

## Perbandingan Metode Ekstraksi Ciri Histogram dan PCA untuk Mendeteksi Stoma pada Citra Penampang Daun *Freycinetia*

### *Comparison of Histogram and PCA as Feature Extraction Methods in Detecting Stoma in Freycinetia Leaf Images*

DONY SATRIA, MUSHTHOFA \*

#### Abstrak

Ekstraksi fitur adalah proses pengambilan ciri sebuah objek yang dapat menggambarkan karakteristik dari objek tersebut. Pada penelitian ini, dua buah metode ekstraksi fitur digunakan, yaitu *Principal Component Analysis* (PCA) dan histogram untuk melakukan deteksi stoma pada gambar penampang daun *Freycinetia*. Penelitian ini menggunakan *frame* berjalan yang melakukan pengolahan bagian citra dan melakukan deteksi kemunculan stoma pada bagian citra tersebut. Untuk memodelkan kemunculan stoma, dibuat tiga kelas *frame*, yaitu *frame* dengan kemunculan stoma penuh, *frame* dengan kemunculan sebagian stoma, dan *frame* tanpa kemunculan stoma. Untuk proses klasifikasi, digunakan pemodelan menggunakan Jaringan Saraf Tiruan (JST) *Backpropagation*. Hasil percobaan menunjukkan bahwa ekstraksi fitur menggunakan PCA menghasilkan akurasi yang lebih baik dibandingkan dengan metode histogram. Nilai *F1-measure* yang terbaik yang didapatkan menggunakan ekstraksi fitur PCA ialah 0.9091.

Kata kunci: deteksi stoma, ekstraksi fitur, *Freycinetia*, histogram, PCA

#### Abstract

*Feature extraction is the process of taking an object identifier which can describe the characteristics of the object. In this study, we use two feature extraction methods, namely Histogram and PCA (Principal Component Analysis) to be used in stoma detection from the Freycinetia sectional leaves images. We used a moving frame to detect the occurrence of stoma in an image. To create a classification model, we distinguish between three frame classes: frames showing full stoma, frames showing parts of stoma, and frames which contain no part of stoma at all. For the classification method, we use the Backpropagation Artificial Neural Network for the classifier. The results of the detection process using Histogram as the feature extraction method will be compared with the results of the detection process using PCA. The best results between the two methods of feature extraction will be used as the first step in the process of species type identification for the genus Freycinetia. The research results show that the PCA feature extraction method is better than the Histogram feature extraction method in detecting the occurrence of stoma on the Freycinetia sectional leaves images. The best f1-measure value that can be achieved by the PCA feature extraction methods is 0.9091.*

*Keywords: feature extraction, Freycinetia, histogram, leaf image, PCA, stoma detection*

## PENDAHULUAN

*Freycinetia* adalah genus terbesar kedua dari famili Pandanaceae yang memiliki jenis spesies yang sangat beragam. Saat ini, diperkirakan terdapat 200-300 jenis spesies *Freycinetia* di seluruh dunia. Spesies-spesies yang termasuk ke dalam genus tersebut memiliki ciri morfologi yang hampir serupa. Karena kesamaan ciri morfologi tersebut, spesies yang satu dengan spesies yang lainnya sangat sulit dibedakan sehingga, ketika ditemukan tumbuhan yang termasuk ke dalam genus *Freycinetia*, jenis spesies dari tumbuhan tersebut sulit untuk dikenali. Pengenalan dan klasifikasi terhadap jenis spesies *Freycinetia* perlu dilakukan untuk mengetahui nilai potensial dan kegunaan, serta penyebaran koleksi plasma nutfah *Freycinetia* pada keanekaragaman sumber daya hayati (Qur'ania 2012).

Parameter paling tepat yang dapat digunakan untuk melakukan pengenalan jenis spesies *Freycinetia* ialah ciri anatomi stoma (Willmer 1983). Karakter anatomi stoma yang dapat diamati ialah bentuk, kerapatan, panjang, lebar, luas, jumlah sel penjaga, ukuran sel epidermis, dan indeks stoma (perbandingan antara jumlah stoma dan jumlah sel epidermis pada luas area tertentu).

Penelitian mengenai pengenalan jenis/varian dari spesies tumbuhan menggunakan ciri morfologi stoma berbasis citra telah dilakukan sebelumnya. Sanyal *et al.* (2008) menggunakan citra *scanning electron microscopy* (SEM) dalam menganalisis fitur morfologis dari stoma daun beberapa spesies tomat untuk mengenali varietasnya. Penelitian menggunakan ciri anatomi stoma pada jenis spesies *Freycinetia* telah dilakukan oleh Pasaribu (2010). Pada penelitian tersebut, rata-rata ukuran stoma, frekuensi stoma, dan indeks stoma masing-masing spesies dihitung secara manual. Perhitungan tersebut memberikan hasil yang berbeda di antara 14 spesies yang digunakan.

Qur'ania (2012) melakukan proses klasifikasi terhadap empat spesies *Freycinetia* menggunakan citra digital anatomi stoma. Proses klasifikasi tersebut menggunakan analisis ekstraksi ciri terhadap komponen nilai RGB dan nilai *grayscale*, serta ekstraksi ciri terhadap nilai dekomposisi Wavelet. Akurasi klasifikasi terbaiknya ialah sekitar 90%. Penelitian tersebut merupakan suatu cara untuk melakukan pengenalan spesies *Freycinetia* secara otomatis, namun data yang digunakan pada penelitian tersebut adalah citra anatomi stoma secara keseluruhan (citra penampang paradermal daun) yang di dalamnya terdapat sel-sel epidermis dan stoma sehingga terdapat kemungkinan bahwa apabila data yang digunakan adalah stoma yang dipisahkan dari sel-sel epidermis, akurasi klasifikasi yang dihasilkan dapat ditingkatkan.

Pemisahan antara stoma dan sel-sel epidermis yang mengelilinginya diharapkan mampu mendeteksi letak atau posisi kemunculan stoma pada citra penampang daun *Freycinetia* sehingga, pada penelitian selanjutnya, dari stoma yang terdeteksi tersebut berbagai analisis seperti penghitungan rata-rata ukuran stoma, frekuensi stoma, kerapatan stoma, jarak antarstoma, dan indeks stoma dapat dihitung untuk menemukan penciri atau karakteristik dari setiap spesies *Freycinetia*.

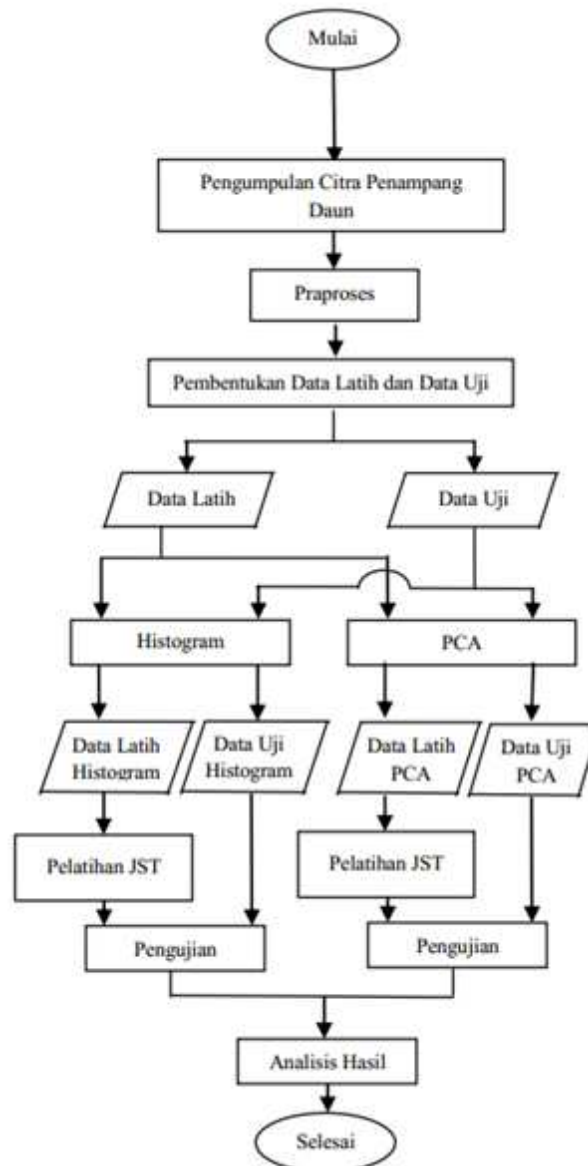
Tujuan utama penelitian ini ialah membandingkan metode ekstraksi ciri PCA dan histogram dalam melakukan pendeteksian stoma pada sebuah gambar penampang daun *Freycinetia*. Metode ekstraksi fitur yang terbaik dapat digunakan pada penelitian selanjutnya untuk pengenalan daun *Freycinetia*.

## **METODE**

Garis besar metode penelitian ini terdiri atas beberapa langkah alur kerja yang dijelaskan pada Gambar 1.

### **Pengumpulan Data**

Data citra penampang daun *Freycinetia* yang digunakan dalam penelitian ini diperoleh dari Pasaribu (2010). Data tersebut terdiri atas tiga jenis spesies, yaitu: *Freycinetia angustifolia*, *F. imbricata*, dan *F. javanica*. Setiap spesies terdiri atas 24 citra penampang daun. Setiap citra penampang daun memiliki dimensi 480 x 640 piksel dengan model warna RGB dan memiliki format penyimpanan JPEG.



Gambar 1 Alur kerja penelitian

### Praproses Data

Tahapan ini terdiri atas dua bagian, yaitu:

#### 1 Konversi RGB menjadi Grayscale

Setelah seluruh data citra penampang daun terkumpul, citra-citra tersebut kemudian diubah ke dalam model *grayscale*. Konversi model citra dari RGB menjadi *grayscale* dilakukan dengan pengubahan komposisi dengan menggunakan Persamaan 1,

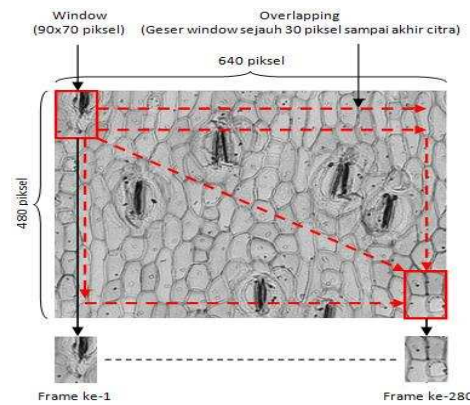
$$Grayscale = \alpha R + \beta G + \delta B \quad (1)$$

dengan nilai  $\alpha=0.299$ ,  $\beta=0.587$  dan  $\delta=0.11$  (Gonzalez dan Woods 2008).

#### 2 Windowing

Setelah diubah ke dalam model *grayscale*, langkah selanjutnya adalah proses *windowing*, yaitu pemotongan citra penampang daun dengan ukuran *window* tertentu secara *overlapping*. *Window* yang digunakan pada penelitian ini memiliki dimensi 90x70 piksel, dengan jarak *overlapping* yang digunakan adalah 30 piksel. Ukuran *window* ini telah disesuaikan dengan

dimensi dari seluruh stoma yang muncul di setiap citra penampang daun. Proses *windowing* tersebut dapat dilihat pada Gambar 2.



Gambar 2 Proses *windowing*

Dari seluruh *frame* yang dihasilkan pada proses *windowing*, terbentuk tiga jenis *frame* dengan karakter yang berbeda. Jenis pertama adalah *frame* yang di dalamnya terlihat stoma secara utuh (*frame* stoma). Jenis kedua adalah *frame* yang di dalamnya terlihat stoma, namun hanya sebagian (*frame* sebagian stoma), dan jenis ketiga adalah *frame* yang di dalamnya tidak terlihat stoma sama sekali (*frame* bukan stoma). Contoh dari ketiga jenis *frame* tersebut dapat dilihat pada Gambar 3.

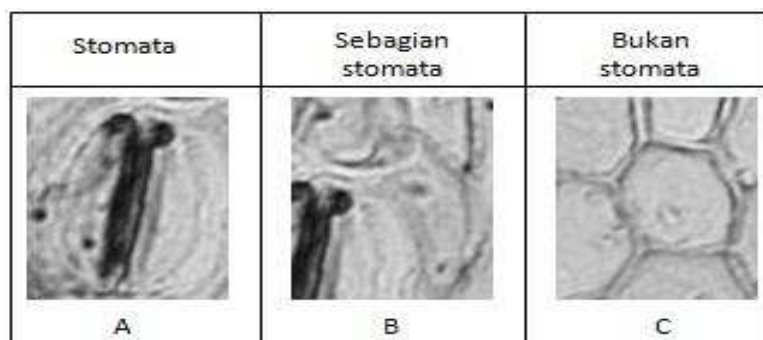
**Pembentukan Data Latih dan Data Uji**

**1 Pemilihan Citra Penampang Daun Latih dan Citra Penampang Daun Uji**

*Frame* yang digunakan sebagai data latih dan data uji berasal dari citra penampang daun yang berbeda. Hal ini karena pada proses pengujian, *input* yang digunakan adalah sebuah citra penampang daun secara utuh sehingga seluruh *frame* yang dihasilkan dari suatu citra penampang daun uji akan digunakan sebagai data uji dan tidak bisa digunakan sebagai data latih. Hal tersebut dilakukan agar dari proses pengujian dapat dilihat hasil deteksi kemunculan stoma terhadap sebuah citra penampang daun uji.

**2 Pemilihan *Frame* dari Citra Penampang Daun Latih dan Citra Penampang Daun Uji**

Untuk citra penampang daun latih, tidak semua *frame* yang terbentuk akan digunakan sebagai data latih. Hal ini karena jumlah *frame* yang dihasilkan terlalu banyak sehingga hanya akan dipilih beberapa *frame* saja. *Frame* yang dipilih tersebut harus bervariasi sehingga dapat mewakili *frame-frame* lainnya yang tidak terpilih. Dari tiap-tiap spesies, *frame* stoma yang terbentuk akan digunakan seluruhnya. Adapun *frame* sebagian stoma yang terbentuk hanya akan digunakan 100 *frame* dan *frame* bukan stoma yang terbentuk hanya akan digunakan 150



Gambar 3 A = frame stoma, B = frame sebagian stoma, dan C = frame bukan stoma

*frame*.

Untuk citra penampang daun uji, seluruh *frame* yang terbentuk akan digunakan sebagai data uji. Hal ini karena pada proses pengujian, data uji yang digunakan adalah berupa citra penampang daun secara utuh dan hasil dari proses pengujian tersebut akan dikembalikan ke citra asal sehingga dapat dilihat daerah-daerah pada citra tersebut yang terdeteksi sebagai *frame* stoma, *frame* sebagian stoma, dan *frame* bukan stoma.

### 3 Transformasi *Frame* dari 2D menjadi 1D

Sebelum memasuki tahap ekstraksi ciri, setiap *frame* yang digunakan sebagai data latih dan data uji harus ditransformasi ke dalam bentuk vektor (*frame* dengan bentuk 1 dimensi). Karena setiap *frame* memiliki dimensi  $90 \times 70$  piksel, setelah melalui proses transformasi, dimensi dari *frame* tersebut berubah menjadi  $1 \times (90 \times 70)$  atau  $1 \times (6300)$  piksel.

## Ekstraksi Ciri

### 1 Histogram

Histogram menunjukkan distribusi piksel berdasarkan intensitas *graylevel* (derajat keabuan) yang dimiliki oleh tiap-tiap piksel. Penggunaan histogram sebagai metode ekstraksi ciri didasarkan pada perbedaan sebaran atau distribusi piksel yang terjadi di antara *frame* stoma, *frame* sebagian stoma, dan *frame* bukan stoma. Pada metode ekstraksi ciri histogram, *bin* merupakan banyaknya batang warna yang akan terbentuk, atau menunjukkan jumlah pembagian rentang warna pada histogram. Jumlah titik ekstraksi ciri yang dihasilkan oleh suatu histogram adalah sama dengan jumlah *bin* yang digunakan pada histogram tersebut. Pada penelitian ini, nilai *bin* yang dicoba adalah 10, 20, 30, 40, dan 50 *bin*.

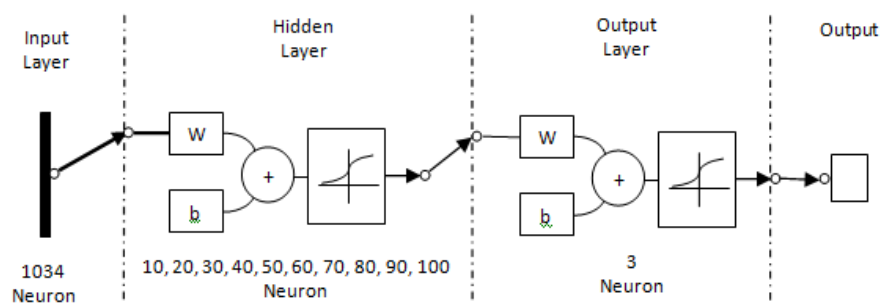
### 2 PCA (*Principal Component Analysis*)

Tujuan dari metode ekstraksi ciri PCA adalah memproyeksikan data latih dari suatu ruang dimensi tertentu ke ruang dimensi yang lebih rendah, tanpa kehilangan banyak informasi yang terkandung (Jolliffe 2002). Metode ekstraksi ciri PCA akan memberikan dua keluaran utama, yaitu matriks penciri dari data latih dan matriks transformasi. Matriks transformasi nantinya akan digunakan untuk mentransformasi data uji agar berada pada ruang dimensi yang sama dengan ruang dimensi dari data latih.

Dengan metode ekstraksi ciri PCA, nilai persentase data yang ingin dibuang, atau data yang dianggap kurang penting, dapat ditentukan. Pada penelitian ini, nilai persentase pembuangan komponen utama yang digunakan adalah 0.04%, 0.05%, 0.06%, 0.07%, 0.08%, 0.09%, dan 0.10%. Semakin besar persentase komponen utama yang dibuang, perhitungan selanjutnya semakin efisien. Namun, hal ini dapat berakibat pada berkurangnya akurasi.

## Pelatihan Jaringan Saraf Tiruan (JST)

Algoritme yang digunakan dalam tahap pelatihan JST adalah *backpropagation*. Fungsi transfer yang digunakan pada lapisan tersembunyi adalah logaritmik *sigmoid*, fungsi transfer yang digunakan pada lapisan *output* adalah logaritmik *sigmoid*, dan fungsi *training* jaringan adalah fungsi Levenberg-Marquardt. Untuk jumlah neuron pada lapisan tersembunyi, akan



Gambar 4 Arsitektur jaringan saraf tiruan yang digunakan

dicoba sepuluh nilai yang berbeda, yaitu 10, 20, 30, 40, 50, 60, 70, 80, 90, dan 100 neuron. Arsitektur dari JST yang digunakan pada penelitian ini dapat dilihat pada Gambar 4.

## Pengujian

Tahap pengujian ini dilakukan beberapa kali sesuai dengan jumlah citra penampang daun uji. *Input* dari setiap proses pengujian adalah sebuah citra penampang daun uji yang telah melewati tahap praproses dan ekstraksi ciri, serta matriks target dan sebuah jaringan terbaik yang dihasilkan pada proses pelatihan JST. Hasil dari proses pengenalan atau pengujian ini, akan dikembalikan ke citra penampang daun yang bersangkutan untuk melihat daerah-daerah pada citra tersebut yang terdeteksi sebagai kelas pertama (*frame* stoma), kelas kedua (*frame* sebagian stoma), dan kelas ketiga (*frame* bukan stoma).

## Analisis Hasil

Langkah pertama dalam tahap analisis adalah menghitung nilai *recall*, *precision*, dan *f<sub>1</sub>-measure*. Penghitungan nilai *recall* dan *precision* membutuhkan suatu matriks yang disebut dengan *confusion matrix*. Matriks tersebut dapat dilihat pada Tabel 1 (Manning *et al.* 2008). Definisi *recall* ditunjukkan pada Persamaan 2, sedangkan definisi *precision* ditunjukkan pada Persamaan 3.

Tabel 1 *Confusion matrix*

	Stoma	Bukan Stoma
Terdeteksi Sebagai Stoma	tp	Fp
Terdeteksi Sebagai Bukan Stoma	fn	tn

Keterangan:

tp: *true positive* (jumlah stoma pada citra penampang daun yang berhasil terdeteksi sebagai stoma).

tn: *true negative* (jumlah bukan stoma pada citra penampang daun yang berhasil terdeteksi sebagai bukan stoma).

fp: *false positive* (jumlah bukan stoma pada citra penampang daun yang terdeteksi sebagai stoma).

fn: *false negative* (jumlah stoma pada citra penampang daun yang terdeteksi sebagai bukan stoma).

$$recall = \frac{tp}{tp + fn} \quad (2)$$

$$precision = \frac{tp}{tp + fp} \quad (3)$$

*F<sub>β</sub>-measure* (Manning *et al.* 2008) adalah ukuran kinerja klasifikasi dalam mendeteksi kemunculan stoma dengan mengombinasikan nilai *recall* dan nilai *precision*. Persamaan dari *f<sub>β</sub>-measure* ditunjukkan pada Persamaan 4.

$$f_{\beta\_measure} = (1 + \beta^2) \cdot \frac{precision \cdot recall}{\beta^2 precision + recall} \quad (4)$$

Pada penelitian ini, nilai  $\beta$  yang digunakan adalah 1 sehingga persamaan *f<sub>β</sub>-measure* di atas berubah sebagaimana ditunjukkan pada Persamaan 5.

$$f_1\_measure = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (5)$$

Nilai 1 tersebut menyatakan bahwa bobot yang diberikan untuk *recall* dan *precision* adalah sama.

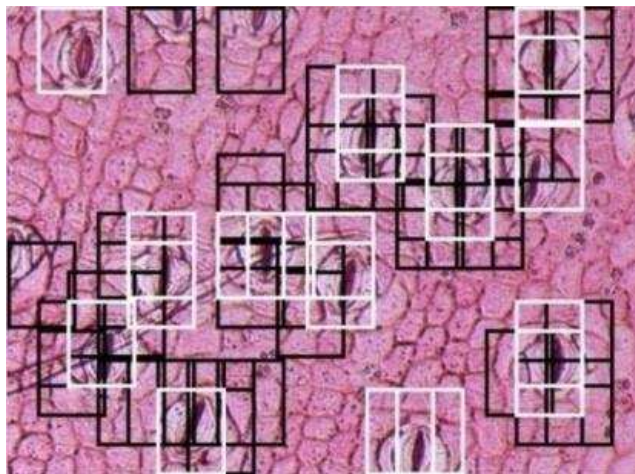
## HASIL DAN PEMBAHASAN

### Pengujian

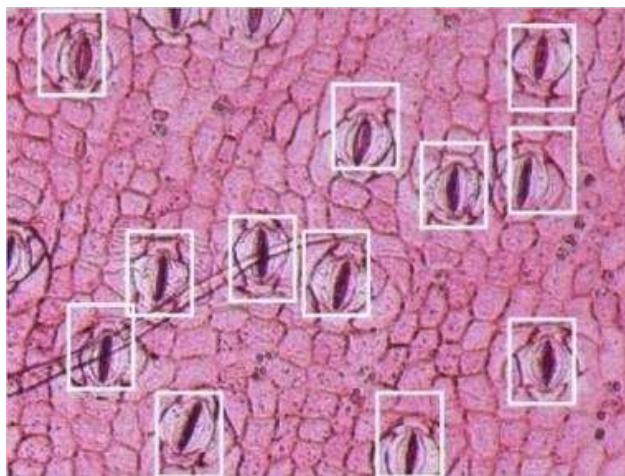
*Input* dari proses pengujian adalah sebuah citra penampang daun uji beserta jaringan terbaik yang diperoleh pada tahap pelatihan JST. Hasil dari proses pengujian ini akan dikembalikan ke citra penampang daun uji yang bersangkutan untuk melihat daerah-daerah pada citra penampang daun uji tersebut yang terdeteksi sebagai kelas pertama (*frame* stoma), kelas kedua (*frame* sebagian stoma), dan kelas ketiga (*frame* bukan stoma). Contoh hasil dari proses pengujian dapat dilihat pada Gambar 5.

Gambar 5 memperlihatkan kotak-kotak berwarna putih dan hitam. Kotak putih menunjukkan bahwa daerah tersebut merupakan daerah yang terdeteksi sebagai *frame* stoma. Kotak hitam menunjukkan bahwa daerah tersebut merupakan daerah yang terdeteksi sebagai *frame* sebagian stoma. Selain kotak putih dan kotak hitam, terdapat juga daerah tanpa kotak yang merupakan daerah yang terdeteksi sebagai *frame* bukan stoma.

Pada Gambar 5, terlihat beberapa stoma yang seharusnya terdeteksi sebagai sebuah stoma, namun muncul beberapa kotak putih yang mengelilinginya. Oleh sebab itu, agar kotak-kotak putih tersebut berkumpul menjadi satu dan tepat berada di wilayah stoma yang benar, perlu adanya proses penggabungan hasil deteksi. Metode penggabungan hasil deteksi yang digunakan pada penelitian ini ialah jarak Euclid (*Euclidean distance*) yang dapat dilihat pada Gambar 6.



Gambar 5 Hasil deteksi awal



Gambar 6 Hasil penggabungan daerah deteksi

### Analisis Hasil Histogram

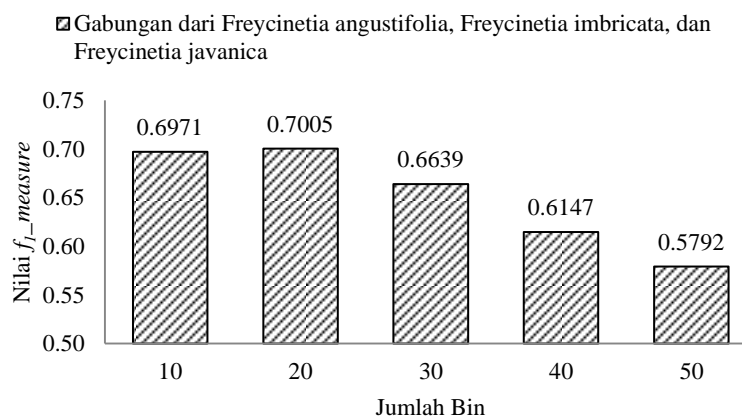
Terdapat lima jaringan terbaik yang dihasilkan histogram berdasarkan jumlah *bin* yang digunakan pada penelitian ini. Nilai *recall*, *precision*, dan *f<sub>1</sub>-measure* yang dihasilkan histogram untuk ketiga jenis spesies dapat dilihat pada Tabel 2.

Tabel 2 Nilai *recall*, *precision*, dan *f<sub>1</sub>-measure* gabungan dari ketiga jenis spesies yang dihasilkan dari metode histogram

Jumlah <i>bin</i>	<i>Recall</i>	<i>Precision</i>	<i>F<sub>1</sub>-measure</i>
10	0.5840	0.8111	0.6791
20	0.6080	0.8261	0.7005
30	0.6320	0.6991	0.6639
40	0.5360	0.7204	0.6147
50	0.5120	0.6667	0.5792

Agar perbandingan nilai-nilai *f<sub>1</sub>-measure* gabungan dari ketiga jenis spesies terlihat lebih jelas, maka nilai-nilai tersebut akan disajikan kedalam bentuk grafik. Grafik nilai *f<sub>1</sub>-measure* gabungan berdasarkan jumlah *bin* yang digunakan dapat dilihat pada Gambar 7.

Berdasarkan grafik pada Gambar 7, dapat dilihat bahwa jumlah *bin* terbaik untuk mendeteksi kemunculan stoma pada citra penampang daun *Freycinetia* adalah 20 *bin* dengan nilai *f<sub>1</sub>-measure* yang dihasilkan adalah 0.7005. Hal ini terjadi karena 20 *bin* adalah jumlah titik ekstraksi terbaik yang mampu merepresentasikan karakteristik yang berbeda bagi ketiga kelas yang ada (stoma, sebagian stoma, dan bukan stoma).



Gambar 7 Nilai *f<sub>1</sub>-measure* gabungan tiga spesies berdasarkan jumlah *bin*

### Analisis Hasil PCA

Terdapat tujuh jaringan terbaik yang dihasilkan PCA berdasarkan persentase pembuangan komponen utama dengan nilai *recall*, *precision*, dan *f<sub>1</sub>-measure* yang dihasilkan PCA untuk ketiga spesies dapat dilihat pada Tabel 3.

Tabel 3 Nilai *recall*, *precision*, *f<sub>1</sub>-measure* gabungan dari ketiga jenis spesies yang dihasilkan PCA berdasarkan persentase pembuangan komponen utama

Komponen utama dibuang jika memiliki kontribusi	<i>Recall</i>	<i>Precision</i>	<i>F<sub>1</sub>-measure</i>
< 0.04%	0.9360	0.7312	0.8211
< 0.05%	0.9360	0.8478	0.8897
< 0.06%	0.9600	0.7947	0.8696
< 0.07%	0.9600	0.8633	0.9091
< 0.08%	0.9280	0.7785	0.8467
< 0.09%	0.9360	0.6802	0.7879
< 0.10%	0.9680	0.8067	0.8800



Persentase pembuangan komponen utama terbaik untuk mendeteksi kemunculan stoma pada citra penampang daun *Freycinetia* ialah 0.07%, dengan nilai  $f_1\_measure$  0.9091. Namun, dapat dilihat juga bahwa rentang nilai  $f_1\_measure$  yang dihasilkan tidak terlalu jauh. Selisih antara nilai  $f_1\_measure$  terbesar dengan nilai  $f_1\_measure$  terkecil hanya 0.1212. Hal ini terjadi karena perbedaan nilai persentase pembuangan komponen utama yang digunakan sangat kecil (hanya berbeda 0.01% di antara nilai persentase digunakan) sehingga jumlah informasi yang dihasilkan hampir sama.

## SIMPULAN

Pada penelitian ini, telah diuji dua metode ekstraksi fitur untuk deteksi kemunculan stoma pada citra mikroskopis penampang daun *Freycinetia*, yaitu: metode histogram dan PCA. Nilai  $f_1\_measure$  terbaik yang mampu dicapai oleh metode ekstraksi ciri Histogram adalah 0.7005, sedangkan metode PCA 0.9091. Secara umum dapat disimpulkan bahwa metode PCA menghasilkan akurasi yang lebih baik dibanding metode histogram, dan mampu memberikan unjuk kerja yang memuaskan sehingga dapat dikembangkan untuk penelitian-penelitian selanjutnya, ataupun untuk digabungkan dengan metode pengenalan spesies *Freycinetia* yang telah ada sebelumnya, seperti pada Qur'ania (2012).

## DAFTAR PUSTAKA

- Gonzalez RC, Woods RE. 2008. *Digital Image Processing*. Ed ke-3. New Jersey (US): Prentice Hall.
- Jolliffe IT. 2002. *Principle Component Analysis*. Ed ke-2. New York (US): Springer-Verlag.
- Manning CD, Raghavan P, Schutze D. 2008. *Introduction to Information Retrieval*. Cambridge (UK): Cambridge University Pr.
- Pasaribu N. 2010. *Freycinetia (Pandanaceae) of Sumatera* [disertasi]. Bogor (ID): Institut Pertanian Bogor.
- Qur'ania A. 2012. *Klasifikasi Freycinetia berbasis citra anatomi stomata menggunakan k-Nearest Neighbor dan Jaringan Saraf Tiruan* [tesis]. Bogor (ID): Institut Pertanian Bogor.
- Sanyal P, Bhattacharya U, Bandyopadhyay SK. 2008. Analysis of SEM images of stomata of different tomato cultivars based on morphological features. Di dalam: *Second Asia International Conference on Modelling & Simulation*; 2008 Mei 13-15; Kuala Lumpur, Malaysia. New York (US): IEEE. hlm 890-894.
- Willmer CM. 1983. *Stoma*. New York (US): Longman.