
Pengenalan Tulisan Tangan Huruf Hangeul Menggunakan Metode *Hidden Markov Model*

Jesen Sandywan¹, Cristin Adiya Chandra², Derry Alamsyah³

^{1,2}STMIK GI MDP; Jl. Rajawali No.14, +62(711) 376400/376300

³Program Studi Teknik Informatika, STMIK GI MDP, Palembang

e-mail: jesen@mhs.mdp.ac.id, cristinchandraa@mhs.mdp.ac.id, derry@mdp.ac.id

Abstrak

Setiap individu manusia memiliki bentuk tulisan yang berbeda-beda. Hal ini yang menyebabkan bentuk karakter tulisan tangan yang dimiliki masing-masing orang unik. Pengenal tulisan tangan (*handwriting*) huruf *hangeul* adalah proses secara otomatis mendeteksi huruf *hangeul* yang terdapat pada foto tulisan tangan. Sistem pengenal tulisan tangan dibentuk melalui dua proses utama yaitu pelatihan (*training*) dan pengujian (*testing*). Penelitian ini menggunakan metode *Hidden Markov Model* (HMM) dengan algoritma *Forward-Backward* untuk mendeteksi tulisan tangan huruf *hangeul*. Tujuan penelitian ini adalah menerapkan dan menguji performa *Hidden Markov Model* pada sistem pengenal tulisan tangan huruf *hangeul*. Hasil penelitian menunjukkan bahwa pengujian secara keseluruhan mendapatkan tingkat akurasi sebesar 22.083%.

Kata kunci— Pengenal Tulisan Tangan, Algoritma *Forward Backward*, *Hidden Markov Model*

Abstract

Every human being has its own unique shape and structure of handwriting. Handwriting recognition of Hangeul is the process that automatically detects Hangeul letters contained from photograph image of handwriting. The handwriting recognition system is formed through two main processes; training and testing. This research used Hidden Markov Model (HMM) method with Forward Backward algorithm to detect the handwriting Hangeul letter. The purpose of this research is to implement and test the performance of the Hidden Markov Model in the handwriting recognition system of Hangeul letters. The results of this study showed that the overall test had an accuracy rate of 22.083%.

Keywords— Handwriting Recognition, Forward Backward Algorithm, Hidden Markov Model

1. PENDAHULUAN

Di Indonesia budaya Korea yang disebut dengan *Korean Wave* ini cukup banyak digemari oleh banyak orang mulai dari anak-anak hingga orang tua. Bahkan beberapa penggemar membuat kelompok yang disebut dengan *fanbase*. *Fanbase* merupakan kumpulan dari banyak orang yang mendukung idolanya. Bahkan, banyak penggemar di Indonesia yang mengikuti sang idola, seperti *fashion*. Tak hanya *fashion* para penggemar juga mengikuti Bahasa yang digunakan oleh sang idola. Bahasa yang digunakan oleh sang idola adalah Bahasa Korea. Bahasa Korea biasa disebut dengan *Hangeul* terbagi menjadi 2 yaitu Vocal dan Konsonan. Bahasa Korea adalah Bahasa resmi yang dapat digunakan di Korea Selatan dan Korea Utara.

Huruf *hangeul* saat ini adalah huruf yang cukup populer di kalangan masyarakat. Dibutuhkan sebuah sistem untuk mengetahui tulisan tangan *hangeul* menjadi alfabet. Banyak penelitian yang menggunakan kecerdasan buatan untuk mengenali suatu objek. Penelitian tentang pengenalan tulisan tangan huruf *hangeul* pernah dilakukan dengan menggunakan metode Jaringan Saraf Tiruan *Backpropagation* oleh Sulysthian tahun 2018 yang mendapatkan tingkat akurasi sebesar 80,4763%. Penelitian tulisan tangan huruf *hangeul* belum pernah dilakukan dengan menggunakan metode *Hidden Markov Model* dan belum diketahui tingkat akurasi yang dihasilkan dari metode tersebut.

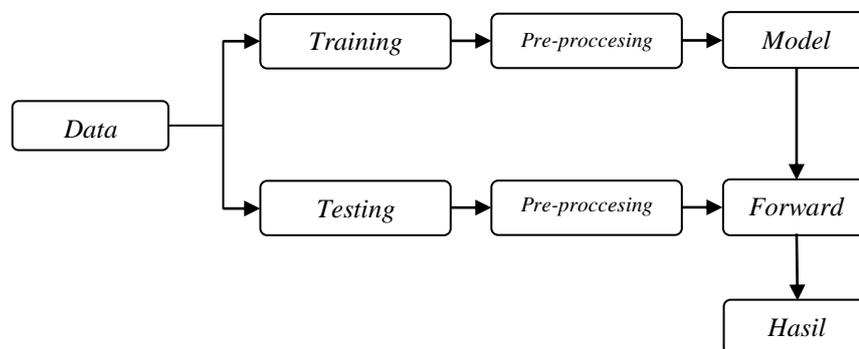
Pada penelitian sebelumnya pengenalan tulisan tangan huruf *Hangeul* dilakukan dengan menggunakan metode Jaringan Saraf Tiruan *Backpropagation* sebagai pemodelan sistem dan *Intensiry of Character* dan *Mark Direction* sebagai metode ekstraksi cirinya. Pada percobaan proses pengenalan tulisan tangan huruf *Hangeul* ini mendapatkan tingkat akurasi sekitar 80,4762%. Percobaan yang dilakukan menggunakan 2 lapisan tersembunyi dengan jumlah neuron pada lapisan tersembunyi 1 sebanyak 100 neuron. dan pada lapisan 2 sebanyak 95 neuron [1].

Pada penelitian sebelumnya pengenalan karakter huruf tulisan tangan *Alfabet* menggunakan metode *Principal Components Analysis* mendapatkan tingkat keberhasilan atau akurasi yang cukup tinggi yaitu mencapai 88,46%. Dengan mempengaruhi banyaknya data training dan data uji yang digunakan untuk mereduksi dimensi yang besar dari ruang data (*observed variables*) menjadi dimensi intrinsik yang lebih kecil dari ruang fitur (*independent variables*), yang dibutuhkan untuk mendeskripsikan data lebih sederhana [2].

Karena belum ditemukan publikasi penelitian pengenalan tulisan tangan huruf *Hangeul* menggunakan metode *Hidden Markov Model* maka penelitian ini penting untuk dilakukan.

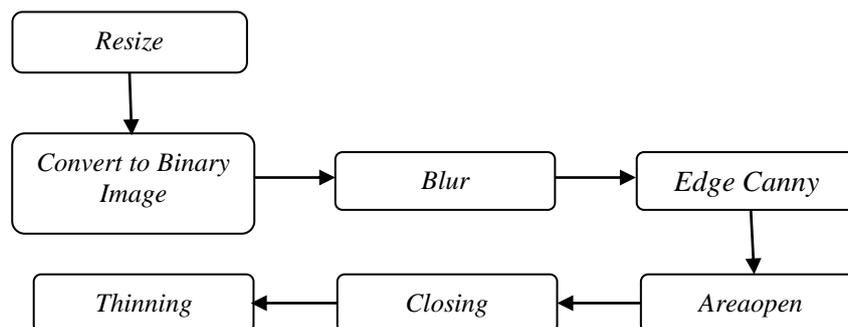
2. METODE PENELITIAN

Pada penelitian ini dilakukan dengan menggunakan beberapa tahapan yang harus dilakukan secara berurutan guna mencapai hasil yang maksimal seperti studi literatur, pengumpulan data, perancangan sistem, implementasi dan evaluasi. Pada tahapan perancangan sistem dalam pengenalan tulisan tangan huruf *hangeul* menggunakan metode *hidden markov model* dapat dilihat pada Gambar 2.1.



Gambar 2.1 Diagram Langkah Penelitian

Pada Gambar 2.1 menjelaskan proses dari metode *hidden markov model*. Penulis melakukan pengumpulan data sebanyak 60 responden dimana masing-masing responden menggores huruf *hangeul* sebanyak 24 huruf, kemudian data yang telah terkumpul dibagi menjadi dua kelompok yaitu *training* dan *testing*. Data *training* dan data *testing* sama-sama melakukan proses *pre-processing*, setelah melewati tahapan *pre-processing*, data *training* membentuk *model ergodic* terhadap masing-masing huruf. Setelah data *testing* melewati tahapan *pre-processing* selanjutnya data *testing* melakukan proses pengujian menggunakan algoritma *forward-backward* dengan mengambil model yang telah dibentuk oleh data *training* selanjutnya menghasilkan output berupa huruf alfabet.



Gambar 2.2 Diagram Alir Preprocessing

Pada tahap *preprocessing*, peneliti melakukan beberapa tahap seperti pada Gambar 2.2 yang akan dijelaskan sebagai berikut:

1. *Resize*
Resize bertujuan agar setiap citra yang diproses memiliki ukuran yang sama, hal ini bertujuan untuk meningkatkan akurasi dari penelitian ini. Pixel yang dibutuhkan pada penelitian ini berukuran 100x100 pixel.
2. *Convert to Binary Image*
 Pada tahap ini dilakukan perubahan terhadap citra RGB menjadi citra biner bertujuan untuk mempermudah pengolahan citra dimana hanya terdapat nilai 0 dan 1 saja.
3. *Blur*
 Pada tahap ini dilakukan pemberian efek kabur terhadap citra RGB yang telah diproses menjadi citra biner bertujuan untuk mereduksi *noise* pada pengolahan citra seperti Gambar 2.3.



Gambar 2.3 Hasil Citra Blur

4. *Edge Canny*

Pada tahap ini adalah proses lanjutan dari hasil tahap blur yang selanjutnya dilakukan pendeteksian tepi dengan operator *canny* hasilnya dapat dilihat pada Gambar 2.4.



Gambar 2.4 Hasil Citra *Edge Canny*

5. *Area Open*

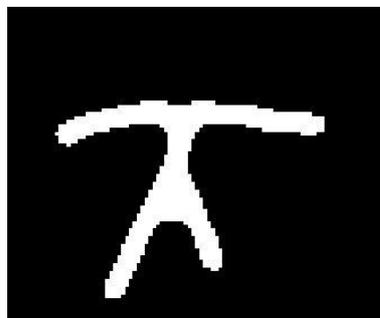
Setelah melewati proses *edge canny* dilakukan proses *areaopen* yang bertujuan untuk menghilangkan pixel yang bukan bagian dari karakter, yaitu dengan cara melihat jumlah pixel tersebut pada suatu area. Apabila kurang dari nilai batas yang ditentukan, maka pixel tersebut dijadikan *background* seperti Gambar 2.5.



Gambar 2.5 Hasil Citra *Area Open*

6. *Closing*

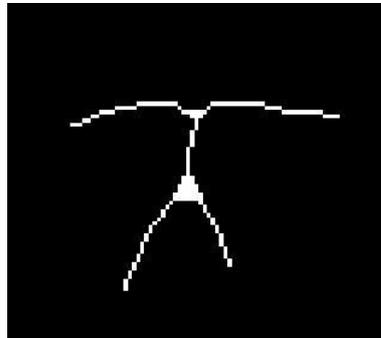
Pada tahap ini dilakukan proses *closing* dari hasil *edge canny* yang bertujuan untuk melakukan penutupan morfologis (pelebaran diikuti oleh erosi) seperti pada pada Gambar 2.6.



Gambar 2.6 Hasil Citra *Closing*

7. *Thinning*

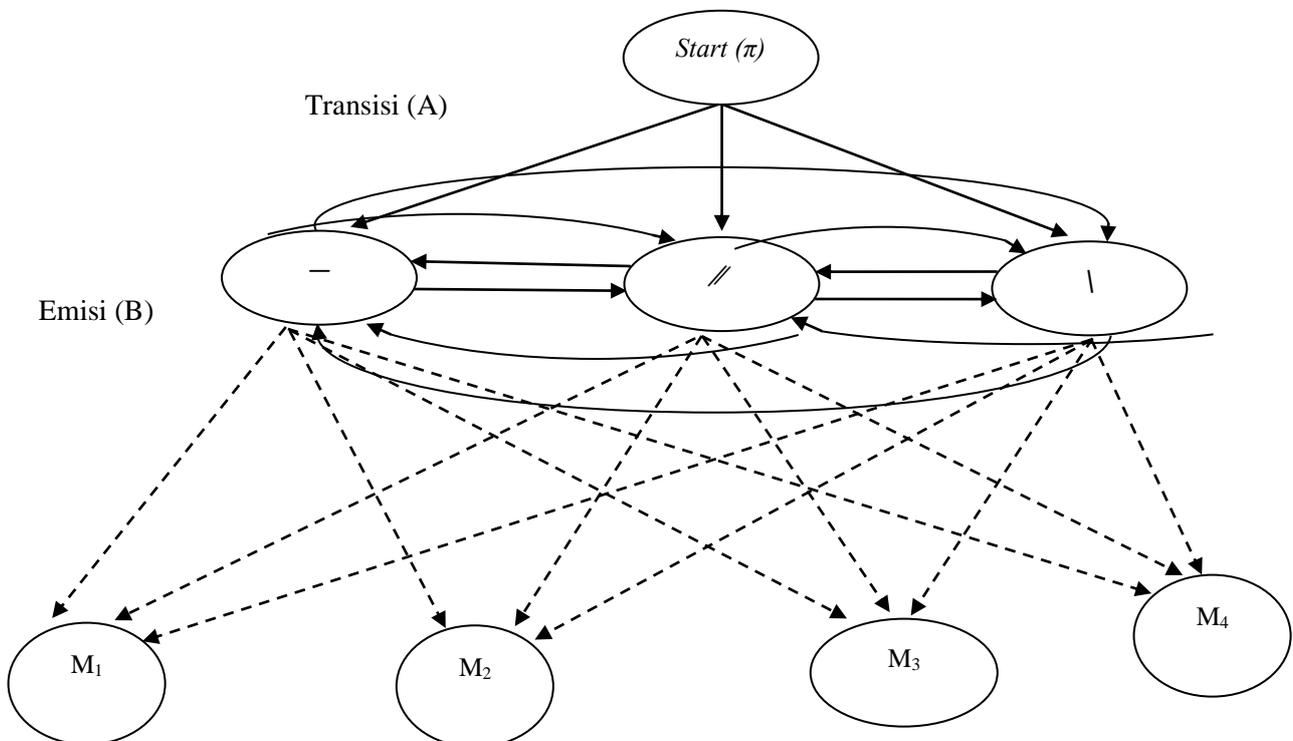
Pada tahap ini dilakukan proses *thinning* dari hasil *closing* yang bertujuan untuk mempertipis suatu citra menjadi satu pixel seperti Gambar 2.7.



Gambar 2.7 Hasil Citra *Thinning*

3. HASIL DAN PEMBAHASAN

Pada tahap ini dilakukan pembentukan HMM yang ditentukan dengan cara melihat *hidden state* (S) dan N adalah jumlah *hidden state* yang terdiri dari inialisasi, transisi dan emisi. Inialisasi didapat dari seluruh goresan yang dilakukan oleh responded, transisi didapat dari jumlah goresan yang digunakan dan emisi terdiri dari 5 *hidden state* yaitu $M_1, M_2, M_3, M_4,$ dan M_5 . Dapat dilihat pada Gambar 3.1.



Gambar 3.1 Model *Ergodic* ☆

Pada Gambar 3.1 menjelaskan model *ergodic* dari huruf *hangeul* ㄸ(CH) yang terdiri dari inisialisasi (π) yang biasa disebut start, transisi (A) dan emisi (B) sebagai *hidden*. Inisialisasi didapat dari jumlah seluruh goresan yang dilakukan oleh responden, transisi didapat dari urutan goresan yang dilakukan oleh responden dan emisi didapat dari gradien yang dihasilkan oleh Matlab.

Tabel 3.1 Nilai Transisi Huruf *Hangeul* ㄸ

Hasil Kondisi	—	/	\
—	30	50	0
/	0	0	50
\	20	0	0

Tabel 3.1 Menjelaskan nilai transisi huruf *hangeul* ㄸ yang didapat dari pengumpulan data set terdiri dari 50 orang dengan melihat urutan goresan yang terdapat pada lembar kuisisioner.

Tabel 3.2 Add One Smoothing Transisi Huruf *Hangeul* ㄸ

Hasil Kondisi	—	/	\
—	0,3735	0,6145	0,0120
/	0,0189	0,0189	0,9623
\	0,9130	0,0434	0,0434

Tabel 3.2 Menjelaskan hasil dari nilai *add one smoothing* transisi huruf *hangeul* ㄸ yang telah didapat dari pengumpulan data set terdiri dari 50 orang, dengan cara menambahkan 1/3 (sesuai dengan banyak *hidden state* yang digunakan).

Tabel 3.3 Nilai Emisi Huruf *Hangeul* ㄸ

Hasil Kondisi	M ₁	M ₂	M ₃	M ₄
—		112	70	
/	36			
\				86

Tabel 3.3 Menjelaskan nilai emisi huruf *hangeul* ㄸ yang didapat dari pengumpulan data set terdiri dari 50 orang dengan melihat nilai yang didapat dari matlab kemudian digolongkan menjadi 5 kelompok gradien.

Tabel 3.4 Add One Smoothing Emisi Huruf Hangeul ㄸ

Hasil \ Kondisi	M ₁	M ₂	M ₃	M ₄
—	0,0054	0,6075	0,3817	0,0054
/	0,9250	0,0250	0,0250	0,0250
\	0,0111	0,0111	0,0111	0,9667

Tabel 3.4 menjelaskan hasil dari nilai *add one smoothing* emisi huruf *hangeul* ㄸ yang telah didapat dari pengumpulan data set terdiri dari 50 orang, dengan cara menambahkan 1/4 (sesuai dengan banyak *hidden state* yang digunakan).

Tabel 3.5 Inisialisasi Huruf Hangeul ㄸ

Hasil \ Kondisi	—	/	\
Start	0,500	0,250	0,250

Tabel 3.5 Menjelaskan hasil dari nilai peluang awal huruf *hangeul* ㄸ yang telah didapat dari pengumpulan data set terdiri dari 50 orang, selengkapnya dapat dilihat pada lampiran.

Pengujian dilakukan pada 10 data test dengan melalui proses *Preprocessing* yaitu, *Resize RGB, Blur, Edge Canny, Area Open, Closing, Thinning*, untuk mendapatkan garis-garis atau titik-titik koordinasi dari setiap obyek. Dapat dilihat seperti Tabel 3.6.

Tabel 3.6 Pengujian Garis untuk Setiap Object

No.	Input	Garis Setiap Obyek	No.	Input	Garis Setiap Obyek
1			6		
2			7		
3			8		
4			9		

No.	Input	Garis Setiap Obyek	No.	Input	Garis Setiap Obyek
5			10		

Pada Tabel 3.6 nomor satu didapatkan beberapa garis dan gradien dari obyek tersebut setelah diproses menggunakan matlab, garis hijau menunjukkan goresan yang dideteksi oleh matlab seperti pada Tabel 3.7.

Tabel 3.7 Gradien dari Object 大

No. Pengujian	Gradien					
	1	2	3	4	5	6
1	1,666667	1,623406	0	-0,04762	-1,16667	-
2	-0,08571	-0,10256	-1,06667	-1,13043	-0,04348	0,9
3	-0,06818	1,411765	-2,08333	0	-	-
4	-0,18182	-0,2	-1,89474	-0,22857	-1,64286	-0,08333
5	-0,06522	-1,53333	-1,93333	-1,66667	-2,23077	-
6	-0,04545	0,038462	-1,3913	0,956522	1,117647	-
7	-0,11765	1,222222	-1,16667	-0,09375	1	-
8	-2,35714	-0,23529	-0,17391	0,95	-	-
9	-0,04348	-0,08824	-0,1	0	-0,12	0
10	0,1	0	-0,05	0,058824	-5,25	-

Tabel 3.8 menunjukkan *gradien* yang terdeteksi di setiap *object*, *gradien* yang telah didepan dikelompokkan menjadi 5 bagian, dengan ketentuan :

1. $M_1 \geq 1$, jika di dapat hasil gradien lebih besar atau sama dengan satu maka garis yang dibentuk adalah garis miring (*slash*).
2. $0 \leq M_2 < 1$, jika di dapat hasil gradien lebih besar atau sama dengan nol dan lebih kecil dari satu maka garis yang dibentuk adalah garis cenderung datar.
3. $-1 < M_3 < 0$, jika di dapat hasil gradien lebih besar dari minus satu dan lebih kecil dari nol maka garis yang dibentuk adalah garis cenderung datar.
4. $M_4 \leq -1$, jika di dapat hasil gradien lebih kecil atau sama dengan minus satu maka garis yang dibentuk adalah garis miring (*backslash*).
5. $M_5 = \text{inf}$, jika di dapat hasil gradien adalah inf maka garis yang dibentuk adalah garis lurus. Maka dapat dikelompokkan seperti Tabel 3.8

Tabel 3.8 Pengelompokkan Gradien

No. Pengujian	Gradien					
	1	2	3	4	5	6
1	M_1	M_1	M_2	M_3	M_4	-
2	M_3	M_3	M_4	M_4	M_3	M_2
3	M_3	M_1	M_4	M_2	-	-
4	M_3	M_3	M_4	M_3	M_4	M_3
5	M_3	M_4	M_4	M_4	M_4	-
6	M_3	M_2	M_4	M_2	M_1	-
7	M_3	M_1	M_4	M_3	M_1	-
8	M_4	M_3	M_3	M_2	-	-
9	M_3	M_3	M_3	M_2	M_3	M_2

No. Pengujian	1	2	3	4	5	6
10	M ₂	M ₂	M ₃	M ₂	M ₄	-

Pengujian dilakukan dengan algoritma *forward* dengan rumus:

1. Inisialisasi

$$\alpha_1(i) = \pi_i b_i(O_1), 1 \leq i \leq 2 \quad (1)$$

2. Induksi

$$\alpha_{t+1}(j) = [\sum_{i=1}^N \alpha_t(t) a_{ij}] b_j[o_{t+1}] \quad (2)$$

Pengujian dilakukan dengan algoritma *backward* dengan rumus:

1. Inisialisasi

$$\beta_T(i) = 1, 1 \leq i \leq 2 \quad (3)$$

2. Induksi

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(o_{t+1}) \beta_{t+1}(j) \quad (4)$$

Untuk mencari nilai akhir dari algoritma *forward-backward* menggunakan *Observation Evaluation*:

$$P(O|\lambda) = \sum_{i=1}^N \alpha_t(i) \beta_t(i), \forall t \quad (5)$$

Tabel 3.9 Nilai Maksimum dari Data Uji Pertama

No.	Input	Nilai Maksimum	Hasil	Keterangan
1	⊖	0,012	⇒	SALAH
2	⊢	0,00001592	⊢	BENAR
3	⊥	0,0011	⊃	SALAH
4	⊣	0,000361	⊘	SALAH
5	⊂	0,0174	⊐	SALAH
6	⊤	0,0001469	⊚	SALAH
7	⊚	0,00057569	⊚	SALAH
8	⊣	0,000002	⊘	SALAH
9	⊐	0,00017936	⊘	SALAH
10	⊥	0,00021676	⊘	SALAH
11	⊚	0,000639	⊣	SALAH
12	⊚	0,0029	⊘	SALAH
13	⋈	0,0031	⊘	SALAH
14	⊣	0,00031	⊚	SALAH
15	○	0,0047	⊣	SALAH
16	⊘	0,000304	⊚	SALAH
17	⋈	0,000154	⊚	SALAH
18	—	0,0182	—	BENAR
19	⋈	0,0001078	⋈	BENAR
20		0,0013		BENAR
21	⇒	0,0000507		SALAH

No.	Input	Nilai Maksimum	Hasil	Keterangan
22	π	0,0033	π	BENAR
23	Ε	0,0011	π	SALAH
24	Ϝ	0,015	Ε	SALAH

Pada Tabel 3.9 dijelaskan hasil dari pengujian no 1 dari Tabel 3.8 didapat lima gradien yang diuji terhadap 24 model *ergodic* yang telah dibentuk dengan pengujian menggunakan algoritma *forward-backward* untuk mengetahui nilai maksimum dari hasil pengujian terhadap masing-masing huruf terhadap 24 model.

Tabel 3.10 Hasil Pengujian Data Precision Perorang

No.	Jumlah Input	Total Benar	Total Salah	Akurasi Precision
1	24	5	19	20,83%
2	24	6	18	25%
3	24	5	19	20,83%
4	24	4	20	16,67%
5	24	6	18	25%
6	24	4	20	16,67%
7	24	5	19	20,83%
8	24	7	17	29,17%
9	24	6	18	25%
10	24	5	19	20,83%
Rata-Rata Akurasi Precision				22,083 %

Pada Tabel 3.10 menjelaskan hasil dari pengujian *precision* yang menghasilkan nilai rata-rata akurasi sebesar 22,083%, dimana data yang diambil sebanyak 240 huruf untuk dilakukan *testing*. Hasil yang didapatkan dari 240 huruf, sebanyak 52 huruf dapat dijawab sistem dengan benar, sedangkan 188 huruf yang diprediksi oleh sistem masih salah. Pengujian ini dilakukan terhadap jumlah data perorangan.

Tabel 3.11 Hasil Pengujian Data Precision Perhuruf

No.	Jumlah Input	Sesuai	Tidak Sesuai	Akurasi Precision
1	10	2	8	20%
2	10	2	8	20%
3	10	0	10	0%
4	10	0	10	0%
5	10	0	10	0%
6	10	2	8	20%
7	10	0	10	0%
8	10	0	10	0%
9	10	0	10	0%
10	10	0	10	0%
11	10	0	10	0%
12	10	1	9	10%
13	10	1	9	10%
14	10	2	8	20%
15	10	0	10	0%
16	10	1	9	10%
17	10	6	4	60%

No.	Jumlah Input	Sesuai	Tidak Sesuai	Akurasi Precision
18	10	10	0	100%
19	10	4	6	40%
20	10	10	0	100%
21	10	0	10	0%
22	10	8	2	80%
23	10	2	8	20%
24	10	2	8	20%
Rata-Rata Akurasi Precision				22,083 %

Pada Tabel 3.11 menjelaskan hasil dari pengujian *precision* yang menghasilkan nilai rata-rata akurasi sebesar 22,083%, dimana data yang diambil sebanyak 240 huruf untuk dilakukan *testing*. Hasil yang didapatkan dari 240 huruf, sebanyak 52 huruf dapat dijawab sistem sesuai dengan model, sedangkan 188 huruf yang diprediksi oleh sistem masih salah. Pengujian ini dilakukan terhadap jumlah data perhuruf.

4. KESIMPULAN

Dari hasil penelitian dan pembahasan yang telah dilakukan, maka dapat disimpulkan bahwa :

1. Metode *Hidden Markov Model* (HMM) dapat digunakan untuk pengenalan tulisan tangan huruf hangeul dengan algoritma *forward-backward*.
2. Persentasi ketepatan *Hidden Markov Model* (HMM) sangat bergantung pada pengelompokan data.
3. Tingkat akurasi yang dihasilkan pada pengujian pengenalan huruf *hangeul* menggunakan algoritma *forward-backward* terhadap 10 orang dan 24 huruf menghasilkan akurasi yang sama dengan rata-rata adalah 22,083%.

5. SARAN

Saran yang dapat direkomendasikan untuk penelitian selanjutnya adalah:

1. Menambahkan deteksi pada huruf hangeul yang berbentuk lingkaran, sehingga diharapkan mampu meningkatkan peluang emisi guna meningkatkan akurasi dari *Hidden Markov Model* (HMM) untuk mendeteksi huruf hangeul.
2. Melakukan lebih banyak tahapan preprocessing untuk meningkatkan ketepatan huruf hangeul yang akan dideteksi.

DAFTAR PUSTAKA

- [1] A. Setiyoaji, L. Muflikhah, and M. A. Fauzi, "Named Entity Recognition Menggunakan Hidden Markov Model dan Algoritma Viterbi pada Teks Tanaman Obat," *Pengemb. Teknol. Inf. dan Ilmu Komput.*, 2017.
- [2] M. M. Faridh, "Pengenalan Karakter Huruf Tulisan Tangan Menggunakan Metode Principal Components Analysis," *Univ. Dian Nuswantoro Semarang*, 2013.

- [3] A. N. Septiawan and T. Setiadi, "Aplikasi Pengenalan Huruf Hijaiyah Berbasis Android," *Univ. Ahmad Dahlan*, 2013.
 - [4] M. Gultom, Mukhlisa, and D. Alamsyah, "Menggunakan Metode Hidden," *STMIK GI MDP*, 2014.
 - [5] Siska Fathia, "Penerapan Jaringan Syaraf Tiruan Dalam Pengenalan Tulisan Tangan Huruf Korea (Hangul) Menggunakan Metode Propagasi Balik," *Univ. Dian Nuswantoro*, 2013.
 - [6] Y. S. Putra, A. Novianty, and N. Anbarsanti, "Perancangan Dan Implementasi Sistem Pengenalan Bahasa Isyarat Indonesia Menggunakan Kombinasi Sensor Depth Image Dan Human Skeleton Kinect Dengan Metode Hmm Design and Implementation Recognition System of Indonesia Sign Language Using Combination Depth Im," *e-Proceeding Eng.*, 2016.
 - [7] E. F. Yuwitaning, B. Hidayat, and N. Andini, "Implementasi Metode Hidden Markov Model untuk Deteksi Tulisan Tangan," *e-Proceeding Eng.*, 2014.
-