

# Identifikasi Kualitas Buah Tomat dengan Metode PCA (*Principal Component Analysis*) dan *Backpropagation*

Umi Murdika<sup>1</sup>, Muhammad Alif<sup>2</sup>, Yessi Mulyani<sup>3</sup>

<sup>1,2,3</sup>Jurusan Teknik Elektro Universitas Lampung, Bandar Lampung

Jl. Prof. Sumantri Brojonegoro No.1 Bandar Lampung 35145

<sup>1</sup>umi.murdika@eng.unila.ac.id

**Intisari** — Analisis komponen utama atau *principal component analysis* merupakan suatu metode yang digunakan dalam menganalisis kumpulan dataset untuk meringkas karakteristik utama mereka. Metode PCA ini mengurangi dimensi dataset dengan memproyeksikan setiap titik data ke hanya beberapa komponen utama pertama untuk mendapatkan data berdimensi lebih rendah sambil mempertahankan sebanyak mungkin variasi data. Pada penelitian ini digunakan metode PCA untuk memproyeksikan data citra sehingga diperoleh data ekstraksi ciri dengan dimensi yang lebih kecil. Selanjutnya metode Backpropagation diterapkan untuk melakukan proses identifikasi. Dataset yang digunakan sebanyak 30 dataset yang terdiri dari 10 buah citra uji dan 20 data latih. Dari simulasi yang dilakukan diperoleh kesimpulan bahwa Metode PCA yang diterapkan berhasil mengurangi dimensi data. Identifikasi kualitas buah tomat menggunakan metode Back-propagation menunjukkan tingkat ketepatan dengan akurasi mencapai 76,7%. Nilai akurasi tersebut menunjukkan bahwa sistem ini telah berjalan dengan baik.

**Kata kunci** — *Back-propagation*, PCA (*Principal Component Analysis*), Pengolahan Citra, JST (Jaringan Syaraf Tiruan).

**Abstract** — Principal component analysis is a method used in analyzing datasets to summarize their main characteristics. This PCA method reduces the dimension of the dataset by projecting each data point onto only the first few principal components to obtain lower dimensional data while maintaining as much variation of the data as possible. In this study, the PCA method was used to project image data in order to obtain feature extraction data with smaller dimensions. Furthermore, the Backpropagation method is applied to carry out the identification process. The dataset used is 30 data consisting of 10 test images data and 20 training data. From the simulation, it can be concluded that the PCA method applied has succeeded in reducing the dimensions of the data. Identification of tomato fruit quality using the Back-propagation method shows the level of accuracy with an accuracy of 76.7%. it indicates that this system has been running well.

**Keywords** — Backpropagation, PCA (Principal Component Analysis), Image Processing, ANN (Artificial Neural Networks).

## I. PENDAHULUAN

Perkembangan teknologi kecerdasan buatan saat ini telah mempengaruhi suatu pekerjaan menjadi lebih efektif. Bidang ilmu yang sering digunakan dalam kecerdasan buatan yaitu Jaringan Syaraf Tiruan (JST) yaitu metode yang populer dan cukup handal digunakan sebagai metode klasifikasi dan identifikasi. Dalam Penerapan metode JST, perlu adanya proses pengolahan citra digital.

Pengolahan citra sudah diterapkan dalam berbagai hal seperti kesehatan, keamanan, pertanian, pengujian dan lain-lainnya. Dalam bidang pertanian, satu diantara penerapannya adalah dalam hal identifikasi kualitas hasil pertanian agar sesuai dengan standar pemasaran. Dalam pengolahan citra, suatu citra digital diolah, dan dianalisis sehingga menghasilkan informasi yang dapat dipahami oleh manusia. Suatu citra digital tersusun dari sebuah matrik yang terdiri dari M kolom dan

N baris dimana perpotongan antara kolom dan baris disebut piksel (*pixel = picture element*), yaitu elemen terkecil dari sebuah citra.

Sejumlah penelitian telah dilakukan tentang penggunaan PCA dan JST di area yang berbeda. Pada referensi [1], mengklasifikasikan kualitas produk makanan dengan feedforward ANN dengan satu lapisan tersembunyi dan algoritma backpropagation. PCA sebagai ekstraktor fitur diterapkan ke data spektral sebelum melatih JST untuk mengurangi jumlah informasi yang berlebihan. Penelitian lainnya [2] menggabungkan PCA dan JST dalam bidang kedokteran menggunakan JST backpropagation untuk mengklasifikasikan pasien menjadi dua kategori, dan PCA untuk mengekstrak beberapa fitur penting prediktif untuk kelangsungan hidup pasien. Pada makalah yang lain [3] menyajikan proses sistem pengenalan wajah menggunakan analisis komponen utama (PCA) dengan jaringan saraf tiruan Back-propagation dimana fitur citra wajah digabungkan dengan menerapkan teknik deteksi wajah dan deteksi tepi.

Pada makalah ini dilakukan identifikasi kualitas buah tomat. Analisis komponen utama (PCA) dan jaringan saraf tiruan (JST) yang diadopsi untuk dapat melakukan proses identifikasi tersebut. Tujuan dari makalah ini adalah membuat sistem identifikasi kualitas buah tomat dengan mengaplikasikan metode Analisis komponen utama (PCA) dan jaringan saraf tiruan (JST).

#### A. Citra Digital

Definisi Citra secara visual adalah bentuk representasi sebuah informasi yang terkandung di dalamnya sehingga mata manusia dapat menganalisis dan menginterpretasikan informasi tersebut sesuai dengan tujuan yang diinginkan. Secara matematis, citra didefinisikan sebagai fungsi dua dimensi  $f(x,y)$ , dimana variabel  $x$  dan  $y$  merupakan koordinat spasial dan  $f$  adalah nilai intensitas warna untuk  $x$  dan  $y$ . Nilai dari variabel  $f$ ,  $x$  dan  $y$  merupakan nilai terhingga. Pada umumnya, citra digital dua dimensi (2D) yang dinyatakan dengan bentuk matriks dengan jumlah yang berhingga. Citra

digital ialah bentuk representasi piksel-piksel pada ruang 2D yang dinyatakan dengan matriks ukuran N baris dan M kolom.

$$F = [f(x,y)] = \begin{bmatrix} f(0,0) & f(0,1) & \dots & f(0,M) \\ f(1,0) & f(1,1) & \dots & f(1,M) \\ \vdots & \vdots & \dots & \vdots \\ f(N,0) & f(N,1) & \dots & f(N,M) \end{bmatrix} \quad (1)$$

Elemen pada sebuah matriks citra memiliki koordinat  $x$  dan  $y$  tertentu yang terdapat nilai di dalamnya. Elemen-elemen matriks citra disebut sebagai piksel. Nilai dari setiap piksel  $f$  pada koordinat  $x$  dan  $y$  merupakan representasi nilai intensitas warna dengan masing-masing kode seperti citra berwarna RGB dengan 24 bit, citra gray-level dengan 8 bit dan citra biner dengan 1 bit [4].

#### B. JST

Jaringan Syaraf Tiruan adalah suatu sistem pemrosesan informasi dimana karakteristiknya menyerupai jaringan-jaringan syaraf biologis. JST merupakan bentuk pemodelan matematis dari jaringan syaraf biologi dengan beberapa asumsi sebagai berikut:

- Pemrosesan informasi terjadi pada banyak elemen sederhana (neuron).
- Sinyal dikirim antar neuron-neuron melalui penghubung-penghubung.
- Penghubung memiliki bobot yang dapat memperkuat atau melemahkan sinyal.

Setiap neuron menggunakan fungsi aktivasi (biasanya bukan fungsi linier) yang dikenakan pada banyaknya input yang diterima untuk menentukan output-nya. Suatu ambang batas akan dibandingkan dengan besarnya output yang dihasilkan. Hal-hal terpenting yang menentukan JST adalah Pola hubungan antar neuron (arsitektur jaringan), metode untuk penentuan bobot penghubung (metode pelatihan) fungsi aktivasi [5].

Metode Backpropagation adalah suatu metode pembelajaran lanjutan yang dikembangkan dari aturan perceptron. Metode ini termasuk dalam golongan jaringan lapisan jamak (MLP)[6]. Tahapan dari metode ini terdiri dari tiga tahap yaitu: feedforward yang diambil dari masukan

perception, tahap penghitungan backpropagation error dan pengaturan bobot [7].

### 1) Propagasi maju

Berikut ialah langkah-langkah algoritma pembelajaran metode backpropagation:

- Inisialisasi bobot (ditentukan dengan bilangan acak antara 0 dan 1).
- Selama kondisi berhenti tidak terpenuhi, lakukan langkah 3 sampai langkah 10.
- Setiap pasangan vektor pelatihan melakukan langkah 3 sampai langkah 8.
- Setiap neuron pada lapisan masukan ( $X_i$ ,  $i=1, 2, \dots, n$ ) menerima sinyal masukan  $x_i$  dan meneruskan sinyal tersebut ke seluruh neuron pada lapisan tersembunyi.
- Bobot dijumlahkan dengan sinyal masukan masing-masing dari setiap neuron dalam lapisan tersembunyi ( $Z_j$ ,  $j=1, 2, \dots, p$ ):

$$Z_{in_j} = v_{0j} + \sum_{i=1}^n x_i v_j \quad (6)$$

- Fungsi aktivasi diterapkan untuk menghitung nilai sinyal keluaran.

$$Z_j = f(Z_{in_j}) \quad (7)$$

- Bobot dijumlahkan dengan sinyal masukan masing-masing dengan lapisan keluaran ( $Y_k$ ,  $k=1, 2, \dots, m$ ):

$$Y_{in_k} = w_{0j} + \sum_{i=1}^p z_j w_k \quad (8)$$

- Fungsi aktivasi diterapkan untuk menghitung nilai sinyal keluaran.

$$Y_k = f(Y_{in_k}) \quad (9)$$

### 2) Propagasi Mundur

- Sebuah pola target diterima setiap neuron pada lapisan keluaran ( $Y_k$ ,  $k=1, 2, \dots, m$ ) yang berhubungan dengan pola masukan pelatihan dan menghitung kesalahan/error-nya.

$$\delta_k = (t_k - y_k) \cdot f'(y_{in_k}) \quad (10)$$

perubahan bobotnya dihitung dan digunakan saat mengubah nilai  $w_{jk}$ .

$$\Delta w_{jk} = \alpha \cdot \delta_k \cdot z_k \quad (11)$$

hitung perubahan-perubahan biasnya (digunakan berikutnya untuk mengubah nilai  $w_{0k}$ ).

$$\Delta w_{0k} = \alpha \cdot \delta_k \quad (12)$$

Nilai *delta* masukannya dijumlahkan (dari *neuron* pada lapisan di atasnya) untuk setiap *neuron* pada lapisan tersembunyi ( $Z_j$ ,  $j=1, 2, \dots, p$ ).

$$\delta_{in_j} = \sum_{k=1}^m \delta_k w_{jk} \quad (13)$$

Kalikan dengan turunan aktivitasnya untuk menghitung nilai kesalahan/error-nya.

$$\delta_j = \delta_{in_j} \cdot f'(z_{in_j}) \quad (14)$$

Perubahan bobotnya dihitung dan digunakan selanjutnya untuk mengubah nilai  $v_{ij}$ ).

$$\Delta v_{ij} = \alpha \cdot \delta_j \cdot x_i \quad (15)$$

kemudian perubahan biasnya dihitung dan digunakan berikutnya untuk mengubah nilai  $v_{0j}$ .

$$\Delta v_{0j} = \alpha \cdot \delta_j \quad (16)$$

### 3) Pengaturan bobot

- Nilai bobot dan biasnya ( $j=0, 1, 2, \dots, p$ ) diganti untuk setiap neuron pada lapisan keluaran ( $Y_k$ ,  $k=1, 2, \dots, m$ ).

$$w_{jk(\text{baru})} = w_{jk(\text{lama})} + \Delta w_{jk} \quad (17)$$

Nilai bobot dan biasnya ( $i=0, 1, 2, \dots, n$ ) diganti untuk setiap neuron pada lapisan tersembunyi ( $Y_k$ ,  $k=1, 2, \dots, m$ ).

$$v_{ij(\text{baru})} = v_{ij(\text{lama})} + \Delta v_{ij} \quad (18)$$

### C. PCA

PCA (*Principal Component Analysis*) adalah formulasi matematis yang digunakan dalam pengurangan dimensi data. PCA merupakan bagian dari keilmuan statistika. Penggunaan PCA tidak hanya pada keperluan statistika saja, tapi juga digunakan pada pengolahan citra yang dalam hal ini bekerja pada bagian feature extraction.

Dalam satu set sampel pelatihan  $N$  di mana setiap sampel adalah seperangkat  $n$ , untuk setiap sampel pelatihan  $x_i$  yang dimiliki. Setiap sampel yang diuji dilambangkan dengan:

$$x_i = (x_{1i}, x_{2i}, \dots, x_{ni}) ; i \in 1, N \quad (19)$$

dimana  $x_{ki}$  merupakan variabel dari  $k_{th}$  dalam sampel pelatihan  $i_{th}$ . *Mean* atau rata-rata dari kumpulan sampel pelatihan secara matematis dinyatakan:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^n x_i \quad (20)$$

Sehingga dapat dihitung Deviasi dari *mean* setiap sampel yang diuji:

$$\delta x_i = x_i - \bar{x} \quad (21)$$

Deviasi ini memperlihatkan seberapa jauh perbedaan setiap sampel dari *mean* pada suatu titik. Matriks *covariance*  $2n \times 2n$  dari  $S$  menunjukkan sejauh mana semua perbedaan berasal dari mean sebagai:

$$S = \frac{1}{N} \sum_{i=1}^n \delta x_i \delta x_i^T \quad (22)$$

Analisis komponen utama dari matriks *covariance* ini menunjukkan banyaknya sampel yang ada. Suatu bentuk sampel dapat diperkirakan sebagai:

$$x_i = \bar{x} + P_w \quad (23)$$

di mana  $P = (p_1, p_2, \dots, p_t)$  adalah matriks dari *eigenvector* pertama, dan  $w = (w_1, w_2, \dots, w_t)^T$  adalah bobot vektor yang sesuai di mana masing-masing nilai bobot mengontrol kontribusi *eigenvector* tertentu.

Nilai yang berbeda dalam  $w$  memberikan kemunculan yang berbeda dari model. Mengingat bahwa perubahan ini dalam batas yang ditentukan, maka model baru akan terlihat mirip dengan bentuk dasar (rata-rata). Ini karena mode variasi dijelaskan oleh (unit) *eigenvector*  $S$ , sebagai :

$$S p_k = \lambda_k p_k \quad (24)$$

di mana  $\lambda_k$  menunjukkan nilai *eigen* dan vektor *eigen* yang mengikuti orthogonalitas matriks.

$$P_k P_k^T = 1 \quad (25)$$

dimana nilai *eigen* dibentuk sehingga  $\lambda_k \geq \lambda_{k+1}$ . Dari persamaan 7, Nilai *eigen* terbesar sesuai dengan mode variasi paling signifikan dalam data. Proporsi *variance* dalam data pelatihan, sesuai dengan masing-masing vektor *eigen*, sebanding dengan nilai *eigen* yang sesuai [7].

## II. METODOLOGI

### A. Persiapan bahan

Bahan pengujian yang digunakan untuk penelitian ini adalah citra buah tomat. Pembagian jenis citra buah tomat untuk pengujian ini dibagi menjadi 2 yaitu citra latih dan citra uji. Resolusi citra yang digunakan adalah 80x80 piksel. Berikut adalah data pembagian citra buah tomat berdasarkan kualitasnya.

Tabel 1. Dataset citra buah tomat berdasarkan kualitas

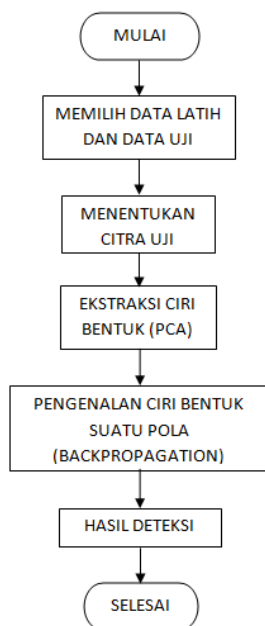
	Tomat kualitas baik	Tomat kualitas buruk	Jumlah data citra
Data Latih	10	10	20
Data uji	5	5	10
			30



Gbr. 1 citra buah tomat untuk pegujian system a. tomat kualitas baik, b. tomat kualitas buruk

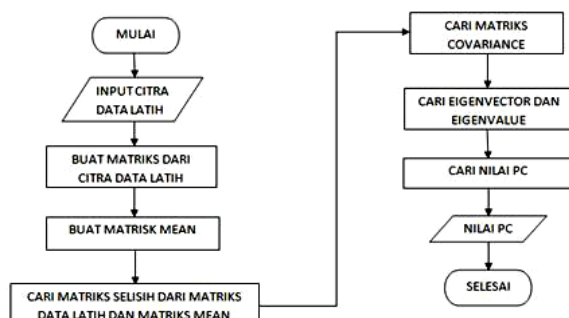
### B. Perancangan Kerja Sistem

Berikut adalah kerja sistem untuk penelitian ini dijelaskan dengan flowchart dibawah ini :



Gbr. 2 Flowchart kerja sistem identifikasi kualitas buah tomat

Data input citra buah tomat diproses melalui metode PCA untuk mendapatkan hasil ekstraksi ciri citra buah tomat tersebut. Selanjutnya dengan metode *backpropagation*, sistem melatih data-data hasil ekstraksi ciri tersebut dan melakukan klasifikasi data menggunakan metode *backpropagation*.



Gbr. 3 Flowchart metode PCA (*Principal Component Analysis*)

### III. HASIL DAN PEMBAHASAN

Dari proses ekstraksi citra dengan metode PCA didapatkan nilai Eigenfaces dengan ukuran yang lebih kecil yaitu dengan dimensi matrik 30x30 pixel yang sebelumnya ukuran matrik citra adalah 80x80 piksel.

Selanjutnya adalah pada simulasi yang dilakukan untuk identifikasi buah tomat dihitung tingkat akurasi dari setiap konfigurasi jaringan yang telah dibuat. Pengujian akurasi konfigurasi jaringan ini dilakukan dengan menggunakan 10 buah gambar citra uji dan 20 citra latih dapat dilihat pada Tabel 2.

Tabel 2. Dataset Citra Tomat Berdasarkan Kualitas

Konfigurasi Jaringan	30_15	30_20_10	45_35_25	5_10_15	10_15_20	15_20_25	10_15_20_25	20_25_30_35	25_30_35_40	30_35_40_45
1	Salah	Benar	Salah	Benar	Benar	Benar	Benar	Benar	Benar	Benar
2	Salah	Benar	Salah	Benar	Benar	Benar	Benar	Benar	Benar	Benar
3	Salah	Benar	Salah	Benar	Benar	Benar	Benar	Benar	Benar	Benar
4	Salah	Benar	Salah	Benar	Benar	Benar	Benar	Benar	Benar	Benar
5	Salah	Benar	Salah	Benar	Benar	Benar	Benar	Benar	Benar	Benar
6	Salah	Benar	Salah	Benar	Benar	Benar	Benar	Benar	Benar	Benar
7	Salah	Benar	Salah	Benar	Benar	Benar	Benar	Benar	Benar	Benar
8	Salah	Benar	Salah	Benar	Benar	Benar	Benar	Benar	Benar	Benar
9	Salah	Benar	Salah	Benar	Benar	Benar	Benar	Benar	Benar	Benar
10	Salah	Benar	Salah	Benar	Benar	Benar	Benar	Benar	Benar	Benar
11tr	Salah	Benar	Salah	Benar	Benar	Benar	Benar	Benar	Benar	Benar
12tr	Salah	Benar	Salah	Benar	Benar	Benar	Benar	Benar	Benar	Benar
13tr	Salah	Benar	Salah	Benar	Benar	Benar	Benar	Benar	Benar	Benar
14tr	Salah	Benar	Salah	Benar	Benar	Benar	Benar	Benar	Benar	Benar
15tr	Salah	Benar	Salah	Benar	Benar	Benar	Benar	Benar	Benar	Benar
16tr	Salah	Benar	Salah	Benar	Benar	Benar	Benar	Benar	Benar	Benar
17tr	Salah	Benar	Salah	Benar	Benar	Benar	Benar	Benar	Benar	Benar
18tr	Salah	Benar	Salah	Benar	Benar	Benar	Benar	Benar	Benar	Benar
19tr	Salah	Benar	Salah	Benar	Benar	Benar	Benar	Benar	Benar	Benar
20tr	Salah	Benar	Salah	Benar	Benar	Benar	Benar	Benar	Benar	Benar
Tingkat Akurasi (%)	56,7	46,7	70	66,7	50	76,7	63,3	43,3	56,7	53,3

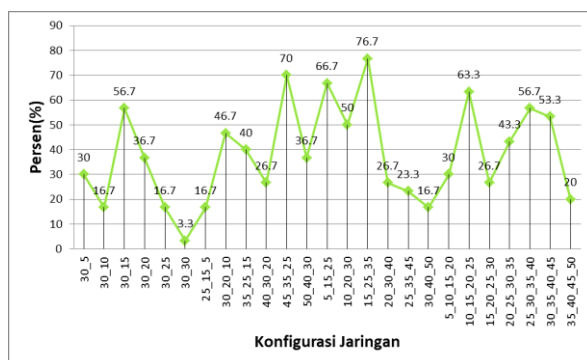
Tabel 3. Keterangan Warna Pada Tabel.

Keterangan	
Salah	Benar
Error	Benar
Benar	Benar

Berdasarkan tabel 2 terdapat 10 konfigurasi jaringan pelatihan yang berbeda. Pengujian berdasarkan pola konfigurasi jaringan pelatihan dengan tingkat akurasi tertinggi adalah sebesar 76,7% pada konfigurasi jaringan 15\_25\_35 yang terdiri dari 3 hidden layer dengan urutan hidden layer 1 sebanyak 15 node, hidden layer 2 sebanyak 25 node dan hidden layer 3 sebanyak 35 node.

Sedangkan untuk konfigurasi jaringan 4 hidden layer dengan tingkat akurasi tertinggi adalah sebesar 63,3% pada konfigurasi jaringan 10\_15\_20\_25 dengan urutan hidden layer 1 sebanyak 10 node, hidden layer 2 sebanyak 15 node, hidden layer 3 sebanyak 20 node dan hidden layer 4 sebanyak 25 node.

Grafik pada gambar 4 menunjukkan bahwa banyaknya node di setiap layer tidak sepenuhnya mempengaruhi peningkatan akurasi pelatihan jaringan. Hal ini mungkin terjadi karena beberapa faktor tertentu seperti kualitas citra pengujian yang kurang baik.



Gbr. 4 Tingkat Akurasi Sistem Identifikasi Kualitas Buah Tomat Berdasarkan Konfigurasi Jaringan

#### IV. KESIMPULAN

Metode PCA yang digunakan untuk tahap Preprocessing membantu untuk mengurangi dimensi citra menjadi 30x30 piksel dan mempercepat kerja program sebelum dilanjutkan ke tahap JST Backpropagation. Pada penelitian ini jaringan syaraf tiruan *backpropagation* yang digunakan menunjukkan tingkat akurasi benar mencapai 76,7% menggunakan 20 sample uji yang dibaca dengan benar sebanyak 23 buah sample. Nilai akurasi tersebut menunjukkan bahwa sistem ini dapat berjalan sesuai tujuan.

#### REFERENSI

- [1] O'Farrella, M., Lewisa, E., Flanagan, C., Lyonsa, W.B., Jackman, N. (2005), "Combining principal component analysis with an artificial neural network to perform online quality assessment of food as it cooks in a large-scale industrial oven", *Sensors and Actuators B*, Vol. 107, pp. 104–112
- [2] Bucinski, A., Baczek, T., Wasniewski, T., and Stefanowicz, M. (2005), "Clinical data analysis with the use of artificial neural networks (ANN) and principal component analysis (PCA) of patients with endometrial carcinoma", *Reports on Practical Oncology and Radiotherapy*, Vol. 10, pp. 239-248.
- [3] R Toufiq, M. Rabiul Islam, (2005), "Face recognition system using PCA-ANN technique with feature fusion method", *International Conference on Electrical Engineering and Information & Communication Technology*, vol. ,pp.1-5.
- [4] Madenda, Sarifuddin, *Pengolahan Citra & Video Digital*, Jakarta: Penerbit Erlangga, 2015
- [5] Drs. Jong Jek Siang, M.Sc., *Jaringan Syaraf Tiruan & Pemrograman Menggunakan MATLAB, 2 ed.* Yogyakarta: Penerbit Andi, 2009.
- [6] Anita Desniani, *Konsep Kecerdasan Buatan*, Yogyakarta: Penerbit Andi, 2006.
- [7] L. Fausett, *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications*. Prentice-Hall, 1994.