

BAB II

LANDASAN TEORI

Bab ini dijelaskan tinjauan pustaka dan tinjauan studi yang berkaitan dengan penelitian yang berjudul Penerapan *Texture Based Extraction* untuk Mendeteksi Massa pada Citra Mammogram.

2.1. Tinjauan Pustaka

Dalam kegiatan penelitian ini ada beberapa pustaka yang berkaitan dengan kegiatan penelitian yang akan dilakukan, antara lain :

(Costa, Humpire-Mamani, & Traina, 2012) melakukan penelitian yang berjudul *An Efficient Algorithm For Fractal Analysis Of Texture*. Penelitian ini tentang membandingkan metode ekstraksi fitur yang biasanya digunakan seperti GLCM dan *gabor filter* dengan SFTA untuk *image retrieval* dan klasifikasi citra. Hasil penelitian tersebut bahwa algoritma SFTA sederhana namun efektif karena ekstraksi SFTA 3,7 kali lebih cepat dari *gabor filter* dan 1,6 lebih cepat dari GLCM.

(Aisyah & Sulisty, 2016) melakukan penelitian yang berjudul *Klasifikasi Kanker Payudara Menggunakan Algoritma Gain Ratio*. Penelitian ini tentang klasifikasi kanker payudara menggunakan metode *decision tree* dengan algoritma *gain ratio*. Hasil menunjukkan nilai *recall*, akurasi dan presisi masing-masing sebesar 92,55%, 95,17% dan 93,76%.

(Biswas et al., 2016) melakukan penelitian yang berjudul *Mammogram Classification Using Gray-Level Co-Occurrence Matrix For Diagnosis Of Breast Cancer*. Penelitian ini tentang klasifikasi citra mammogram normal atau abnormal menggunakan GLCM dan klasifikasi menggunakan SVM, KNN, ANN. Hasil metode tersebut memberikan akurasi 95%, sensitivitas 100% dan spesifisitas 90%.

(Helwan & Abiyev, 2016) melakukan penelitian yang berjudul *Shape And Texture Features For The Identification Of Breast Cancer*. Penelitian ini tentang mengidentifikasi kanker payudara dengan berdasarkan bentuk dan fitur tekstur menggunakan GLCM menggunakan klasifikasi *Neural Network*. Hasilnya menunjukkan identifikasi sebesar 92%.

(Mahdaouy, Gaussier, & Alaoui, 2016) melakukan penelitian yang berjudul *Mammogram Classification Using Curvelet GLCM Texture Features and GIST*

Features. Penelitian ini tentang klasifikasi gambar mammogram dengan CGLCM dan GIST dengan klasifikasi SVM. Penggabungan fitur menghasilkan hasil yang lebih baik dengan akurasi 92,39% dibandingkan dengan 77,97% dan 91% untuk GIST dan CGLCM.

(**Junita, 2017**) melakukan penelitian yang berjudul *Ekstraksi Fitur Dan Klasifikasi Menggunakan Metode Glcm Dan Svm Pada Citra Mammogram Untuk Identifikasi Kanker Payudara*. Penelitian ini tentang mendeteksi kanker payudara menggunakan ekstraksi fitur tekstur GLCM dan klasifikasi SVM pada citra mammogram. Hasil yang didapatkan dari metode tersebut dengan probabilitas 60%.

(**Hariraj et al., 2017**) melakukan penelitian yang berjudul *An Efficient Data Mining Approaches For Breast Cancer Detection And Segmentation In Mammogram*. Penelitian ini tentang deteksi dini kanker payudara menggunakan klasifikasi KNN dan SVM. Segmentasi yang digunakan *K-Means Clustering* dan ekstraksi ciri menggunakan GLCM. Hasilnya SVM memiliki nilai akurasi, sensitivitas, dan spesifitas dibandingkan KNN sebesar 96,5%, 94,3%, 78%.

(**Novar Setiawan & Suwija Putra, 2018**) melakukan penelitian yang berjudul *Klasifikasi Citra Mammogram Menggunakan Metode K-Means, GLCM, dan Support Vector Machine (SVM)*. Penelitian ini tentang klasifikasi citra mammogram dengan hasil normal atau abnormal dengan *K-Means* untuk segmentasi, GLCM untuk ekstraksi fitur tekstur, dan SVM untuk klasifikasi. Hasil akurasi untuk metode tersebut mencapai 80%.

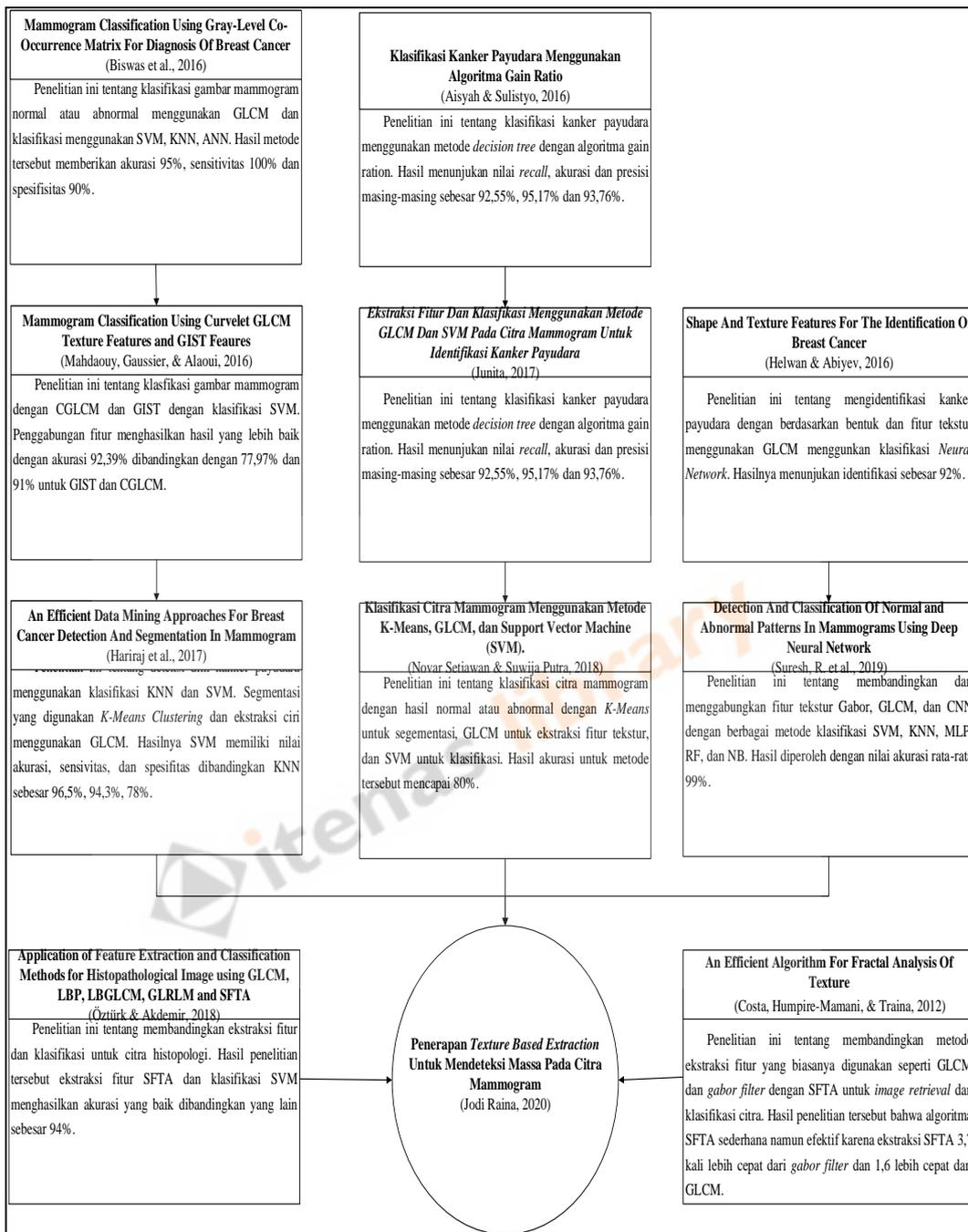
(**Öztürk & Akdemir, 2018**) melakukan penelitian yang berjudul *Application of Feature Extraction and Classification Methods for Histopathological Image using GLCM, LBP, LBGLCM, GLRLM and SFTA*. Penelitian ini tentang membandingkan ekstraksi fitur dan klasifikasi untuk citra histopologi. Hasil penelitian tersebut ekstraksi fitur SFTA dan klasifikasi SVM menghasilkan akurasi yang baik dibandingkan yang lain sebesar 94%.

(**Suresh et al., 2019**) melakukan penelitian yang berjudul *Detection And Classification Of Normal And Abnormal Patterns In Mammograms Using Deep Neural Network*. Penelitian ini tentang menggunakan metode *hybrid* untuk melakukan ekstraksi fitur dengan GLCM dan HOG dengan klasifikasi DNN. Hasil

yang diperoleh dengan akurasi sebesar 98,8%. Hasil tersebut lebih besar 3-9% dari metode sebelumnya.

Keterhubungan penelitian yang pertama yaitu penelitian (Costa et al., 2012) memiliki keterhubungan dengan penelitian ini karena penelitian tersebut menjelaskan secara *detail* algoritma dari SFTA mulai dari *input* hingga *output*. Penelitian (Aisyah & Sulisty, 2016) melakukan penelitian dengan obyek yang sama yaitu klasifikasi kanker payudara namun menggunakan metode *gain ratio*. Penelitian (Biswas et al., 2016) juga melakukan penelitian tentang klasifikasi citra mammogram namun menggunakan metode GLCM dan klasifikasi SVM, KNN, dan ANN. Penelitian (Helwan & Abiyev, 2016) melakukan penelitian yang sama berkaitan dengan klasifikasi kanker payudara berdasarkan ekstraksi fitur tekstur GLCM. Penelitian (Mahdaouy et al., 2016) melakukan penelitian yang serupa yaitu klasifikasi citra mammogram namun menggunakan GLCM dan GIST. Penelitian (Junita, 2017) melakukan penelitian dengan obyek yang sama berkaitan dengan identifikasi kanker payudara namun menggunakan metode GLCM. Penelitian (Hariraj et al., 2017) bahwa klasifikasi SVM memiliki akurasi yang lebih baik dari KNN maka dari itu penelitian ini menggunakan klasifikasi SVM untuk mendeteksi massa pada citra mammogram. Penelitian (Novar Setiawan & Suwija Putra, 2018) menggunakan segmentasi *k-means* sebelum ekstraksi fitur tekstur GLCM maka dari itu penelitian ini menggunakan segmentasi *k-means*. *K-means* bertujuan sebagai *color quantization* yang dapat mengurangi jumlah warna pada citra mammogram. Penelitian oleh (Suresh et al., 2019) metode ekstraksi fitur *hybrid* GLCM dan HOG, bahwa HOG menghitung orientasi gradien dan iluminasi dari batas mammogram sama seperti SFTA yang menghitung nilai fraktal pada batas sebuah gambar. Penelitian (Öztürk & Akdemir, 2018) keterhubungan dengan penelitian ini yaitu alasan menggunakan ekstraksi fitur SFTA memiliki akurasi yang tertinggi dari ekstraksi fitur tekstur yang lainnya.

Berdasarkan tinjauan pustaka tersebut, maka Gambar 2.1 berikut menampilkan pemetaan pustaka yang menunjang penelitian ini.



Gambar 2.1 Pemetaan pustaka

2.2. Tinjauan Studi

Dalam kegiatan penelitian ini ada teori - teori yang berkaitan dengan kegiatan penelitian yang dilakukan, antara lain:

2.2.1. Citra Digital

Citra digital adalah suatu nilai yang menciptakan suatu imitasi dari obyek seperti sebenarnya. Komputer membaca citra berisi *array* dengan bilangan *real* yang direpresentasikan dengan deretan *bit*. Nilai ini berisi koordinat x dan y yaitu baris dan kolom yang disebut piksel. Nilai piksel ini bisa mewujudkan warna, titik, garis, dan bidang (Mirah, 2018). Gambar 2.2 merupakan contoh representasi koordinat dan piksel pada citra yang dibaca oleh komputer.

(X,Y)	(0,0)	(3,0) → X			
(0,0)	(80, 70, 78)	(123, 121, 118)	(100, 100, 106)	(15, 59, 22)	(3,0)
	(88, 81, 53)	(112, 86, 51)	(6, 77, 78)	(59, 133, 99)	
	(156, 110, 210)	(65, 90, 168)	(139, 127, 126)	(115, 150, 51)	
	(192, 70, 120)	(80, 70, 78)	(214, 184, 184)	(68, 145, 95)	
(0,4) Y	(221, 189, 230)	(167, 203, 200)	(149, 64, 164)	(168, 65, 314)	(3,4)

Gambar 2.2 Koordinat dan piksel pada citra

Piksel pada citra ditunjukkan oleh kolom dari koordinat (0,0) hingga (3,4) dengan isi sejumlah bilangan. Koordinat dimulai dari titik (0,0) pada sudut kiri atas dari sebuah citra. Citra dibagi menjadi 3 jenis dalam pengolahan citra digital yaitu citra warna (RGB), citra *grayscale*, dan citra biner (Kadir & Adhi, 2013).

1. Citra Berwarna (RGB)

Citra RGB memiliki nilai yang tiap pikselnya diwakili oleh 3 komponen yaitu komponen R (merah), G (hijau), dan B (biru). Rentang nilai tiap komponen dari 0 hingga 255 yang berbasis 8 bit. Sehingga warna yang dihasilkan adalah $255 \times 255 \times 255$ atau sebanyak 16.581.375 warna.



Gambar 2.3 Citra RGB

Sumber : (Pemrograman Matlab, 2017)

2. Citra *Grayscale*

Citra *grayscale* merupakan citra yang intensitasnya diwakili oleh nilai keabuan tiap-tiap piksel dan memiliki nilai dengan kisaran 0 sampai 255. Nilai 0 merupakan nilai terendah dengan warna hitam, sementara nilai 255 merupakan nilai tertinggi dengan warna putih. Semakin kecil nilai piksel maka warna yang dihasilkan maka akan semakin gelap, sedangkan semakin tinggi maka warna akan semakin terang.



Gambar 2.4 Citra *grayscale*

Sumber : (Pemrograman Matlab, 2017)

3. Citra Biner

Citra biner adalah citra yang hanya memiliki 2 nilai intensitas piksel yaitu 0 dan 255. Nilai 0 untuk warna hitam dan nilai 255 untuk warna putih. Citra biner sering juga disebut citra *black and white*.



Gambar 2.5 Citra biner

Sumber : (Pemrograman Matlab, 2017)

2.2.2. Kanker Payudara

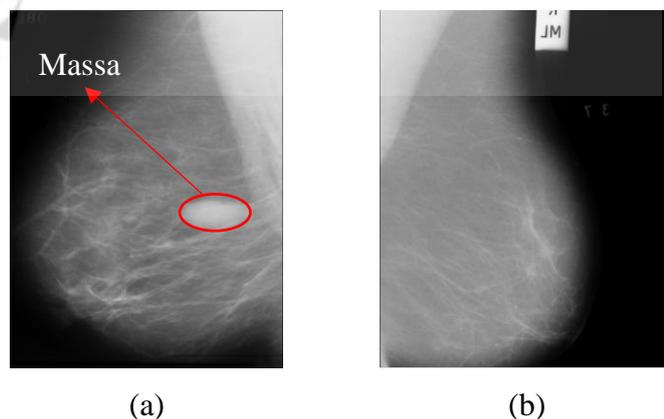
Kanker payudara merupakan salah satu penyakit yang mematikan. Kanker payudara terjadi karena pertumbuhan yang berlebih dan perkembangan

sel jaringan yang tidak terkontrol dan terbentuklah sebuah benjolan. Kanker payudara biasanya tumbuh pada bagian sel lobulus yaitu kelenjar yang menghasilkan susu juga pada salurannya (Biswas et al., 2016).

Deteksi kanker payudara dapat dilakukan dengan teknologi pencitraan medis yang saat ini berkembang. Teknologi tersebut memungkinkan membantu bidang kedokteran dalam melakukan penafsiran dan diagnosa penyakit, seperti pada kasus ini mendeteksi kanker payudara. Citra medis tersebut dapat dihasilkan dari berbagai macam peralatan yang digunakan di bidang kedokteran, seperti *Ultra Sound Graphic* (USG), MRI, dan Mammogram.

2.2.3. Mammogram

Mammografi adalah proses skrining dan pemeriksaan radiologi khusus menggunakan sinar-X dengan dosis rendah yang digunakan untuk mengidentifikasi adanya keabnormalan pada payudara seperti kanker (Hariraj et al., 2017). Hasil dari mammografi disebut mammogram. Keabnormalan atau kelainan pada payudara dapat dilihat dari keberadaan massa yang mengindikasikan kanker atau tumor. Massa adalah area terdapatnya pola tekstur dengan bentuk serta batas tertentu pada mammogram (Junita, 2017).



Gambar 2.6 Citra mammogram

a) Abnormal, b) Normal

Sumber : (Suckling, et al., 1994)

2.2.4. Grayscale

Citra *grayscale* atau citra keabuan adalah citra yang memiliki intensitas pikselnya dari nilai 0 sampai 255. Nilai 0 mewakili warna hitam dan nilai 255 warna putih (Kadir & Adhi, 2013). *Grayscale* adalah proses konversi citra menjadi citra *grayscale* dengan cara mengurangi dimensi yang dimiliki citra yaitu citra yang semula memiliki tiga komponen RGB (*Red, Green, Blue*) dikonversi hanya memiliki satu komponen warna abu dengan *range* piksel 0 sampai 255. Konversi tersebut digunakan perhitungan menggunakan Persamaan 2.1 sebagai berikut (Listia, Harjoko, Mada, & Utara, 2014):

$$\text{Grayscale} = 0.21 * R + 0.72 * G + 0.07 * B \quad (2.1)$$

Dimana,

R = intensitas piksel komponen *red*

G = intensitas piksel komponen *green*

B = intensitas piksel komponen *blue*

2.2.5. Segmentasi

Segmentasi adalah proses membagi gambar menjadi segmen dengan tujuan memisahkan obyek yang akan diteliti dengan *background* seperti *k-means clustering* dan *thresholding*.

2.2.5.1. K-Means Clustering

Algoritma *k-means clustering* digunakan untuk segmentasi, memisahkan daerah sebagai obyek dan bukan. *K-means* mengelompokkan obyek sebanyak n (jumlah data) dengan atribut ke dalam kelompok k (*cluster*), dimana $k < n$ (Novar Setiawan & Suwija Putra, 2018). Mencari jarak terdekat dengan *centroid* secara berulang. Langkah – langkahnya sebagai berikut.

- a. Tentukan jumlah k (*cluster*). Nilai k ini adalah banyak segmen yang akan nanti dihasilkan dimana k adalah bilangan positif.
- b. Tentukan posisi *centroid* awal.
- c. Hitung jarak *centroid* dengan obyek lain menggunakan *euclidean distance*. Berikut Persamaan 2.2 (Novar Setiawan & Suwija Putra, 2018) *euclidean distance*.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.2)$$

Dimana,

$d(x,y)$ = hasil *euclidean distance*

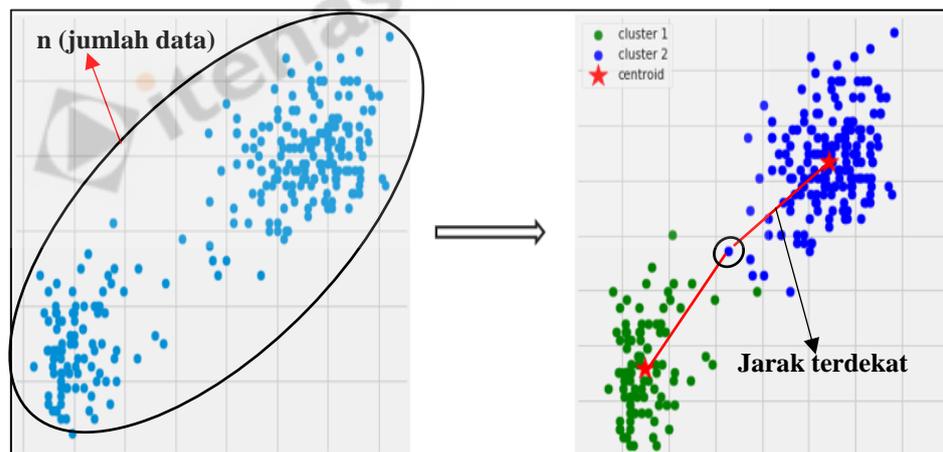
n = jumlah data

x_i = data x ke-i

y_i = data y ke-i

- d. Mengelompokkan dengan menetapkan piksel citra ke dalam setiap *cluster* terdekat antara k.
- e. Perbarui *centroid* dengan menghitung rata-rata piksel yang ditetapkan dalam *cluster* yang sesuai. Bandingkan dengan *centroid* sebelumnya jika *centroid* berubah maka ulangi proses c dan d hingga *centroid* stabil. Stabil pada proses ini yaitu nilai *centroid* sebelum dan sesudah memiliki nilai yang sama.

Berikut Gambar 2.7 merupakan contoh sebelum dan sesudah proses *k-means clustering*.



Gambar 2.7 Ilustrasi *k-means clustering*

Sumber : (Dabbura, 2018)

2.2.5.2. Thresholding

Thresholding merupakan metode segmentasi yang digunakan untuk memisahkan obyek dengan *background* berdasarkan kecerahannya (Li et al., 2018). *Thresholding* digunakan sebagai pembatas nilai piksel pada citra. Pembatas tersebut terdiri dari nilai batas atas dan bawah. Hasil citra dari

thresholding akan berupa citra biner yaitu piksel bernilai 255 atau 0. Gambar 2.8 merupakan contoh proses *thresholding*.

104	223	50	$T=100$ \Rightarrow	1	1	0
12	255	0		0	1	0
46	120	100		0	1	1

Gambar 2.8 Ilustrasi *thresholding*

Berikut Persamaan 2.3 yang digunakan untuk *thresholding* (Wijaya, Chamidah, & Santoni, 2019).

$$g(x) \begin{cases} 255, & \text{jika } f(x, y) \geq T \\ 0, & \text{jika tidak } < T \end{cases} \quad (2.3)$$

Dimana,

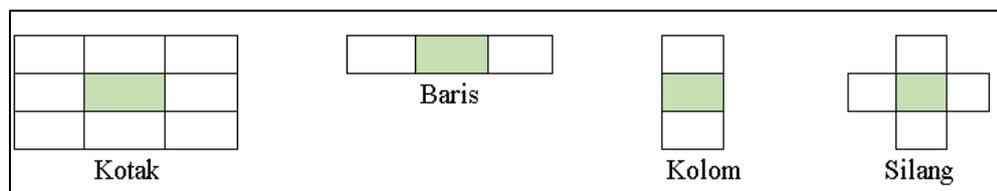
$g(x)$ = citra biner

$f(x, y)$ = piksel citra

T = *threshold*

2.2.6. Morfologi

Morfologi adalah teknik pengolahan citra berdasarkan bentuk segmen citra yang terdiri dari dua *array*. *Array* pertama sebagai masukan citra yang akan dilakukan proses *morfologi* sedangkan yang kedua berisi *strel* atau *structure element* (SE) (Luthfi, Amin, & Cahyono, 2019). Gambar 2.9 merupakan bentuk *strel* yang biasanya digunakan dalam proses morfologi.



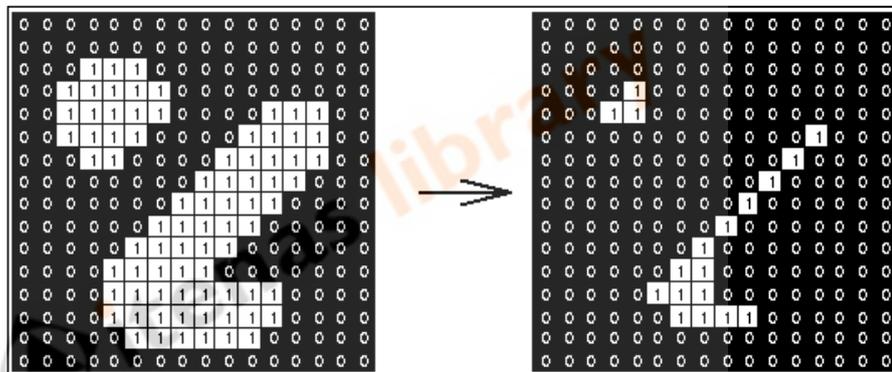
Gambar 2.9 Bentuk *strel*

Bentuk operasi *morfologi* yang digunakan dalam pemrosesan citra digital yaitu dilasi, erosi, *opening*, dan *closing* (Kadir & Adhi, 2013). Pada penelitian ini yang digunakan yaitu *opening*.

2.2.6.1. Opening

Opening adalah proses *morfologi* yang digunakan untuk citra biner. Tahapan *opening* yang dilakukan erosi terlebih dahulu lalu selanjutnya dilakukan dilasi. Prosesnya *strel* atau SE yang sudah ditentukan akan dicocokkan dengan tiap piksel masukan citra biner. *Strel* dengan titik poros akan diletakkan tiap pikselnya. Untuk penjelasannya berikut ini.

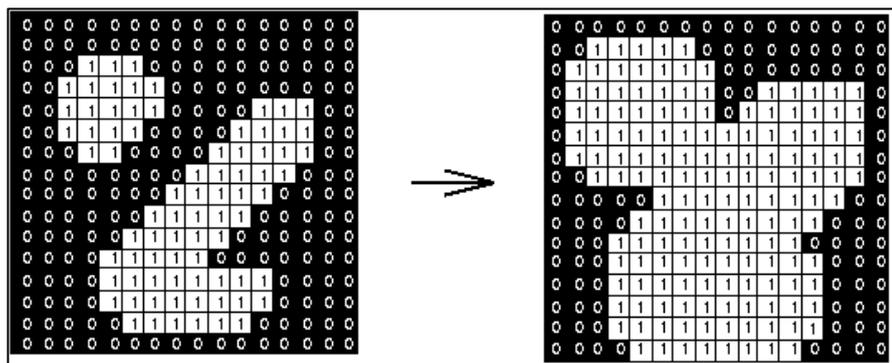
- a. Erosi adalah proses menipiskan atau mengecilkan tepi dari sebuah piksel. Prosesnya jika *strel* berada pada bagian luar masukan citra maka titik poros yang ditempelkan pada piksel dengan koordinatnya akan bernilai 0 atau berwarna hitam. Gambar 2.10 merupakan contoh proses erosi pada citra.



Gambar 2.10 Ilustrasi erosi

Sumber : (Auckland University, 2010)

- b. Dilasi adalah kebalikan dari erosi, menebalkan atau memperbesar di tepian piksel. Prosesnya piksel yang tertimpa dengan *strel* maka piksel akan bernilai 255 atau berwarna putih. Gambar 2.10 merupakan contoh proses erosi pada citra.

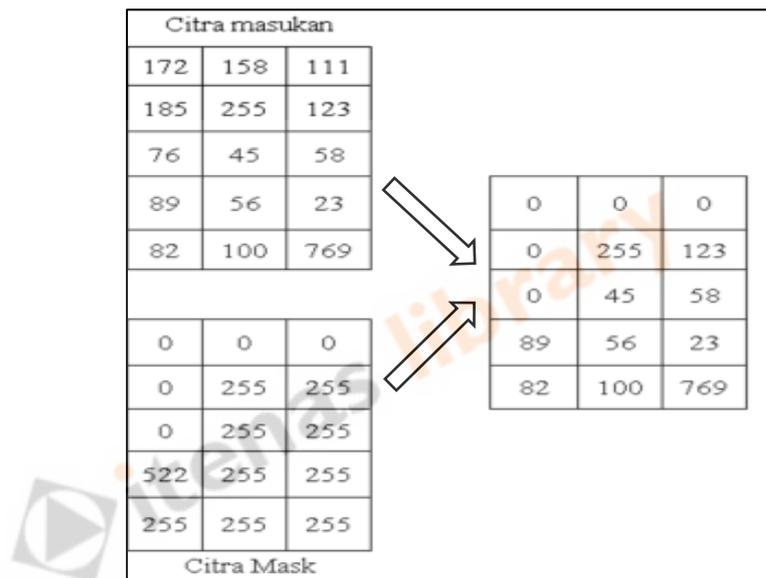


Gambar 2.11 Ilustrasi dilasi

Sumber : (Auckland University, 2010)

2.2.6.2. Masking

Masking adalah proses menggabungkan dua obyek yang dimana obyek yang mirip akan ditimpa sehingga menghasilkan obyek yang semula. Pada umumnya bertujuan untuk mengembalikan piksel citra awal dari hasil segmentasi yang berupa citra biner.

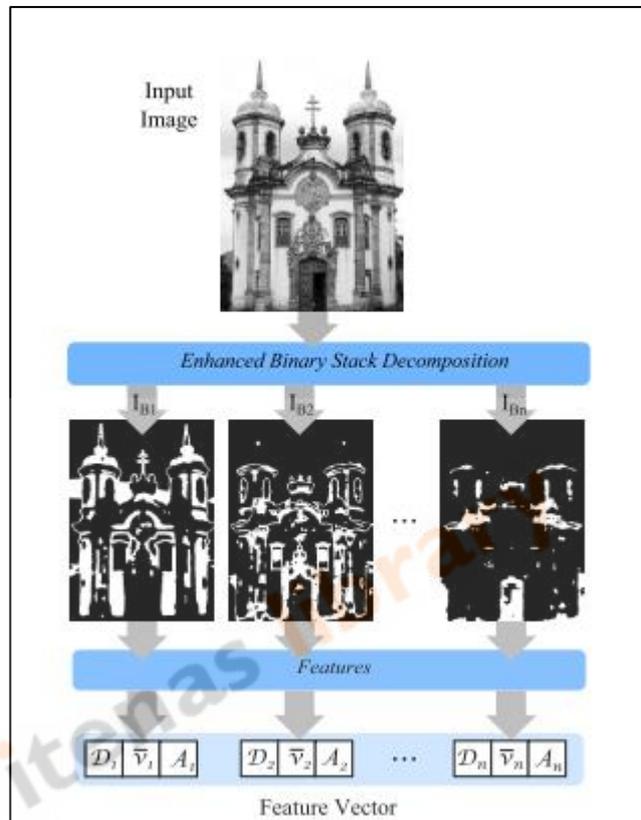


Gambar 2.12 Ilustrasi *masking*

2.2.7. Segmentation-based Fractal Texture Analysis (SFTA)

SFTA adalah metode yang digunakan untuk ekstraksi fitur fraktal (Ergen & Baykara, 2014). Metode ini terdiri dari dua bagian, yang pertama proses *two threshold binary decomposition* (TTBD) pada gambar keabuan dan lalu diubah menjadi gambar biner yang berbeda, dan tahap kedua tahap ekstraksi properti dari setiap gambar biner yaitu menghitung *fractal dimension* dari batas tepi citra biner, *mean* dan *area* dari setiap citra biner (Öztürk & Akdemir, 2018). Dalam algoritma TTBD, pertama nilai batas ambang *T* dihitung menggunakan *multilevel otsu thresholding* yang meminimalkan nilai *within class variance*. Algoritma *otsu* diterapkan pada setiap citra hingga jumlah ambang mencapai nilai *nt* (*number of thresholding*), di mana *nt* adalah

parameter yang ditentukan pengguna. Nilai ambang dipilih berdasarkan nilai *within class variance* terkecil (Paramkusham, Rao, & Prabhakar Rao, 2018). Berikut Gambar 2.13 Merupakan proses metode dari SFTA.



Gambar 2.13 Alur proses SFTA

Sumber : (Costa et al., 2012)

Berikut Persamaan 2.4 digunakan untuk mencari *within class variance* terkecil (Yousefi, 2015).

$$\text{Within Class Variance } (\sigma_w^2) = (W_b \cdot \sigma_b^2) + (W_f \cdot \sigma_f^2) \quad (2.4)$$

Dimana :

- W_b adalah nilai dari *weight background*. Nilai dari *weight background* didapatkan dari Persamaan 2.5 berikut.

$$W_b(t) = \sum_{i=1}^t P(i) \quad (2.5)$$

- μ_b adalah nilai dari *mean background*. Nilai dari *mean background* didapatkan dari persamaan 2.6 berikut.

$$\mu_b(t) = \frac{\sum_{i=1}^t i \cdot P(i)}{W_b(t)} \quad (2.6)$$

- W_f adalah nilai dari *weight foreground*. Nilai dari *weight foreground* didapatkan dari persamaan 2.7 berikut.

$$W_f(t) = \sum_{i=t+1}^l P(i) \quad (2.7)$$

- μ_f adalah nilai dari *mean foreground*. Nilai dari *mean foreground* didapatkan dari persamaan 2.8 berikut.

$$\mu_f(t) = \frac{\sum_{i=t+1}^l i * P(i)}{W_f(t)} \quad (2.8)$$

- σ_b^2 adalah nilai dari *variance background*. Nilai dari *variance background* didapatkan dari persamaan 2.9 berikut.

$$\sigma_b^2(t) = \frac{\sum_{i=1}^l (i - \mu_b(t))^2 * P(i)}{W_b(t)} \quad (2.9)$$

- σ_f^2 adalah nilai dari *variance foreground*. Nilai dari *variance foreground* didapat dari persamaan 2.10 berikut.

$$\sigma_f^2(t) = \frac{\sum_{i=t+1}^l (i - \mu_f(t))^2 * P(i)}{W_f(t)} \quad (2.10)$$

Dimana,

$P(i)$ = nilai propabilitas histogram dari citra *grayscale*.

t = nilai *threshold*.

l = nilai piksel tertinggi pada histogram.

i = nilai piksel terendah pada histogram.

Berikut Persamaan 2.11 merupakan set pertama dan Persamaan 2.12 merupakan set kedua untuk menghasilkan citra biner (Krishna M, Edwin, & Hariharan, 2017).

$$I_{b1}(x, y) = \begin{cases} 1, & \text{jika } I(x, y) \geq t_l \\ 0, & \text{jika tidak} \end{cases} \quad (2.11)$$

$$I_{b2}(x, y) = \begin{cases} 1, & \text{jika } t_l < I(x, y) \leq t_u \\ 0, & \text{jika tidak} \end{cases} \quad (2.12)$$

Dimana,

x = baris matriks

y = kolom matriks

t_l = batas bawah keabuan

t_u = batas atas keabuan

$I(x,y)$ = gambar keabuan

$I_b(x,y)$ = gambar biner

Berdasarkan Persamaan 2.11 dan 2.12 maka hasil citra biner merupakan 2 kali nt (*number of thresholding*). Persamaan 2.13 digunakan untuk menentukan batas dari citra biner (Krishna M et al., 2017).

$$\Delta(x,y) = \begin{cases} 255, & \text{jika ya } \exists(x',y') \in N_8[(x,y)] : I_b(x',y') = 0 \cap 255 \\ 0, & \text{jika tidak} \end{cases} \quad (2.13)$$

Dimana,

x = baris matriks

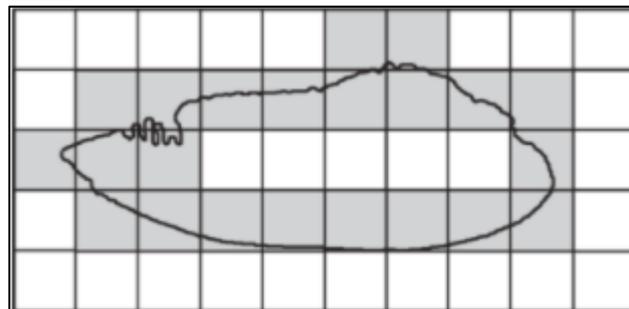
y = kolom matriks

$N_8[(x,y)]$ = 8 piksel yang terhubung ke (x,y) (ketetanggaan)

$I_b(x,y)$ = gambar biner

$\Delta(x,y)$ = batas gambar biner

Batas dari citra biner akan digunakan untuk mencari nilai *fractal dimension*. Nilai tersebut didapatkan dengan menggunakan cara *box counting*. *box counting* yaitu menutup citra masukan dengan bentuk kotak dan hitung berapa banyak bentuk kotak yang menutupi citra dan ulangi proses tersebut sampai sebuah struktur pola terbentuk (Krishna et al ., 2017). Struktur pola terbentuk mengikuti tepi dari sebuah citra. Gambar 2.14 merupakan ilustrasi dari *box counting* pada sebuah batas.



Gambar 2.14 Ilustrasi *box counting*

Berikut Persamaan 2.15 yang digunakan untuk menghitung nilai *fractal dimension* (Krishna et al ., 2017).

$$D = \frac{\log(N)}{\log(r)} \quad (2.14)$$

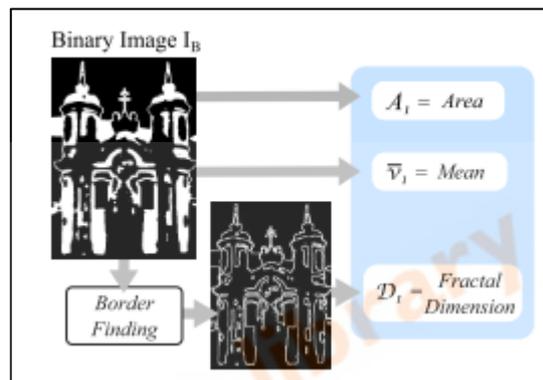
Dimana,

$D = \text{fractal dimension}$

$N = \text{banyak kotak yang menutupi citra}$

$r = \text{kebalikan dari ukuran kotak}$

Hasil ekstraksi fitur SFTA menghasilkan nilai *fractal dimension*, *mean*, dan *area* yang seperti ditunjukkan pada Gambar 2.16.



Gambar 2.15 Ekstraksi Fitur SFTA tiap citra biner

Sumber : (Costa et al., 2012)

2.2.8. Support Vector Machine (SVM)

SVM adalah metode *machine learning* yang digunakan untuk proses klasifikasi atau regresi dengan tujuan menemukan *hyperplane* terbaik yang memisahkan dua buah *class* pada *input space* (Novar Setiawan & Suwija Putra, 2018). SVM banyak digunakan sebagai klasifikasi untuk di berbagai penelitian seperti tentang diagnosis sebuah penyakit. SVM juga dapat mengatasi masalah klasifikasi dan regresi dengan *linear* maupun *non linear*.

a. Klasifikasi *linear*

Klasifikasi yang menggunakan data variabel dengan target kategori. SVM menggunakan *hyperplane* satu dimensi (garis) yang memisahkan dua kelas berdasarkan kategori yang ditentukan.

b. Klasifikasi *non linear*

SVM juga dapat menangani masalah klasifikasi *non linear* dengan cara menggunakan fungsi kernel bertujuan memetakan data ke ruang dimana

hyperplane digunakan dalam melakukan pemisahan. Fungsi kernel juga dapat mengubah data menjadi dimensi ruang yang lebih tinggi.

Berikut Persamaan 2.15 untuk mencari *hyperplane* untuk data latih (Kowalczyk, 2017).

$$h(x) = wx + b \quad (2.15)$$

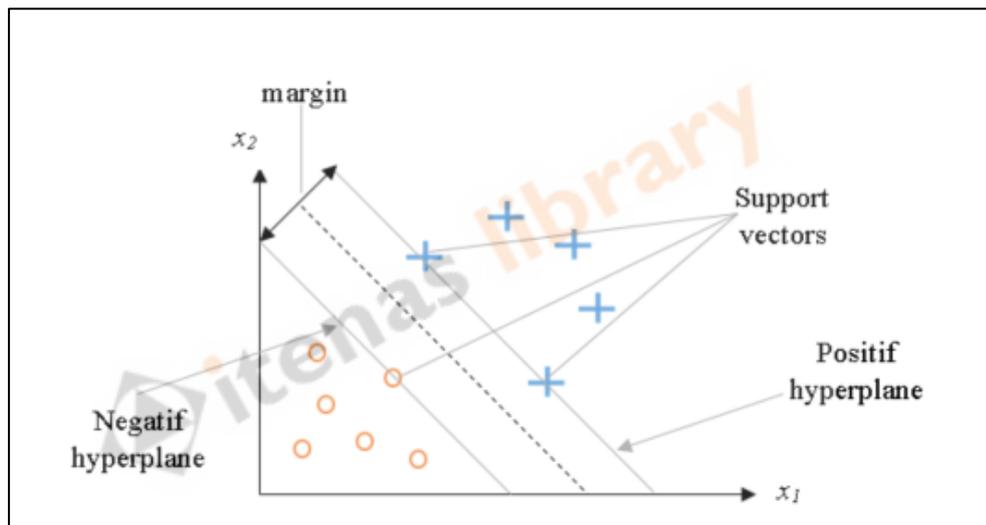
Dimana,

$h(x)$ = *hyperplane*

x = fitur

w = normal bidang

b = posisi bidang relatif terhadap pusat koordinat



Gambar 2.16 *Hyperplane* yang memisahkan dua kelas

Hyperlane digunakan untuk membagi data *input* menjadi dua buah kelas. Data *input* berupa nilai fitur vektor disimbolkan dengan $x = (x_1, x_2)$, dimana vektor berkoordinat dengan sumbu x_1 dan x_2 . SVM mencari *hyperplane* ini berdasarkan *support vectors* dan margin. *Support vectors* merupakan seluruh vektor data yang berjarak paling mendekati *hyperplane*, sedangkan margin merupakan lebar dari *separating hyperplane*.

2.2.9. Pengukuran Kinerja Sistem

Pengukuran kinerja sistem dilakukan dengan menghitung akurasi, presisi, dan *recall*. Pengujian menghitung akurasi, presisi, *recall*, dan *fi-score*. Akurasi adalah ketepatan prediksi benar (positif dan negatif) dari keseluruhan data.

Preisisi adalah ketepatan prediski benar positif dari keseluruhan hasil yang diprediksi positif. *Recall* adalah ketepatan prediksi dari keseluruhan data yang benar positif. *F1-score* adalah ketepatan prediksi negatif dari keseluruhan data negatif. Pengukuran kinerja sistem menggunakan persamaan berikut (Fajrin, Nugroho, & Soesanti, 2015):

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+FN+TN} \times 100\% \quad (2.16)$$

$$\text{Presisi} = \frac{TP}{TP+FP} \times 100\% \quad (2.17)$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \quad (2.18)$$

$$\text{F1-score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \times 100\% \quad (2.19)$$

Dimana,

TP = *True Positive*

FP = *False Positive*

TN = *True Negative*

FN = *False Negative*

Berkaitan pada penelitian mendeteksi massa citra mamogram dijelaskan bahwa *True Positive* (TP) adalah terklasifikasi secara benar untuk citra mammogram abnormal, sedangkan *False Positive* (FP) adalah terklasifikasi salah, misalnya citra mammogram tidak terdapat massa namun terklasifikasi sebagai citra abnormal. *True Negative* (TN) adalah terklasifikasi secara benar untuk citra mammogram normal, sedangkan *False Negative* (FN) adalah terklasifikasi salah, misalnya citra mammogram ada keberadaan massa tetapi, terklasifikasi sebagai citra normal.