
Klasifikasi Citra Digital Berbasis Ekstraksi Ciri Berdasarkan Tekstur Menggunakan GLCM Dengan Algoritma K-Nearest Neighbor

Dani Rohpandi¹, Asep Sugiharto², M Yoga Sukma Jati³

Teknik Informatika, STMIK Tasikmalaya

Jl. R.E. Martadinata No. 272 A Tasikmalaya, Telp. (0265) 310830

e-mail: danirtms@gmail.com¹, asepsugiharto@yahoo.com², yogasukmajati@gmail.com³

Abstrak

Pemrosesan citra adalah ilmu untuk memanipulasi gambar, yang melingkupi teknik-teknik untuk memperbaiki atau mengurangi kualitas gambar, menampilkan bagian tertentu dari gambar, membuat sebuah gambar yang baru dari beberapa bagian gambar yang sudah ada, dan beberapa teknik manipulasi gambar lainnya. Suatu citra yang mempunyai kontras rendah dapat dihasilkan dari sumber citra dengan proses pencahayaan atau penerangan yang rendah atau karena adanya kesalahan setting pada saat pengambilan citra berlangsung. Metode GLCM (Gray Level Co-occurrence Matrix) adalah satu cara mengekstrak fitur tekstur statistik orde kedua, matriks GLCM mampu menangkap sifat tekstur tetapi tidak secara langsung dapat digunakan sebagai alat analisi, misalnya membandingkan dua tekstur. Fitur-fitur yang dihasilkan oleh metode ekstraksi ciri kemudian akan digunakan sebagai masukan proses klasifikasi. K-NN adalah sebuah metode klasifikasi yang tangguh terhadap data training yang noisy dan efektif apabila datanya besar. Metode penelitian yang digunakan adalah metode eksperimen, yang bertujuan untuk mengidentifikasi hubungan sebab akibat dari satu atau lebih variabel terikat dengan melakukan manipulasi variabel bebas pada suatu keadaan yang terkendali. Hasil yang didapatkan, klasifikasi dengan menggunakan ekstraksi ciri GLCM 4 arah dimana menggunakan Energi dengan nilai rata-rata 55,99%, Metode K-Nearest Neighbor dapat menghasilkan akurasi yang diharapkan pada sistem klasifikasi.

Kata kunci: Klasifikasi, Citra Digital, Ekstraksi, GLCM, K-Nearest Neighbor.

Abstract

Image processing is a science for manipulating images, which encompasses techniques for improving or reducing image quality, displaying certain parts of an image, creating a new image from several parts of an existing image, and some other image manipulation techniques. An image with low contrast can be generated from the image source by low lighting or lighting process or due to errors setting when image capture takes place. The GLCM method (Gray Level Co-occurrence Matrix) is one way of extracting second-order statistical texture features, GLCM matrices capable of capturing textural properties but not directly applicable as analytical tools, eg comparing two textures. The features produced by characteristic extraction methods will then be used as classification inputs. K-NN is a powerful method of classification of noisy and effective training data when the data is large. The research method used is the experimental method, which aims to identify the causal relationship of one or more dependent variables by manipulating the independent variables in a controlled state. The results obtained, classification by using 4-way GLCM characteristic extraction where using Energy with an average value of 55.99%, K-Nearest Neighbor method can produce the expected accuracy in the classification system.

Keywords: Classification, Digital Image, Extraction, GLCM, K-Nearest Neighbor.

1. PENDAHULUAN

Pemrosesan citra adalah ilmu untuk memanipulasi gambar, yang melingkupi teknik-teknik untuk memperbaiki atau mengurangi kualitas gambar, menampilkan bagian tertentu dari gambar, membuat sebuah gambar yang baru dari beberapa bagian gambar yang sudah ada, dan beberapa teknik manipulasi gambar lainnya. Suatu citra yang mempunyai kontras rendah dapat dihasilkan dari sumber citra dengan proses pencahayaan atau penerangan yang rendah atau karena adanya kesalahan *setting* pada saat pengambilan citra berlangsung. Pengetahuan dan pemanfaatan citra digital berkembang pesat, tidak hanya di bidang kedokteran, industri, dan kesehatan, pertanian, dan lain-lain[1].

Semakin pesatnya peningkatan data gambar pada internet memberikan peluang bagi peneliti untuk menggarap bidang pencarian dan pengklasifikasian isi suatu gambar. Google Image merupakan salah satu contoh dari *tool* dalam manajemen gambar dan biasa digunakan untuk melakukan *image mining* serta *image searching* dalam internet. Begitu pentingnya pencarian informasi gambar dalam konteks web menimbulkan banyak komunitas peneliti yang menghasilkan banyak metode algoritma dengan *tool* untuk pengambilan dan pengelompokan gambar, yang mana teknik ini termasuk didalam bidang teknik pengambilan gambar berbasis konten atau *Content Based Image Retrieval (CBIR)*[2].

Metode GLCM (*Gray Level Co-occurrence Matrix*) adalah satu cara mengekstrak fitur tekstur statistik orde kedua. GLCM (yang disebut juga *Grey Tone Spatical Depedency Matrix*) adalah tabulasi mengenai frekuensi atau seberapa seringnya kombinasi nilai kecerahan piksel yang berbeda posisinya dalam suatu citra dengan tingkat keabuan (*Grey level*) pada jarak $d=1$ dan arah 0° . Matriks GLCM mampu menangkap sifat tekstur tetapi tidak secara langsung dapat digunakan sebagai alat analisi, misalnya membandingkan dua tekstur. Data ini harus disarikan lagi agar didapatkan angka-angka yang bisa digunakan untuk mengklasifikasi tekstur. Haralick pada tahun 1973 mengusulkan 14 ukuran (atau ciri/fitur), tetapi Connors dan Harlow pada tahun 1980-an mengkaji bahwa dari 14 fitur yang diusulkan Haralick tersebut, hanya 5 diantaranya yang biasa digunakan. Kelima fitur itu adalah: energi, entropi, korelasi, homogenitas, dan inersia[3].

Fitur-fitur yang dihasilkan oleh metode ekstraksi ciri kemudian akan digunakan sebagai masukan proses klasifikasi. K-NN adalah sebuah metode klasifikasi yang tangguh terhadap data *training* yang noisy dan efektif apabila datanya besar. Algoritma K-NN memiliki konsistensi yang kuat dan ketika jumlah data mendekati tak hingga, algoritma ini menjamin *error rate* yang tidak lebih dari dua kali *Bayes error rate (error rate minimum untuk distribusi data tertentu)*[4].

Oleh karena itu penulis mengambil topik **“KLASIFIKASI CITRA DIGITAL BERBASIS EKSTRAKSI CIRI BERDASARKAN TEKSTUR MENGGUNAKAN GLCM (*Grey Level Co-occurrence Matrix*) DENGAN ALGORITMA K-NEAREST NEIGHBOR”**.

Mengingat luasnya kemungkinan pembahasan yang diambil, maka penulis membatasi ruang lingkup pembahasan agar tujuan penelitian terarah dan dapat tercapai. Diantaranya :

- 1) Citra yang diolah adalah citra berwarna dengan format *.jpg.
- 2) *Tool* yang digunakan untuk mengekstraksi ciri dan mengklasifikasi adalah Matlab.
- 3) Sample citra data uji sebanyak 150 citra diantaranya, 30 columns, 30 flowers, 30 horse, 30 model, 30 sea sebagai kualifikasi.
- 4) Mode warna citra yang dipakai dalam ekstraksi ciri adalah RGB dan Greyscale.
- 5) Ekstraksi ciri menggunakan metode GLCM (*Gray Level Co-occurrence Matrix*) 16 arah dan klasifikasi menggunakan K-Nearest Neighbour dengan $k=1$.
- 6) Data latih sebanyak 5 kelas.

2. METODE PENELITIAN

Pada penelitian ini menggunakan metode penelitian Eksperimen. Metode penelitian eksperimen merupakan suatu penelitian yang dengan sengaja peneliti melakukan manipulasi terhadap satu atau lebih variabel dengan suatu cara tertentu sehingga berpengaruh pada satu atau lebih variabel lain yang diukur. Selain itu, metode penelitian eksperimental merupakan satu-satunya metode penelitian yang dapat diuji secara benar hipotesis menyangkut hubungan kausal

(sebab akibat). Dalam penelitian eksperimen dilakukan manipulasi paling sedikit satu variabel, mengontrol variabel lain yang relevan dan mengobservasi atau pengaruhnya terhadap satu atau lebih variabel terikat[4]. Penelitian ini dilakukan dengan cara mengumpulkan bahan-bahan penelitian dengan secara acak atau *random* yang berhubungan dengan data-data penelitian, dari data-data tersebut digunakan untuk mencari hasil yang diharapkan dengan metode-metode tertentu.

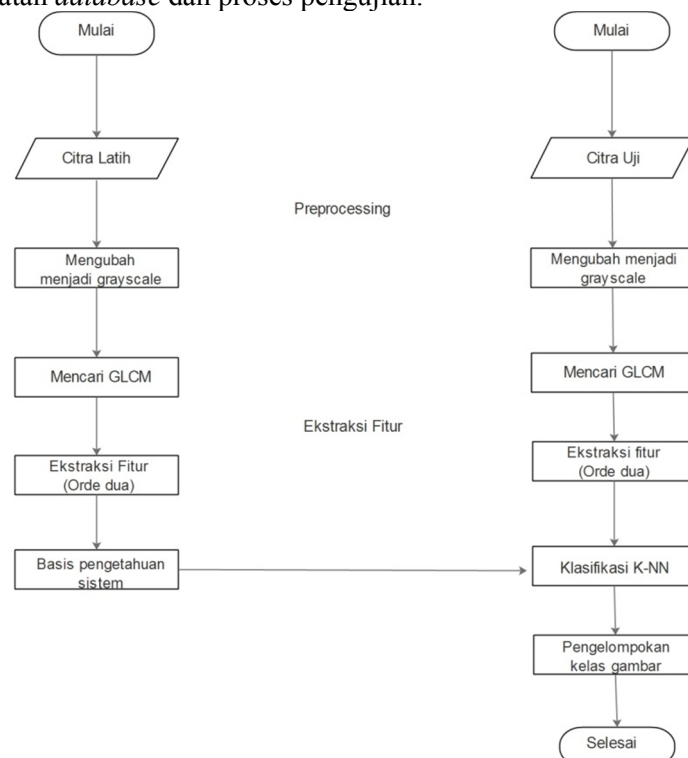
Adapun metode pengumpulan data yang digunakan yaitu melalui studi literatur. Literatur merupakan data pelengkap, dapat berbentuk dokumen tertulis atau tidak tertulis. Dengan dokumentasi peneliti dapat memperoleh informasi macam-macam sumber tertulis dari dokumen, catatan hasil observasi, wawancara dengan narasumber, dan data-data yang diperlukan yang berhubungan dengan permasalahan.

Setelah melakukan analisis masalah, maka Penulis dapat menjabarkan masalah yang terdapat pada klasifikasi citra digital yaitu sebagai berikut :

- 1) Ekstraksi ciri pada setiap citra dengan menggunakan GLCM (*Gray Level Co-occurrence Matrix*).
- 2) Mengklasifikasikan citra berdasarkan kelasnya dengan metode K-Nearest Neighbor.
- 3) Menghitung tingkat akurasi dari pengujian citra setiap kelasnya.

2.1. Model dan Perancangan

Proses klasifikasi citra digital terbagi menjadi dua skenario, yaitu pembuatan *database* dan pengujian sistem klasifikasi. Prinsipnya citra latih dan citra uji yang telah di ekstraksi dibandingkan jarak terdekatnya dengan K-Nearest Neighbor. Jarak yang terdekat menunjukkan kemiripan suatu citra. Hasil dari kemiripan tersebut berupa pengelompokan citra terhadap kelasnya pada keluaran. Berikut adalah diagram alir dari sistem klasifikasi citra digital yang terdiri dari pembuatan *database* dan proses pengujian.



Gambar 1. Diagram Alir Klasifikasi Citra Digital.

Database dari sistem adalah ciri dari masing-masing kelas citra pada citra latih yang di simpan dalam file *.mat. dan *.dat. Melalui database ciri latih tersebut, maka digunakan untuk membandingkan citra latih dengan ciri citra uji dalam pengestraksian fitur citra untuk mencari nilai rata-rata matriks dengan GLCM (*Gray Level Co-occurrence Matrix*).

Pengujian pada sistem klasifikasi adalah pengujian berdasarkan parameter uji yang digunakan. Dalam fase pengujian akan dilakukan uji pencarian nilai GLCM dengan statistik orde kedua, dan metode pengukuran jarak pada K-Nearest Neighbor untuk pengelompokan citra berdasarkan kelasnya.

Dari hasil ekstraksi pembentukan GLCM untuk mendapatkan data vektor sebagai data ujian klasifikasi dengan K-NN, maka data hasil ekstraksi dan klasifikasi dapat dilihat pada Tabel 1 dan Tabel 2.

Tabel 1. Nilai rata-rata hasil ekstraksi dari pembentukan GLCM dengan orde dua.

Kelas	Columns15
Kontras	2,19105000000000
Korelasi	0,776787300000000
Energi	0,080381200000000
Homogeniti	0,711770700000000
Entropi	7,53425400000000

Dari Tabel diatas, dapat dilihat nilai dari Energi mendekati nilai 1 yaitu jarak kemiripan dengan piksel tetangga yang bernilai derajat keabuan sejenis.

Tabel 2. Data hasil klasifikasi dengan K-NN

Columns15	1,70450000000000	1
	2,02100000000000	2
	2,35520000000000	5
	2,78150000000000	4
	3,06820000000000	3

Dari Tabel 2 dapat dilihat nilai yang sering muncul terdapat pada baris ketiga dengan kemunculan 5 dan kemunculan yang paling sedikit berada pada baris pertama dengan kemunculan 1. Maka dari hasil tersebut dapat dikelompokkan sesuai kelasnya berdasarkan hasil dari data latih.

2.2. Pengubahan Mode Warna Citra



Gambar 2 Diagram Proses Pengubahan Mode Warna Citra.

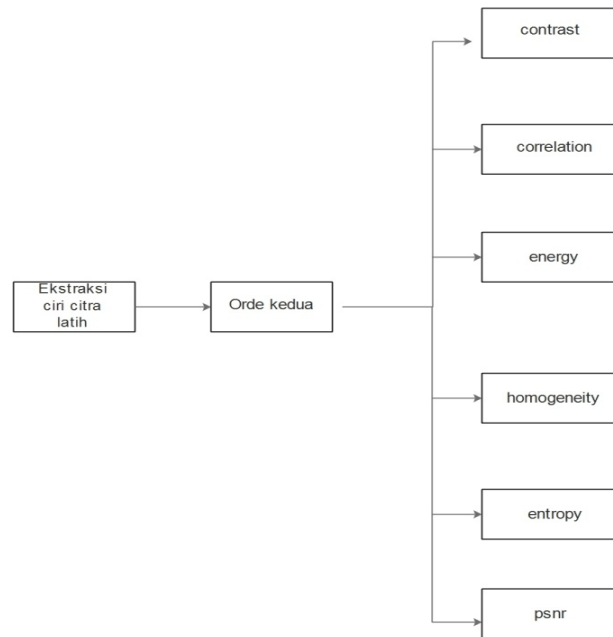
Citra uji yang masuk dalam sistem merupakan citra RGB berukuran 128 x 85 piksel dengan format JPEG. Citra tersebut diubah ke 1 mode warna, yaitu grayscale. Pengubahan warna ini dilakukan untuk mencari nilai GLCM yaitu statistik orde kedua.

2.3. Ekstraksi Ciri

Ekstraksi ciri dilakukan untuk mendapatkan ciri tekstur dari citra yang akan di uji. Citra yang akan diekstrak untuk *database* ciri latih sebanyak 30 citra 5 kelas berbeda, sedangkan citra yang diekstrak untuk pengujian sebanyak 50 citra. Setelah mengalami proses pengubahan warna, citra diekstrak dengan ekstraksi ciri statisik orde kedua dengan teknik GLCM. Ciri orde kedua terdiri dari enam parameter ciri, yaitu *contrast*, *correlation*, *energy*, *homogeneity*, *entropy*, *psnr*. Ciri latih yang disimpan akan dibandingkan dengan ciri citra uji pada tahapan klasifikasi.

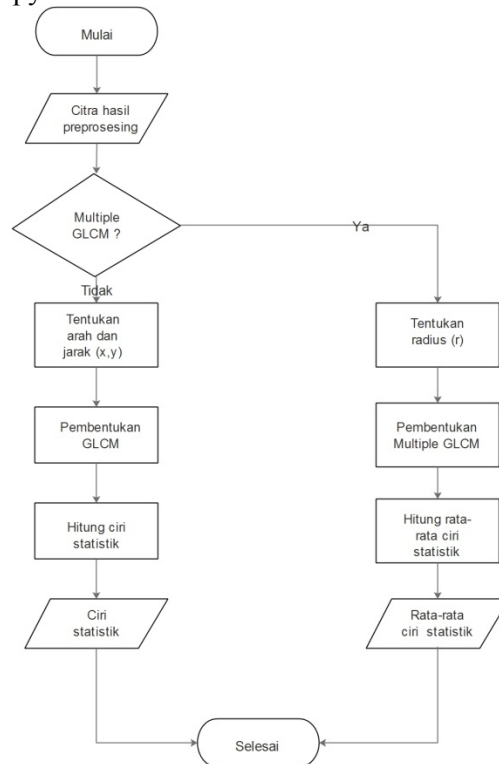
Langkah selanjutnya adalah membentuk GLCM. Tahap pembentukan GLCM untuk 4 arah (0° , 45° , 90° , 135°) $d=1$ dan $d=2$, maka akan ditentukan koordinat (x,y) . Diagram alir pembentukan GLCM ditunjukkan pada Gambar 4.4. sebagai contoh, untuk arah 0° dan jarak $d=1$ artinya koordinat (x,y) adalah $(1,0)$. Setelah menentukan arahnya, selanjutnya membentuk matriks kookurensi dengan cara menghitung frekuensi kemunculan pasangan nilai keabuan piksel referensi dan piksel tetangga pada jarak dan arah yang ditentukan.

Selanjutnya menjumlahkan semua elemen untuk menghitung probabilitas setiap elemen dengan cara membagi setiap elemen GLCM dengan total jumlah semua elemen.



Gambar 3 Blok Diagram Ekstraksi Ciri

Langkah terakhir adalah menghitung ciri statistik GLCM yaitu contrast, correlation, energy, homogeneity, entropy.



Gambar 4 Diagram Alir pembentukan GLCM

2.4. Klasifikasi dengan K-Nearest Neighbor

K-Nearest Neighbor (K-NN) adalah metode pengukuran kemiripan yang sederhana. Penelitian karya ilmiah ini menggunakan satu cara dalam pengukuran kemiripannya, yaitu Euclidean. Euclidean tergolong dalam metode pengukuran kemiripan berdasarkan geometrik.

Analisis yang dilakukan K-NN adalah pengaruh penggunaan pengukuran kemiripan dan nilai k yang digunakan terhadap akurasi sistem dalam mengklasifikasi citra digital. Nilai k yang

di uji adalah 1. Dipilihnya nilai k yang ganjil agar mengurangi kesalahan algoritma jika peluang kemiripannya sama.

2.5. Analisis Performansi Sistem

Untuk parameter akurasi sistem, ditentukan dengan perbandingan banyaknya pengujian tepat dengan banyaknya seluruh pengujian, dengan rumus sebagai berikut :

$$A = \frac{\sum p}{\sum p} \frac{-t_i}{-t_i} \times 100\%$$

3. HASIL DAN PEMBAHASAN

3.1. Hasil Pengujian

Berdasarkan hasil pengujian yang dilakukan pada tool matlab dengan mengambil sampel 150 image best retrieval pada kualifikasi columns sebanyak 30 image, flower sebanyak 30 image, horse sebanyak 30 image, model sebanyak 30 image dengan menggunakan ekstraksi ciri berdasarkan tekstur menggunakan GLCM dan algoritma K-Nearest Neighbor dapat diinterpretasikan sebagai berikut :

3.1.1. Hasil Pengujian Ekstraksi Ciri

Pengujian ini menggunakan parameter sudut GLCM ($0^\circ+45^\circ+90^\circ+135^\circ$) dan jarak (d) sebesar 2 piksel digunakan secara bersamaan untuk mencari nilai rata-rata dari seluruh sudut. Hasil pengujian ditunjukkan pada Tabel 3.

Tabel 3. Hasil Ekstraksi Ciri pada berkas citra Columns04.jpg, flower01.jpg, horse01.jpg, model01.jpg, sea02.jpg.

Parameter	Columns04	flower01	horse01	model01	sea02
kontras	1,536126	0,3159	1,706871	1,024615	0,359466
korelasi	0,848219	0,890854	0,763711	0,871445	0,871302
energy	0,748537	0,142233	0,087065	0,166944	0,226058
homogenitas	0,799931	0,868092	0,710498	0,799931	0,869883
entropi	7,669435	7,142536	7,301199	6,978275	6,575693

Dari pengujian Tabel 3, dapat dilihat bahwa setiap parameter pada masing-masing citra menunjukkan nilai-nilai yang berbeda dengan jarak yang relatif jauh. Oleh sebab itu, pengenalan dengan metode ini bisa dilakukan.

3.1.2. Nilai Pelatihan

Pada saat melakukan pelatihan nilai-nilai petatihan tersebut disimpan secara otomatis di tempat penyimpanan sementara sesuai dengan citra yang dilatih. Nilai-nilai tersebut merupakan nilai statistik orde yang menampilkan *Contrast*, *Correlation*, *Energy*, *Homogeneity*, *Entropy*. Nilai-nilai tersebut dapat dilihat pada tabel 4.

Tabel 4. Nilai Pelatihan

Sampel	Nilai Statistik Ciri Orde Dua				
	Contrast	Correlation	Energy	Homogeneity	Entropy
columns51	9.030510e-01	7.764814e-01	6.547791e-02	7.268423e-01	7.446351e+00
flower56	7.804798e-01	8.673591e-01	9.191023e-02	8.121131e-01	7.458586e+00
horse20	5.143758e-01	8.321978e-01	1.465210e-01	8.247788e-01	7.049677e+00
model43	5.542993e-01	8.959432e-01	1.093802e-01	8.445628e-01	7.278015e+00
sea48	7.438701e-01	9.130953e-01	1.123766e-01	8.051691e-01	7.056606e+00

3.1.3. Nilai Rata-rata Ekstraksi Fitur GLCM

Dari hasil ekstraksi dengan menggunakan GLCM, maka dapat di ambil nilai dari rata-rata sudut yang diuji, data yang diambil untuk menentukan tingkat akurasi berdasarkan *Energy* seperti pada Tabel 5.

Tabel 5. Hasil Akurasi Statistik Nilai Energy

Kelas	Jumlah Data	Total Benar	Akurasi (%)
Columns	30	16	53,33
Flower	30	17	56,66
Horse	30	13	43,33
Model	30	20	66,66
Sea	30	18	60

Dari hasil ekstraksi ciri menggunakan GLCM dengan nilai rata-rata pada semua sudut dengan statistik dari nilai *Energy*, maka didapat nilai tertinggi sebesar 66,66% dari kelas Model dan nilai terkecil didapat dari kelas *Horse* sebesar 43,33%.

3.1.4. Prediksi dengan K-Nearest Neighbor

Tabel 6. Prediksi Akurasi Sistem dengan K-Nearest Neighbor.

Kelas	k	Jumlah Data	Total Benar	Akurasi (%)
Columns	1	30	21	70
Flower	1	30	5	16,66
Horse	1	30	3	10
Model	1	30	13	43,33
Sea	1	30	2	6,66

Dapat dilihat dari Tabel 6 hasil klasifikasi dengan *K-Nearest Neighbor* menggunakan nilai dari rata-rata ekstraksi ciri GLCM. Hasil akurasi tertinggi didapatkan dengan menggunakan metode *Euclidean Distance* dengan nilai 70% pada kelas *Columns*. Nilai $k=1$ sebagai vektor berdekatan yang digunakan sebagai sebagai pembanding, dengan nilai tersebut maka sudah di wakili vektor ciri dari berbagai kelas. Sedangkan akurasi terendah didapat pada kelas *Sea* dengan nilai 6,66%. Jadi klasifikasi citra dengan menggunakan ciri energi, korelasi, kontras, homogenitas, entropi dapat dilakukan dengan metode pengukuran kemiripan berdasarkan geometrik yang terbukti dengan hasil akurasi yang lebih tinggi.

4. KESIMPULAN

Berdasarkan hasil percobaan dan analisis Penulis dapat mengambil kesimpulan sebagai berikut :

- 1) Metode ekstraksi ciri statistik orde dua menggunakan GLCM dan K-Nearest Neighbor dapat digunakan untuk mengklasifikasikan citra digital berdasarkan kelasnya masing-masing.
- 2) Kalasifikasi dengan menggunakan ekstraksi ciri GLCM 4 arah dimana menggunakan Energi dengan nilai rata-rata 55,99%
- 3) Metode *K-Nearest Neighbor* dapat menghasilkan akurasi yang diharapkan pada sistem klasifikasi.

5. SARAN

Penelitian ini masih dapat dikembangkan atau dapat dicoba dengan data yang lebih besar serta dapat menggunakan metode lain untuk mandapat hasil prosentase yang lebih baik.

UCAPAN TERIMA KASIH

Terima kasih atas dukungan dari keluarga, serta bapak ibu dosen pembimbing yang telah banyak memberikan dukungan serta arahan selama penulisan jurnal.

DAFTAR PUSTAKA

- [1] Debby Permatasari, "Sistem Klasifikasi Kualitas Biji Jagung Berdasarkan Tekstur Berbasis Pengolahan Citra Digital", Tugas Akhir, Institute Teknologi Telkom, Bandung, 2012.
 - [2] Dani Rohpandi, "Image Clustering Menggunakan Algoritma K-Means Dengan Memakai Metode Non-Negative Matrix Factorization Dalam Histogram Warna Pada Content Base Image Retrieval", Thesis, Universitas Dian Nuswantoro, Semarang, 2014.
 - [3] Ingrid Nurtanio, I Ketut Eddy Purnama, Mauridhi Hery Purnomo Cucun Very Angkoso, "Analisa Tekstur Untuk Membedakan Kista Dan Tumor Pada Citra Panoramik Rahang Gigi Manusia", *The 12th Seminar Intelligent Technology And Its Applications*, Institute Teknologi Sepuluh November, Surabaya, 2011.
 - [4] Ahmad Dahlan. (2016) EUREKA PENDIDIKAN. [Online].
<http://www.eurekapedidikan.com/2015/11/metode-penelitian-eksperimen.html>
-