



Tersedia online di www.sciencedirect.com

Sains Langsung

Ilmu Material Procedia 10 (2015) 433 – 440

Procedia
Materials Science

www.elsevier.com/locate/procedia

Konferensi Internasional ke-2 tentang Nanomaterial dan Teknologi (CNT 2014)

Deteksi Kanker Paru Dengan Klasifikasi K-NN

Menggunakan Algoritma Genetika

P . Bhuvaneswaria*, Dr.A.Brintha Therese

*A Raja Rajeswari College of Engg, Bangalore, Research Scholar, VIT University, Chennai, India
Universitas VIT, Chennai, India*

Abstrak

Tulisan ini berfokus pada deteksi kanker paru stadium dini. Algoritma Genetic K-Nearest Neighbor (GKNN) diusulkan untuk deteksi yang merupakan metode non parametrik. Algoritma optimasi ini memungkinkan dokter untuk mengidentifikasi nodul yang ada pada gambar CT paru-paru pada tahap awal yang menyebabkan kanker paru-paru. Karena interpretasi manual terhadap gambar CT kanker paru-paru memakan waktu dan sangat kritis, untuk mengatasi kesulitan ini metode Algoritma Genetika dikombinasikan dengan algoritma K-Nearest Neighbor (K-NN) yang akan mengklasifikasikan gambar kanker dengan cepat dan efektif. MATLAB Implementasi berbasis kotak alat pemrosesan gambar dilakukan pada gambar CT paru-paru dan klasifikasi gambar-gambar ini dilakukan. Ukuran kinerja seperti tingkat klasifikasi dan tingkat positif palsu dianalisis. Dalam algoritma K-NN tradisional, awalnya jarak antara semua sampel pengujian dan pelatihan dihitung dan K-tetangga dengan jarak yang lebih jauh diambil untuk klasifikasi. Dalam metode yang diusulkan ini, dengan menggunakan Algoritma Genetika, jumlah sampel K (50-100) dipilih untuk setiap iterasi dan akurasi klasifikasi 90% dicapai sebagai kebugaran. Akurasi tertinggi dicatat setiap kali.

© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under the responsibility of the Organizing Committee of the 2nd International Conference on Nanomaterials and Technology (CNT 2014).

Tinjauan sejawat di bawah tanggung jawab Konferensi Internasional tentang Bahan dan Teknologi Nano (CNT 2014)

Kata kunci: Algoritma genetika; Filter Gabor; K-Tetangga Terdekat;

1. Perkenalan

Di alam, penyakit paru-paru mempunyai peranan yang besar dalam permasalahan kesehatan. Dalam segala bentuk penyakit paru-paru, pernapasan terutama terkena dampaknya, berikut adalah beberapa bentuk penyakit paru-paru yang umum. Bronkitis Akut, Asma, Paru Obstruktif Kronik

* Penulis yang sesuai. Telp: 09448394177; faks: +91 80 2843 7373.

Alamat email: bhuvanasamuel@gmail.com

Penyakit (COPD), bronkitis kronis, Emfisema, Sindrom gangguan pernapasan akut (ARDS) dan kanker paru-paru. Sebagai per Laporan kanker dunia tahun 2014 kanker paru-paru adalah penyebab paling umum kematian terkait kanker pada pria dan wanita, dan bertanggung jawab atas 1,56 juta kematian setiap tahunnya, pada tahun 2012. Penyebab utama penyakit paru-paru adalah merokok, menghirup obat-obatan, merokok dan bahan alergi. Gambar computerized tomography (CT) membantu mendeteksi penyakit paru-paru yang ekstrem. Untuk analisis metode yang diusulkan, gambar CT sudah cukup dan visibilitas jaringan lunak lebih baik. Ada beberapa jenis kanker paru-paru, dan ini dibagi menjadi dua kelompok utama: Kanker paru-paru sel kecil dan kanker paru-paru non-sel kecil yang memiliki tiga sub tipe: Karsinoma, Karsinoma Adeno dan Karsinoma Sel Skuamosa [1].

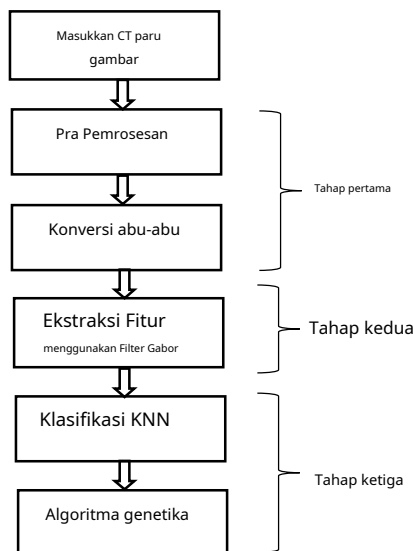
Deteksi berbasis Neural Ensemble (NED) diusulkan oleh Zhi-Hua Zhou et al [2] untuk mengidentifikasi gambar kanker paru-paru di mana pembelajaran bertahap satu kali dilakukan oleh pengklasifikasi saraf adaptif dengan kecepatan dan akurasi tinggi. Mokhled S. AL-TARAWNEH mengusulkan teknik untuk mendeteksi fitur gambar yang akurat dengan membandingkan persentase piksel dan pelabelan topeng dan faktor waktu dipertimbangkan untuk menemukan masalah kelainan yang ada pada gambar target. Penilaian kualitas gambar serta tahap peningkatan diadopsi pada teknik seperti filter Gabor dalam aturan Gaussian. Sebuah algoritma diusulkan oleh KAGUdeshani, RGNMeegama, dan TGIFernando [3] yang menggunakan dua langkah pemrosesan gambar. Pada langkah pertama pemisahan paru-paru dan pada langkah kedua jaringan saraf dilatih berdasarkan dua jenis input berbasis piksel dan berbasis fitur statistik serta tingkat pengenalan ditabulasi dan dibandingkan. Kerry A. Seitz et.al menunjukkan efektivitas dan efisiensi pengambilan gambar berbasis konten yang dapat ditingkatkan dengan algoritma genetika [4].

Dalam usulan pekerjaan untuk menghilangkan noise dan peningkatan kontras, gambar diproses terlebih dahulu untuk mendapatkan gambar yang disempurnakan secara akurat. Filter Gabor digunakan dalam ekstraksi Fitur. Nilai keluaran filter Gabor diberikan ke K-NN (Kernel Nearest Neighborhood) yang dioptimasi dengan GA (Genetic Algorithm)

Makalah ini disusun sebagai berikut: Bagian 2 memfokuskan metodologi yang diusulkan. Bagian 3 menjelaskan klasifikasi K-NN, Bagian 4 membahas tentang Algoritma Genetika dan Bagian 5 menyajikan implementasi GA dengan K-NN beserta analisis kinerjanya dan Bagian 6 membahas faktor kinerja untuk optimalisasi gambar yang lebih baik.

2. Metodologi

Metode yang diusulkan melibatkan tiga tahap ditunjukkan pada Gambar.1. Awalnya gambar CT paru-paru telah diproses sebelumnya dan disegmentasi. Tahap selanjutnya adalah ekstraksi fitur yang dilakukan dengan filter Gabor, tahap ketiga adalah klasifikasi dengan K-NN, dan optimasi dengan Algoritma Genetika.



Gambar. 1 Tahapan Algoritma yang Diusulkan

Pada algoritma yang diusulkan, peningkatan kontras gambar masukan melalui metode pra-pemrosesan dilakukan dengan terlebih dahulu mengubah gambar masukan yang diberikan menjadi gambar skala abu-abu. Setelah meningkatkan kontras gambar, diterapkan ke filter Gabor untuk mengekstrak kontras fitur. Filter Gabor yang lebih banyak digunakan dalam pengolahan citra adalah analisis tekstur. Prinsip Ketidakpastian digunakan dan memberikan lokasi frekuensi waktu yang tepat. Baik dalam domain spasial maupun frekuensi, filter ini dapat dioperasikan dan respons impulsnya ditentukan oleh gelombang sinusoidal dikalikan dengan fungsi Gaussian. Konvolusi transformasi Fourier fungsi Gaussian adalah transformasi Fourier respons filter Gabor. Respons Filter Gabor berisi komponen nyata dan imajiner yang mewakili arah ortogonal yang ditunjukkan oleh persamaan di bawah ini.

Kompleks

$$g_{x,kamu}, \dots, \text{pengalaman} \frac{X_2}{2^2} \text{pengalaman} 2 \frac{X}{2} \quad (1)$$

Nyata

$$g_{x,kamu}, \dots, \text{pengalaman} \frac{X_2}{2^2} \text{Karena } 2 \frac{X}{2} \quad (2)$$

Imajiner

$$g_{x,kamu}, \dots, \text{pengalaman} \frac{X_2}{2^2} \text{Dosa } 2 \frac{X}{2} \quad (3)$$

Di mana

$$X \quad x \cos \quad ya \quad (4)$$

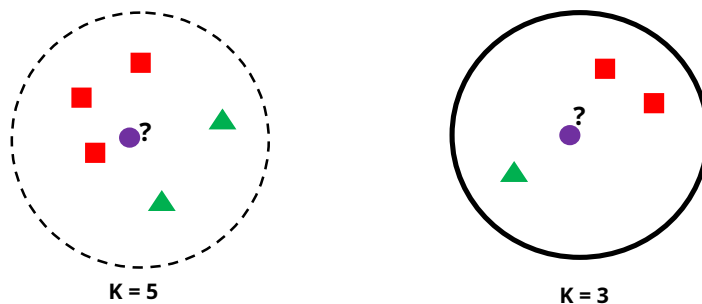
$$kamuxsin \quad kamu \quad (5)$$

Konvolusi citra masukan dengan kernel Gabor-filter untuk seluruh kombinasi orientasi dan seluruh fase-offset dengan citra masukan dihitung dengan filter Gabor yang menghasilkan matriks 4D berisi koordinat citra, offset fasa dan orientasi [5]. filter Gabor diberikan untuk klasifikasi KNN yang dioptimasi dengan Algoritma Genetika.

3. Klasifikasi K-Tetangga Terdekat

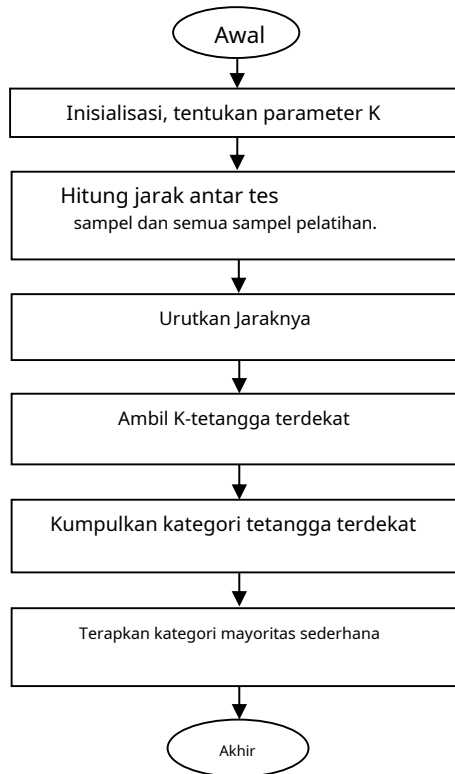
Dalam pengenalan pola, K-Algoritma Nearest Neighbor (K-NN) adalah metode non-parametrik yang digunakan untuk klasifikasi dan regresi. Dalam kedua kasus tersebut, masukan terdiri dari K contoh pelatihan terdekat di ruang fitur. K-NN merupakan salah satu jenis pembelajaran berbasis instance.

Pada Klasifikasi K-NN, keluarannya berupa keanggotaan kelas. Klasifikasi dilakukan dengan suara terbanyak dari tetangga. Jika K = 1, maka kelas tersebut merupakan tetangga terdekat tunggal [6].



Gambar .2. Klasifikasi sampel uji

Klasifikasi sampel uji ditunjukkan pada Gambar 2 di atas. Misalkan sampel uji adalah titik besar yang terletak di dalam lingkaran yang diklasifikasikan ke dalam segitiga kelas satu atau kotak kelas kedua. Jika $K=5$ (lingkaran garis putus-putus) dimasukkan ke kelas kedua karena di dalam lingkaran tersebut terdapat 3 kotak dan 2 segitiga. Jika $K=3$ (lingkaran garis padat) dimasukkan ke kelas kedua karena di sini terdapat 2 kotak dan 1 segitiga di dalam lingkaran tersebut. Hal ini akan berguna jika kontribusi bobot dari tetangga dipertimbangkan karena tetangga terdekat memberikan kontribusi lebih besar dibandingkan tetangga jauh. Misalnya, dalam skema pembobotan umum, masing-masing tetangga diberi bobot $1/d$ jika d adalah jarak ke tetangga. Jarak terpendek antara dua tetangga selalu berupa garis lurus dan jarak tersebut dikenal dengan jarak Euclidean [7]. Keterbatasan algoritma K-NN adalah sensitif terhadap konfigurasi data lokal. Proses mengubah data masukan menjadi sekumpulan fitur dikenal sebagai ekstraksi fitur. Di ruang Fitur, ekstraksi dilakukan pada data mentah sebelum menerapkan algoritma K-NN. Gambar 3 menceritakan langkah-langkah yang terlibat dalam algoritma K-NN.



Gambar. 3. Algoritma Klasifikasi K-NN

4. Algoritma Genetika

Algoritma genetika termasuk dalam kelas algoritma evolusioner yang lebih besar, yang menghasilkan solusi untuk masalah optimasi menggunakan teknik yang terinspirasi oleh evolusi alam, seperti pewarisan, mutasi, seleksi, dan persilangan. [11]

4.1. Pilihan

Di alam, seleksi individu dilakukan berdasarkan survival of the fittest. Jika individu lebih beradaptasi dengan lingkungan, maka peluangnya lebih besar untuk bertahan hidup dan menghasilkan keturunan sehingga dapat mentransfer gennya ke populasi berikutnya. Dalam Algoritma Evolusioner, pemilihan individu terbaik didasarkan pada evaluasi kebugaran

fungsi. Contoh fungsi kebugaran tersebut adalah jumlah kesalahan kuadrat antara respons sistem yang diinginkan dan respons sistem sebenarnya; jarak kutub sistem loop tertutup ke kutub yang diinginkan, dll. Jika masalah optimasi adalah masalah minimalisasi, maka individu dengan nilai fungsi kebugaran yang kecil akan memiliki peluang lebih besar untuk rekombinasi dan masing-masing menghasilkan keturunan.

4.2. Rekombinasi

Langkah pertama dalam proses reproduksi adalah rekombinasi (crossover). Di dalamnya gen orang tua digunakan untuk membentuk kromosom yang sama sekali baru. Rekombinasi tipikal untuk GA adalah operasi yang membutuhkan dua orang tua, namun skema dengan lebih banyak orang tua juga dimungkinkan. Dua algoritma yang paling banyak digunakan adalah Conventional (Scattered) Crossover dan Blending (Intermediate) Crossover [9]

4.2.1. Crossover Konvensional (Tersebar).

Dalam tipe rekombinasi ini, orang tua menukar gen yang sesuai untuk membentuk seorang anak. Crossover dapat berupa titik tunggal atau titik banyak seperti ditunjukkan pada Gambar 4. (a) dan (b). Untuk rekombinasi sedikit Masker digunakan. Persamaan yang menggambarkan proses tersebut adalah:

$$C1 = \text{Masker1} \& P1 + \text{Masker2} \& P2 \quad (6)$$

$$C2 = \text{Masker2} \& P1 + \text{Masker1} \& P2 \quad (7)$$

Di mana

P1, P2 - Kromosom Induk;

C1, C2 - Kromosom Anak (individu keturunan); Mask1,

Mask2 – sedikit Masker;

Mask2 = NOT (Mask1)) & operasi bit “AND”

Misalnya pada Gambar (4.b);

Topeng 1 = [1 1 1 0 1 1 0 0 0];

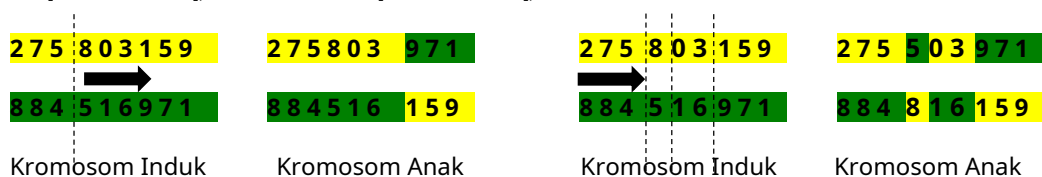
P1 = [2 7 5 8 0 3 1 5 9];

Topeng 2 = TIDAK (Topeng1) = [0 0 0 1 0 0 1 1 1];

P2 = [8 8 4 5 1 6 9 7 1];

(8)

(9)



Gambar. 4(a): Persilangan Titik Tunggal

Gambar. 4(b): Persilangan Multi Titik

4.3. Mutasi

Keturunan yang dihasilkan melalui seleksi dan persilangan populasi selanjutnya dapat diterapkan pada mutasi. Dalam istilah Algoritma Genetika, mutasi berarti perubahan acak pada nilai suatu gen dalam suatu populasi, juga beberapa elemen DNA berubah. Perubahan tersebut terutama disebabkan oleh kesalahan pada proses penyalinan gen induk. [10]

4.4. Penerapan

4.4.1. Representasi String

Di sini, pengkodean kromosom dengan bilangan real; Di setiap kromosom, jumlah gen mewakili sampel dalam set pelatihan. Setiap gen akan memiliki k jumlah gen dan 4 digit untuk indeks vektor. Misalnya, jika K = 4, kromosom sampel dapat ditulis sebagai berikut:

0001 0102 0204 0302 0401 0500 0601 0702 0802 0901 1002 1101 1201

Pada string di atas, dua digit pertama mewakili atribut, dua digit sisanya mewakili nilai. Demikian pula seluruh gen (4 digit) dikodekan dan sekarang populasi awal dihasilkan sehingga kita dapat menerapkan operator genetika. Dengan tetangga K ini, jarak antara masing-masing sampel dalam set pengujian dihitung dan keakuratannya disimpan sebagai nilai kebugaran kromosom ini.

4.4.2. Pilihan

Dalam proses seleksi, setiap kromosom dipilih dari kelompok perkawinan sesuai dengan arah konsep sistem genetika alami yang paling sesuai. Teknik umum yang menerapkan strategi seleksi proporsional adalah pemilihan roda Roulette [12]. Dalam strategi seleksi proporsional, sebuah kromosom diberi sejumlah salinan, yang sebanding dengan kesesuaiannya dalam populasi yang akan dimasukkan ke dalam kelompok perkawinan untuk diadopsi operasi genetika lebih lanjut.

4.4.3 Persilangan

Crossover adalah proses probabilitas yang menukarkan informasi antara dua kromosom induk untuk menghasilkan dua kromosom anak. Crossover titik tunggal dengan probabilitas crossover tetap G digunakan dalam tulisan ini. Untuk kromosom dengan panjang S, bilangan bulat acak, yang disebut titik persilangan, dihasilkan dalam rentang $[1, S-1]$. Kromosom yang terletak di bagian kanan titik persilangan dipertukarkan untuk menghasilkan dua keturunan.

4.4.3 Mutasi

Setiap kromosom mengalami mutasi dengan probabilitas tetap G. Untuk representasi biner kromosom, posisi bit (atau gen) dimutasi hanya dengan membalik nilainya. Karena dalam makalah ini bilangan real dipertimbangkan, posisi acak dipilih dalam kromosom dan diganti dengan bilangan acak antara 0-9. Setelah operator genetika diterapkan, nilai kebugaran maksimum lokal dihitung dan dibandingkan dengan maksimum global. Jika maksimum lokal lebih besar dari maksimum global maka maksimum global ditetapkan dengan maksimum lokal, dan iterasi selanjutnya dilanjutkan dengan populasi baru. Titik cluster akan direposisi sesuai dengan kromosom yang memiliki maksimum global. Jika tidak, maka iterasi berikutnya dilanjutkan dengan populasi lama yang sama. Proses ini diulangi sebanyak N iterasi. Dari bagian berikut, terlihat bahwa algoritma di bawah ini meningkatkan kualitas cluster.

1. Pilih jumlah K sampel dari set pelatihan untuk menghasilkan populasi awal (P1).
2. Hitung jarak antara set pelatihan di setiap kromosom dan sampel pengujian, sebagai nilai kebugaran.
3. Pilih kromosom dengan nilai fitness tertinggi, simpan sebagai Global maksimum (Gmax), untuk nilai iterasi 1 hingga L

Lakukan reproduksi

Terapkan operator crossover.

Lakukan mutasi dan dapatkan populasi baru. (P2)

Hitung maksimum lokal (Lmax).

Jika $G_{max} < L_{max}$ maka

Tetapkan $G_{max} = L_{max}$;

$P1 = P2$;

Mengulang

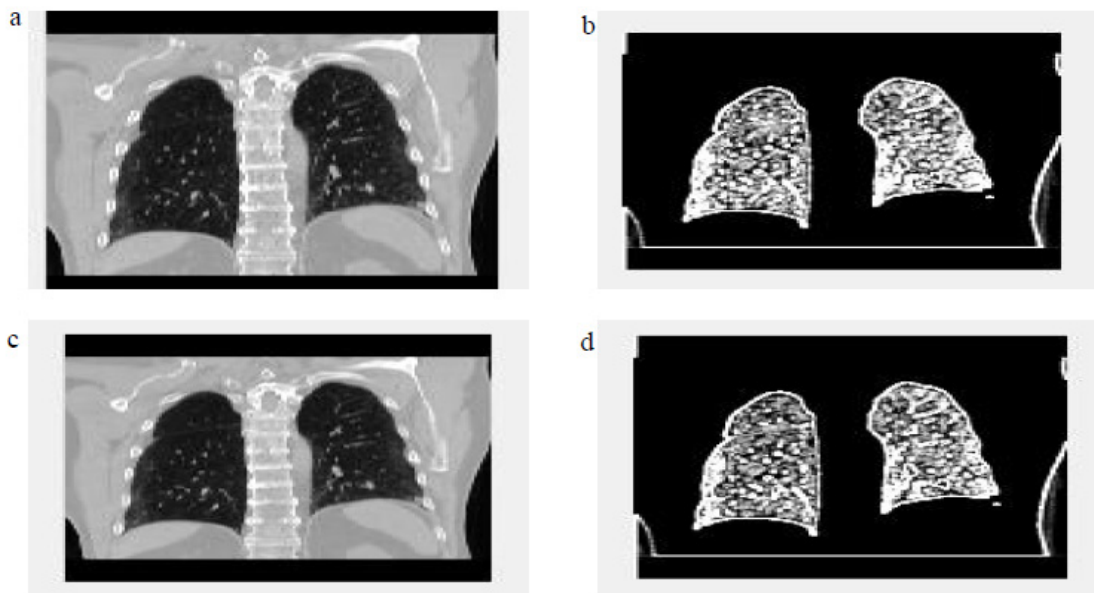
4. Output – kromosom yang memperoleh Gmax memiliki K-tetangga optimal dan label yang sesuai merupakan hasil klasifikasi. [8]

5. Hasil dan Pembahasan

Dalam metode yang diusulkan ini, kami telah berhasil mengembangkan solusi untuk mendeteksi nodul kanker paru-paru menggunakan algoritma pemrosesan gambar dan jaringan saraf. Algoritma ini diuji untuk lima set gambar CT paru-paru kanker dan non-kanker dan ditunjukkan pada Gambar 5 & Gambar 6 di bawah ini. Gambar masukan dan gambar uji serta filter Gabornya

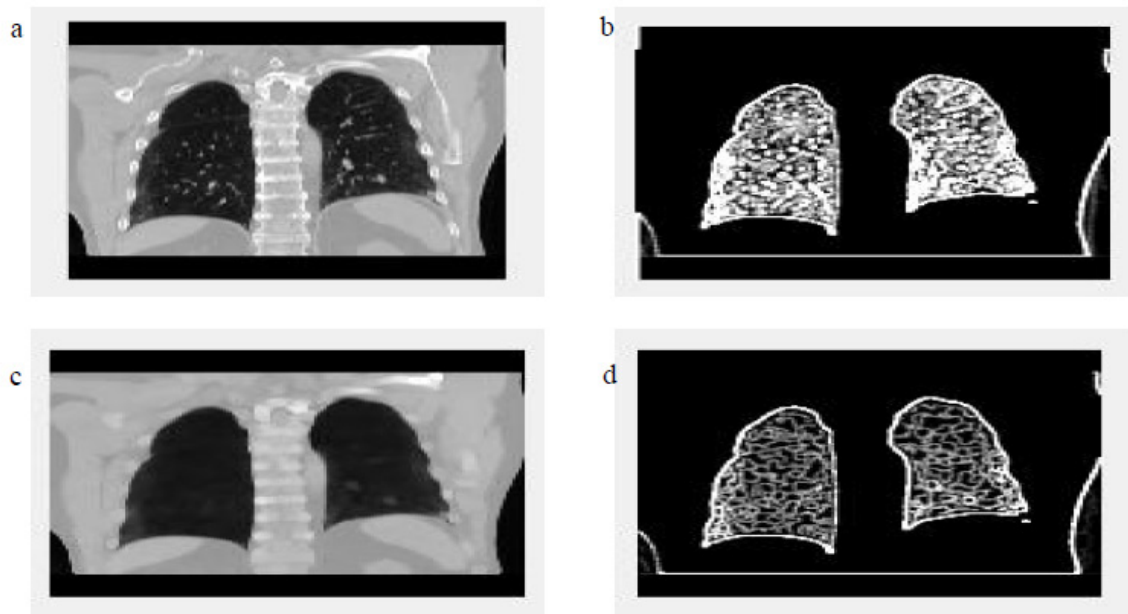
output ditampilkan untuk kedua kasus. Nilai K, waktu eksekusi dan akurasi dihitung dan ditabulasi. Lihat Tabel. 1

Gambar Terdeteksi Kanker:



Gambar .5. (a) Gambar masukan (b) Keluaran Filter Gabornya (c) Gambar Uji (d) Keluaran Filter Gabor gambar uji

Gambar yang Tidak Terdeteksi Kanker:



Gambar. 6. (a) Gambar masukan (b) Keluaran Filter Gabornya (c) Gambar Uji (d) Keluaran Filter Gabor gambar uji

Tabel 1. Ukuran Kinerja K, Waktu Eksekusi dan Akurasi Gambar CT Paru

Gambar CT paru	K	Waktu Eksekusi (Detik)	Ketepatan%
Kanker-1	52	3.65	89
Kanker-2	51	3.80	88
Kanker-3	53	3.60	88
Non Kanker-1	50	4.16	90
Non Kanker-2	49	4.25	89
Non Kanker-3	51	4.00	90

6. Kesimpulan

Untuk mengatasi keterbatasan K-NN tradisional, metode baru untuk meningkatkan kinerja klasifikasi K-NN menggunakan Algoritma Genetika (GA) diusulkan dalam makalah ini. Pengklasifikasi G-KNN yang diusulkan diterapkan untuk klasifikasi dan k-tetangga serupa dipilih pada setiap iterasi untuk klasifikasi dengan menggunakan GA, sampel uji diklasifikasikan dengan tetangga ini dan akurasi dihitung untuk jumlah nilai K yang berbeda untuk mendapatkan akurasi tinggi; sehingga waktu komputasi K-NN berkurang dari hasil yang diperoleh dalam metode ini. Implementasi berbasis kotak peralatan pemrosesan gambar MATLAB dilakukan pada gambar CT paru-paru dan dilakukan klasifikasi gambar-gambar tersebut. Nilai k, waktu eksekusi dan akurasi dihitung dan ditabulasi. Deteksi dini seperti itu mungkin bermanfaat bagi dokter.

Ucapan Terima Kasih

Terima kasih banyak kepada Sri.ACShanmugam, Ketua dan Dr.MSBhagyaShekar, Kepala Sekolah Teknik Rajarajeswari, Bangalore yang telah memberikan kesempatan dan dorongan untuk mempresentasikan karya ini. Ucapkan terima kasih juga kepada Rekan Kerja dan anggota keluarga atas dukungannya yang baik.

Referensi

- Mokhled S. Al-Tarawneh 2012 Deteksi Kanker Paru Menggunakan Teknik Pemrosesan Gambar Leonardo Electronic Journal of Practices and Teknologi Edisi 20. hlm.147-158
- Zhi-Hua Zhou, Yuan Jiang, Yu-Bin Yang, Shi-Fu Chen 2002 Identifikasi Sel Kanker Paru-paru Berdasarkan Ansambel Jaringan Syaraf Tiruan Pengobatan Kecerdasan Buatan Elsevier .Vol 24.Edisi 1.pp 25-36
- KAGUdeshani,RGNMeegama,TGIFernando 2011 Pendekatan Jaringan Syaraf Berbasis Fitur Statistik untuk Deteksi Kanker Paru-Paru pada Gambar Rontgen Dada, International Journal of Image Processing.Vol 5 .Issue 4.pp 425-434
- Kerry A. Seitz, Jr.a, Anne-Marie Giucab, Jacob Furstc, Daniela Raicuc . 2012. Pembelajaran Kesamaan Nodul Paru Menggunakan Algoritma Genetika Proses. SPIE 8315,
- W. Li, Kezhi Mao, Tianyou Chai. Pemilihan filter Gabor untuk ekstraksi fitur tekstur yang ditingkatkan. Konferensi IEEE ke-17 2010 tentang Pemrosesan Gambar, hal 361-364
- Saravanan Thirumuruganathan, Pengantar Mendetail Algoritma K-Nearest Neighbor (KNN). 2010.Wordpress.com
- M. Akhil jabbara, BL Deekshatulub dan Priti Chandrac,. 2013. Klasifikasi Penyakit Jantung Menggunakan K- Nearest Neighbor dan Algoritma Genetika", *Procedia Technology*, Vol.10
- N.Suguna,Dr.K.Thanushkodi 2010 Klasifikasi K-Nearest Neighbor yang ditingkatkan menggunakan algoritma genetika IJCSI Edisi 4, Vol-7 hal 18-21 Shital Shah dan Andrew Kusiak, 2007. Pencarian gen kanker dengan penambahan data dan algoritma genetika", *Komputer dalam Biologi dan Kedokteran*, Jilid 37, Edisi 2,
- Temesguen Messaya, , Russell C. Hardiea, Steven K. Rogersb,. 2010 Sistem CAD baru yang efisien secara komputasi untuk nodul paru deteksi dalam citra CT, *Analisis Citra Medis*, Vol 14, edisi 3 hal 390-406.
- Algoritma Genetika oleh David E. Goldberg,2006. Pendidikan Pearson
- J.Aruna Devi, Dr.V.Rajamani. 2011 Klasifikasi multi label Evolusi menggunakan Associative Rule Mining untuk Preferensi Spasial, Pendekatan Novel AIT IJCA dan Aplikasi Praktis