

COMPUTATIONAL INTELLIGENT

**Computer-Aided Diagnosis (CAD) Of Breast Cancer Using Artificial Neural
Networks : Comparison of Backpropagation and Genetic Algorithms**



Oleh :

Cathrine Nicea Folamauk

**Magister Ilmu Komputer dan Elektronika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Gajah Mada
2016**

TOPIK :

Computer-Aided Diagnosis (CAD) Of Breast Cancer Using Artificial Neural Networks : Comparison of Backpropagation and Genetic Algorithms

LATAR BELAKANG

Kanker payudara merupakan salah satu masalah utama yang menyebabkan kematian pada para wanita. Meskipun pemeriksaan menggunakan mammography telah menunjukkan hasil yang efektif dengan mendeteksi kanker payudara dilakukan secara dini. Perbedaan antara kanker ganas dan jinak menjadi tantangan tersendiri untuk para ahli radiologi. Skema Computer-Aided Diagnosis (CAD) untuk mendiagnosis adanya kanker payudara telah diteliti dengan harapan system CAD dapat memberikan “pendapat kedua (Solusi pendukung)” bagi para ahli radiologi sehingga hasil diagnosis yang diberikan oleh para ahli menjadi lebih akurat, efisien dan secara signifikan dapat dikembangkan. Saat ini, banyak struktur CAD telah diteliti untuk mengidentifikasi daerah-daerah pada payudara yang dicurigai secara spesifik memiliki kondisi yang tidak normal dengan mammogram. Skema CAD tersebut cenderung lebih berfokus pada temuan-temuan pada mamografi dan beberapa temuan secara fisik yang diperoleh dari data riwayat pasien yang dikumpulkan untuk melakukan diagnosis kanker payudara.

Skema CAD digunakan untuk menentukan probabilitas terhadap kemunculan kanker payudara menggunakan Artificial Neural Networks (ANNs) dengan proses pelatihan menggunakan algoritma Backpropagation (BP) atau Genetik Algoritma (GA). Pada penelitian menggunakan 418 data pasien yang telah diverifikasi sebelumnya pada database klinik dan kemudian diacak kedalam 2 kumpulan data yaitu data latih dan data uji. Selama melakukan pelatihan, BP dan GA secara bebas diterapkan untuk mengoptimalkan dan untuk mengembangkan interkoneksi bobot dari ANNs. Proses pelatihan BP dan GA pada kinerja system CAD dibandingkan menggunakan analisis receiver-operating characteristics (ROC). Pada kumpulan data latih, proses latih BP dan GA pada struktur system CAD menghasilkan daerah yang terletak dibawah kurva ROC masing-masing 0.91 dan 0.93. Sedangkan pada data uji, proses uji BP dan GA menghasilkan daerah yang terletak dibawah kurva ROC sekitar 0.83. Hasil tersebut menunjukkan kinerja GA sedikit lebih baik dibandingkan BP, meskipun tidak signifikan.

KOMPONEN DAN METODE ANNs

Digunakan sebanyak 418 kasus pasien pada sebuah database klinik, yang telah dikumpulkan sebelumnya dan diverifikasi oleh para ahli radiologi dengan laporan pathology. Semua 418 kasus telah diakuisisi selama pemeriksaan payudara yang dilakukan oleh University of Pittsburgh Medical Center (UPMC) dan dengan rumah sakit yang menjadi mitra dengan UPMC tersebut. Pada database, sebanyak 92 kasus positif kanker payudara, sedangkan 326 kasus menunjukkan hasil yang negative. Kasus “negative” diverifikasi oleh para ahli radiologi selama 2 tahun terakhir mengikuti temuan negative pada mammogram. Setiap kasus terdiri dari 5 fitur pada riwayat pasien, 4 temuan secara fisik, dan 4 temuan menggunakan memografi yang telah dikumpulkan dan digunakan sebagai inputan pada skema CAD untuk menghasilkan probabilitas kemunculan kanker payudara.

Pada gambar 1 merupakan skema diagram algoritma CAD yang digunakan untuk mendiagnosis kanker payudara menggunakan ANNs dengan proses pelatihan menggunakan Backpropagation atau algoritma genetic. Semua node inputan (selain microcalcification dan masa) pada ANN ditetapkan dengan angka 1 (Ya) dan 0 (Tidak). Terlepas dari sejumlah daerah yang dicurigai digambarkan dalam mammogram, node-node inputan untuk microcalcification atau masa pada ANN ditetapkan dengan angka 1, jika menunjukkan level 4-5; 0,5 jika menunjukkan level 1-3 atau 0. Secara subjektif range level ditentukan oleh para ahli radiologi, dimana pada level 5 (sangat mencurigakan) dan 1 (paling sedikit dicurigai sebagai kanker ganas) pada microcalcification atau masa. Target node output ANN ditetapkan dengan angka 1 (Positif Kanker Payudara) dan 0 (Negative Kanker Payudara).

ANN yang digunakan pada penelitian ini adalah multi layer, feed-forward Neural Network dengan interkoneksi bobot. Dengan fungsi :

$$O_p = \frac{1}{\sum_t W_j O_p \theta_j} x 1/e$$

Dimana :

O_p = Output sebanyak j unit yang dihasilkan

O_p = Bobot per unit i,j

O_p = Nilai Threshold sebanyak j unit

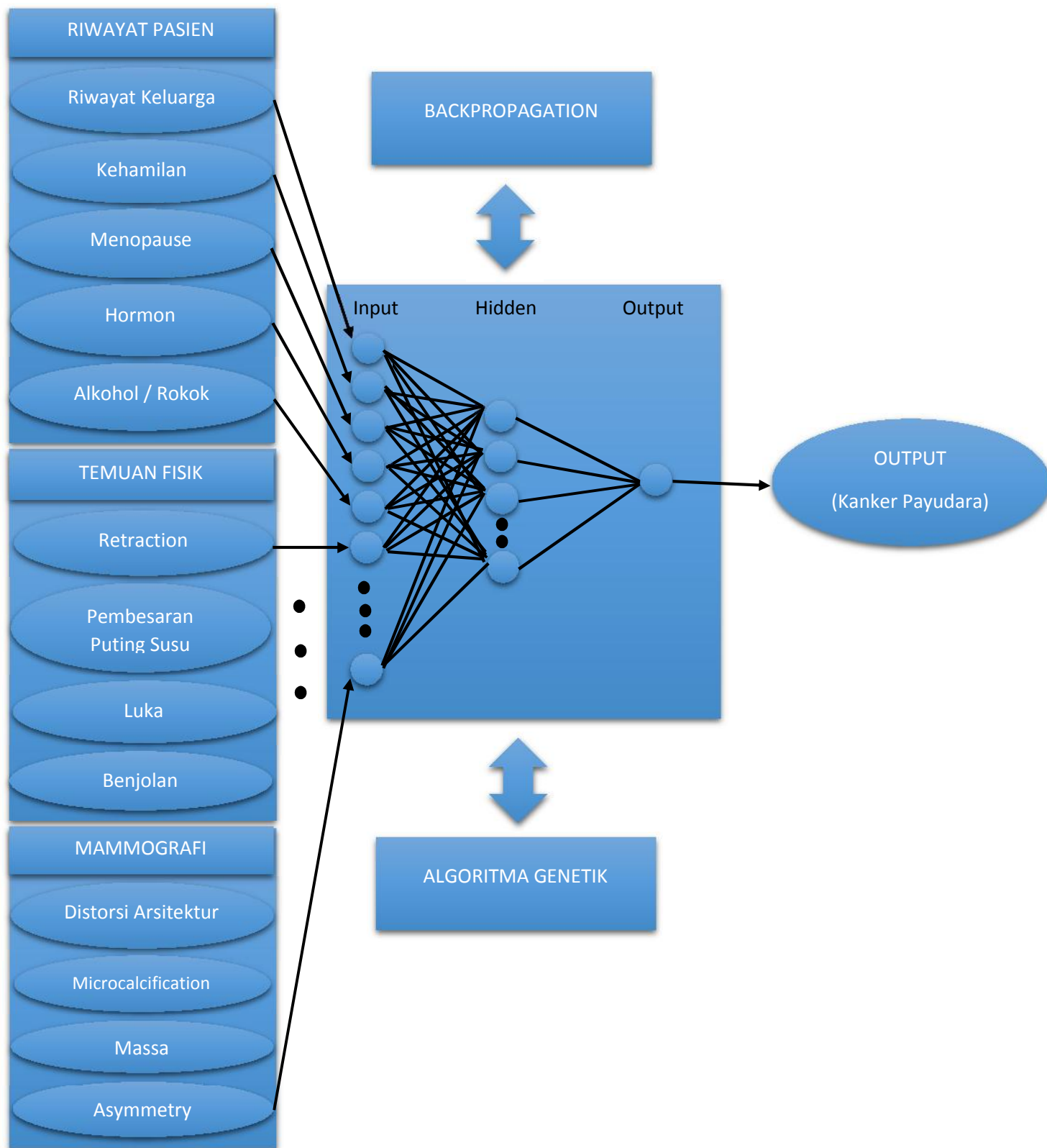
Menghitung error antara target tujuan dan nilai output sebenarnya, disimpulkan :

$$E_p = \frac{1}{2} \sum_j (t_p - O_p)^2$$

Definisi Fitur :

Kategori	Deskripsi Node	Value
Diagnosis	Kanker Payudara	Present, Absent
Riwayat Pasien	Apakah terdapat anggota keluarga yang memiliki riwayat kanker?	Yes, No
	Apakah Sudah Pernah Hamil?	Yes, No
	Apakah Telah Memasuki Masa Menopause?	Yes, No
	Apakah Sudah Melakukan Pengecekan Hormon ?	Yes, No
	Apakah Anda Minum Alkohol atau Merokok?	Yes, No
Temuan Secara Fisik	Adanya Perubahan Kulit (Retraction) ?	Yes, No
	Adanya Pembesaran Pada Puting Susu ?	Yes, No
	Adaya Luka pada Payudara ?	Yes, No
	Adanya Benjolan pada Payudara?	Yes, No
Temuan Secara Mammografi	Distorsi Arsitektur ?	Present, Absent
	Microcalcifications	Level : 4 – 5 Level : 1 – 3 Level : 0 (Absent)
	Massa	Level : 4 – 5 Level : 1 – 3 Level : 0 (Absent)
	Asymmetry	Present, Absent

Selama $E = \sum E_p$ merupakan keseluruhan error dihitung dengan menjumlahkan keseluruhan error pada kumpulan data training. Pada awal proses, 3 layer ANN termaksud 13 input unit, 4 hidden unit dan 1 output unit akan dipilih. Bobot nilai dimulai dari 0 sampai 1 pada unit output digunakan untuk mengidentifikasi probabilitas kemunculan kanker payudara. Backpropagation dan algoritma genetic digunakan untuk mengoptimalkan, atau untuk mengembangkan interkoneksi bobot dalam ANN.



Gambar 1. Skema Diagram CAD kanker payudara menggunakan ANN dengan proses pelatihan Backpropagation atau Algoritma Genetik.

ALGORITMA BACKPROPAGATION

Algoritma backpropagation menjadi salah satu algoritma yang sering digunakan untuk proses pelatihan pada ANNs. Pada proses ini, bobot andta unit proses perulangan disesuaikan sehingga keseluruhan perhitungan error E dapat diperkecil, menggunakan formula :

$$\Delta W_j(n+1) = \eta(\delta_p O_p) + \alpha \Delta W_j(n)$$

Dimana :

η = Learning Rate

n = Banyaknya iterasi yang dilakukan

α = Momentum

δ_p = Sinyal Error

Meskipun algoritma backpropagation dapat digunakan untuk menyediakan sebuah solusi untuk masalah pada pelatihan ANNs, algoritma tersebut mungkin saja dapat menemukan hasil minimum dalam solusi sub-optimal

ALGORITMA GENETIK

Algoritma genetic dapat digunakan untuk menentukan bobot interkoneksi ANN. Langkah-langkah penggunaan algoritma genetic pada ANN :

- Langkah 1 :** Kromosom dibuat dalam bentuk vektor keseluruhan bobot interkoneksi ANN. Inisialisasi populasi dari kromosom untuk setiap bobot antara -1 dan 1
- Langkah 2 :** Evaluasi fitness untuk setiap kromosom dalam populasi. Pada penelitian ini, maksimum fitness ekuivalen dengan perhitungan keseluruhan error E pada kumpulan data latih. Proses seleksi dilakukan menggunakan metode Roulette Wheel
- Langkah 3 :** Lakukan operasi crossover menggunakan 2 parent kromosom terpilih pada tahap 2 untuk menghasilkan 2 offspring/keturunan. Pertama, copy semua bobot menuju unit output offspring 1 dari parent 1; dan offspring 2 dari parent 2. Kemudian copy semua bobot genap dan ganjil pada hidden unit untuk offspring 1 (crossover parent 1 dan 2). Sebagai alternative, copy semua bobot genap dan ganjil pada hidden unit untuk offspring 2 (crossover parent 2 dan 1).
- Langkah 4 :** Lakukan operasi mutase dengan melakukan pemilihan secara random pada non-input unit dan untuk setiap bobot yang masuk kedalam unit, tambahkan nilai random antara -1 dan 1.

Langkah 5 : Hapus member dari parent untuk memberikan ruangan bagi offspring kromosom. Evaluasi offspring kromosom tersebut dan masukan kedalam populasi

Langkah 6 : Meningkatkan generasi dengan melakukan perulangan sebanyak n (batasan generasi yang diinginkan), dengan mengulangi langkah ke-2 hingga langkah ke-5.

GAs digunakan untuk mengembangkan bobot interkoneksi pada ANN, meskipun GAs hanya dapat menghasilkan solusi optimal terdekat, dikarenakan GAs memiliki ruang pencarian yang besar (multidimensional error).

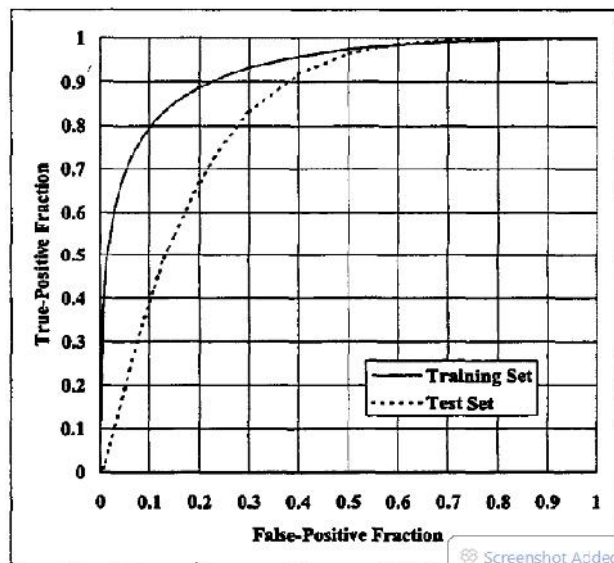
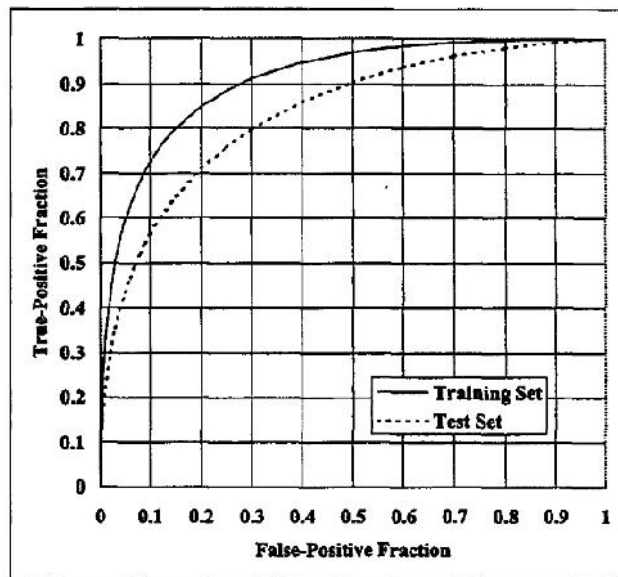
PROSES PELATIHAN DAN PENGUJIAN

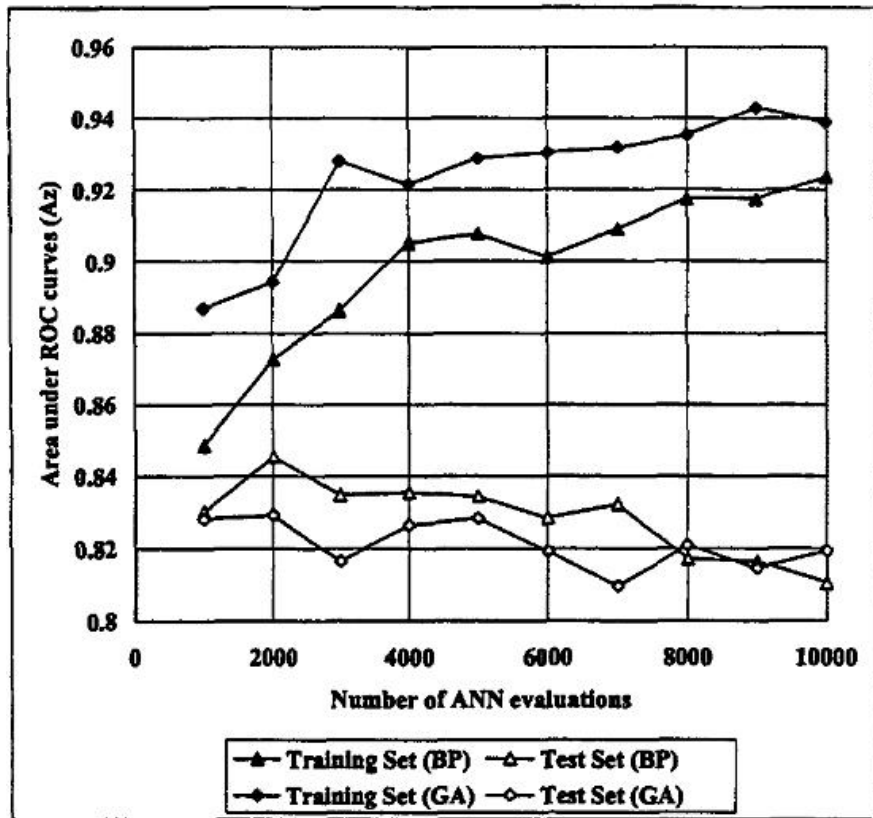
Untuk menilai apakah ANN dapat belajar dari kasus-kasus yang diketahui dan mengoreksi klasifikasi kasus yang tidak diketahui, maka digunakan 92 data positive kanker payudara dan 326 data negative kanker payudara pada database klinik yang digunakan dan dipartisi secara random ke dalam 2 bagian yaitu data latih dan data uji yaitu sebanyak 46 data positif dan 163 data negative. Data training digunakan untuk menentukan kumpulan bobot interkoneksi ANN yang optimal selama dilakukan pengujian. Proses pengujian dilakukan untuk mengevaluasi performa dari pelatihan ANN yang sudah dilakukan sebelumnya. Selama proses pelatihan jaringan, pada ANN menggunakan 2 jenis algoritma backpropagation yaitu forward backpropagation dan backward error propagation untuk setiap kromosom. Untuk itu, parameter pada Backpropagation dan algoritma genetic dipilih dari total jumlah evaluasi ANN yang mengidentifikasi perbedaan (contoh : total iterasi $n \times 2$, backpropagation = total jumlah populasi \times total jumlah generasi GA).

HASIL

Pada gambar 2 dan 3 menunjukkan kurva ROC untuk pelatihan backpropagation dan algoritma genetic ANN. Parameter untuk pelatihan backpropagation ditentukan berdasarkan : $\eta = 0.01$, momentum $\alpha = 0.8$ dan jumlah iterasi $n = 2500$. Sedangkan pada algoritma genetic, parameter yang digunakan pada proses pelatihan ditentukan berdasarkan : jumlah populasi = 20 dan jumlah generasi = 250. Melihat bahwa jumlah evaluasi ANN untuk backpropagation dan algoritma genetika adalah sama yaitu sebanyak 5000 evaluasi ANN. Proses pelatihan backpropagation ANN pada kurva ROC menghasilkan 0.91 ± 0.018 dari kumpulan data latih dan 0.83 ± 0.028 dari kumpulan data uji. Sedangkan pada proses pelatihan algoritma genetic ANN pada kurva ROC menghasilkan 0.93 ± 0.016 pada kumpulan data latih dan 0.83 ± 0.021 pada kumpulan data uji.

Pada gambar 4 menunjukkan kurva ROC untuk pelatihan backpropagation dan algoritma genetic ANN berdasarkan pada seleksi jumlah evaluasi ANN. Seperti yang diharapkan, kurva ROC untuk pelatihan meningkat seperti pada jumlah evaluasi ANN yang juga mengalami peningkatan. Tetapi kurva ROC untuk data uji menurun seperti pada jumlah evaluasi ANN yang meningkat. Dengan jumlah evaluasi yang Kurang dari 3000, proses pelatihan pada ANN mungkin menurun. Selama lebih dari 700 evaluasi ANN, ANN akan mengalami proses pelatihan yang berlebih. Optimal pada performa dari skema CAD dapat dicapai jika menggunakan 5000 evaluasi ANN.





Sumber :

Yuan-Hsiang .dkk, "Computer-Aided Diagnosis Of Breast Cancer Using Artificial Neural Network : Comparison Of Backpropagation and Genetic Algorithms", Department of Radiology, University Of Pittsburgh, Pittsburgh, PA 15261-001, IEEE Transaction