

Penerapan Forward Selection Pada Support Vector Machine Untuk Klasifikasi Kanker Payudara

Prihananto Joko Tri Laksono

masaan.dinus06@gmail.com

<http://www.prihanantojoko.net>

Lisensi Dokumen:

Seluruh dokumen di iTSc RSUP Dr. Kariadi dapat digunakan, dimodifikasi dan disebarakan secara bebas untuk tujuan bukan komersial (nonprofit), dengan syarat tidak menghapus atau merubah atribut penulis dan pernyataan copyright yang disertakan dalam setiap dokumen. Tidak diperbolehkan melakukan penulisan ulang, kecuali mendapatkan ijin terlebih dahulu dari iTSc RSUP Dr. Kariadi.

Kanker payudara adalah kanker yang paling sering didiagnosis dan merupakan penyebab utama kematian pada wanita di seluruh dunia, sekitar 23% (1,38 juta) total kasus kanker payudara baru dan 14% (458.400) dari total kematian akibat kanker pada tahun 2008. Keakuratan ahli radiologi dalam melakukan klasifikasi hasil citra mammogram termasuk tumor jinak (*benign*) atau tumor ganas/ kanker (*malignant*) bervariasi, antara 65%-85%, ini dipengaruhi oleh beberapa faktor, misalnya kualitas citra yang kurang baik dan kelelahan mata. Untuk mendukung keputusan ahli radiologi dalam melakukan klasifikasi jenis tumor pada payudara, maka digunakan metode klasifikasi *machine learning* atau *datamining*. *Support Vector Machine (SVM)* adalah salah satu metode klasifikasi yang sering digunakan, dengan akurasi klasifikasi yang tinggi dan dapat dengan baik melakukan klasifikasi pada data yang tidak terpisah secara linier. Namun, fitur- fitur pada dataset yang digunakan tidak semuanya memiliki kontribusi terhadap peningkatan akurasi klasifikasi, oleh karena itu fitur- fitur pada dataset harus diseleksi menggunakan *forward feature selection*, sehingga fitur yang digunakan pada proses klasifikasi hanya fitur- fitur yang relevan dan memiliki kontribusi terhadap peningkatan akurasi klasifikasi. Penerapan seleksi fitur *forward selection* pada SVM terbukti berpengaruh terhadap peningkatan akurasi klasifikasi yang signifikan, peningkatan klasifikasi rata- rata sebesar 10.44%. Tanpa penerapan *forward selection* pada SVM untuk klasifikasi dataset MIAS akurasi rata- rata tercatat hanya 77.14%, namun setelah penerapan *forward selection* pada klasifier SVM, akurasi rata-rata meningkat menjadi 87.57%. Penerapan FS-SVM pada klasifikasi dataset WBCD memiliki akurasi yang lebih baik, peningkatan rata- rata adalah 2% daripada LS- SVM pada porsi percobaan *training-testing* : 80-20, 70-30 dan 50- 50.

Keyword : Kanker Payudara, mammogram, benign, malignant, SVM, Forward Feature Selection

1. Pendahuluan

Metode SVM telah diterima baik oleh peneliti komputer beberapa tahun terakhir, dan digunakan pada banyak aplikasi. Pada penelitian yang dilakukan oleh Polat , penerapan metode SVM untuk klasifikasi *Winconsin Breast Cancer Dataset (WBCD)*, menunjukkan akurasi klasifikasi yang tinggi, yaitu 98.53 %.

Namun SVM memiliki beberapa kelemahan, salah satunya adalah bagaimana memilih fitur untuk input yang optimal. Pada dasarnya SVM adalah bekerja pada data yang terpisah secara linier, ketidak linieran data menghasilkan *hyperplane* SVM yang kurang optimal, *hyperplane* yang optimal dapat ditentukan dengan mencari nilai margin yang paling maksimal. Besar nilai margin ditentukan oleh jarak data antar kelas yang berbeda, semakin jauh jarak data antar kelas yang berbeda, maka akan semakin besar atau maksimal nilai margin. Dengan kata lain semakin besar nilai margin, maka semakin maksimal keterpisahan data.

Memaksimalkan keterpisahan data antar kelas, sangat mempengaruhi akurasi klasifikasi, caranya dapat dilakukan dengan pemilihan fitur data yang relevan untuk SVM. Salah satu teknik dapat dilakukan dengan teknik Seleksi Fitur (SF). SF sangat banyak digunakan untuk reduksi dimensi dan menghilangkan fitur yang tidak relevan. Hasil dari proses SF adalah subset, yang berisi sedikit jumlah fitur dari dataset yang memiliki kontribusi terhadap akurasi klasifikasi.

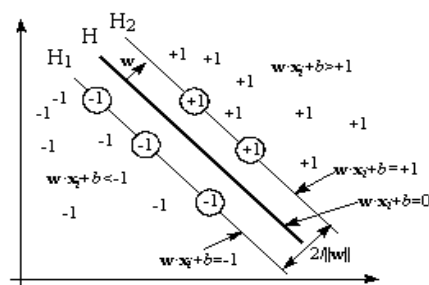
Salah satu metode SF yang banyak diusulkan peneliti dalam data mining adalah *forward selection*. *Forward selection* telah teruji sebagai metode pencarian yang efektif untuk masalah dimensi fitur dataset yang besar. Pada penelitian yang dilakukan Reif menunjukkan akurasi klasifikasi meningkat di angka 98% saat *forward selection* diterapkan, penelitian Miche menunjukkan peningkatan akurasi klasifikasi di angka 95% dan penelitian Wang menunjukkan peningkatan akurasi klasifikasi di angka 86%.

Pada penelitian ini, model yang diusulkan untuk proses klasifikasi adalah SVM dan *forward selection* untuk seleksi fitur pada pemilihan input yang tepat untuk SVM.

1.1. Support Vector Machine

Support Vector Machine (SVM) adalah suatu teknik yang relatif baru (1995) untuk melakukan prediksi, baik dalam kasus klasifikasi, yang sangat populer belakangan ini. Klasifikasi adalah pengelompokan beberapa obyek kedalam suatu kategori yang telah ditentukan, yang disebut dengan kelas. Yang, mendefinisikan SVM sebagai algoritma klasifikasi bertujuan untuk mengetahui fungsi pemetaan antara input fitur x dan nilai keanggotaan dalam sebuah kelas. Input pada proses klasifikasi disebut dengan fitur, disebut fitur karena mereka ditentukan sehingga mereka mewakili masing-masing kelas atau data yang dimiliki oleh kelas yang berbeda yang dipisahkan ke dalam suatu ruang input.

Ide dasar dari SVM adalah bagaimana mencari fungsi garis pemisah (*hyperplane*) yang mampu memisahkan antara dua kelas secara optimal. Optimal disini memiliki arti bahwa *hyperplane* mampu memisahkan kedua kelas dengan *margin* yang maksimal, sehingga biasa disebut dengan istilah *Optimal Hyperplane*. *Margin* adalah jarak antara garis *hyperplane* dengan anggota-anggota terdekat dari kedua kelas. Ilustrasi ditunjukkan pada gambar 1.1 di bawah ini.



Gambar 1.1 : Ilustrasi Support Vector Machine. Sumber : Optimum Separation Hyperplane.
<http://www.support-vector-machines.org/>

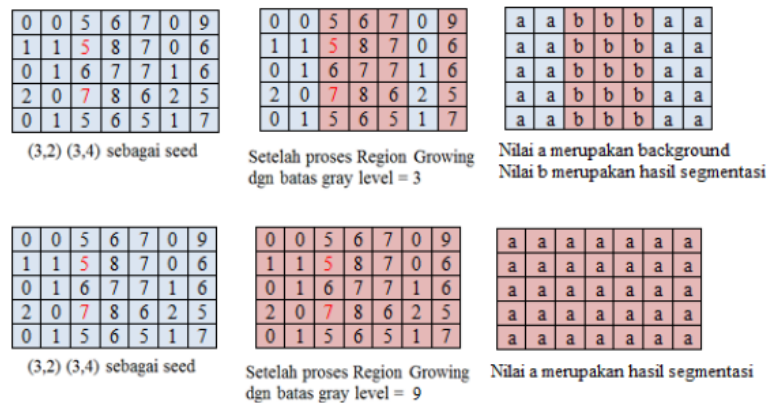
Pada gambar 1.1 di atas *hyperplane* ditunjukkan oleh H, H1 dan H2. H adalah *hyperplane* pemisah 2 (dua) kelas data yang berbeda, kelas +1 dan kelas -1, yang memenuhi syarat persamaan $wx+b=0$. H1 adalah *hyperplane* yang bersinggungan langsung dengan data- data yang ada pada kelas -1, yang memenuhi persamaan $wx+b=-1$, sedangkan H2 adalah *hyperplane* yang bersinggungan langsung dengan data- data pada kelas +1, yang memenuhi persamaan $wx+b=+1$. Data- data yang bersinggungan langsung dengan H1 pada kelas -1 dan H2 pada kelas +1 disebut dengan *support vector*, pada gambar 1.1 di atas ditandai dengan data yang dilingkari.

1.2. Feature Selection

Feature Selection adalah suatu kegiatan yang umumnya bisa dilakukan secara preprocessing dan bertujuan untuk memilih fitur yang berpengaruh dan mengesampingkan fitur yang tidak berpengaruh dalam suatu kegiatan pemodelan atau penganalisaan data. Ada banyak alternatif yang bisa digunakan dan harus dicoba-coba untuk mencari yang cocok. Secara garis besar ada dua kelompok besar dalam pelaksanaan feature selection: Ranking Selection dan Subset Selection.

1.3. Seed Region Growing

Seed Region Growing (SRG) merupakan metode segmentasi citra yang menggunakan teknik berbasis region, piksel yang berdekatan pada daerah yang sama memiliki fitur visual yang sama seperti level keabuan, nilai warna, atau teksturnya. Prinsip dari SRG adalah dengan memulainya dengan penentuan *seed points* (lokasi titik awal). Dari *seed* tersebut akan dikembangkan area-area dengan penambahan terhadap setiap *seed* dimana piksel tetangga memiliki kemiripan karakteristik dengan *seed*. Jika *seed* telah diketahui, SRG akan mencoba menemukan segmentasi citra yang akurat ke dalam daerah dengan properti dimana setiap komponen daerah yang saling berhubungan bertemu dengan salah satu.



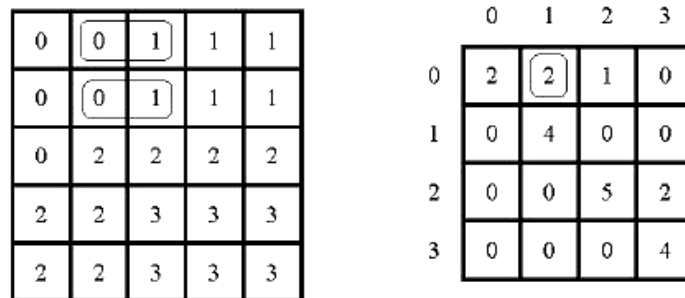
Gambar 1.2. : Ilustrasi Proses Region Growing

1.4. Gray Level Co-occurrence Matrix (GLCM)

Salah satu metode analisis tekstur yang paling banyak digunakan adalah *gray level*

co-occurrence matrix yang didasarkan pada fungsi statistika orde kedua. *Co-occurrence matrix* ini diperkenalkan pertama kali oleh Haralick untuk mengekstrak fitur-fitur yang digunakan sebagai analisis citra hasil penginderaan

jauh. *Co-occurrence* didefinisikan sebagai distribusi gabungan dari tingkat keabuan (*gray level*) dua piksel yang terpisah jarak dan arah tertentu ($\Delta x, \Delta y$). Metode GLCM adalah salah satu cara mengekstrak fitur tekstur statistik orde-kedua, GLCM (yang disebut juga *Grey Tone Spatial Dependency Matrix*) adalah tabulasi mengenai frekuensi atau seberapa seringnya kombinasi nilai kecerahan piksel yang berbeda posisinya terjadi dalam suatu citra.



Gambar 1.3 : Ilustrasi pembentukan GLCM atas citra dengan 4 tingkat keabuan (*gray level*). Sumber : Analisa Tekstur Untuk Membedakan Kista Dan Tumor Pada Citra Panoramik Rahang Gigi Manusia, ITS

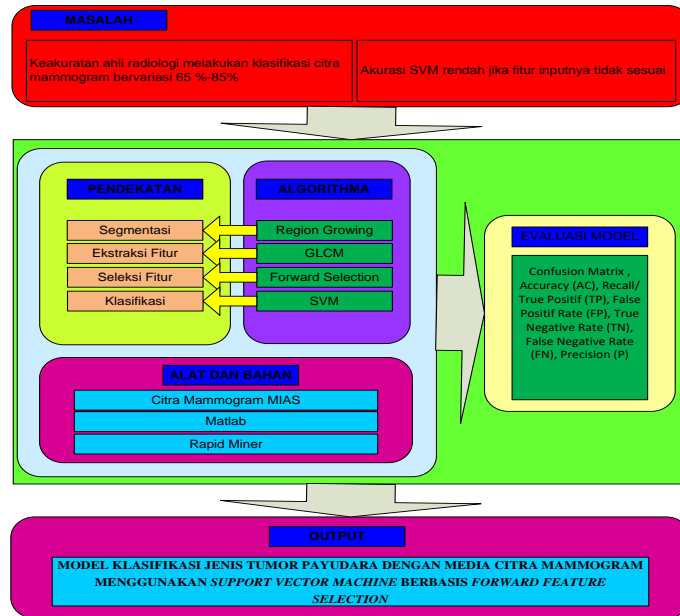
Matriks GLCM mampu menangkap sifat tekstur tetapi tidak secara langsung dapat digunakan sebagai alat analisis, misalnya membandingkan dua tekstur. Data ini harus disarikan lagi agar didapatkan angka-angka yang bisa digunakan untuk mengklasifikasi tekstur, Haralick pada tahun 1973 mengusulkan 13 ukuran atau fitur, yaitu *Homogeneity (Angular Second Moment), Contrast, Correlation, Sum of Squares, Inverse Difference Moment, Sum Average, Sum Variance, Sum Entropy, Entropy, Difference Variance, Difference Entropy, Information Measures of Correlation 1, Information Measures of Correlation 2*. Variabel- variabel tersebut di notasikan seperti di bawah ini :

$p(i,j)$ adalah masukan *gray-tone spatial dependence matrix*, $=P(i,j)/R$

$p_x(i)$ adalah masukan ke i di dalam *marginal- probability matrix* yang diperoleh dengan menjumlahkan baris $p(i,j)$

N_g adalah jumlah *gray level* di dalam citra

1.5. Kerangka Pemikiran



2. Teori Dan Hipotesis

Memaksimalkan keterpisahan data antar kelas, sangat mempengaruhi akurasi klasifikasi, caranya dapat dilakukan dengan pemilihan fitur data yang relevan untuk SVM. Salah satu teknik dapat dilakukan dengan teknik seleksi fitur (SF). SF sangat banyak digunakan untuk reduksi dimensi dan menghilangkan fitur yang tidak relevan atau tidak memiliki kontribusi terhadap peningkatan akurasi. Pada dataset, tidak semua fitur berperan dalam peningkatan akurasi klasifikasi, beberapa fitur yang tidak relevan akan menjadi *noise* yang mengganggu kerja klasifier dalam melakukan klasifikasi, sehingga akurasi klasifikasi menjadi menjadi rendah. Fitur- fitur yang menjadi menjadi *noise* tersebut harus dihilangkan sebelum dataset digunakan untuk proses klasifikasi. Seleksi fitur akan menghilangkan fitur- fitur yang tidak memiliki kontribusi dalam peningkatan akurasi klasifier dan akan mempertahankan fitur yang memiliki kontribusi terhadap peningkatan akurasi klasifier. Selanjutnya data dengan fitur- fitur terpilih dari proses seleksi tersebut akan disimpan ke dalam subset. Subset inilah yang akan digunakan untuk proses klasifikasi.

Tabel 2.1 : Penelitian Terkait Seleksi Fitur

No	Judul Penelitian	Peneliti	Tahun	Metode	Akurasi (Tanpa FS)	Akurasi (Dengan FS)
1	Advantages of Using Feature Selection Techniques on Steganalysis Schemes	Yoan Miche	2007	Feature Selection -kNN	86.65%	93.20%

2	Feature Selection for Human Resource Selection Based on Affinity Propagation and SVM Sensitivity Analysis	Qiangwei Wang	2009	Feature Selection-SVM	84.98%	86.27%
---	---	---------------	------	-----------------------	--------	--------

Salah satu metode SF yang banyak diusulkan peneliti dalam data mining adalah *forward selection*. *Forward selection* telah teruji sebagai metode pencarian yang efektif untuk masalah dimensi fitur dataset yang besar. Pada penelitian yang dilakukan Reif menunjukkan akurasi klasifikasi meningkat di angka 98% saat *forward selection* diterapkan, penelitian Miche menunjukkan peningkatan akurasi klasifikasi di angka 95 % dan penelitian Wang menunjukkan peningkatan akurasi klasifikasi di angka 86 %.

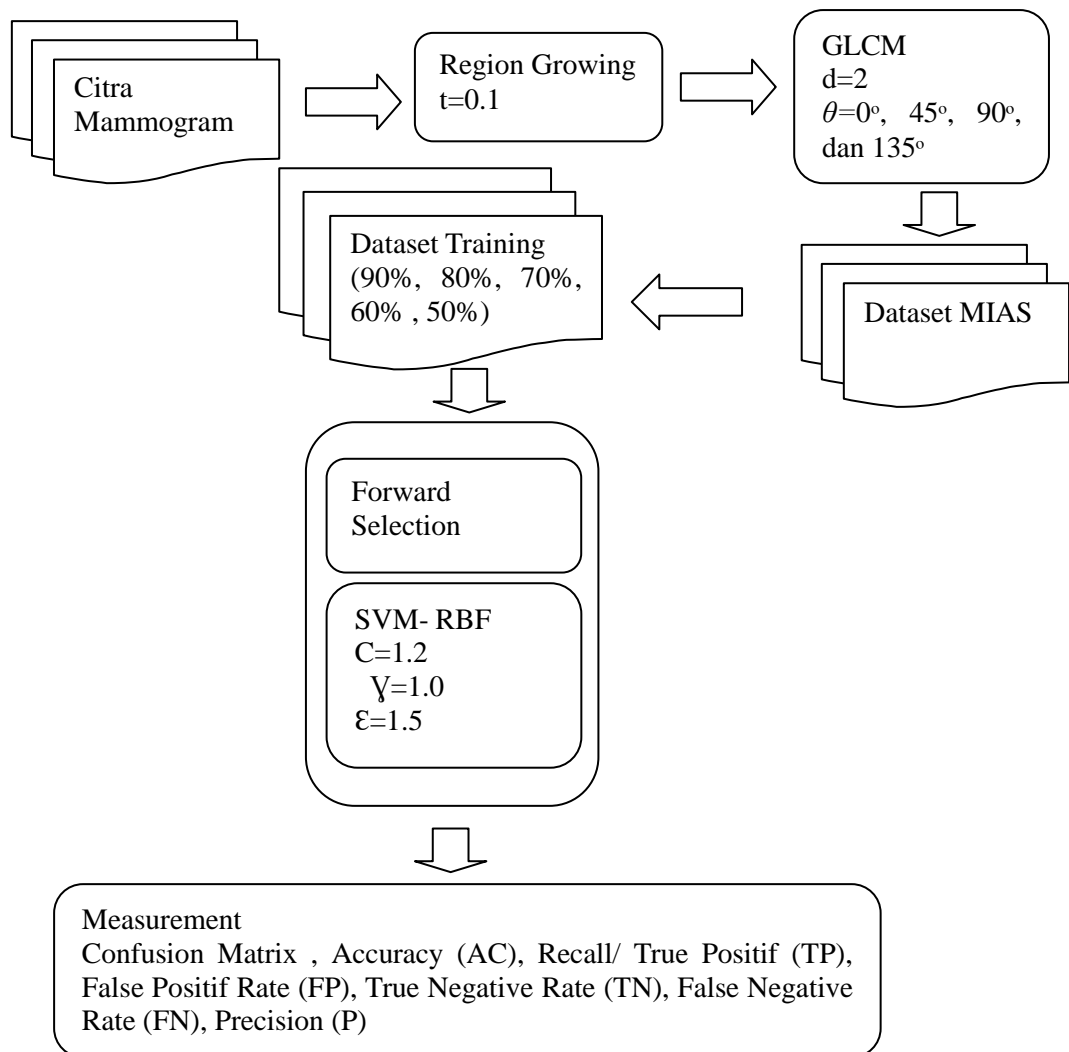
Penulis memiliki keyakinan bahwa pengaplikasian *feature selection* pada SVM dapat meningkatkan akurasi klasifikasi, karena hanya atribut- atribut pada dataset yang berpengaruh terhadap klasifikasi saja yang akan digunakan, sedangkan atribut yang tidak signifikan atau atribut yang menyebabkan data tidak terpisah secara linier akan dibuang, sehingga akurasi klasifikasi SVM akan meningkat.

3. METODE PENELITIAN

Pada penelitian ini, peneliti menggunakan data berupa citra mammogram digital yang bersumber dari MIAS, dengan ukuran citra 1024 x 1024 pixel, dan dilakukan digitasi pada 200 mikron. Data sebanyak 67 citra mammogram digital, dengan rincian kasus tumor jinak sejumlah 42 data dan kasus tumor ganas (kanker) sebanyak 25 data. Citra mammogram tersebut kemudian diubah menjadi sebuah dataset dengan cara mengekstraksi fiturnya menggunakan metode GLCM, sehingga diperoleh beberapa atribut, yang selanjutnya data tersebut akan disebut sebagai "Dataset MIAS" Informasi detail Dataset MIAS akan digunakan :

- Jumlah Data : 67 citra mammogram, terdiri dari 42 kasus tumor ganas (*malignant*) dan 25 kasus tumor jinak (*benign*)
- Jumlah Atribut : 13 atribut, yaitu Homogeneity (Angular Second Moment), Contrast, Correlation, Sum of Squares, Inverse Diference Moment, Sum Average, Sum Variance, Sum Entropy, Entropy, Diference Variance, Difference Entropy, Information Measures of Correlation 1, Information Measures of Correlation 2.
- Jumlah Class (Label) : 1 class, yaitu Severity, dengan nilai B menunjukkan *benign* dan M menunjukkan *malignant*.

4. Tahapan Eksperimen



Gambar 4.1 : Tahapan Eksperimen

Pada tahap segmentasi metode yang digunakan adalah *region growing*. Nilai parameter t yang digunakan adalah 0.1, dihasilkan dari *trial* dan *error* sehingga ditemukan *threshold* yang ideal.

Pada tahap ekstraksi fitur menggunakan gray level co-occurrence matrix (GLCM), akan menghasilkan atribut : Homogeneity (Angular Second Moment), Contrast, Correlation, Sum of Squares, Inverse Diference Moment, Sum Average, Sum Variance, Sum Entropy, Entropy, Difference Variance, Difference Entropy, Information Measures of Correlation 1, Information Measures of Correlation 2. Nilai

pada parameter distance d untuk menentukan jarak pixel tetangga adalah 2, nilai 2 merupakan jarak yang ideal untuk membentuk GLCM. Sedangkan sudut θ yang digunakan adalah 0° , 45° , 90° , dan 135° , yang kemudian hasil GLCM dari beberapa variasi sudut θ akan dihitung rata-ratanya.

Metode klasifikasi yang digunakan dalam penelitian ini adalah *Support Vector Machine* (SVM) dengan fungsi kernel *Radial Basis Function* (RBF). Pada SVM terdapat beberapa parameter yang harus ditentukan nilainya, yaitu $C=1.2$, $\gamma=1.0$ dan $\epsilon=1.5$, nilai ketiga parameter tersebut dihasilkan dari proses seleksi parameter menggunakan *Rapid Miner* sehingga ditemukan nilai-nilai parameter yang ideal.

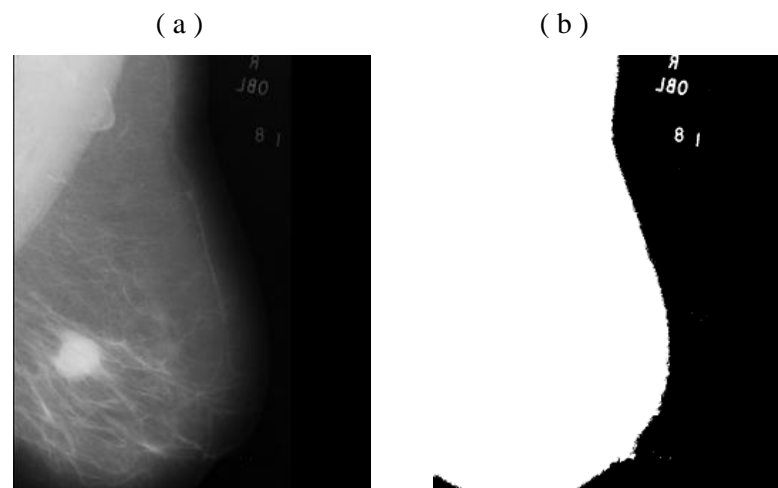
4.1. Segmentasi Mammogram

Segmentasi citra mammogram merupakan tahap dari pre prosesing, yang mana terdapat tiga tahap segmentasi pada proses ini. Segmentasi pertama adalah segmentasi untuk membuang area-area yang bukan merupakan area payudara seperti label-label pada citra mammogram. Segmentasi kedua bertujuan untuk membuang otot pektoral atau otot dada. Segmentasi ketiga bertujuan untuk mendeteksi area massa tumor pada citra mammogram. Massa adalah gumpalan dari pertumbuhan sel yang berlebihan terdiri jinak dan ganas. Pada citra mammogram massa memiliki kontras yang lebih tinggi daripada area-area lainnya dan berbentuk cenderung bulat dengan diameter yang berbeda-beda. Alat yang digunakan untuk melakukan proses segmentasi adalah MATLAB.

a. Segmentasi Area Payudara

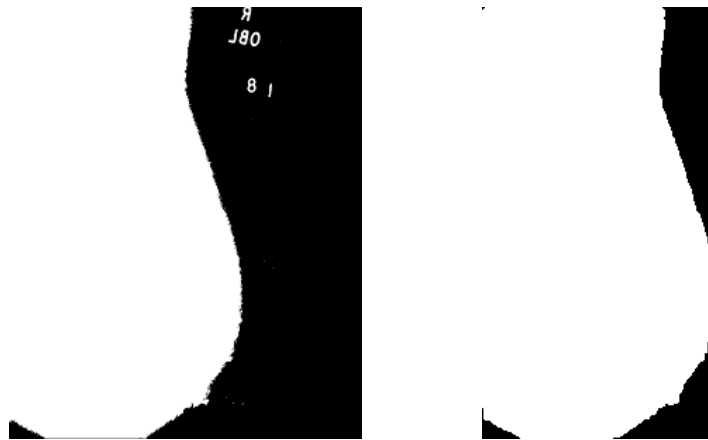
Tujuan dari proses ini adalah melakukan segmentasi area payudara dan membuang area-area pada citra mammogram yang tidak diperlukan, seperti label atau identitas pada citra mammogram. Metode yang digunakan adalah binerisasi citra. Algoritma :

1. Konversi citra ke dalam bentuk binari menggunakan fungsi *im2bw* dengan *level threshold* 0.1, sehingga menjadi citra biner.



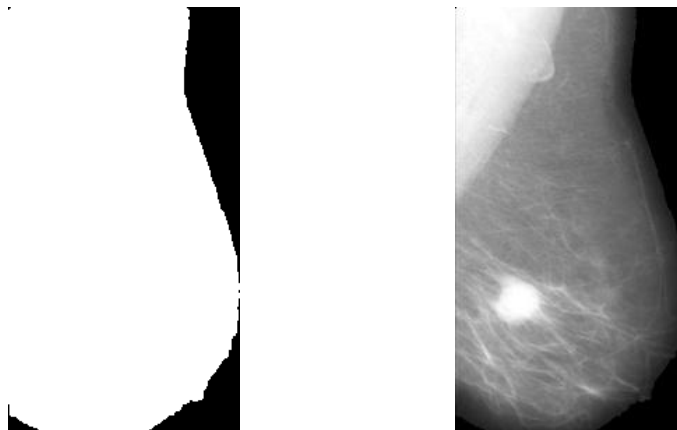
Gambar 4.2.: Gambar (a) citra mammogram asli, (b) hasil binerisasi

2. Menandai citra biner payudara dan background pada citra menggunakan fungsi *bwlabel* . Pada proses ini citra masih bercampur antara area payudara dan area- area yang bukan merupakan area payudara, pada step selanjutnya akan dibuang area- area tersebut sehingga hanya akan tersisa area payudara.
3. Pencarian area payudara dari citra biner yang telah ditemukan pada step kedua dengan menghitung luas area- area pada citra menggunakan fungsi *regionprops*, dan mencari nilai yang terbesar dari hasil kalkulasi.
4. Ekstraksi area terbesar yang ditemukan menggunakan fungsi *bwareaopen*. Fungsi ini akan menghilangkan semua area, kecuali area yang terbesar, yaitu area payudara.



Gambar 4.3 : Gambar (a) citra biner mammogram, (b) segmen area payudara

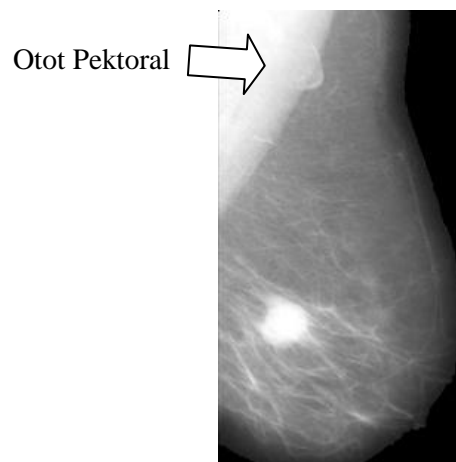
5. Operasi *morphological* untuk mengurangi distorsi pada hasil ekstraksi, fungsi yang digunakan adalah *bwmorph* dengan parameter *operation 'clean'*. Kemudian lakukan operasi *morphological* lagi menggunakan *bwmorph* dengan parameter *operation 'majority'* dan *n* dengan nilai 3, operasi ini akan menghaluskan tepi citra.
6. Menutup lubang- lubang yang mungkin terdapat pada citra biner menggunakan fungsi *imholes*.
7. Proses *masking* antara citra segmen payudara dan citra asli mammogram untuk mendapatkan citra yang tersegmentasi.



(a) (b)
Gambar 4.1: Gambar (a) citra segmen payudara, (b) hasil segmentasi

b. Segmentasi Untuk Membuang Otot Pektoral (Pectoral Muscle)

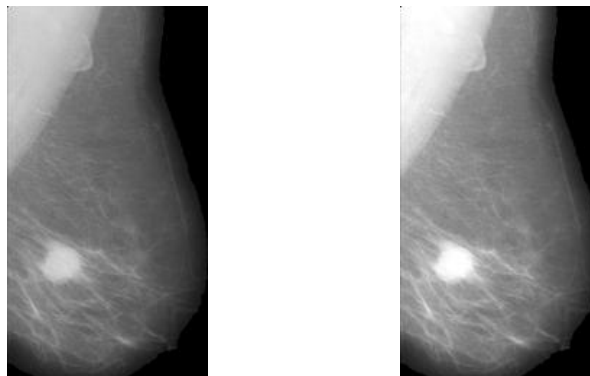
Otot pektoral adalah otot dada, pada citra mammogram otot ini memiliki warna yang cenderung kontras. Lokasi otot pektoral ditunjukkan pada gambar 4.5 di bawah ini :



Gambar 4.5 : Lokasi Otot Pektoral

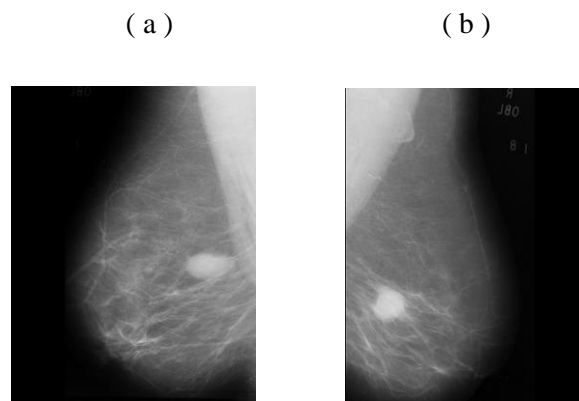
Hasil segmentasi tahap awal menghasilkan citra area payudara yang masih terdapat area otot pektoral. Bagian otot pektoral jika tidak dihilangkan akan sangat mempengaruhi ketepatan segmentasi massa tumor karena memiliki kemiripan warna dengan sel tumor pada citra mammogram, karena itu harus dihilangkan untuk meminimalkan kesalahan segmentasi area masa tumor. Metode yang digunakan adalah *Region Growing*. Algoritma :

1. Perbaiki kontras pada citra mammogram menggunakan fungsi *imadjust* dan fungsi *stretchlim* untuk mencari nilai *threshold* yang akan diaplikasikan ke dalam fungsi *imadjust*. Perbaikan kontras ditujukan untuk memperjelas area otot pektoral



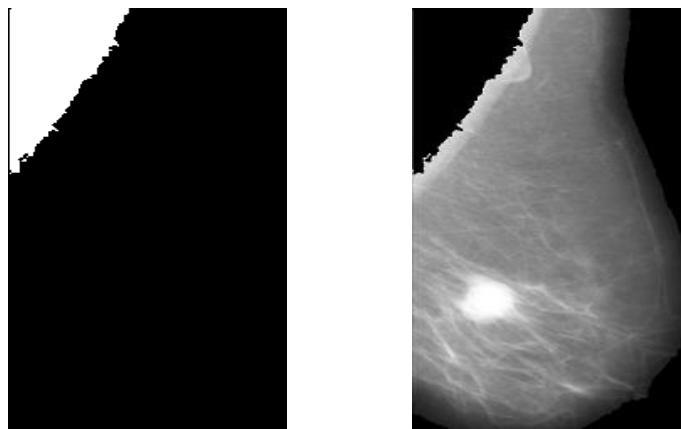
(a) (b)
Gambar 4.6 : Gambar (a) citra mammogram sebelum perbaikan kontras, (b) setelah perbaikan contrast

2. Mencari orientasi payudara, apakah payudara mengarah ke kanan atau ke kiri, dengan cara menghitung penjumlahan 5 pixel terkiri ke bawah (PL), dan 5 pixel terkanan ke bawah (PR). Jika nilai $PR > PL$ maka payudara menghadap ke kiri, jika $PR < PL$ maka payudara menghadap ke kanan. Step ini dilakukan untuk menentukan lokasi pixel sebagai *initial seed* pada *Region Growing*.



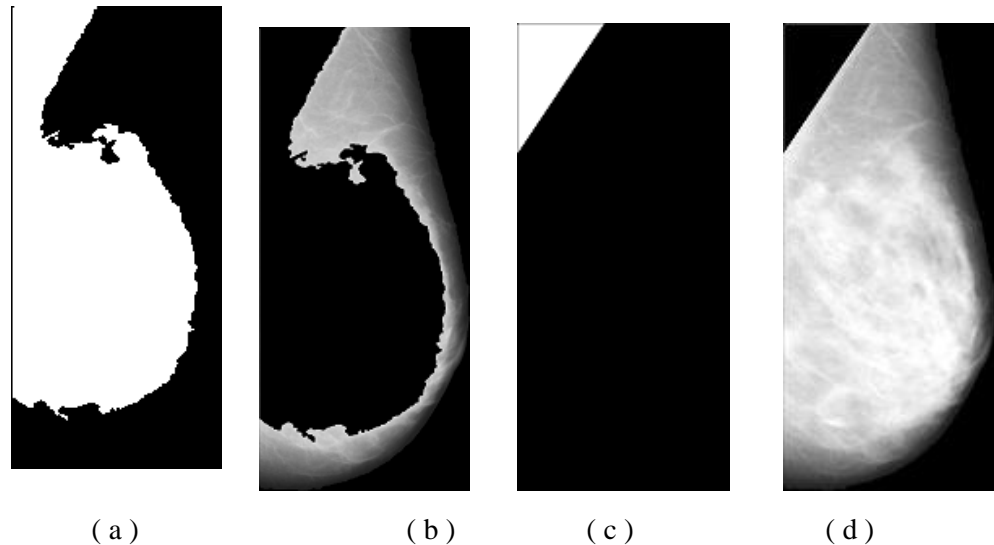
Gambar 4.7 : Gambar (a) payudara menghadap ke kiri, (b) payudara menghadap ke kanan

3. Mencari area otot pektoral menggunakan *Region Growing*. Jika payudara menghadap ke kanan *initial seed* menggunakan pixel pada baris ke 5 dan kolom ke 5 dari kiri. Jika payudara menghadap ke kiri *initial seed* menggunakan pixel pada baris ke 5 kolom ke 5 dari kanan.



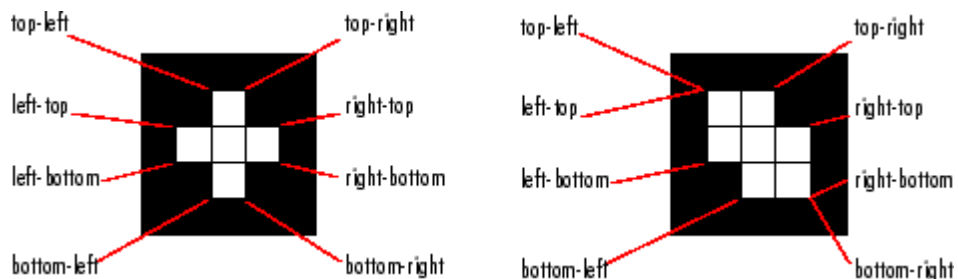
(a) (b)
Gambar 4.8 : Gambar (a) segmen otot pektoral, (b) segmentasi payudara

4. Pada beberapa kasus, warna pada otot pektoral dan area payudara memiliki kemiripan warna, sehingga dapat menyebabkan kesalahan segmentasi, seperti yang terjadi pada citra di bawah ini.



Gambar 4.2: Gambar (a) segmen otot pektoral yang salah, (b) segmentasi payudara yang salah, (c) segmen otot pektoral yang benar, (d) segmentasi payudara yang benar

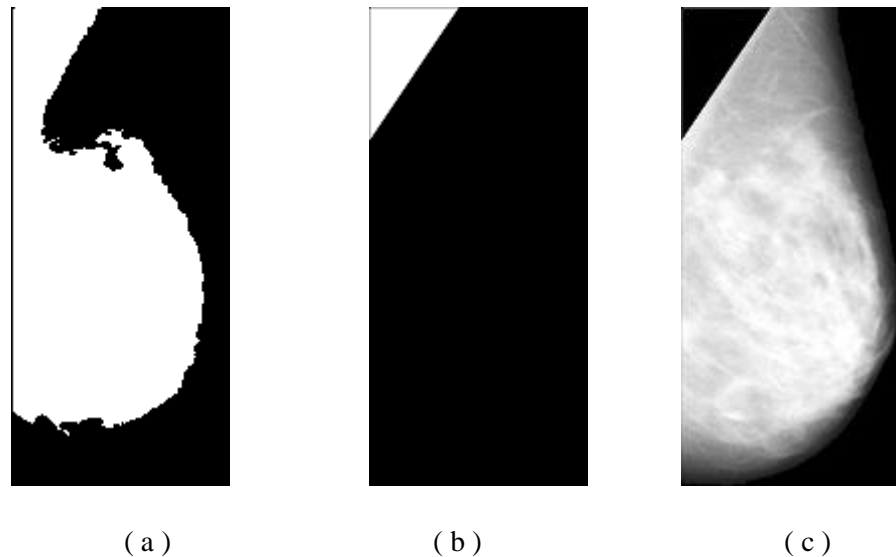
Pada gambar 4.10 (a) terlihat kesalahan segmentasi otot pektoral, untuk mengatasi hal tersebut maka harus ditarik garis lurus dari tepi- tepi area otot pectoral menggunakan persamaan garis lurus $y = mx + b$, kemudian dibentuk poligon untuk melakukan segmentasi pada area otot pektoral. Koordinat awal pembentukan poligon pada segmen otot pektoral adalah dengan mencari *extrema point (EP)* dari segmen otot pektoral. *Extrema point* adalah titik- titik atau pixel terluar dari suatu obyek. *Extrema point* dicari dengan menggunakan fungsi *regionprops* dengan nilai parameter *properties 'Extrema'*. *Extrema point* dapat diilustrasikan seperti di bawah ini :



Gambar 4.10 : Ilustrasi Extrema Point

Untuk citra payudara yang menghadap ke kanan, urutan koordinat poligon (baris, kolom) adalah $(1, EP \text{ top right}) - (y, 1) - (1, 1) - (1, EP \text{ top right})$, sedangkan untuk citra payudara yang menghadap ke kiri urutan koordinat poligon (baris, kolom) adalah $(1, EP \text{ top left}) - (y, \text{ lebar citra}) - (1, \text{ lebar citra}) - (1, EP \text{ top left})$.

5. Segmentasi pada area otot pektoral sesuai dengan koordinat poligon yang telah ditentukan, menggunakan fungsi *roipoly*.
6. Proses *masking* antara citra segmen pektoral dan citra asli mammogram untuk mendapatkan citra yang tersegmentasi.



Gambar 4.11 : Gambar (a) segmentasi otot pektoral, (b) segmentasi otot pektoral setelah ditarik garis lurus, (c) hasil akhir segmentasi

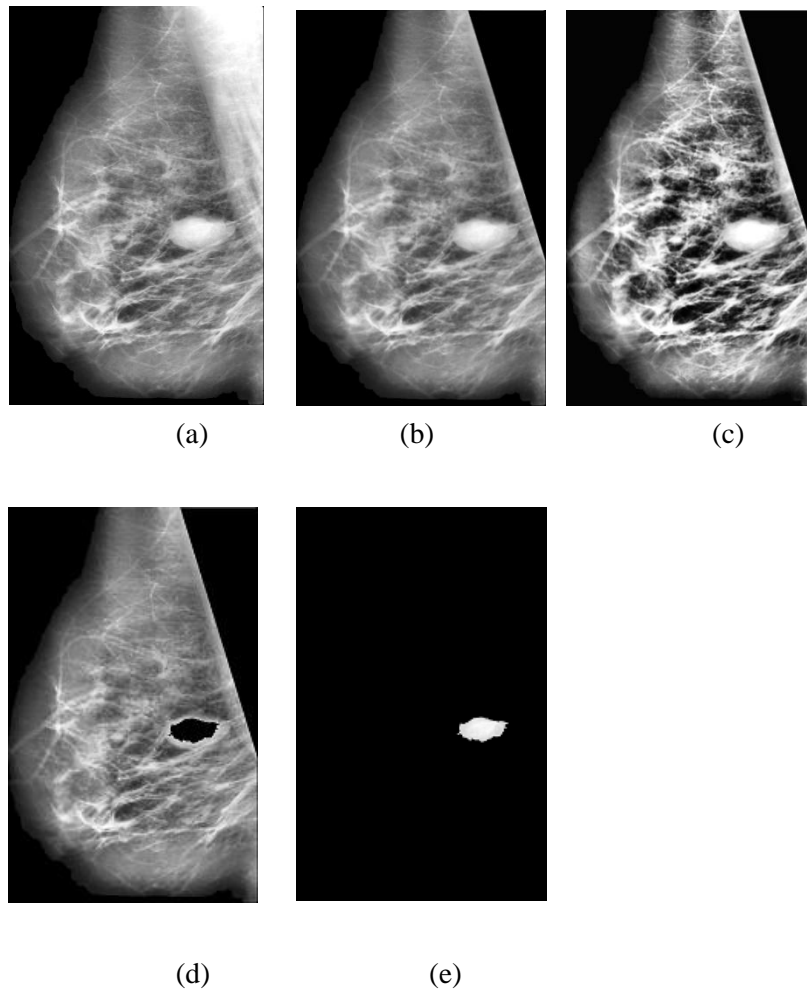
c. Pencarian *Region Of Interest (ROI)* Massa Tumor

Melakukan deteksi masa tumor adalah bagian yang paling sulit karena beberapa faktor, antara lain karena bentuk dan ukuran masa tumor yang berbeda- beda, kemiripan warna antara massa tumor dan area payudara di sekelilingnya, dan ketajaman warna massa tumor yang kurang. Metode untuk melakukan deteksi terhadap massa tumor, metode yang digunakan pada penelitian ini adalah segmentasi *Region Growing*. Tugas yang sulit untuk menentukan *initial seed pixel* untuk segmentasi, karena harus mencari area- area yang dimungkinkan itu adalah area massa tumor secara otomatis. Algoritma :

1. Mentransformasikan citra menggunakan metode *contrast-limited adaptive histogram equalization (CLAHE)* untuk menambah kontras pada citra, fungsi pada MATLAB yang digunakan adalah *adapthisteq*.
2. Mentransformasikan citra ke dalam bentuk biner, dengan kriteria pixel yang memiliki nilai *grayscale* kurang dari 230, set nilainya menjadi 0.
3. Operasi *morphological* untuk mengurangi distorsi pada hasil ekstraksi, fungsi yang digunakan adalah *bwmorph* dengan parameter *operation 'clean'*. Kemudian lakukan

- operasi *morphological* lagi menggunakan *bwmorph* dengan parameter *operation* 'majority' dan *n* dengan nilai 3, operasi ini akan menghaluskan tepi citra.
4. Mencari objek pada citra biner yang memiliki bentuk menyerupai bentuk bulat, dengan cara mencari rasio dari *bounding* tiap- tiap obyek biner pada citra, pada penelitian ini nilai rasio minimal adalah 0.4, sehingga obyek dengan rasio *bounding-box* lebih dari atau sama dengan 0.4 akan di pertahankan, dan obyek dengan rasio *bounding-box* di bawah 0.4 akan dihilangkan. Fungsi yang digunakan untuk mencari rasio *bounding-box* adalah *regionprops* dengan parameter 'Extent'.
 5. Mencari dan pertahankan obyek yang ditemukan pada step 4 dengan luas area yang paling besar, menggunakan fungsi *regionprops* dengan parameter 'Area', dan hilangkan obyek- obyek lainnya.
 6. Mencari titik tengah dari obyek yang ditemukan pada step 5 menggunakan fungsi *regionprops* dengan parameter 'Centroid'. Ini merupakan lokasi initial seed point yang akan digunakan untuk segmentasi *region growing*.
 7. Segmentasi *region growing* dengan *initial seed point* yang telah ditemukan pada step 6 dan nilai *threshold* 0.3.

Contoh hasil dari proses segmentasi ditunjukkan pada gambar 4.12 di bawah ini, dengan contoh kasus pada citra *mdb025_CIRC_2.PGM*:



Gambar 4.12 : (a) Citra mammogram, (b) mammogram tanpa otot pektoral, (c) transformasi CLAHE (d) mammogram tanpa massa tumor (e) massa tumor

4.2. Ekstraksi Fitur Massa Tumor

4.2.1. Ekstraksi Fitur Dengan GLCM

Tahap ini masih merupakan tahap pre- prosesing dari proses klasifikasi, ekstraksi fitur bertujuan untuk mendapatkan nilai- nilai fitur dari citra mammogram untuk membentuk dataset, hasil dari proses ini adalah dataset MIAS. Metode yang digunakan pada ekstraksi fitur adalah *GLCM (Gray Level Co-occurrence Matrix)*. *Co-occurrence* didefinisikan sebagai distribusi gabungan dari tingkat keabuan (*gray level*) dua piksel yang terpisah jarak dan arah tertentu ($\Delta x, \Delta y$). Metode GLCM adalah salah satu cara mengekstrak fitur tekstur statistik orde-kedua, GLCM (yang disebut juga *Grey Tone Spatial Dependency Matrix*) adalah tabulasi mengenai frekuensi atau seberapa seringnya kombinasi nilai kecerahan piksel yang berbeda posisinya terjadi dalam suatu citra. Ilustrasi pembentukan GLCM atas citra dengan 4 tingkat keabuan (*gray level*) pada jarak $d=1$ dan sudut 0° adalah seperti 3.13.

0	0	1	1	1
0	0	1	1	1
0	2	2	2	2
2	2	3	3	3
2	2	3	3	3

	0	1	2	3
0	2	2	1	0
1	0	4	0	0
2	0	0	5	2
3	0	0	0	4

Gambar 4.13 : Ilustrasi pembentukan GLCM atas citra dengan 4 tingkat keabuan (*gray level*)

Matriks GLCM mampu menangkap sifat tekstur tetapi tidak secara langsung dapat digunakan sebagai alat analisis, misalnya membandingkan dua tekstur. Data ini harus disarikan lagi agar didapatkan angka-angka yang bisa digunakan untuk mengklasifikasi tekstur, fitur- fitur yang digunakan pada penelitian ini adalah *Homogeneity (Angular Second Moment)*, *Contrast*, *Correlation*, *Sum of Squares*, *Inverse Diference Moment*, *Sum Average*, *Sum Variance*, *Sum Entropy*, *Entropy*, *Diference Variance*, *Diference Entropy*, *Information Measures of Correlation 1*, *Information Measures of Correlation 2*.

Terdapat 2 (dua) parameter pada GLCM yang menentukan hasil ekstraksi fitur, yaitu jarak (*distance*) d dan sudut (*angle*) θ , pada penelitian ini nilai d adalah 2 dan nilai θ yang digunakan adalah $0^\circ, 45^\circ, 90^\circ, \text{ dan } 135^\circ$, dan kemudian dicari rata- rata hasil *GLCM* dari beberapa sudut tersebut.

Di bawah adalah contoh nilai fitur- fitur dari citra pada mdb025_CIRC_2.PGM. Fitur- fitur ini diekstraksi menggunakan *GLCM* dengan parameter jarak (distance) $d=2$ dan $\theta=0^\circ, 45^\circ, 90^\circ,$ dan 135° .

Tabel 4.1 : Hasil Ekstraksi Fitur Citra Mammogram

No	Fitur	Nilai
1	<i>Homogeneity</i>	0.932
2	<i>Contrast</i>	0.194
3	<i>Correlation</i>	0.971
4	<i>Sum of Squares</i>	16.007
5	<i>Inverse Difference Moment</i>	0.997
6	<i>Sum Average</i>	7.136
7	<i>Sum Variance</i>	38.780
8	<i>Sum Entropy</i>	2.098
9	<i>Entropy</i>	2.208
10	<i>Difference Variance</i>	0.194
11	<i>Difference Entropy</i>	0.414
12	<i>Information Measures of Correlation 1</i>	-0.748
13	<i>Information Measures of Correlation 2</i>	0.963

4.2.2. Normalisasi Dataset MIAS

Pada tahap ini, proses yang dilakukan adalah melakukan normalisasi terhadap dataset MIAS, agar fitur- fitur memiliki jangkauan yang sama. Metode yang digunakan untuk proses normalisasi adalah normalisasi linier.

Tabel 4.2 : Sampel Dataset MIAS sebelum dinormalisasi

No	Homogeneity	Contrast	Correlation	Sum of Squares	Inv.Diff. Moment	Sum Average	Sum Variance	Sum Entropy	Entropy	Diff. Variance	Diff. Entropy	IMC1	IMC2	Label
1	0.946	0.146	0.984	17.436	0.998	7.170	43.001	2.240	2.327	0.146	0.358	-0.801	0.977	B
2	0.957	0.118	0.988	14.880	0.998	6.313	38.376	2.030	2.105	0.118	0.307	-0.824	0.973	B
3	0.949	0.162	0.981	19.574	0.998	7.884	51.368	1.981	2.057	0.162	0.341	-0.791	0.966	B
4	0.938	0.157	0.983	23.048	0.998	8.615	58.455	2.281	2.376	0.157	0.386	-0.781	0.976	B
5	0.946	0.139	0.982	18.214	0.998	7.601	46.711	2.004	2.086	0.139	0.351	-0.783	0.965	B
6	0.918	0.231	0.974	22.650	0.997	8.570	57.598	2.241	2.366	0.231	0.465	-0.723	0.965	B
7	0.932	0.261	0.978	23.946	0.997	8.532	61.732	2.336	2.443	0.261	0.417	-0.774	0.977	B
8	0.935	0.184	0.980	18.942	0.997	7.580	48.182	2.144	2.246	0.184	0.401	-0.766	0.969	B
9	0.938	0.203	0.978	17.760	0.997	7.302	44.830	2.126	2.224	0.203	0.389	-0.776	0.970	B
10	0.945	0.175	0.982	22.889	0.998	8.523	60.236	2.127	2.208	0.175	0.359	-0.790	0.972	M

Di atas adalah contoh data pada dataset MIAS yang belum dilakukan normalisasi, pada kolom yang berwarna hijau menunjukkan fitur- fitur yang memiliki nilai jangkauan yang besar dari fitur- fitur lainnya, sehingga harus dilakukan normalisasi dataset.

Tabel 4.3 : Sampel Dataset MIAS yang setelah dinormalisasi

No	Homo geneity	Contr ast	Correlati on	Sum of Squares	Inv.Diff. Moment	Sum Average	Sum Variance	Sum Entropy	Entrop y	Diff. Varian ce	Diff.E ntropy	IMC1	IMC2	Lab el
1	0.649	-1.021	0.975	-0.862	1.016	-0.931	-0.943	0.671	0.583	-1.021	-0.592	-0.821	1.052	B
2	1.957	-1.815	1.789	-1.712	2.087	-2.125	-1.480	-1.201	-1.302	-1.815	-2.044	-1.732	0.389	B
3	1.035	-0.574	0.145	-0.151	0.720	0.063	0.028	-1.638	-1.715	-0.574	-1.081	-0.449	-0.974	B
4	-0.220	-0.709	0.664	1.005	0.636	1.080	0.851	1.033	1.003	-0.709	0.189	-0.025	0.792	B
5	0.736	-1.228	0.481	-0.603	1.269	-0.331	-0.512	-1.434	-1.464	-1.228	-0.792	-0.100	-1.029	B
6	-2.541	1.358	-1.254	0.872	-1.993	1.018	0.752	0.678	0.919	1.358	2.424	2.267	-1.072	B
7	-0.903	2.176	-0.416	1.303	-2.103	0.965	1.232	1.532	1.570	2.176	1.071	0.253	0.949	B
8	-0.543	0.023	0.129	-0.361	-0.222	-0.361	-0.342	-0.184	-0.106	0.023	0.605	0.549	-0.430	B
9	-0.203	0.553	-0.441	-0.754	-0.575	-0.748	-0.731	-0.343	-0.295	0.553	0.281	0.156	-0.257	B
10	0.536	-0.222	0.463	0.952	0.289	0.953	1.058	-0.339	-0.430	-0.222	-0.564	-0.409	0.067	M

Di atas adalah contoh data pada dataset MIAS yang sudah dilakukan normalisasi, dapat dilihat bahwa nilai – nilai pada fitur dataset memiliki jangkauan nilai yang relatif sama.

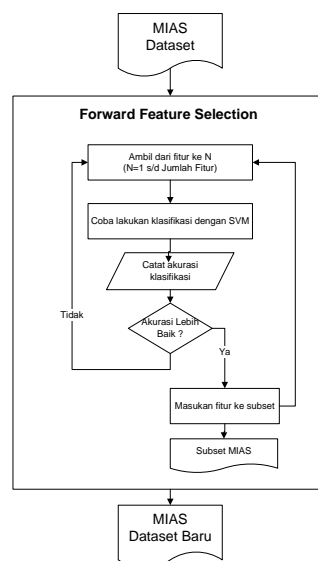
4.3. Klasifikasi Mammogram

Pada tahap pre prosesing telah di dapatkan dataset hasil ekstraksi fitur dari citra mammogram, pada tahap ini adalah melakukan klasifikasi terhadap dataset tersebut. Pada proses klasifikasi ini terdapat dua tahap, yaitu seleksi fitur dataset dan proses klasifikasi itu sendiri. Proses seleksi fitur adalah proses mencari fitur- fitur yang paling relevan untuk proses klasifikasi, sehingga dengan adanya proses seleksi fitur ini hasil akurasi klasifikasi lebih tinggi jika dibandingkan melakukan klasifikasi tanpa adanya proses seleksi fitur. Pada proses seleksi fitur, metode yang digunakan adalah *forward selection*, dan untuk proses klasifikasi metode yang digunakan adalah *support vector machine (SVM)* dengan fungsi *kernel radial basis function (RBF)*. Alat yang digunakan untuk proses seleksi fitur dan klasifikasi adalah software *Rapid Miner*.

4.3.1. Seleksi Fitur Forward Selection

Forward selection adalah seleksi fitur dimulai dari menambahkan fitur satu- persatu, pada setiap langkah ditambahkan fitur yang menurunkan *error* paling banyak, sampai semua *error* dihilangkan, fitur- fitur yang relevan hasil seleksi ini yang nantinya akan digunakan sebagai input pada proses klasifikasi. Pada Rapid Miner, untuk melakukan seleksi fitur metode *forward selection* menggunakan operator *Forward Selection*, fitur- fitur yang relevan akan ditandai menggunakan bobot (*weight*) 1. Contoh di bawah ini adalah contoh hasil seleksi fitur menggunakan *forward selection* pada Rapid Miner, pada gambar 4.15 terlihat bahwa fitur yang relevan untuk klasifikasi adalah *Sum Average* dan *Sum Variance* yang di tandai dengan bobot (*weight*) bernilai 1. Fitur- fitur hasil seleksi akan selalu berbeda setiap proses, tergantung dari jumlah data pada dataset dan nilai yang terdapat dalam fitur- fitur di dataset.

Percobaan menggunakan bahan sample dari dataset MIAS, alat yang digunakan adalah Rapid Miner. Pada proses ini, feature selection akan mencari fitur yang memiliki kontribusi terhadap peningkatan akurasi klasifikasi, fitur yang memiliki kontribusi terhadap akurasi akan di set *weight* (bobot) dengan nilai 1, yang tidak berkontribusi akan di set dengan nilai 0. Hasil dari proses ini adalah dataset baru yang telah terseleksi fiturnya, atau disebut dengan subset. Alur dari proses seleksi fitur ditunjukkan oleh flowchart di bawah ini :



Gambar 4.3: Alur proses forward feautre selection

Hasil seleksi fitur ditunjukkan oleh gambar di bawah ini :

attribute	weight
Homogeneity	0
Contrast	0
Correlation	0
Sum of Squares	0
Inv.Diff.Moment	0
Sum Average	1
Sum Variance	1
Sum Entropy	0
Entropy	0
Diff. Variance	0
Diff.Entropy	0
IMC1	0
IMC2	0

Gambar 4.15 : Seleksi fitur dataset Mammogram

Berikut adalah contoh perbandingan hasil akurasi klasifikasi, antara klasifikasi yang menggunakan seleksi fitur dan klasifikasi tanpa menggunakan seleksi fitur.

Table View Plot View

accuracy: 42.86%

	true B	true M	class precision
pred. B	2	3	40.00%
pred. M	1	1	50.00%
class recall	66.67%	25.00%	

Gambar 4.16 : Akurasi SVM tanpa seleksi fitur

Table View Plot View

accuracy: 100.00%

	true B	true M	class precision
pred. B	4	0	100.00%
pred. M	0	2	100.00%
class recall	100.00%	100.00%	

Gambar 4.17 : Akurasi SVM dengan seleksi fitur

Pada gambar 4.16 dan 3.17 di atas terlihat sekali perbedaan akurasi, klasifikasi yang tidak menggunakan seleksi fitur hanya menghasilkan akurasi 42.86%, sedangkan klasifikasi dengan menggunakan seleksi fitur, dapat menghasilkan akurasi hingga 100%.

4.3.2. Klasifikasi Dengan Support Vector Machine

Tujuan dari proses klasifikasi adalah menentukan kategori dari sebuah data, pada penelitian ini klasifikasi digunakan untuk menentukan data pada dataset MIAS termasuk ke dalam kategori tumor ganas (kanker) disimbolkan dengan “M” atau tumor jinak yang disimbolkan dengan “B”.

Pada penelitian ini klasifikasi data menggunakan alat bantu software RapidMiner. Urutan proses klasifikasi pada Rapid Miner adalah sebagai berikut:

1. Membaca dataset MIAS menggunakan operator *ReadExcel*.
2. Seleksi fitur dataset menggunakan operator *ForwardSelection*.
3. Membagi porsi dataset training dan dataset testing menggunakan operator *Validation*.
4. Pembelajaran (training) klasifier menggunakan operator *SVM (LibSVM)*, nilai parameter *svm type= C-SVC, kernel type = rbf, gamma = 1, C =1.2, epsilon 1.5*.
5. Menguji coba (*testing*) klasifier menggunakan operator *Apply Model* dan *Performance* untuk mengukur akurasi dari klasifier.

Eksperimen akan menggunakan beberapa kombinasi proporsi untuk data training dan data testing, yaitu 90%, 80%, 70%, 60%, 50% proporsi untuk data training, sedangkan proporsi untuk data testing menyesuaikan. Untuk mengukur performa dari model klasifikasi digunakan metode *Confusion Matrix*, penjabaran secara lengkap adalah :

a. Confusion Matrix

Tabel 4.4 : Confusion Matrix

		Prediksi	
		Negative (B)	Positive (M)
Aktual	Negative (B)	a	b
	Positive (M)	c	d

a adalah jumlah dari prediksi yang **benar**, nilai aktual **Negative (B)** dan hasil prediksi **Negative (B)**

b adalah jumlah dari prediksi yang **salah**, nilai aktual **Negative (B)** tetapi hasil prediksi **Positive (M)**

c adalah jumlah dari prediksi yang **salah**, nilai aktual **Positive (M)** tetapi hasil prediksi **Negative (B)**

d adalah jumlah dari prediksi yang benar **benar**, nilai aktual **Positive (M)** dan hasil prediksi **Positive (M)**

b. Accuray (AC)

Adalah proporsi dari jumlah total dari prediksi yang benar.

$$AC = \frac{a + d}{a + b + c + d}$$

c. Recall atau True Positive Rate (TP)

Adalah proporsi kasus malignant yang benar diidentifikasi sebagai malignant

$$TP = \frac{d}{c + d}$$

d. False Positive Rate (FP)

Adalah proporsi kasus benign yang salah diidentifikasi sebagai malignant

$$FP = \frac{b}{a + b}$$

e. True Negative Rate (TN)

Adalah proporsi kasus benign yang benar diidentifikasi sebagai benign

$$TN = \frac{a}{a + b}$$

f. False Negative Rate (FN)

Adalah proporsi kasus malignant yang salah diidentifikasi sebagai benign

$$FN = \frac{c}{c + d}$$

g. Precision (P)

Adalah proporsi dari prediksi yang benar pada kasus malignant

$$P = \frac{d}{b + d}$$

5. HASIL PENELITIAN DAN PEMBAHASAN

5.1. Hasil Klasifikasi Pada Dataset MIAS

Rangkuman hasil percobaan adalah sebagai berikut, kolom dengan label A adalah hasil klasifikasi tanpa menggunakan seleksi fitur *forward selection*, sedangkan kolom dengan label B adalah hasil klasifikasi dengan menggunakan seleksi fitur *forward selection*.

Pada percobaan yang dilakukan, dataset dibagi menjadi 2 (dua) bagian, dataset *training* dan dataset *testing*. Masing- masing akan dibagi menjadi beberapa porsi dataset training – testing : 90% - 10 %, 80% - 20%, 70% - 30%, 60%- 40%, 50% - 50% dan terakhir menggunakan validasi *10 Cross Validation*. Kombinasi beberapa porsi dataset dilakukan untuk menguji kehandalan klasifier dalam melakukan klasifikasi.

Kemampuan model klasifikasi dapat diukur dari nilai dari *accuracy (AC)*, *recall (TP)* dan *precision (P)*. Pada hasil percobaan klasifikasi yang ditunjukkan pada tabel 5.1 di bawah menunjukkan kemampuan model klasifikasi nilai *accuracy* menghasilkan nilai antara 76% -100%, *recall* menghasilkan nilai 56%- 100% dan *precision* menunjukkan nilai 66.7% - 100%. Berdasarkan hasil percobaan pada penelitian ini, penerapan seleksi fitur *forward selection* terbukti berpengaruh terhadap hasil klasifikasi yang signifikan, peningkatan klasifikasi pada tiap percobaan rata- rata sebesar 10%.

Pada percobaan no 1 s/d no 5, menunjukkan model klasifier memiliki *accuracy*, *recall* dan *precision* yang tinggi yaitu berkisar antara nilai 81% -100%. Pada percobaan no 6 menggunakan *10 Cross Validation*, akurasi hanya mencapai angka 76.1% dengan *recall* hanya 56%. Jika dilihat dari data hasil percobaan no 6, klasifier banyak salah dalam melakukan klasifikasi pada kasus kanker (label M), contoh pada percobaan no 6, dari 10 kasus kanker, klasifier salah mengklasifikasikan data kasus kanker ke dalam kasus tumor jinak sebanyak 11 data, hanya 14 data yang benar diklasifikasi sebagai kasus kanker. Kesalahan tersebut yang menyebabkan nilai *recall* menurun. Namun disisi lain nilai *True Negative* (kasus tumor jinak (label B) yang benar dideteksi sebagai tumor jinak) berdasarkan data dari percobaan no 6, menunjukkan nilai yang tinggi yaitu 88.10%, ini menunjukkan bahwa klasifier mampu mengklasifikasi kasus tumor jinak dengan baik. Kurang tepatnya klasifier dalam melakukan klasifikasi kasus kanker pada percobaan no 6, dapat disebabkan karena jumlah sample data kasus kanker masih terlalu sedikit, yaitu hanya 25 record.

Tabel 5.1 : Hasil Klasifikasi

No	Porsi Dataset (Training-Testing) (%)	AC (%)		Recall(%)		FP(%)		TN(%)		FN(%)		P(%)	
		A	B	A	B	A	B	A	B	A	B	A	B
1	90 - 10	100	100	100	100	0.00	0.00	100	100	0.00	0.00	100	100
2	80 - 20	76.9	92.3	60	80	12.5	0.00	87.5	100	40	20	75	100
3	70 - 30	75	90	42	71.4	7.7	0.00	92.3	100	57.1	28.6	75	100
4	60 - 40	74	85.2	30	70	0.00	5.88	100	94.1	70	30	100	87.5
5	50 - 50	69.7	81.8	25	100	4.76	28.6	95.2	71.4	75	0.00	75	66.7
6	10 Cross Validation	67.2	76.1	56	56	26.1	11.9	73.8	88.1	44	44	56	73.7

5.2. Hasil Pengujian Set Atribut

Dari hasil pengujian klasifikasi, kemudian dilakukan pengujian atribut terhadap hasil akurasi pada beberapa kombinasi proporsi data training. Pengujian ini dilakukan untuk mendapatkan set atribut yang memiliki hasil akurasi paling baik terhadap beberapa kombinasi data training. Metode pengujian ini dengan cara mencari nilai akurasi klasifikasi dengan mengaplikasikan set atribut pada model klasifier, dengan porsi data training- testing 90-10,80-20,70-30,60-40,50-50. Hasil ditunjukkan pada table 5.2 di bawah, set atribut yang memiliki rata- rata akurasi paling tinggi terhadap kombinasi data training- testing adalah set atribut nomor 3, yaitu sebesar 86.8 %.

Tabel 5.2 : Hasil Pengujian Set Atribut

No	Set Atribut	Proporsi Dataset (%)					Rata-rata (%)
		90-10	80-20	70-30	60-40	50-50	
1	Homogeneity, Contrast, Correlation, Sum of Squares, Inv.Diff.Moment, IMC2	100	77	80	77.8	75.8	82.1
2	Homogeneity, Contrast, Correlation, Diff.Entropy	100	84.62	85	74.1	78.8	84.5
3	Homogeneity, Contrast, Inv.Diff.Moment, Sum Entropy, Entropy, Diff. Variance, IMC1	100	92.3	85	77.8	78.8	86.8
4	Homogeneity, Contrast, Sum Entropy, Diff. Variance, Diff.Entropy, IMC1	100	84.6	80	77.8	81.82	84.8
5	Sum Entropy, Diff.Entropy	100	76.9	90	81.5	75.8	84.8
	Rata-Rata	100	83.084	84	77.8	78.204	

6. KESIMPULAN DAN PENUTUP

Penerapan seleksi fitur *forward selection* pada SVM terbukti berpengaruh terhadap peningkatan akurasi klasifikasi yang signifikan, peningkatan klasifikasi rata-rata sebesar 10.43%. Tanpa penerapan *forward selection* pada SVM untuk klasifikasi dataset MIAS akurasi rata-rata tercatat hanya 77.13%, setelah penerapan *forward selection* pada klasifier SVM, akurasi rata-rata meningkat menjadi 87.57%. Akurasi klasifikasi FS- SVM secara lengkap ditunjukkan pada tabel 5.3 di bawah ini :

Tabel 5.3 : Akurasi Klasifikasi FS- SVM

No	Porsi Dataset (Training- Testing) (%)	Akurasi (%)		Peningkatan Akurasi (%)
		SVM	FS- SVM	
1	90 - 10	100	100	0
2	80 - 20	76.9	92.3	15.4
3	70 - 30	75	90	15
4	60 - 40	74	85.2	11.2
5	50 - 50	69.7	81.8	12.1
6	10 Cross Validation	67.2	76.1	8.9
	Rata- Rata	77.13	87.57	10.43

Referensi

- [1] Jemal, A. et al. "Global Cancer Statistics," *World*, vol. 61, 2011 , pp. 69-90.
- [2] Meenalosini, S. et al , "Segmentation Of Cancer Cells In Mammogram Using Region Growing Method And Gabor Features," *International Journal of Engineering*, vol. 2, 2012, pp. 1055-1062.
- [3] Okun, Oleg and H. Priisalu, "Dataset complexity in gene expression based cancer classification using ensembles of k-nearest neighbors," *Gene Expression*, 2009.
- [4] Cruz-ramírez, Nicrando. et al, "Diagnosis of breast cancer using Bayesian networks : A case study," *Computers in Biology and Medicine*, vol. 37, 2007, pp. 1553 - 1564.
- [5] Akay, Mehmet Fatih, "Support vector machines combined with feature selection for breast cancer diagnosis," *Expert Systems with Applications*, vol. 36, 2009, p. 101016/jeswa200801009.
- [6] Chen, Hui-ling et al, "Support vector machine based diagnostic system for breast cancer using swarm intelligence Abstract :," *Knowledge Creation Diffusion Utilization*.
- [7] Palanivel, J. , "An Efficient Breast Cancer Screening System Based on Adaptive Support Vector Machines with Fuzzy C-Means Clustering," *European Journal of Scientific Research*, vol. 51, 2011, pp. 115-123.
- [8] Mu, Tingting, "Breast cancer detection from FNA using SVM with different parameter tuning systems and SOM – RBF classifier," *Journal of the Franklin Institute*, vol. 344, 2007, pp. 285-311.
- [9] Polat, K. and S. Gunes, "Breast cancer diagnosis using least square support vector machine," *Digital Signal Processing*, vol. 17, 2007, pp. 694-701.

- [10] Cord, P.C. Matthieu, "Machine Learning Techniques for Multimedia: Case Studies on Organization and Retrieval," *Springer*, 2008.
- [11] Hindman, M., "The Myth of Digital Democracy," *Princeton University Press*, 2010.
- [12] Zhang, G.P., "Neural networks for classification: a survey," *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 30, 2000, pp. 451-462.
- [13] Reif, Matthias and F. Shafait, "Efficient feature size reduction via predictive forward selection," *Pattern Recognition*, vol. 47, 2014, pp. 1664-1673.
- [14] Miche, Yoan et al, "Advantages of Using Feature Selection Techniques on Steganalysis Schemes," *Techniques*, 2007, pp. 606-613.
- [15] Wang, Qiangwei et al, "Feature Selection for Human Resource Selection Based on Affinity Propagation and SVM Sensitivity Analysis," *Computing*, 2009, pp. 31-36.
- [16] Santosa, Budi, "Tutorial Support Vector Machine," 1995, pp. 1-23.
- [17] S.L., Michae, "Advances in Pattern Recognition," *Analysis*.
- [18] Yang, Z.R. "Machine Learning Approaches to Bioinformatics," *World Scientific Publishing Co. Pte. Ltd.*, 2010.
- [19] Gonzalez, R.C., "Digital Image Processing Third Edition."
- [20] Haralick, Robert M et al, "Textural Features for Image Classification," *Earth*, 1973.
- [21] Gadkari, Dhanashree, "Image Quality Analisis Using GLCM," *University of Pune*, 2000 *A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Modeling and Simulation in the College of Arts and Sciences at the University of Central Florida*, 2004.

- [22] Mousa, Rafayah et al, " Breast cancer diagnosis system based on wavelet analysis and fuzzy-neural," *Expert Systems with Applications*, 2005, pp. 713-723.
- [23] Moayedi , Fatemeh et al, " Contourlet-based mammography mass classification using the SVM family," *Computers in Biology and Medicine*, 2010, pp. 373-383.

Biografi Penulis



Prihananto Joko Tri Laksono, lahir di Grobogan, 14 Juli 1988, merupakan alumni S1 Teknik Informatika Universitas Dian Nuswantoro Semarang dan alumni Magister Teknik Informatika, konsentrasi Intelligent System di Universitas Dian Nuswantoro. Pekerjaan saat ini sebagai Staff IT di Instalasi SIRS & Kom RSUP Dr. Kariadi Semarang, Programmer.