

Klasifikasi Citra Mammogram Berbasis Tekstur Menggunakan SVM

Lussiana ETP¹, Suryarini Widodo², Karmilasari³, Matrisnya H.⁴, Faisal Reza⁵

¹Program Studi Sistem Informasi, STMIK Jakarta STI&K, Jakarta

E-mail : lussiana@jak-stik.ac.id

^{2,3,4}Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Gunadarma, Depok 16424

E-mail : {srini, karmila, matrisnya}@staff.gunadarma.ac.id

⁵Fakultas Teknologi Industri Universitas Gunadarma, Depok 16424

E-mail : ichza@student.gunadarma.ac.id

ABSTRAK

Latar Belakang : Kanker payudara adalah penyakit yang ditandai oleh terjadinya pertumbuhan jaringan yang berlebihan, atau perkembangan sel-sel jaringan payudara yang tidak terkontrol. Salah satu cara untuk mendeteksi adanya jaringan tersebut adalah dengan menggunakan pemeriksaan mammografi. Selanjutnya untuk menganalisis citra hasil mammografi (mammogram) sangat bergantung pada dokter ahli radiologi, sehingga pengalaman dan keahlian dalam menganalisis citra mammogram sangat mempengaruhi keakuratan diagnosis. Dengan demikian untuk membantu dan memudahkan dokter dalam menganalisis citra mammogram perlu adanya alat bantu pendiagnosis citra mammogram yang dapat menganalisis dan mengidentifikasi adanya jaringan yang dicurigai secara akurat. Tujuan penelitian adalah melakukan klasifikasi citra mammogram dengan menggunakan metode Support Vector Machine (SVM).

Metode : Tahap penelitian diawali dengan pencarian ROI, dilanjutkan dengan melakukan segmentasi citra, ekstraksi citra, dan klasifikasi citra dengan menggunakan SVM.

Hasil : Berdasarkan hasil percobaan menunjukkan bahwa SVM mampu mengklasifikasi citra mammogram dengan akurasi 91% untuk data kelas normal dan 87,667% untuk data kelas abnormal

Kata kunci : Citra, Mammogram, Tekstur, SVM

PENDAHULUAN

Menurut data WHO 8-9% wanita mengalami kanker payudara pada masa hidupnya. Berdasarkan data yang ada di situs <http://globocan.iarc.fr> sejak tahun 2008, diestimasi bahwa di Indonesia terdapat sebanyak 39.831 kasus yang terjadi pada wanita dan hanya setengah dari penderita yang selamat dari penyakit ini. Kanker payudara merupakan penyakit yang disebabkan oleh terjadinya pertumbuhan jaringan yang berlebihan, atau perkembangan sel-sel jaringan payudara yang tidak terkontrol. Berkaitan dengan pengidentifikasian dan diagnosis kanker payudara, para dokter ahli menggunakan bantuan rekomendasi dokter radiologi melalui hasil analisis citra dari USG, MRI, atau X-RAY Mammography.¹ Hal penting sebagai langkah awal dalam menganalisis citra mammogram adalah mengetahui atau mengidentifikasi letak jaringan yang dicurigai. Untuk menentukan letak jaringan tersebut dapat dilakukan dengan menganalisis tekstur citra mammogram, selanjutnya dilakukan segmentasi yang bertujuan untuk memisahkan objek yang dicurigai terhadap jaringan lainnya. Beberapa metode yang telah dilakukan antara lain adalah dengan menentukan *Region of Interest* (ROI) berbasis kluster, dan fitur *Wavelet*. *Wavelet* sendiri digunakan dalam penelitian mengenai pendeteksian tumor pada mammogram *digital* menggunakan *wavelet*.²

Tumor adalah benjolan tidak normal akibat pertumbuhan sel yang terjadi secara terus menerus. Menurut teori genetika, penyebab terbentuknya tumor adalah terjadinya penyimpangan genetika sedemikian rupa sehingga menyebabkan pembelahan sel menjadi berlebihan dan tidak terkendali. Tumor ini dapat bersifat jinak maupun ganas. Tumor ganas atau kanker dapat menembus dan menghancurkan jaringan tubuh yang sehat. Oleh karena kanker membawa risiko kematian bagi

penderitanya, usaha untuk mengurangi faktor risiko dapat dilakukan dengan melakukan pemeriksaan dini, salah satunya adalah⁶ : pemeriksaan radiologi yang merupakan pemeriksaan lanjutan untuk melengkapi pemeriksaan dokter berupa mamografi. Mamografi adalah suatu metode pendeskripsian dengan menggunakan sinar X berkead rendah⁷. Citra yang hasil mamografi disebut mammogram. Hal ini membuktikan bahwa teknologi pengolahan citra telah banyak diaplikasikan ke berbagai bidang, salah satunya adalah bidang kedokteran. Pada teknik pengolahan citra, cara untuk menentukan daerah atau lokasi yang diinginkan dikenal dengan metode segmentasi.

Segmentasi citra bertujuan untuk membagi wilayah-wilayah yang homogen, sehingga dapat membedakan objek dengan latar belakang. Pengelompokan citra berdasarkan kesamaan-kesamaan sifat atau kriteria yang ada dapat dilakukan dengan segmentasi berbasis *clustering*, yaitu pengelompokan piksel citra ke dalam beberapa cluster. Keberhasilan proses segmentasi berbasis kluster ditentukan oleh keberhasilan dalam mengelompokkan fitur-fitur yang berdekatan ke dalam satu cluster. Metode-metode dalam segmentasi berbasis cluster salah satunya adalah iterasi, *K-means*. *K-means* mengelompokkan objek menjadi K kluster. Metode ini akan mencari pusat kluster dan batas-batas kluster melalui proses perulangan (*iterative*). Kedekatan atau kemiripan (*similarity*) suatu objek dengan objek lain atau dengan pusat kluster dihitung dengan menggunakan fungsi jarak. Pada umumnya *K-means* menggunakan jarak *Euclidean* untuk menghitung kemiripan tersebut. Langkah pertama dari metode *K-means* adalah menentukan inisialisasi sejumlah K pusat kluster. Secara iteratif, pusat kluster tersebut akan diperbaiki sehingga merepresentasikan pusat-pusat dari K kluster.

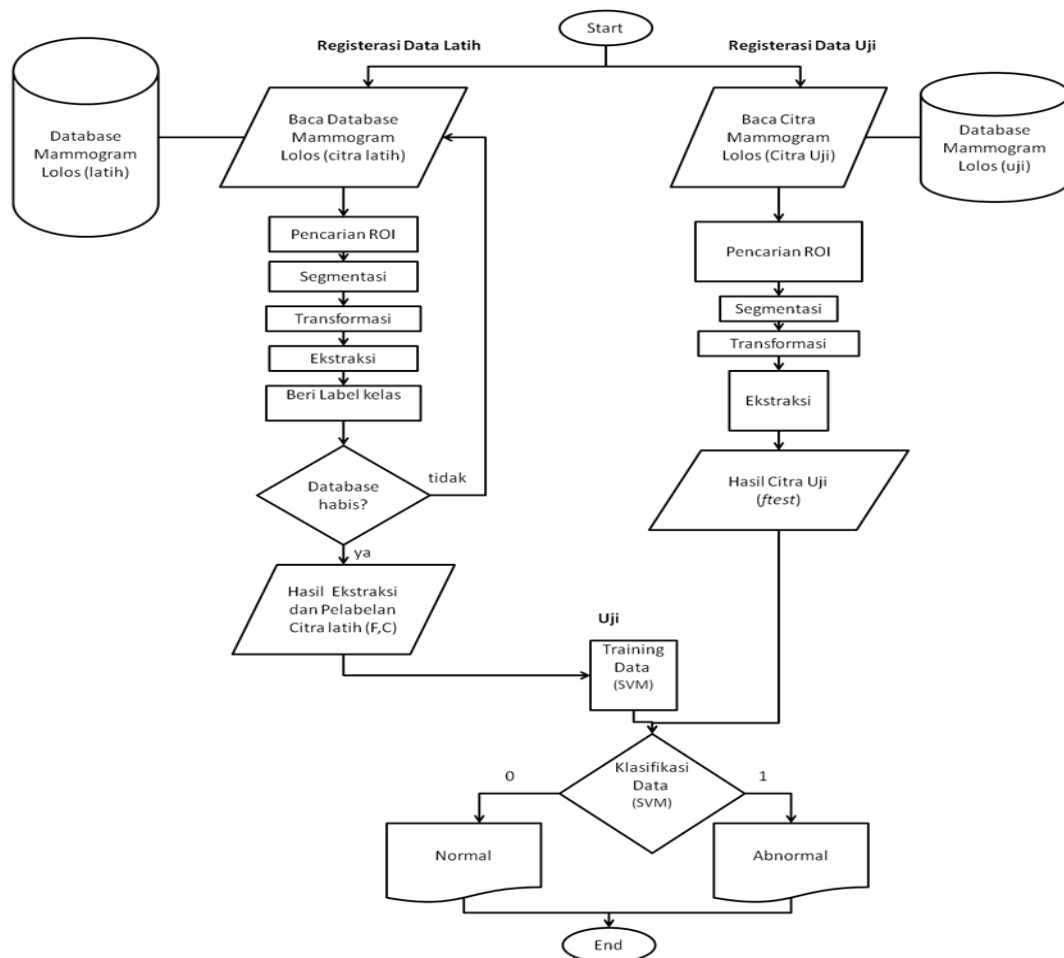
Ekstraksi adalah sebuah proses yang dilakukan untuk mendapatkan nilai – nilai yang digunakan untuk proses selanjutnya. Dalam citra digital proses ekstraksi dapat menghasilkan nilai-nilai (fitur) seperti energi, *entropy*, dan standar deviasi. Selain nilai intensitas, citra juga dapat dikenali melalui tekstur citra.

Klasifikasi merupakan proses pengelompokan piksel ke dalam kelas-kelas tertentu. Dalam perkembangannya, proses klasifikasi dapat dilakukan dengan mempelajari fitur-fitur yang disajikan oleh hasil ekstraksi. Metode yang banyak digunakan untuk melakukan klasifikasi adalah dengan kecerdasan buatan atau *machine learning*. Salah satu metode *machine learning* adalah *Support Vector Machine*. *Support vector machine* (SVM) adalah suatu teknik untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi, yang sangat populer belakangan ini.⁸

KLASIFIKASI CITRA MAMMOGRAM MENGGUNAKAN SVM

Data yang digunakan dalam penelitian ini adalah data citra mammogram yang tersedia pada situs mammogram¹ . Data tersebut koleksi *Mammographic Image Analysis Society* (*MIAS*) yang relatif berukuran kecil, dengan format PGM (Portable Gray Map). Jumlah data yang digunakan dalam penelitian ini berjumlah 322 citra mammogram, yang terbagi menjadi 3 kategori utama, yaitu : Dense-glandular, *Fatty*, dan *Fatty-glandular*. Dari setiap kategori terbagi menjadi 3 klasifikasi, yaitu : Normal, *Benign* dan *Malignant*. Metode penelitian yang dilakukan pada penelitian ini dapat dilihat pada gambar 1.

¹http://abacus.ee.cityu.edu.hk/imagedb/cgi-bin/ibrowser/ibrowser.cgi?folder=Medical_Image/mammogram/



Gambar 1. Bagan Klasifikasi Citra Mammogram

Tahap registrasi citra bertujuan untuk pembentukan database citra mammogram. Pada tahap ini terdapat tahapan-tahapan seperti tahap pencarian ROI, segmentasi, transformasi dan ekstraksi citra. Sebelum melakukan segmentasi, terlebih dahulu menentukan daerah yang dicurigai atau Region of Interest (ROI) dengan menggunakan Algoritma K-Means. Dalam penelitian ini digunakan sebanyak 3 cluster.

Transformasi citra merupakan fungsi untuk menggeser nilai-nilai citra. Dengan menggunakan transformasi dapat dihasilkan citra baru dengan informasi yang sudah terkelompok dengan metode tertentu. Pada penelitian ini digunakan *wavelet daubechies* orde 6. Hasil transformasi berupa citra yang terdekomposisi merupakan citra-citra yang selanjutnya diekstraksi. Ekstraksi citra merupakan tahap untuk menghasilkan nilai-nilai yang menjadi parameter klasifikasi citra mammogram.

Tahap pengenalan citra merupakan proses pengenalan hasil pengumpulan data kepada mesin pengenal (*machine learning*). Selanjutnya data dilatih dengan menggunakan metode *Support Vector Classification* (SVC). Data yang dilatih merupakan hasil proses registrasi data latih. Seperti diketahui bahwa hasil proses registrasi data adalah variabel F dan variabel C. Variabel F merupakan variabel yang berisikan nilai-nilai fitur (fitur energi), sedangkan variabel C merupakan variabel yang berisikan informasi klasifikasi citra pada fitur yang berada dalam satu urutan baris.

Tahap akhir adalah tahap pengujian. Pengujian data merupakan kegiatan membandingkan data uji dengan struktur data latih yang dihasilkan dari proses

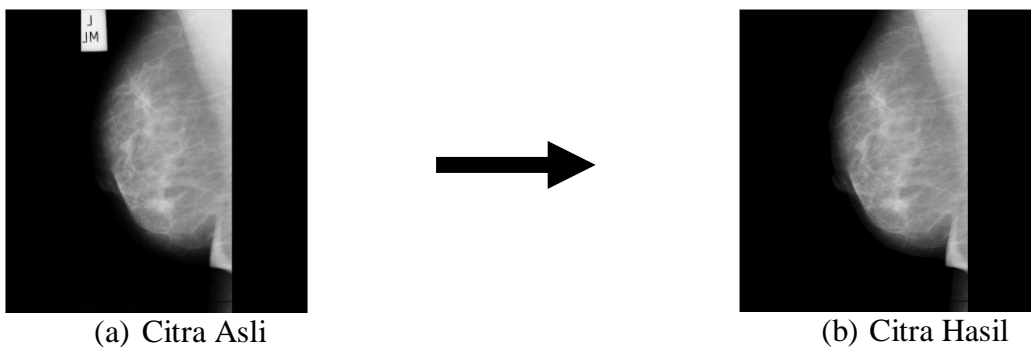
pengenalan (*training*). Proses ini menghasilkan keluaran sama seperti label yang diberikan pada fitur data latih. Label yang ada pada data latih adalah nilai 0 atau 1, dimana 0 menunjukkan kondisi citra mammogram untuk kategori normal dan 1 menunjukkan citra mammogram yang dicurigai (*suspect*) atau kategori abnormal.

Beberapa penelitian yang telah dilakukan antara lain³ melakukan ekstraksi tekstur mammogram dengan menggunakan metode *wavelet* dan *gray-level*. Lalu, Manoharan⁴ melakukan klasifikasi terhadap mikro klasifikasi pada mammogram dengan menggunakan metode SVM. Selain itu, Leonardo de Oliviera⁵ melakukan deteksi massa dalam mammogram digital menggunakan K-means.

Penelitian ini membahas mengenai identifikasi jaringan yang dicurigai dan dilanjutkan dengan pengklasifikasian citra mammogram. Sebagai citra uji digunakan citra mammogram dengan kategori sebagai berikut : *Dense-glandular*, *Fatty*, *Fatty-Glandular*. Tujuan penelitian ini antara lain: Mengidentifikasi dan mensegmentasi citra mammogram berdasarkan pada tekstur citra dengan menggunakan K-means dan melakukan klasifikasi citra dengan menggunakan metode SVM, yaitu suatu teknik pengenalan menggunakan mesin pemisah.

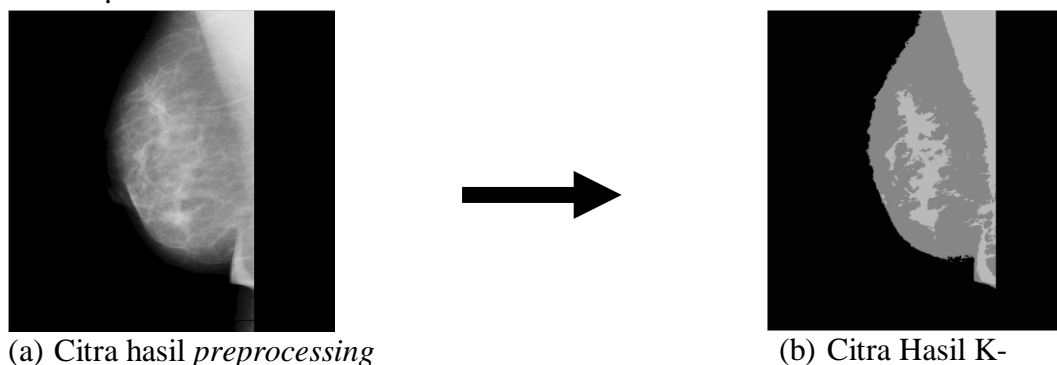
HASIL DAN PEMBAHASAN

Tahap awal yang dilakukan adalah tahap *preprocessing* yang bertujuan untuk menghilangkan informasi yang tidak diperlukan, dalam hal ini penghilangan label pada citra mammogram, hasil tahap ini dapat dilihat pada gambar 2.



Gambar 2. Perubahan citra asli menjadi citra hasil *preprocessing*

Tahap kedua adalah tahap pencarian ROI dilakukan dengan menggunakan algoritma *K-Means* dengan $K=3$. ROI yang dihasilkan oleh tahap ini disajikan pada Gambar 3



Means

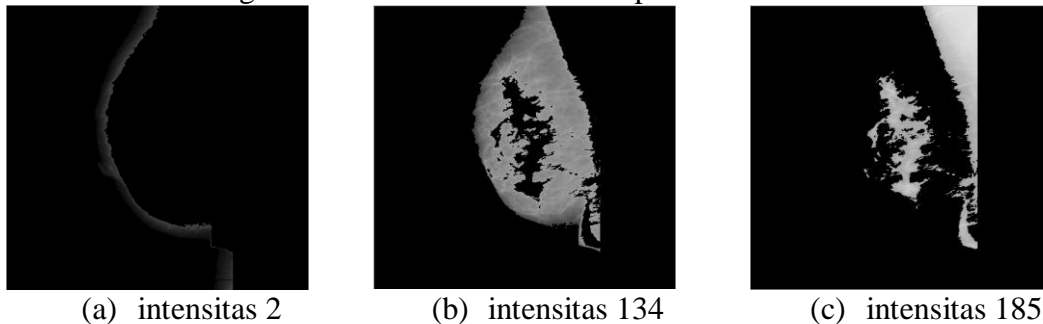
Gambar 3. Citra hasil pengolahan *K-means*

Gambar 3(b) menunjukkan citra segmentasi menggunakan *K-means* dengan $K=3$. Dapat dilihat bahwa dengan pengolahan *K-means* diperoleh 3 kelompok kumpulan piksel yang homogen, yang memiliki intensitas dan jumlah setiap intensitas pada setiap *pixel*, seperti terlihat pada tabel 1 :

Table 1. Intensitas Pixel Setelah diproses Menggunakan *K-means*

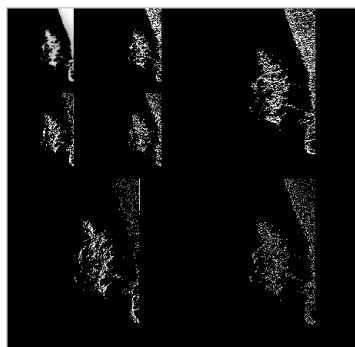
Intensitas <i>pixel</i>	Jumlah <i>pixel</i>
2	756256
134	194032
185	98288

Setelah pencarian ROI, kemudian dilakukan proses segmentasi. Hasil dari segmentasi citra berdasarkan tiga intensitas tersebut adalah seperti Gambar 4 berikut :

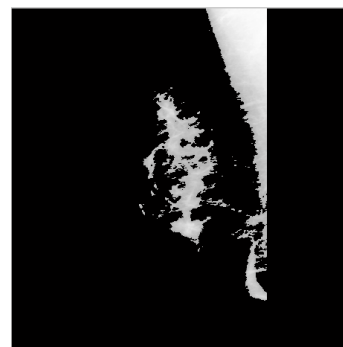


Gambar 4. Hasil segmentasi beberapa Intensitas

Berdasarkan hasil segmentasi dengan mudah dapat ditentukan daerah yang dicurigai, yaitu daerah dengan nilai intensitas tertinggi yang menunjukkan adanya kandungan kapur. Hasil proses segmentasi tersebut kemudian ditransformasikan menggunakan Wavelet Daubechies 6 sampai dengan dekomposisi level 2. Hasil tahap transformasi tampak pada Gambar 5, sedangkan hasil rekonstruksi ditunjukkan oleh Gambar 6.



Gambar 5. Dekomposisi Citra Mammogram



Gambar 6. Rekonstruksi Citra Mammogram

Tahap akhir adalah proses klasifikasi yang dilakukan dengan menggunakan metode *Support Vector Classification* (SVC). Hasil klasifikasi citra normal dan abnormal berdasarkan hasil uji coba terhadap 120 citra mammogram dapat dilihat pada Tabel 2.

Tabel 2 Persentase rata-rata hasil uji coba klasifikasi

Uji Coba	Normal(%)		Abnormal(%)	
	Benar	Salah	Benar	Salah
<i>Dense glandular – Dense glandular</i>	93	7	83,5	16,5
<i>Dense glandular – Fatty</i>	92,5	7,5	85,5	14,5
<i>Dense glandular – Fatty Glandular</i>	92	8	84	16
<i>Fatty Glandular - Fatty Glandular</i>	88	12	98,5	1,5
<i>Fatty Glandular - Dense Glandular</i>	88,5	11,5	97,5	2,5
<i>Fatty Glandular - Fatty</i>	90,5	9,5	99	1
<i>Fatty-Fatty</i>	96,5	3,5	82,5	17,5
<i>Fatty - Dense Glandular</i>	97,5	2,5	84,5	15,5
<i>Fatty - Fatty Glandular</i>	92,5	7,5	80	20
Rata-rata	91	9	87,67	12,33

Dari Tabel 2 dapat dinyatakan bahwa SVM dapat mengklasifikasikan citra mammogram dengan akurasi 91% untuk kategori normal dan 87.67 % untuk abnormal.

5. PENUTUP

Klasifikasi citra mammogram sangat penting untuk memudahkan mengidentifikasi jaringan yang dicurigai. Berdasarkan pada hasil pengujian menunjukkan bahwa SVM berhasil mengklasifikasikan dengan baik, dengan tingkat akurasi 91 % untuk kategori normal dan 87,67 untuk abnormal.

DAFTAR PUSTAKA

- [1] Anonim. 2008. *10 Facts on Cancer* dalam <http://www.who.int/features/factfiles/cancer/facts/en/index.html> akses terakhir tanggal 10 Oktober 2012.
- [2] Stojić, Tomislav dan Reljin, Branimir. “*Enhancement of Microcalcifications in Digitized Mammograms: Multifractal and Mathematical Morphology Approach*”, Belgrade : University of Belgrade, 2010.
- [3] Kulkarni, D.A., Bhagyashree.S.M, G.R.Udupi”Texture Analysis of Mammographic images” *International Journal of Computer Applications* (0975 – 8887) , Volume 5– No.6, August 2010
- [4] Manoharan, C and N.V.S.Sree Rathna Lakshmi. “Classification of Micro Calcifications in Mammogram using Combined Feature Set with SVM”. *International Journal of Computer Applications* 11(10):30–34, December 2010.
- [5] Oliveira, L, Geraldo Braz Junior, Aristófanés Correa Silva, Anselmo Cardoso de Paiva, Marcelo Gattass, “Detection of Masses in Digital Mammograms using K-Means and Support Vector Machine”, *Electronic Letter on Computer Vision and Image Analysis*, Vol 8 (2), 2009
- [6] Anonim. *Pengetahuan Dasar Tentang Kanker*. <http://www.hompedin.org/artikel1.php> akses terakhir tanggal 10 Oktober 2012.
- [7] Malagelada, A.O.I, 2007. *Automatic Mass Segmentation in Mammographics Images*, Thesis in Department of Electronics, Computer Science and Automatic Control, Universitat de Girona, Girona.S

- [8] Gonzales R.C, Woods & Eddin. 2004. *Digital Image Processing using Matlab*. Gatesmark Publishing. USA.
- [9] Dane Kurnia Putra, dkk. *Identifikasi Keberadaan Kanker Pada Citra Mammografi Menggunakan Metode Wavelet Haar*. Semarang: Universitas Diponegoro.
- [10] Diana Purwitasari, dkk. *Implementasi Adaptive Support Vector Machine untuk Membantu Identifikasi Kanker Payudara*. Surabaya: Kampus ITS.
- [11] Jain, Ramesh, "Machine Vision", McGraw-Hill, 1995.
- [12] S, Meenalosini, dkk, "A Novel Approach in Malignancy Detection of Computer Aided Diagnosis", Chennai : Vel Tech Dr. RR and Dr. SR Technical University, 2012.
- [13] Zhang, Xinsheng, dkk, "Micro Calcification Clusters Detection by Using Gaussian Markov Random Fields Representation", Xi'an : University of Architecture and Technology, 2012.