordings from DROZY database equipped with Karolinska Sleepiness Scale (KSS). The system input features were extracted based on time domain and frequency domain methods. Drowsiness levels were classified based on simplified KSS: drowsy and awake. Random Forest was validated using the Out-of-Bag (OOB) method. The effects of the number of estimators and max features variations were evaluated. The features were sorted based on their importance and combined for the system input with various ECG segmentation lengths. The best performance of the drowsiness detection system was with an average accuracy of 94.61%, sensitivity of 96.67%, and specificity of 91.67% from 40 s segmentations.

Intisari—Kantuk merupakan salah satu penyebab utama kecelakaan dalam lalu lintas, industri manufaktur, maupun pada bidang lain. Untuk itu, sistem yang dapat mendeteksi kantuk secara dini merupakan hal yang sangat penting dalam rangka mengurangi angka kecelakaan akibat kantuk. Kantuk dapat dianalisis melalui *Heart Rate Variability* (HRV) dari sinyal EKG yang menunjukkan perubahan aktivitas saraf otonom. Pengklasifikasi *Random Forest* diketahui berkinerja sangat baik serta kuat terhadap *overfitting*. Oleh karena itu, pada makalah ini dikembangkan sistem deteksi kantuk menggunakan sinyal elektrokardiogram (EKG) dan *Random Forest*. Sistem deteksi kantuk dilatih menggunakan rekaman EKG dari database DROZY yang dilengkapi *Karolinska Sleepiness Scale* (KSS). Fitur masukan sistem diekstraksi berdasarkan metode ranah waktu dan ranah frekuensi. Tingkat kantuk diklasifikasikan berdasarkan KSS yang disederhanakan menjadi dua kelas, yaitu kantuk dan terjaga. *Random Forest* divalidasi dengan metode *Out-of-Bag* (OOB). Efek dari variasi jumlah *estimator* dan *max feature* terhadap kinerja sistem dievaluasi. Fitur diurutkan berdasarkan kepentingannya dan dikombinasikan sebagai masukan sistem dengan berbagai panjang segmentasi EKG. Kinerja terbaik sistem deteksi kantuk yaitu rata-rata akurasi 94,61%, sensitivitas

96,67%, dan *specificity* 91,67%, yang diperoleh dengan segmentasi 40 detik.

Kata Kunci--Kantuk, Elektrokardiogram, KSS, *Random Forest*.

I. PENDAHULUAN

Kematian dan luka akibat kecelakaan lalu lintas menjadi masalah yang serius secara global. Menurut data *Global Status Report on Road Safety 2018*, angka kematian akibat kecelakaan lalu lintas terus bertambah, mencapai 1,35 juta jiwa pada tahun 2016 [1]. Kantuk merupakan salah satu penyebab meningkatnya risiko kecelakaan lalu lintas. AAA Foundation for Traffic Safety telah melakukan survei pada lebih dari 14.000 kecelakaan dari tahun 2009 hingga 2013 dan memperkirakan bahwa kantuk terlibat dalam 21% kecelakaan fatal di Amerika [2].

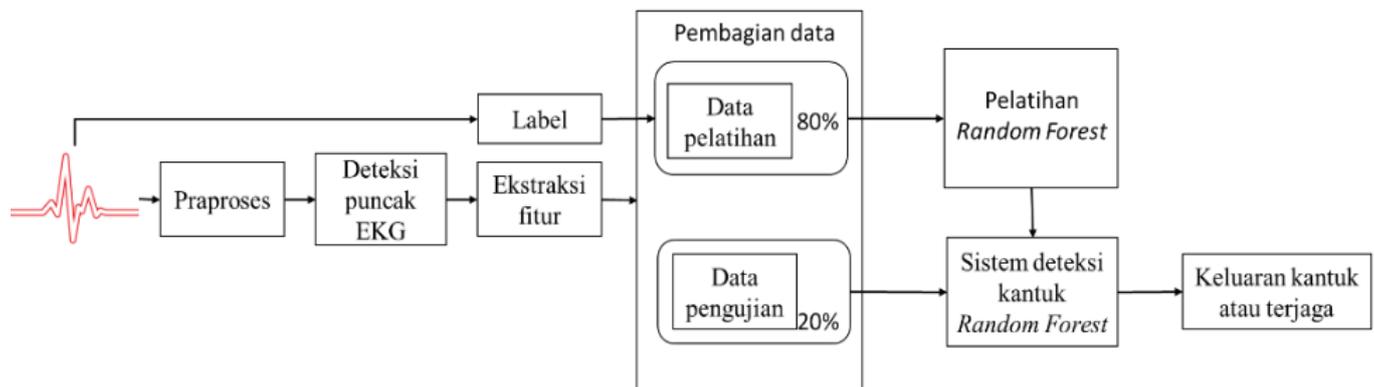
Kantuk didefinisikan sebagai keadaan ketika kesadaran terganggu karena kecenderungan untuk tidur [3]. Kinerja mengemudi yang tidak stabil sering terjadi pada pengemudi dalam keadaan mengantuk [2]. Tingkat kantuk dapat dikelompokkan berdasarkan *Karolinska Sleepiness Scale* (KSS) [4]. Dengan skala KSS, subjek berada pada tingkatan yang mencerminkan kondisi psikofisik yang dirasakan [5].

Keadaan kantuk dan terjaga berkorelasi dengan aktivitas saraf otonom simpatetik dan parasimpatetik tubuh. Keadaan terjaga ditunjukkan dengan meningkatnya aktivitas simpatetik dan/atau menurunnya aktivitas parasimpatetik. Aktivitas saraf otonom dapat dianalisis menggunakan *Heart Rate Variability* (HRV) dari sinyal elektrokardiogram (EKG) [6]. Analisis HRV tersebut dapat dilakukan, baik pada ranah waktu maupun ranah frekuensi [7]. EKG merupakan rekaman dari aktivitas listrik yang dihasilkan oleh sel jantung yang mencapai permukaan tubuh [8].

Beberapa penelitian mengenai deteksi kantuk menggunakan sinyal EKG telah dilakukan sebelumnya, di antaranya menggunakan analisis diskriminan linier [6], jaringan Saraf Tiruan (JST) dengan *Radial Basis Function* (RBF) [9], dan JST-RBF dengan optimasi menggunakan *Particle Swarm Optimization* (PSO) [10].

Pada makalah ini dipaparkan sistem deteksi kantuk menggunakan *Random Forest*. Dipilihnya *Random Forest* pada sistem deteksi ini adalah karena keunggulan-keunggulan yang dimiliki oleh *Random Forest*, di antaranya unggul dalam penanganan *overfitting*. *Overfitting* merupakan masalah yang banyak terjadi pada *machine learning*. Di samping itu, *Random Forest* juga belum banyak dieksplorasi untuk sistem deteksi kantuk, khususnya yang menggunakan EKG. *Random Forest*

¹ Program Studi Fisika, F. MIPA, Universitas Sebelas Maret, Jl. Ir. Sutami 36A, Surakarta, INDONESIA (Tlp: 0271-669017; e-mail: nuryani@mipa.uns.ac.id)



Gbr. 1 Rancangan pengembangan sistem deteksi kantung.

tersusun dari sejumlah *decision tree*. *Random Forest* diketahui berkinerja sangat baik dibandingkan dengan banyak pengklasifikasi lainnya serta kuat terhadap *overfitting* [11]. Pada makalah ini digunakan fitur EKG pada ranah waktu dan ranah frekuensi.

II. METODOLOGI

Metode untuk deteksi kantung menggunakan EKG dan *Random Forest* ditunjukkan pada Gbr. 1, yang menjelaskan proses pengembangan sistem deteksi kantung. Dalam metode ini ada beberapa tahap, yaitu praproses, ekstraksi fitur, pelatihan *Random Forest*, dan pengujian sistem.

A. Persiapan Data

Data rekaman EKG dari partisipan kantung/terjaga terdiri atas 34 rekaman. Tiap rekaman memiliki durasi 600 detik dengan *sampling rate* 512 Hz. Rekaman EKG dilengkapi dengan tingkat kantung berdasarkan KSS. Pada makalah ini, tingkat kantung disederhanakan menjadi dua tingkat, yaitu keadaan kantung dengan skala KSS 1-5 dan keadaan terjaga dengan skala KSS 6-9. Deskripsi skala 5 pada KSS yaitu “tidak terjaga dan tidak kantung” sedangkan pada skala 6 KSS yaitu “muncul tanda mengantuk” sehingga skala 5 dan 6 dapat dijadikan batasan antara kelas terjaga dan kantung. Tingkat kantung digunakan sebagai target keluaran atau label pada pelatihan *Random Forest*. Data EKG yang digunakan adalah seperti yang digunakan pada penelitian sebelumnya [12], yaitu *ULg Multimodality Drowsiness Database* atau dinamai dengan *database DROZY*. Rincian rekaman EKG yang digunakan pada makalah ini disajikan pada Tabel I, yang menunjukkan bahwa diperoleh dua puluh rekaman kelas kantung dan empat belas rekaman kelas terjaga.

B. Praproses

Tahap praproses pengolahan EKG adalah tahap pertama sebelum tahap ekstraksi fitur. Pada tahap praproses dilakukan segmentasi, filtrasi, deteksi *R-peaks*, dan analisis spektrum dengan metode Welch. Tiga macam panjang segmen diteliti, yaitu 20 detik, 30 detik, dan 40 detik. Jumlah segmen adalah 1.020 segmen untuk segmentasi 20 detik, 680 segmen pada segmentasi 30 detik, dan 510 segmen pada segmentasi 40 detik. Proses filtrasi dilakukan untuk mengurangi *noise* yang sering terjadi pada data mentah EKG. Filter yang digunakan adalah

TABEL I
RINCIAN REKAMAN EKG YANG DIGUNAKAN

Rekaman	KSS	Label	Rekaman	KSS	Label
1-1.edf	3	0	7-3.edf	9	1
1-2.edf	6	1	8-1.edf	2	0
1-3.edf	7	1	8-2.edf	6	1
2-2.edf	7	1	8-3.edf	8	1
2-3.edf	6	1	9-2.edf	6	1
3-1.edf	2	0	9-3.edf	8	1
3-2.edf	3	0	10-1.edf	3	0
3-3.edf	4	0	10-3.edf	7	1
4-1.edf	4	0	11-1.edf	4	0
4-2.edf	8	1	11-2.edf	7	1
4-3.edf	9	1	11-3.edf	7	1
5-1.edf	3	0	12-1.edf	2	0
5-2.edf	7	1	13-1.edf	6	1
5-3.edf	8	1	13-2.edf	3	0
6-2.edf	3	0	14-1.edf	5	0
6-3.edf	7	1	14-2.edf	7	1
7-2.edf	4	0	14-3.edf	8	1

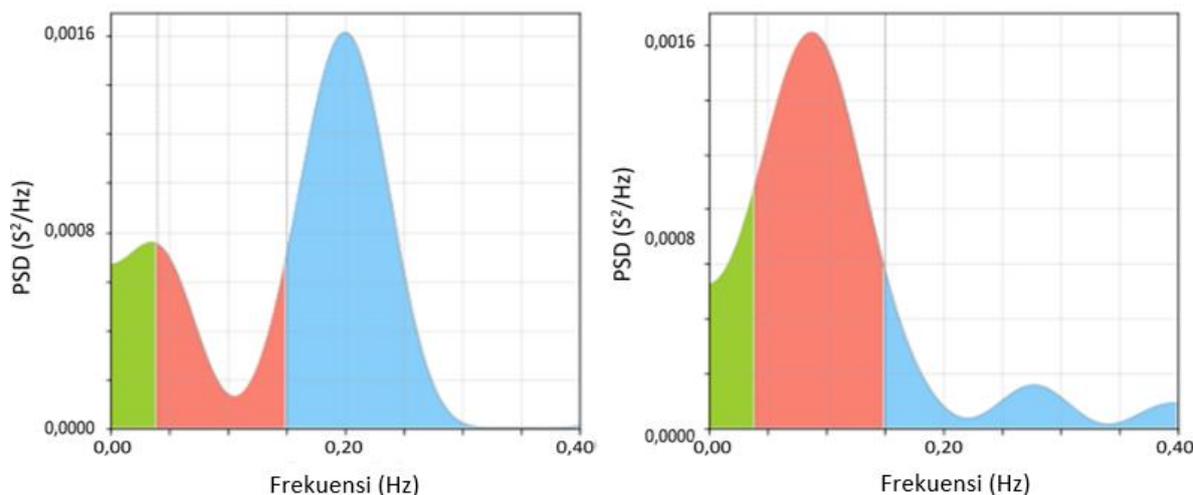
Keterangan label: 0 = terjaga
1 = kantung

filter *Finite Impulse Response* (FIR). Selanjutnya, dicari *R-peaks* EKG (puncak gelombang QRS) menggunakan pendekatan Hamilton [13]. *R-peaks* yang diperoleh digunakan sebagai dasar untuk menghitung *Normal-to-Normal Interval* (NNI), *Heart Rate* (HR) dan *Power Spectral Density* (PSD). NNI merupakan jarak antara dua *R-peaks* yang berurutan, sedangkan PSD dihitung dari NNI menggunakan metode Welch. Pita frekuensi dibagi pada batasan *Very Low Frequency* (VLF) (0,00 Hz – 0,04 Hz), *Low Frequency* (LF) (0,04 Hz – 0,15 Hz), dan *High-Frequency* (HF) (0,15 Hz – 0,40 Hz).

C. Ekstraksi Fitur

Perbedaan panjang NNI pada EKG terjaga dan kantung menjadi dasar untuk ekstraksi fitur. Pita LF berkorelasi dengan aktivitas simpatetik dan pita HF berkorelasi dengan aktivitas parasimpatetik pada analisis HRV [6], sehingga analisis ranah frekuensi ini digunakan sebagai fitur.

PSD atau *spectral power* mewakili kekuatan sinyal masukan pada suatu rentang frekuensi. Perhitungan PSD berdasarkan



Gbr. 2 Grafik PSD rekaman, (a) kelas kantuk, (b) terjaga.

TABEL II
FITUR RANAH FREKUENSI

Fitur Ranah Frekuensi	Pita Frekuensi	Satuan
Puncak frekuensi	VLF, LF, dan HF	Hz
Daya absolut	VLF, LF, dan HF	ms ²
Daya relatif	VLF, LF, dan HF	ms ²
Daya logaritmik	VLF, LF, dan HF	log
Daya ternormalisasi	LF, dan HF	-
Total daya semua pita frekuensi	-	ms ²

metode Welch melibatkan pembagian sinyal menjadi beberapa segmen, mengambil *periodogram* yang dimodifikasi dari segmen ini, dan dirata-rata [14]. Parameter ranah frekuensi dihitung berdasarkan pita frekuensi dengan rentang standar untuk pita VLF (0,00 Hz – 0,04 Hz), LF (0,04 Hz – 0,15 Hz), dan HF (0,15 Hz – 0,40 Hz) [15]. Dari proses ekstraksi fitur, seperti pada Tabel II, diperoleh enam belas fitur ranah frekuensi berdasarkan penelitian sebelumnya [15]. Gbr. 2 menunjukkan hasil ekstraksi metode ranah frekuensi. Terlihat dalam gambar bahwa rekaman kelas kantuk mempunyai komponen HF yang lebih tinggi, sedangkan rekaman kelas terjaga mempunyai komponen LF yang lebih tinggi.

Parameter statistik diturunkan dari seri NNI, seri perbedaan NNI berurutan (Δ NNI), dan seri HR [15]. NNI pada rekaman kelas kantuk terlihat lebih panjang daripada NNI pada rekaman kelas terjaga, seperti ditunjukkan pada Gbr. 3. Tabel III menunjukkan parameter statistik NNI dan detak jantung yang merupakan parameter EKG ranah waktu, yang terdiri atas delapan belas fitur.

Pada semua fitur diberlakukan normalisasi menggunakan *minmax scaller* sehingga fitur berada pada skala antara 0 dan 1 berdasarkan (1). Normalisasi bertujuan untuk menyamakan standar fitur dalam rentang tertentu. Fitur harus disamakan pada skala karena satuan dan amplitudo fitur-fitur yang digunakan dalam klasifikasi sangat berbeda antara satu dengan yang lain [16].

$$f_{i,N} = \frac{f_i - \min(f_i)}{\max(f_i) - \min(f_i)} \quad (1)$$

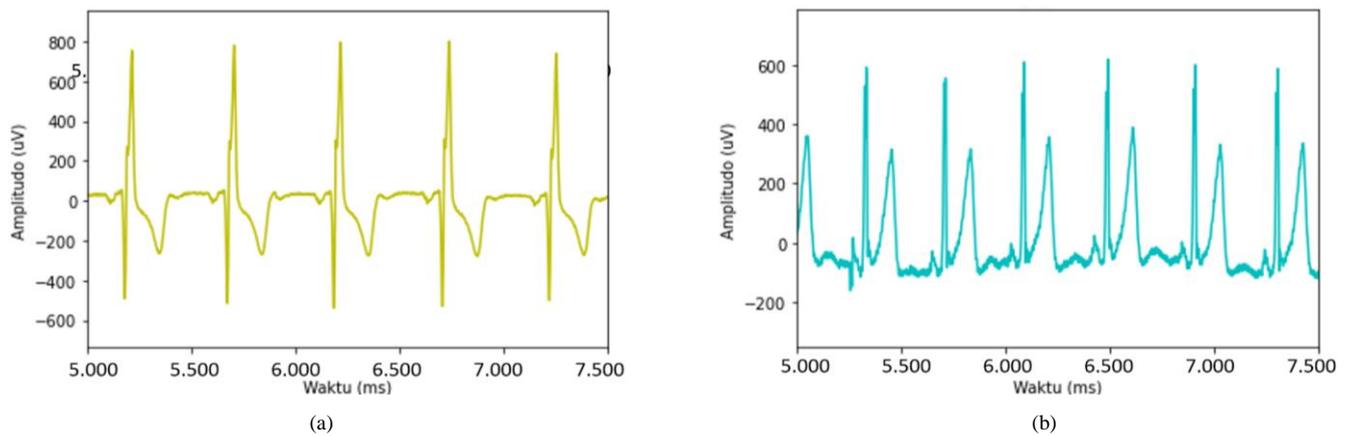
TABEL III
FITUR STATISTIK PARAMETER EKG RANAH WAKTU

Fitur Ranah Waktu	Keterangan	Satuan
NNI <i>counter</i>	jumlah NNI	-
NNI <i>mean</i>	<i>mean</i> NNI	ms
NNI <i>min</i>	NNI minimal	ms
NNI <i>max</i>	NNI maksimal	ms
NNI <i>diff mean</i>	<i>mean</i> Δ NNI	ms
NNI <i>diff min</i>	Δ NNI minimal	ms
NNI <i>diff max</i>	Δ NNI maksimal	ms
HR <i>mean</i>	<i>mean</i> detak jantung	bpm
HR <i>min</i>	detak jantung minimal	bpm
HR <i>max</i>	detak jantung maksimal	bpm
HR <i>std</i>	standar deviasi seri detak jantung	bpm
SDNN	standar deviasi NNI	ms
RMSSD	akar rata-rata dari Δ NNI kuadrat	ms
SDSD	standar deviasi Δ NNI	ms
NN50	jumlah Δ NNI yang lebih besar dari 50 ms	-
pNN50	rasio antara NN50 dengan jumlah total NNI	-
NN20	jumlah Δ NNI yang lebih besar dari 20 ms	-
pNN20	rasio antara NN20 dengan jumlah total NNI	-

dengan $f_{i,N}$ adalah fitur ke- i yang telah dinormalisasi, f_i adalah fitur ke- i , sedangkan $\min(f_i)$ dan $\max(f_i)$ adalah nilai minimal dan maksimal fitur ke- i [16].

D. Pelatihan dan Pengujian Random Forest

Random Forest adalah kumpulan atau ansambel dari *Classification and Regression Trees* (CART) yang dilatih pada set data dengan ukuran yang sama dengan set pelatihan, yang disebut *bootstrap* dan dibuat dari pengambilan sampel ulang secara acak pada set pelatihan itu sendiri [17]. Pelatihan *Random Forest* dilakukan dengan menggunakan 80% data dari setiap variasi segmen waktu. Sampel data pelatihan diambil dari 80% segmen secara acak dari setiap rekaman, sehingga



Gbr. 3 Grafik NNI rekaman, (a) kelas kantuk, (b) terjaga.

data pelatihan dan pengujian mempunyai rasio kelas terjaga dan kantuk yang sama. Tahap pelatihan terdiri atas tahap-tahap sebagai berikut [11].

1. Pembuatan pohon sejumlah N estimator dari data pelatihan.
2. Penumbuhan pohon klasifikasi yang tidak dipotong pada setiap sampel *bootstrap*. Pada setiap nodus, daripada memilih pemisahan terbaik di antara semua prediktor, diambil sampel m_{try} acak dari prediktor dan dipilih pemisah terbaik di antara variabel tersebut.
3. Prediksi data baru dengan mengumpulkan prediksi dari semua pohon (yaitu suara terbanyak untuk klasifikasi).

Random Forest divalidasi menggunakan akurasi *Out-of-Bag* (OOB), yaitu data yang tidak termasuk dalam sampel *bootstrap*. Pengujian *Random Forest* dilakukan dengan data pengujian sejumlah 20% dari data keseluruhan yang tidak digunakan dalam proses pelatihan.

Random forest dapat digunakan untuk menentukan peringkat fitur berdasarkan pengaruhnya dalam klasifikasi. Kepentingan fitur dinilai oleh indeks kriteria *Gini impurity*. Indeks Gini adalah ukuran kekuatan prediksi variabel pada regresi atau klasifikasi. Gini yang rendah memiliki arti sebuah fitur mempunyai peran yang bagus dalam memisahkan data ke dalam dua kelas. Oleh karena itu, indeks Gini dapat digunakan untuk membuat peringkat kepentingan fitur pada klasifikasi [17].

Kinerja sistem deteksi ini ditentukan menggunakan tiga parameter kinerja, yaitu akurasi, sensitivitas, dan *specificity*. Akurasi menunjukkan kedekatan antara nilai prediksi dengan nilai aktual, yaitu tingkat kesuksesan *Random Forest* mengklasifikasi kelas terjaga dan kantuk. Sensitivitas menunjukkan *True Positive Rate* (TPR) atau tingkat kesuksesan *Random Forest* mendeteksi kelas kantuk. *Specificity* menunjukkan *False Positive Rate* (FPR) atau tingkat kesuksesan *Random Forest* mendeteksi kelas terjaga. Akurasi diperoleh dari (1), sensitivitas diperoleh dari (2), dan *specificity* diperoleh dari (3).

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

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

TABEL IV
KINERJA SISTEM DETEKSI DENGAN VARIASI JUMLAH ESTIMATOR PADA RANDOM FOREST

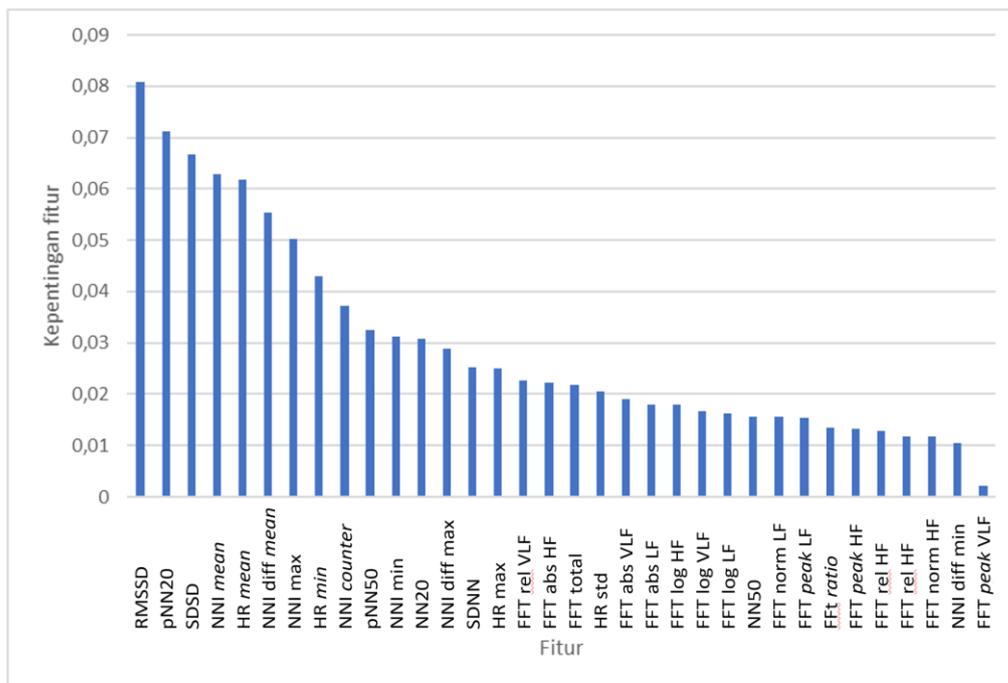
Jumlah Estimator	Akurasi (%)	Sensitivitas (%)	Specificity (%)	OOB (%)
10	84,74	88,91	78,79	70,40
20	85,20	91,09	76,79	75,55
30	86,21	93,59	75,67	76,84
40	86,76	93,75	76,79	78,49
50	86,76	93,75	76,79	77,94
60	85,66	92,50	75,89	79,23
70	85,29	91,25	76,79	78,68
80	85,66	91,25	77,68	79,04
90	86,03	91,88	77,68	80,33
100	86,03	91,25	78,57	79,04

$$Specificity = \frac{TN}{FP+TN} \times 100\% \quad (4)$$

dengan *TP* adalah *true positive*, yaitu jumlah data EKG kantuk terdeteksi benar; *FP* adalah *false positive*, yaitu jumlah data EKG kantuk terdeteksi salah; *FN* adalah *false negative*, yaitu jumlah data EKG terjaga terdeteksi salah; dan *TN* adalah *true negative*, yaitu jumlah data EKG terjaga terdeteksi benar [18].

III. HASIL DAN PEMBAHASAN

Pada makalah ini dipaparkan sistem deteksi kantuk menggunakan elektrokardiogram dan algoritme *Random Forest*. Prinsip dari algoritme *Random Forest* adalah ansambel yang terdiri atas sejumlah sistem klasifikasi individu, yaitu *decision tree* yang masing-masing menentukan hasil klasifikasi individu, dengan hasil akhir dari model ansambel *Random Forest* ditentukan oleh suara terbanyak dari masing masing *decision tree*. Struktur *Random Forest* mempunyai parameter yang perlu ditentukan sebelum dilakukannya pelatihan. Pada makalah ini, dilakukan variasi jumlah estimator (jumlah *decision tree* yang digunakan), seperti pada Tabel IV, serta variasi *max feature* (jumlah maksimal fitur acak yang dipertimbangkan di setiap nodus), seperti pada Tabel V, untuk menentukan parameter terbaik. Kualitas percabangan pada



Gbr. 4 Grafik kepentingan fitur.

TABEL V
HASIL VARIASI MAX FEATURES PADA RANDOM FOREST

Max Feature	Akurasi (%)	Sensitivitas (%)	Specificity (%)	OOB (%)
0,2	86,76	90,00	82,14	80,70
0,3	87,87	93,13	80,36	79,60
0,4	86,40	91,88	78,57	79,78
0,5	87,50	91,25	82,14	78,31

pohon ditentukan dengan *Gini impurity*. Parameter lain seperti urutan kepentingan fitur serta kombinasi gabungan fitur ditentukan dalam makalah ini. Kinerja model dinilai menggunakan parameter akurasi, sensitivitas, *specificity*, dan akurasi OOB. Akurasi, sensitivitas, dan *specificity* yang digunakan merupakan rata-rata dari pelatihan dan pengujian, sedangkan akurasi OOB diperoleh dari hasil pengujian.

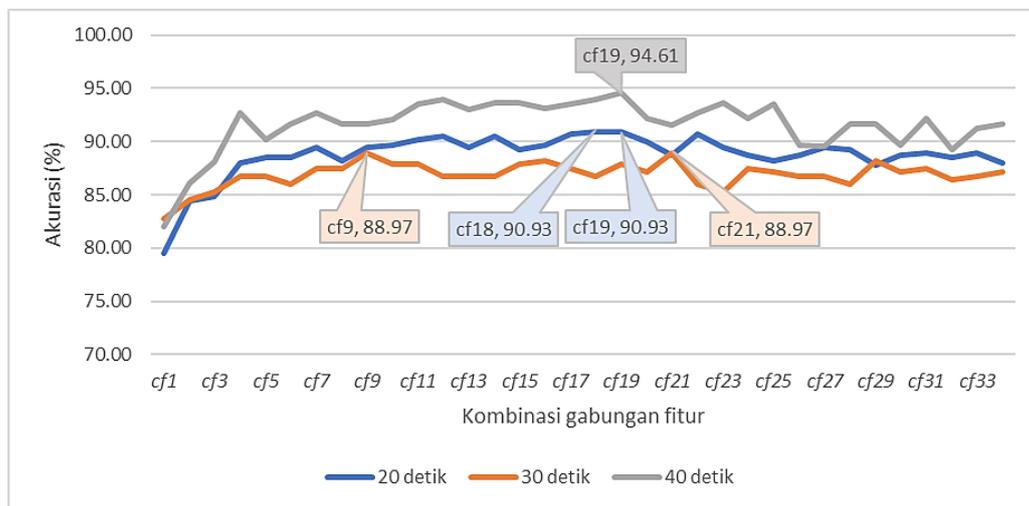
Jumlah *estimator* menunjukkan banyaknya pengklasifikasi pohon (*decision tree*) yang digunakan untuk membangun ansambel *Random Forest*. Variasi ini bertujuan untuk menemukan jumlah *estimator* yang tepat untuk sistem deteksi kantuk. Hasil variasi jumlah *estimator* disajikan pada Tabel IV. Kinerja terbaik rata-rata akurasi, sensitivitas, dan *specificity* diperoleh pada jumlah *estimator* 40 dan 50 dengan nilai akurasi, sensitivitas, dan *specificity* berturut-turut sebesar 86,76%, 93,75%, dan 76,79%. Dari hasil tersebut, terlihat bahwa semakin banyak pohon yang digunakan, semakin baik kinerja *Random Forest*. Namun, setelah variasi jumlah *estimator* 50, tidak terdapat kenaikan kinerja.

Nilai *max features* menunjukkan banyaknya nilai maksimal fitur acak yang dapat diambil pada setiap titik pemisah setiap individu pohon *Random Forest*. Dari Tabel V diperoleh kinerja terbaik pada akurasi, sensitivitas, dan *specificity* pada nilai 0,3,

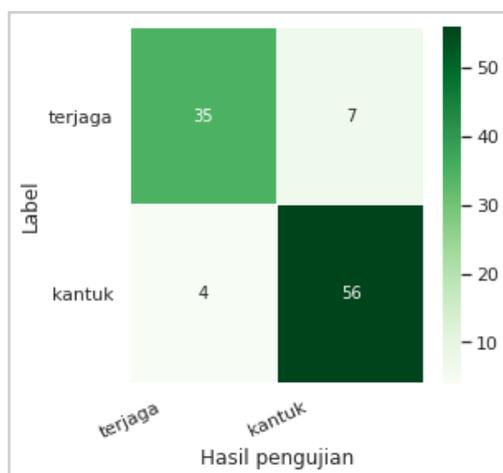
dengan nilai akurasi, sensitivitas, dan *specificity* berturut-turut 87,87%, 93,13%, dan 80,36%. *Max features* 0,3 mengandung arti bahwa setiap individu pohon hanya dapat mempertimbangkan maksimal 30% fitur pada setiap nodus untuk dijadikan pemisah.

Urutan kepentingan fitur bertujuan untuk menilai kontribusi setiap fitur dalam menentukan hasil klasifikasi. Semua fitur ranah waktu dan ranah frekuensi dijadikan masukan *Random Forest*. Jumlah *estimator* atau pohon yang digunakan sebanyak lima puluh dan *max feature* 0,3. Pada Gbr. 4 ditunjukkan fitur yang berperan paling penting sampai paling tidak penting dalam menentukan hasil deteksi. Nilai kepentingan fitur didasarkan pada tingkat kesuksesan fitur tersebut mengurangi *Gini impurity*. Pengurutan fitur-fitur berdasarkan kepentingannya bermanfaat untuk menentukan fitur yang lebih efisien dalam menjalankan klasifikasi. Dari eksperimen diketahui bahwa fitur yang mempunyai tingkat kepentingan paling tinggi adalah RMSSD dan fitur yang mempunyai tingkat kepentingan paling rendah adalah FFT *peak* VLF. Tingkat kepentingan fitur lainnya berada di antara kedua fitur tersebut, seperti ditunjukkan pada Gbr. 4.

Akurasi sistem deteksi yang menggunakan variasi kombinasi fitur dan panjang segmen yang berbeda ditunjukkan pada Gbr. 5. Kombinasi ini dilakukan berdasarkan urutan kepentingan fitur seperti yang ditunjukkan Gbr. 4. Kombinasi *cf1* terdiri atas satu fitur, yaitu fitur dengan kepentingan tertinggi, RMSSD. Kombinasi *cf2* terdiri atas dua fitur, yaitu fitur dengan kepentingan tertinggi (RMSSD) dan fitur dengan kepentingan tertinggi kedua (pNN20). Dengan cara yang sama, *cf3* terdiri atas tiga fitur dengan kepentingan tertinggi pertama sampai ketiga, yaitu RMSSD, pNN20, dan SDDSD. Kombinasi *cf4* sampai dengan *cf33* disusun dengan cara yang serupa.



Gbr. 5 Akurasi sistem deteksi pada kombinasi fitur dan panjang segmen.



Gbr. 6 Grafik confusion matrix model *Random Forest* terbaik.

Dengan demikian, *cf33* merupakan kombinasi dari semua fitur. Akurasi dengan panjang segmen (20, 30, dan 40 detik) pada semua kombinasi tersebut juga dipaparkan pada Gbr. 5.

Seperti ditunjukkan pada Gbr. 5, terlihat bahwa dengan menggabungkan fitur dengan urutan kepentingan tertinggi, dihasilkan kenaikan akurasi. Namun, akurasi mengalami penurunan pada tingkat tertentu. Akurasi terbaik untuk panjang segmen 20 detik adalah 90,93%, pada kombinasi fitur *cf18* dan *cf19*. Akurasi terbaik pada segmen 30 detik adalah 88,97% pada *cf9* dan *cf21*, sedangkan akurasi terbaik pada segmen 40 detik adalah 94,61% pada *cf19*. Dari hasil kombinasi gabungan fitur dan panjang segmentasi, diperoleh sistem deteksi kantuk dengan kinerja terbaik adalah pada *cf19*, yang terdiri atas sembilan belas fitur dan segmentasi 40 detik. Kinerja sistem deteksi keadaan ini adalah akurasi 94,61%, sensitivitas 96,67%, dan *specificity* 91,67%. Akurasi menunjukkan kemampuan model dalam mendeteksi kelas terjaga dan kantuk dengan benar. Dengan demikian, model dengan akurasi terbaik adalah kombinasi gabungan fitur *cf19* dan segmentasi 40.

Pada panjang segmen 40 detik terdapat 510 segmen data, sehingga didapatkan data pelatihan 408 segmen dan data pengujian 102 segmen, dengan perbandingan 80% dan 20%.

Hasil pengujian ditunjukkan pada Gbr. 6, yaitu diperoleh 56 *true positive*, 35 *true negative*, tujuh *false positive*, dan empat *false negative* dari total data pengujian 102 segmen. Hasil ini menunjukkan bahwa sistem deteksi dapat mengidentifikasi keadaan kantuk dengan benar sebanyak 56 kali, mengidentifikasi keadaan terjaga dengan benar sebanyak 35 kali, salah mendeteksi kelas kantuk sebanyak empat kali, dan salah mendeteksi kelas terjaga sebanyak tujuh kali.

IV. KESIMPULAN

Pada makalah ini telah dibuat suatu model sistem deteksi kantuk menggunakan ansambel *Random Forest* dengan EKG. Masukan sistem deteksi kantuk berasal dari fitur yang diekstraksi dari rekaman EKG dengan metode ranah waktu dan ranah frekuensi. Sebelum dilakukan ekstraksi fitur, rekaman EKG difiltrasi menggunakan filter FIR dan kompleks QRS dideteksi menggunakan algoritme *Hamilton segmenter*. Hasil deteksi kompleks QRS digunakan untuk menentukan NNI yang diperlukan untuk ekstraksi fitur.

Pengembangan model deteksi kantuk *Random Forest* dilakukan dengan optimasi variasi struktur *Random Forest*. Variasi yang dilakukan adalah variasi jumlah *estimator*, jumlah *max features*, mengurutkan fitur berdasarkan kepentingannya dalam menentukan keputusan *Random Forest*, serta variasi masukan kombinasi gabungan fitur. Pelatihan *Random Forest* menghasilkan model dengan kinerja terbaik dengan segmentasi 40 detik dan kombinasi gabungan fitur *cf19*, dengan rata-rata akurasi 94,61%, sensitivitas 96,67%, dan *specificity* 91,67%.

REFERENSI

- [1] WHO, *Global Status Report on Road Safety 2018*, Geneva, Switzerland: World Health Organization, 2018.
- [2] J. He, W. Choi, Y. Yang, J. Lu, X. Wu, dan K. Peng, "Detection of Driver Drowsiness Using Wearable Devices: A Feasibility Study of the Proximity Sensor," *Applied Ergonomics*, Vol. 65, hal. 473-480, Nov. 2017.
- [3] J.D. Slater, "A Definition of Drowsiness: One Purpose for Sleep?" *Med. Hypotheses*, Vol. 71, No. 5, hal. 641-644, Nov. 2008.
- [4] K. Kaida, M. Takahashi, T. Åkerstedt, A. Nakata, Y. Otsuka, T. Haratani, dan K. Fukasawa, "Validation of the Karolinska Sleepiness Scale against

- Performance and EEG Variables,” *Clin. Neurophysiol.*, Vol. 117, No. 7, hal. 1574-1581, Jul. 2006.
- [5] A. Shahid, K. Wilkinson, S. Marcu, dan C.M. Shapiro, Eds., *STOP, THAT and One Hundred Other Sleep Scales*, New York, USA: Springer Science & Business Media, 2012.
- [6] J. Vicente, P. Laguna, A. Barta, dan R. Bailón, “Drowsiness Detection Using Heart Rate Variability,” *Med. Biol. Eng. Comput.*, Vol. 54, No. 6, hal. 927-937, Jan. 2016.
- [7] B. Singh dan N. Bharti, “Software Tools for Heart Rate Variability Analysis,” *Int. J. Recent Sci. Res.*, Vol. 6, No. 4, hal. 3501-3506, Apr. 2015.
- [8] G.S. Wagner dan D.G. Strauss, *Marriott’s Practical Electrocardiography*, Philadelphia, USA: LWW, 2013.
- [9] O. Maftukhaturrizqoh, “Deteksi Kantuk Menggunakan Jaringan Saraf Tiruan dengan Fitur Statistik RR Elektrokardiogram,” Skripsi, Universitas Sebelas Maret, Surakarta, Indonesia, Jul. 2018.
- [10] A. Shalihah, “Sistem Deteksi Kantuk Menggunakan Jaringan Saraf Tiruan (JST) dan Particle Swarm Optimization,” Skripsi, Universitas Sebelas Maret, Surakarta, Indonesia, Sep. 2019.
- [11] A. Liaw dan M. Wiener, “Classification and Regression by randomForest,” *R News*, Vol. 2/3, hal. 18-22, Des. 2002.
- [12] Q. Massoz, T. Langohr, C. François, dan J.G. Verly, “The ULg Multimodality Drowsiness Database (Called DROZY) and Examples of Use,” *2016 IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, 2016, hal. 1-7.
- [13] P. Hamilton, “Open Source ECG Analysis,” *Comput. Cardiol.*, 2002, hal. 101-104.
- [14] K.K. Parhi dan M. Ayinala, “Low-complexity Welch Power Spectral Density Computation,” *IEEE Trans. Circuits Syst. I: Regul. Pap.*, Vol. 61, No. 1, hal. 172-182, Jun. 2013.
- [15] P. Gomes, P. Margaritoff, dan H. Silva, “pyHRV: Development and Evaluation of an Open-source Python Toolbox for Heart Rate Variability (HRV),” *Proc. Int. Conf. Electric., Electron. Comput. Eng. (IcETAN)*, 2019, hal. 822-828.
- [16] Y. Isler, “Discrimination of Systolic and Diastolic Dysfunctions Using Multi-layer Perceptron in Heart Rate Variability Analysis,” *Comput. Biol. Med.*, Vol. 76, hal. 113-119, Sep. 2016.
- [17] A. Sarica, A. Cerasa, dan A. Quattrone, “Random Forest Algorithm for the Classification of Neuroimaging Data in Alzheimer's Disease: A Systematic Review,” *Front. Aging Neurosci.*, Vol. 9, hal. 1-12, Okt. 2017.
- [18] M. Ghaderzadeh, R. Fein, dan A. Standing, “Comparing Performance of Different Neural Networks for Early Detection of Cancer from Benign Hyperplasia of Prostate,” *Appl. Med. Inform.*, Vol. 33, No. 3, hal. 45-54, Sep. 2013.