

Penerapan Algoritma KMP dan Boyer-Moore untuk Deteksi Pola Abnormal pada Representasi Simbolik Sinyal ECG

Ramadhian Nabil Firdaus Gumay - 13524126

Program Studi Teknik Informatika

Sekolah Teknik Elektro dan Informatika

Institut Teknologi Bandung, Jalan Ganesha 10 Bandung

E-mail: ramadhianx12@gmail.com , 13524126@std.stei.itb.ac.id

Abstrak—Elektrokardiogram (ECG) merupakan alat diagnostik penting untuk memantau aktivitas listrik jantung. Rekaman ECG jangka panjang menghasilkan data dalam jumlah besar sehingga pemeriksaan manual menjadi tidak praktis. Penelitian ini mengusulkan pendekatan berbasis pencocokan string untuk mendeteksi pola beat abnormal pada sinyal ECG dengan memanfaatkan anotasi beat-by-beat dari MIT-BIH Arrhythmia Database. Setiap rekaman ECG direpresentasikan sebagai string simbolik menggunakan label anotasi seperti N (normal), V (PVC), dan A (atrial premature beat), sehingga permasalahan deteksi anomali dapat direduksi menjadi permasalahan pencocokan string. Dua algoritma dibandingkan dalam penelitian ini, yaitu Knuth-Morris-Pratt (KMP) dan Boyer-Moore, dengan menggunakan lima pattern klinis (VV, VVV, NVN, AV, LR) pada lima rekaman MIT-BIH yang dipilih secara representatif. Metrik evaluasi yang digunakan meliputi jumlah perbandingan karakter, waktu eksekusi, dan jumlah kemunculan pattern. Dari 25 eksperimen yang dilakukan, Boyer-Moore lebih unggul dalam jumlah perbandingan karakter pada 22 eksperimen dengan keunggulan yang meningkat seiring panjang pattern. Namun, KMP secara konsisten lebih cepat dalam waktu eksekusi akibat overhead operasi Boyer-Moore yang lebih besar pada alfabet kecil. Hasil penelitian menunjukkan bahwa pendekatan berbasis pencocokan string efektif untuk deteksi pola abnormal ECG, dengan pilihan algoritma yang bergantung pada karakteristik pattern dan data yang digunakan.

Kata Kunci—ECG, pencocokan string, KMP, Boyer-Moore, MIT-BIH, deteksi aritmia

I. PENDAHULUAN

Elektrokardiogram (ECG) merupakan salah satu alat diagnostik yang paling banyak digunakan dalam dunia medis untuk memantau aktivitas listrik jantung. Sinyal ECG merekam pola detak jantung secara periodik dan menghasilkan gelombang yang memiliki karakteristik tertentu, yaitu gelombang P, kompleks QRS, dan gelombang T. Setiap penyimpangan dari pola normal pada gelombang-gelombang tersebut dapat mengindikasikan adanya gangguan kardiovaskular, seperti aritmia, fibrilasi atrium, atau infark miokard. Karena itulah, kemampuan untuk mendeteksi pola abnormal dalam sinyal ECG secara cepat dan akurat menjadi sangat penting, terutama dalam konteks pemantauan jantung secara *real-time*.

Salah satu cara untuk menyederhanakan permasalahan ini adalah dengan merepresentasikan sinyal ECG dalam bentuk simbolik, di mana setiap detak jantung diwakilkan oleh sebuah simbol berdasarkan jenisnya. Dataset MIT-BIH Arrhythmia Database dari PhysioNet menyediakan rekaman ECG yang sudah dilengkapi anotasi beat-by-beat oleh ahli kardiologi, dengan label seperti N untuk detak normal, V untuk *Premature Ventricular Contraction*, A untuk *Atrial Premature Beat*, dan sebagainya. Dari anotasi ini, seluruh rekaman ECG dapat langsung direpresentasikan sebagai string simbolik, misalnya "NNNVNNNANNN...", tanpa memerlukan transformasi sinyal numerik tambahan. Dengan representasi seperti ini, permasalahan deteksi pola abnormal dapat direduksi menjadi permasalahan pencocokan string, yaitu mencari kemunculan pola tertentu seperti "VV" atau "NVN" di dalam string yang panjang.

Di sinilah algoritma pencocokan string menjadi relevan. Dua algoritma yang paling dikenal dalam bidang ini adalah Knuth-Morris-Pratt (KMP) dan Boyer-Moore. Algoritma KMP bekerja dengan memanfaatkan fungsi pinggir (*border function*) untuk menghindari pengulangan perbandingan karakter yang sudah dilakukan sebelumnya. Sementara itu, Boyer-Moore menggunakan heuristik *looking-glass technique* dan *character-jump technique* yang memungkinkan lompatan lebih jauh saat terjadi ketidakcocokan, dan pada kasus rata-rata cenderung lebih cepat dari KMP terutama untuk alfabet berukuran besar.

Penerapan kedua algoritma ini pada data ECG menarik untuk dikaji karena karakteristik string anotasi yang panjang tetapi menggunakan alfabet yang relatif kecil, yaitu hanya sekitar 20 jenis label beat. Kondisi ini secara teoritis akan mempengaruhi performa Boyer-Moore dibandingkan KMP, karena Boyer-Moore lebih diuntungkan pada alfabet yang besar. Sebuah rekaman MIT-BIH berdurasi 30 menit dapat mengandung ribuan hingga puluhan ribu anotasi beat, sehingga efisiensi pencarian pola tetap menjadi faktor yang kritis untuk dianalisis.

Makalah ini bertujuan untuk mengkaji penerapan algoritma KMP dan Boyer-Moore dalam konteks deteksi pola abnormal pada representasi simbolik sinyal ECG berbasis anotasi MIT-BIH Arrhythmia Database. Pembahasan mencakup pembentukan string simbolik dari anotasi beat, formulasi deteksi

pola abnormal sebagai masalah pencocokan string, implementasi dan perbandingan kinerja kedua algoritma, serta analisis kompleksitas masing-masing dalam konteks ini. Dengan penelitian ini, diharapkan diperoleh gambaran yang jelas mengenai efektivitas pendekatan berbasis pencocokan string untuk mendeteksi kelainan jantung sekaligus perbandingan praktis antara KMP dan Boyer-Moore pada dataset medis nyata.

II. LANDASAN TEORI

A. Sinyal ECG dan Klasifikasi Beat

Elektrokardiogram (ECG) adalah rekaman aktivitas listrik jantung yang dihasilkan oleh elektroda yang ditempelkan pada permukaan tubuh. Sinyal ECG yang normal terdiri dari tiga komponen utama: gelombang P yang merepresentasikan depolarisasi atrium, kompleks QRS yang merepresentasikan depolarisasi ventrikel, dan gelombang T yang merepresentasikan repolarisasi ventrikel. Gangguan pada salah satu komponen ini dapat mengindikasikan kondisi patologis tertentu pada jantung.

MIT-BIH Arrhythmia Database yang tersedia di PhysioNet merupakan salah satu dataset ECG paling banyak digunakan dalam penelitian. Dataset ini terdiri dari 48 rekaman ECG dua kanal dengan durasi masing-masing sekitar 30 menit, direkam pada frekuensi sampling 360 Hz. Setiap rekaman dilengkapi dengan tiga jenis file, yaitu file .dat yang berisi data sinyal mentah, file .hea yang berisi informasi header rekaman, dan file .atr yang berisi anotasi beat-by-beat hasil pemeriksaan dua orang ahli kardiologi secara independen. Tabel I menunjukkan beberapa jenis label anotasi yang umum ditemukan dalam dataset ini.

TABLE I. JENIS LABEL ANOTASI MIT-BIH

Simbol	Keterangan
N	Normal beat
V	Premature Ventricular Contraction (PVC)
A	Atrial Premature Beat
L	Left Bundle Branch Block Beat
R	Right Bundle Branch Block Beat
/	Paced Beat
E	Ventricular Escape Beat
F	Fusion of Ventricular and Normal Beat

Dengan mengekstrak urutan label dari file .atr, seluruh rekaman ECG dapat direpresentasikan sebagai sebuah string simbolik. String inilah yang selanjutnya menjadi masukan bagi algoritma pencocokan string.

B. Pencocokan String

Pencocokan String adalah permasalahan untuk menemukan lokasi kemunculan sebuah pattern di P di dalam Teks T. Secara formal, diberikan teks T dengan panjang n karakter dan pattern P dengan panjang m karakter (dengan asumsi $m \ll n$), tujuannya adalah menemukan semua posisi i dalam T sedemikian sehingga $T[i..i+m-1] = P[0..m-1]$.

T: the rain in spain stays **mainly** on the plain
P: **main**

Gambar. 1. Ilustrasi Pencocokan String

Source:

[https://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2025-2026/23-Pencocokan-string-\(2026\).pdf](https://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2025-2026/23-Pencocokan-string-(2026).pdf)

Pencocokan string memiliki banyak aplikasi nyata, mulai dari fitur pencarian pada text editor, mesin pencari web, analisis citra, hingga bioinformatika untuk pencocokan rantai asam amino pada DNA. Dalam konteks ini, teks T adalah anotasi ECG dan pattern P adalah pola beat abnormal yang ingin dicari, seperti “VV” untuk dua PVC berturut-turut atau “NVN” untuk PVC yang terisolasi diapit beat normal.

C. Algoritma Knuth-Morris-Pratt (KMP)

Algoritma KMP dikembangkan oleh Donald Knuth, Vaughan Pratt, dan James Morris sebagai penyempurnaan dari algoritma *brute force*, tetapi menggunakan informasi dari kegagalan pencocokan sebelumnya untuk menghindari perbandingan yang redundan.

Kunci dari algoritma KMP adalah fungsi pinggiran (*border function*), yang sering juga disebut *failure function*. Fungsi pinggiran $b(k)$ didefinisikan sebagai ukuran prefix terpanjang dari $P[0..k]$ yang juga merupakan suffix dari $P[1..k]$. Fungsi ini dihitung terlebih dahulu dari pattern sebelum pencarian dimulai.

<i>j</i>	0	1	2	3	4	5
<i>P[j]</i>	a	b	a	a	b	a

<i>k</i>	0	1	2	3	4
<i>b(k)</i>	0	0	1	1	2

Gambar. 2. Ilustrasi Border Function dan Tabel Nilainya untuk Pattern P = “abaaba”

Source:

[https://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2025-2026/23-Pencocokan-string-\(2026\).pdf](https://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2025-2026/23-Pencocokan-string-(2026).pdf)

Ketika terjadi ketidakcocokan pada posisi j di pattern, nilai $b(j-1)$ digunakan untuk menentukan posisi baru j tanpa harus menggeser pattern satu per satu. Dengan cara ini, pointer pada teks T tidak perlu mundur, yang menjadi keunggulan KMP untuk pemrosesan data stream.

Kompleksitas waktu KMP terdiri dari dua tahap, yaitu perhitungan *border function* sebesar $O(m)$ dan proses pencarian sebesar $O(n)$, sehingga total kompleksitas adalah $O(m+n)$.

D. Algoritma Boyer-Moore

Algoritma Boyer-Moore dikembangkan oleh Robert Boyer dan J Strother Moore pada tahun 1977. Berbeda dari KMP, Boyer-Moore melakukan perbandingan karakter dari kanan ke kiri pada pattern, dan menggunakan dua teknik utama untuk melakukan lompatan yang lebih efisien saat terjadi ketidakcocokan.

Teknik pertama adalah *looking-glass technique*, yaitu pencocokan dilakukan mulai dari karakter paling kanan pattern menuju ke kiri. Teknik kedua adalah *character-jump technique*, yang menentukan seberapa jauh pattern digeser berdasarkan karakter di teks yang menyebabkan ketidakcocokan. Terdapat tiga kasus dalam *character-jump*:

1. Kasus 1: Jika karakter yang menyebabkan mismatch (sebut x) ada di dalam pattern di sebelah kiri posisi j, geser pattern sehingga kemunculan terakhir x pada pattern sejajar dengan posisi x di teks.
2. Kasus 2: Jika x ada di pattern, tetapi di sebelah kanan posisi j, geser pattern satu posisi ke kanan.
3. Kasus 3: Jika x sama sekali tidak ada di dalam pattern, geser pattern sehingga P[0] sejajar dengan T[i+1].

Untuk mendukung *character-jump*, Boyer-Moore melakukan *preprocessing pattern* dengan membangun *last occurrence function* L(x), yang menyimpan indeks kemunculan terakhir setiap karakter dalam pattern. Jika karakter x tidak ada di dalam pattern, maka L(x) = -1.

L() Example

- A = {a, b, c, d}
- P: "abacab"

P	a	b	a	c	a	b
	0	1	2	3	4	5

x	a	b	c	d
L(x)	4	5	3	-1

Gambar. 3. Ilustrasi Tabel Last Occurrence
Source:

[https://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2025-2026/23-Pencocokan-string-\(2026\).pdf](https://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2025-2026/23-Pencocokan-string-(2026).pdf)

Kompleksitas waktu Boyer-Moore pada kasus terburuk adalah $O(nm + |A|)$ dengan $|A|$ adalah ukuran alfabet, tetapi pada kasus rata-rata untuk teks dengan alfabet besar, algoritma ini bekerja jauh lebih cepat dari KMP. Sebaliknya, untuk alfabet kecil seperti string anotasi ECG yang hanya memiliki sekitar 20 jenis simbol, keunggulan Boyer-Moore berkurang karena peluang terjadinya mismatch yang menghasilkan lompatan jauh menjadi lebih kecil.

E. Perbandingan KMP dan Boyer-Moore

Kedua algoritma memiliki karakteristik yang berbeda dan cocok untuk situasi yang berbeda pula. Tabel II merangkum perbandingan keduanya.

TABLE II. PERBANDINGAN KMP DAN BOYER-MOORE

Aspek	KMP	Boyer-Moore
Arah pencarian pada pattern	Kiri ke kanan	Kanan ke kiri
Preprocessing	Border function $O(m)$	Last occurrence function $O(m + A)$
Kompleksitas worst case	$O(m+n)$	$O(nm + A)$

Aspek	KMP	Boyer-Moore
Kompleksitas average case	$O(m+n)$	$O(n/m)$ sublinear
Performa pada alfabet kecil	Stabil	Menurun
Performa pada alfabet besar	Stabil	Lebih cepat
Cocok untuk data stream	Ya (pointer teks tidak mundur)	Tidak ideal

Dari tabel di atas, terlihat bahwa dalam konteks string anotasi ECG yang beralfabet kecil, KMP berpotensi memberikan performa yang lebih konsisten, sementara Boyer-Moore mungkin tidak mendapatkan keuntungan maksimalnya.

III. METODOLOGI

A. Dataset dan Persiapan Data

Dataset yang digunakan dalam penelitian ini adalah MIT-BIH Arrhythmia Database yang tersedia secara publik melalui platform PhysioNet. Dataset ini dipilih karena merupakan standar de facto dalam penelitian deteksi aritmia dan sudah digunakan secara luas oleh komunitas riset di bidang ini. Dataset MIT-BIH terdiri dari 48 rekaman ECG dua kanal dengan durasi masing-masing sekitar 30 menit, direkam pada frekuensi sampling 360 Hz. Setiap rekaman dilengkapi anotasi beat-by-beat yang dibuat oleh dua orang ahli kardiologi secara independen, sehingga label yang tersedia dapat dianggap sebagai ground truth yang kredibel.

Dari 48 rekaman yang tersedia, penelitian ini memilih 5 rekaman yang dianggap representatif berdasarkan variasi komposisi beat-nya. Pemilihan ini bertujuan agar pengujian mencakup kondisi yang bervariasi, mulai dari rekaman yang didominasi beat normal hingga rekaman yang mengandung banyak aritmia. Tabel III menunjukkan rekaman yang dipilih beserta komposisi beat aktualnya setelah ekstraksi.

TABLE III. REKAMAN MIT-BIH YANG DIGUNAKAN

Rekaman	Deskripsi	Total Beat	N	V	A	Lainnya
100	Mayoritas normal	2273	2239	1	33	0
106	Banyak PVC	2027	1507	520	0	0
108	Campuran aritmia	1762	1739	17	4	2
200	Banyak PVC dan aberrant	2601	1743	826	30	2
210	Campuran kompleks	2628	2423	194	0	11

B. Ekstraksi dan Pembentukan String Simbolik

Setiap rekaman MIT-BIH terdiri dari tiga jenis file, yaitu file .dat berisi sinyal ECG mentah dalam format biner, file .hea berisi informasi header rekaman, dan file .atr berisi anotasi beat-by-beat. Dalam penelitian ini, hanya file .atr yang digunakan karena informasi yang dibutuhkan adalah urutan label beat, bukan sinyal numeriknya.

Proses ekstraksi dilakukan menggunakan library wfdb versi 4.3.1 pada Python, yang merupakan library resmi yang dikembangkan oleh PhysioNet untuk membaca format file WFDB. Data diunduh secara otomatis dari server PhysioNet menggunakan fungsi wfdb.rdann(). Setelah anotasi berhasil dimuat, dilakukan proses filtering untuk mempertahankan hanya simbol beat yang relevan secara klinis, yaitu N (normal), V (PVC), A (atrial premature), L (left bundle branch block), R (right bundle branch block), / (paced beat), E (ventricular escape), dan F (fusion beat). Simbol-simbol lain yang bukan merupakan label beat individual, seperti + (rhythm change) dan ~ (signal quality), dibuang karena tidak merepresentasikan jenis detak jantung.

Hasil dari proses ini adalah sebuah string simbolik per rekaman yang merepresentasikan urutan seluruh beat dalam rekaman tersebut. Sebagai contoh, rekaman 106 menghasilkan string sepanjang 2027 karakter yang di dalamnya terdapat karakter N dan V sesuai urutan kemunculannya. String inilah yang selanjutnya menjadi input teks (T) bagi algoritma pencocokan string.

C. Penentuan Pattern

Pattern yang digunakan dalam penelitian ini dipilih berdasarkan relevansi klinisnya sebagai indikator kondisi jantung tertentu. Penelitian ini menggunakan lima pattern dengan panjang yang bervariasi antara 2 hingga 3 karakter, sebagaimana ditunjukkan pada Tabel IV.

TABLE IV. PATTERN YANG DIGUNAKAN DALAM PENGUJIAN

Pattern	Panjang	Makna Klinis
VV	2	Dua PVC berturut-turut, indikator potensi takikardia ventrikel
VVV	3	Tiga PVC berturut-turut, kondisi lebih berbahaya
NVN	3	PVC terisolasi diapit beat normal, pola PVC paling umum
AV	2	Atrial premature beat diikuti PVC
LR	2	Kombinasi left dan right bundle branch block

Variasi panjang pattern dipilih secara sengaja untuk mengamati bagaimana panjang pattern memengaruhi jumlah perbandingan karakter pada masing-masing algoritma.

D. Implementasi Algoritma

Kedua algoritma diimplementasikan dari awal menggunakan Python tanpa menggunakan library pencocokan string bawaan seperti str.find() atau modul re. Hal ini dilakukan agar jumlah perbandingan karakter dapat dihitung secara eksplisit pada setiap langkah algoritma.

Implementasi KMP terdiri dari dua fungsi utama. Pertama, fungsi compute_border_function() yang menghitung tabel border function dari pattern sebelum pencarian dimulai. Kedua, fungsi kmp_search() yang melakukan pencarian pada teks menggunakan tabel tersebut. Counter perbandingan diinkrementasi setiap kali terjadi perbandingan karakter antara

pattern dan teks, baik yang menghasilkan kecocokan maupun ketidakcocokan.

Implementasi Boyer-Moore juga terdiri dari dua fungsi utama. Pertama, fungsi build_last_occurrence() yang membangun tabel last occurrence function dari pattern dan alfabet. Kedua, fungsi boyer_moore_search() yang melakukan pencarian dengan arah kanan ke kiri pada pattern menggunakan mekanisme character-jump. Pergeseran dihitung menggunakan rumus $s += \max(1, j - lo)$ di mana j adalah posisi mismatch pada pattern dan lo adalah nilai last occurrence dari karakter yang menyebabkan mismatch. Untuk mendukung pencarian kemunculan yang saling tumpang tindih, ketika pattern ditemukan, window digeser satu posisi ke kanan alih-alih digeser sebesar panjang pattern.

E. Pengukuran dan Metrik Evaluasi

Untuk membandingkan performa kedua algoritma, tiga metrik utama diukur pada setiap eksperimen:

Pertama, jumlah perbandingan karakter, yaitu total berapa kali algoritma membandingkan satu karakter pattern dengan satu karakter teks. Metrik ini merupakan ukuran paling langsung dari efisiensi algoritma yang tidak bergantung pada kecepatan perangkat keras. Kedua, waktu eksekusi dalam milidetik, diukur menggunakan fungsi time.perf_counter(). Ketiga, jumlah kemunculan pattern yang ditemukan, yang berfungsi sebagai validasi kebenaran kedua algoritma.

Selain itu, dihitung juga comparison ratio, yaitu perbandingan jumlah perbandingan KMP terhadap Boyer-Moore, dan time ratio yaitu perbandingan waktu eksekusi KMP terhadap Boyer-Moore. Nilai rasio lebih dari 1 berarti KMP membutuhkan lebih banyak perbandingan atau waktu lebih lama dibandingkan Boyer-Moore, dan sebaliknya.

IV. HASIL DAN ANALISIS

A. Hasil Pencocokan String

Pengujian dilakukan terhadap 5 rekaman MIT-BIH dengan 5 pattern, menghasilkan total 25 eksperimen. Seluruh hasil menunjukkan bahwa kedua algoritma menemukan posisi kemunculan yang identik pada semua eksperimen, yang membuktikan kebenaran implementasi keduanya. Tabel V merangkum jumlah kemunculan pattern yang ditemukan pada masing-masing rekaman.

TABLE V. JUMLAH KEMUNCULAN PATTERN PER REKAMAN

Rekaman	VV	VVV	NVN	AV	LR
100	0	0	1	0	0
106	97	1	327	0	0
108	0	0	15	0	0
200	57	8	718	1	0
210	22	8	155	0	0

Dari tabel di atas, terlihat bahwa pattern NVN memiliki jumlah kemunculan paling banyak di semua rekaman yang

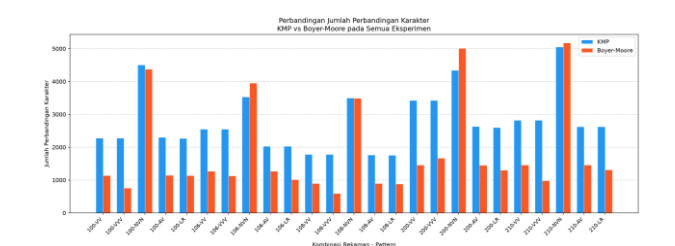
mengandung PVC. Hal ini sesuai dengan karakteristik klinis PVC, di mana PVC yang terisolasi dan diapit beat normal merupakan pola yang paling umum terjadi pada pasien aritmia. Rekaman 200 yang memiliki proporsi PVC tertinggi (826 dari 2601 beat) juga menunjukkan jumlah kemunculan NVN tertinggi, yaitu 718 kali. Sebaliknya, pattern LR tidak ditemukan di rekaman manapun karena tidak ada rekaman yang mengandung kombinasi left dan right bundle branch block secara bersamaan dalam dataset yang dipilih.

B. Perbandingan Jumlah Perbandingan Karakter

Metrik utama yang dianalisis adalah jumlah perbandingan karakter karena metrik ini mencerminkan efisiensi algoritma secara teoritis tanpa dipengaruhi faktor perangkat keras. Tabel VI menunjukkan jumlah perbandingan karakter KMP dan Boyer-Moore untuk setiap kombinasi rekaman dan pattern.

TABLE VI. JUMLAH PERBANDINGAN KARAKTER KMP VS BOYER-MOORE

Rekam an	Patter n	Panjang g	Panjang g Teks	KMP	BM	Rasio KMP/ BM
100	VV	2	2273	2274	1136	2.00
100	VVV	3	2273	2274	757	3.00
100	NVN	3	2273	4511	4377	1.03
106	VV	2	2027	2574	1270	2.01
106	VVV	3	2027	2547	1129	2.26
106	NVN	3	2027	3533	3953	0.89
108	VV	2	1762	1779	899	1.98
108	VVV	3	1762	1779	591	3.01
108	NVN	3	1762	3500	3492	1.00
200	VV	2	2601	3427	1459	2.35
200	VVV	3	2601	3427	1667	2.06
200	NVN	3	2601	4343	5007	0.87
210	VV	2	2628	2822	1462	1.93
210	VVV	3	2628	2822	985	2.87
210	NVN	3	2628	5050	5176	0.98



Gambar. 4. Bar Chart Perbandingan Jumlah Perbandingan Karakter KMP vs BM

Dari Tabel VI, terdapat dua pola yang cukup jelas. Pertama, untuk pattern yang tidak atau jarang ditemukan dalam teks (VV, VVV, AV, LR), Boyer-Moore secara konsisten melakukan lebih sedikit perbandingan dibandingkan KMP. Rasionya berkisar antara 1.93 hingga 3.01, yang artinya KMP melakukan hampir dua hingga tiga kali lebih banyak perbandingan dari Boyer-Moore. Ini terjadi karena ketika pattern tidak banyak

ditemukan, Boyer-Moore dapat melakukan lompatan jauh setiap kali terjadi *mismatch* berkat mekanisme *character-jump*.

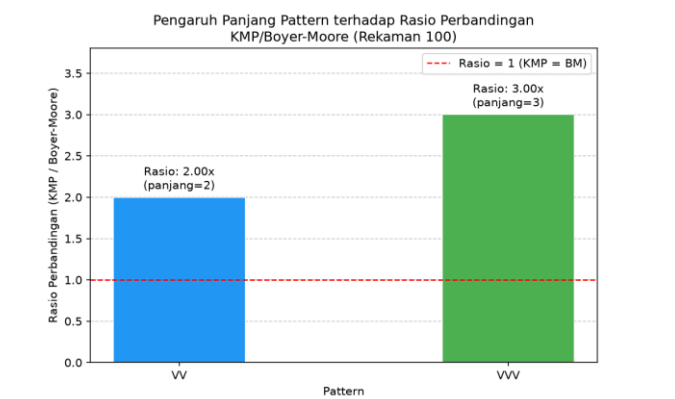
Kedua, untuk pattern NVN yang sangat sering ditemukan, situasinya berbalik. Boyer-Moore justru melakukan lebih banyak atau hampir sama banyaknya perbandingan dengan KMP, bahkan melebihi KMP di rekaman 106 dan 200 dengan rasio 0.89 dan 0.87. Hal ini terjadi karena ketika pattern sering ditemukan, window Boyer-Moore sering hanya bergeser satu posisi, sehingga keuntungan lompatan jauh dari *character-jump* tidak bisa dimanfaatkan.

C. Pengaruh Panjang Pattern terhadap Efisiensi Boyer-Moore

Salah satu aspek menarik yang terlihat dari data adalah pengaruh panjang pattern terhadap selisih perbandingan antara KMP dan Boyer-Moore. Tabel VII menampilkan rasio perbandingan berdasarkan panjang pattern, diambil dari rekaman yang pattern-nya tidak ditemukan, sehingga faktor frekuensi kemunculan tidak mempengaruhi hasil.

TABLE VII. PENGARUH PANJANG PATTERN TERHADAP RASIO PERBANDINGAN (REKAMAN 100)

Pattern	Panjang	KMP	BM	Rasio KMP/BM
VV	2	2274	1136	2.00
VVV	3	2274	757	3.00



Gambar. 5. Bar Chart Rasio KMP/BM untuk Pattern VV dan VVV di Rekaman 100

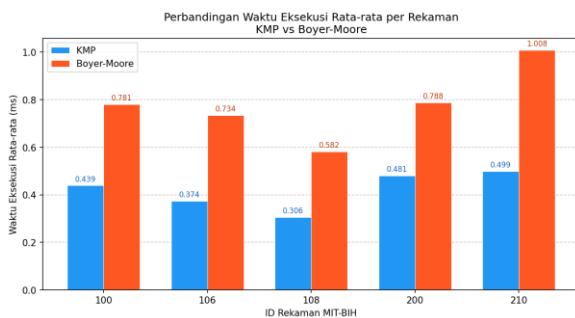
Dari Tabel VII terlihat bahwa untuk pattern panjang 2 karakter, Boyer-Moore melakukan sekitar setengah dari jumlah perbandingan KMP. Ketika panjang pattern bertambah menjadi 3 karakter, Boyer-Moore hanya melakukan sepertiga dari jumlah perbandingan KMP. Pola ini sesuai dengan teori Boyer-Moore, di mana semakin panjang pattern, semakin besar lompatan yang bisa dilakukan saat terjadi *mismatch* pada karakter pertama yang diperiksa dari kanan. Dengan kata lain, keunggulan Boyer-Moore atas KMP dalam hal jumlah perbandingan cenderung meningkat seiring bertambahnya panjang pattern.

D. Perbandingan Waktu Eksekusi

Selain jumlah perbandingan karakter, waktu eksekusi juga diukur untuk melihat performa aktual kedua algoritma. Tabel VIII merangkum waktu eksekusi rata-rata per rekaman.

TABLE VIII. WAKTU EKSEKUSI RATA-RATA PER REKAMAN (MS)

Rekaman	KMP(ms)	BM(ms)	Rasio KMP/BM
100	0.439	0.781	0.562
106	0.374	0.734	0.510
108	0.306	0.582	0.526
200	0.481	0.788	0.610
210	0.499	1.008	0.495
Rata-rata	0.420	0.779	0.539



Gambar. 6. Grouped Bar Chart Waktu Eksekusi per Rekaman

Hasil waktu eksekusi menunjukkan hasil yang sedikit kontradiktif dengan hasil jumlah perbandingan. Meskipun Boyer-Moore melakukan lebih sedikit perbandingan karakter pada sebagian besar kasus, KMP justru secara konsisten lebih cepat dari sisi waktu eksekusi pada seluruh 25 eksperimen. Rata-rata waktu KMP adalah 0.420 ms sedangkan Boyer-Moore rata-rata 0.779 ms, atau sekitar 1.85 kali lebih lambat.

Hal ini dapat dijelaskan oleh perbedaan *overhead* per iterasi antara kedua algoritma. Dalam setiap langkah ketika terjadi *mismatch*, Boyer-Moore perlu melakukan lookup ke tabel last occurrence function `last.get(x, -1)`, menghitung nilai lompatan, dan baru menggeser pointer. Sementara KMP hanya perlu mengakses satu elemen array border function `b[j-1]`. Pada string dengan alfabet kecil seperti anotasi ECG yang hanya memiliki sekitar 8 jenis simbol, *overhead* ini menjadi relatif signifikan karena lompatan yang dihasilkan pun tidak terlalu besar, sehingga frekuensi operasi lookup tetap tinggi.

E. Analisis Kasus Pattern NVN

Pattern NVN layak mendapat perhatian tersendiri karena menunjukkan perilaku yang paling berbeda dibandingkan pattern lainnya. Pada rekaman 200 misalnya, NVN ditemukan sebanyak 718 kali dari total 2601 karakter, yang berarti hampir setiap tiga karakter sekali terdapat kemunculan pattern ini. Dalam kondisi seperti ini, Boyer-Moore tidak bisa memanfaatkan lompatan jauh karena setiap kali menemukan karakter N atau V, kemungkinan besar karakter tersebut ada

dalam pattern, sehingga nilai last occurrence-nya tinggi dan lompatan yang dihasilkan kecil.

Konsekuensinya, Boyer-Moore justru melakukan lebih banyak perbandingan dibandingkan KMP pada kasus ini. Rasio perbandingan KMP/BM untuk NVN berkisar antara 0.87 hingga 1.03, yang artinya keduanya hampir setara atau Boyer-Moore bahkan sedikit lebih boros. Ini merupakan contoh nyata dari kelemahan teoritis Boyer-Moore pada teks dengan alfabet kecil dan pattern yang sering muncul, yang sebelumnya hanya dibahas secara teori di Landasan Teori.

Tabel IX merangkum perbandingan keseluruhan antara KMP dan Boyer-Moore berdasarkan seluruh eksperimen yang telah dilakukan.

TABLE IX. RINGKASAN PERBANDINGAN KMP DAN BOYER-MOORE

Aspek	KMP	Boyer-Moore	Pemenang
Rata-rata perbandingan karakter	2771.4	1809.2	Boyer-Moore
Waktu eksekusi rata-rata	0.420 ms	0.779 ms	KMP
Konsistensi pada alfabet kecil	Stabil	Menurun	KMP
Keunggulan pada pattern panjang	Tidak signifikan	Signifikan	Boyer-Moore
Keandalan hasil	25/25	25/25	Seri

V. KESIMPULAN

Penelitian ini menunjukkan bahwa representasi simbolik berbasis anotasi beat dari MIT-BIH Arrhythmia Database dapat digunakan secara efektif sebagai input untuk algoritma pencocokan string dalam mendeteksi pola abnormal pada sinyal ECG. Dari 25 eksperimen yang dilakukan, kedua algoritma menghasilkan posisi kemunculan yang identik, yang membuktikan kebenaran implementasi keduanya. Dalam hal jumlah perbandingan karakter, Boyer-Moore lebih unggul pada 22 dari 25 eksperimen dengan keunggulan yang semakin besar seiring bertambahnya panjang pattern. Namun dari sisi waktu eksekusi, KMP justru lebih cepat di seluruh eksperimen karena *overhead* operasi Boyer-Moore per iterasi yang lebih besar, terutama pada alfabet kecil seperti string anotasi ECG yang hanya memiliki sekitar 8 jenis simbol aktif.

Secara keseluruhan, pemilihan algoritma untuk kasus serupa sebaiknya mempertimbangkan karakteristik data yang digunakan. Boyer-Moore lebih cocok ketika pattern yang dicari panjang dan jarang muncul dalam teks, sementara KMP lebih stabil dan konsisten untuk berbagai kondisi, khususnya pada domain dengan alfabet kecil. Untuk penelitian selanjutnya, pengujian dapat diperluas dengan menambah jumlah rekaman, menggunakan pattern yang lebih panjang, atau membandingkan dengan algoritma lain seperti Rabin-Karp untuk mendapatkan gambaran yang lebih komprehensif.

REPOSITORY LINK AT GITHUB

<https://github.com/Rmadhian/kmp-bm-ecg-detection>

ACKNOWLEDGMENT

Puji dan syukur penulis panjatkan ke hadirat Allah SWT. atas rahmat dan karunia-Nya sehingga makalah ini dapat diselesaikan dengan baik. Penulis menyampaikan terima kasih kepada Dr. Ir. Rinaldi Munir, M.T., Dr. techn. Muhammad Zuhri Catur Candra, S.T., M.T., Tricya Esterina Widagdo, S.T., M.Sc., dan Dr. Fazat Nur Azizah, S.T., M.Sc., selaku dosen mata kuliah IF2211 Strategi Algoritma, atas bimbingan dan ilmu yang diberikan selama semester ini sehingga penulis dapat menyelesaikan penelitian ini.

REFERENCES

- [1] R. Munir, "Pencocokan String (String/Pattern Matching)," Bahan Kuliah IF2211 Strategi Algoritma, Program Studi Teknik Informatika STEI-ITB, 2026. [Online]. Available: [https://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2025-2026/23-Pencocokan-string-\(2026\).pdf](https://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2025-2026/23-Pencocokan-string-(2026).pdf)
- [2] G. B. Moody and R. G. Mark, "The MIT-BIH Arrhythmia Database on CD-ROM and software for use with it," *Computers in Cardiology*, vol. 17, pp. 185–188, 1990.
- [3] R. G. Mark, P. S. Schluter, G. B. Moody, P. H. Devlin, and D. Chernoff, "An annotated ECG database for evaluating arrhythmia detectors," *IEEE Transactions on Biomedical Engineering*, vol. 29, no. 8, p. 600, 1982.

- [4] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. C. Ivanov, R. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Bandung, 19 Juni 2026



Ramadhian Nabil Firdaus Gumay, 13524126