

Pengenalan Tulisan Tangan Bahasa Arab Menggunakan Metode *Probabilistic Neural Network*

Vidia

*Program Teknik Informatika, Fakultas Ilmu Komputer,
Universitas Nahdlatul Ulama Sidoarjo
Jln. Mongonsidi Kav. DPR Sidoklumpuk Sidoarjo
vidia9981@gmail.com*

Abstrak

Perkembangan teknologi yang sangat pesat saat ini tidak membuat pendokumentasian data atau informasi dengan tulisan tangan dihilangkan. Bahkan dalam berbagai penelitian, menulis dengan tangan terbukti memiliki efek yang sangat baik untuk kecerdasan otak. Dalam penelitian ini akan dibangun aplikasi pengenalan tulisan tangan menggunakan metode Probabilistic Neural Network dengan data training. Data dengan masing-masing huruf terdiri dari 10 sampel yang digunakan untuk pelatihan sistem. Dalam pengenalan pola, data training memiliki peran yang sangat penting dalam menentukan akurasi sistem, selain itu ukuran dimensi pixel baris dan kolom citra juga sangat berpengaruh. Proses pengenalan citra tulisan tangan ini melalui beberapa tahapan yaitu data training dan data testing. Pada tahap data training ada beberapa data yang harus di training sebelum data tersebut di testing. Setelah di training data akan di testing kemudian diubah menjadi RGB dalam arah keabuan. Data yang di testing awalnya tersambung agar bisa terpisah, maka data tersebut dilakukan cropping (pemotongan). Kemudian data akan di atur bagian ekstraksi fiturnya untuk menghasilkan data training dan data testing. Dalam penelitian ini, matriks pixel citra akan diubah dalam bentuk vektor.

Kata kunci: *Pengenalan Tulisan Tangan, Ekstraksi Fitur, Probabilistic Neural Network (PNN)*

Abstract

Technological developments are very rapid at this time does not make documentation of data or information with handwriting removed. Even in various studies, writing by hand has proven to have a very good effect on brain intelligence. In this research, handwriting recognition applications will be built using the Probabilistic Neural Network method with training data. Data with each letter consists of 10 samples used for system training. In pattern recognition, training data has a very important role in determining the accuracy of the system, besides the size of the pixel dimensions of rows and columns of image is also very influential. This handwriting image recognition process goes through several stages, namely training data and testing data. In the training data phase there are some data that must be trained before the data is tested. After training the data will be tested and then converted to RGB in the gray direction. The data tested is initially connected so that it can be separated, then the data is cropped. Then the data extraction feature will be set to produce training data and testing data. In this study, the image pixel matrix will be changed in vector form.

Keywords: *Handwriting Recognition, Feature Extraction, Probabilistic Neural Network (PNN).*

1. PENDAHULUAN

Huruf yaitu lafal yang tidak dapat membentuk makna tersendiri karena harus dirangkai dengan huruf yang lainnya (tersambung). Kumpulan huruf yang dapat membentuk arti biasanya 3 huruf, oleh karena itu pada bentuk-bentuk tertentu ada satu huruf yang sudah mempunyai arti. Sedangkan hijaiyah yaitu disebut juga alfabet arab. Kata alfabet itu sendiri berasal dari bahasa arab alif, ba', ta'. Namun ada beberapa orang yang mempunyai pendapat berbeda tentang definisi huruf hijaiyah. Huruf hijaiyah dimulai dari huruf alif dan di akhiri dengan huruf ya' secara terpisah-pisah.

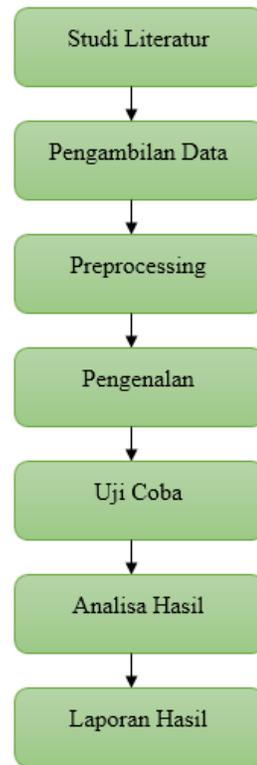
Pengenalan pola adalah mengelompokkan data numerik dan simbolik (seperti citra) secara otomatis oleh mesin (komputer). Tujuan pengenalan pola yaitu untuk mengenali suatu objek yang ada dalam citra digital. Beberapa manusia telah mengenali objek yang dilihatnya karena otak manusia telah belajar memproses satu objek dengan objek lainnya. Komputer hanya bisa menerima inputan data berupa citra objek yang akan diidentifikasi sehingga proses citra tersebut akan mengeluarkan hasil berupa informasi atau deskripsi objek di dalam citra tersebut.

Pengenalan pola huruf hijaiyah dilakukan dalam tugas akhir ini memiliki dua tahap, yaitu ekstraksi dan klasifikasi. Terdapat beberapa metode dalam menentukan ekstraksi seperti Direction Feature (DF), Gabor filter, dan lain-lain. Kemudian ada beberapa metode dalam mengklasifikasi, seperti Support Vector Machine (SVM), Jaringan Syaraf Tiruan (JST), dan Naive Bayes Classifier (NBC). Metode ini melakukan pendekatan dengan cara melakukan deteksi tulisan dari segmen-segmen garis pembentuk karakter dan pixel-pixel, namun kemudian digantikan dengan nilai arah yang sesuai (Morwati, 2014).

Dari berbagai hasil penelitian dan penjelasan tersebut, menulis dengan tangan terbukti memiliki efek yang lebih baik daripada mengetik. Namun dengan banyaknya dokumen yang ditulis menggunakan tangan pada media kertas menimbulkan sebuah permasalahan dalam pemrosesannya, karena untuk dapat memproses data yang ditulis pada media kertas dengan komputer. Sehingga data akan diketik kembali dalam bentuk teks digital, hal yang seperti ini bisa menyusahkan seseorang karena harus melakukan pekerjaan yang berulang-ulang. Untuk mengatasi masalah tersebut dibutuhkan sebuah aplikasi yang dapat mengenali karakter tulisan tangan sehingga pemrosesan data teks tulisan tangan bisa dikerjakan dengan lebih cepat dan efisien. Dalam penelitian ini akan mencoba menginputkan sebuah citra tulisan tangan yang diperoleh menggunakan scanner dengan output berupa teks digital dari citra yang diinputkan. Pada penelitian ini akan mencoba mengklasifikasi citra tulisan tangan pada huruf hijaiyah dengan menggunakan metode Probabilistic Neural Network (PNN).

2. METODE PENELITIAN

Dalam penelitian ini, terdapat 4 tahapan utama, yaitu perancangan dan pengumpulan data, desain sistem, implementasi dan uji coba sistem. Blok diagram tahapan penelitian ditunjukkan pada Gambar berikut.



Gambar 1. Tahapan Penelitian

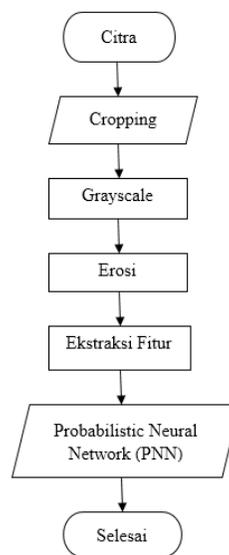
Secara umum, penelitian ini dilakukan dalam beberapa tahap yaitu dimulai dari studi literatur, perumusan masalah dan hipotesis, perancangan metode dan implementasi, serta uji coba dan evaluasi. Sedangkan penulisan laporan penelitian dimulai dari awal sampai akhir penelitian. Sebagai berikut :

- a. *Studi literatur*, mempelajari berbagai literatur tentang tulisan tangan bahasa arab dalam citra digital menggunakan metode Probabilistik Neural Network.
- b. *Perumusan masalah*, merumuskan masalah yang akan diteliti dan mencari solusi dari penelitian tersebut.
- c. *Perancangan metode dan implementasi*, merancang dan mengimplementasikan metode penyelesaian dari masalah yang telah dirumuskan berdasarkan penelitian yang diperoleh dari studi literatur.

Dalam pembangunan aplikasi pengenalan tulisan tangan ini terdapat dua macam data yang dibutuhkan, yaitu data training dan data testing. Data training berupa tulisan tangan yang diperoleh dari beberapa responden dari berbagai umur dalam sebuah kalimat yang mengandung huruf *ba'* - *qaf* yang tidak bersambung, media yang digunakan untuk menulis tangan yaitu kertas HVS putih. Kemudian data akan di-scan dengan scanner dalam bentuk citra yang berformat JPEG. Data citra tersebut kemudian akan di-crop per karakter huruf dan disimpan dalam sebuah folder data training. Setiap

huruf dalam data training akan mewakili satu kelas dengan jumlah huruf 18 sampel per-huruf. Citra tulisan tangan masuk ke dalam ekstraksi fitur untuk di ambil matriksnya dengan format “.mat”. Data testing juga berupa data tulisan tangan yang ditulis pada media kertas HVS. Kemudian diperoleh dari beberapa responden yang nantinya akan di-scan dengan scanner dalam bentuk citra yang berformat JPEG. Dan dikenali oleh aplikasi pada saat uji coba sistem untuk mengetahui apakah dapat berjalan sesuai tujuan. Untuk mengukur akurasi metode *Probabilistik Neural Network* dalam mengenali tulisan tangan. Data testing yang dapat dikenali dapat berupa citra huruf, kata satu baris kalimat yang berhuruf kecil dan tidak bersambung.

- d. *Uji coba dan evaluasi*, melakukan pengujian dan evaluasi terhadap penelitian yang telah dirancang dengan beberapa laporan. Uji coba dilakukan berdasarkan tingkat kemudahan dalam mengenali tulisan tangan bahasa arab menggunakan citra dengan metode Probabilistik Neural Network :



Gambar 2. proses pengenalan data citra tulisan tangan.

- 1) *Input data* : Data yang akan di input berupa tulisan tangan yang akan diproses dalam aplikasi Matlab.
- 2) *Cropping* : Data citra asli dari hasil scan akan di crop satu per satu. Agar mempermudah untuk mengklasifikasi tiap huruf hijaiyah.
- 3) *Grayscale* : Data citra RGB kemudian di ubah menjadi Grayscale sehingga menjadi warna abu-abu.
- 4) *Erosi* : Dari citra biner kemudian di erosi untuk menentukan nilai vektor 1s dan 0s. Dengan perintah “strel”, maka blok akan menerapkan setiap objek ke seluruh matriks.
- 5) *Ekstraksi Fitur* : Data citra asli akan di ekstraksi menggunakan matlab kemudian akan ditentukan akurasi dari tiap nilai huruf hijaiyah .

- e. *Penyusunan laporan*, penyusunan laporan dilakukan mulai dari awal hingga akhir penelitian.

3. HASIL DAN PEMBAHASAN

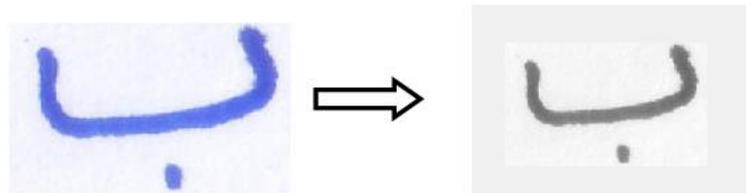
3.1. Implementasi

Dari semua rancangan yang akan dibangun di atas, maka ada beberapa tahapan yang akan di implementasi terhadap hasil rancangan yang sudah dibuat. Tahapan implementasi yaitu penerapan algoritma *Probabilistik Neural Network* ke dalam aplikasi, dalam pembuatan tampilan antarmuka dan pembuatan fungsi – fungsi dengan menggunakan aplikasi MATLAB.

Tabel 1. Spesifikasi data citra yang diambil

No	Spesifikasi
1.	Citra huruf berformat .jpg
2.	Citra huruf hijaiyah hanya diambil 10 huruf saja
3.	Data citra huruf hijaiyah latih berjumlah sebanyak 100 gambar

Pertama konversi RGB ke Grayscale program akan membaca citra asli lebih dulu. Kemudian pada tahap ini gambar asli yang berwarna akan di ubah menjadi warna abu-abu. Seperti gambar dibawah ini

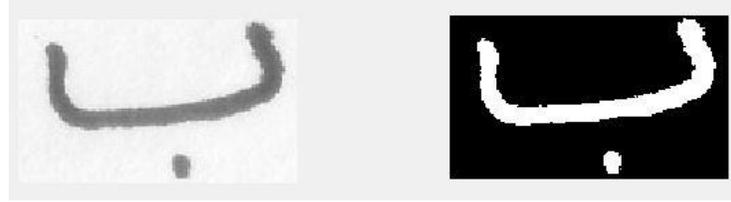


Gambar 3. Konversi RGB ke Grayscale

Kemudian konversi grayscale ke biner, Pada tahapan ini dari citra *grayscale* kemudian di ubah ke biner. Dimana citra yang awalnya berwarna abu-abu di ubah menjadi biner. Pada citra biner akan ditentukan nilai pikselnya yang berupa nilai 0 atau 1. Seperti gambar di bawah ini :



Gambar 4. Konversi Grayscale ke Biner



Gambar 5. Konversi *Grayscale* ke komplemen

Pada tahapan selanjutnya, konversi biner ke erosi, dari citra komplemen yang kemudian diubah menjadi erosi sehingga menjadi citra huruf yang tipis. Seperti gambar di bawah ini :



Gambar 6. Konversi Biner ke Erosi

3.2. Pengujian akurasi sistem dengan metode *K-fold cross validation*

K-fold cross validation adalah teknik yang dapat digunakan apabila memiliki jumlah data yang terbatas (jumlah instance tidak banyak). *K-fold cross validation* merupakan salah satu metode yang digunakan untuk mengetahui rata-rata keberhasilan. Dari suatu sistem dengan cara melakukan perulangan dengan mengacak atribut masukan sehingga sistem tersebut teruji untuk beberapa atribut input yang acak.

K-fold cross validation diawali dengan membagi data sejumlah *n-fold* yang diinginkan. Dalam proses *cross validation* data akan dibagi dalam *n* buah partisi dengan ukuran yang sama selanjutnya proses testing dan training dilakukan sebanyak *n* kali. Dalam iterasi ke-*i* partisi *Di* akan menjadi data testing dan sisanya akan menjadi data training. Untuk penggunaan jumlah *fold* terbaik untuk uji validitas, dianjurkan menggunakan *10-fold cross validation*. Keseluruhan total data yang akan kita gunakan yaitu 100 data, yang terdiri dari 10 huruf hijaiyah dan data berformat “jpg”. Keseluruhan data bisa dilihat pada lampiran. Untuk pengujian akurasi menggunakan data 100 dan dibagi dengan beberapa *k-fold* yang berubah untuk mendapatkan akurasi.

Pada pengenalan citra huruf akan dilakukan pengelompokan data menjadi data latih dan data uji. Pengelompokan data ini terdapat permasalahan dalam menentukan model seleksinya dan hasil estimasi. Dari salah satu teknik pengelompokan ini adalah menggunakan *cross validation* yang terbagi menjadi 3 macam yaitu mengelompokan data secara acak (*Random Subsampling*), *K-fold cross validation*, *leave-one-out cross validation*. *K-fold cross validation* digunakan dalam metode pembelajaran pada data latih untuk menentukan keakuratan suatu algoritma sehingga dapat menguji data yang tidak terlatih (data uji).

3.3. Hasil klasifikasi

Klasifikasi menggunakan data sebanyak 100 data citra, dengan mengubah parameter k-fold. Berikut ini adalah hasil uji coba dengan nilai k-fold yang berbeda :

Tabel 2 Hasil Uji coba nilai K-fold

K-fold	Akurasi
7	90.1461%
3	89.8148%
5	90%
9	89.8355%

Pada pengujian ini terdapat tingkat nilai akurasi yang paling tinggi pada perhitungan k-fold 7 dengan akurasi 90.1461%.

4. KESIMPULAN

Pada penelitian ini memperoleh kesimpulan bahwa pengujian terhadap jarak. Jika semakin kecil jaraknya maka akurasi yang didapat semakin banyak. Pada jarak piksel menggunakan nilai 1. Pada penelitian ini didapatkan akurasi tertinggi dengan nilai k-fold 7 dengan prosentase 90.14615%.

5. SARAN

Untuk pengembangan aplikasi pengenalan tulisan tangan ini dikemudian hari diperlukan beberapa perbaikan dan tambahan untuk mendapatkan hasil yang lebih maksimal, diantaranya :

1. Memperbanyak objek yang dapat dikenali, misalnya dengan menambahkan huruf hijaiyah.
2. Penelitian dapat dikembangkan dengan objek berupa tulisan tangan huruf hijaiyah yang memiliki nilai berbeda – beda.
3. Ada 10 anak yang menulis huruf hijaiyah digunakan sebagai citra inputan yang ditulis dengan cara manual kemudian di scan.
4. Melakukan penelitian dengan cara mencoba berbagai macam metode yang lebih mudah dan beragam.

UCAPAN TERIMA KASIH

Terimakasih kepada pihak para rekan yang turut serta membantu dalam perancangan dan penyusunan penelitian dan jurnal ini.

DAFTAR PUSTAKA

- Morwati, M. (2014). *Pengenalan Citra Huruf Alphabet Tulisan Tangan Menggunakan Metode Naïve Bayes Classifier* (Doctoral dissertation, Universitas Islam Negeri Maulana Malik Ibrahim).
- Kanta, I. A. (2013). *Pengenalan Pola Huruf Hijaiyah Tulisan Tangan Menggunakan Logika Fuzzy dengan Jaringan Syaraf Tiruan Backpropagation* (Doctoral dissertation, Universitas Muhammadiyah Surakarta).
- Hara, E. (2016). *Sistem Pengenalan Tulisan Tangan Aksara Lampung dengan Metode Deteksi Tepi (Canny) Berbasis Jaringan Syaraf Tiruan Backpropagation* (Doctoral dissertation, Fakultas Teknik). Carol seefelt, 2015. *Definisi Huruf Hijaiyah*.
- Ardianto Elvinaro, 2007. *Definisi Citra Digital Dengan Menggunakan Aplikasi Matlab*.
- Putra, D. 2010. *Pengolahan Citra Digital*. Yogyakarta : Andi.
- M. C, Wijaya dan A. Prijono, 2007. *Pengolahan Citra Digital Menggunakan Matlab*. Bandung : Informatika.
- Gilewski, J., Phil Philips, S. Yanushkevich, D. Popel. 1997. "Education Aspects: Handwriting Recognition – Neural Networks – Fuzzy Logic". Proceedings of the IAPR International Conference on Pattern Recognition and Information Processing.