

Pelabelan Sinyal Elektrokardiogram (EKG) pada Klasifikasi Fibrilasi Atrium Menggunakan *Hidden Markov Model*

Zakia Firdha Razak¹, Adiwijaya², Dody Qori Utama³

^{1,2}Prodi S1 Ilmu Komputasi, Fakultas Informatika, Universitas Telkom

³Prodi S1 Teknik Informatika, Fakultas Informatika, Universitas Telkom

¹zakiafirdha@students.telkomuniversity.ac.id, ²adiwijaya@telkomuniversity.ac.id,

³dodyqori@telkomuniversity.ac.id

Abstrak

Penyakit jantung telah menjadi salah satu penyebab kematian terbesar di dunia. Di antaranya yang paling sering terjadi adalah Fibrilasi Atrium, yaitu kondisi yang meliputi aktivitas sangat cepat dan tidak beraturan dalam atria serta menimbulkan gejala-gejala seperti jantung berdebar, sakit kepala, kehilangan kesadaran, sesak nafas dan rasa letih. Untuk mengenali penyakit ini, akan dikembangkan sebuah sistem pengenalan gelombang EKG. Hidden Markov Model yang menggunakan pendekatan probabilitas telah terbukti menghasilkan performansi yang bagus dalam pengenalan sinyal suara. Oleh karena itu, dalam tugas akhir ini akan diujikan metode HMM dalam kasus elektrokardiogram (EKG). Dalam sistem pendeteksian penyakit jantung terdapat 2 proses yaitu pemodelan dan pengenalan. Pada proses pemodelan akan dibuat suatu model pelabelan sinyal EKG dengan menghitung parameter-parameter HMM, yaitu distribusi inisial state, distribusi probabilitas transisi antar state, dan distribusi probabilitas symbol observasi pada suatu state. Setelah diperoleh ketiga nilai tersebut yang optimal, maka terbentuklah suatu model HMM untuk jenis penyakit jantung. Sedangkan proses pengenalan penyakit jantung dilakukan pada tiap data EKG dengan menghitung likelihood dari data testing yang akan dikenali terhadap semua model data EKG yang telah dilatih sebelumnya. Dengan pelabelan yang teliti dan penentuan nilai probabilitas observasi yang optimal, HMM dapat digunakan untuk mengenali penyakit jantung. Hasil dari pengujian menunjukkan, nilai probabilitas observasi yang didapat dari nilai random selalu menghasilkan akurasi yang berubah-ubah, berbeda jika disama ratakan dengan jumlah state dan jumlah pengamatan. Selain itu, perbandingan jumlah state serta proporsi data juga berpengaruh terhadap akurasi. Akurasi terbesar adalah 100% dengan 2 dan 4 state ketika proporsi data *training* sebanyak 75% dan nilai probabilitas observasi diatur random.

Kata kunci : hidden markov model, elektrokardiogram, distribusi inisial state, distribusi probabilitas transisi antar state, distribusi probabilitas simbol observasi

Abstract

Heart disease has become one of the biggest causes of death in the world. Among the most common is Atrial Fibrillation, a condition that includes very fast and irregular activities in atria and causes many symptoms such as palpitations, headaches, loss of consciousness, shortness of breath and tiredness. To recognize this disease, an ECG wave recognition system will be developed. Hidden Markov Model that use a probability approach has been proven to make good performance in speech signal recognition. Therefore, in this final project the HMM method will be tested on electrocardiogram (ECG) signal. In the detection system for heart disease there are 2 processes; modeling and recognition. In the modeling process, ECG signal labeling model will be made by calculating HMM parameters, namely the initial state distribution, the state transition probability distribution, and the observation symbol probability distribution. After obtaining these three optimal values, an HMM then formed for each type of heart disease. The recognition process of heart disease is carried out on each ECG data by calculating likelihood from testing data that will be identified by all models of ECG record after they have been previously trained. With appropriate labeling and determination of optimal observation probability values, HMM can be used to identify heart disease. The results showed that the probability value of the observations obtained from random values always results in varying accuracy. The comparison of the state number and the data proportion also affects accuracy. The highest accuracy is 100% with 2 states when the proportion of training data is 75% and the probability value is arranged randomly.

Keywords: hidden markov model, electrocardiogram, initial state distribution, state transition probability distribution, observation symbol probability distribution

1. Pendahuluan

Latar Belakang

Dewasa ini, persoalan kesehatan jantung telah menjadi salah satu faktor terbesar penyebab kematian di berbagai negara di dunia dan dapat makin membesar pada masa yang akan datang. Di antara penyakit jantung yang paling sering terjadi adalah Fibrilasi Atrium atau biasa disebut AF (*Atrial Fibrillation*). Penyakit ini tergolong dalam jenis *Arrhythmia*, yaitu gangguan irama jantung.

AF merupakan takiaritmia yang ditandai dengan tidak terkoordinasinya aktivitas atrium akibat kerusakan mekanik atrium (Fuster, dkk., 2001). Biasanya, AF lebih sering terjadi seiring dengan meningkatnya usia dan prevalensinya naik dua kali lipat setiap pertambahan usia satu dekade (sepuluh tahun) setelah tahun kelima dalam kehidupan [1]. AF juga menyebabkan peningkatan mortalitas dan morbiditas, termasuk stroke, gagal jantung serta penurunan kualitas hidup. Pasien dengan AF memiliki resiko stroke lima kali serta resiko gagal jantung tiga kali lebih tinggi dibandingkan dengan pasien tanpa AF [2].

Elektrokardiografi adalah tes yang digunakan untuk mengukur aktivitas listrik detak jantung. Hasil rekamannya disebut dengan elektrokardiogram (EKG), yang mana akan digunakan sebagai objek penelitian pada makalah ini. Karakteristik EKG pada AF ditandai dengan *heart rate* yang sangat cepat dan tidak beraturan sehingga gelombang P di dalam EKG akan sulit atau bahkan tidak dapat dilihat [3]. Dalam pembacaan hasil elektrokardiografi sangat dibutuhkan sifat tenang dan konsentrasi. Jika tidak, bahkan ahli kardiologi dapat memberikan analisa yang salah [4].

Sehubungan dengan permasalahan di atas, perlu adanya sistem klasifikasi sinyal EKG untuk membantu tenaga medis dalam mendeteksi penyakit AF yang mungkin terjadi pada pasien secara lebih akurat. Klasifikasi sinyal pada umumnya terdiri dari tiga langkah yaitu *pre-processing*, ekstraksi ciri, dan klasifikasi. Untuk tahapan *pre-processing* akan dilakukan dengan metode *resampling*, kemudian ekstraksi ciri menggunakan teknik pelabelan sinyal, dan *classifiernya* adalah *Hidden Markov Model* (HMM) dengan pengamatan diskrit.

HMM telah secara luas diterapkan pada masalah analisis sinyal seperti pemrosesan ucapan (*speech processing*) sejak pertengahan tahun 1970. Popularitas HMM utamanya disebabkan oleh algoritma estimasi parameter otomatis yang ditemukan oleh Baum dan Eagon [5] pada akhir 1960-an, lalu selanjutnya diaplikasikan pada *speech processing* [6],[7]. Model ini akhirnya menarik perhatian para peneliti setelah dipelopori oleh Coast, dkk. Dalam analisisnya pada penyakit aritmia menggunakan HMM [8]. Pada 2010, Jondri, dkk., juga telah melakukan penelitian tentang deteksi otomatis kelainan jantung menggunakan HMM dan memperoleh akurasi hingga 95%. Oleh karena itu, metode inilah yang akhirnya dipilih untuk mengklasifikasikan penyakit AF. Sedangkan dalam ekstraksi ciri, sudah banyak teknik canggih serta transformasi diajukan demi menentukan metode yang cepat dan akurat. Ekstraksi menggunakan cara pelabelan berdasarkan bentuk sinyal EKG (keberadaan gelombang PQ, QRS, ST, Baseline) tergolong mudah diterapkan untuk mengubahnya dari data yang awalnya berupa barisan kontinu ke dalam simbol-simbol diskrit, dengan tetap menghasilkan akurasi yang baik.

Topik dan Batasannya

Dalam penelitian ini diangkat beberapa isu/permasalahan berdasarkan latar belakang yang telah dijelaskan sebelumnya, di antaranya adalah bagaimana cara merepresentasikan sinyal EKG ke dalam state HMM dengan teknik pelabelan berdasarkan keberadaan gelombang, tahapan membuat suatu pemodelan HMM yang mampu mengklasifikasi penyakit AF, serta cara melakukan pengenalan terhadap penyakit AF.

Dari tiga permasalahan tersebut, penelitian ini juga memiliki beberapa batasan masalah dalam proses pengerjaannya meliputi jumlah kelas dari sinyal yang diuji, yaitu hanya dua kelas saja dari penyakit AF dan NSR menggunakan data sekunder yang diperoleh dari situs internet PhysioNet MIT-BIH *Atrial Fibrillation Database* untuk sinyal AF (EKG penderita fibrilasi atrium) dan MIT-BIH *Normal Sinus Rythm Database* untuk sinyal NSR (EKG dengan detak jantung normal). Sistem ini menggunakan teknik *resampling* pada tahap *pre-processing* guna menyamakan frekuensi dari dua kelas yang berbeda, pelabelan sinyal pada ekstraksi ciri ke dalam simbol-simbol diskrit berdasarkan bentuk gelombang EKG, lalu diklasifikasikan menggunakan HMM setelah dibuatkan model untuk masing-masing kelas. Perlu diperhatikan bahwa analisis sinyal jantung ini tidak dilakukan untuk tujuan medis, tetapi hanya secara pengolahan sinyal dan probabilitas model saja.

Tujuan

Adapun beberapa tujuan yang hendak dicapai dalam penelitian ini di antaranya adalah merepresentasikan sinyal EKG ke dalam state HMM dengan cara pelabelan, membuat suatu pemodelan HMM yang mampu mengklasifikasi penyakit AF, dan melakukan proses pengenalan terhadap penyakit AF.

Organisasi Tulisan

Pada proses penelitian ini, telah dilakukan beberapa tinjauan pustaka atau studi berkenaan dengan klasifikasi sinyal EKG pada AF. Tinjauan tersebut meliputi dasar elektrokardiogram itu sendiri, penyakit AF dalam kaitannya dengan detak jantung normal dan abnormal, serta tahapan klasifikasi dan metode-metode yang

diambil. Selanjutnya dilakukan pengumpulan data untuk membangun sistem, sehingga hasilnya dapat dievaluasi dan ditarik kesimpulan.

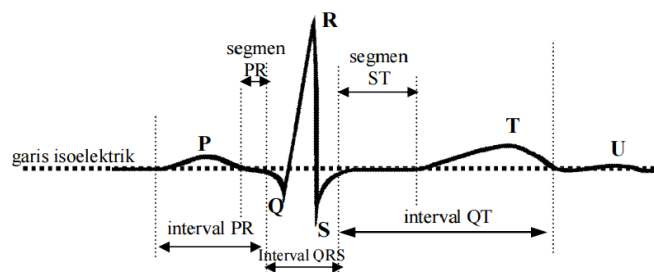
2. Studi Terkait

2.1 Elektrokardiogram (EKG)

Elektrokardiogram (EKG) adalah suatu gambaran dari potensial listrik yang dihasilkan oleh aktivitas listrik otot jantung (Thaler, 2000). Pemeriksaan EKG telah digunakan dalam berbagai studi selama lebih dari 40 tahun, dan saat ini EKG merupakan prosedur diagnostik yang paling sering dilakukan untuk mendiagnosa kelainan jantung karena sifatnya yang aman, mudah, *reproducible*, dan tidak mahal [9].

Otot jantung terbentuk dari serabut-serabut otot yang bermuatan listrik, dikarenakan adanya aliran ion Natrium dari dan ke dalam sel. Ion inilah yang membuat jantung mengalami siklus depolarisasi-repolarisasi secara kontinu sehingga membentuk pola denyutan jantung. Bioelektrik jantung dibangkitkan oleh *Sinoatrial Node (SA Node)* dan *Atrioventricular Node (AV Node)* kemudian menjaral melalui sel konduksi yang disebut berkas HIS atau serat Purkinje, dan mengalir ke seluruh bagian jantung sehingga membentuk kompleks sinyal EKG di permukaan tubuh.

Dalam Ilmu Kardiologi dikenal 12 sadapan EKG standar yang digunakan untuk memperoleh tafsiran kondisi jantung. Tidak seluruh bagian rekaman EKG memiliki arti klinis dalam penafsirannya. Hanya bagian-bagian tertentu yang dipakai sebagai dasar penentuan suatu kondisi jantung, seperti terlihat pada Gambar 2.1 [10]



Gambar 2-1 Grafik bentuk EKG^[11]

2.2 Fibrilasi Atrium/Atrial Fibrillation (AF)

Fibrilasi Atrium/*Atrial Fibrillation (AF)* merupakan suatu kondisi ritme jantung abnormal yang paling sering terjadi dan menyerang jutaan manusia di seluruh dunia. Kondisinya meliputi aktivitas yang sangat cepat dan tidak beraturan dalam atria serta menimbulkan gejala-gejala seperti jantung berdebar, sakit kepala, kehilangan kesadaran, sesak nafas dan rasa letih. AF juga merupakan penyebab umum dari stroke yang rata-rata menyerang 5% orang yang memiliki AF setiap tahunnya (Pendidikan, Dosen, 2014).

AF adalah takiaritmia supraventrikular yang khas, dengan aktivasi atrium yang tidak terkoordinasi mengakibatkan perburukan fungsi mekanis atrium. Pada elektrokardiogram (EKG), ciri dari AF adalah tiadanya konsistensi gelombang P, yang digantikan oleh gelombang getar (fibrilasi) yang bervariasi amplitudo, bentuk dan durasinya. Pada fungsi NAV yang normal, AF biasanya disusul oleh respons ventrikel yang juga ireguler, dan seringkali cepat [12].

Ciri-ciri AF pada gambaran EKG umumnya sebagai berikut [13]:

1. EKG permukaan menunjukkan pola interval RR yang ireguler.
2. Tidak dijumpainya gelombang P yang jelas pada EKG permukaan. Kadang-kadang dapat terlihat aktivitas atrium yang ireguler pada beberapa sadapan EKG.
3. Interval antara dua gelombang aktivasi atrium tersebut biasanya bervariasi, umumnya kecepatannya melebihi 450 kali per menit.

Berikut ini adalah rekaman EKG AF (Gambar 2.3) dibandingkan dengan EKG NSR dan Aritmia (Gambar 2.4 dan 2.5) yang didapat dari situs Physionet.



Gambar 2-3 Sinyal EKG AF



Gambar 2-4 Sinyal EKG NSR

2.3 Hidden Markov Model (HMM)

Hidden Markov Model (HMM) adalah rantai markov yang statenya tidak dapat diamati secara langsung. Modelnya diasumsikan sebagai suatu Proses Markov dengan parameter yang tak diketahui, dan tantangannya adalah menentukan parameter tersembunyi dari parameter yang dapat diamati. Parameter-parameter yang

ditentukan kemudian dapat digunakan untuk analisis yang lebih jauh, salah satunya seperti dalam pengenalan sinyal jantung. *Hidden Markov Model* sangat populer diaplikasikan di bidang pengolahan sinyal suara dan bioinformatika.

HMM merupakan pendekatan dengan menggunakan kumpulan *state* dimana *state* ini memiliki hubungan satu sama lain yang menggambarkan suatu kejadian di dunia nyata, dengan layer observasi sebagai hal yang dapat diamati, dan layer tersembunyi sebagai layer yang hanya bisa diamati melalui layer observasi. HMM ini dapat memodelkan persoalan – persoalan di dunia nyata yang sifatnya probabilistik [6].

HMM mempunyai beberapa elemen, yaitu N buah state $\{S_1, S_2, S_3, \dots, S_n\}$, M buah pengamatan untuk setiap state, misalkan $\{V_1, V_2, \dots, V_m\}$, matriks peluang transisi $A = [a_{ij}]$ dimana $a_{ij} = P(x_{n+i}=S_j | x_n=S_i)$, matriks peluang emisi $B = [b_{i,k}]$, dengan $b_{i,k} = P(O_i=V_k | x_i=S_i)$ dan distribusi inisial $\pi = [\pi_i(0)]$, dimana $\pi_i(0) = P(X_0 = S_i)$ [14].

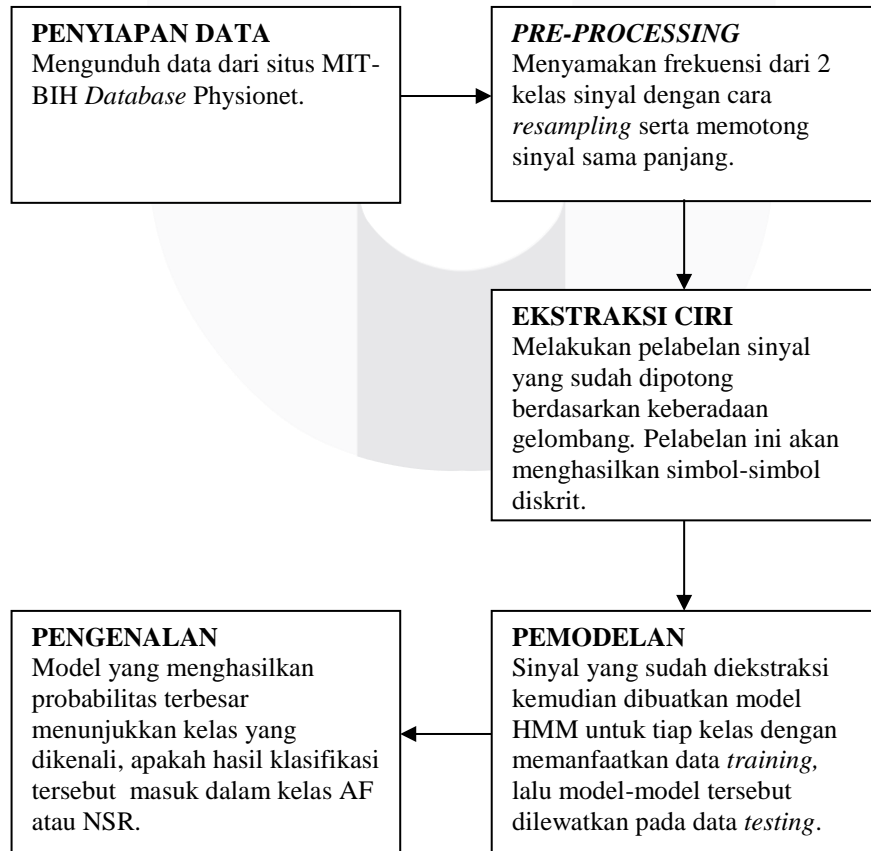
Matriks peluang transisi A dan matriks peluang emisi B untuk masing-masing sinyal EKG yang akan dikenali ditentukan terlebih dahulu dengan menggunakan data *training*. Matriks A dan B hasil *training* ini akan digunakan dalam proses klasifikasi sinyal EKG untuk deteksi AF [14].

Hasil *training* berupa model HMM $\lambda = (A, B, \pi)$ dengan A adalah matriks peluang transisi antar *state*, B matriks peluang simbol observasi, dan π peluang *state* inisial. Pada penelitian ini jenis HMM yang digunakan merupakan diskrit ergodic, dimana parameter-parameter HMM seperti matriks A, B, π dibangkitkan secara random dengan nilai yang dinormalisasi ke satu. Nilai-nilai matriks A, B, π tersebut kemudian di-estimasi ulang melalui proses pelatihan untuk mendapatkan nilai parameter yang optimal.

Keluaran dari re-estimasi parameter HMM adalah nilai baru elemen-elemen matriks A, B, π . Iterasi untuk re-estimasi dihentikan apabila iterasi sudah maksimal atau jika model baru tidak memberikan perubahan yang berarti. Setelah proses re-estimasi berakhir sistem akan menyimpan model $\lambda = (A, B, \pi)$. Model ini nantinya akan digunakan pada saat *testing* dan jumlah model yang disimpan sama dengan jumlah kelas yang dikenali. Tiap data *testing* dilewatkan pada model yang telah dibuat pada masing-masing kelas tadi, lalu probabilitas observasi yang paling tinggi akan menunjukkan kelas yang dikenali.

3. Sistem yang Dibangun

Serangkaian proses klasifikasi AF berdasarkan sinyal EKG pada penelitian ini memiliki beberapa tahapan seperti yang telah disebutkan sebelumnya. Tahapan tersebut akan dilakukan sesuai dengan diagram blok pada gambar di bawah ini.



Gambar 3-1 Diagram blok perancangan sistem klasifikasi sinyal EKG.

3.1 Penyiapan Data

Data yang akan digunakan dalam penelitian ini adalah hasil rekaman EKG tipe AF (berjumlah 20 data) dan NSR (berjumlah 12 data), didapat dari situs internet MIT-BIH *Atrial Fibrillation Database* dan MIT-BIH *Normal Sinus Rhythm Database* – PhysioNet yang berekstensi .mat dengan durasi seluruh data adalah 10 detik. Meski durasinya sama, kedua tipe sinyal ini memiliki *sampling rate* yang berbeda. Tipe AF memiliki frekuensi 250 Hz sedangkan tipe NSR hanya 128 Hz. Oleh karena itu, diperlukan penyamaan frekuensi pada tahap *pre-processing*.

3.2 Pre-processing

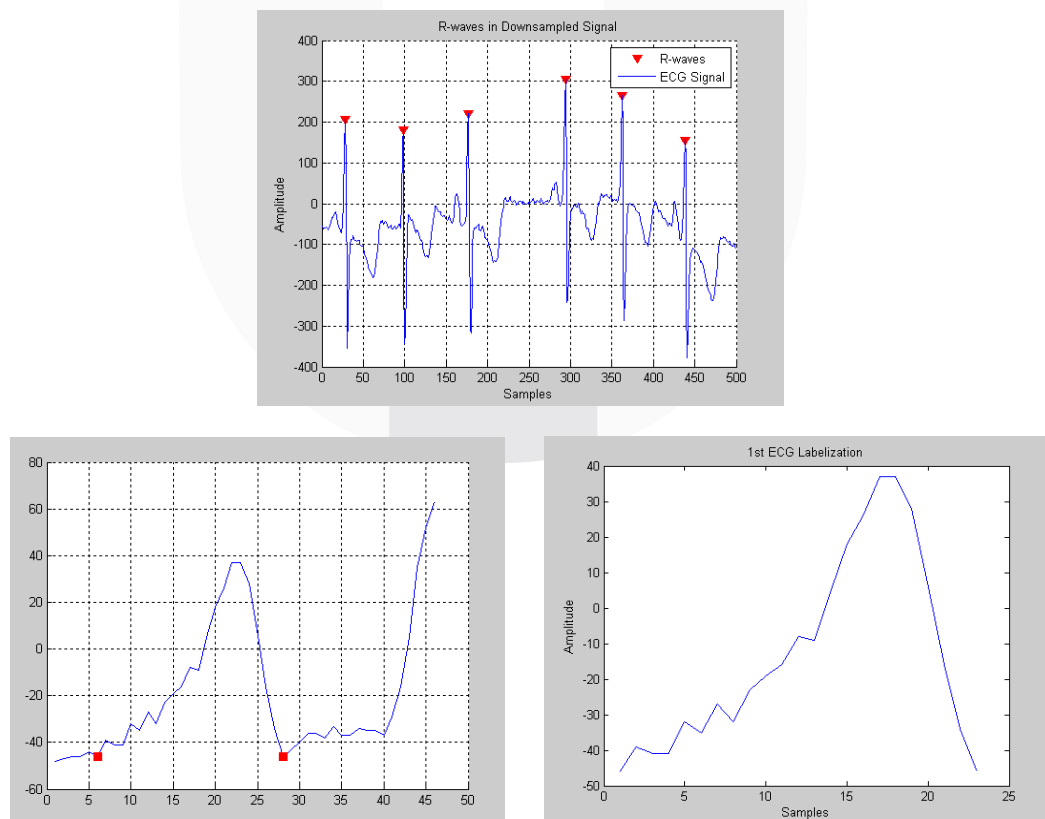
Kedua tipe sinyal dengan frekuensi *sampling* yang berbeda akan disamakan dahulu dengan metode *resampling* sehingga sama-sama memiliki jumlah *sample* sepanjang 2500. Metode ini dipakai untuk melakukan *sampling* ulang dari data asli agar diperoleh sekumpulan data baru dengan jumlah yang diinginkan (bisa lebih banyak atau lebih sedikit).

Selain melakukan *resampling*, di tahap *pre-processing* ini juga dilakukan pemotongan sinyal yang sebelumnya berjumlah 2500 *sample* menjadi hanya 650 *sample* saja agar tidak terlalu panjang dan mempermudah pengolahan data. Tidak ada alasan khusus mengapa dipilih 650 *sample*, pada dasarnya sinyal dapat dipotong menjadi berapapun kemungkinan jumlah *sample* dengan tetap mempertahankan ciri khas bentuk sinyal EKG.

3.3 Ekstraksi Ciri

Setelah *pre-processing* dilakukan, tahap selanjutnya adalah ekstraksi ciri menggunakan cara pelabelan. Berdasarkan penggolongannya, gelombang sinyal EKG termasuk ke dalam golongan sinyal *quasiperiodic*, jadi sinyal jenis ini hampir selalu berulang secara periodik. Hal tersebut dapat dilihat dari bentukan sinyal EKG yang berurutan dari P-QRS-T dan kembali lagi ke P. Oleh karena itu, pada penelitian ini gelombang-gelombang tersebut dilabeli ke dalam 4 bagian, yaitu PQ, QRS, ST, dan *baseline* (*node* yang menghubungkan T dan P).

Segmentasi dari 4 bagian gelombang EKG ini dilakukan dengan bantuan deteksi puncak dan lembah, dan selalu diawali dari gelombang P. Untuk kasus HMM dengan 4 *state*, gelombang PQ menunjukkan *state* 1 dilabeli dengan 1, QRS dilabeli dengan 2, ST dilabeli dengan 3, dan *baseline* dilabeli dengan 4. Pada akhirnya, sinyal EKG yang merupakan barisan kontinu tadi akan berubah menjadi barisan diskrit sehingga klasifikasi menggunakan HMM dapat dilakukan.



Gambar 3-2 Proses segmentasi sinyal EKG dengan deteksi puncak dan lembah sehingga masing-masing gelombang dapat dilabeli.

3.4 Tiga Permasalahan Dasar pada HMM dan Solusinya

Dalam HMM, ada tiga permasalahan dasar yang harus dipecahkan agar model tersebut dapat diterapkan dalam permasalahan nyata, yaitu:

1. Masalah Pertama (*Evaluation*)
Diberikan deretan observasi $O = \{o_1, o_2 \dots o_T\}$ dan model $\lambda = (A, B, \pi)$, bagaimana menghitung probabilitas deretan observasi berdasarkan model yang telah diberikan ($P(O/\lambda)$) secara efisien?
2. Masalah Kedua (*Decoding*)
Diberikan deretan observasi $O = \{o_1, o_2 \dots o_T\}$ dan model λ , bagaimana memilih deretan state $q = (q_1, q_2, q_3, \dots, q_T)$ yang optimal?
3. Masalah Ketiga (*Training*)
Bagaimana mengatur parameter model $\lambda = (A, B, \pi)$ yang dapat memaksimalkan $P(O/\lambda)$?
Namun dalam kasus klasifikasi sinyal EKG ini, permasalahan yang harus dipecahkan adalah masalah pertama (*evaluation*) dan masalah ketiga (*training*) saja karena berkenaan dengan pemodelan dan pengenalan. Sedangkan masalah kedua (*decoding*) tidak diperlukan.

3.4.1 Training (Solusi Masalah Ketiga) atau Pemodelan

Pada tahap *training*, data dari tiap kelas dibagi menjadi 25%, 50%, dan 75% dari total keseluruhan untuk digunakan sebagai masukan sebelum menguji apakah sistem mengenalinya dengan benar. Data yang sudah dipakai untuk *training* tidak dipakai lagi untuk *testing*. Proses *training* ini dilakukan untuk membentuk matriks peluang transisi A dan matriks peluang emisi B [14].

Pada proses ini dihasilkan 2 model yang merepresentasikan setiap kelasnya. Model tersebut dapat dinotasikan $\lambda = (A, B, \pi)$ sebagai model parameter dimana A menunjukkan matriks peluang transisi, B menunjukkan matriks peluang observasi, dan π menunjukkan matriks peluang suatu kejadian pada awal state.

Dengan *training*, akan didapatkan model baru yang telah dilatih. Sebenarnya terdapat beberapa algoritma yang bisa digunakan untuk pelatihan, tetapi pada penelitian ini yang akan digunakan yaitu algoritma Baum-Welch. Tujuannya adalah melakukan estimasi parameter model guna memaksimalkan $P(O/\lambda)$.

Disini terdapat beberapa skenario dengan menggunakan 2, 3, 4, 5, dan 6 hidden state. Hasil dari proses ini berupa matriks peluang transisi dan matriks peluang transmisi oleh masing-masing kelas. Sedangkan inialisasinya berupa matriks random dengan jumlah baris sama dengan 1. Kemudian inialisasi matriks emisi dilakukan dengan dua cara, yaitu saat nilainya disama-ratakan dan saat nilainya random. Adapun matriks peluang state awal diasumsikan bernilai 1 sebagai peluang pertama kali kemunculan state, karena pada penelitian ini memang segmentasi serta pelabelan sinyal selalu dimulai dari gelombang P yang direpresentasikan oleh simbol diskrit 1.

Setelah inialisasi model $\lambda = (A, B, \pi)$ dibangkitkan, perlu dilakukan perhitungan $\alpha_t(i)$ dan $\beta_t(i)$ dengan menggunakan algoritma *forward* dan *backward*. $\alpha_t(i)$ dapat dihitung secara induktif melalui tiga tahap menggunakan *forward algorithm* yaitu [15]:

1. Inialisasi
$$\alpha_1 = \pi_i b_i(O_1), \quad 1 \leq i \leq N \quad (3.1)$$

2. Induksi
$$\alpha_{t+1}(j) = [\sum_{i=1}^N \alpha_t(i) a_{ij}] b_j(O_{t+1}),$$
$$1 \leq t \leq T - 1$$
$$1 \leq j \leq N \quad (3.2)$$

3. Terminasi
$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i) \quad (3.3)$$

Penjelasan dari tiga poin di atas yaitu; pada langkah pertama peluang *forward* diinisialisasi sebagai peluang *joint* untuk memiliki observasi O_1 pada state awal i untuk semua kemungkinan state. Langkah kedua, $\alpha_{t+1}(j)$, peluang *joint* dari sebagian barisan observasi hingga $t + 1$ dan state j saat $t + 1$ dihitung dari $\alpha_t(j)$, peluang sebagian barisan observasi yang sebelumnya, secara rekursif.

Sedangkan $\beta_t(i)$ dapat kita hitung secara induktif menggunakan algoritma *backward* sebagai berikut.

1. Inialisasi
$$\beta_T(i) = 1, \quad 1 \leq i \leq N \quad (3.4)$$

2. Induksi
$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j),$$
$$t = T - 1, T - 2, \dots, 1, 1 \leq i \leq N \quad (3.5)$$

Selanjutnya proses re-estimasi parameter HMM dapat dilakukan dengan formula berikut.

$$\pi_i = \frac{\hat{a}_1(i)\hat{\beta}_1(i)}{\sum_{j=1}^N \hat{a}_t(j)} \tag{3.6}$$

$$a_{ij} = \frac{\sum_{t=1}^{T-1} \hat{a}_t(i)a_{ij}b_j(O_{t+1})\hat{\beta}_{t+1}(j)}{\sum_{t=1}^{T-1} \sum_{j=1}^T \hat{a}_t(i)a_{ij}b_j(O_{t+1})\hat{\beta}_{t+1}(j)} \tag{3.7}$$

$$b_j(k) = \frac{\sum_{t=1}^{T-1} \hat{a}_t(i)\hat{\beta}_t(i)\delta(O_t, V_k)}{\sum_{t=1}^{T-1} \hat{a}_t(i)\hat{\beta}_t(i)} \tag{3.8}$$

3.4.2 Evaluation/Testing (Solusi Masalah Pertama) atau Pengenalan

Pada proses pengenalan, deretan observasi dari data testing yang akan dikenali kemudian dilewatkan pada model-model HMM yang sudah disimpan. Untuk setiap model HMM akan dihitung peluang bahwa suatu penyakit jantung yang akan dikenali itu benar dibangkitkan oleh model HMM tersebut dan model HMM dengan peluang terbesar (*maximum likelihood*) akan dipilih sebagai model yang dianggap menghasilkan penyakit jantung tersebut.

4. Evaluasi

4.1 Hasil Pengujian

Pada pelabelan yang dilakukan saat ekstraksi, sinyal EKG berhasil diubah ke dalam simbol-simbol diskrit berdasarkan keberadaan puncak dan jumlah *state*. Hasil pelabelan ditunjukkan pada tabel di bawah ini.

Tabel 4-1 Pelabelan sinyal EKG ke dalam simbol diskrit berdasarkan jenis dan letak gelombang

State	Hasil Pelabelan Menjadi Simbol Diskrit Berdasarkan Jenis Gelombang							
	PQ	QRS	ST	Baseline	PQ	QRS	ST	Baseline
2	1	2	1	2	1	2	1	2
3	1	2	3	1	2	3	1	2
4	1	2	3	4	1	2	3	4
5	1	2	3	4	5	1	2	3
6	1	2	3	4	5	6	1	2

Sinyal yang telah dilabeli ke dalam simbol-simbol tersebut kemudian masuk ke tahap pemodelan untuk menghasilkan HMM pada masing-masing kelas. Berikut ini matriks A dan B hasil re-estimasi dari nilai-nilai awal yang ditebak secara random. Hasil re-estimasi tersebut diperoleh setelah melalui tahap *training* menggunakan algoritma *forward-backward*.

Tabel 4-2 Contoh nilai baru parameter A dan B pada masing-masing jenis EKG setelah reestimasi

EKG	A				B			
AF	0.984	0.016	0	0	7.78E-217	3.17E-239	0.594	0.406
	0	2.00E-05	0.999	0	0.842	0	0	0.158
	0	0	0.982	0.018	0.999	3.20E-201	0	2.08E-250
	0.039	0	0	0.961	8.09E-132	0.999	7.59E-144	0
NSR	0.988	0.012	0	0	0.727	1.60E-253	3.70E-298	0.273
	0	0.958	0.042	0	9.09E-135	0.999	7.03E-189	0
	0	0	0.974	0.026	0	1.79E-71	0.999	0
	0.026	0	0	0.974	0	0	0.999	5.20E-172

Adapun untuk tahap akhir pengujian (*testing*), skenario yang diterapkan adalah dengan tiga cara berikut.

1. Mengubah nilai parameter B, yaitu ketika nilainya disama ratakan dan ketika nilainya diatur random. Mengubah-ubah tebakan nilai awal seperti ini sebenarnya bisa diterapkan pula terhadap parameter A dan π , namun nilai awal parameter A dalam penelitian ini telah ditetapkan berdasarkan hasil pemotongan fitur saat ekstraksi, sedangkan nilai π diasumsikan bernilai 1 seperti telah dijelaskan sebelumnya.
2. Mengubah jumlah state mulai dari berjumlah 2, 3, 4, 5, hingga 6 state untuk melihat berapakah jumlah state yang paling optimal digunakan pada penelitian ini.

3. Mengubah perbandingan jumlah data *training* dan data *testing*, yaitu 25:75, 50:50, dan 75:25, untuk mengetahui apakah proporsi data akan memberikan pengaruh yang berarti dalam proses klasifikasi.

Dari hasil pengujian dengan data *testing* yang telah dilabelisasi dan melewati proses yang telah dibahas pada sub bab-sub bab sebelumnya, maka didapatkan suatu sistem pengenalan penyakit jantung. Performansi untuk nilai B awal random menghasilkan akurasi yang selalu berubah-ubah atau tidak tentu, namun dalam penelitian ini akan dicantumkan nilai yang paling tinggi dari semua nilai akurasi yang pernah keluar dari tiap *running*. Perbedaan akurasi dalam dua percobaan tersebut selain dipengaruhi oleh nilai B, juga dikarenakan perbedaan jumlah state serta proporsi data *training* dan *testing*.

Tabel 4-3 Tabel hasil akurasi menggunakan proporsi data 25:75 (*training:testing*)

Jumlah State	Proporsi Data 25:75					
	Akurasi		Sensitivitas		Spesifisitas	
	B Random	B Samarata	B Random	B Samarata	B Random	B Samarata
2	91.67%	91.67%	93.33%	93.33%	88.89%	88.89%
3	66.67%	66.67%	66.67%	66.67%	66.67%	66.67%
4	79.16%	75.00%	80.00%	80.00%	77.78%	66.67%
5	79.16%	70.83%	86.67%	60.00%	66.67%	88.89%
6	66.67%	62.50%	73.33%	100.00%	55.56%	0.00%
Rata-rata	76.67%	73.33%	80.00%	80.00%	71.11%	62.22%

Tabel 4-4 Tabel hasil akurasi menggunakan proporsi data 50:50 (*training:testing*)

Jumlah State	Proporsi Data 50:50					
	Akurasi		Sensitivitas		Spesifisitas	
	B Random	B Samarata	B Random	B Samarata	B Random	B Samarata
2	81.25%	81.25%	100.00%	100.00%	50.00%	50.00%
3	81.25%	81.25%	90.00%	90.00%	66.67%	66.67%
4	87.50%	87.50%	100.00%	100.00%	66.67%	66.67%
5	75.00%	37.50%	90.00%	0.00%	50.00%	100.00%
6	87.50%	62.50%	100.00%	90.00%	66.67%	50.00%
Rata-rata	82.50%	70.00%	96.00%	76.00%	60.00%	66.67%

Tabel 4-5 Tabel hasil akurasi menggunakan proporsi data 75:25 (*training:testing*)

Jumlah State	Proporsi Data 75:25					
	Akurasi		Sensitivitas		Spesifisitas	
	B Random	B Samarata	B Random	B Samarata	B Random	B Samarata
2	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
3	87.50%	87.50%	80.00%	80.00%	100.00%	100.00%
4	100%	100%	100.00%	100.00%	100.00%	100.00%
5	100.00%	37.50%	100.00%	0.00%	100.00%	100.00%
6	87.50%	75.00%	100.00%	100.00%	66.67%	33.33%
Rata-rata	95.00%	80.00%	96.00%	76.00%	93.33%	86.67%

Pada ketiga tabel di atas dapat dilihat bahwa pengubahan jumlah proporsi data berpengaruh terhadap kenaikan akurasi, yaitu semakin banyak data *training* maka semakin besar akurasi yang didapat. Sedangkan perbedaan jumlah state tidak selalu memberikan peningkatan performansi secara teratur, dengan akurasi paling tinggi didapat saat *state* berjumlah 2. Kemudian untuk nilai parameter B, akurasi cenderung lebih tinggi ketika nilainya diatur random.

4.2 Analisis Hasil Pengujian

1. Skenario Nilai Parameter B

Melakukan percobaan terhadap nilai B menghasilkan akurasi yang berbeda dengan hasil uji menggunakan nilai B random terbukti lebih tinggi akurasinya dibandingkan jika nilainya disama-ratakan. Akurasi dengan nilai B random mencapai rata-rata 84.72% sedangkan akurasi dengan nilai B sama-rata hanya 74.44%.

Tabel 4-6 Tabel hasil perbandingan akurasi saat B diatur sama rata dan random

B	25:75	50:50	75:25	Rata-rata
Sama Rata	73.33%	70%	80%	74.44%
Random	76.67%	82.50%	95%	84.72%

Hal ini disebabkan karena nilai random menghasilkan akurasi yang selalu berubah-ubah dan tidak tentu, sehingga banyak kemungkinan nilai yang terjadi. Adapun akurasi yang dicantumkan hanya satu yang tertinggi dari beberapa kali *running* memakai data random.

2. Skenario Variasi Jumlah State

Jumlah state yang coba diujikan adalah 2, 3, 4, 5, dan 6. Akurasi tertinggi didapat oleh pengujian dengan 2 state, yaitu 90.97%, disusul 4 state dengan akurasi 88.19%, 3 state dengan 78.47%, 6 state dengan 73.61%, dan terakhir 5 state dengan 66.67%.

Tabel 4-7 Tabel hasil perbandingan akurasi saat jumlah state diubah-ubah

Jumlah State	25:75	50:50	75:25	Rata-rata
2	91.67%	81.25%	100.00%	90.97%
3	66.67%	81.25%	87.50%	78.47%
4	77.08%	87.50%	100.00%	88.19%
5	75.00%	56.25%	68.75%	66.67%
6	64.59%	75.00%	81.25%	73.61%

Hal ini menandakan bahwa tidak selalu bertambahnya state menjadikan akurasi lebih tinggi. Dalam penelitian ini, akurasi cenderung tinggi saat jumlah state adalah genap karena sejak awal pelabelan memang dilakukan dengan segmentasi sinyal ke dalam beberapa bagian berjumlah genap sehingga sesuai dengan bentuk EKG-nya.

3. Skenario Proporsi Data *Training* dan Data *Testing*

Dalam penelitian ini proporsi data dibagi menjadi tiga, yaitu 75%, 50%, dan 25% data *training*. Data yang sudah dipakai untuk melakukan pemodelan ini tidak dipakai lagi pada proses testing. Dilihat dari akurasi yang didapat oleh ketiga proporsi data, akurasi cenderung meningkat seiring bertambahnya jumlah data training yang diambil secara acak. Akurasi tertinggi didapat oleh proporsi data 75:25 yaitu 87.50%, kemudian proporsi data 50:50 sebanyak 76.25%, dan terakhir proporsi data 25:75 dengan akurasi 75.00%.

Tabel 4-8 Tabel hasil perbandingan akurasi berdasarkan proporsi data

Jumlah State	25:75	50:50	75:25
2	91.67%	81.25%	100.00%
3	66.67%	81.25%	87.50%
4	77.08%	87.50%	100.00%
5	75.00%	56.25%	68.75%
6	64.59%	75.00%	81.25%
Rata-rata	75.00%	76.25%	87.50%

Adapun faktor yang paling mempengaruhi sistem klasifikasi tersebut adalah ketepatan dalam pelabelan EKG, nilai B yang random sehingga selalu berubah-ubah, serta perbandingan jumlah state. Akurasi tertinggi yang didapat adalah 100%, terjadi sebanyak lima kali dan semuanya ada pada proporsi data 75:25. Hal ini dikarenakan semakin banyak data yang dilatih, maka semakin dalam pula sistem itu 'belajar'. Oleh karena itu, state yang optimal dalam penelitian ini adalah 2 dan 4, dengan nilai B random serta menggunakan proporsi data training sebesar 75%.

5. Kesimpulan

HMM dapat dipakai untuk melakukan klasifikasi pada sinyal EKG, meskipun penggunaannya kalah populer dengan penerapan HMM terhadap kasus *speech recognition* dan bioinformatika. Teknik pelabelan biasa terbukti mampu merepresentasikan barisan sinyal EKG ke dalam simbol-simbol diskrit dengan cara yang mudah. Hasil pemodelan mampu mendeteksi penyakit AF dengan akurasi yang cukup baik, yaitu mencapai 100% dari hasil yang tertinggi.

HMM yang digunakan dalam tugas akhir ini merupakan HMM standar, untuk penelitian lebih lanjut dapat dicoba menggunakan metode HMM lainnya seperti *factorial* atau *continuous* guna memperoleh akurasi pengenalan ECG yang lebih baik. Selain itu, dapat diubah jumlah state yang digunakan dan cara merepresentasikan sinyal EKG ke dalam state untuk mendapatkan representasi HMM yang benar-benar dapat mengenali HMM dengan tepat.

Daftar Pustaka

- [1] Rampengan, S. H. 2015. *Kardioversi Pada Fibrilasi Atrium*. Jakarta: Badan Penerbit FKUI.
- [2] Perhimpunan Dokter Spesialis Kardiovaskular Indonesia. 2014. *Pedoman Tata Laksana Fibrilasi Atrium*. Jakarta: Centra Communications.
- [3] Damayanti, B. P., Limantoro, Charles. 2014. *Hubungan Antara Hipertensi dan Hipertrofi Ventrikel Kiri pada Pasien Lansia dengan Atrial Fibrilasi*. Diponegoro University Institutional Repository.
- [4] Hasanah, U., dkk. 2016. *Perbandingan Metode SVM, Fuzzy-KNN, Dan BDT-SVM untuk Klasifikasi Detak Jantung Hasil Elektrokardiografi*. Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK), Vol. 3, No. 2,.
- [5] Baum, L. E., Eagon J. A. *An Inequality with Applications to Statistical Prediction for Functions of Markov Processes and to A Model for Ecology*. Bull. Amer. Math. Soc., vol. 73. pp. 360-363, 1967.
- [6] Baker, I. K. *Stochastic Modeling for Automatic Speech Understanding*.
- [7] Jelinek, F. *Continuous Speech Recognition by Statistical Methods*. Proc. IEEE, vol. 64, pp. 532-556, Apr. 1976.
- [8] Coast, D. A., Stern R. M. "An Approach to Cardiac Arrhythmia Analysis Using Hidden Markov Models. IEEE Trans. Biomed. Eng., vol. 37, no. 9, 1990.
- [9] Kadish, A. H., Buxton, A. E., Kennedy, H. L., Knight, B. D., Mason, J. W., Schuger, C. D., et al. 2001. *ACC/AHA Clinical Competence and Ambulatory on Electrocardiography*. J Am Coll Cardiol; 38: 2091-100.
- [10] Sudjadi, Warsito, A., Nugroho, E. S. *Pengenalan Pola Sinyal Elektrokardiograf (EKG) dengan Jaringan Syaraf Tiruan Backpropagation untuk Diagnosa Kelainan Jantung Manusia*. Semarang: Jurusan Teknik Elektro Universitas Diponegoro.
- [11] Meurs-Arntzenius. 1990. *Elektrokardiografi Praktis*. Hipokrates.
- [12] Bellet, S. 1971. *Clinical Disorders of the Heart Beat*. Philadelphia: Lea & Febiger.
- [13] S, Wijaya. 1990. *EKG Praktis*. Jakarta: Binarupa Aksara.
- [14] Jondri, Rizal, A. *Deteksi Otomatis Kelainan Jantung Menggunakan Hidden Markov Model (HMM)*. Bandung: Biomedical Signal Processing & Instrumentation Research Group (BioSPIN) Fakultas Elektro & Komunikasi Institut Teknologi Telkom.
- [15] Rabiner, Lawrence R. 1989. *A Tutorial on Hidden Markov Model and Selected Applications in Speech Recognition*. Proc. IEEE, vol. 77, no. 2, February 1989.