

Klasifikasi Kanker Paru-Paru Menggunakan Pengolahan Citra

Noor Maulida, Dinda F. Paramitha, Ekky A. Sukarno dan Agus Z. Arifin
Jurusan Teknik Informatika, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh Nopember (ITS)
Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia
e-mail: agusza@cs.its.ac.id

Abstrak— Kanker paru-paru adalah salah satu jenis kanker yang paling umum dan mematikan. Tingkat kematian karena kanker paru-paru bertambah dua kali lipat setiap dekade. Saat ini, sebagian besar prosedur pendeteksian penyakit kanker paru-paru masih dilakukan secara manual oleh tenaga laboratorium. Pendeteksian manual akan menghasilkan diagnosis yang subjektif dikarenakan rendahnya kualitas citra *X-ray* sehingga sering kali bagian kanker tersamarkan oleh struktur anatomi lainnya. Penelitian ini mengembangkan teknik komputerisasi citra untuk melakukan proses klasifikasi kanker paru-paru. Metode yang digunakan adalah segmentasi citra paru-paru dengan nilai *threshold* yang didapatkan dari korelasi nilai Euler. Selanjutnya dilakukan ekstraksi fitur menggunakan GLCM. Hasil ekstraksi fitur digunakan sebagai masukan untuk diklasifikasi dengan *Artificial Neural Network* (ANN). Dari 40 data citra *X-ray* bagian dada yang digunakan didapatkan hasil klasifikasi kanker paru-paru dapat membedakan antara kanker paru-paru ganas dan jinak dengan akurasi 87,5%.

Kata Kunci—ANN, Citra *X-ray* bagian dada, Kanker paru-paru, Segmentasi Euler.

I. PENDAHULUAN

PENGOLAHAN citra digital saat ini digunakan dalam berbagai bidang riset. Salah satunya di bidang kedokteran. Citra medis yang akan diolah pada penelitian ini adalah citra yang diambil menggunakan teknologi *Computed Tomography Scan* (CT Scan) paru-paru. Teknologi CT Scan dikembangkan melalui proses pemindaian setiap lapisan-lapisan jaringan tubuh dari berbagai arah dengan menggunakan radiasi sinar X (Dougherty, 2009). Hasil dari pemindaian dengan menggunakan teknologi ini disebut sebagai citra *X-ray*. Pemilihan paru-paru sebagai objek yang akan diolah pada penelitian ini adalah dikarenakan tingginya jumlah kematian yang disebabkan oleh penyakit kanker paru-paru. Menurut data tahun 2008 dari situs resmi WHO (*World Health Organization*), kanker paru-paru adalah penyebab kematian paling besar di antara jenis kanker lainnya yaitu sebesar 1,4 juta kematian dibandingkan dengan jenis kanker lainnya seperti kanker perut (740.000 kematian), kanker hati (700.000 kematian), kanker kolorektal (610.000 kematian) dan kanker payudara (460.000 kematian) [1]. Hasil setiap citra irisan *X-ray* secara manual akan diperiksa oleh ahli radiologi dengan menggunakan layar baca citra rontgen. Bagian yang paling penting dan sulit bagi dokter dan ahli radiologi adalah ketika

mendeteksi bagian kanker dari citra medis *X-ray*. Salah satu penyebabnya adalah rendahnya kualitas citra *X-ray* sehingga sering kali bagian kanker tersamarkan oleh struktur anatomi lainnya dan kurang akuratnya penentuan kriteria variabel keputusan. Bahkan, ahli radiologi yang telah berpengalaman mengalami kesulitan dalam membedakan bagian kanker pada percabangan pembuluh darah di bagian paru-paru atau mendeteksi nodul halus yang menunjukkan kanker paru-paru pada citra *X-ray*. Penelitian sebelumnya menunjukkan bahwa dokter dan ahli radiologi gagal mendeteksi bagian kanker sejumlah 30% dari kasus positif [2]. Implementasi sistem CAD untuk melakukan klasifikasi dan mendeteksi bagian kanker telah dikembangkan, namun hasil yang didapatkan dari implementasi tersebut adalah masih terdapat banyak kesalahan dalam hasil klasifikasi [3] [4]. Oleh karena itu, penelitian ini akan mengembangkan teknik komputerisasi citra untuk melakukan proses klasifikasi kanker paru-paru. Dengan adanya penelitian ini diharapkan algoritma yang dikembangkan dapat membantu dokter dan ahli radiologi untuk mendeteksi kanker dalam waktu singkat dengan lebih akurat.

II. METODE PENELITIAN

Implementasi klasifikasi kanker paru-paru pada penelitian ini menggunakan citra digital sejumlah 20 citra kanker jinak dan 20 citra kanker ganas. Total citra digital yang digunakan berjumlah 40 citra, setiap citra berukuran 512 X 512 piksel. Data citra yang digunakan berupa citra *X-ray* bagian dada yang diambil dari database umum JSRT (*Japanese Society of Radiological Technology*).

Tahap pertama dalam pengolahan citra adalah tahap pemrosesan awal. Setiap sampel citra dimasukkan ke dalam program dan disimpan dalam ukuran 512 X 512 piksel. Citra yang sudah disimpan pada ukuran 512 X 512 piksel dilakukan proses ekualisasi histogram, perbaikan citra dan transformasi intensitas seperti pada Gambar 1. Ekualisasi histogram adalah proses di mana distribusi nilai derajat keabuan pada suatu citra dibuat rata. Kualitas citra digital *X-ray* yang buram diperbaiki dengan menggunakan filter Wiener dan median. Setelah itu, dilakukan transformasi intensitas dari citra skala keabuan menggunakan fungsi *imadjust*. Fungsi *imadjust* akan memetakan nilai intensitas dalam gambar skala keabuan dari

citra awal dengan nilai-nilai intensitas baru sedemikian rupa.



Gambar 1. Hasil pengolahan citra X-ray tahap pemrosesan awal. Bagian kiri merupakan citra X-ray asli dan bagian kanan merupakan citra yang telah dilakukan tahap pemrosesan awal.

Setelah tahap pemrosesan awal tahap selanjutnya adalah tahap segmentasi. Tujuan yang ingin dicapai dalam segmentasi adalah untuk menentukan objek yang spesifik atau sebuah area dalam citra untuk pencocokan atau identifikasi. Dalam citra X-ray bagian yang perlu dilakukan segmentasi adalah bagian kanan dan kiri paru-paru.

Prosedur 1. Segmentasi Korelasi Nilai Euler	
Input: PREIMG	
1	Inisialisasi $level \leftarrow$ skala keabuan
2	Inisialisasi $value \leftarrow 1:level+1$
3	FOR EACH $value[i]$ DO
4	$T[i] \leftarrow value[i] / level$
5	END FOR
6	FOR EACH $value[i]$ DO
7	$V \leftarrow value[1,i]$
8	$Result[i,1] \leftarrow$ Euler(gambar hitam
9	putih dari PREIMG dengan nilai
10	threshold diambil dari nilai V)
11	END FOR
12	Inisialisasi $Max \leftarrow$ Nilai maksimum dari
13	$Result$
14	Inisialisasi $Min \leftarrow$ Nilai minimum dari
15	$Result$
16	Inisialisasi $M \leftarrow 0$
17	Inisialisasi $N \leftarrow 0$
18	Inisialisasi $MaxSum \leftarrow 0$
19	Inisialisasi $MinSum \leftarrow 0$
20	FOR EACH $value[i]$ DO
21	IF $Result[i,1] == Max$
22	Increase M
23	$MaxSum \leftarrow MaxSum + i$
24	ELSE IF $Result[i,1] == Min$
25	Increase N
26	$MinSum \leftarrow MinSum + i$
27	END IF
28	END FOR
29	$threshold_euler \leftarrow (MaxSum+MinSum) / (M+N)$
Output: $threshold_euler$	

Gambar 2. Pseudocode segmentasi menggunakan korelasi nilai Euler.

Gambar 2 menjelaskan proses segmentasi menggunakan nilai $threshold$ yang diperoleh dari korelasi nilai Euler sejumlah level skala keabuan yang ditentukan [5]. Pada baris keenam hingga kesebelas dilakukan perulangan sebanyak level skala keabuan untuk mendapatkan nilai Euler dari setiap kemungkinan gambar hitam dan putih dari hasil pemrosesan awal dengan nilai $threshold$ sementara diambil dari nilai perulangan dibagi dengan skala level keabuan. Selanjutnya dari semua nilai Euler yang telah didapatkan diambil nilai maksimum dan minimumnya untuk menentukan skala mana saja yang mencapai nilai maksimum dan minimum. Sehingga nilai $threshold$ akhir didapatkan dari persamaan

$$threshold_euler = \frac{(maxsum + minsum)}{(m + n)} \quad (1)$$

dengan nilai $maxsum$ diambil dari total nilai skala yang mencapai nilai maksimum Euler, nilai $minsum$ diambil dari total nilai skala yang mencapai nilai minimum Euler, nilai m diambil dari berapa skala yang mencapai nilai maksimum dan nilai n diambil dari berapa skala yang mencapai nilai minimum. Hasil dari segmentasi dengan $threshold$ menggunakan korelasi nilai Euler ditunjukkan pada Gambar 3.



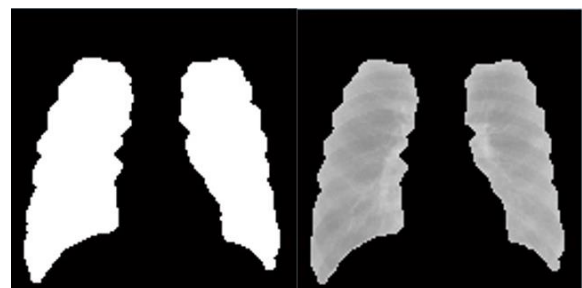
Gambar 3. Hasil segmentasi menggunakan korelasi nilai Euler.

Citra yang telah disegmentasi menggunakan korelasi nilai Euler selanjutnya dikomplemenkan dan dipadukan dengan $mask$ berupa dua titik yang merepresentasikan bagian kanan dan kiri paru-paru. Gambar 4 merupakan contoh dari $mask$ yang digunakan. $Region$ hasil segmentasi menggunakan korelasi nilai Euler yang sesuai dengan titik $mask$ akan dipilih dan hasilnya adalah bagian kanan dan kiri paru-paru telah tersegmentasi.



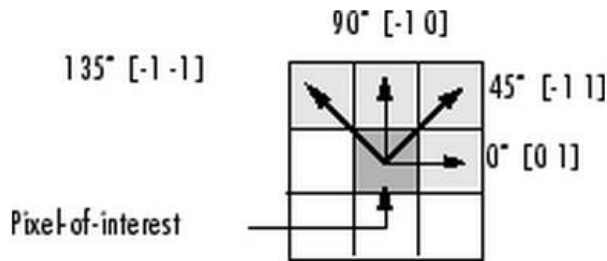
Gambar 4. Contoh $mask$ berupa dua titik untuk mendapatkan bagian kanan dan kiri paru-paru.

Setelah bagian kanan dan kiri paru-paru berhasil disegmentasi, dilakukan proses morfologi berupa dilasi, erosi, *closing*, *opening*, *regionprops* dan *hole filling* untuk memperhalus gerigi, menghilangkan bagian-bagian kecil di luar objek utama, menghubungkan garis yang terputus dan mengisi lubang kosong di dalam objek utama. Setelah itu hasil akhir didapatkan dengan menggabungkan hasil morfologi dengan citra asli. Hasil dari proses morfologi dapat dilihat pada Gambar 5.



Gambar 5. Hasil akhir segmentasi.

Untuk mengambil informasi dari citra yang telah berhasil dilakukan segmentasi digunakan fitur ekstraksi, yaitu fitur tekstur. Fitur tekstur yang digunakan dalam penelitian ini adalah *average gray level*, standar deviasi, *smoothness*, *third moment*, uniformitas, entropi, kontras dan energi dihitung berdasarkan Gray Level Co-Occurrence Matrix (GLCM) dari hasil akhir segmentasi.



Gambar 6 Jarak d dan sudut θ pada GLCM.

GLCM adalah matriks yang menggambarkan frekuensi munculnya pasangan dua piksel dengan intensitas tertentu dalam jarak d dan orientasi arah dengan sudut θ tertentu dalam citra [6] seperti yang digambarkan pada Gambar 6. Contoh dari matriks GLCM dijelaskan pada Gambar 7 yang menggunakan jarak $d = 1$ dan sudut $\theta = 0^\circ$. Dari matriks awal dicari nilai yang paling maksimum, lalu matriks GLCM dibentuk dengan ukuran nilai maksimum X nilai maksimum. Sehingga dari matriks awal pada Gambar 7 yang mempunyai nilai maksimum 8 didapatkan matriks GLCM dengan ukuran 8 X 8. Jarak $d = 1$ dan sudut $\theta = 0^\circ$ menunjukkan kedekatan matriks yang akan dihitung hanya dengan yang berjarak 1 dan sudutnya mengarah ke tetangga piksel arah kanan, misal pada baris pertama yang akan dihitung nilai frekuensinya adalah nilai 1 dengan 1, nilai 1 dengan 5, nilai 5 dengan 6 dan nilai 6 dengan 8. dengan Sebagai contoh nilai frekuensi munculnya 1 dan 1 sebanyak 1 kali ditempatkan nilai 1 kali tersebut di matriks GLCM baris 1 kolom 1, begitu juga nilai frekuensi munculnya 1 dan 2 sebanyak 2 kali ditempatkan nilai 2 kali tersebut di matriks GLCM baris 1 kolom 2, dan seterusnya.

I	1	1	5	6	8	GLCM	1	2	3	4	5	6	7	8
	2	3	5	7	1		1	0	0	0	1	0	0	0
	4	5	7	1	2		2	0	0	0	0	1	0	0
	8	5	1	2	5		0	0	0	0	1	0	0	0
							4	0	0	0	0	1	0	0
							5	1	0	0	0	0	1	2
							6	0	0	0	0	0	0	1
							7	2	0	0	0	0	0	0
							8	0	0	0	0	1	0	0

Gambar 7. Contoh matriks GLCM.

Pada penelitian ini, jarak yang digunakan adalah 1 dan sudut θ diambil dari nilai rata-rata dari 0° , 45° , 90° dan 135° . Citra yang dijadikan matriks GLCM untuk diambil nilai-nilainya adalah citra hasil akhir segmentasi yang telah dilakukan proses morfologi dan penggabungan dengan citra asli.

Setelah nilai-nilai dari fitur tekstur diambil, nilai tersebut akan dijadikan masukan ke dalam pemodelan ANN. ANN adalah suatu sistem pemodelan yang meniru cara kerja syaraf manusia yang memiliki kemampuan belajar dan menyesuaikan model dengan data-data terbaru meskipun data yang dimiliki terbatas Keuntungan dari permodelan dengan ANN adalah [7]:

- Model ini dapat dibuat hanya berdasar data historis proses masukan-keluaran.
- Dapat memberikan simulasi hubungan antara banyak masukan-banyak keluaran.

Penelitian ini dilakukan pada 40 data citra *X-ray* bagian dada, yang terdiri dari 20 data citra kanker jinak dan 20 citra kanker ganas. Pembagian data sebesar 24 data latih dan 16 data uji. Algoritma pelatihan yang digunakan adalah algoritma *back-propagation* dengan menggunakan metode Levenberg-Marquardt. Lapisan pertama dari ANN adalah lapisan masukan berupa 8 *node* (8 hasil fitur tekstur), lapisan kedua adalah layer tersembunyi sejumlah 8 *node* (h_1 sampai h_8) dan lapisan terakhir adalah layer keluaran sejumlah 2 *node* (kanker jinak dan ganas).

Untuk setiap pola pelatihan pada data uji yang menghasilkan suatu keluaran tertentu akan dihitung kesalahan pada perkiraan hasil tersebut. Algoritma *back-propagation* mengacu pada penyebaran kesalahan *node* dari lapisan keluaran ke *node* di lapisan tersembunyi. Kesalahan ini digunakan untuk memperbarui bobot jaringan untuk masukan selanjutnya [8]. Eksperimen awal dalam klasifikasi dimulai dengan menggunakan 5 *node* lapisan tersembunyi dan 25% dari total data, yaitu 10 sampel. Dari setiap kategori kanker diambil acak 5 sampel dengan perbandingan data 60% data latih dan 40% data uji. Fungsi transfer antara lapisan masukan dengan lapisan tersembunyi menggunakan tipe *tansigmoid* dan antara lapisan tersembunyi dengan lapisan keluaran menggunakan tipe linear. Algoritma *back-propagation* yang digunakan untuk melatih data uji menggunakan nilai 1 pada *learning rate* dan toleransi kesalahan 0.005. Hasil yang didapatkan dari eksperimen awal tersebut adalah nilai 40% untuk performa klasifikasi. Hasil yang lebih baik didapatkan ketika jumlah data latih ditambah menjadi 24 sampel (masing-masing kategori berjumlah 12 sampel citra), data uji ditambah menjadi 16 sampel (masing-masing kategori berjumlah 8 sampel citra) dan menambah jumlah *node* pada lapisan tersembunyi menjadi 6 *node*. Jaringan ANN 8:6:2 tersebut menghasilkan hasil performa klasifikasi 87,50% dengan 20 iterasi.

III. HASIL DAN DISKUSI

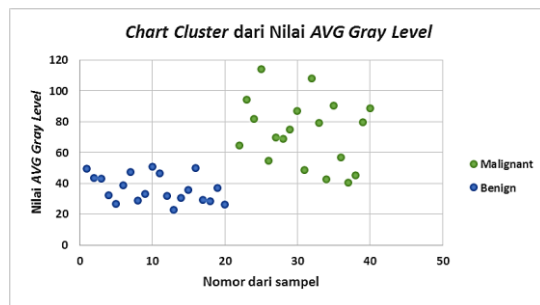
Hasil segmentasi dari korelasi nilai Euler sebagai *threshold* mendapatkan hasil yang cukup baik dan mampu melakukan segmentasi bagian kanan dan kiri paru-paru. Kekurangan dari hasil segmentasi ini adalah hasil yang bergerigi dikarenakan citra *X-ray* bagian paru-paru tertutup samar-samar oleh tulang rusuk. Hasil tersebut telah diperbaiki dengan proses morfologi, salah satunya dengan menggunakan elemen struktur morfologi berbentuk garis dengan sudut 140° dan 230° sesuai dengan bentuk dari paru-paru.

Hasil dari pengujian pada data latih didapatkan nilai rata-rata untuk setiap ciri tekstur dari citra latih seperti pada Tabel 1.

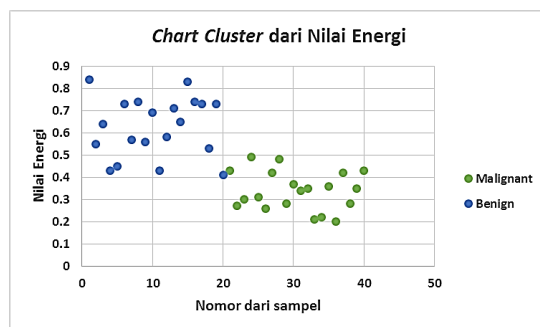
Tabel 1.
Nilai rata-rata fitur tekstur GLCM

Nilai rata-rata	Kelompok citra	
	Jinak	Ganas
AVG Gray Level	36.77	72.88
Standar deviasi	56.51	59.51
Smoothness	0.16	0.22
Third moment	1.52	1.86
Uniformitas	0.56	0.61
Entropi	2.77	4.52
Kontras	0.45	0.41
Energi	0.34	0.61

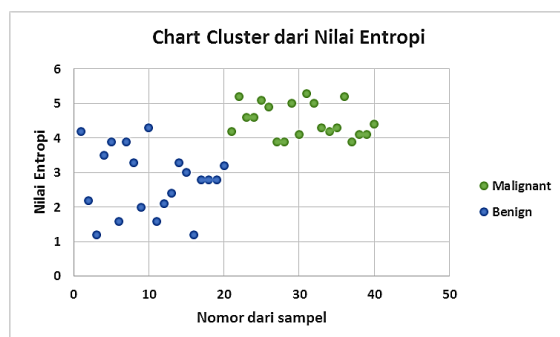
Berdasarkan hasil dari nilai rata-rata fitur ekstraksi dengan GLCM pada Tabel 1, dapat disimpulkan bahwa nilai yang terlihat paling jelas perbedaannya antara kanker jinak dan ganas adalah fitur *AVG gray level*, energi dan entropi. Sehingga ketiga fitur itulah yang paling berkontribusi dalam proses klasifikasi citra. Representasi grafik dari nilai-nilai *AVG gray level*, energi dan entropi pada semua sampel digambarkan pada Gambar 8, 9 dan 10.



Gambar 8. Chart cluster dari nilai AVG Gray Level.

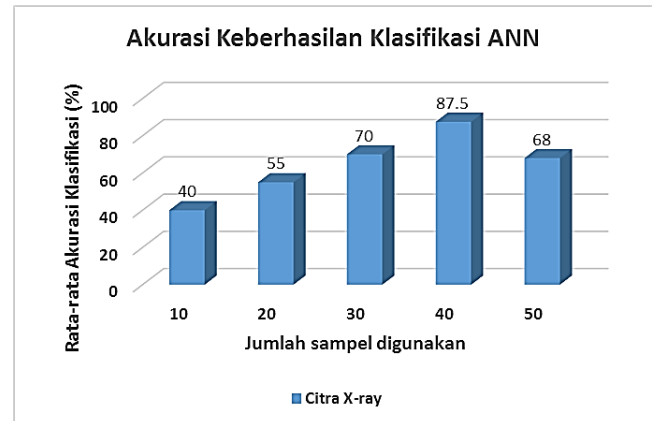


Gambar 9. Chart cluster dari nilai energi.



Gambar 10. Chart cluster dari nilai entropi.

Performa klasifikasi dari hasil pengujian kepada 40 data citra X-ray (20 data citra kanker jinak dan 20 data citra kanker ganas) menghasilkan persentase 87,50% dengan iterasi sebanyak 20. Akurasi klasifikasi semakin baik jika data latih terus ditambah. Hasil dari penelitian ini juga menunjukkan bahwa algoritma *back-propagation* handal dalam menangani kesalahan keluaran dengan memperbaiki bobot jaringan untuk masukan selanjutnya. Representasi grafik untuk akurasi klasifikasi ditunjukkan pada Gambar 11.



Gambar 11. Akurasi keberhasilan klasifikasi ANN.

IV. KESIMPULAN/RINGKASAN

Dari penelitian ini dapat disimpulkan bahwa pengembangan teknik komputerisasi citra segmentasi menggunakan korelasi nilai Euler berhasil mengambil bagian kanan dan kiri paru-paru dan memisahkannya dengan struktur anatomi lain dari citra X-ray. Keluaran dari klasifikasi program yang dikembangkan ini dapat membedakan citra X-ray yang merepresentasikan kanker jinak dan ganas.

DAFTAR PUSTAKA

- [1] "WHO Media Centre," World Health Organization, [Online]. Available: <http://www.who.int/mediacentre/factsheets/fs297/en/>. [Accessed 27 May 2014].
- [2] R. S. Fontana, R. Sanderson, L. B. Woolner, W. F. Taylot, W. E. Miller, and J. R. Muhm, "Lung Cancer Screening: The Mayo Program", J. Occupat. Med., Vol. 28, pp. 746-750, 1986.
- [3] M. L. Giger, K. Doi, H. Mac Mahon, C. E. Metz, and F. F. Yin, "Image feature analysis and computer-aided diagnosis in digital radiography.
- [4] J. S. Lin, P. A. Ligomenides, Y. M. F. Lure, M. T. Freedman,
- [5] and S. K. Mun, "Application of neural networks for improvement of lung nodule detection in radiographic images," in Proc. Symp. Compute. Assist. Radiol (S/CAR'92), pp. 108 – 115, 1992. M. J. Carreira, M. G. Penedo, D. Cabello, and J. M. Pardo, "Computer-aided lung nodule detection in chest radiography," in Lecture Notes in Computer Science: Image Analysis Applications and Computer Graphics, vol. 1024. Berlin, Germany: Springer-Verlag, pp. 331 – 338, 1996
- [6] J. Carreira and C. Sminchisescu, "CPMC: Automatic Object Segmentation Using Constrained Parametric Min-Cuts," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012.
- [7] Masfran, Ananda dan E. S. Nugroho, "Segmentasi Tepi Citra CT Scan Paru-paru Menggunakan Metode Chain Code dan Operasi Morfologi," Jurnal Teknik Informatika Politeknik Caltex Riau, Vol. I, 2012.
- [8] Z. G. Che, T. A. Chiang and Z. H. Che, "Feed-Forward Neural Network Training: A Comparison Between Genetic Algorithm and Back-Propagation Learning Algorithm," International Journal of Innovative Computing, Information and Control, Vol. 7, October 2011.