

Klasifikasi Massa pada Citra Mammogram Berdasarkan *Gray Level Cooccurrence Matrix (GLCM)*

Refta Listia*¹, Agus Harjoko²

¹Prodi S2/S3 Ilmu Komputer FMIPA UGM, Yogyakarta

²Jurusan Ilmu Komputer dan Elektronika (JIKE), Fakultas Matematika dan Ilmu Pengetahuan
Alam, Universitas Gadjah Mada, Sekip Utara, Bulaksumur, Yogyakarta

e-mail: *refta_listia@mail.ugm.ac.id, aharjoko@ugm.ac.id

Abstrak

Kanker payudara adalah penyakit yang paling umum diderita oleh wanita pada banyak negara. Pemeriksaan kanker payudara dapat dilakukan dengan menggunakan mamografi. Pada penelitian ini, pendekatan yang diusulkan bertujuan untuk mengklasifikasi mammogram berdasarkan tiga kelas yaitu kelas normal, tumor jinak, dan tumor ganas. Sistem yang diusulkan terdiri dari empat langkah utama yaitu preprosesing, segmentasi, ekstraksi fitur dan klasifikasi. Pada tahap preprosesing akan dilakukan grayscale, interpolasi, amoeba mean filter dan segmentasi. Ekstraksi ciri menggunakan Gray Level Cooccurrence Matrix (GLCM) dan akan dihitung ciri-ciri statistik pada 4 arah ($d=1$ dan $d=2$), GLCM 8 arah ($d=1$) dan GLCM 16 arah ($d=2$). Fitur yang digunakan ada 5 yaitu kontras, energi, entropi, korelasi dan homogenitas. Langkah terakhir adalah klasifikasi menggunakan Backpropagation. Beberapa parameter penting divariasikan dalam proses ini seperti learning rate dan jumlah node dalam lapisan tersembunyi. Hasil penelitian menunjukkan bahwa fitur ekstraksi GLCM 4 arah ($0^\circ, 45^\circ, 90^\circ, 135^\circ$) dengan jarak $d=1$ memiliki akurasi terbaik dalam mengklasifikasi mammogram yaitu sebesar 81,1% dan khusus pada arah 0° akurasi klasifikasi diperoleh sebesar 100%.

Kata kunci—Mammogram, GLCM, Backpropagation

Abstract

Breast cancer is the most common disease in women in many countries. Breast cancer can be performed using mammography. In this work, an approach is proposed to classify mammogram based on three classes such as normal, benign, and malignant. The proposed system consist of four major steps : preprocessing, segmentation, feature extraction and classification. In preprocessing grayscale, interpolation, amoeba mean filter and segmentation are applicated. Feature extraction using Gray level Cooccurrence Matrix (GLCM) and the features will be calculated in 4 angles ($d=1$ and $d= 2$), GLCM 8 angles and GLCM 16 angles. The 5 features are contrast, energy, entropy, correlation and homogeneity. The final step is classification using Backpropagation. Some of important parameters will be variated in this process such as learning rate and the number of node in hidden layer. The research result suggest that extraction feature in 4 angles ($0^\circ, 45^\circ, 90^\circ, 135^\circ$) and $d=1$ is the best accuracy for classifying mammogram based on classes 81,1% and especially in 0° accuracy is 100%.

Keywords— mammogram, GLCM, Backpropagation

1. PENDAHULUAN

Kanker payudara adalah salah satu penyakit yang paling ditakuti oleh kaum wanita. Salah satu metode yang paling efektif untuk mendeteksi dan mengidentifikasi kanker payudara adalah melalui pemeriksaan mammografi dengan menggunakan sinar X. Citra yang dihasilkan disebut mammogram. Mammogram digunakan sebagai *screening* untuk menemukan perubahan pada payudara, apabila pasien sendiri tidak merasakan adanya perubahan apa-apa. Selain itu, mammogram digunakan sebagai alat diagnosa apabila dokter mencurigai adanya suatu perubahan pada payudara maka pasien mungkin memerlukan analisis mammogram. Analisis terhadap mammogram ini masih dilakukan secara manual oleh dokter atau radiologis [1].

Untuk mengatasi kekurangan tersebut maka dibutuhkan sistem pengolahan citra digital. Pengolahan citra digital merupakan bidang yang berkembang pesat dan mengalami kemajuan dalam bidang kedokteran ketika ditemukannya tomografi terkomputerisasi (*computerized tomography/ CT*) pada tahun 1970-an [2]. Pengolahan citra telah banyak dilakukan khususnya di bidang kedokteran sebagai alat bantu mendiagnosis suatu penyakit termasuk untuk mendeteksi keberadaan penyakit pada citra mammogram. Hal ini dikarenakan kualitas mammogram mampu memperlihatkan struktur anatomi pada payudara sehingga kelainan dapat terdeteksi.

Beberapa penelitian sebelumnya meneliti tentang kelainan payudara pada citra mammogram menggunakan metode GLCM untuk mengekstraksi ciri yaitu [3] mengatakan fitur-fitur dengan GLCM menggunakan 4 arah ($0^\circ, 45^\circ, 90^\circ, 135^\circ$) dan jarak = 1 dapat digunakan untuk membedakan antara massa kistik dan massa non kistik meliputi citra mioma dan citra tumor padat pada citra ultrasonografi. Sedangkan penelitian yang dilakukan oleh [4] dengan tujuan untuk mengklasifikasi kanker payudara menggunakan Backpropagation. Tahap preprosesing yang dilakukan adalah konversi citra asli ke grayscale kemudian melakukan contrast adaptive histogram equalization (CLAHE), segmentasi dan morfologi. Hasil penelitian ini memiliki akurasi sebesar 98,60%. Selanjutnya penelitian yang dilakukan oleh [5] yaitu membandingkan metode ekstraksi ciri. Metode yang dibandingkan adalah intensitas histogram, GLCM, dan intensitas berdasarkan fitur. Data yang digunakan diperoleh dari *Digital Database for Screening* (DDMS). Dari hasil penelitian tersebut, ekstraksi ciri dengan menggunakan GLCM adalah metode ekstraksi yang paling baik.

Berdasarkan alasan diatas, pada penelitian ini penulis mengusulkan penelitian yang bertujuan untuk mengklasifikasi citra mammogram ke dalam 3 kelas citra yaitu kelas citra normal, citra tumor jinak dan citra tumor ganas. Tahap preprosesing yang dilakukan adalah konversi citra asli ke grayscale, interpolasi untuk *resample* citra, amoeba mean filter untuk melakukan proses filtering lebih adaptif dengan kondisi citra dan hasil yang didapatkan diharapkan dapat lebih maksimal daripada mean filter biasa [6] kemudian melakukan segmentasi dengan menentukan nilai ambang untuk memisahkan objek yang akan diteliti dengan *background*. Sedangkan metode ekstraksi ciri yang digunakan yaitu *Gray Level Co-occurrence Matrix* (GLCM) dimana akan dibandingkan 4 jenis GLCM yaitu GLCM 4 arah ($0^\circ, 45^\circ, 90^\circ$ dan 135°) jarak 1 piksel tetangga ($d=1$), GLCM jarak 2 piksel tetangga ($d=2$), GLCM 8 arah dan GLCM 16 arah. Ciri statistik (fitur) GLCM yang digunakan sebanyak 5 fitur yaitu korelasi, homogenitas, entropi, kontras dan energi. Pada GLCM 4 arah $d=1$ dan $d=2$ akan diambil ciri statistiknya pada masing-masing arah yaitu $0^\circ, 45^\circ, 90^\circ$ dan 135° sedangkan untuk GLCM 8 arah dan GLCM 16 arah akan diambil rata-rata ciri statistiknya. Nilai ciri statistik tersebut akan dijadikan input pada tahap klasifikasi. Tahap klasifikasi menggunakan metode jaringan syaraf backpropagation (JST-BP). Alasan peneliti menggunakan JST-BP dikarenakan dari beberapa metode JST yang lain, backpropagation adalah metode yang paling sesuai untuk klasifikasi. Hal ini diperkuat dengan penelitian sebelumnya yang menunjukkan bahwa dengan klasifikasi menggunakan backpropagation diperoleh hasil akurasi klasifikasi yang tinggi. Metode backpropagation merupakan metode yang sangat baik dalam menangani masalah pengenalan pola-pola kompleks [7]. Backpropagation melatih jaringan untuk mendapatkan

keseimbangan antara kemampuan jaringan untuk mengenali pola yang digunakan selama pelatihan serta kemampuan jaringan untuk memberikan respon yang benar terhadap pola masukan yang serupa dengan pola yang dipakai selama pelatihan. Cara kerja jaringan backpropagation ini yaitu gradien error unit-unit tersembunyi diturunkan dari penyiaran kembali error-error yang diasosiasikan dengan unit-unit output. Hal ini karena target untuk unit-unit tersembunyi tidak diberikan.

Backpropagation memiliki beberapa keunggulan pada segi kekonvergenan dan lokasi lokal minimumnya yang sangat peka terhadap pemilihan inisialisasi awal serta perbaikan pembobotnya dapat terus dilakukan hingga diperoleh nilai hasil yang hampir sama dengan target dimana error yang dihasilkan mendekati nol. Namun masalah utama yang dihadapi dalam backpropagation adalah lamanya iterasi yang harus dilakukan karena backpropagation tidak memberikan kepastian terhadap berapa epoch yang harus dilakukan untuk mencapai kondisi yang diinginkan [8]. Oleh karena itu, peneliti berusaha mencari parameter jaringan yang tepat agar menghasilkan jumlah iterasi yang sedikit namun menghasilkan akurasi yang baik.

Berdasarkan uraian pada latar belakang, maka dapat dirumuskan suatu permasalahan bagaimana mengklasifikasi massa pada citra mammogram, bagaimana mengesktraksi ciri menggunakan GLCM jarak 1 dan 2 piksel tetangga masing-masing pada orientasi arah $0^\circ, 45^\circ, 90^\circ$ dan 135° , GLCM 8 arah dan GLCM 16 arah, bagaimana kinerja penentuan kelas pada citra mammogram menggunakan JST-backpropagation.

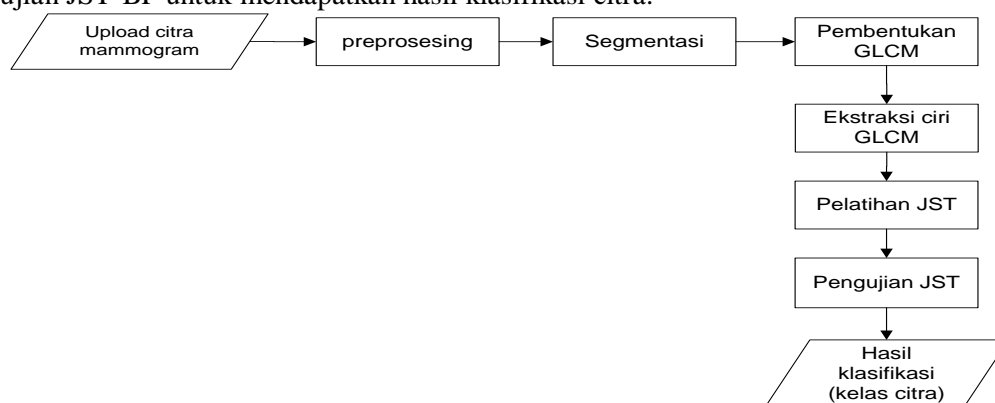
2. METODE PENELITIAN

2.1 Bahan Penelitian

Data yang digunakan dalam penelitian ini berupa citra mammogram dari database MIAS (*Mammographic Image Analysis Society*). Objek dari penelitian ini adalah kedua payudara wanita yang diambil dari posisi kanan dan kiri (RCC dan LCC) dengan ukuran 1024 piksel x 1024 piksel yang disimpan dalam bentuk bitmap. Citra yang digunakan dalam penelitian ini berjumlah 103 citra yang terdiri dari 62 data latih (20 citra normal, 20 citra tumor ganas, 22 citra tumor jinak) dan 41 citra data uji (15 citra normal, 7 citra tumor ganas, 19 citra tumor jinak). Setiap citra akan diekstraksi menggunakan *Gray Level Cooccurrence Matrix* (GLCM) sehingga untuk orientasi 4 arah ($0^\circ, 45^\circ, 90^\circ$ dan 135°) jarak $d=1$ dan $d=2$ total citra untuk pelatihan sebanyak 248 dan total citra uji sebanyak 164 citra. Sedangkan untuk ekstraksi ciri citra dengan menggunakan GLCM 8 dan 16 arah citra yang digunakan sebanyak 103 citra.

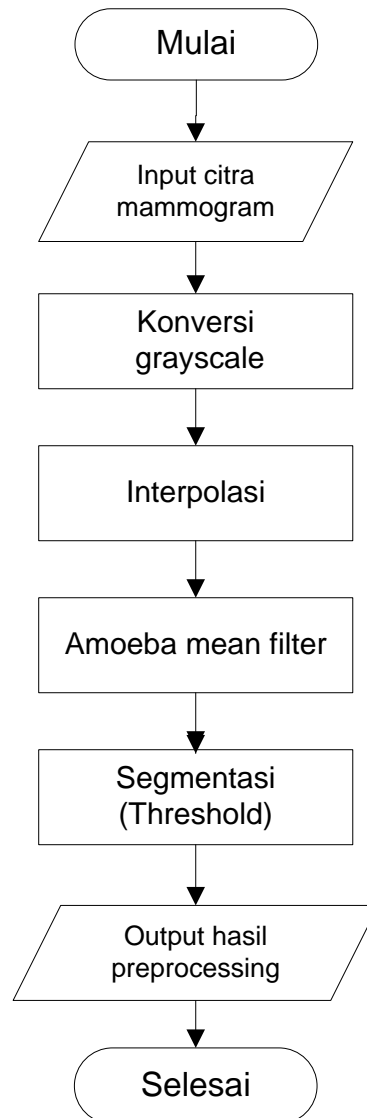
2.2 Arsitektur Sistem

Berdasarkan analisis sistem maka untuk membangun sistem klasifikasi citra dibutuhkan rancangan proses. Gambar 1 menunjukkan arsitektur sistem yang secara garis besar meliputi input citra, preprosesing, segmentasi, pembentukan GLCM, ekstraksi GLCM, pelatihan dan pengujian JST-BP untuk mendapatkan hasil klasifikasi citra.



Gambar 1 Arsitektur Sistem

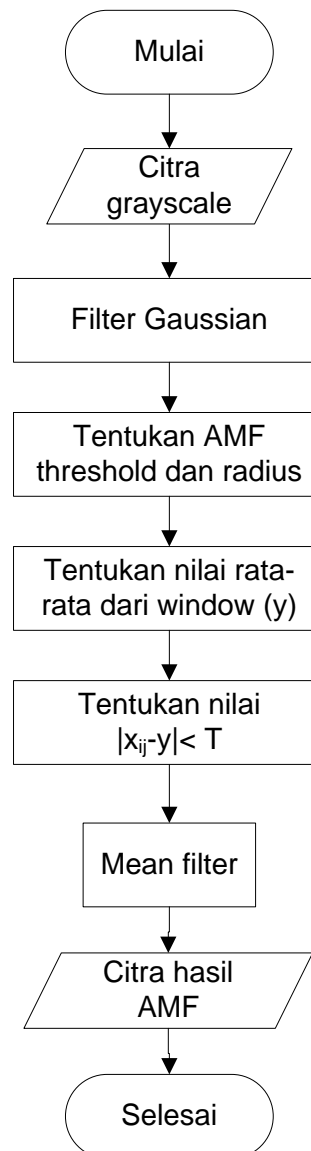
Gambar 1 menunjukkan langkah-langkah dari sistem yang dirancang. Preprocessing bertujuan untuk mendapatkan citra grayscale yang selanjutnya citra siap digunakan untuk proses selanjutnya. Tahap preprocessing yang dilakukan adalah mengkonversi citra asli menjadi citra grayscale sehingga untuk setiap citra hanya akan terdapat satu nilai keabuan. Langkah selanjutnya melakukan interpolasi dengan tujuan untuk *resample* citra kemudian menghilangkan *noise* dengan menggunakan amoeba mean filter. Gambar 2 menunjukkan diagram alir preprocessing secara umum.



Gambar 2 Preprocessing

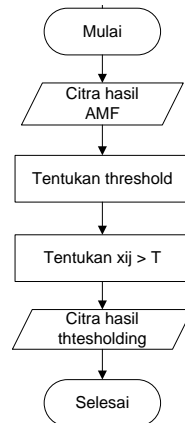
Pada tahap preprocessing setelah citra grayscale yang di interpolasi selanjutnya dilakukan penerapan amoeba mean filter dengan tujuan untuk mengurangi *noise*. Pada penelitian ini menggunakan amoeba mean filter dikarenakan metode *filtering* umumnya menggunakan bentuk kernel yang statis sehingga terkadang hasil yang didapatkan masih belum maksimal. Metode – metode baru mulai bermunculan dengan konsep kernel yang lebih adaptif dengan kondisi struktur elemen pada gambar salah satunya amoeba mean filter. AMF diadaptasi dari mean filter yang kemudian dikembangkan dan beradaptasi sesuai struktur gambar dengan fleksibilitas yang maksimum.

Metode ini terdiri dari 2 fase pengerjaan, yaitu konstruksi *amoeba* dan menerapkan mean filter sesuai dengan bentuk *amoeba* yang telah dibentuk. Dengan adanya konsep *amoeba* ini, bentuk *kernel* yang digunakan untuk melakukan proses *filtering* lebih adaptif dengan kondisi citra sehingga hasil yang didapatkan lebih maksimal daripada mean filter biasa. Pada tahap ini, langkah awal adalah mengaplikasikan Gaussian Blur yang telah tersedia pada library C#. Kemudian menentukan threshold amoeba mean filter dan radius. Radius digunakan untuk menentukan ukuran *window*. Langkah selanjutnya membaca piksel ke- ij (x_{ij}) kemudian menentukan selisih piksel x_{ij} dengan nilai rata-rata pada *window*. Apabila selisih tersebut kurang dari threshold maka nilai piksel tersebut dijadikan tetangga dan akan digunakan untuk perhitungan mean filter. Gambar 3 menunjukkan diagram alir amoeba mean filter.



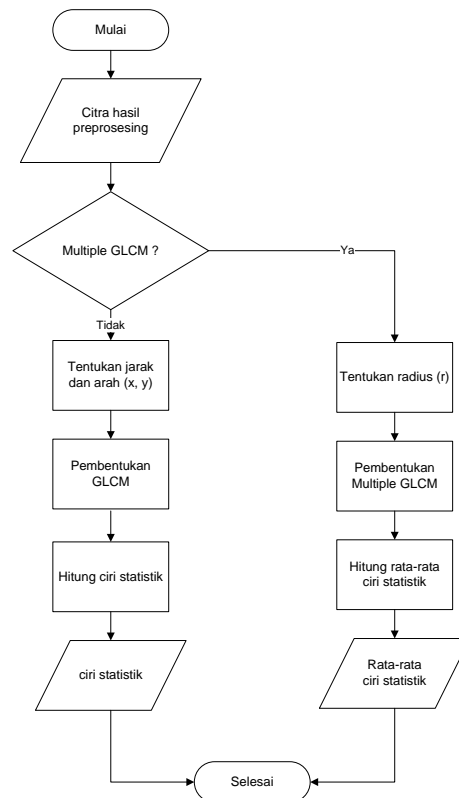
Gambar 3 Amoeba Mean Filter

Setelah menerapkan amoeba mean filter selanjutnya melakukan segmentasi. Gambar 4 menunjukkan diagram alir segmentasi.



Gambar 4 Segmentasi

Pada Gambar 4, tujuan dari segmentasi adalah untuk menentukan objek yang akan diteliti dari citra mammogram. Langkah awal segmentasi dengan menentukan nilai ambang dari citra. Apabila nilai piksel (x_{ij}) tersebut lebih kecil dari nilai ambang yang ditentukan maka diberi warna hitam dan tidak akan digunakan sebagai objek penelitian sedangkan nilai piksel yang sama atau lebih besar dari nilai threshold akan digunakan sebagai objek sehingga akan didapatkan *region of interest* (ROI) dari citra mammogram. Langkah selanjutnya adalah membentuk GLCM. Tahap pembentukan GLCM, untuk GLCM 4 arah ($0^\circ, 45^\circ, 90^\circ, 135^\circ$) $d=1$ dan $d=2$ maka akan ditentukan koordinat arah (x,y). Diagram alir pembentukan GLCM ditunjukkan pada Gambar 5. Sebagai contoh, untuk arah 0° dan jarak $d=1$ artinya koordinat (x,y) adalah (1,0). Setelah menentukan arahnya, selanjutnya membentuk matriks korelasi dengan cara menghitung frekuensi kemunculan pasangan nilai keabuan piksel referensi dan piksel tetangga pada jarak dan arah yang ditentukan.

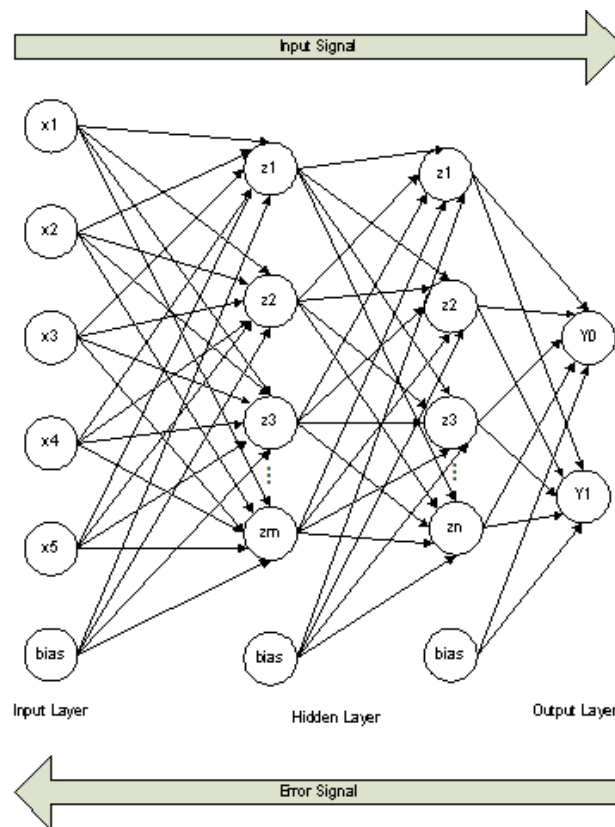


Gambar 5 Diagram Alir Pembentukan GLCM

Selanjutnya menjumlahkan semua elemen untuk menghitung probabilitas setiap elemen dengan cara membagi setiap elemen GLCM dengan total jumlah semua elemen. Langkah terakhir adalah menghitung ciri statistik GLCM yaitu kontras, energi, entropi, korelasi dan homogenitas. Gambar 3 menunjukkan pembentukan GLCM.

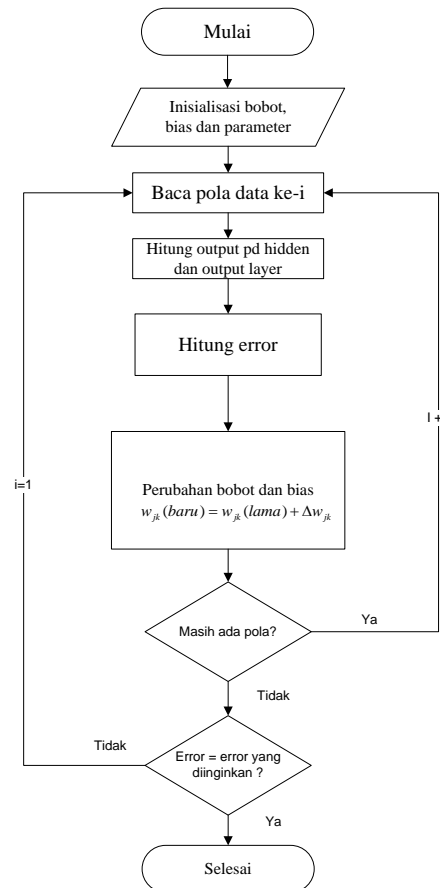
Sedangkan untuk GLCM 8 arah dan 16 arah, langkah awal adalah menentukan radius (r) sehingga jumlah arah akan terbentuk, untuk radius (r) = 1 maka jumlah arah yang akan dihitung sebanyak 8 arah sedangkan untuk radius (r) = 2 maka jumlah arah yang akan dihitung sebanyak 16 arah. Selanjutnya membentuk GLCM dengan cara menghitung frekuensi kemunculan pasangan nilai keabuan piksel referensi dan piksel tetangga pada jarak dan arah yang ditentukan. Selanjutnya menjumlahkan semua elemen untuk menghitung probabilitas setiap elemen dengan cara membagi setiap elemen GLCM dengan total jumlah semua elemen. Langkah selanjutnya menghitung ciri statistik GLCM yaitu kontras, energi, entropi, korelasi dan homogenitas untuk setiap arah yang terbentuk kemudian langkah terakhir menghitung rata-rata ciri statistik pada semua arah yang terbentuk [9].

Ciri statistik yang diperoleh dari ekstraksi ciri menggunakan GLCM selanjutnya digunakan sebagai input pada pelatihan JST-BP. Gambar 6 adalah arsitektur jaringan syaraf tiruan backpropagation yang digunakan. Pada penelitian ini, peneliti akan menggunakan layer tersembunyi dimulai dari 1 layer tersembunyi hingga 2 layer tersembunyi hingga diperoleh hasil klasifikasi yang paling baik menggunakan berapa layer tersembunyi.



Gambar 6 Arsitektur JST-Backpropagation

Selanjutnya melakukan pelatihan menggunakan JST-BP. Diagram alir pelatihan ditunjukkan pada Gambar 7. Pada tahap pelatihan dilakukan setting parameter JST yaitu *learning rate*, jumlah node pada layer tersembunyi, *sum of squared error* (SSE), dan momentum. Selanjutnya membaca data pertama kemudian menghitung nilai jaringannya ke layer tersembunyi dan menghitung nilai keluaran di layer keluaran. Langkah selanjutnya adalah menghitung error pada jaringan output, kemudian melakukan perhitungan mundur untuk menghitung error di setiap jaringan.



Gambar 7 Pelatihan JST-Backpropagation

Langkah selanjutnya adalah menghitung perubahan bobot dari layer keluaran ke layer tersembunyi (w_{kj}) dan perubahan bobot dari layer tersembunyi ke layer masukan (v_{ji}). Setelah melakukan perubahan bobot selanjutnya sistem akan melakukan pengecekan apakah data masih ada atau tidak ada. Jika masih ada data maka proses akan dilanjutkan untuk membaca data berikutnya ($i++$) sedangkan apabila tidak ada data maka dilakukan pengecekan terhadap error dengan cara menghitung nilai SSE, apabila SSE hitung $<$ SSE input maka proses dihentikan, jika tidak maka proses akan diteruskan untuk membaca data pertama lagi ($i=1$) hingga syarat SSE hitung $<$ SSE input terpenuhi.

Pada tahap pengujian hanya melakukan fase maju dimana akan dilakukan perhitungan semua keluaran di layer tersembunyi dan di layer keluaran. Langkah awal adalah meng-upload citra yang akan diuji, kemudian *load* parameter dan bobot yang sudah tersimpan. Selanjutnya jaringan akan menghitung output tiap layer. Apabila hasil output sesuai dengan target artinya klasifikasi benar sedangkan jika tidak sesuai target maka klasifikasi masih salah.

3. HASIL DAN PEMBAHASAN

Sebelum melakukan pengujian, dilakukan beberapa pelatihan dengan memvariasikan beberapa parameter, baik parameter pada tahap preprosesing maupun pada tahap klasifikasi. Dari beberapa pengujian yang telah dilakukan dapat disimpulkan bahwa jaringan mampu mengklasifikasi citra sesuai dengan kelasnya yaitu pada GLCM $d=1$ (arah 0°) dengan $T = 120$ dan $T_{AMF} = 25$ yaitu sebesar 100% dan pada GLCM $d=2$ (arah 90°) dengan $T = 120$ dan $T_{AMF} = 25$ yaitu sebesar 75,61%. Oleh karena itu dibutuhkan pengujian sensitivitas dan spesifisitas [10] untuk melihat kemampuan jaringan yang ditunjukkan pada Tabel 1.

Tabel 1 Pengujian Sensitivitas dan Spesivitas

Citra GLCM	Arah	Sensitivitas	Spesifisitas	Akurasi
d=1 $\alpha = 0,001$	0°	100%	100%	100%
d=2 $\alpha = 0,005$	90°	60%	86,9%	75,61%

Berdasarkan Tabel 1, citra d=1 dengan arah 0° memiliki sensitivitas, spesivitas dan akurasi sebesar 100%. Hal ini menunjukkan jaringan mampu mengenali semua kelas citra. Sedangkan pada citra d=2 dengan arah 90° memiliki sensitivitas, spesivitas dan akurasi masing-masing sebesar 60%, 86,9% dan 75,61%, artinya jaringan masih melakukan kesalahan dalam mengenali kelas normal sebagai kelas sebenarnya dan kelas abnormal (tumor ganas dan tumor jinak) sebagai kelas sebenarnya.

Dari penjelasan diatas dan untuk penelitian ini maka dapat disimpulkan bahwa klasifikasi dengan menggunakan metode ekstraksi ciri GLCM dengan jarak d=1 arah 0° dimana menggunakan 2 layer tersembunyi dan jumlah node sebanyak 19 node pada layer tersembunyi, *learning rate*=0,001 lebih baik dibandingkan dengan GLCM 8 arah d=1, dan jarak d=2 [11].

4. KESIMPULAN

1. Klasifikasi dengan menggunakan metode ekstraksi ciri GLCM 4 arah dan d=1 dimana menggunakan 2 layer tersembunyi dengan 19 node pada layer tersembunyi, *learning rate* (α) = 0,001 memiliki rata-rata akurasi sebesar 81,1%.
2. Klasifikasi dengan menggunakan metode ekstraksi ciri GLCM dengan jarak d=1 dimana menggunakan 2 layer tersembunyi dengan 19 node pada layer tersembunyi, *learning rate* (α) = 0,001 memiliki sensitivitas, spesifisitas dan akurasi sebesar 100% yaitu pada orientasi arah 0°.
3. Metode esktraksi ciri GLCM 4 arah dengan d=1 mampu mengklasifikasi citra mammogram ke dalam kelas normal, tumor ganas dan tumor jinak.

5. SARAN

1. Preprosesing dengan menggunakan metode morfolgi karena citra abnormal dapat dikenali dari bentuknya.
2. Pengujian dengan jumlah data citra yang lebih banyak dapat melihat kemampuan sistem dalam mengklasifikasi kelas citra.

DAFTAR PUSTAKA

- [1] Sjamsuhidajat, R. dan Jong, W., 1997, Buku ajar Ilmu Bedah. Kedokteran EGC, Jakarta.
- [2] Putra, D., 2009, Pengolahan Citra Digital, Penerbit Andi, Yogyakarta.
- [3] Wibawanto, H., 2011, Analisis Tekstur untuk Diskriminasi Massa Kistik dan Non Kistik pada Citra Ultrasonografi. *Disertasi*. Fakultas Ilmu Teknik UGM.

-
- [4] Singh, S., Guptaz, P. R., 2011, Breast Cancer detection and Classification using Neural Network, *International Journal Of Advanced Engineering Sciences And Technologies*. Vol No.6 Issue No. 1, 004 – 009.
- [5] Nithya,R. and Santi, B., 2011, Comparative Study on Feature Extraction Method for Breast Cancer Classification, *J. Theoretical and Applied Information Technology*, Vol. 33 No.2.
- [6] Lerallut, R., et al. 2005. *Image Filtering Using Morphological Amoebas*. Centre de Morphologie Mathématique, École des Mines de Paris.
- [7] Puspitaningrum, D. 2006, Pengantar Jaringan Saraf Tiruan. Penerbit Andi. Yogyakarta.
- [8] Siang, J.J, 2004, Jaringan Syaraf Tiruan dan Pemrogramannya. Penerbit Andi. Yogyakarta.
- [9] Gonzales, R. C., Woods, R. E., 2008, *Digital Image Processing, 3rd ed*, Prentice Hall, Upper Sadle River, New Jersey, USA
- [10] Garcia, M., Sanchez, C. I., Poza, J., Lopez, M. I., 2009, Hornero, R., Detection of Hard Exudates in Retinal Images Using a Radial Basis Function Classifier., *Journals of Biomedical Engineering*, No. 7, Vol. 37, 1448-1463.
- [11] Listia, R., 2013, Klasifikasi Massa pada Citra Mammogram Berdasarkan *Gray Level Cooccurrence Matrix* (GLCM). *Tesis*. Jurusan Ilmu Komputer FMIPA UGM, Yogyakarta.