

SISTEM PENGENALAN POLA KARAKTER HURUF KOREA MENGUNAKAN METODE *PRINCIPAL COMPONENT ANALYSIS* DAN JARINGAN SYARAF TIRUAN - *BACK PROPAGATION*.

Delsavonita¹⁾ dan Feri Candra²⁾

¹⁾Mahasiswa Program Studi Teknik Informatika S1 dan ²⁾Dosen Teknik Informatika
Fakultas Teknik Universitas Riau

Kampus Bina Widya Jl.HR.Soebrantas Km 12.5 Simpang Baru, Panam, Pekanbaru 28293

<https://www.unri.ac.id>. Telp. : 0761-63272. Fax: 0761-566821

Email: delsavonita.d@student.unri.ac.id

ABSTRACT

This paper describes about a recognition sistem for Korean characters into Latin form by using Principle Component Analysis Method and Artificial Intelligence - Back Propagation Approach. This sistem uses image of Korean characters as input data with 65x65 pixel of original image size and is processed by image preprocessing in the form of pixel size conversion into 15x15 pixel binary image. Every image is then extracted to produce an image feature. The features are processed first using Principle Component Analysis to reduce image feature before they enter classification stage by using Artificial Intelligence - Back Propagation Approach. This study uses 10 sample data of Korean vowel letters, obtained from 25 different font types and each font consists of normal and bold sample. Total data reaches 500 samples divided into 70 training data and 30 testing data. Architecture of this artificial intelligence uses 3 hidden layers . Each hidden layer consists of 20, 20 and 5 neurons, one output neuron. The result of this sistem research is obtained an accuracy of 95%.

Keyword : Korean Letters a.k.a Hangul, Principle Components Analysis, Artificial Intelligence - Back Propagation

1. PENDAHULUAN

Latar Belakang

Banyaknya bangsa dan negara di dunia, menimbulkan keanekaragaman bahasa dan budaya. Ada pula bangsa dan negara yang saling bekerjasama demi kemajuan negara tersebut, baik dalam bidang ekonomi, budaya, pendidikan, politik, kemanusiaan dan lainnya. Dengan adanya hubungan antar bangsa tersebut, akan terjadi komunikasi dengan bangsa yang memiliki tutur bahasa yang berbeda satu dengan yang lainnya. Dengan demikian ditetapkanlah sebuah bahasa Internasional. Tetapi, selain

dengan bahasa Internasional, ada pula yang ingin mempelajari lebih dalam tentang bahasa dari negara lain, salah satunya yang sedang berkembang pesat di Indonesia adalah bahasa Korea.

Karakter huruf Korea atau sering disebut dengan huruf “Hangul” terdiri dari 21 huruf vokal (10 vokal tunggal dan 11 vokal ganda) dan 19 huruf konsonan yang diinput dari keyboard. Huruf hangul sendiri berupa simbol-simbol yang mempunyai tulisan latin masing-masingnya.

Terdapat berbagai cara manusia dalam mempelajari bahasa asing, selain dengan

mempelajarinya dengan cara manual seperti kursus dan berbicara aktif, juga dapat mempelajari bahasa asing secara digital, autodidak, melalui internet dan komputer. Maka dibuatlah sistem pengenalan karakter huruf Korea yang dapat membantu mempelajari bahasa Korea dasar, dimana sistem ini akan membantu pemula dalam mengenali tulisan latin dari karakter huruf Korea yang diinputkan.

Kecerdasan buatan atau *Artificial Intelligence* adalah salah satu bidang ilmu yang mempelajari sebuah komputer agar dapat bertindak sebaik dan layaknya seorang manusia. Bagian yang paling penting dalam teori kecerdasan buatan adalah sistem pengenalan polanya (*Pattern Recognition*) yang merupakan implementasian dari jaringan syaraf tiruan. Sistem pengenalan pola merupakan elemen penting dalam proses komputer meniru kebiasaan manusia, terutama penglihatan.

Teknik pengenalan pola (*Pattern Recognition*) mengalami banyak kemajuan dan semakin dipakai untuk memecahkan suatu masalah. Teknik pengenalan pola dipakai untuk mengenali tulisan tangan, tanda tangan, gambar dan sebagainya. Berbeda dengan disiplin ilmu pengolahan citra yang dibatasi oleh penggunaan citra sebagai masukan maupun keluarannya, suatu aplikasi pengenalan pola bertujuan untuk melakukan proses pengenalan terhadap suatu objek (misalnya citra) kedalam salah satu kelas tertentu, berdasarkan pola yang dimilikinya. Jaringan syaraf tiruan memiliki beberapa metode atau algoritma yang dapat digunakan untuk

mengklasifikasikan atau mengenali suatu tipe pola khususnya dalam pengenalan pola tulisan tangan, salah satunya adalah *Back-Propagation*. Algoritma ini melakukan dua tahap~perhitungan, yaitu: perhitungan maju untuk menghitung galat (*error*) antara keluaran aktual dan target; dan perhitungan mundur yang mempropagasikan balik galat (*error*) tersebut untuk memperbaiki bobot-bobot~sinaptik~pada~semua neuron yang ada. Dengan kata lain, kita dapat memasukkan acuan~yang~akan~digunakan sesuai~kebutuhan.

Pada pengenalan pola karakter huruf Korea ini akan menggunakan metode *Principal Component Analysis* (PCA), dimana ekstraksi fitur citra akan di kurangkan terlebih dahulu sebelum memasuki tahapan selanjutnya, yakni Jaringan Syaraf Tiruan - *Back Propagation* (JST-BP). Dengan tujuan agar fitur yang diinputkan merupakan fitur yang dianggap berperan besar terhadap sistem pengenalan pola ini. JST-BP merupakan bagian dari kecerdasan buatan yang menirukan cara kerja dari syaraf manusia, yang dapat mempelajari sebuah pengetahuan baru yang di dapat dari proses *training* dan *testing*. Dengan menggunakan metode ini citra inputan dapat diklasifikasi sesuai dengan kelas masing-masing.

Pada sistem ini, tahapan proses yang akan dilakukan adalah pengambilan data *input* berupa huruf-huruf Korea dengan berbagai bentuk penulisan dalam bentuk citra, kemudian dilakukan *preprocessing* yakni *cropping* dengan tujuan citra yang akan dikenali lebih terfokus pada objek citranya kemudian dengan mengubah

citra tersebut kedalam bentuk *grayscale* dan *biner* dan kemudian menerapkan proses PCA untuk mengurangi fitur yang tidak digunakan, setelahnya akan dilakukan proses ekstraksi fitur berupa pola *biner* dimana matrik dari citra *biner* di buat kedalam bentuk vektor. Selanjutnya menggunakan metode klasifikasi JST-BP terhadap fitur-fitur yang ada, dan pada akhirnya akan menghasilkan keluaran berupa pelafalan huruf Korea tersebut.

Tujuan Penelitian

Berikut ini merupakan beberapa objektif yang akan dicapai dalam pembuatan sistem pengenalan pola karakter huruf Korea menggunakan *PrincipalComponent Analysis* dan Jaringan Syaraf Tiruan – *Back Propagation*.

1. Terciptanya sistem pengenalan pola dalam proses pengenalan karakter huruf korea
2. Menghasilkan model jaringan syaraf tiruan yang dapat mendeteksi karakter huruf korea.
3. Mendapatkan tingkat akurasi yang baik dari pemilihan dimensi fitur citra menggunakan PCA

Batasan Masalah

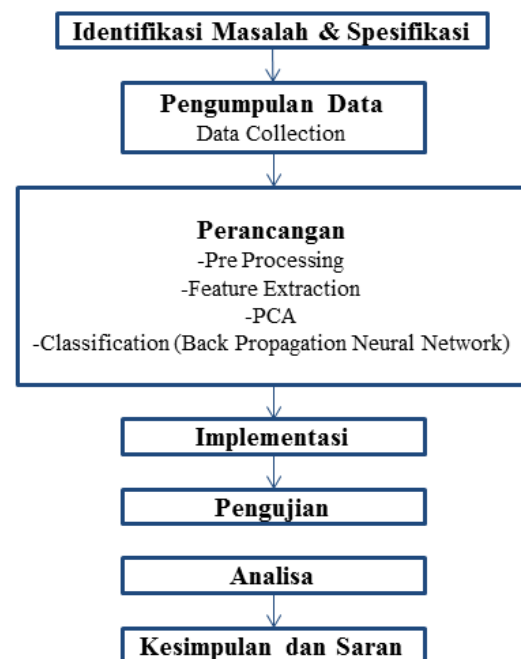
Dalam pembuatan penelitian ini akan disertai beberapa batasan topik penelitian berupa :

1. Sistem hanya menerima inputan untuk huruf Korea dalam bentuk citra
2. Hanya mengenali satu huruf per satu kali percobaan dan tidak menggabungkan beberapa jenis huruf Korea.

3. Keluaran hanya menampilkan tulisan latin dari huruf Korea yang diinputkan.
4. Terdiri dari 10 jenis karakter huruf vokal korea
5. Sampel menggunakan 25 jenis font yang berbeda serta tingkat ketebalan normal dan bold

2. METODOLOGI

Berikut ini ditampilkan diagram dari tahapan metodologi penelitian Sistem Pengenalan Pola Karakter Huruf Korea Menggunakan *Principal Component Analysis* Dan Jaringan Syaraf Tiruan - *Back Propagation*.



Gambar 2.1 Tahapan Metodologi Penelitian

Identifikasi Masalah & Justifikasi

Tahapan identifikasi masalah adalah tahapan dimana akan dicarinya sebuah permasalahan yang dialami ketika mempelajari sebuah bahasa, dalam pembahasan ini adalah bahasa korea, yang mana para pemula tidak

memiliki dasar dari pengetahuan huruf latin korea tersebut. Peneliti akan mencari berbagai informasi maupun referensi yang berkaitan erat dengan penelitian tentang sistem pengenalan pola baik secara umum, pengenalan pola tentang wajah, tulisan jawa, maupun pengenalan pola *alphabet* dan *numeric*, yang bersumber dari buku, jurnal, *e-book* dan lainnya. Setelah mendapatkan informasi dari penelitian sebelumnya, kemudian akan dipelajari tentang permasalahan apa yang belum terselesaikan, baik dari segi objek yang diteliti, metode yang digunakan, maupun jumlah sampel. Bercermin dari permasalahan yang ada, akan dibuat suatu sistem pengenalan pola huruf korea yang akan menampilkan bahasa latin dari huruf korea yang dicari.

Dengan menggunakan ekstraksi fitur *principal component analysis* yang dinilai dapat mengurangi dimensi dari variable fitur yang akan diolah dalam tahap klasifikasi menggunakan metode jaringan syaraf tiruan – *back propagation*. Dengan dibuatnya penelitian tentang sistem pengenalan pola pengenalan karakter huruf Korea ini diharapkan dapat membantu dalam proses pembelajaran bahasa Korea bagi pemula.

Pengumpulan Data

Tahapan pengumpulan data yang dilakukan adalah pengumpulan data sampel, dimana akan dijelaskan bagaimana cara dan tahapan dalam pengambilan data sampel, serta dari mana data sampel akan diambil. data

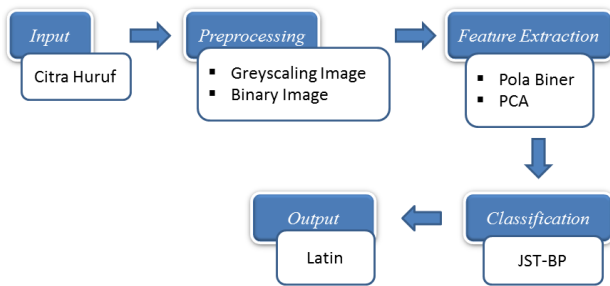
sampel merupakan data yang akan digunakan dalam sistem yang akan dibuat. Data sampel ini akan diperoleh dari hasil screenshot 10 jenis karakter huruf korea dari PC, dimana setiap jenis terdiri dari 25 font yang berbeda dan juga masing-masing font memiliki tingkat ketebalan yaitu normal dan bold, sehingga akan menghasilkan sampel sebanyak 500 sampel yang akan dibagi menjadi data training dan data testing pada proses jaringan syaraf tiruan.

Perancangan

perancangan yang dilakukan pada penelitian ini adalah perancangan proses mengenai tahap-tahap pengenalan pola karakter huruf dan perancangan sistem untuk implementasi yang dilakukan.

Perancangan proses yang dilakukan dalam penelitian ini berupa Perancangan *preprocessing*, *feature extraction* dan *classification*. Berikut akan dijelaskan tahapan proses perancangan sistem.

Tahapan perancangan dibuat untuk menjelaskan tahap apa yang akan dilakukan setelah semua tahap pada proses analisa selesai dilakukan. Perancangan yang akan dilakukan pada sistem adalah membuat *flowchart* dan perancangan tampilan aplikasi (*Interface*). *Flowchart* berfungsi untuk menggambarkan alur proses dari sistem yang akan dibangun. Perancangna tampilan aplikasi dibuat untuk menjadi panduan dalam pembuatan tampilan sistem yang akan dibangun.



Gambar 2.2 Tahapan Perancangan Sistem.

Preprocessing

Tahapan *preprocessing* merupakan tahapan kedua dari proses pengenalan pola. Pada tahapan ini akan dilakukan *preprocessing* terhadap citra karakter huruf Korea dengan mengubah citra RGB menjadi biner, kemudian akan dilakukan proses *cropping* dan *resize* data citra sehingga memiliki keseragaman ukuran, yakni 15x15 *pixel*. Hasil dari tahapan *preprocessing* ini akan dijadikan data berupa vektor yang akan digunakan dalam tahapan *processing* dan *classification*.



Gambar 2.3.Tahapan *Preprocessing*.

Feature Extraction

Setelah tahapan *preprocessing* selesai, maka akan dilakukan tahapan *Feature Extraction*, dimana akan dilakukan pengekstraksian fitur dari citra biner menjadi sebuah matrik, kemudian mengubah matrik tersebut ke dalam bentuk vektor. Proses *feature extraction* terdiri dari 2 tahapan, yaitu :

1. Pemetaan Area Sampel Ke Dalam Matriks (*Pixel Mapping*)

Setelah citra karakter telah melewati tahapan *preprocessing*, maka saatnya pola karakter tersebut dipetakan ke dalam matrik misalnya menjadi 15x15 *pixel*. Pada matriks tersebut, bilangan 1 untuk area berwarna putih, dan bilangan 0 untuk area berwarna hitam.

$$\begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Gambar 2.4 Pemetaan area sampel ke dalam matrix

2. Pembuatan Vektor (*Creating Vector*)

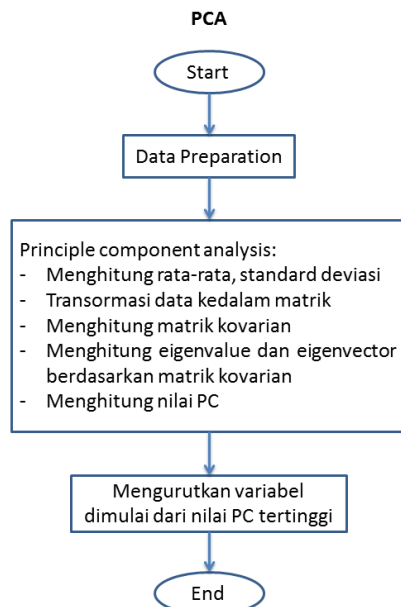
Setelah diperoleh sampel dari area bilangan biner (area putih dan hitam), maka dilakukan pengumpulan bilangan biner pada matriks citra setiap barisnya kemudian digabungkan.

$$\begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \rightarrow [1 \ 0 \ 1 \ 1 \ 0 \ 1 \ 1 \ 1 \ 1]$$

Gambar 2.5 Pembuatan *vector*

Principal Component Analysis (PCA)

Setelah tahapan *feature extraction* selesai dilakukan, akan didapati keluaran berupa vektor dengan panjang 255 variabel. Data vektor yang terlalu besar tersebut akan diproses menggunakan *principal component analysis*, agar variabel yang tidak diperlukan dapat dihilangkan, sehingga nantinya pada proses *classification* dapat meningkatkan waktu pengklasifikasian karakter.



Gambar 2.6 Flowchart PCA

Berikut adalah hasil akhir dari tahapan proses PCA, dimana akan didapatkan hasil pengurangan *variable* yang kurang berpengaruh.

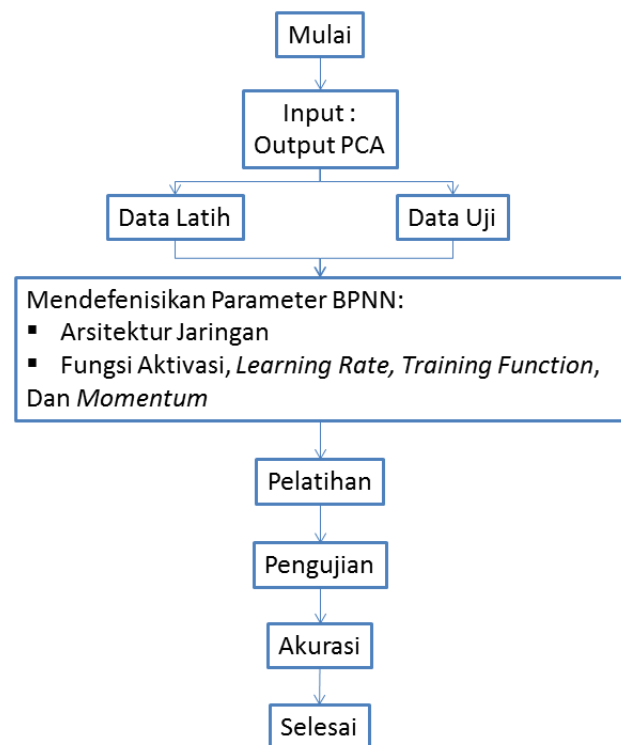
A	B	C	D	E		B	C	D
1	0	1	0	1	→	0	1	0
1	1	0	0	1		1	0	0
1	0	0	1	1		0	0	1
1	0	0	0	1		0	0	0
1	1	0	1	1		1	0	1

Gambar 2.7 Reduksi fitur

Classification

Tahapan *classification* yang dilakukan pada proses ini adalah menggunakan metode jaringan syaraf tiruan - *back propagation*. Algoritma jaringan syaraf tiruan *back propagation* merupakan algoritma pembelajaran yang terawasi dan biasanya digunakan oleh *perceptron* dengan banyak lapisan untuk mengubah bobot-bobot yang terhubung dengan *neuron* yang ada pada lapisan tersembunyi. Algoritma jaringan syaraf tiruan - *back propagation* menggunakan nilai *error output* untuk mengubah nilai bobotnya dalam arah

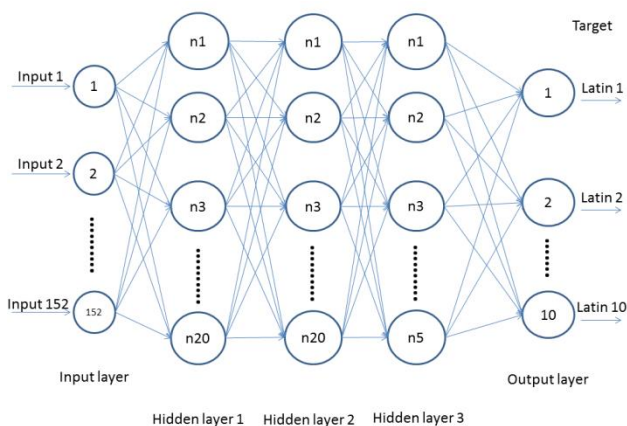
mundur. Untuk mendapatkan *error* ini, tahapan perambatan maju harus dikerjakan terlebih dahulu. Setiap unit yang berhubungan dengan lapisan masukan (*Input*) akan berhubungan dengan setiap unit yang ada pada lapisan tersembunyi. Begitu pula dengan lapisan tersembunyi, setiap unit yang ada pada lapisan tersembunyi akan berhubungan dengan semua unit yang ada pada lapisan keluaran (*Output*).



Gambar 2.8 Tahapan Jaringan Syaraf Tiruan *Back Propagation*

Tahapan awal untuk membangun sebuah BPNN adalah menentukan arsitektur jaringannya, diantaranya jumlah hidden layer dan jumlah neuron dalam suatu hidden layer. Dalam penelitian ini digunakan 5 lapisan layer, dimana layer pertama disebut input layer, layer kedua, ketiga dan keempat disebut hidden layer yang masing-masing layer memiliki 20 neuron untuk *hidden layer* pertama 20 neuron untuk *hidden layer* kedua dan 5 nueron untuk *hidden*

layer ketiga dan yang terakhir disebut *output layer*. Dalam penentuan jumlah hidden layer digunakan rumus $I < m < 2I$, dimana I merupakan jumlah inputan, sehingga jika jumlah inputan adalah 30, maka neuron untuk hidden layer tidak boleh kecil dari 30 maupun besar dari 60, dimana pada penelitian ini digunakan 20 neuron hidden layer pertama, 20 neuron hidden layer kedua dan 5 neuron hidden layer ketiga yang dipilih secara acak.



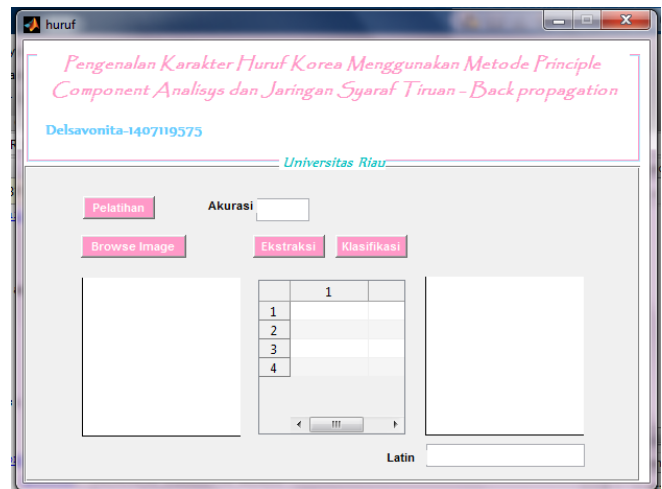
Gambar 2.9 Arsitektur jaringa BPNN

Dalam BPNN, untuk mengatur perubahan bobot dala proses pembelajaran, diperlukan sebuah fungsi aktvasi .fungsi aktivasi yang digunakan dalam penelitian ini adalah log-sigmoid (logsig) yang menghitung normalisasi *variable* menggunakan rumus ‘min-max’ :

$$x = \frac{x_i - x_{min}}{x_{max} - x_{min}}$$

3. HASIL DAN PEMBAHASAN

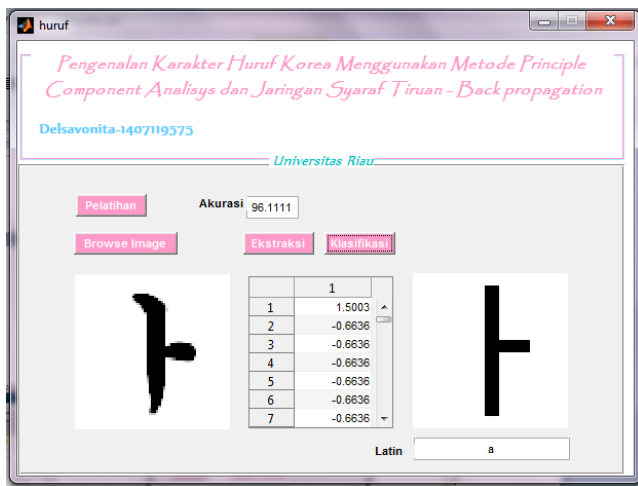
Berikut implemenatsi hasil GUI matlab pada sistem yang dirancang :



Gambar 3.1 Tampilan GUI Matlab

Pertama kali jaringan akan detraining menggunakan data citra yang telah diekstraksi dan direduksi menggunakan metode *principal component analysis* untuk mendapatkan bobot yang akan digunakan sebagai acuan jaringan untuk mengenali pola huruf yang akan diuji menggunakan jaringan syaraf tiruan. Untuk menilai apakah bobot yang dihasilkan mampu mengenali data pengujian, maka digunakan perhitungan akurasi pelatihan dengan membandingkan keluaran output dengan target data. Perhitungan ini dihitung dalam satuan persen (%). Biasanya perhitungan akurasi yang dianggap cukup untuk pelatihan data diatas 85%. Pada sistem ini menghasilkan persenan akurasi yang lebih tinggi, yakni 96.1111%, dimana jaringan ini layak digunakan untuk mengenali pola data pengujian.

Berikut rangkuman hasil pengenalan pola pada keseluruhan data uji dari jaringan yang telah di training



Gambar 3.2 Pengujian sistem

Table 1 Hasil pengujian terhadap sistem pengenalan pola karakter huruf vokal korea

No	Karakter	Hasil	Persentase
1	ㅏ	Benar = 14	100 %
		Salah = 0	
2	ㅡ	Benar = 12	85.71 %
		Salah = 2	
3	ㅑ	Benar = 12	85.71 %
		Salah = 2	
4	ㅣ	Benar = 14	100 %
		Salah = 0	
5	ㅓ	Benar = 14	100 %
		Salah = 0	
6	ㅕ	Benar = 14	100 %
		Salah = 0	
7	ㅗ	Benar = 13	92.85 %
		Salah = 1	
8	ㅛ	Benar = 14	100 %
		Salah = 0	
9	ㅜ	Benar = 14	100 %
		Salah = 0	

10	ㅟ	Benar = 13	92.85 %
		Salah = 1	

Maka total akurasi data pengujian adalah :

$$\bar{x} = \frac{100 \% + 85.71 \% + 85.71 \% + 100 \% + 100 \% + 100 \% + 92.85 \% + 100 \% + 100 \% + 92.85 \%}{10}$$

$$\bar{x} = \frac{957.12 \%}{10} = 95.71 \%$$

Sehingga didapatkan akurasi dari data pengujian sebesar 95.71 %.

4. PENUTUP

Berdasarkan hasil pengujian sistem pengenalan huruf korea menggunakan metode *Principal Component Analysis* Dan Jaringan Syaraf Tiruan – *Back Propagation* dengan pengaturan parameter sedemikian rupa sehingga menghasilkan tingkat akurasi diatas 95%, maka sistem ini dapat digunakan sebagai pengembangan penelitian dibidang jaringan syaraf tiruan maupun *principal component analysis*, khususnya bagi peneliti maupun peserta didik untuk pengembangan teknologi selanjutnya.

5. DATAR PUSTAKA

Faridh H., M. M (2013).*Pengenal Karakter Huruf Tulisan Tangan Menggunakan Metode Principle Component Analysis*. 6.

Fathia, S. (2013).*Penerapan Jaringan Syaraf Tiruan Dalam Pengenalan Tulisan Tangan Huruf Korea (Hangul) Menggunakan Metode Propagasi Balik*. 9.

Handoyo, E. D., & Susanto, L. W. (2011).*Penerapan Jaringan Syaraf Tiruan Metode Propagasi Balik Dalam Pengenalan*

Tulisan Tangan Huruf Jepang Jenis Hiragana Dan Katakana. Jurnal Informatika. 16.

Hastari, I. N., Hidayat, B., & Saepudin, D. (2010). Pengenalan Pola Korea Menggunakan Jaringan Syaraf Tiruan Propagasi Balik Dan Pengalihan Bahasa Korea-Indonesia. 9.

Nurmila, N., Sugiharto, A., & Sarwoko, E. A. (2010). Algoritma Backpropagation Neural Network Untuk Pengenalan Pola Karakter Huruf Jawa. 10.

Periasamy, Azha, Et Al. (2012). Identifying The Character By Applying Pca Method Using MATLAB. International Journal Of Computer Application, 4.

Pal, A., & Singh, D. (2010). Handwritten English Character Recognition Using Neural Network. International Journal Of Computer Science & Communication, 4.

Prasojo, A. (2011). Pengenalan Karakter Alphabet Menggunakan Jaringan Syaraf Tiruan. 7.

Puspitaningrum, D., Sari, D.K, & Susilo, B. (2014). Dampak Reduksi Sample Menggunakan Principle Component Analysis (PCA) Pada Pelatihan Jaringan Syaraf Tiruan Terawasi (Studi Kasus : Pengenalan Angka Tulisan Tangan). Pseudocode. 7.

Sharma, A., & Chaudhary, D.R. (2013). Character Recognition Using Neural Network. International Journal Of Engineering Trends And Technology, 6.

Sholahuddin, M.A. (2002). Penerapan Neural Network Tentang Metode Backpropagation Pada Pengenalan Pola Huruf. Kommit.

Sinurat, S. (2014). Analisa Sistem Pengenalan Wajah Berbentuk Citra Digital Dengan Algoritma Principal Component Analysis. 11.

Utari, W.A. (2011). Pengenalan Pola Dengan Menggunakan Metode Backpropagation Menggunakan s. 14.