

## ABSTRAK

Pengenalan huruf atau teks merupakan kajian penelitian yang banyak dikembangkan dalam beberapa dasawarsa terakhir ini, terutama pengenalan huruf Latin. Terbukti dengan banyaknya produk teknologi yang sudah menggunakan pengenalan huruf Latin seperti *handphone*, *notebook*, iPad, dsb. Sedangkan pengenalan huruf Arab belum banyak dikembangkan menjadi produk teknologi, padahal milyaran penduduk muslim dunia menggunakan bahasa Arab dalam kehidupan sehari-harinya. Penelitian ini membahas tentang pengenalan huruf Arab tulisan tangan secara *c<sub>j</sub>f-line*, dimana gambar tulisan diambil melalui *scanner*. Huruf Arab tulisan tangan yang dikenali yaitu huruf Arab tunggal (*isolatea*).

Metode ekstraksi yang digunakan yaitu *Principal Component Analysis* (PCA), sedangkan metode klasifikasinya yaitu jaringan syaraf tiruan menggunakan pelatihan *Backpropagation* dan metode *k-Nearest Neighbor* (kNN) sebagai pembanding. Data tulisan berjumlah 12000 gambar berukuran 40 × 50 piksel yang diperoleh dari 80 penulis. Huruf Arab yang berjumlah 30 huruf diklasifikasikan ke dalam dua model, yaitu 30 kelas dan 18 kelas. Hasil percobaan menunjukkan PCA berhasil mereduksi dimensi data input dari 2000 dimensi menjadi 78 dimensi. Model klasifikasi 30 kelas menunjukkan *backpropagation* mampu mengenali tulisan dengan tingkat akurasi 98,4% untuk data *training* dan 63% untuk data *testing*, sedangkan kNN mampu mengenali tulisan dengan tingkat akurasi 100% untuk data *training* dan 51% untuk data *testing*. Untuk klasifikasi 18 kelas, *backpropagation* mampu mengenali tulisan dengan tingkat akurasi 99,6% untuk data *training* dan 76,4% untuk data *testing*, sedangkan kNN mampu mengenali tulisan dengan tingkat akurasi 100% untuk data *training* dan 67% untuk data *testing*. Hasil ini menunjukkan metode *backpropagation* lebih baik dibandingkan kNN.

Kata kunci: sistem *c<sub>j</sub>f-line*, pengenalan huruf Arab, pelatihan *backpropagation*, *k-nearest neighbor* (kNN), *principal component analysis* (PCA).

## ABSTRACT

The character or text recognition is one of study research which developed in recent decades, especially recognizing of Latin characters. Evidenced by many technology products that have been using Latin characters recognition on mobile phones, notebooks, iPad, etc. Otherwise, recognizing of Arabic characters have not been developed into products of technology, whereas there are billion Muslim in the world who use Arabic in their daily life. This study discusses the Arabic handwriting recognition with off-line system, where the images were taken by means of the scanner. Handwriting of Arabic characters which are recognized is single character (isolated).

The feature extraction is using Principal Component Analysis (PCA) and the classification method is using Artificial Neural Network (ANN) with Backpropagation training, while *k*-Nearest Neighbor (kNN) as comparison. There are 12000 images of handwriting data which have measured  $40 \times 50$  pixels who are collected from 80 authors. The Arabic characters, which number in 30 letters are classified into two models, that are 30 classes and 18 classes. The experimental results showed that PCA reduce dimension of input data successfully from 2000 dimensions to 78 dimensions. Classification model of 30 classes showed that Backpropagation is able to recognize the character with an accuracy rate reached 98,4% for data training and 63% for data testing, while kNN is able to recognize the character with an accuracy rate reached 100% for data training and 51% for data testing. Meanwhile, for classification on 18 classes, backpropagation is able to recognize the character with an accuracy rate reached 99,6% for data training and 76,4% for data testing, in other side, K-NN is able to recognize the character with an accuracy rate reached 100% for data training and 67% for data testing. This results show that backpropagation better than kNN.

Keyword : off-line system, Arabic characters recognition, backpropagation training, k-nearest neighbor (kNN), principal component analysis (PCA).