

PENGAMBILAN FITUR ANGKA JAWA MENGUNAKAN *SHADOW FEATURE* *EXTRACTION*

Angsorul Anam¹, Susijanto Tri Rasmana², Madha Christianto Wibowo³

Jurusan Sistem Komputer
Institut Bisnis dan Informatika Surabaya
Surabaya, Indonesia

angSORULA6@gmail.com¹, susyanto@stikom.edu², Madha@stikom.edu³

Abstract—Many Applications developed to recognition handwritten called Optical Character Recognition (OCR), who generally it only presented alphabet recognition. Javanese numbers (aksara wilangan) or Javanese characters are culture of Indonesia from great grandmother and must be known by rising generation. The final project presented “Get Character Feature Handwriting Javanese Numbers Used Shadow Feature Extraction Method and Multi Layer Perceptron (MLP). The shadow feature method used to recognition characteristic of handwritten before it classification by MLP. Application test have two stage that is a sample training test 100 of data set and sample testing test 50 of data set. Percentage succeeded pattern concerning training samples 99% and error recognition 0,10%, whereas testing samples 90,8% and error recognition 9,2%.

Keywords—Javanese Numerals; Optical Character Recognition; Shadow Feature; MLP

I. LATAR BELAKANG

Tulisan aksara Jawa sebagai salah satu kebudayaan yang harus terus dilestirikan untuk dipelajari oleh anak cucu kita. di negara - negara lain misalnya Jepang, Korea mereka menggunakan tulisan mereka sendiri untuk diaplikasikan pada komputer, smartphone, dan sebagainya.

Banyak aplikasi-aplikasi yang dikembangkan untuk mengenali penulisan tangan. Pada umumnya pengenalan dilakukan pada huruf *alphabet* saja. Angka Jawa (aksara wilangan) sebagai salah satu bagian dari aksara Jawa kebudayaan peninggalan leluhur yang harus terus dilestirikan keberadaannya. Karakter angka Jawa terdiri dari beberapa angka dasar dari angka 0-9.

Dalam pengerjaan Tugas Akhir ini dibuat aplikasi untuk pengambilan fitur menggunakan *shadow feature extraction*. Metode ini digunakan untuk mengenali pola (ciri) dari citra tulisan tangan sebelum dilakukan klasifikasi karakter oleh MLP. *Shadow feature* merupakan metode pengenalan pola dengan menghitung panjang bayangan dari citra. Penggunaan metode ekstraksi ciri *shadow feature* berfungsi untuk memperkecil kemungkinan klasifikasi atau pengenalan karakter, disebabkan penulisan tangan masing-masing orang berbeda. Dikatakan demikian karena pola penulisan satu orang dengan orang lainnya akan memiliki perbedaan yang lebih sedikit dibandingkan citra *input* yang langsung dikenali, proses

ini dinamakan ekstraksi ciri atau pengenalan pola. Pola atau ciri pada metode *shadow feature* yang dimaksud adalah panjang bayangan pada tiap area yang dibagi seperti yang telah dijelaskan sebelumnya. MLP berfungsi untuk mengklasifikasikan karakter setelah mendapatkan *input* data-data hasil ekstraksi *shadow feature*.

A. Angka Jawa

Carakan (Abjad/Aksara Jawa) adalah huruf yang digunakan dalam ejaan bahasa Jawa yang pada dasarnya terdiri atas dua puluh aksara pokok (*nglegena*). Selain dua puluh aksara pokok tersebut, Aksara Jawa juga memiliki kelompok aksara kapital (*murda*), vokal (*swara*), rekaan (*rekan*), pengubah bunyi (sandangan), penanda gugus konsonan, penutup konsonan (*pasangan*), *pangon*, tanda baca, dan angka.

ᮘ	ᮙ	ᮚ	ᮛ	ᮜ
1	2	3	4	5
ᮝ	ᮞ	ᮟ	ᮠ	ᮡ
6	7	8	9	0

Gambar 1 Angka Jawa

B. Shadow Feature

Shadow feature adalah metode ekstraksi fitur yang digunakan untuk mengenali suatu pola tulisan, angka, atau simbol pada citra dengan cara menghitung jumlah fitur bayangan dari citra (Nibarandak, 2006). Citra yang dapat digunakan dalam metode ini hanya citra biner.

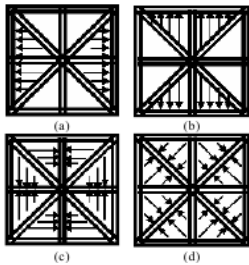
C. Multi Layer Perceptron (Mlp)

Multi layer perceptron (MLP) adalah perkembangan dari algoritma *perceptron* yang memiliki *layer* lebih dari sekedar *input layer* dan *output layer*. Pada MLP, terdapat *layer-layer* tambahan yang disebut dengan *hidden layer*. Dengan adanya *hidden layer* ini MLP dimungkinkan untuk dapat mengenali pola yang kompleks yang tidak dapat dikenali oleh struktur

jaringan *perceptron* konvensional. Struktur jaringan MLP dapat dilihat pada Gambar 4.

II. METODE

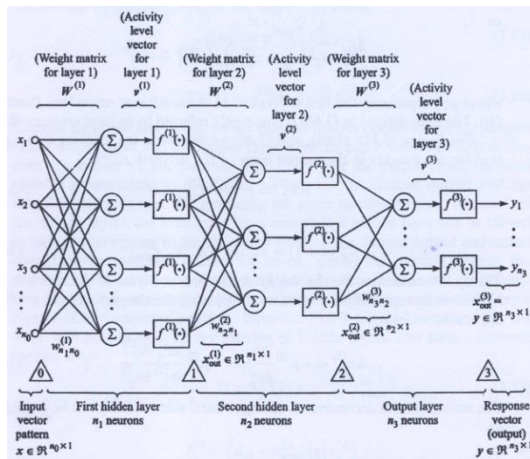
Model penelitian yang akan dilakukan adalah model penelitian pengembangan. Untuk mempermudah dalam memahami sistem yang akan dibuat dapat dijelaskan melalui blok diagram pada Gambar 5.



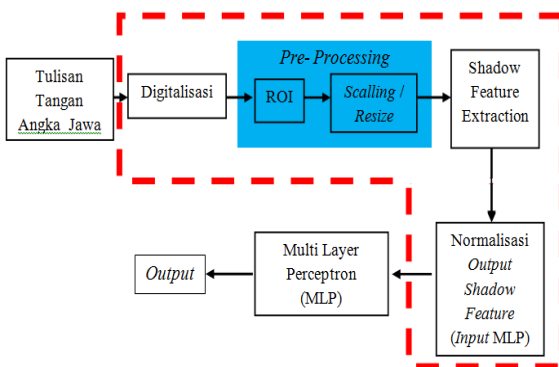
Gambar 2 ilustrasi untuk *Shadow Feature* (a - d).



Gambar 3 Hasil Citra Fitur Bayangan (Basu dkk, 2005).



Gambar 4 Struktur Jaringan MLP (Ham, Kostanic, 2001)



Gambar 5 Blok Diagram Sistem

Gambar 5 adalah blok diagram sistem keseluruhan perangkat lunak yang dibuat, akan tetapi blok di dalam garis putus-putus berwarna merah merupakan fokus utama dari pengerjaan Tugas Akhir ini.

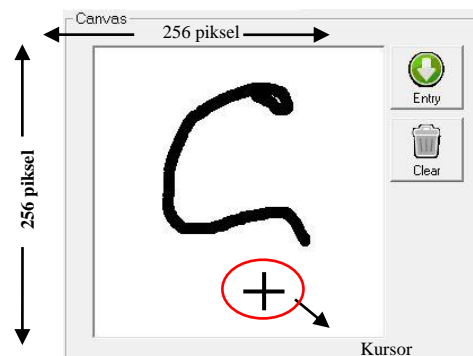
1. Tulisan tangan angka Jawa diperoleh dari 150 partisipan dari kanvas yang tersedia maupun citra yang didapat dari *tablet*.
2. Digitalisasi adalah proses untuk membaca nilai piksel citra jika piksel berwarna hitam maka nilai piksel yang dibaca "0" jika piksel berwarna putih nilai piksel yang dibaca "255".
3. *Preprocessing* terdiri dari 2 proses yaitu:
 - a. *Region of interest*(ROI) adalah proses yang digunakan untuk mengabaikan bagian dari kanvas yang tidak perlu agar proses ekstraksi terfokus pada citra. Dalam hal ini baik citra yang diambil dari kanvas maupun citra yang diambil dari *tablet* dilakukan pada kanvas yang disediakan.
 - b. *Scaling / resize* ukuran citra menjadi 64 x 64 piksel.
4. *Shadow feature extraction* merupakan pokok dari perangkat lunak yang dibuat yaitu proses perhitungan panjang fitur bayangan. Normalisasi *output shadow feature* dimaksudkan untuk menghindari *overflows* saat proses pelatihan pada MLP.

III. PERANCANGAN PERANGKAT LUNAK

A. Desain Kanvas

Pengambilan pola penulisan angka jawa dari partisipan dilakukan pada media interaktif berupa kanvas yang digunakan untuk menangkap pola tulisan tangan dari partisipan. Kanvas tersebut akan dibuat dengan *software visual basic* menggunakan *tool picture box*. Resolusi dari kanvas sebesar 256 x 256 piksel.

Dalam penggunaan kanvas tersebut partisipan cukup menggerakkan (*drag*) kursor sesuai pola angka jawa yang akan digambar pada kanvas menggunakan *mouse/mouse Pen* seperti pada Gambar 6 berikut.



Gambar 6 Kanvas

Pola penulisan angka jawa dari partisipan disimpan dengan format *bitmap* karena *bitmap* merupakan format gambar yang bisa dibaca oleh program grafis manapun.

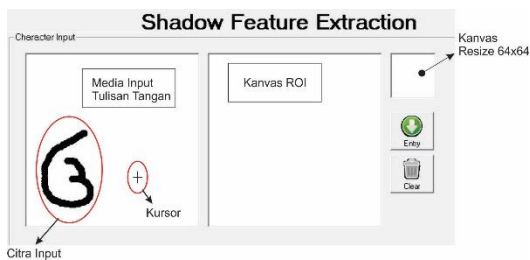
B. Digitalisasi

Digitalisasi adalah proses pembacaan warna piksel pada kanvas. Kanvas berwarna putih nilai biner 255 dan citra tulisan

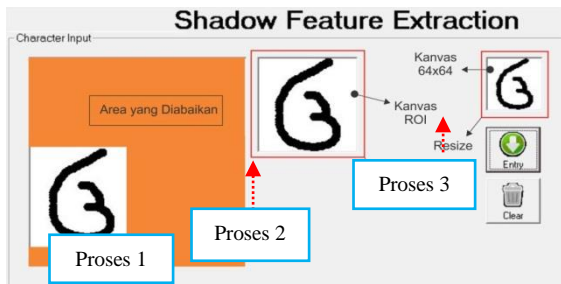
berwarna hitam nilai biner 0. Citra yang telah digambar di scan per baris untuk dibaca warna pikselnya. Dalam scanning ini dilakukan pengklasifikasian terhadap piksel yang diambil, Koordinat piksel-piksel tersebut disimpan ke dalam array dua dimensi dengan format baris dan kolom.

C. ROI (Region Of Interest) dan Scalling / Resize

Proses ROI berfungsi untuk mengabaikan bagian (piksel) dari kanvas yang dirasa tidak perlu untuk diolah datanya. Selain itu setiap orang bisa berbeda-beda dimensi tulisan tangannya. Dalam program sebelum dilakukan proses ROI terlebih dahulu dilihat berapakah ukuran yang digambar oleh partisipan. Media input (picturebox) berukuran 256 x 256 (apabila citra diambil dari komputer) diambil kemudian digambar kembali ke kanvas ROI yang berukuran sama seperti Gambar 7, namun pada proses tersebut kanvas (picturebox) akan menyesuaikan dimensi dari tulisan tangan partisipan agar dapat terlihat perbedaan sebelum citra melalui ROI. Selanjutnya citra di-resize ke ukuran 64 x 64 piksel seperti Gambar 8.



Gambar 7 Citra yang Akan Melewati Proses ROI dan Scalling / Resize



Gambar 8 Proses ROI dan Resize

Langkah-langkah untuk melakukan ROI dalam program digambarkan secara umum dalam flowchart Gambar 9 berikut ini.

D. Ekstraksi Ciri Shadow Feature

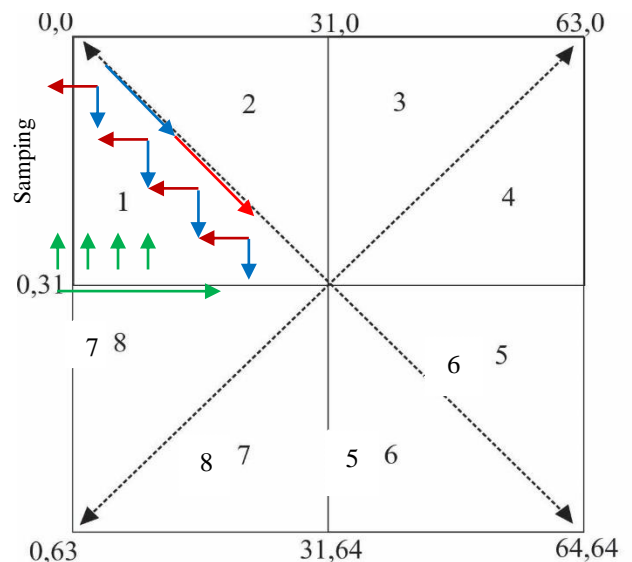
Citra penulisan tangan dari user yang sudah melalui ROI (Region Of Interest) dan scalling/ resize baru dapat diolah untuk diekstrak. Shadow feature extraction adalah pengenalan pola dari suatu citra dengan cara menghitung panjang bayangan dari bagian-bagian citra tertentu. Mula-mula citra digambar/dicetak kembali pada kanvas (picturebox) dimensi 64x64, kemudian kanvas tersebut dibagi menjadi delapan bagian yang masing-masing bagian mempunyai 3 sisi, jadi total fitur bayangan yang didapat berjumlah 24 fitur seperti Gambar 11.

Pada Gambar 10 telah tercantum koordinat-koordinat piksel kanvas (picturebox), nomer area, dan arah gerak proyeksi

bayangan pada ke tiga sisi pada area 1. Bagian piksel dari citra yang berada di dalam area-area pada gambar diatas akan dihitung panjang bayangannya terhadap sisi yang menghadapnya.



Gambar 9 Flowchart Preprocessing (ROI dan Resize)

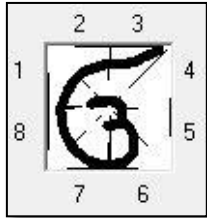


Gambar 10 Ilustrasi Pembagian Area Citra

Arah proyeksi fitur bayangan ke tiap sisi dari area 1 menunjukkan letak bayangan yang terproyeksi. Cara menghitung fitur bayangan dilakukan dengan scanning pada tiap sisi-sisinya yaitu sisi samping, alas dan diagonal. Scanning

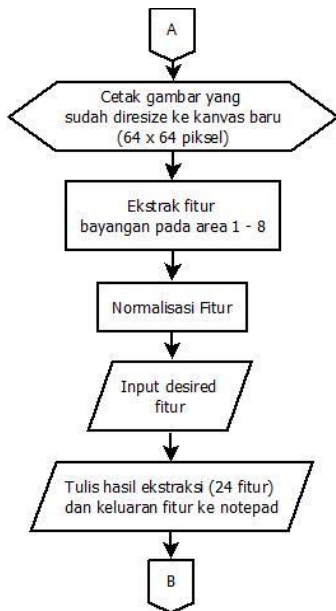
berhenti jika scan pertama kali menemukan piksel yang berwarna hitam pada tiap baris atau kolom tergantung arah proyeksi bayangannya. Pergerakan dan arah *scanning* pada area 1 diilustrasikan oleh anak panah yang berwarna merah, biru dan hijau.

Ekstraksi fitur bayangan samping area 1 *scanning* bergerak sepanjang sisi diagonal ke bawah, sedangkan arah *scanning* dari sisi diagonal ke kiri (panah warna merah). Ekstraksi fitur alas pergerakan *scanning* sepanjang sisi diagonal ke bawah dan arah *scanning* dari sisi diagonal ke bawah (panah warna merah). Fitur bayangan diagonal pergerakan *scanning* dari samping ke kanan dan arah *scanning*nya dari alas ke atas (panah warna hijau). *Seterusnya* sampai dengan area 8 dengan menyesuaikan hadap masing area (segitiga) diatas.



Gambar 11 Tampilan Dalam Software Citra Yang Sudah Diekstrak

Langkah-langkah ekstraksi dijelaskan dengan *flowchart* pada Gambar 3.11 berikut:



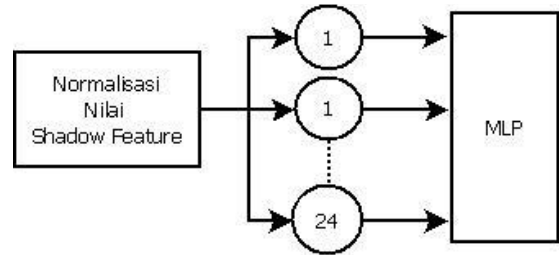
Gambar 12 Flowchart Preprocessing (ROI dan Resize)

E. Input Multi Layer Perceptron (MLP)

Keluaran dari ekstraksi *shadow feature* adalah panjang bayangan fitur bayangan dari tiap-tiap sisi masing-masing area yang telah di normalisasi. Nilai fitur bayangan dari tiap area ini kemudian di normalisasi dengan cara dibagi dengan panjang total masing-masing sisi.

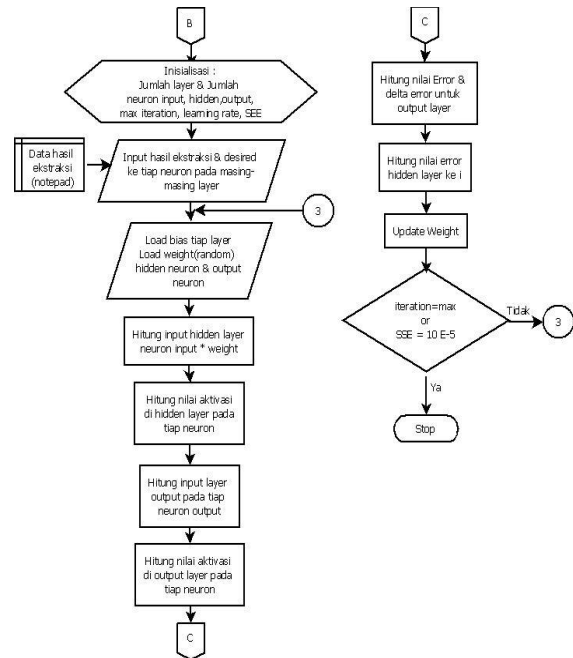
Jumlah nilai normalisasi berjumlah 24 buah begitu juga dengan *neuron input* dari MLP seperti pada Gambar 13. Nilai

normalisasi tersebut dimasukkan ke MLP untuk dilakukan pembelajaran.



Gambar 13 Data ekstraksi diumpukan pada MLP

Langkah meng-*input*-kan nilai output ekstraksi dan pengujian MLP serta pengolahan data ekstraksi diperlihatkan dalam *flowchart* Gambar 14 dan Gambar 15 berikut ini.



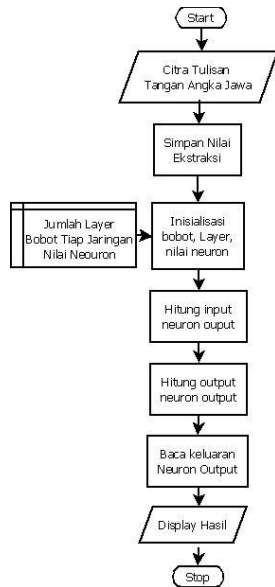
Gambar 14 Flowchart Training MLP

IV. TEKNIK PENGUMPULAN & ANALISIS DATA

A. Pengumpulan Data

Data sampel primer untuk 1 *set* (10 pola angka) diperoleh dengan mengumpulkan hasil tulisan tangan dari orang yang berbeda. Karena ada 100 data sampel pembelajaran dan 50 data sampel uji untuk masing-masing angka, diperlukan 150 orang yang masing-masing memberikan 1 *set* sampel tulisan tangannya dalam Angka Jawa.

Pengumpulan data sekunder dilakukan dengan melakukan percobaan di laboratorium setelah perangkat lunak sudah selesai dibuat dan sudah diperoleh bobot yang optimal seperti yang dijelaskan pada sub bab 3.3. Sistem MLP dengan bobot optimal tersebut akan diuji menggunakan pola Angka, baik dari sampel yang digunakan untuk pembelajaran, pengujian maupun citra yang baru digambarkan. Hasil pengujian antara sampel pelatihan, pengujian, dan data baru akan saling dibandingkan.



Gambar 15 Flowchart Testing MLP

B. Pelatihan MLP

Aplikasi ini akan menghasilkan 2 buah file log untuk bobot yang dihasilkan. Bobot hasil akhir pembelajaran yang tercapai apabila syarat error minimum atau iterasi maksimum terpenuhi, akan disimpan ke dalam file *weight.dat*. Selain itu, ada file bernama *weight2.dat* untuk menyimpan informasi bobot setiap 100 iterasi. Hal ini diperlukan sebagai tindakan preventif bila terjadi hal-hal yang dapat menghentikan aplikasi sebelum mencapai akhir pembelajaran, yang bisa disebabkan karena komputer *hang* atau putusnya aliran listrik.

Proses pembelajaran dilakukan dengan konfigurasi sebagai berikut:

- Jumlah *hidden layer* : 3
- Jumlah *neuron input* : 24
- Jumlah *neuron layer 1* : 40
- Jumlah *neuron layer 2* : 40
- Jumlah *neuron layer 3* : 40
- Jumlah *neuron output* : 10
- Laju pembelajaran (μ) : 0.1
- *Error* minimum : $10E-4$
- Iterasi maksimum : $10E+5$

Proses pembelajaran dilakukan sebanyak 133.500 iterasi dengan menggunakan 100 set sampel input dengan nilai *error* akhir sebesar $6.76168169753287E-04$. Ini berarti bahwa proses pembelajaran selesai hingga target *error* minimum terhadap data sampel terpenuhi. Bobot yang dihasilkan akan disimpan pada file *weight.dat*

C. Analisis Data

Data yang sudah terkumpul baik citra maupun hasil normalisasi ekstraksi (*notepad*) akan dilatihkan terhadap MLP. Akan dilakukan pengujian kembali pada data sampel yang diperoleh untuk menghitung nilai *error* pada perangkat lunak sesudah pembelajaran.

Setelah menganalisa data sampel untuk pelatihan berjumlah 100 set tulisan angka jawa, proses selanjutnya adalah analisa

data sampel pengujian, sampel pengujian adalah data yang belum pernah dilatihkan pada MLP sebelumnya. Data tersebut diperoleh dari partisipansebanyak 50 set tulisan angka jawa, dan sudah disimpan hasil normalisasi ekstraksinya. Kemudian pengujian secara langsung dengan citra tulisan angka jawa yang baru diluar data sampel pelatihan dan sampel uji. Ada dua parameter yang akan diperhatikan dalam pengujian yang akan dihasilkan pada tiap-tiap pengujian, yaitu apakah pengklasifikasian yang dihasilkan benar atau salah.

V. HASIL DAN PENGUJIAN

Tujuan dari pengujian ini adalah untuk mengetahui apakah aplikasi ekstraksi dan pengenalan angka Jawaan MLP dapat berjalan dengan baik. Pengujian yang dilakukan dilakukan pada *set* sampel pelatihan dan *set* sampel uji.

A. Sampel Training

Dari hasil pengujian ini jumlah data pada masing-masing sampel angka (0 – 9) sebanyak 100 set data, Presentase keberhasilan pengenalan pola terhadap sampel *training* sebesar 99%.

Tabel 1 Pengujian Sampel Pelatihan

Angka	Jumlah Data (buah)	Data Yang Berhasil Dikenali (buah)	Kesalahan Pengenalan (buah)	Presentase Keberhasilan (%)
0	100	100	0	100,00 %
1	100	100	0	100,00 %
2	100	100	0	100,00 %
3	100	100	0	100,00 %
4	100	100	0	100,00 %
5	100	100	0	100,00 %
6	100	100	0	100,00 %
7	100	99	1	99,00 %
8	100	100	0	100,00 %
9	100	100	0	100,00 %
Persentase Rata-rata				99,90%

Pada tabel 1 tingkat keberhasilan pengenalan aplikasi setelah pelatihan adalah sebesar 99,90%.

B. Sampel Testing

Tabel 2 Pengujian Sampel Test

Angka	Jumlah Data (buah)	Data Yang Berhasil Dikenali (buah)	Data Yang Tidak Berhasil Dikenali (buah)	Presentase Keberhasilan (%)
0	50	50	0	100,00 %
1	50	50	0	100,00 %
2	50	49	1	98,00 %
3	50	42	8	84,00 %
4	50	49	1	98,00 %
5	50	40	10	80,00 %
6	50	32	18	64,00 %
7	50	48	2	96,00 %
8	50	48	2	96,00 %
9	50	46	4	92,00 %

Persentase Rata-rata	90,80%
----------------------	--------

Dari hasil pengujian ini jumlah data pada masing-masing sampel angka (0 – 9) sebanyak 50 *set* data, dan berhasil dikenali 90,8% saja.

Tabel 3 Analisa Kesalahan Tiap Karakter

Angka	Dikenali Sebagai Angka									Presentase Error	
	0	1	2	3	4	5	6	7	8		9
0											0%
1											0%
2					1						2,00%
3	6	2									16,00%
4						1					2,00%
5	4	2	1	3							20,00%
6	7				11						36,00%
7									1	1	4,00%
8								1		1	4,00%
9	1						2	1			8,00%
Rata-Rata Persentase Error											9,20%

Kesalahan pengenalan pada angka “6” adalah yang paling banyak dengan presentase error sebesar 36%, dan angka “5” sebesar 20% dari 50 sampel data. Kemungkinan yang terjadi kesalahan tersebut dikarenakan pola dari citra tersebut mirip atau memang pola dari citra memang tidak sesuai dengan pola Angka Jawa yang sesuai.

VI. KESIMPULAN

Adapun kesimpulan yang dapat dituliskan setelah melakukan analisa kesalahan yang terjadi dikarenakan beberapa angka Jawa ada kemiripan ciri. seperti pada angka “0” dan angka “4” (angka Jawa), juga karena citra sampel yang tidak memenuhi pola penulisan angka Jawa yang sesuai.

Tingkat keberhasilan pengenalan aplikasi pada sampel *training* sebesar 99,90%, sedangkan pada sampel *testing* tingkat keberhasilannya sebesar 90,8% dengan rata-rata persentase error sebesar 9%.

DAFTAR PUSTAKA

[1] Basu, S., dkk., 2005, Handwritten ‘Bangla’ Alphabet Recognition using an MLP Based Classifier, Proceeding of the 2nd National Conference on Computer Processing of Bangla, hal. 285 – 291. Dhaka.

[2] Basu, S., dkk., 2005, An MLP Based Approach for Recognition of Handwritten ‘Bangla’ Numerals, Proceeding 2nd Indian International Conference on Artificial Intelligence, hal. 407 – 417. Pune.

[3] Chaudhuri, B.B. dan Bhattacharya, U., 2000, Efficient Training and Improved Performance of Multilayer Perceptron in Pattern Classification, *Neurocomputing*, vol. 34, hal. 11-27.

[4] Darusuprta, dkk., 2002, *Pedoman Penulisan Aksara Jawa*, Yayasan Pustaka Nusantara, Yogyakarta.

[5] Das, N., dkk., 2006, Handwritten Arabic Numeral Recognition using a *Multi layerPerceptron*, Proceeding National Conference on Recent Trends in Information Systems, hal. 200 – 203.

[6] Das, N., dkk., 2010, Handwritten Bangla Basic and Compound Character Recognition Using MLP and SVM, *Journal of Computing*, Vol. 2, No. 2, hal. 109 – 115.

[7] Fausett, L. 2006. *Fundamentals of Neural Networks*. Prentice-Hall, New York.

[8] Fiset, R., dkk., 1998, Map-Image Matching Using a *Multi-Layerperceptron*: The Case of the Road Network, *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 53, hal. 76-84.

[9] Ham, F.M. dan Kostanic, I., 2001, *Principles of Neurocomputing for Science & Engineering*. McGraw-Hill, New York.

[10] Hasibuan, F.M., 2011, Desain dan Implementasi Sistem Penerjemah Aksara Jawa Ke Huruf Latin Berbasis Pengolahan Citra Digital dan Jaringan Syaraf Tiruan Self-Organizing Map (SOM), Tugas Akhir Mahasiswa Institut Teknologi Telkom.

[11] Kiong, L.V., 2006, *Visual Basic 6 Made Easy*, ISBN: 141962895X, 2006

[12] Mukherjee, S., 2010, Recognition of Handwritten Bengali Character Based On Character Features, (M.Tech IT Thesis), Jadavpur University, Kolkata, India.

[13] McCulloch, W.S. dan Pitts, W., 1943, A Logical Calculus of the Ideas Immanent in Nervous Activity, *Bulletin of Mathematical Biophysics*, vol 5: hal. 115-133.

[14] Nurtanio, Ingrid, Astuti, E. R., Purnama, I. K. E., Hariadi, M., Purnomo, M. H., *Classifying Cyst and Tumor Lesion Using Support Vector Machine Based on Dental Panoramic Images Texture Features*, IAENG International Journal of Computer Science, vol. 40, no. 1, Feb. 2013.

[15] Rosenblatt, F., 1958, The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain, *Psychological Review*, vol. 5: hal 368-408.

[16] Rumelhart, D.E. dkk., 1986, Learning Representations by Back-Propagating Errors. *Nature*, vol. 323: hal. 533-536.

[17] Wibowo, M.C. dan Wirakusuma, S., 2013, *Pengenalan Pola Tulisan Tangan Aksara Jawa “Ha Na Ca Ra Ka” Menggunakan Multi Layer Perceptron*, Prosiding Seminar Nasional Teknologi Informasi (SNASTI) 2013, Oktober 2013, Surabaya, hal. ICCS-27 – ICCS-32

[18] Zhang, Z. dkk., 1998, Comparison Between Geometry-Based and Gabor-Wavelets-Based Facial Expression Recognition Using Multi-Layer Perceptron. *Automatic Face and Gesture Recognition. Third IEEE International Conference*. Naara

[19] Trier O.D., Anil K.J., dan Torfinn, T., ” Feature Extraction methods for Character Recognition –A survey “, *Pattern Recognition*, Vol29, No.4, pp-641-662, 1996