**Bab III**

**Perancangan Pengujian**

## III.1 Perancangan Awal Sistem dan Eksperimen Pengenalan Huruf Arab

Dalam tesis ini akan dibuat sebuah perangkat lunak pengenalan huruf Arab yang bisa mengenali huruf Arab terisolasi dan dalam rangkaian kata. Masukan dari sistem ialah berupa citra berisikan kalimat tulisan Arab kemudian melalui proses binerisasi. Untuk pengenalan huruf Arab terisolasi setelah melalui proses binerisasi akan dilanjutkan ke tahap penipisan, kemudian fitur dari tiap huruf akan di ekstrak melalui chain code, jumlah titik dan posisi titik. Setelah fitur di ekstrak akan dilakukan klasifikasi menggunakan neural network dan hidden markov model. Sedangkan pada pengenalan huruf Arab dalam kalimat setalah proses binerisasi akan melalui tahap segemntasi, kemudian dilanutkan pada tahap penipisan, lalu akan di ekstrak fitur setiap huruf dengan chain code, umlah titik dan posisi titik. Setelah fitur di ekstrak akan dilakukan klasifikasi menggunakan neural network dan hidden markov model. Gambaran umum tahapan proses pengenalan huruf Arab dapat dilihat pada Gambar III.1



Gambar III.1 Alur proses pengenalan huruf Arab

Gambar III. 1 Diagram blok utama dalam sistem pengenalan huruf Arab

Dalam penelitian ini dikenal data latih dan data uji. Berikut penjelasan dari kedua jenis data tersebut.

1. Data latih

Data latih merupakan kumpulan data huruf arab sesuai sesuai posisinya pada kalimat yakni terisolasi, di awal, di tengah, dan di akhir yang akan dilatih untuk kemudian disimpan data latih huruf Arab.

1. Data uji

Data uji terbagi dua, yaitu data uji huruf Arab terisolasi dan kumpulan data kalimat Arab yang akan dimasukan dalam tahap pengujian dan dibandingkan data latih untuk ditentukan hasil pengenalan huruf Arab tersebut.

Jumlah huruf arab yang akan dilatih sebanyak 31 huruf arab, 1 huruf arab memiliki 4 posisi yang berbeda-beda, yaitu posisi terpisah, diawal, ditengah dan diakhir. Berikut teknik yang digunakan dalam pengambilan data latih.

1. Ketik huruf 31 huruf Arab menggunakan Microsoft Word , huruf arab ditulis sesuai posisinya pada kalimat, yakni posisi terisolasi, di awal, di tengah dan di akhir.
2. Potong setiap huruf menggunakan Snipping Tool
3. Hasil pemotongan huruf disimpat sebagai data latih huruf arab

Saat pengambilan data uji, dilakukan dalam dua tingkat kesulitan, yaitu:

1. Pengujian huruf Arab terisolasi

Saat pengujian huruf ini dilakukan sebanyak 31 huruf arab terisolasi yang terdapat pada tabel II.1. Pangujian huruf Arab terisolasi dilakukan dalam waktu 13 detik.

1. Pengujian kalimat

Saat pengujian kalimat dilakukan dengan memasukkan 10 kalimat yang berbeda kedalam sistem. Pengujian kalimat dilakukan dalam waktu 11 detik. Berikut 10 kalimat Arab yang akan diujikan kedalam sistem :

1. الاحترام المتبادل بين الاديان
2. لقمر تبدو جميلة جدا
3. ف كهة دوريان ا لا ذو اق جيدة
4. تتخلي ابدا في الحياة
5. انتظر اي محا كمة
6. لحفاظ علي صحتك حياة طيبة
7. المعلمين يعلمو ن
8. جدةالارز المطبو خ
9. يتكلم ببطء مفهومة حتي
10. الحفا ظ علي صحتك حيا ة طيبة

**III.2 Perancangan Sistem dan Eksperimen**

Pada penelitian ini terdapat 5 tahap yang dilakukan untuk membuat sistem pengenalan huruf Arab, berikut adalah penjelasan masing – masing tahap.

**III.2.1 Eksperimen Binerisasi**

Binariisasi gambar adalah proses pengubahan gambar menjadi biner yang memiliki nilai 0 dan 1. Citra grayscale akan berubah menjadi hitam putih. Proses binerisasi diperlukan untuk melakukan langkah selanjutnya pada pengenalan huruf dan kalimat Arab. Cara melakukannya adalah dengan melakukan threshold pada setiap channel warna. Ambang yang digunakan adalah 150. Jika saluran warna kurang dari 150 maka akan dikonversi menjadi hitam, dan jika warnanya lebih dari 150 akan berubah menjadi putih.



Sebelum Binerisasi Sesudah Binerisasi

Gambar III.2 Perbandingan citra sebelum binerisasi dan sesudah binerisasi

**III.2.2 Perancangan Sistem Penipisan**

Setelah diperoleh citra biner, dilakukan proses penipisan untuk mendapatkan tulang dari huruf. Proses penipisan dilakukan dengan algoritma Stentiford. Langkah – langkah dari algoritma Stentiford telah dijelaskan pada bab II.4.2.

**III.2.3 Eksperimen Segmentasi**

Pada pengenalan huruf Arab dalam kalimat melalui tahap segmentasi terlebih dahulu sebelum tahap penipisan.

Segmentasi dilakukan dengan menggunakan algoritma Zidouri [7]. Langkah pertama segmentasi adalah menentukan beberapa parameter yang digunakan sebagai segmen segmentasi. Setelah parameter dilakukan tahap segmentasi yang dilakukan. Kemudian akan dipilih guide band sebagai referensi segmentasi karakter. Untuk memilih panduan band yang benar beberapa fitur diekstraksi dari masing-masing guide band.

Ada empat aturan yang digunakan algoritma saat ini calon band giude:

1. Pilih guide band jika lebar relatifnya paling besar dan F4 = 1
2. Pilih guide band jika F2 > Ls dan F4 = 1
3. Pilih guide band jika F2 < Ls dan F3 > Lsa dan pita bukan pita terakhir
4. Pilih guide band jika F1 lebih dari Lm dan F4 = 1

dengan F1 = lebar guide band, F2 = jarak guide band ke guideband sebelah kanannya, F3 = jarak guide band ke guide band kedua sebelah kanannya, dan F4 = posisi ditemukannya guide band, bernilai satu di atas lini basis dan nol jika di bawah (Zidouri, 2010).

**III.2.4 Eksperimen Ekstraksi Fitur**

Ekstraksi fitur dilakukan dalam 3 tahap yakni sebagai berikut :

1. Chain code

Setelah melalui tahap penipisan fitur chain code di ekstrak untuk setiap huruf Arab kemudian dialkukan normalisasi pada setiap chain code sehingga panjang chain code hanya 10 untuk setiap huruf. Langkah – langkah untuk mengambil chain code pada suatu objek telah dijelaskan pada bab II.4.1, dan langkah – langkah untuk melakukan normalisais chain code telah dijelaskan pada tahap II.4.1.1

1. Jumlah titik

Setelah fitur chaincode fitur jumlah titik merupakan fitur yang penting karena beberapa huruf arab memiliki kemiripan bentuk namun dapat di bedakan berdasarkan jumlah titik, jumlah titik terbagi menjadi 4 , yakni 0, 1, 2 dan 3.

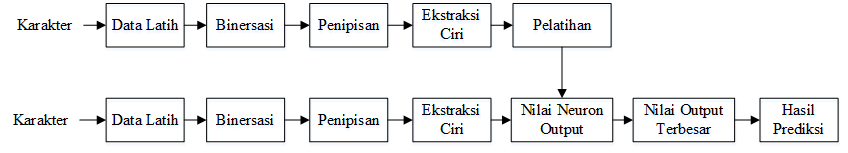
1. Posisi titik

Setelah fitur chaincode dan jumlah titik, fitur posisi titik juga merupakan fitur yang penting karena beberapa huruf arab memiliki kemiripan bentuk namun dapat di bedakan berdasarkan posisi titik, posisi titik terbagi menjadi 3, yaitu di atas, ditengah dan di bawah.

**III.3 Perancangan Sistem dan Eksperimen Pengklasifikasi**

**III.3.1 Tahap Pelatihan dan Pengujian dengan Metode Neural Network**

Proses pengenalan huruf Arab pada tahap klasifikasi dengan metode neural network dapat dilihat pada gambar di bawah ini.



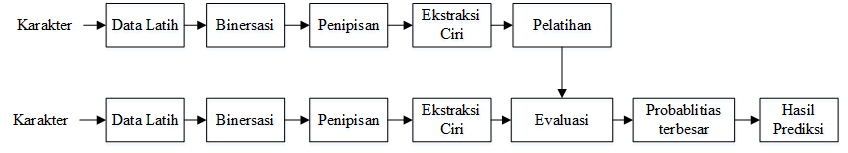
Gambar III.3 Proses pengenalan huruf Arab dengan metode neural network

Tahap pelatihan yang dibangun metode neural network menggunakan metode Backpropagation yang merupakan salah satu metode learning dalam neural network. Pada Backpropagation terdapat cara update bobot secara khusus,bobot di update secara terus menerus sampai output neuron mendekati sama dengan target. Update bobot berhenti saat epoch telah mencapai batasnya.

Pada penelitian ini jenis arsitektur neural network yang digunakan adalah jaringan layer plural dengan 1 hidden layer. Fungsi aktivasi yang digunakan untuk hidden layer adalah fungsi aktivasi sigmoid, dan fungsi aktivasi yang digunakan untuk layer output adalah fungsi aktivasi softmax. Masukan neuron untuk setiap sampel adalah 12, neuron pertama adalah jumlah titik, neuron kedua adalah posisi titik-titik dan neuron ketiga adalah chain code yang telah dinormalisasi.

**III.3.2 Tahap Pelatihan dan Pengujian dengan Metode Hidden Markov Model**

Proses pengenalan huruf Arab pada tahap klasifikasi dengan metode hidden markov model dapat dilihat pada gambar di bawah ini.



Gambar III.4 Proses pengenalan huruf Arab dengan metode hidden markov model

Tahap pelatihan dilakukan dengan menentukan parameter estimasi, sehingga terbentuk hidden markov model berupa λ = (A, B, ∏), dilakukan dengan algoritma Baum Welch.

Data latih yang disimpan di ekstrak fiturnya menggunakan chain code, jumlah titik dan posisi titik. Keluaran dari ekstraksi fitur adalah jumlah titik, posisi titik dan chain code yang telah ternormalisasi menjadi 10 angka untuk setiap huruf. Hasil ekstraksi fitur ini menjadi masukan oleh proses pelatihan dengan algoritma Baum Welch. Hidden state dalam penelitian ini adalah label id atau nomor urut dari huruf Arab. Observed sequence adalah jumlah titik, posisi titik dan chain code yang telah dinormalisasi.

Pada proses Pengujian atau evaluasi menggunakan algoritma forward karen akan dicari nilai peluang yang paling tinggi dan menentukan kecocokan antara data uji dan data latih. Nilai peluang yang paling tinggi akan dipilih sebagai huruf hasil pengenalan dari masukan huruf yang diujikan.