

KLASIFIKASI DAUN HERBAL MENGGUNAKAN METODE *NAÏVE BAYES CLASSIFIER* DAN *K-NEAREST NEIGHBOR*

Febri Liantoni¹⁾, Hendro Nugroho²⁾

^{1,2}Prodi Teknik Informatika, Fakultas Teknologi Informasi,
Institut Teknologi Adhi Tama Surabaya
Jl. Arief Rachman Hakim no. 100, Surabaya
Email: febri.liantoni@gmail.com, dosh3ndro@gmail.com

ABSTRAK

Perkembangan ilmu tanaman telah mengalami kemajuan yang pesat, khususnya ilmu mengenai tanaman herbal. Tanaman herbal memiliki banyak manfaat bagi kehidupan manusia yaitu sebagai penyediaan oksigen, bahan makanan, obat-obatan, maupun bahan kosmetika. Untuk mengetahui jenis-jenis tanaman herbal dapat dilakukan dengan proses klasifikasi. Klasifikasi tanaman herbal dapat dilakukan dengan cara mengidentifikasi bentuk citra daun dari tanaman tersebut. Proses klasifikasi tanaman herbal memerlukan ekstraksi fitur dari bentuk dari tanaman tersebut. Pada penelitian ini, fitur *invariant moment* dan fitur geometri digunakan untuk ekstraksi fitur daun herbal. Fitur yang digunakan berjumlah 10 fitur. Ada beberapa macam metode klasifikasi yang biasa digunakan. Pada penelitian ini metode klasifikasi yang digunakan adalah metode *Naïve Bayes Classifier* dan *K-Nearest Neighbor* (KNN). Metode *Naïve Bayes Classifier* merupakan metode *Bayesian Learning* yang paling cepat dan sederhana. Sedangkan metode KNN dapat melakukan klasifikasi dengan cepat berdasarkan jarak terdekat diantara objek data. Berdasarkan hasil uji coba yang dilakukan, penggunaan metode *Naïve Bayes Classifier* didapatkan nilai akurasi sebesar 75%, sedangkan dengan menggunakan metode *K-Nearest Neighbor* didapatkan nilai akurasi sebesar 70,83%. Hal ini menunjukkan bahwa kinerja metode *Naïve Bayes Classifier* lebih baik dibandingkan metode KNN.

Kata Kunci: *Fitur Invariant Moment, Fitur Geometri, Naïve Bayes Classifier, K-Nearest Neighbor, Bayesian Learning.*

ABSTRACT

Science of the plant has made progress, particularly knowledge about herbs. Herb has many benefits for human life as provision of oxygen, foodstuffs, pharmaceuticals, and cosmetics. To determine the types of herbs with the classification process. Classification of herbs conducted by identifying the shape of the image of the leaves of these plants. Herbal plant classification process requires the extraction of features from the shape of plant. In this study, moment invariant features and geometrical feature is used for feature extraction of herbal leaves. Features used amounted to 10 features. There are several kinds of commonly used classification method. In this study, the classification method used is the method Naïve Bayes classifier and K-Nearest Neighbor (KNN). Naïve Bayes classifier is Bayesian Learning method of the most rapid and simple. While the KNN method can perform fast classification is based on the shortest distance between data objects. Based on the results of tests conducted, the use of methods Naïve Bayes Classifier accuracy values obtained by 75%, while using K-Nearest Neighbor value obtained an accuracy of 70.83%. These results indicate that the performance of Naïve Bayes Classifier method is better than KNN method.

Keywords: *Invariant Moment Feature, Geometrical Feature, Naïve Bayes Classifier, K-Nearest Neighbor, Bayesian Learning.*

PENDAHULUAN

Tanaman herbal memiliki banyak manfaat bagi kehidupan manusia sebagai penyedia oksigen, bahan makanan, obat-obatan, maupun bahan kosmetika. Proses pengenalan tanaman herbal dapat dikenali dengan cara pengambilan gambar daun dari tanaman herbal kemudian dilakukan pengenalan pola daun dengan cara mengenali karakteristik struktural daun seperti bentuk dan tekstur daun tersebut [1][2][3]. Klasifikasi tanaman herbal dapat dilakukan dengan mengidentifikasi jenis dari tanaman herbal berdasarkan bentuk daun tersebut.

Proses klasifikasi tanaman herbal dilakukan berdasarkan struktural daun yang kemudian dikenali dengan metode pengolahan citra digital. Proses ini dilakukan dengan mengidentifikasi citra masukan untuk menganalisa karakteristik dari struktural daun tersebut. Berbagai teknik pengolahan citra digital telah dikembangkan untuk mempermudah pekerjaan manusia, baik sebagai pengolah citra, analisis citra maupun penggunaan citra untuk berbagai tujuan dan keperluan lainnya. Untuk mendapatkan bentuk tepian dan ciri struktural dari masing-masing daun digunakan ekstraksi fitur dengan pengenalan fitur morfologi digital [1]. Pada penelitian ini metode ekstraksi fitur yang digunakan adalah fitur geometri dan *invariant moment*.

Pada penelitian [4] dilakukan identifikasi spesies tanaman berdasarkan fitur bentuk daun, pembuluh daun, margin daun (bentuk gerigi pada daun terluar dan tekstur daun. Penelitian ini tidak menggunakan fitur warna. Pada penelitian [5] dilakukan ekstraksi fitur citra daun menggunakan fitur bentuk dan warna daun. Penelitian ini hanya membahas beberapa metode ekstraksi fitur untuk meningkatkan performa sistem temu kembali.

Proses seleksi fitur yang dilakukan pada penelitian [6] menggunakan *Genetic Algorithm* (GA) dan Kernel PCA dengan tujuan memilih subset terbaik. Proses penggabungan GA dan KPCA tersebut digunakan untuk

melakukan klasifikasi dengan menggunakan *Support Vector Machine* (SVM). Hasil penelitian tersebut dengan menggunakan seleksi fitur lebih efektif secara komputasi dan dapat meningkatkan hasil akurasi.

Beberapa penelitian yang telah dilakukan tentang pengelompokan bentuk tanaman, umumnya menggunakan metode jaringan saraf tiruan. [1][2][3][7][8]. Metode jaringan saraf tiruan banyak digunakan karena metode ini dikenal lebih cepat secara substansial. Akan tetapi penentuan jumlah hidden layer yang digunakan akan berpengaruh besar pada uji coba yang dihasilkan, selain itu jumlah parameter epoch yang besar mengakibatkan tingkat komputasi yang lebih tinggi.

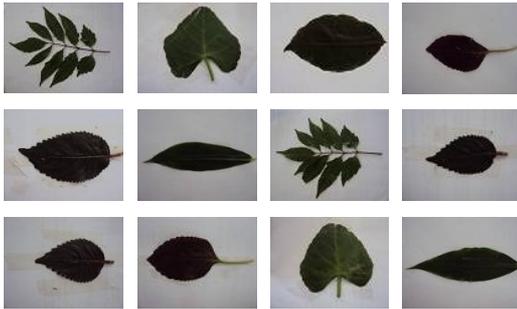
Pada penelitian ini akan dilakukan klasifikasi tanaman herbal berdasarkan fitur bentuk tepi daun. Metode klasifikasi yang digunakan adalah *Naïve Bayes Classifier* dan *K-Nearest Neighbor*. Sebelum dilakukan tahapan pengelompokan, maka terlebih dahulu dilakukan tahapan praproses citra dan ekstraksi fitur citra tepi daun agar didapatkan nilai masukan yang tepat untuk tahapan klasifikasi spesies daun berdasarkan citra daun.

METODE

Morfologi Tanaman

Morfologi tanaman merupakan ilmu yang mempelajari bentuk dan susunan tubuh tanaman, dipisahkan menjadi morfologi luar atau morfologi saja dan morfologi dalam atau anatomi tanaman. Salah satunya diambil morfologi tanaman berdasarkan bentuk tepi daun. Secara konsep macam-macam pembagian bentuk daun [9] yaitu Integer (pinggir rata), Divisus atau bertoreh (pinggir tidak rata), Tepi daun dengan toreh yang merdeka (tidak mempengaruhi bentuk umum), Serratus (pinggir bergerigi atau bergergaji), Dentatus (pinggir bergigi), Creatus (pinggir beringgit), Tepi daun dengan toreh tidak merdeka (mempengaruhi bentuk umum),

Lobatus (pinggir berlekuk menyirip dan menjari), Fissus / vidus (pinggir bercanggap menyirip dan menjari), Paritus / diversifolia (pinggir berbagi menyirip dan menjari). Pada penelitian ini data yang digunakan berupa gambar daun dengan *background* berwarna putih. Contoh gambar yang digunakan seperti ditunjukkan pada Gambar 1.



Gambar 1. Contoh data daun

Ekstraksi Fitur

Ekstraksi fitur merupakan suatu proses untuk menghasilkan informasi morfologi dari nilai-nilai fitur berupa vektor fitur dari citra biner daun yang telah dilakukan deteksi tepi. Informasi morfologi digital ini merupakan informasi morfologi fisik atau bentuk dari citra daun. Fitur morfologi digital atau *Digital Morphological Feature* (DMF) ini terdiri dari dua fitur utama, yaitu fitur bentuk geometri daun atau *Geometrical Feature* (GF) dan fitur momen invarian gambar atau *Invariant Moment Feature* (MF). Pada penelitian ini ekstraksi fitur yang digunakan adalah fitur geometri dan fitur *invariant moment*.

Fitur Geometri

Fitur Geometri atau *Geometrical Feature* merupakan fitur yang berasal dari bentuk geometris dari citra daun. Beberapa fitur geometri yang menjadi dasar penghitungan fitur geometri dari citra daun antara lain diameter, *physiological length*, *physiological width*, luas daun, keliling daun [8]. Dari dari fitur geometri dasar sebuah daun tersebut dapat dihitung fitur geometri utamanya. Fitur geometri dari suatu daun yang akan

digunakan dalam penelitian ini antara lain *perimeter ratio of convexity*, *circularity* dan *eccentricity*.

Fitur Invariant Moment

Proses pengenalan sebuah obyek di dalam sebuah citra setelah proses segmentasi, sering terbentur pada permasalahan posisi obyek, rotasi sumbu obyek, dan perubahan skala dari obyek. Posisi obyek yang bergeser atau berputar maupun ukurannya yang lebih kecil atau lebih besar daripada dapat menyebabkan kesalahan dalam pengenalan atau identifikasi obyek.

Secara tradisional, *invariant moment* dihitung berdasarkan informasi yang diberikan oleh *boundary* bentuk dan daerah interiornya. Moment digunakan untuk membentuk *invariant moment* yang didefinisikan secara kontinu namun untuk implementasi praktis, moment dihitung secara diskrit.

Pada penggunaan perhitungan nilai 2 (dua) dimensi momen sample gambar $M \times M$ dari fungsi kontinu $f(x,y), (x,y=0, \dots, M-1)$ didapatkan Persamaan 1.

$$m_{pq} = \sum_{x=0}^{M-1} \sum_{y=0}^{M-1} (x)^p \cdot (y)^q \cdot f(x,y) \quad (1)$$

Momen dapat menggambarkan suatu obyek dalam hal area, posisi, orientasi dan parameter terdefinisi lainnya. Dengan mendapatkan sejumlah informasi momen, baik momen tingkat ke nol (m_{00}) dan kesatu (m_{10} dan m_{01}) atau momen sentral, dan momen pada tingkat ≥ 2 atau moment invariant dari sebuah obyek, maka obyek tersebut dapat diidentifikasi sekalipun telah mengami pergeseran (translasi), perputaran (rotasi) maupun perubahan skala [10].

Dari moment $f(x,y)$ akan ditranslasikan dengan nilai (a,b) sehingga didapatkan perhitungan baru seperti Persamaan 2.

$$m_{pq} = \sum_x \sum_y (x + a)^p \cdot (y + b)^q \cdot f(x,y) \quad (2)$$

Dari central moment utama yaitu m_{pq} atau μ_{pq} dikomputasi melalui proses substitusi terhadap nilai $a = -x$ dan nilai $b = -y$ maka akan didapatkan perhitungan pada Persamaan 3.

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p \cdot (y - \bar{y})^q f(x, y) \quad (3)$$

Untuk nilai $\bar{x} = \frac{m_{10}}{m_{00}}$ dan $\bar{y} = \frac{m_{01}}{m_{00}}$

Ketika proses normalisasi maka nilai penskalaan yang digunakan dalam perhitungan pada Persamaan 4.

$$\eta_{pq} = \mu_{pq} / \mu_{pq}^{\gamma} \quad (4)$$

Untuk nilai $\gamma = [(p + q) / 2] + 1$

Kemudian dari proses tersebut maka didapatkan nilai seven moment invariant dengan perhitungan pada Persamaan 5.

$$\begin{aligned} Hu_1 &= \eta_{20} + \eta_{02} \\ Hu_2 &= (\eta_{20} + \eta_{02})^2 + 4\eta_{11}^2 \\ Hu_3 &= (\eta_{20} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\ Hu_4 &= (\eta_{20} + \eta_{12})^2 + (3\eta_{21} + \eta_{03})^2 \\ Hu_5 &= (\eta_{20} + 3\eta_{12})(\eta_{20} + \eta_{12})[(\eta_{20} + \eta_{12})^2 - 3(\eta_{12} + \eta_{03})^2] + (3\eta_{21} + \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{20} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\ Hu_6 &= (\eta_{20} - \eta_{12})[(\eta_{21} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{21} + \eta_{12})(\eta_{21} + \eta_{03}) \\ Hu_7 &= (3\eta_{21} - \eta_{20})(\eta_{20} + \eta_{12})[(\eta_{20} + \eta_{12})^2 - 3(\eta_{12} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{20} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \end{aligned} \quad (5)$$

Naive Bayesian Classifier

Naive Bayes Classifier adalah metode Bayesian Learning yang paling cepat dan sederhana. Hal ini berasal dari teorema Bayes dan hipotesis kebebasan, menghasilkan klasifier statistik berdasarkan peluang. Metode Bayes merupakan pendekatan statistic untuk melakukan inferensi induksi pada persoalan klasifikasi.

Probabilitas X di dalam Y adalah probabilitas inteseksi X dan Y dari

probabilitas Y , atau dengan bahasa lain $P(X/Y)$ adalah prosentase banyaknya X di dalam Y . Penghitungan pada Naive Bayes dapat dirumuskan seperti pada Persamaan 6.

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(X)} \quad (6)$$

K-Nearest Neighbor Classifier

K-Nearest Neighbor Classifier (k-NN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Data pembelajaran diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data.

Ruang dimensi dibagi menjadi bagian-bagian berdasarkan klasifikasi data pembelajaran. Nilai k yang terbaik untuk algoritma ini tergantung pada data; secara umum nilai k yang tinggi akan mengurangi efek noise pada klasifikasi, akan tetapi membuat batasan antara setiap klasifikasi menjadi lebih kabur. Nilai k yang bagus dapat dipilih dengan optimasi parameter, misalnya dengan menggunakan *cross-validation*. Kasus khusus di mana klasifikasi diprediksikan berdasarkan data pembelajaran yang paling dekat (dengan kata lain, $k = 1$) yang biasanya disebut algoritma *nearest neighbor*.

Perancangan Sistem

Pada penelitian ini, sistem klasifikasi citra daun herbal menggunakan metode Naive Bayes Classifier dan K-Nearest Neighbor terdiri dari 3 proses yaitu praproses, ekstraksi fitur daun dan tahap klasifikasi citra daun herbal.

Praproses

Pada tahap praproses dilakukan beberapa langkah pengolahan citra daun. Data masukan berupa citra daun berwarna atau RGB yang kemudian diubah menjadi citra *grayscale*. Dari citra *grayscale* citra dikonversi menjadi citra biner dengan operasi *thresholding*.

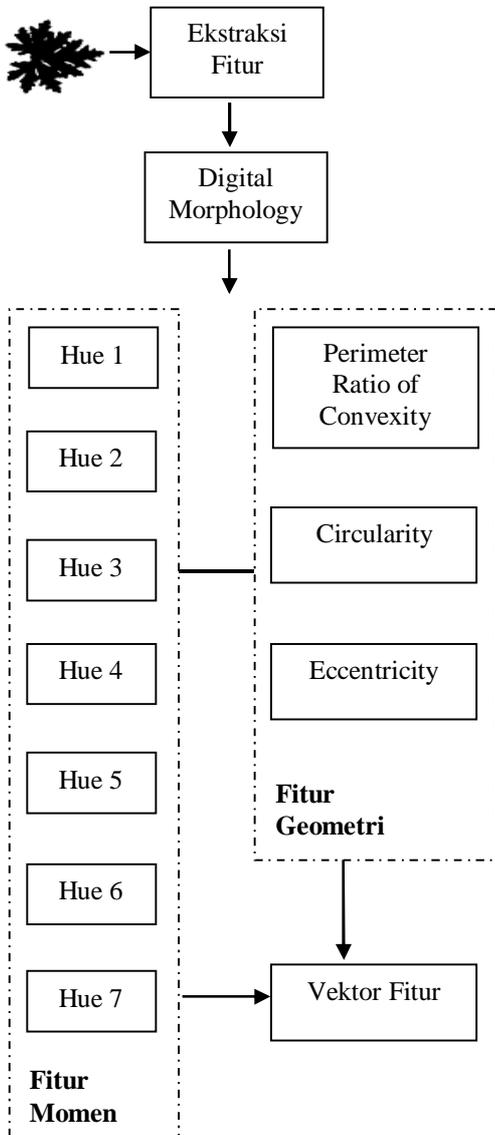
Untuk konversi citra grayscale digunakan yang ditunjukkan pada Persamaan 7.

$$gray = \frac{red \times 299 + green \times 587 + blue \times 114}{1000} \quad (7)$$

Ekstraksi Fitur

Pada penelitian ini ekstraksi fitur yang digunakan adalah fitur geometri dan fitur *invariant moment*. Fitur yang digunakan berjumlah 10 fitur yang terdiri dari 3 fitur geometri yaitu *perimeter ratio of convexity*, *circularity* dan *eccentricity*, sedangkan pada *invariant moment* berjumlah 7 fitur.

Tahapan yang dilakukan pada proses ekstraksi fitur ditunjukkan pada Gambar 2.

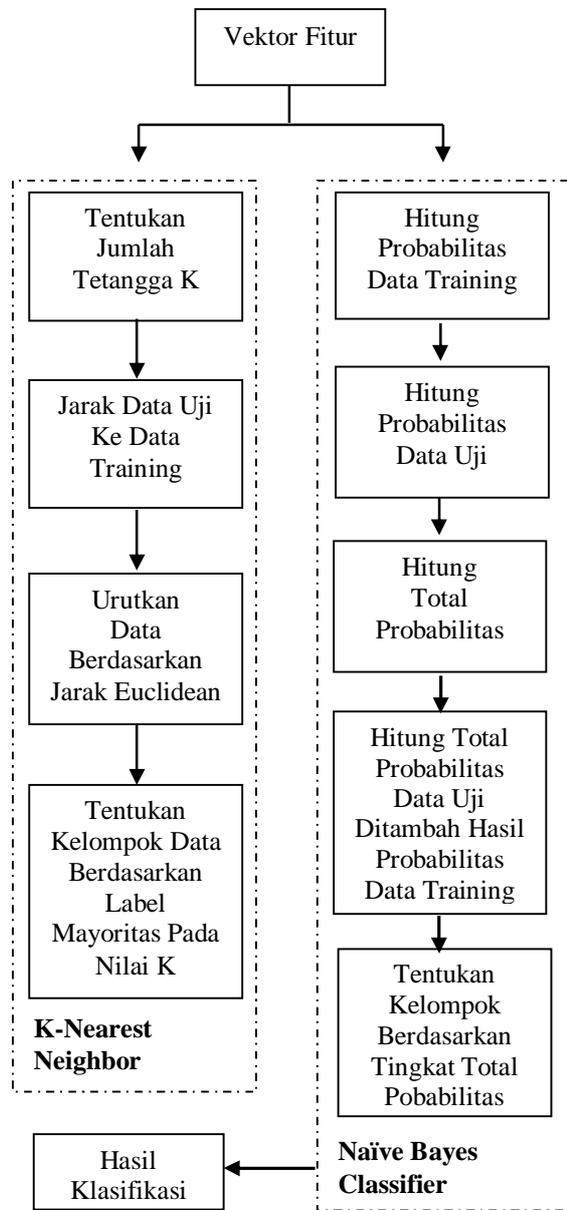


Gambar 2. Proses ekstraksi fitur citra

Klasifikasi

Tahapan terakhir penelitian ini adalah proses klasifikasi pola citra daun herbal. Metode *Naïve Bayes Classifier* dan *K-Nearest Neighbor* digunakan untuk proses klasifikasi. Hasil dari kedua metode akan dibandingkan untuk mengetahui tingkat akurasi terbaik dari metode yang digunakan.

Tahapan yang dilakukan pada proses klasifikasi daun ditunjukkan pada Gambar 3.



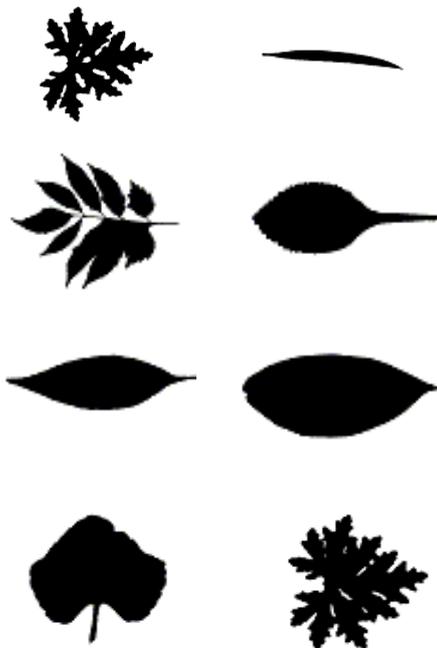
Gambar 3. Proses klasifikasi citra daun

HASIL DAN PEMBAHASAN

Pengujian dilakukan dengan tujuan untuk mengetahui keberhasilan sistem dalam klasifikasi daun herbal. Skenario uji coba yang dilakukan dengan jumlah data gambar daun sebanyak 120 gambar. Pengujian dilakukan dengan cara memisahkan data menjadi dua bagian yaitu 80% atau 96 gambar digunakan sebagai data latih dan 20% atau 24 gambar digunakan sebagai data uji. Data latih digunakan sebagai data rujukan klasifikasi yang sesuai, sedangkan data uji coba digunakan untuk menguji ketepatan sistem dalam melakukan klasifikasi daun. Hasil klasifikasi tersebut akan dicatat dan dibandingkan dengan klasifikasi yang sebenarnya. Dari hasil klasifikasi kemudian dihitung akurasi.

Hasil Praproses

Pada tahap praproses sistem akan mengubah gambar dalam ruang warna RGB ke dalam grayscale, kemudian dilakukan thresholding untuk mendapatkan citra biner. Contoh beberapa gambar hasil dari praproses ditunjukkan pada Gambar 4.



Gambar 4. Hasil praproses

Hasil yang diperoleh dari praproses seperti yang ditunjukkan pada Gambar 4 kemudian akan dilakukan ekstraksi fitur untuk mendapatkan nilai fitur dari tiap-tiap gambar daun herbal.

Hasil Ekstraksi Fitur

Pada tahap ekstraksi fitur akan dihasilkan vektor fitur yang berisi 10 nilai dari fitur geometri dan *invariant moment*. Contoh nilai pada ekstraksi fitur dari *Invariant Moment* yang dihasilkan oleh sistem ditunjukkan pada Tabel 1.

Tabel 1. Contoh hasil data ekstraksi fitur *Invariant Moment*

	<i>Invariant Moment</i>						
	<i>hue 1</i>	<i>hue 2</i>	<i>hue 3</i>	<i>hue 4</i>	<i>hue 5</i>	<i>hue 6</i>	<i>hue 7</i>
da	220,	3,71	0,00	0,06	1,46	0,00	-
un	368	688	840	515	E-06	386	4,5E
1	1	4	9	3			-07
da	246,	5,19	0,01	0,03	3,97	0,00	6,89
un	823	986	583	419	E-07	1236	E-07
2	1	6	5	1			
da	230,	4,45	0,01	0,06	1,09	0,00	-
un	367	463	263	351	E-06	2117	1,4E
3	8	1	1	6			-06
da	284,	7,21	0,02	0,19	1,31	0,01	-
un	242	99	789	455	E-05	1321	5,9E
4	9		4	3			-06
da	260,	6,42	0,02	0,05	-	-	-2E-
un	884	989	935	173	6,7E	0,00	06
5	9	3	3	3	-08	2	
						4	

Contoh nilai pada ekstraksi fitur geometri yang dihasilkan oleh sistem ditunjukkan pada Tabel 2.

Tabel 2. Contoh hasil data ekstraksi fitur geometri

	Fitur Geometri		
	<i>convexity</i>	<i>circularity</i>	<i>eccentricity</i>
daun1	0,032212	0,658284	153,014
daun2	0,039413	0,672468	143,5035
daun3	0,034993	0,670392	149,2146
daun4	0,043663	0,678573	131,1427
daun5	0,042272	0,686155	138,6381

Hasil Klasifikasi

Pengujian klasifikasi dilakukan terhadap 24 citra daun herbal. Data uji akan diklasifikasikan terhadap data training. Proses klasifikasi yang dilakukan menggunakan metode *Naïve Bayes Classifier* dan *K-Nearest Neighbor* (KNN).

Berdasarkan hasil pengujian menggunakan metode *Naïve Bayes Classifier* didapatkan 18 data yang terklasifikasi dan terdapat 6 data kesalahan. Sedangkan pengujian menggunakan metode *K-Nearest Neighbor* didapatkan 17 data yang terklasifikasi dan terdapat 7 data kesalahan.

Dari hasil pengujian tersebut didapatkan akurasi untuk metode *Naïve Bayes Classifier* sebesar $18/24 = 75\%$. Sedangkan akurasi untuk *K-Nearest Neighbor* sebesar $17/24 = 70,83\%$.

SIMPULAN

Dari hasil uji coba klasifikasi citra daun herbal didapatkan kesimpulan yaitu pada penggunaan metode *Naïve Bayes Classifier* didapatkan nilai akurasi klasifikasi sebesar 75%, sedangkan dengan menggunakan metode *K-Nearest Neighbor* didapatkan nilai akurasi klasifikasi sebesar 70,83%. Hal ini menunjukkan bahwa kinerja metode *Naïve Bayes Classifier* lebih baik dibandingkan metode *K-Nearest Neighbor*.

SARAN

Pada penelitian ini hanya membandingkan dua metode klasifikasi, untuk selanjutnya bisa dibandingkan menggunakan metode lain seperti jaringan saraf tiruan.

DAFTAR PUSTAKA

- [1] Z. Husin, A.Y.M. Shakaff, A. H. A. Aziz, R.S.M. Farook, M.N. Jaafar, U. Hashim, and A. Harun, "Embedded Portable Device For Herb Leaves Recognition Using Image Processing Techniques And Neural Network Algorithm", *Science Direct on Computers and Electronics in Agriculture*, pp. 18–29, 2012.
- [2] K. Abdul, E.N Lukito, and N. Adhi, "Leaf Classification Using Shape, Color, and Texture Features", *International Journal of Computer Trends and Technology*, July to Aug, 2011.
- [3] L.N. Pradany, A.Y. Wijaya, and R. Soelaiman, "Identifikasi Parameter Optimal Jaringan Syaraf Tiruan Multi Layer Perceptron pada Pengenalan Pola Daun: Studi Kasus Tanaman Herbal", *Jurnal Teknik Pomits*, vol. 2, no. 1, 2014.
- [4] Cope, and S. James, "Plant species identification using digital morphometrics: A review", *Expert Systems with Applications*, vol. 39, no.8, pp. 7562-7573, 2012.
- [5] Y.A. Sari, R.K. Dewi, and C. Fatichah, "Seleksi Fitur Menggunakan Ekstraksi Fitur Bentuk, Warna, Dan Tekstur Dalam Sistem Temu Kembali Citra Daun", *Juti*, vol. 12, no. 1, pp. 1-8, 2014.
- [6] N. Valiammal, and Geethalaksmi, "An Optimal Feature Subset Selection for Leaf Analysis", *International Journal of Computer and Communication Engineering*, vol. 6, 2012.
- [7] J . Chaki, and R. Parekh, "Plant Leaf Recognition using Shape based Features and Neural Network Classifiers", *International Journal of Advanced Computer Science and Applications*, vol. 2, no. 10, 2011.
- [8] S. G. Wu, F. S. Bao, E. Y. Xu, Y. -X. Wang, Y. -F. Chang, and Q. -L. Xiang, "A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network", *IEEE International Symposium*, pp. 1-6, July, 2007.
- [9] T. Gembong, *Morfologi Tanaman*, Gadjah Mada University, Yogyakarta, 2005.
- [10] B. Xiao, J. -T. Cui, H. -X. Qin, W. -S. Li, and G. -Y. Wang, "Moments

and Moment Invariants in the Radon Space”, *Elsevier Pattern Recognition*, vol. 8, no. 9, pp. 2772-2784, 2015.