Jurnal Ilmu Komputer Agri-Informatika

Tersedia secara online di:
http://journal.ipb.ac.id/index.php.jika

Volume 2 Nomor 1 halaman 38-46
ISSN: 2089-6026

Pengenalan Genus Diatom Menggunakan *Principal Component Analysis* dan Jaringan Saraf Tiruan Propagasi Balik Sebagai *Classifier*

Identification of the Diatoms Genus Using Principal Component Analysis and Backpropagation Neural Network as Classifier

SILVIA RAHMI[1*], TOTO HARYANTO[1], NIKEN T. M PRATIWI[2]

Abstrak

Diatom merupakan suatu mikroalga unisel (kadang berkoloni) yang mempunyai peranan penting dalam dunia riset dan penelitian. Identifikasi diatom merupakan pekerjaan yang rumit. Hal ini dikarenakan diatom memiliki ratusan taksa dengan banyak variasi bentuk dan karakteristik biologi yang menyebabkan proses identifikasinya tidak mudah bahkan bagi seorang pakar. Penelitian ini menerapkan *Principal Component Analysis* (PCA) untuk reduksi data dan Jaringan Saraf Tiruan (JST) untuk identifikasi diatom. Proporsi PCA yang digunakan ialah 80% dan 90%. JST yang digunakan adalah propagasi balik dengan satu *hidden layer*. Data yang dipakai pada penelitian ini adalah citra diatom berformat JPG yang diambil menggunakan mikroskop elektrik. Hasil penelitian menunjukkan bahwa generalisasi terbaik sebesar 90% diperoleh pada percobaan menggunakan proporsi PCA 90% dengan persentase data latih 80%.

Abstract

Diatoms are microalgae uniselular (sometimes colonized) that have an important role in the world of research and study. Identification of diatoms is a complex task. This is because diatoms have hundreds of taxa with many variation forms and biological characteristics so that the process of identification is not easy even for an expert. This study applied Principal Component Analysis (PCA) for data reduction and Artificial Neural Network (ANN) for identification of diatoms. The proportion of PCA used in this study was 80% and 90% and Backpropagation ANN is used with one hidden layer. The data used in this research isthe image of diatoms in JPG format.This data was obtained by sampling using electric microscope. The results showed that the best generalization rate is 90% was obtained in an experiment using 90% PCA proportion and80%training data.

PENDAHULUAN

Diatom (*Bacillariophyceae*) merupakan suatu mikroalga unisel (kadang berkoloni) dengan ukuran berkisar antara2µm sampai 4 mm. Diatom dapat ditemukan di ekosistem perairan tawar maupun ekosistem laut dan secara umum hidup pada tempat yang lembab. Diatom mempunyai peranan penting dalam dunia riset dan penelitian. Peranan tersebut di antaranya ialah sebagai indikator kualitas air, untuk pembuatan kapsul obat dan penentuan umur fosil (Forero *et al.* 2004).

Struktur sel diatom berbeda dari alga lainnya karena diatom memiliki cangkang yang terbuat dari silika yang disebut *frustul*, yang terdiri atas dua bagian. Karakteristiknya dijadikan sebagai kunci identifikasi diatom. Identifikasi diatom biasanya dilakukan secara manual dengan membandingkan pengamatan melalui mikroskop dengan gambar diatom yang terdapat pada buku identifikasi (Tomas 1997), namun hal ini cukup rumit dan membutuhkan waktu. Diatom memiliki ratusan taksa (nama yang diberikan kepada sekelompok taksonom dalam sistem nomenklatur) dengan banyak variasi bentuk dan karakteristik biologisyang menyebabkan proses identifikasinya tidak mudah (Forero *et al.* 2004). Oleh karena itu, sistem

identifikasi diatom berbasis citra digital untuk mempermudah identifikasi diatom secara otomatis dan cepat diperlukan.

Budiman (2008) menggunakan Jaringan Saraf Tiruan (JST) *Backpropagation* sebagai teknik identifikasi spesies nematoda puru akar melalui karakteristik morfologi ekor. Analisis komponen utama (PCA*)* digunakan sebagai metode ekstraksi ciri menghasilkan akurasi sebesar 83.3%. Tingginya akurasi yang diperoleh pada penelitian tersebut melatarbelakangi penelitian ini, yaitu JST Propagasi Balik dan ekstraksi ciri PCA untuk identifikasi genusdiatom.

METODE

Data citra diatom yang digunakan adalah citra digital berformat JPG. Sampel mikrolofi dipotret menggunakan mikroskop elektrik. – Laboratorium Biomikro, Fakultas Perikanan dan Ilmu Kelautan, Institut Pertanian Bogor – Citra yang dipakai merupakan citra diatom dengan subordo, family, dan genus diatom yang terdapat pada Tabel 1. Jumlah keseluruhan famili diatom yang digunakan dalam penelitian berjumlah 6 famili yang terdiri atas 10 genus. Dalam penelitian ini digunakan 10 citra sehingga total citra adalah 100 citra.

Tabel 1 Rincian genus diatom yang digunakan dalam penelitian

Subordo	Famili	Genus
Coscinodisceae	Thalassiosiraceae *	Thalassiosira, Cyclotella, Lauderia, Skeletonema
	Melosiraceae*	Melosira
	Coscinodisceae*	Coscinodiscus
Biddulphianeae	Hemiaulaceace*	Eucampi,
Fragillariineae	Fragillariaceae*	Fragillaria, Asterionela
	Surirelaceae**	Surirela

Keterangan :
* Tomas (1997).
** Presscott (1970).

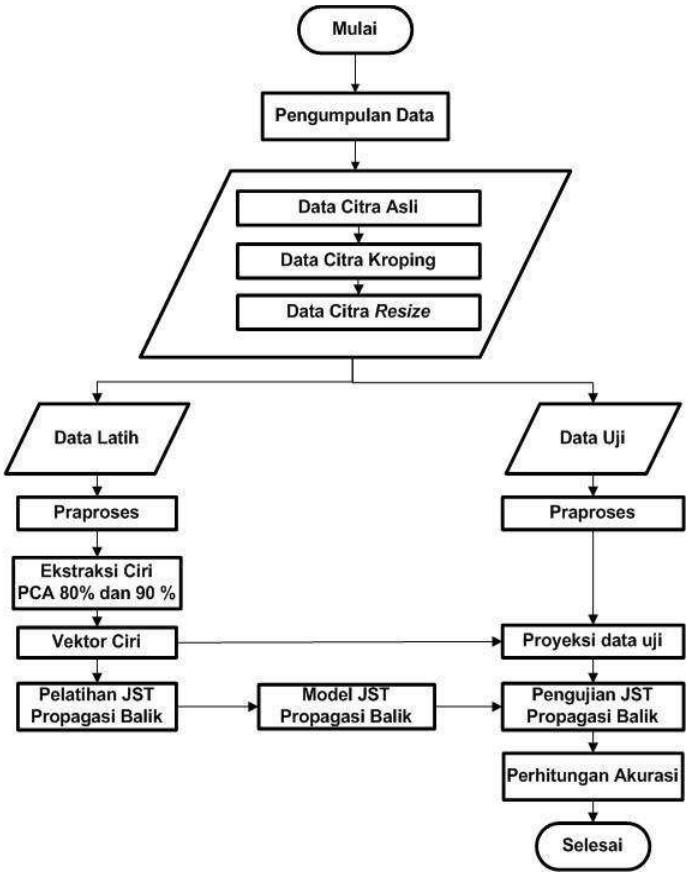
Proses Pengenalan Diatom

Pengenalan diatom menggunakan model klasifikasi JST *Backpropagation* dan ektraksi ciri PCA diawali dengan proses pengumpulan data diatom yang kemudian dijadikan sebagai data latih dan data uji dengan proporsi tertentu. Proses ekstraksi ciri dilakukan menggunakan PCA yang menghasilkan vektor ciri yang kemudian diproyeksikan terhadap data uji sehingga didapatkan dimensi yang sama agar dapat dilakukan proses klasifikasi dengan menggunakan JST (Gambar 1)

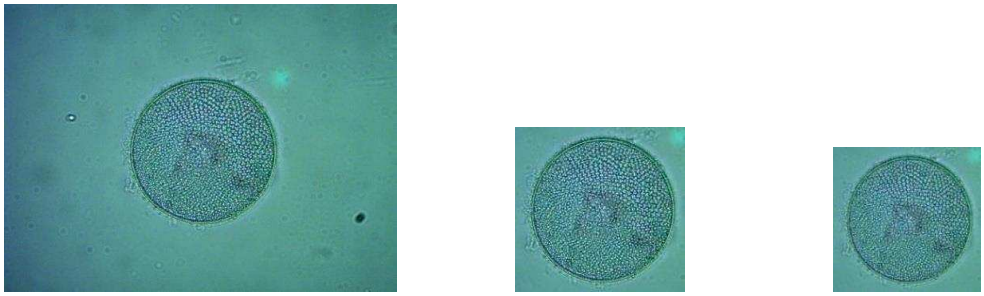
a Fase pengumpulan data merupakan tahap pengambilan data citra menggunakan mikroskop elektrik. Citra diatom awal yang berukuran 2560x1920 piksel selanjutnya dipotong dengan ukuran 1127x1097 piksel dan diperkecil menjadi 60x60 piksel (Gambar 2).
b Fase pembagian data membagi data citra menjadi 2 bagian, yaitu data latih dan data uji. Persentase yang digunakan terdiri atas dua skenario percobaan. Pada skenario pertama, data dibagi menjadi 60% data latih dan 40% data uji dan pada skenario kedua data dibagi menjadi 80% data latih dan 20% data uji.
c Fase praproses mengubah citra *red–green–blue* (RGB) hasil proses pemasukan data citra menjadi citra *grayscale*. Proses ini dilakukan untuk mengubah 3 layer matriks warna RGB menjadi 1 layer matriks citra keabuan. Hal ini berguna untuk mempercepat pengolahan citra pada fase ekstraksi ciri menggunakan PCA. Pada tahap praproses ini juga dilakukan normalisasi data untuk menskalakan masukan dan target sehingga berada dalam rentang tertentu.
d Tahap ekstraksi ciri PCA mereduksi dimensi data latih menggunakan proporsi PCA 80% dan 90%. Matriks kovarian dari citrayang telah terskala berukuran 3600x3600 piksel akan menghasilkan komponen utama dan nilai Eigen. Hasil dari komponen utama berupa vektor

ciri yang akan diproyeksikan terhadap data latih dan data uji. Proporsi PCA 80% berarti mengambil nilai kolom matriks yang merepresentasikan komponen utama sebesar 80%. Masukan matriks yang akan masuk pada tahap pelatihan JST *Backpropagation* merupakan hasil proyeksi vektor ciri terhadap citra latih. Untuk proses pengenalan, suatu citra uji memiliki dimensi yang sama dengan citra latih telah disajikan ke sistem. Citra uji tersebut diekstraksi ciri dengan cara dikalikan dengan vektor ciri citra latih dan akan menghasilkan vektor ciri berisikan komponen utama yang memiliki dimensi yang sama dengan komponen utama data latih yang masuk pada pelatihan JST.

e Data latih yang didapat pada tahap ekstraksi ciri dengan PCA akan masuk ke proses pelatihan JST. Hasil pelatihan akan menghasilkan model klasifikasi dan dilakukan pengujian menggunakan data uji yang juga mengalami proses *grayscale* dan normalisasi. Pada tahap ini, klasifikasi menggunakan JST *Backpropagation* akan dilakukan dalam beberapa percobaan dengan mengombinasikan persentase pembagian data latih dan data uji, parameter proporsi PCA, *hidden neuron*, dan toleransi kesalahan yang digunakan sehingga menghasilkan parameter optimal yang menghasilkan akurasi terbaik.



Gambar 1 Tahapan Proses Pengenalan Diatom dengan JST *Backpropagation*



(a) Ukuran asli 2560 x 2920 piksel (b) Pemotongan 1127 x 1097 piksel (c) Penskalaan 60 x 60 piksel

Gambar 2 Citra diatom

Struktur JST *Backpropagation*

Identifikasi diatom dilakukan dengan menggunakan JST metode pelatihan *Backpropagation* dengan struktur JST menggunakan satu *hidden layer* dengan sepuluh *neuron* (Tabel 2).

Tabel 2 Struktur JST Backpropagation

Karakteristik	Spesifikasi
Arsitektur	1 *hidden layer*
Input neuron	Dimensi PCA 80% dan 90%
Hidden neuron	10, 20, 30, 40, 50, 60, 70, 80, 90, dan 100
Output neuron	Banyaknya kelas target
Inisialisasi bobot dan bias	Nguyen-Widrow*
Fungsi aktivasi	Sigmoid bipolar, Sigmoid biner
Algoritme pelatihan	Traingdx
Toleransi galat	10^{-2}, 10^{-3}, dan 10^{-4}
Laju pembelajaran	10^{-1}

*(Siang 2009)

Kelas target pada penelitian ini berjumlah 10 dan setiap target mewakili satu genus dari diatom yang direpresentasikan dengan nilai 0 dan 1 (Tabel 3).

Tabel 3 Kelas target JST *Backpropagation*

Kelas	Target
Coscinodiscus	1000000000
Asterionella	0100000000
Fragillaria	0010000000
Eucampia	0001000000
Melosira	0000100000
Skeletonema	0000010000
Surirella	0000001000
Cyclotella	0000000100
Lauderia	0000000010
Thalassiosira	0000000001

Parameter Pengenalan Diatom

Parameter yang digunakan untuk mengetahui tingkat keberhasilan proses identifikasi diatom menggunakan JST *Backpropagation* ialah konvergensi dan generalisasi. Konvergensi adalah tingkat kecepatan jaringan mempelajari pola masukan yang dinyatakan dengan satuan waktu atau satuan *epoch*. Satu *epoch* ialah lamanya jaringan mempelajari satu kali seluruh pola pelatihan.

Maksimum *epoch* pada penelitian ini dibatasi sebanyak 5000. Generalisasi adalah tingkat pengenalan jaringan dalam mengenali sejumlah pola yang diberikan. Secara matematis generalisasi dapat dituliskan seperti pada Persamaan 1 (Sarbini *et al.* 2002).

$$Generalisasi = \frac{jumlah\_pengenalan\_benar}{jumlah\_seluruhnya} \text{x}100\%$$

Lingkungan Pengembangan

Perangkat keras dan perangkat lunak yang digunakan untuk penelitian ini memiliki prosesor Intel® CoreTM 2 Duo, memory 1 GB, *harddisk* 250 GB, sistem operasi Microsoft Windows 7, dan Matlab 7.7 (R2008b)

HASIL DAN PEMBAHASAN

Penelitian ini terdiri atas empat percobaan, yaitu kombinasi dua buah pembagian data latih dengan dua proporsi PCA. Pada setiap percobaan digunakan kombinasiparameter JST dengan toleransi kesalahan 10^{-2}, 10^{-3}, dan 10^{-4} serta *hidden neuron* 10, 20, 30, 40, 50, 60, 70, 80, 90, dan 100.

PCA proporsi 80% dengan pembagian data latih 60% dan data uji 40%

Matriks kovarian berukuran 3600x3600 piksel yang berasal dari citra hasil normalisasi (60x3600 piksel) menghasilkan nilai Eigen yang mewakili 3600 kolom. Proporsi 80% menghasilkan komponen utama berdimensi 27, yang berarti data sebanyak 27 kolom mewakili sebesar 80% data citra. Komponen utama dari proporsi 80% berupa matriks berukuran 3600 x 27.Matriks PCA yang menjadi masukan pada JST merupakan hasil proyeksi dari matriks citra latih hasil normalisasi dengan komponen utama sehingga dimensi matriks masukan JST pada percobaan ini ialah 60 x 27.

a Generalisasi

Toleransi kesalahan 10^{-2}, 10^{-3}, dan 10^{-4} masing-masing akan dikombinasikan dengan semua *hidden neuron* sehingga akan didapatkan parameter optimal. Dari ketiga toleransi kesalahan didapat akurasi terbaik sebesar 80%, yaitu pada percobaan menggunakan toleransi kesalahan 10^{-2} (*hidden neuron* 50) dan pada percobaan menggunakan toleransi kesalahan 10^{-4} (*hidden neuron* 30).

b Konvergensi

Waktu dan jumlah iterasi pelatihan terkecil pada percobaan ini ialah pada toleransi kesalahan 10^{-2}, yaitu sebesar 3.41 detik dengan 298 iterasi. Toleransi kesalahan 10^{-3} menghasilkan akurasi maksimum dengan lama waktu latih 7.6 detik dan 786 iterasi. Adapun pada toleransi 10^{-4}, waktu latih yang dibutuhkan 21.3 detik dengan 2210 iterasi.

PCA proporsi 90% dengan pembagian data latih 60% dan data uji 40%

Percobaan menggunakan proporsi PCA 90% menghasilkan komponen matriks berdimensi 60 x 38. Matriks hasil ekstraksi inilah yang akan menjadi masukan untuk klasifikasi diatom menggunakan JST *Backpropagation.*

a Generalisasi

Pada Percobaan 2 terlihat bahwa akurasi lebih tinggi dibandingkan dengan percobaan sebelumnya. Akurasi terbaik sebesar 82.5 % yang didapatkan pada toleransi galat $10^{-4.}$

b Konvergensi

Waktu pelatihan dan jumlah iterasi pada toleransi 10^{-2} ialah 3.1 detik dengan 298 *epoch*. Pada toleransi kesalahan 10^{-3} memerlukan waktu latih 5.1 detik dengan501 iterasi, sedangkan pada toleransi kesalahan 10^{-4} waktu latih tercatat sebesar 9.3 detik dengan 780 iterasi.

PCA proporsi 80% dengan pembagian data latih 80% dan data uji 20%

Percobaan 3menghasilkan komponen utama berdimensi 80 x 29. Hasil pengukuran parameter kinerja untuk percobaan ini adalah sebagai berikut:

a Generalisasi

Akurasi maksimum pada Percobaan 3 dengan toleransi galat 10^{-2} adalah 85% yang di dapat pada *hidden neuron* 80. Akurasi maksimum sebesar 85% juga diperoleh ketika toleransi galat diturunkan menjadi 10^{-3}, yang terdapat pada *hidden neuron* 100. Pada toleransi galat 10^{-4} akurasi maksimum 85% dihasilkan pada *hidden neuron* 20.

b Konvergensi

Pada toleransi galat 10^{-2} dibutuhkan waktu latih sebesar 1.7 detik dengan 119 *epoch*. Sementara percobaan menggunakan toleransi galat 10^{-3} dan 10^{-4} membutuhkan waktu latih dan jumlah iterasi masing-masing sebesar 6.9 detik dengan 590 iterasi dan 29.3 detik dengan 2880 iterasi.

PCA proporsi 90% dengan pembagian data latih 80% dan data uji 20%

Percobaan ini menghasilkan komponen utama berdimensi 80 x 42. Hasil pengukuran parameter kinerja untuk percobaan ini adalah sebagai berikut:

a Generalisasi

Akurasi maksimum pada toleransi galat 10^{-2} adalah sebesar 90% yang didapat pada *hidden neuron* 100.Adapun pada toleransii galat 10^{-3}, akurasi maksimum hanya sebesar 80% diperoleh pada *hidden neuron* 30, 40, 50, 60, 70, 80, dan 90. Pada toleransi galat 10^{-4} menghasilkan akurasi maksimum 80% pada *hidden neuron* 20, 30, 90, dan 100.

b Konvergensi

Waktu latih sebesar 3.03 detik dengan 232 *epoch* diperoleh pada percobaan dengan menggunakan toleransi galat 10^{-2}. Sementara itu, percobaan menggunakan toleransi galat 10^{-3} dan 10^{-4} membutuhkan waktu latih masing-masing sebesar 5.8 detik dengan 564 iterasi dan 21.5 detik dengan 2181 iterasi.

Perbandingan Empat Percobaan

a Generalisasi

Setelah dilakukan empat buah percobaan yang merupakan kombinasi dari dua buah persentase pembagian data dan dua buah proporsi PCA, didapatkan akurasi maksimum pada Percobaan 4 yaitu sebesar 90%. Hal ini menunjukkan bahwa semakin besar komponen utama yang dipakai untuk pelatihan JST *Backpropagation*akan memudahkan JST dalam melakukan pengenalan. Tabel 4 menunjukkan perbandingan akurasi maksimum empat percobaan dengan ketiga toleransi galat.

Tabel 4 Akurasi maksimum empat percobaan

Toleransi Galat	Generalisasi (%)			
	P1	P2	P3	P4
10^{-2}	80.0%	80.0%	85.0%	90.0%
10^{-3}	75.0%	80.0%	85.0%	80.0%
10^{-4}	80.0%	82.5%	85.0%	80.0%

Keterangan : P1 = Percobaan 1, P2 = Percobaan 2, P3 = Percobaan 3, P4 = Percobaan 4

b Konvergensi

Dalam hal konvergensi terlihat bahwa perbedaan persentase pembagian data latih dan uji serta peningkatan proporsi PCA tidak mempengaruhi lamanya waktu pelatihan dan jumlah *epoch*.Tabel 5 menunjukkan waktu pelatihan untuk akurasi maksimum masing-masing percobaan pada tiga toleransi galat.

Berdasarkan data pada Tabel 5, terlihat bahwa waktu pelatihan terkecil dari keempat percobaan terdapat pada percobaan 3 dengan proporsi PCA 80% dan toleransi kesalahan 10^{-2} yaitu sebesar 1.7 detik. Jumlah *epoch*terkecil juga terdapat pada Percobaan 3 menggunakan toleransi kesalahan 10^{-2}.Tabel 6 menunjukkan banyaknya *epoch* untuk akurasi maksimum masing-masing percobaan pada ketiga toleransi galat.

Tabel 5 Waktu pelatihan untuk akurasi maksimum empat percobaan

Toleransi Galat	Waktu Latih (detik)			
	P1	P2	P3	P4
10^{-2}	3.41	3.10	1.70	3.03
10^{-3}	7.60	5.10	6.90	5.80
10^{-4}	21.30	9.30	29.30	21.50

Keterangan : P1 = Percobaan 1, P2 = Percobaan 2, P3 = Percobaan 3, P4 = Percobaan 4

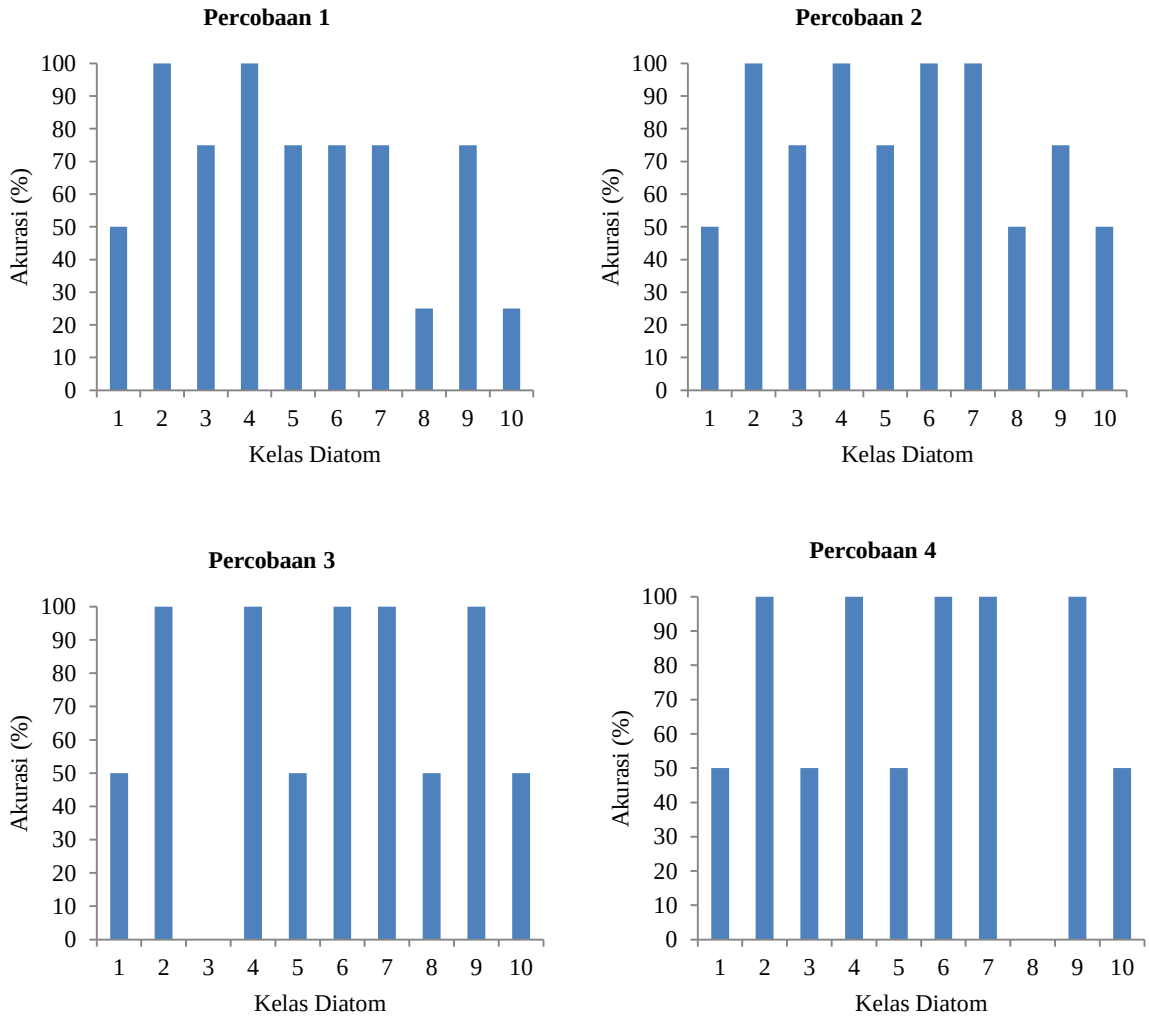
Tabel 6 Jumlah *epoch* untuk akurasi maksimumempat percobaan

Toleransi Galat	Banyaknya Epoch			
	P1	P2	P3	P4
10^{-2}	298	298	118	232
10^{-3}	786	501	590	564
10^{-4}	2210	780	2888	2181

Keterangan : P1 = Percobaan 1, P2 = Percobaan 2, P3 = Percobaan 3, P4 = Percobaan 4

Analisis *Error*

Secara umum, penerapan metode reduksi dimensi PCA dan *classifier* JST *Backpropagation* menghasilkan akurasi yang baik, namun perlu dilakukan analisis*error* untuk mengetahui akurasi setiap genus pada masing-masing percobaan.Akurasi per genus ditunjukkan pada Gambar 3.



Keterangan:

1	2	3	4	5	6	7	8	9	10
Coscinodiscus	*Asterionella*	*Fragillaria*	*Eucampia*	*Melosira*	*Skeletonema*	*Surirella*	*Cyclotella*	*Lauderia*	*Thalassiosira*

Gambar 3 Perbandingan akurasi per genus diatom

Secara umum kesalahan klasifikasi terjadi pada diatom yang memang memiliki kemiripan bentuk dilihat dari penampang atas. Tabel 7 menunjukkan kelas citra dengan

akurasi rendah dan salah teridentifikasi. Kelas yang cenderung memiliki akurasi tinggi pada keempat percobaan tersebut ialah kelas 2, 4, 6, 7 dan 9. Kelas tersebut cenderung memiliki tingkat kemiripan yang rendah dengan kelas-kelas lainnya.

Tabel 7 Kelas citra dengan akurasi rendah

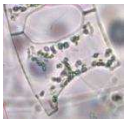
Kelas Awal	Kelas Hasil Identifikasi
1	8
3	6
5	7, 8, 10
8	1
10	6, 8

Gambar 4 menunjukkan citra yang cenderung menghasilkan akurasi tinggi.Umumnya citra diatom tersebut memang unik sehingga mudah untuk diklasifikasikan.


(a) *Asterionella*


(b) *Eucampia*

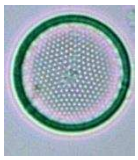

(c)) *Skeletonema*


(d) *Surirella*

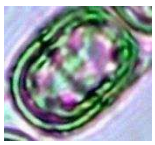

(e) *Lauderia*

Gambar 4 Kelas citra dengan akurasi tinggi

Kelas yang cenderung memiliki akurasi rendah adalah kelas 1, 3, 5, 8 dan 10. Gambar 5 menunjukkan citra dengan akurasi rendah yang sering teridentifikasi ke kelas lain.


(a) *Coscinodiscus*


(b) *Fragillaria*


(c) *Melosira*


(d) *Cyclotella*


(e) *Thalassiosira*

Gambar 5 Citra dengan akurasi rendah

Rendahnya akurasi disebabkan karena citra untuk lima genus ini cenderung memiliki tingkat kemiripan yang tinggi dengan genus lainnya. Selain itu posisi pengambilan sel diatom yang dilakukan hanya dari posisi atas sel. Hal ini juga memberikan pengaruh besarterhadap besarnya akurasi. Beberapa genus cenderung mirip dari posisi atas, tapi sangat jauh berbeda dari sisi lateral.

SIMPULAN

Model Jaringan Saraf Tiruan dapat digunakan untuk pengenalan beberapa genus diatom. Penggunakan persentase data latih sebesar 80% memberikan hasil akurasi yang lebih baik dibandingkan dengan persentase data latih 60%. PCA dengan proporsi 90% pada toleransi kesalahan 10^{-2} menggunakan persentase pembagian data latih 80% menghasilkan kinerja JST optimal dengan akurasi 90% pada *hidden neuron* 100.

DAFTAR PUSTAKA

Budiman R. 2008. Pengenalan spesies nematoda puru akar melalui karakteristik morfologi ekor menggunakan jaringan syaraf tiruan [Skripsi]. Bogor(ID): Institut Pertanian Bogor.

Sarbini, Rachmaniah M, Buono A. 2002. *Perbandingan Metode Eigen pada Pengenalan Wajah*. Bogor (ID): Institut Pertanian Bogor.

Forero MG, Sroubek F, Flusser J, Redondo R, Cristobal G. 2004. Automatic screening and multifocus fusion methods for diatom identification. Di dalam: Presscott GW. 1970. *The Freshwater Algae*. Cincinnati (US): Brown Company Publishers.

Siang JJ. 2009. *Jaringan Syaraf Tiruan dan Pemrogramannya Menggunakan MATLAB.*Yogyakarta (ID): Andi Offset.

Tomas CR. 1997. *Identifying Marine Phytoplankton.* Florida (US): Florida Marine Research Institute.