

**PENINGKATAN HASIL DIAGNOSIS IDC (*INVASIVE DUCTAL
CARCINOMA*) DARI HASIL CITRA HISTOPATOLOGI
MENGUNAKAN METODE EKSTRAKSI CIRI DAN
KLASIFIKASI**

SKRIPSI

untuk memenuhi salah satu persyaratan
mencapai derajat Sarjana S1



Disusun oleh:

VERA GIYANING TIYAS

16524079

**Jurusan Teknik Elektro
Fakultas Teknologi Industri
Universitas Islam Indonesia
Yogyakarta**

2021

LEMBAR PENGESAHAN

**PENINGKATAN HASIL DIAGNOSIS IDC (*INVASIVE DUCTAL CARCINOMA*) DARI
HASIL CITRA HISTOPATOLOGI MENGGUNAKAN METODE EKSTRAKSI CIRI
DAN KLASIFIKASI**

TUGAS AKHIR

**Diajukan sebagai Salah Satu Syarat untuk Memperoleh
Gelar Sarjana Teknik
pada Program Studi Teknik Elektro
Fakultas Teknologi Industri
Universitas Islam Indonesia**

Disusun oleh:

**Vera Giyaning Tiyas
16524079**

Yogyakarta, 24-02-2021

Menyetujui,

Pembimbing

318091

**Elvira Sukma Wahyuni, S.Pd, M.Eng
155231301**

LEMBAR PENGESAHAN

SKRIPSI

PENINGKATAN HASIL DIAGNOSIS IDC (*INVASIVE DUCTAL CARCINOMA*) DARI HASIL CITRA HISTOPATOLOGI MENGGUNAKAN METODE EKSTRAKSICIRI DAN KLASIFIKASI

Dipersiapkan dan disusun oleh:

Vera Giyaning Tiyas

16524079

Telah dipertahankan di depan dewan penguji

Pada tanggal: 10 Maret 2021

Susunan dewan penguji

Ketua Penguji : Elvira Sukma Wahyuni, S.Pd, M.Eng,

Anggota Penguji 1: Dr.Eng. Hendra Setiawan, ST., MT.

Anggota Penguji 2: Dzata Farahiyah, ST, M.Sc

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana

Tanggal: 1 April 2021

Ketua Program Studi Teknik Elektro



Yusuf Aziz Amrulloh, S.T., M.Eng., Ph.D

045240101

PERNYATAAN

PERNYATAAN

Dengan ini Saya menyatakan bahwa:

1. Skripsi ini tidak mengandung karya yang diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan Saya juga tidak mengandung karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.
2. Informasi dan materi Skripsi yang terkait hak milik, hak intelektual, dan paten merupakan milik bersama antara tiga pihak yaitu penulis, dosen pembimbing, dan Universitas Islam Indonesia. Dalam hal penggunaan informasi dan materi Skripsi terkait paten maka akan diskusikan lebih lanjut untuk mendapatkan persetujuan dari ketiga pihak tersebut diatas.

Yogyakarta, 24 Februari 2021



Vera Giyaning Tiyas

KATA PENGANTAR

Bismillahirrahmanirrahim

Assalamu'alaikum Warahmatullahi Wabarakatuh

Dengan memanjatkan puji syukur kehadiran Allah SWT. Atas limpahan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Peningkatan Hasil Diagnosis IDC (*Invasive Ductal Carcinoma*) Dari Hasil Citra Histopatologi Menggunakan Metode Ekstraksi Ciri dan Klasifikasi”. Tidak lupa *shalawat* serta salam kita panjatkan pada Nabi Muhammad SAW yang telah membimbing umatnya menjadi pribadi yang lebih baik. Skripsi ini disusun dalam rangka memenuhi prasyarat akademis untuk mencapai gelar kesarjanaan Strata 1 pada Jurusan Teknik Elektro, Fakultas Teknologi Industri, Universitas Islam Indonesia.

Dalam menyelesaikan skripsi ini penulis telah mendapatkan banyak bantuan, dorongan, bimbingan, dan arahan dari berbagai pihak. Hal ini dikarenakan keterbatasan ilmu dan pengalaman yang penulis miliki. Oleh karena itu, dengan rasa hormat dan terimakasih kepada semua yang telah mendukung penulis untuk menyelesaikan skripsi, terutama kepada:

1. Orang Tua saya yang selalu memberikan dukungan, doa, dorongan, kasih sayang, nasihat selama proses penelitian untuk menyelesaikan laporan tugas akhir ini hingga selesai, serta bantuan baik material maupun spiritual yang telah diberikan kepada saya selama ini.
2. Adikku tersayang, Anwar Halim Nurhadi. Terimakasih atas doa dan dukungan yang telah diberikan dalam menyelesaikan skripsi ini.
3. Ibu Elvira Sukma Wahyuni, S.Pd, M.Eng, selaku Dosen Pembimbing, yang telah memberi banyak bantuan dan pengarahan hingga terselesaikan laporan Skripsi ini.
4. Bapak Yusuf Aziz Amrullah, S.T., M.Eng., Ph.D., selaku Ketua Jurusan Teknik Elektro, Universitas Islam Indonesia.
5. Seluruh Dosen Jurusan Teknik Elektro, terima kasih atas bimbingan selama menempuh kuliah dari semester pertama hingga akhir di Jurusan Teknik Elektro.
6. Seluruh staf Jurusan Teknik Elektro yang sudah mendampingi, membimbing, dan memberi bantuan selama menempuh kuliah di Jurusan Teknik Elektro Universitas Islam Indonesia.
7. Teman-teman saya Retno Paras Rasmi, Nabilah Haptriani, Annisa Christyanti, Rachel Kurnia, Yunira Alifah Dzakiyah, Ryan Andri yang selalu mendampingi dan membantu saya selama menempuh kuliah maupun menyelesaikan laporan skripsi ini.

8. Teman-teman mahasiswa Jurusan Teknik Elektro Universitas Islam Indonesia angkatan 2016.
9. Terakhir untuk segenap pihak yang telah membantu penulis, yang tidak bisa disebutkan satu persatu, terima kasih telah memberikan semangat dan dorongan hingga akhirnya skripsi ini dapat selesai dengan baik dan sesuai harapan. Semoga Allah membalas kebaikan kalian semua.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Yogyakarta, 24 Februari 2021



Vera Giyaning Tiyas



ARTI LAMBANG DAN SINGKATAN

σ	= Standar deviasi citra histogram
σ_x	= Standar deviasi elemen kolom
σ_y	= Standar deviasi elemen baris
σ^2	= Tingkat keragaman pada citra histogram (<i>variance</i>)
μ	= Nilai rata-rata citra histogram (<i>mean</i>)
μ_x	= Nilai rata-rata elemen kolom
μ_y	= Nilai rata-rata elemen baris
b	= Normalisasi <i>blue</i>
B	= <i>Blue</i>
CNN	= <i>Convolutional Neural Network</i>
DCIS	= <i>Ductal Carcinoma In Situ</i>
FN	= <i>False Negative</i>
FP	= <i>False Positive</i>
g	= Normalisasi <i>green</i>
G	= <i>Green</i>
GLCM	= <i>Gray-Level Co-Occurrence Matrix</i>
H	= <i>Hue</i>
HSV	= <i>Hue Saturation Value</i>
i	= Baris pada piksel
IDC	= <i>Invasive Ductal Carcinoma</i>
ILC	= <i>Invasive lobular carcinoma</i>
j	= Kolom pada piksel
$p(i,j)$	= Piksel baris ke- i kolom ke- j
PNG	= <i>Portable Network Graphics</i>
r	= Normalisasi <i>red</i>
R	= <i>Red</i>
RGB	= <i>Red Green Blue</i>
S	= <i>Saturation</i>
SVM	= <i>Support Vector Machine</i>
TN	= <i>True Negative</i>
TP	= <i>True Positive</i>
V	= <i>Value</i>

ABSTRAK

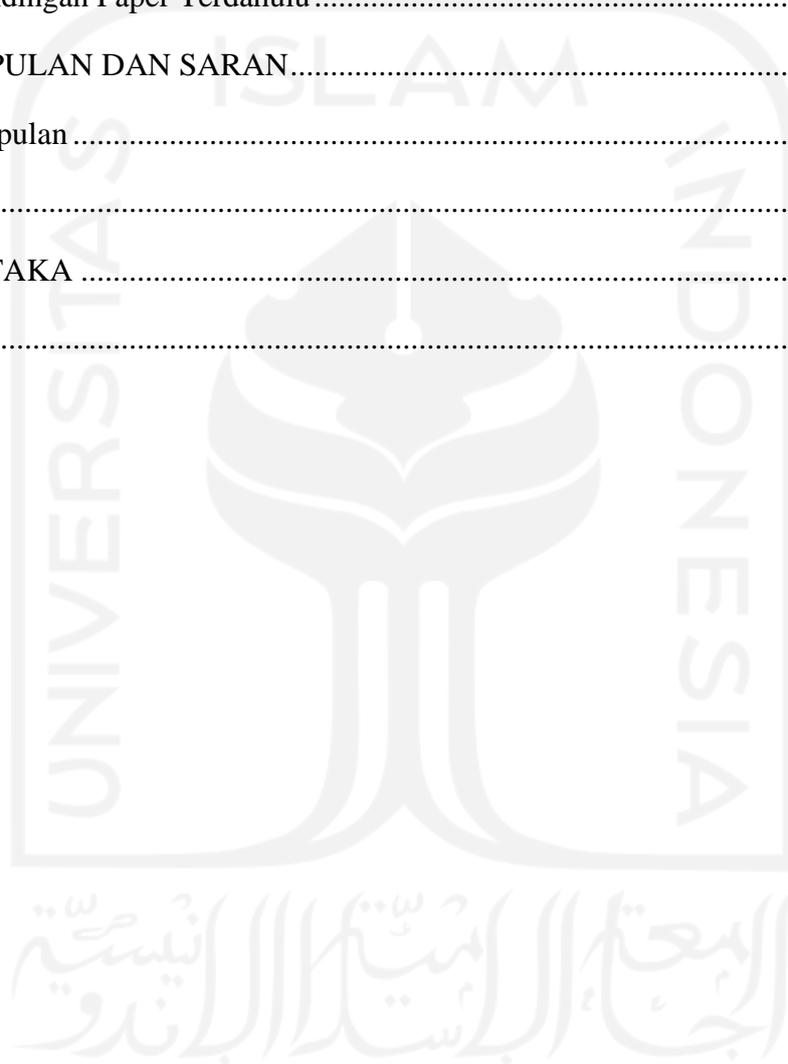
IDC (*Invasive Ductal Carcinoma*) merupakan salah satu dari jenis kanker payudara yang paling umum terjadi, hampir 70-80% dari semua diagnosis kanker payudara. Pendeteksian kanker payudara merupakan upaya untuk pencegahan dan pengendalian terjadinya kanker payudara yang dapat dilakukan menggunakan berbagai metode seperti menggunakan *x-ray* untuk mengambil gambar *mammogram* dan pemindaian menggunakan *ultrasound*. Langkah pertama dalam melakukan pengobatan kanker payudara adalah pemeriksaan yang tepat, biasanya dilakukan dengan gambar biomedis seperti *mammogram* dan histopatologi. Adapun metode yang efektif dalam mendeteksi adanya kanker payudara yaitu dengan pemeriksaan jaringan tubuh (biopsi) dan melakukan analisis histopatologi. Pada penelitian ini, deteksi citra histopatologi dilakukan dengan beberapa tahapan yaitu input data histopatologi, ekstraksi fitur, dan klasifikasi. Ekstraksi fitur merupakan tahapan paling penting pada proses klasifikasi. Pada penelitian ini akan dilakukan sebuah ekstraksi dengan menggabungkan beberapa ekstraksi fitur diantaranya yaitu GLCM (*Gray Level Co-occurrence Matrix*), RGB (*Red Green Blue*), HSV (*Hue Saturation Value*), dan Histogram. Hasil terhadap ekstraksi dari fitur tersebut akan dilakukan klasifikasi menggunakan *classifier backpropagation* dan SVM (*Support Vector Machine*). *Dataset* yang digunakan berupa *breast histopathology images*, didapat pada situs Kaggle yang tersedia untuk umum berjumlah 1080 citra IDC (540 IDC negatif dan 540 IDC positif). Dari uji coba yang telah dilakukan didapatkan akurasi tertinggi yaitu 96,94% berdasarkan gabungan semua fitur dengan *classifier backpropagation*.

Kata Kunci : Ekstraksi fitur, IDC, Histopatologi, GLCM, RGB, HSV, Histogram.

DAFTAR ISI

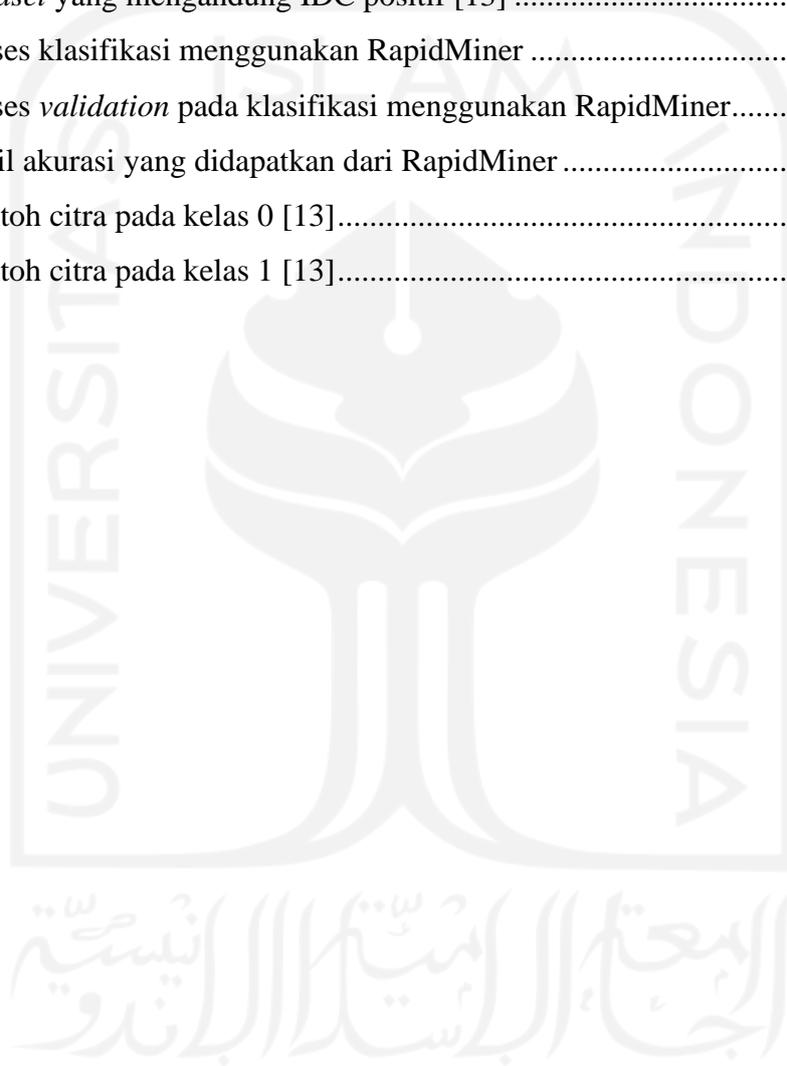
LEMBAR PENGESAHAN.....	i
PERNYATAAN.....	iii
KATA PENGANTAR.....	iv
ARTI LAMBANG DAN SINGKATAN	vi
ABSTRAK	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR	x
DAFTAR TABEL.....	xi
BAB 1 PENDAHULUAN	12
1.1 Latar Belakang Masalah	12
1.2 Rumusan Masalah.....	13
1.3 Batasan Masalah	13
1.4 Tujuan Penelitian	14
1.5 Manfaat Penelitian	14
BAB 2 TINJAUAN PUSTAKA	15
2.1 Studi Literatur.....	15
2.2 Tinjauan Teori.....	16
2.2.1 Kanker Payudara.....	16
2.2.2 Citra Digital.....	17
2.2.3 Ekstraksi Fitur	17
2.2.4 <i>Backpropagation</i>	18
2.2.5 SVM (<i>Support Vector Machine</i>).....	19
BAB 3 METODOLOGI.....	20
3.1 Alur penelitian	20
3.1.1 <i>Dataset</i> (Histopatologi).....	21

3.1.2 Ekstraksi Fitur	22
3.1.3 Proses Klasifikasi	26
3.2 Hasil dan Unjuk Kerja	27
BAB 4 HASIL DAN PEMBAHASAN.....	30
4.1 Hasil Ekstraksi	30
4.2 Hasil Klasifikasi.....	33
4.3 Perbandingan Paper Terdahulu	38
BAB 5 KESIMPULAN DAN SARAN.....	40
5.1 Kesimpulan	40
5.2 Saran	40
DAFTAR PUSTAKA	41
LAMPIRAN	43



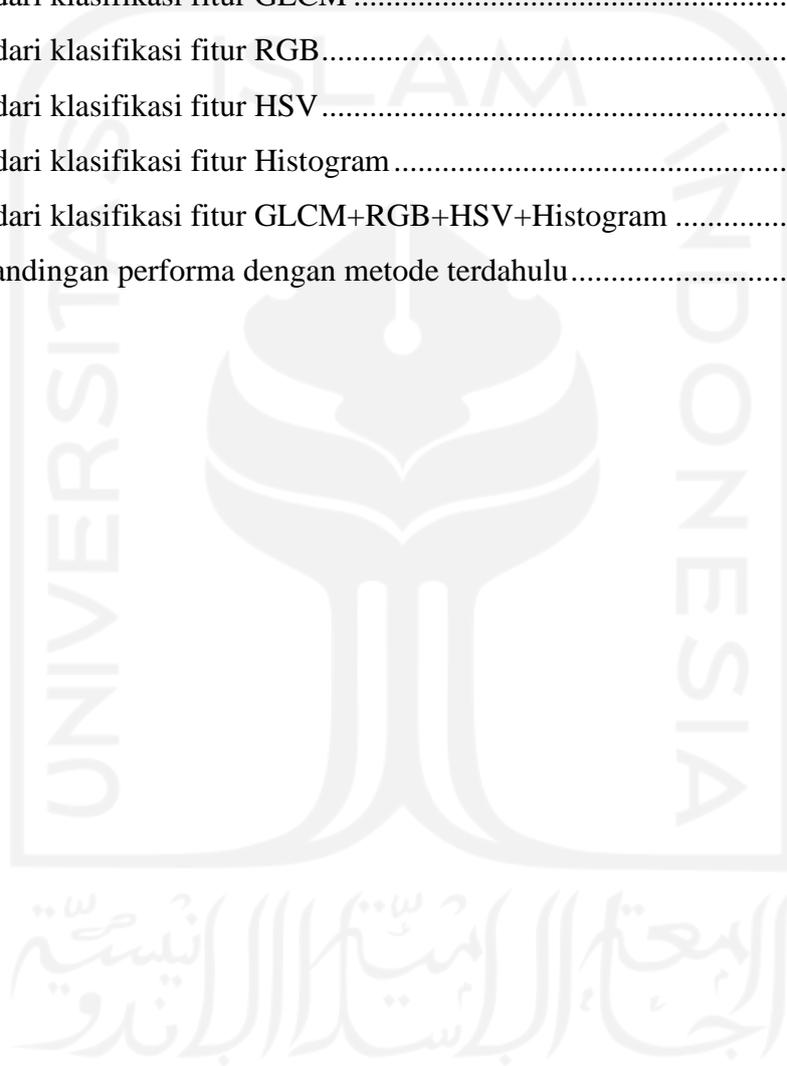
DAFTAR GAMBAR

Gambar 2.1 Arsitektur jaringan <i>Backpropagation</i> [11]	18
Gambar 2.2 Ilustrasi <i>Support Vector Machine</i> [12]	19
Gambar 3.1 Diagram blok alur penelitian yang dilakukan.	20
Gambar 3.2 <i>Dataset</i> yang mengandung IDC negatif [13]	21
Gambar 3.3 <i>Dataset</i> yang mengandung IDC positif [13]	22
Gambar 3.4 Proses klasifikasi menggunakan RapidMiner	27
Gambar 3.5 Proses <i>validation</i> pada klasifikasi menggunakan RapidMiner.....	27
Gambar 3.6 Hasil akurasi yang didapatkan dari RapidMiner	28
Gambar 4.1 Contoh citra pada kelas 0 [13].....	30
Gambar 4.2 Contoh citra pada kelas 1 [13].....	30



DAFTAR TABEL

Tabel 3.1 <i>Confusion matrix</i> [12]	28
Tabel 4.1 Hasil ekstraksi dari GLCM	31
Tabel 4.2 Hasil ekstraksi dari RGB	31
Tabel 4.3 Hasil ekstraksi dari HSV	32
Tabel 4.4 Hasil ekstraksi dari Histogram	33
Tabel 4.5 Hasil dari klasifikasi fitur GLCM	34
Tabel 4.6 Hasil dari klasifikasi fitur RGB	35
Tabel 4.7 Hasil dari klasifikasi fitur HSV	36
Tabel 4.8 Hasil dari klasifikasi fitur Histogram	36
Tabel 4.9 Hasil dari klasifikasi fitur GLCM+RGB+HSV+Histogram	37
Tabel 4.10 Perbandingan performa dengan metode terdahulu	38



BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Kanker payudara adalah salah satu penyakit yang paling umum diderita oleh perempuan. Berdasarkan perkiraan *World Health Organization* (WHO) berdampak pada 2,1 juta wanita setiap tahun dan juga menyebabkan jumlah terbesar kematian terkait kanker di antara wanita. Pada tahun 2018, diperkirakan 627.000 wanita meninggal karena kanker payudara yaitu sekitar 15% dari semua kematian akibat kanker di kalangan wanita [1]. IDC merupakan salah satu dari jenis kanker payudara yang paling umum terjadi, hampir 70-80% dari semua diagnosis kanker payudara. Menurut *American Cancer Society*, di negara Amerika Serikat terdapat lebih dari 18.000 wanita setiap tahunnya yang diketahui bahwa menderita kanker payudara. Sebagian besar dari mereka didiagnosis sebagai kanker payudara jenis IDC [2].

Pendeteksian kanker payudara merupakan upaya untuk pencegahan dan pengendalian terjadinya kanker payudara yang dapat dilakukan menggunakan berbagai metode seperti menggunakan *x-ray* untuk mengambil gambar *mammogram* dan pemindaian menggunakan *ultrasound*. Langkah pertama dalam melakukan pengobatan kanker payudara adalah pemeriksaan yang tepat, biasanya dilakukan dengan gambar biomedis seperti *mammogram* dan histopatologi. Adapun metode yang efektif dalam mendeteksi adanya kanker payudara yaitu dengan pemeriksaan jaringan tubuh (biopsi) dan melakukan analisis histopatologi. Pemeriksaan secara manual yang dilakukan oleh seorang ahli pun masih ada kemungkinan membuat kesalahan serta memerlukan waktu yang relatif lama dalam melakukan pemeriksaan gambar tersebut [3].

Pemrosesan citra pada bidang kesehatan sudah umum dilakukan, salah satunya yaitu mendeteksi kanker payudara pada histopatologi yang telah menjadi topik pembahasan selama beberapa waktu. Hal ini tentunya menjadi tantangan, terutama bagi para peneliti untuk mendapatkan akurasi yang baik untuk deteksi IDC. Para ahli patologi dapat melakukan pemeriksaan klasifikasi secara otomatis dengan lebih konsisten dan produktif dengan bantuan komputer pada gambar histopatologi. Proses klasifikasi secara otomatis dilakukan dengan menganalisa pola tertentu yang terdapat pada citra bergantung dengan ada atau tidaknya sel yang menyebabkan kanker pada jaringan tubuh yang diperiksa [3]. Pada penelitian ini, deteksi citra histopatologi dilakukan dengan beberapa tahapan yaitu input data histopatologi, ekstraksi fitur, dan klasifikasi. Ekstraksi fitur merupakan tahapan paling penting pada proses klasifikasi.

Fitur merupakan ciri pada suatu citra yang dapat memberikan informasi tentang citra, tingkat akurasi pada klasifikasi bergantung pada ekstraksi fitur. Dalam penelitian tentang kanker payudara, *deep learning* dengan metode CNN (*Convolutional Neural Network*) merupakan metode yang paling banyak digunakan [4]. IDC merupakan jenis kanker payudara yang telah banyak dilakukan oleh beberapa peneliti dalam penerapan deteksi objek dengan CNN. Salah satu penelitian sebelumnya dengan *dataset* yang sama dari gambar histopatologi menggunakan ekstraksi fitur CNN dengan *pooling layer* baru yaitu *accept-reject pooling*.

Dari beberapa uraian yang telah dijelaskan, pada penelitian ini akan dilakukan sebuah ekstraksi dengan menggabungkan beberapa ekstraksi fitur di antaranya yaitu GLCM, RGB, HSV, dan Histogram. Berdasarkan gabungan dari beberapa fitur tersebut akan menghasilkan pengaruh terhadap nilai akurasi, sehingga membuat hasil yang didapatkan menjadi lebih akurat dibandingkan ketika dilakukan hanya dengan satu fitur. Dari empat ciri tersebut menghasilkan 15 ciri gabungan. Hasil terhadap ekstraksi dari fitur tersebut akan dilakukan klasifikasi menggunakan *classifier backpropagation* dan SVM yang bertujuan untuk mengetahui metode yang paling efektif dalam melakukan klasifikasi terhadap citra histopatologi.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah yang akan diambil diantaranya adalah sebagai berikut :

1. Bagaimana cara melakukan ekstraksi fitur dengan menggunakan GLCM, RGB, HSV, dan Histogram pada citra histopatologi?
2. Bagaimana hasil ekstraksi fitur dengan menggunakan GLCM, RGB, HSV, dan Histogram pada citra histopatologi?
3. Bagaimana model terbaik dan hasil dari klasifikasi kanker payudara pada citra histopatologi menggunakan *backpropagation* dan SVM?

1.3 Batasan Masalah

1. *Dataset* yang digunakan adalah *breast histopathology mages* yang diambil dari situs Kaggle.
2. *Dataset* histopatologi yang digunakan berjumlah 1080 citra, dengan 540 IDC negatif dan 540 IDC positif.
3. Dalam penelitian ini tidak dibahas proses pengambilan data dan *pre-processing*.

4. Hasil klasifikasi kanker payudara dibagi menjadi dua kategori yaitu IDC negatif dan IDC positif.

1.4 Tujuan Penelitian

1. Untuk memperoleh fitur citra histopatologi dengan menggunakan metode ekstraksi GLCM, RGB, HSV, dan Histogram.
2. Mengevaluasi unjuk kerja dari metode *backpropagation* dan SVM dalam mengklasifikasi fitur yang didapatkan.

1.5 Manfaat Penelitian

Dilakukannya ekstraksi ciri fitur menggunakan metode yang telah ditentukan, maka diharapkan dapat meningkatkan akurasi dalam mengklasifikasi citra histopatologi pada kanker payudara. Sehingga membantu para ahli untuk mendapatkan hasil deteksi yang lebih baik.

BAB 2

TINJAUAN PUSTAKA

2.1 Studi Literatur

Penelitian terkait ekstraksi citra histopatologi pada kanker payudara jenis IDC sebelumnya telah banyak dilakukan, diantaranya oleh Romano, dkk. Pada penelitian ini mereka membahas terkait peningkatan pendekatan *deep learning* untuk memprediksi IDC dari gambar histopatologi. Penelitian ini menggunakan *pre-processing* dan *augmentasi* data yang artinya *dataset* dapat diperbanyak hingga 10 kali dari data *original* untuk mencegah terjadinya *overfitting*. Menggunakan fitur CNN dengan arsitektur yang mengandung *convolutional layer*, *pooling layer* baru yaitu *accept-reject pooling*, *dropout*, dan *fully connected layer* dengan tujuan untuk mendapatkan performa yang lebih baik dan hasil yang signifikan. *Dataset* yang digunakan sebanyak 90.000 citra (64.583 untuk IDC negatif dan 25.417 untuk IDC positif). Hasil eksperimen pada penelitian ini mampu meraih nilai akurasi yaitu 85,41% [4].

Penelitian lain yang berfokus pada kanker payudara histopatologi dilakukan oleh Romero, dkk yang membahas terkait klasifikasi sel IDC dalam *slide* histopatologi dengan model yang dibuat dari arsitektur *insepsi* dengan metode CNN model yaitu *convolution*, *max pooling*, *dropout*, dan *fully connected layer*. Protokol pelatihan juga dilakukan dengan strategi pemilihan parameter yaitu *padding* dengan gambar *patch* berukuran 50x50 piksel. Selanjutnya yaitu normalisasi, yang bertujuan menurunkan variabilitas sehingga meningkatkan stabilitas pelatihan. Kemudian *validation* dan *batch size* digunakan untuk meningkatkan *rendering* mereka juga menerapkan filter *gaussian*. *Dataset* yang digunakan pada penelitian ini berjumlah 227.522, dengan data *train* sebanyak 94.543, validasi berjumlah 31.514 data, dan 151.465 untuk *dataset*. Hasil penelitian mendapatkan tingkat akurasi 89% [5].

Selain itu penelitian yang dilakukan oleh Fina Afifana mengusulkan menggunakan metode *deep learning* dengan ekstraksi fitur CNN. Metode yang digunakan untuk mengklasifikasi IDC positif dan negatif dari citra histopatologi menggunakan model VGG16 dan MobileNet. Arsitektur model VGG16 dan MobileNet memiliki performa yang cukup baik dalam klasifikasi citra. Kontribusi pada penelitian ini adalah optimalisasi *hyperparameter* pada model MobileNet dan VGG16 menggunakan *tunning hyperparameter* yang terlibat dalam kinerja metode untuk dapat mengurangi *loss* dan mencegah *overfitting*. Hasil penelitian menghasilkan model VGG16 memberikan hasil akurasi terbaik dengan akurasi 87,9% [6].

Dari beberapa studi literatur diketahui bahwa penggunaan CNN secara umum dilakukan sebagai ekstraksi fitur dengan berbagai jenis model. Pada penelitian ini bertujuan untuk meningkatkan akurasi pada diagnosis kanker payudara pada citra histopatologi. Upaya yang dilakukan dalam untuk peningkatan dengan melakukan penggabungan beberapa ciri yaitu GLCM, RGB, HSV, dan Histogram. Klasifikasi dilakukan dengan dua *classifier* yaitu *backpropagation* dan SVM.

2.2 Tinjauan Teori

Pemrosesan citra digital merupakan teknik pengolahan gambar atau citra menggunakan berbagai algoritma. Citra berupa foto yang tampak sedikit gelap dapat diolah menjadi lebih terang merupakan salah satu contoh proses yang dapat dilakukan melalui pemrosesan citra digital. Pemrosesan citra digital telah banyak diterapkan pada berbagai bidang, salah satunya digunakan untuk pencitraan histopatologi. Tidak hanya sebatas mengatur revolusi suatu gambar atau meningkatkan kecerahan foto, tetapi juga mampu menganalisis hal-hal yang hanya bisa dilihat oleh kamera namun tidak dapat dilihat secara kasat mata. Tujuan utama dari pemrosesan citra adalah untuk memperoleh gambar yang berkualitas tinggi [7].

2.2.1 Kanker Payudara

Kanker payudara adalah jenis kanker yang dimulai di payudara, kanker dimulai ketika sel mulai tumbuh di luar kendali dari sel-sel kelenjar dan salurannya, sehingga merusak organ atau saluran sekitar dan dapat melakukan penyebaran ke bagian tubuh yang lainnya. Sel kanker payudara biasanya membentuk tumor yang sering terlihat pada *x-ray*. Jaringan payudara tersebut terdiri dari kelenjar susu (kelenjar pembuat air susu), saluran kelenjar (saluran air susu), dan jaringan penunjang payudara [8].

Karsinoma payudara adalah jenis kanker payudara yang paling banyak ditemukan. Kanker ini berasal dari kulit atau jaringan lain yang melapisi atau menutupi organ dalam. Ada berbagai jenis karsinoma payudara yang dibedakan dan diberi nama berdasarkan lokasi dan cara penyebarannya. DCIS (*Ductal Carcinoma In Situ*) merupakan jenis kanker payudara yang dimulai di dalam saluran susu dan tidak menyebar melebihi aliran susu. Oleh karena itu, memiliki tingkat kesembuhan yang tinggi. IDC menyumbang hampir 70-80% dari semua kasus kanker payudara. Sama seperti DCIS, IDC (*Invasive Ductal Carcinoma*) juga dimulai pada saluran susu. Tapi IDC dapat menyebar ke kelenjar getah bening dan bagian tubuh lainnya serta memiliki banyak *subtype*, salah satunya tubular *carcinoma* yang cenderung *low-grade* dan

sangat responsif terhadap pengobatan. Selanjutnya yaitu ILC (*Invasive Lobular Carcinoma*) merupakan jenis kanker yang dimulai dari lobulus, yang dapat menyebar ke kelenjar betah bening dan bagian tubuh lainnya [9]. Pada penelitian ini menggunakan data histopatologi pada jenis kanker IDC.

Histopatologi adalah salah satu cabang dari ilmu patologi yang mempelajari perubahan dari suatu jaringan. Pemeriksaan histopatologi ditujukan untuk mengetahui perubahan yang terjadi pada jaringan baik yang disebabkan mikroorganisme patogen, senyawa, maupun agen kimia [8].

2.2.2 Citra Digital

Citra digital merupakan citra yang disimpan dalam format digital dan dapat diolah oleh komputer. Ditinjau dari sudut pandang matematis, citra merupakan fungsi kontinu dari intensitas cahaya pada bidang dua dimensi. Sumber cahaya menerangi objek, kemudian objek memantulkan kembali sebagian dari berkas cahaya tersebut. Pantulan cahaya tersebut ditangkap oleh alat-alat optik, misalnya mata manusia, kamera, *scanner*, dan sebagainya, sehingga bayangan objek yang disebut citra tersebut terekam [10].

2.2.3 Ekstraksi Fitur

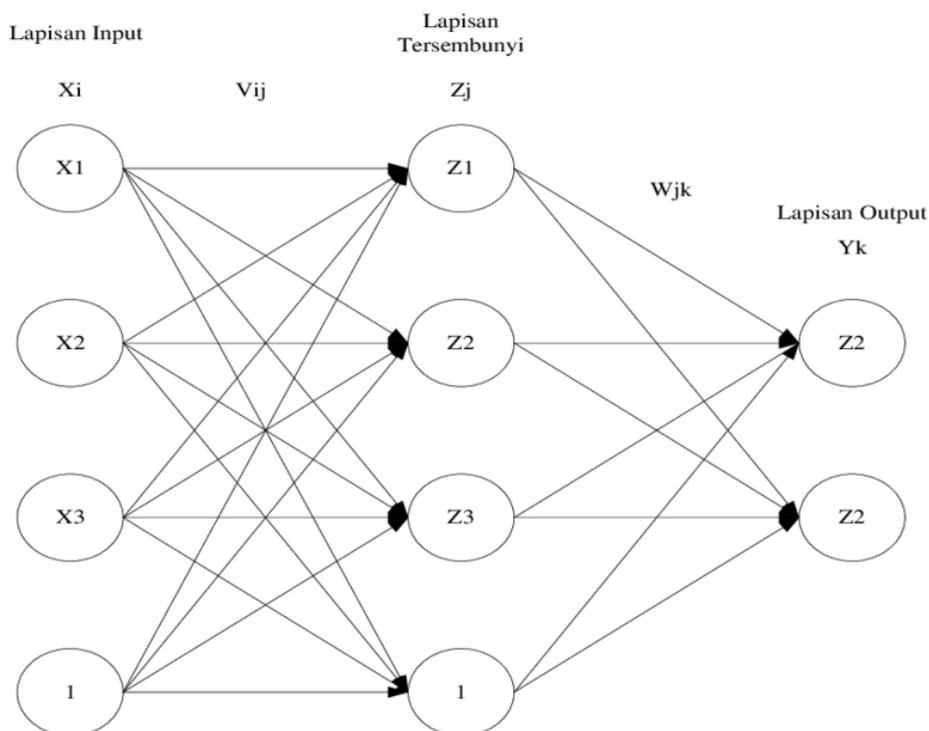
Ekstraksi fitur merupakan bagian dari teknik pengenalan pola (*pattern recognition*) yang bertujuan untuk mengambil atau mengesktraksi nilai-nilai unik dari suatu obyek yang membedakan dengan obyek yang lainnya. Analisis tekstur biasanya dimanfaatkan sebagai proses antara untuk melakukan klasifikasi dan interpretasi citra. Suatu proses klasifikasi citra berbasis analisis tekstur dan warna [8]. Adapun ekstraksi fitur yang digunakan pada penelitian ini adalah GLCM (*Gray-Level Co-Occurrence Matrix*), RGB (*Red Green Blue*), HSV (*Hue Saturation Value*), dan Histogram.

GLCM merupakan salah satu metode dalam menganalisa nilai suatu tekstur. Teknik ini digunakan untuk menganalisa antar piksel bersebelahan yang dapat ditentukan berdasarkan sudut yang digunakan, dengan ekstraksi ciri yang digunakan adalah *contrast*, *correlation*, *energy* dan *homogeneity*. RGB merupakan citra yang tersusun oleh tiga kanal warna yaitu kanal merah, kanal hijau, dan kanal biru. Citra RGB memiliki nilai intensitas yang merupakan kombinasi dari nilai R, G, B. HSV merupakan karakteristik yang membedakan antara satu warna dengan warna yang lain berupa *hue*, *saturation*, dan *value*. Histogram sebagai ekstraksi ciri yang didasarkan pada perbedaan sebaran atau distribusi piksel berdasarkan intensitas *gray*

level (derajat keabuan) yang dimiliki tiap-tiap piksel di masing-masing gambar. Ekstraksi citra yang digunakan berupa *mean*, *variance*, *standard deviation*, *skewness*, dan *entropy*.

2.2.4 Backpropagation

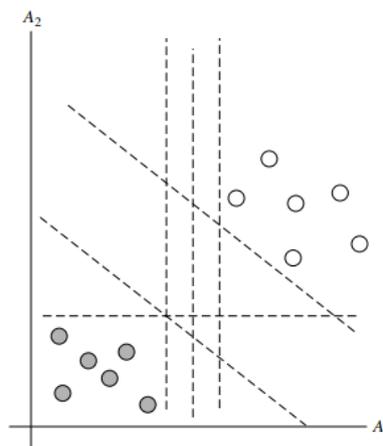
Backpropagation merupakan salah satu jenis jaringan syaraf tiruan yang terdiri dari beberapa *layer* (*multilayer*) dengan metode pembelajaran terawasi (*supervised learning*) [11]. Setiap unit jaringan *backpropagation* yang berada di lapisan input terhubung dengan setiap unit yang ada di lapisan tersembunyi (*hidden layer*) terhubung dengan setiap unit yang ada di lapisan *output*. Ketika jaringan diberikan pola masukan sebagai pola pelatihan, maka pola tersebut menuju unit-unit lapisan tersembunyi untuk selanjutnya keluaran akan memberikan respon sebagai keluaran jaringan syaraf tiruan. Saat hasil keluaran tidak sesuai dengan yang diharapkan, maka keluaran akan dijalankan mundur (*backward*) pada lapisan tersembunyi kemudian dari lapisan tersembunyi menuju lapisan masukan. Setiap unit di dalam *layer* input pada jaringan *backpropagation* selalu terhubung dengan setiap unit yang berada pada *layer* tersembunyi, demikian juga setiap unit pada *layer* tersembunyi selalu dari banyak lapisan [8]. Arsitektur jaringan *backpropagation* dapat dilihat pada Gambar 2.1.



Gambar 2.1 Arsitektur jaringan *Backpropagation* [11]

2.2.5 SVM (*Support Vector Machine*)

Support vector machine merupakan sebuah metode untuk klasifikasi data linier dan *non* linier, yang dapat melakukan pengenalan pola dan regresi berdasarkan teori pembelajaran statistik dan prinsip minimalisasi risiko struktural. SVM bertujuan untuk mencari fungsi garis pemisah (*hyperplane*) satu set contoh positif dan satu set contoh negatif dengan margin maksimum. Margin adalah jarak antara garis pemisah dengan anggota-anggota terdekat dari kedua kelas [12]. Gambar 2.2 menunjukkan bidang pemisah linier dalam SVM.



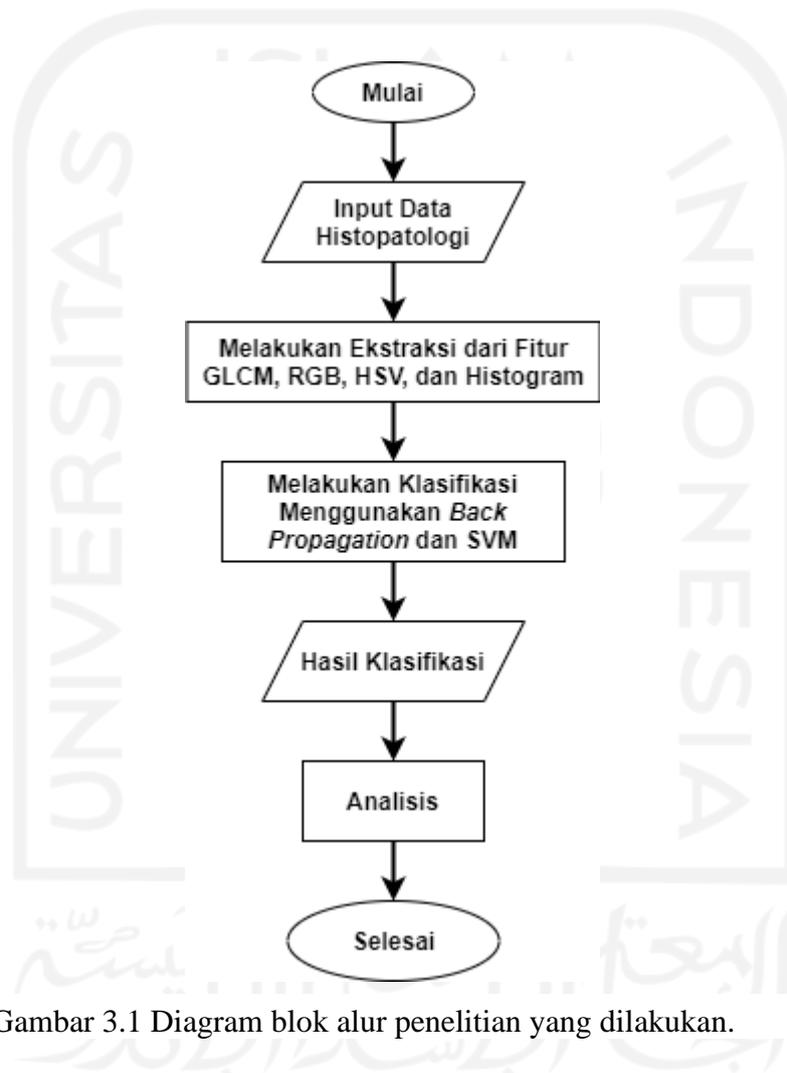
Gambar 2.2 Ilustrasi *Support Vector Machine* [12]

BAB 3

METODOLOGI

3.1 Alur penelitian

Berikut merupakan diagram blok alur penelitian yang dilakukan, ditunjukkan pada Gambar 3. 1.

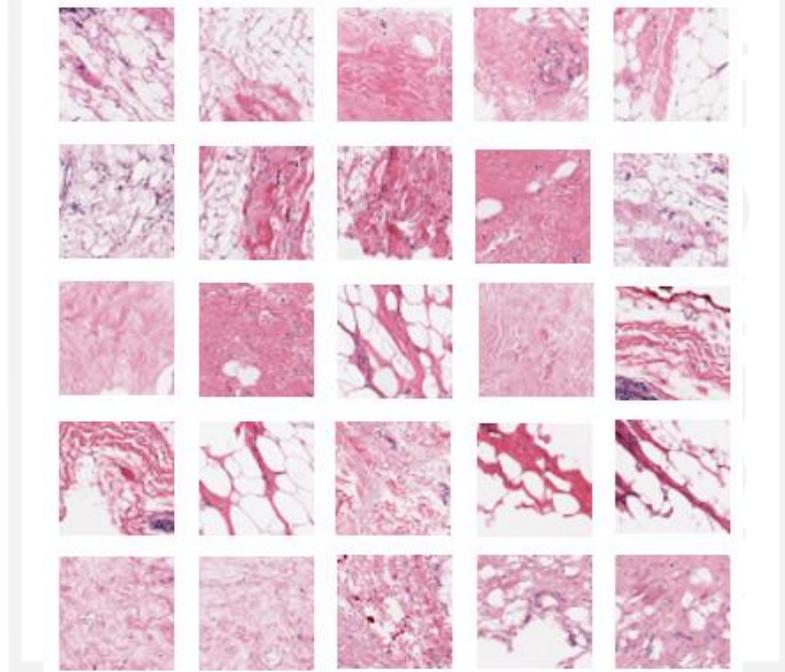


Gambar 3.1 Diagram blok alur penelitian yang dilakukan.

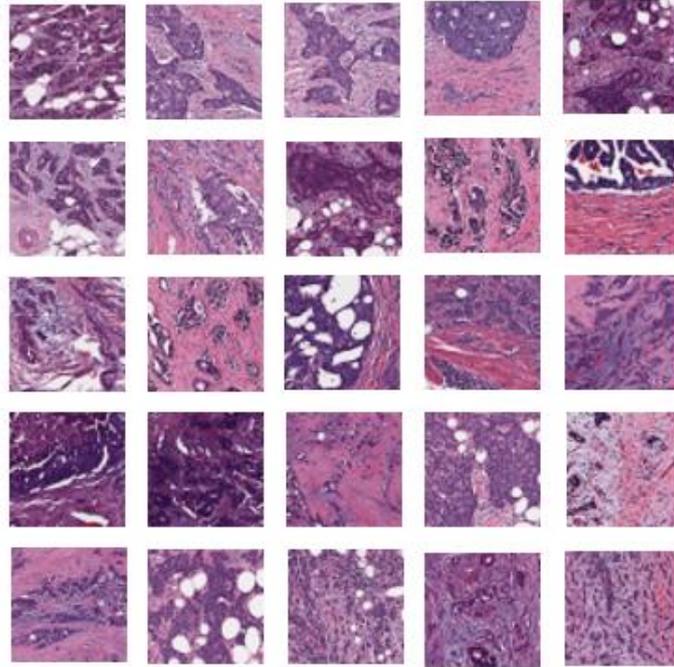
Proses perbaikan citra (*pre-processing*) bersifat opsional, artinya jika tanpa melalui proses perbaikan kualitas citra, citra hasil akuisisi sudah dapat tersegmentasi dengan baik, maka perbaikan kualitas citra boleh tidak dilakukan. Begitu pula dengan segmentasi citra juga bersifat subjektif. Pada penelitian ini, *dataset* berukuran 50x50 (RGB) piksel dan bentuk citra yang sulit untuk dilakukan pemisahan obyek yang dikehendaki dan tidak dikehendaki karena bentuk penyebaran sel yang tidak beraturan, maka diputuskan tidak dilakukan *pre-processing* dan segmentasi citra.

3.1.1 Dataset (Histopatologi)

Penelitian ini menggunakan data sekunder berupa data *breast histopathology images*, didapat pada situs Kaggle yang tersedia untuk umum [13]. *Dataset* berasal dari 162 gambar payudara dengan jumlah citra yang dihasilkan adalah 227.524 gambar IDC dengan masing-masing kelas yaitu 0 untuk kelas IDC negatif berjumlah 198.738 gambar dan 1 untuk kelas IDC positif berjumlah 78.786 gambar, yang terdiri dari 279 folder ID pasien dalam format PNG dan masing-masing berukuran 50x50 piksel. Dari 279 folder ID tersebut, gambar yang diambil dari setiap folder (\pm) 3-4 gambar dengan jumlah 1080 citra IDC (540 IDC negatif dan 540 IDC positif). Gambar 3.2 dan Gambar 3.3 merupakan contoh citra histopatologi yang mengandung IDC negatif dan IDC positif.



Gambar 3.2 *Dataset* yang mengandung IDC negatif [13]



Gambar 3.3 *Dataset* yang mengandung IDC positif [13]

3.1.2 Ekstraksi Fitur

Ekstraksi fitur merupakan bagian dari teknik pengenalan pola (*pattern recognition*) yang bertujuan untuk mengambil atau mengesktraksi nilai-nilai unik dari suatu obyek yang membedakan dengan obyek yang lainnya. Analisis tekstur biasanya dimanfaatkan sebagai proses antara untuk melakukan klasifikasi dan iterpretasi citra. Suatu proses klasifikasi citra berbasis analisis tekstur dan warna [8]. Adapun ekstraksi fitur yang digunakan pada penelitian ini adalah GLCM, RGB, HSV, dan Histogram.

a. GLCM (*Gray-Level Co-Occurrence Matrix*)

GLCM merupakan salah satu metode yang dapat menganalisa nilai suatu tekstur. Teknik ini digunakan untuk menganalisa antar piksel bersebelahan yang dapat ditentukan berdasarkan sudut yang digunakan [8]. GLCM juga digunakan untuk mengetahui seberapa seringnya kombinasi dari nilai kecerahan piksel yang berbeda posisinya yang terjadi dalam suatu citra.

Ciri fitur GLCM yang digunakan diantaranya adalah sebagai berikut:

- *Contrast*

Contrast merupakan fitur yang digunakan untuk mengukur kekuatan perbedaan intensitas pada citra. Ukuran variasi antara derajat keabuan pada suatu daerah citra dapat mempengaruhi nilai *contrast*. Nilai *contrast* akan membesar apabila

variasi intensitas dalam citra tinggi dan akan mengecil apabila variasi rendah. Persamaan yang digunakan yaitu :

$$Contrast = \sum_i \sum_j (i - j)^2 p(i, j) \quad (3.1)$$

Dengan $p(i, j)$ menunjukkan piksel baris ke- i kolom ke- j .

- *Correlation*

Correlation suatu citra menunjukkan ukuran ketergantungan linier derajat keabuan pada citra sehingga dapat menunjukkan adanya struktur linier dalam citra. Persamaan yang digunakan yaitu :

$$Correlation = \sum_i \sum_j \frac{\{(ij)p(i, j)\} - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (3.2)$$

dengan

$p(i, j)$ = Piksel baris ke- i kolom ke- j

μ_x = Nilai rata-rata elemen kolom

μ_y = Nilai rata-rata elemen baris

σ_x = Standar deviasi elemen kolom

σ_y = Standar deviasi elemen baris

- Energi

Energi merupakan fitur yang digunakan untuk mengukur konsentrasi pasangan intensitas pada matriks koakurensi. Nilai energi akan semakin besar apabila piksel yang memenuhi syarat matriks intensitas koakurensi berkumpul atau terkonsentrasi pada beberapa koordinat dan mengecil apabila letaknya menyebar.

$$Energi = \sum_i \sum_j p^2(i, j) \quad (3.3)$$

Dengan $p(i, j)$ menunjukkan piksel baris ke- i kolom ke- j .

- *Homogeneity*

Homogeneity merupakan fitur yang digunakan untuk mengukur keseragaman dari suatu citra atau fitur yang berkebalikan dengan *contrast*. Nilai *homogeneity* akan semakin besar apabila variasi identitas pada citra rendah dan akan semakin kecil apabila variasi identitas pada citra tinggi. Persamaan yang digunakan yaitu :

$$Homogeneity = \sum_i \sum_j \frac{p(i, j)}{1 + |i - j|} \quad (3.4)$$

Dengan $p(i,j)$ menunjukkan piksel baris ke- i kolom ke- j .

b. *Red Green Blue* (RGB)

Pada model ini untuk merepresentasikan gambar menggunakan tiga buah komponen warna yaitu merah, hijau, dan biru. Jika dilakukan pencampuran ketiga unsur warna tersebut akan menghasilkan warna lain. Pengaturan format file grafik menyimpan warna RGB sebagai 24-bit dengan komponen r , g , dan b masing-masing 8 bit. Warna untuk tiap piksel ditentukan dengan kombinasi intensitas merah, hijau, dan biru [7]. Berikut adalah persamaan yang digunakan:

$$r = \frac{R}{R+G+B} \quad (3.5)$$

$$g = \frac{G}{R+G+B} \quad (3.6)$$

$$b = \frac{B}{R+G+B} \quad (3.7)$$

Dengan r = normalisasi *red*, g = normalisasi *green*, b = normalisasi *blue*.

R = Nilai komponen *red*

G = Nilai komponen *green*

B = Nilai komponen *blue*

c. *Hue Saturation Value* (HSV)

HSV adalah salah satu model yang digunakan untuk analisis warna pada pengolahan citra selain RGB. *Hue* untuk menyatakan warna yang sebenarnya, seperti merah, hijau, dan violet yang digunakan untuk menentukan kemerahan (*redness*), kehijauan (*greenness*), dan lainnya. *Saturation* merupakan tingkat kemurnian maupun kekuatan warna atau seberapa banyak cahaya putih yang tercampur dengan *hue*. Setiap warna murni bersaturasi 100% tanpa mengandung cahaya putih, sedangkan warna murni yang bercampur dengan warna putih memiliki saturasi antara 0-100%. *Value* atau kadang disebut *brightness* menyatakan kecerahan warna dan memiliki nilai yang berkisar 0-100%. Jika *value* bernilai 0 maka berwarna hitam, semakin besar nilai *value* maka akan semakin cerah warnanya dan akan muncul variasi-variasi baru [14]. Untuk mendapatkan nilai H, S, V berdasarkan R, G, dan B. Adapun persamaan yang digunakan yaitu:

$$H = \tan\left(\frac{3(G - B)}{(R - B) + (R - B)}\right) \quad (3.8)$$

$$S = 1 - \frac{\min(R,G,B)}{V} \quad (3.9)$$

$$V = \frac{R+G+B}{3} \quad (3.10)$$

Dengan $H = Hue$, $S = Saturation$, dan $V = Value$.

R = Nilai komponen *red*

G = Nilai komponen *green*

B = Nilai komponen *blue*

d. Histogram

Ekstraksi fitur Histogram termasuk ke dalam kelompok metode statistik orde 1 karena pengambilan ciri yang didasarkan pada karakteristik histogram citra. Pada umumnya metode ini digunakan untuk mengekstraksi ciri tekstur dari suatu citra. Penggunaan histogram sebagai ekstraksi ciri didasarkan pada perbedaan sebaran atau distribusi piksel berdasarkan intensitas *gray level* (derajat keabuan) yang dimiliki tiap-tiap piksel di masing-masing gambar [15].

Berikut ini merupakan fitur Histogram yang digunakan:

- *Mean*

Mean menunjukkan ukuran rata-rata intensitas suatu citra.

$$\mu = \sum_i \sum_j (i, j) p(i, j) \quad (3.11)$$

Dengan

$p(i, j)$ = Piksel baris ke- i kolom ke- j

μ = Nilai rata-rata citra histogram (*mean*)

- *Variance*

Variance merupakan tingkat keragaman suatu citra pada data histogram. Citra akan memiliki variansi yang kecil apabila memiliki transisi derajat keabuan kecil.

$$\sigma^2 = \sum_i \sum_j (i - \mu)^2 p(i, j) \quad (3.12)$$

Dengan

$p(i, j)$ = Piksel baris ke- i kolom ke- j

σ^2 = Tingkat keragaman pada citra histogram (*variance*)

μ = Nilai rata-rata citra histogram (*mean*)

- *Satandard Deviation*

Standard deviation atau dapat disebut simpangan baku merupakan fitur yang menunjukkan tingkat keragaman yang berbeda. *Standard deviation* merupakan ukuran rata-rata kontras pada citra dan akar dari nilai *variance*.

$$\sigma = \sqrt{\sum_i \sum_j (i - \mu)^2 p(i, j)} \quad (3.13)$$

Dengan

$p(i, j)$ = Piksel baris ke- i kolom ke- j

σ = Standar deviasi citra histogram

μ = Nilai rata-rata citra histogram (*mean*)

- *Skewness*

Skewness menunjukkan tingkat kemiringan relatif kurva pada citra histogram dan dapat didefinisikan sebagai berikut:

$$Skewness = \frac{1}{\sigma^3} \sum_i \sum_j (i - \mu)^3 p(i, j) \quad (3.14)$$

Dengan

$p(i, j)$ = Piksel baris ke- i kolom ke- j

μ = Nilai rata-rata citra histogram (*mean*)

σ = Standar deviasi citra histogram

- *Entropy*

Entropy merupakan ukuran ketidakteraturan atau keacakan distribusi intensitas pada histogram suatu citra dapat didefinisikan sebagai berikut:

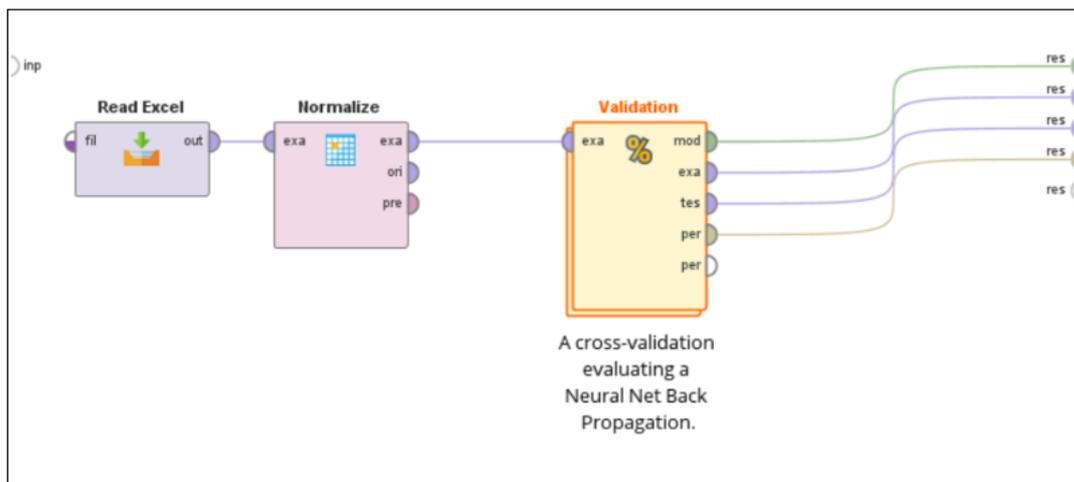
$$Entropy = -\sum_i \sum_j p(i, j) \log_2 p(i, j) \quad (3.15)$$

Dengan $p(i, j)$ menunjukkan piksel baris ke- i kolom ke- j .

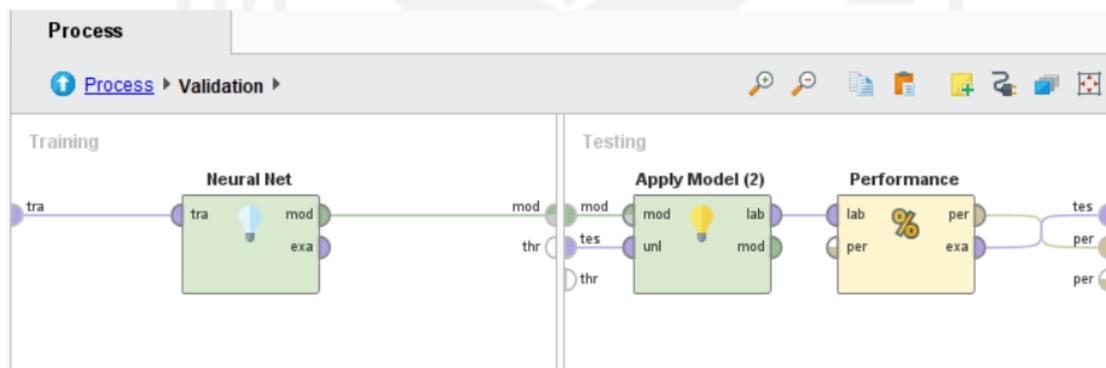
3.1.3 Proses Klasifikasi

Klasifikasi merupakan proses untuk menemukan model atau fungsi yang dapat menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan suatu kelas. Klasifikasi ini juga merupakan proses awal dari pengelompokan data. Hasil ekstraksi fitur yang telah didapatkan pada penelitian ini, digunakan untuk melakukan klasifikasi dimana kelas pada klasifikasi dibedakan menjadi dua yaitu IDC negatif dan IDC positif dengan menggunakan metode klasifikasi *backpropagation* dan SVM dengan

bantuan *tool* pada RapidMiner. Sebelum tahap klasifikasi, proses yang harus dilakukan adalah melakukan normalisasi terhadap *dataset* histopatologi, agar fitur-fitur memiliki jangkauan yang sama. Metode yang digunakan dalam proses normalisasi adalah normalisasi linier. Gambar 3.4 merupakan proses klasifikasi yang dilakukan menggunakan RapidMiner dan Gambar 3.5 merupakan proses *validation* pada proses klasifikasi pada RapidMiner.



Gambar 3.4 Proses klasifikasi menggunakan RapidMiner



Gambar 3.5 Proses *validation* pada klasifikasi menggunakan RapidMiner

3.2 Hasil dan Unjuk Kerja

Hasil yang didapat yaitu berupa nilai akurasi dari klasifikasi dengan metode *neural net backpropagation* menggunakan *tool machine learning* pada RapidMiner. Data pelatihan dan data pengujian dipilih secara acak dengan parameter nilai *fold* 5, 7, dan 10. Untuk mengetahui reliabilitas dari *classifier* terhadap hasil dari ekstraksi fitur yang telah dilakukan, dapat digunakan *confussion matrix* yang berguna dalam menentukan nilai dari akurasi, sensitivitas,

dan spesifisitas. Gambar 3.6 merupakan salah satu contoh hasil klasifikasi yang dilakukan, berupa nilai akurasi. Tabel 3.1 merupakan tabel *confusion matrix* yang digunakan.

accuracy: 96.76% +/- 1.53% (micro average: 96.76%)

	true 0	true 1	class precision
pred. 0	527	22	95.99%
pred. 1	13	518	97.55%
class recall	97.59%	95.93%	

Gambar 3.6 Hasil akurasi yang didapatkan dari RapidMiner

Tabel 3.1 *Confusion matrix* [12]

	<i>Actual True</i>	<i>Actual False</i>
<i>Predicted True</i>	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
<i>Predicted False</i>	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

Keterangan:

TP : *True Positive* (jumlah citra sebenarnya IDC negatif yang berhasil terdeteksi sebagai IDC negatif).

FP : *False Positive* (jumlah citra sebenarnya IDC negatif yang terdeteksi sebagai IDC positif).

TN : *True Negative* (jumlah citra sebenarnya IDC positif yang berhasil terdeteksi IDC positif).

FN : *False Negative* (jumlah citra sebenarnya IDC positif yang terdeteksi IDC negatif).

Hasil unjuk kerja pada sistem dihitung berdasarkan :

1. *True Positive, False Positive, True Negative, dan False Negative.*
2. Akurasi

Akurasi menunjukkan seberapa tepat dan akurat model dapat mengklasifikasikan hasil pengukuran dengan nilai yang benar atau diterima dari kuantitas besaran yang diukur [12]. Nilai akurasi dapat diperoleh berdasarkan persamaan (3.16).

$$\text{Akurasi} = \frac{\text{true positive} + \text{true negative}}{\text{true positive} + \text{true negative} + \text{false positive} + \text{false negative}} \times 100 \% \quad (3.16)$$

3. Sensitivitas

Sensitivitas atau *recall* merupakan proporsi orang yang benar-benar sakit dalam populasi yang juga diidentifikasi sebagai orang sakit oleh tes *screening*. Sensitivitas dapat dihitung dengan cara membagi citra yang benar yang bernilai positif (TP) dengan

hasil penjumlahan dari citra benar yang bernilai positif (TP) dan citra yang salah bernilai negatif (FN) [12]. Nilai sensitivitas diperoleh berdasarkan persamaan (3.17).

$$\text{Sensitivitas} = \frac{\text{true positive}}{\text{true positive} + \text{false negative}} \times 100 \% \quad (3.17)$$

4. Spesifisitas

Spesifisitas merupakan proporsi orang yang benar-benar tidak sakit dan tidak sakit pula pada saat diidentifikasi dengan tes *screening*. Ini adalah ukuran dari kemungkinan benar mengidentifikasi orang tidak sakit dengan tes *screening* [12]. Nilai spesifisitas dapat diperoleh berdasarkan persamaan (3.18)

$$\text{Spesifisitas} = \frac{\text{true negative}}{\text{true negative} + \text{false positive}} \times 100 \% \quad (3.18)$$

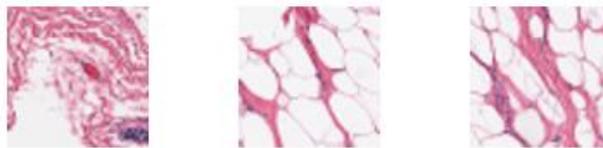


BAB 4

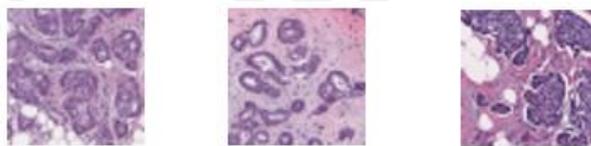
HASIL DAN PEMBAHASAN

4.1 Hasil Ekstraksi

Pada tahapan ini merupakan hasil ekstraksi ciri fitur yang telah dilakukan di Matlab sebelumnya. Penelitian ini menggunakan empat fitur ekstraksi dengan masing-masing ciri. Ekstraksi fitur yang digunakan yaitu, GLCM yang terdiri dari 4 ciri (*contrast*, *correlation*, *energy*, dan *homogeneity*), RGB yang terdiri dari 3 ciri (*red*, *green*, dan *blue*), HSV yang terdiri dari 3 ciri (*hue*, *saturation*, dan *value*), dan Histogram yang terdiri dari 5 ciri (*mean*, *variance*, *standard deviation*, *skewness*, dan *entropy*). Gambar 4.1 merupakan citra pada kelas 0 dan Gambar 4.2 merupakan citra kelas 1 yang digunakan untuk sampel pada masing-masing ekstraksi fitur. Pada Tabel 4.1, 4.2, 4.3, dan 4.4 merupakan contoh hasil ekstraksi dengan masing-masing fitur yang dilakukan, nomor citra 1-3 merupakan contoh citra kelas 0 dan citra no 4-6 merupakan contoh citra kelas 1.



Gambar 4.1 Contoh citra pada kelas 0 [13]



Gambar 4.2 Contoh citra pada kelas 1 [13]

Tabel 4.1 Hasil ekstraksi dari GLCM

No Citra	<i>Contrast</i>	<i>Correlation</i>	<i>Energy</i>	<i>Homogeneity</i>
1	0,945814	0,791031	0,098319	0,73544
2	0,73353	0,75346	0,131014	0,778987
3	0,882436	0,697462	0,107111	0,745002
4	1,177364	0,408467	0,126847	0,701057
5	1,380427	0,391702	0,132794	0,679181
6	1,583222	0,227472	0,088186	0,636498

Berdasarkan tabel 4.1 diatas merupakan hasil nilai dari ekstraksi fitur yang didapatkan pada fitur GLCM dengan menggunakan Matlab. Hasil ekstraksi tersebut akan diklasifikasi menggunakan metode *backpropagation* dan SVM di RapidMiner untuk memperoleh hasil akurasi yang maksimal. Dari hasil percobaan ekstraksi fitur GLCM diperoleh beberapa variasi nilai, dapat dilihat bahwa nilai ciri *contrast* dan *homogeneity* relatif lebih tinggi jika dibandingkan dengan ciri *correlation* maupun *energy*. Nilai *homogeneity* ini dipengaruhi oleh *dataset* pada gambar histopatologi yang digunakan. Semakin kasar bentuk teksturnya maka akan semakin naik turun atau fluktuatif, sehingga dapat mengurangi nilai *homogeneity*.

Dapat diketahui hasil nilai yang didapatkan pada fitur GLCM, jika nilai dari *contrast* kurang dari satu maka citra tersebut dikategorikan ke dalam kelas IDC negatif. Sedangkan jika nilai *contrast* melebihi nilai satu maka citra akan dikategorikan ke dalam kelas IDC positif.

Tabel 4.2 Hasil ekstraksi dari RGB

No Citra	<i>Red</i>	<i>Green</i>	<i>Blue</i>
1	235,1568	211,3772	228,8872
2	228,0308	190,9264	216,262
3	224,7616	208,0456	220,734
4	139,706	99,5916	150,4156
5	139,684	98,8364	148,8952
6	143,2368	89,9032	146,4324

Berdasarkan tabel 4.2 diatas merupakan hasil nilai dari ekstraksi fitur yang didapatkan pada fitur RGB dengan menggunakan Matlab. Hasil ekstraksi tersebut akan diklasifikasi menggunakan metode *backpropagation* dan SVM di RapidMiner untuk memperoleh hasil akurasi yang maksimal. Nilai standar fitur RGB yaitu 0-255. Semakin hitam citra yang digunakan maka nilai ekstraksi fitur RGB yang dihasilkan akan mendekati 0, sedangkan semakin mendekati putih citra yang digunakan maka nilai ekstraksi fitur RGB semakin membesar mendekati 255.

Tabel 4.3 Hasil ekstraksi dari HSV

No Citra	Hue	Saturation	Value
1	0,777543	0,114177	0,924784
2	0,823602	0,177605	0,896218
3	0,712666	0,094564	0,885274
4	0,878240911	0,298800159	0,715363137
5	0,878456431	0,422947391	0,588065882
6	0,855774764	0,427223039	0,52340549

Berdasarkan tabel 4.3 diatas merupakan hasil nilai dari ekstraksi fitur yang didapatkan pada fitur HSV dengan menggunakan Matlab. Hasil ekstraksi tersebut akan diklasifikasi menggunakan metode *backpropagation* dan SVM di RapidMiner untuk memperoleh hasil akurasi yang maksimal. Semakin hitam citra yang digunakan maka nilai ekstraksi fitur HSV yang dihasilkan akan mendekati 0, sedangkan citra yang berwarna cerah maka akan semakin besar pula nilai *value* yang dihasilkan.

Diketahui perbedaan hasil nilai ekstraksi fitur HSV kelas IDC negatif dan IDC positif tidak terlalu signifikan, oleh karena itu hasilnya hanya bisa diklasifikasikan dengan RapidMiner.

Tabel 4.4 Hasil ekstraksi dari Histogram

No Citra	Mean	Variance	Standard Deviation	Skewness	Entropy
1	220,4668	1224,976098	34,99965854	1,235473967	6,024770159
2	204,906	1541,607564	39,26331066	0,729955481	6,64310338
3	214,4844	1638,684157	40,48066398	1,493444146	5,491263448
4	136,3388	917,6544146	30,29281127	0,197493805	6,871254292
5	135,9768	968,4970618	31,12068543	0,439037854	6,881375375
6	126,9544	771,5739206	27,77721945	0,268194608	6,73934879

Berdasarkan tabel 4.4 diatas merupakan hasil nilai dari ekstraksi fitur yang didapatkan pada fitur Histogram dengan menggunakan Matlab. Hasil ekstraksi fitur Histogram kemudian dijadikan parameter dalam mengklasifikasikan citra histopatologi ke dalam dua kelas IDC, yaitu IDC positif dan IDC negatif menggunakan metode *backpropagation* dan SVM di RapidMiner untuk memperoleh hasil akurasi yang maksimal.

Dari tabel 4.1, 4.2, 4.3, dan 4.4 merupakan beberapa hasil dari ekstraksi fitur yang dilakukan menggunakan Matlab. Untuk hasil keseluruhan dari ekstraksi GLCM, RGB, HSV, dan Histogram dapat dilihat pada lembar lampiran 1, 2, 3, dan 4.

4.2 Hasil Klasifikasi

Proses klasifikasi pada penelitian ini menggunakan RapidMiner dari hasil ekstraksi ciri fitur yang telah dilakukan sebelumnya. Dari keempat ekstraksi fitur GLCM, RGB, HSV, dan Histogram kemudian diklasifikasi menggunakan metode *backpropagation* dan SVM dengan tujuan untuk menentukan dua kelas IDC positif dan IDC negatif. Setelah itu, percobaan dilakukan dengan membagi *dataset* menjadi dua bagian yaitu, *dataset* pelatihan dan *dataset* pengujian. Proses yang dilakukan untuk memisahkan antara *dataset* pelatihan dan pengujian dapat dilakukan menggunakan *K-Fold Cross Validation*. Tujuan dari metode ini adalah untuk memperoleh hasil akurasi yang maksimal dengan melakukan prediksi model dan memperkirakan seberapa akurat sebuah model ketika dijalankan. Nilai *fold* yang digunakan pada *dataset* pelatihan dan pengujian adalah 5, 7, dan 10 dengan masing-masing presentase 80%-20%, 85,71%-14,29%, dan 90%-10%. Pada Tabel 4.5, 4.6, 4.7, 4.8, dan 4.9 merupakan

contoh hasil dari klasifikasi masing-masing fitur dengan menggunakan dua *classifier backpropagation* dan SVM.

Tabel 4.5 Hasil dari klasifikasi fitur GLCM

Metode	Number Of Fold	TP	TN	FN	FP	Sensitivitas (%)	Spesifisitas (%)	Akurasi (%)
<i>Back-propagation</i>	5	414	455	85	126	82,96	78,31	80,46
	7	415	440	100	125	80,58	77,87	79,17
	10	422	452	88	118	82,74	79,29	80,92
SVM	5	427	445	95	113	81,8	79,74	80,75
	7	427	443	97	113	81,48	79,67	80,56
	10	427	442	98	113	81,33	79,63	80,46

Berdasarkan tabel 4.5 hasil pengujian klasifikasi pada fitur GLCM dengan *fold* yang berbeda yaitu 5, 7, dan 10. Tingkat sensitivitas tertinggi terdapat pada pengujian dengan metode *backpropagation* untuk *fold* 5 dengan nilai 82,96%. Dilanjutkan dengan tingkat spesifisitas tertinggi terdapat pada pengujian dengan metode SVM untuk *fold* 5 dengan nilai 79,74%. Sedangkan untuk tingkat akurasi tertinggi diperoleh dengan metode *backpropagation* untuk nilai *fold* 10 dengan nilai 80,92%. Nilai *homogeneity* pada fitur GLCM ini dapat mempengaruhi jumlah TP (citra normal dideteksi citra normal) yang didapatkan, semakin besar nilai *homogeneity* dari hasil ekstraksi maka semakin besar pula jumlah TP yang dihasilkan. Hal ini berbanding terbalik dengan nilai *energy*, semakin besar nilai *energy* dari hasil ekstraksi maka akan mengurangi jumlah TP yang dihasilkan.

Dengan *classifier backpropagation*, diketahui bahwa fitur GLCM mampu mendeteksi IDC negatif dideteksi IDC negatif (TP) yaitu 422 data dari 540 citra IDC negatif. Sedangkan untuk data yang terdeteksi sebagai IDC positif terdeteksi IDC positif (TN) 452 data dari 540 data IDC positif. Dengan nilai sensitivitas 82,74% kesalahan untuk kasus sebagai terdiagnosa IDC positif terdeteksi IDC negatif (FN) mencapai 17,26%.

Tabel 4.6 Hasil dari klasifikasi fitur RGB

Metode	Number Of Fold	TP	TN	FN	FP	Sensitivitas (%)	Spesifisitas (%)	Akurasi (%)
<i>Back-propagation</i>	5	521	517	23	19	95,77	96,45	96,11
	7	522	517	23	18	95,77	96,63	96,2
	10	520	517	23	20	95,76	96,27	96,01
SVM	5	519	512	28	21	94,88	96,06	95,46
	7	520	513	27	20	95,06	96,24	95,64
	10	520	513	27	20	95,06	96,24	95,64

Berdasarkan tabel 4.6 hasil pengujian klasifikasi pada fitur RGB dengan *fold* yang berbeda yaitu 5, 7, dan 10. Tingkat sensitivitas tertinggi terdapat pada pengujian dengan metode *backpropagation* untuk *fold* 5 dan 7 dengan nilai 95,57%. Dilanjutkan dengan tingkat spesifisitas tertinggi terdapat pada pengujian dengan metode *backpropagation* untuk *fold* 7 dengan nilai 96,63%. Sedangkan untuk tingkat akurasi tertinggi diperoleh dengan metode *backpropagation* untuk nilai *fold* 5 dengan nilai 96,11%. Akurasi akan semakin besar apabila nilai ciri R, G, B mendekati nilai 255.

Dengan *classifier backpropagation*, diketahui bahwa fitur RGB mampu mendeteksi IDC negatif dideteksi IDC negatif (TP) yaitu 521 data dari 540 citra IDC negatif. Sedangkan untuk data yang terdeteksi sebagai IDC positif terdeteksi IDC positif (TN) 517 data dari 540 data IDC positif. Dengan nilai sensitivitas 95,77% kesalahan untuk kasus sebagai terdiagnosa IDC positif terdeteksi IDC negatif (FN) mencapai 4,23%.

Tabel 4.7 Hasil dari klasifikasi fitur HSV

Metode	Number Of Fold	TP	TN	FN	FP	Sensitivitas (%)	Spesifisitas (%)	Akurasi (%)
<i>Back-propagation</i>	5	527	516	24	13	95,64	97,54	96,57
	7	523	516	24	17	95,61	96,81	96,2
	10	523	515	25	17	95,43	96,8	96,11
SVM	5	527	516	24	13	95,64	97,54	96,57
	7	525	515	25	15	95,45	97,16	96,3
	10	525	515	25	15	95,45	97,16	96,3

Berdasarkan tabel 4.7 hasil pengujian klasifikasi pada fitur HSV dengan *fold* yang berbeda yaitu 5, 7, dan 10. Tingkat sensitivitas, spesifisitas, dan akurasi tertinggi terdapat pada pengujian dengan metode keduanya yaitu, *backpropagation* dan SVM untuk *fold* 5 dengan nilai 96,64%, 97,54%, dan 96,57%. Dengan *classifier backpropagation* dan SVM, diketahui bahwa fitur HSV mampu mendeteksi IDC negatif dideteksi IDC negatif (TP) yaitu 527 data dari 540 citra IDC negatif. Sedangkan untuk data yang terdeteksi sebagai IDC positif terdeteksi IDC positif (TN) 515 data dari 540 data IDC positif. Dengan nilai sensitivitas 95,64% kesalahan untuk kasus sebagai terdiagnosa IDC positif terdeteksi IDC negatif (FN) mencapai 4,36%.

Tabel 4.8 Hasil dari klasifikasi fitur Histogram

Metode	Number Of Fold	TP	TN	FN	FP	Sensitivitas (%)	Spesifisitas (%)	Akurasi (%)
<i>Back-propagation</i>	5	491	506	34	49	93,52	91,17	92,31
	7	490	504	36	50	93,15	90,97	92,03
	10	492	497	43	48	91,96	91,19	91,57
SVM	5	495	490	50	45	90,82	91,58	91,2
	7	497	493	47	43	91,36	91,97	91,67
	10	495	493	47	45	91,32	91,63	91,48

Berdasarkan tabel 4.8 hasil pengujian klasifikasi pada fitur Histogram dengan *fold* yang berbeda yaitu 5, 7, dan 10. Tingkat sensitivitas dan akurasi tertinggi terdapat pada pengujian dengan metode *backpropagation* untuk *fold* 5 dengan nilai 93,52% dan 92,31%. Sedangkan untuk tingkat spesifisitas tertinggi diperoleh dengan metode *backpropagation* untuk nilai *fold* 7 dengan nilai 90,97%. Dengan *classifier backpropagation*, diketahui bahwa fitur Histogram mampu mendeteksi IDC negatif dideteksi IDC negatif (TP) yaitu 491 data dari 540 citra IDC negatif. Sedangkan untuk data yang terdeteksi sebagai IDC positif terdeteksi IDC positif (TN) 506 data dari 540 data IDC positif. Dengan nilai sensitivitas 93,52% kesalahan untuk kasus sebagai terdiagnosa IDC positif terdeteksi IDC negatif (FN) mencapai 6,48%.

Tabel 4.9 Hasil dari klasifikasi fitur GLCM+RGB+HSV+Histogram

Metode	Number Of Fold	TP	TN	FN	FP	Sensitivitas (%)	Spesifisitas (%)	Akurasi (%)
<i>Back-propagation</i>	5	526	518	22	14	95,98	97,36	96,67
	7	526	521	19	14	96,51	97,38	96,94
	10	527	518	22	13	95,99	97,55	96,75
SVM	5	525	511	29	15	94,76	97,14	95,92
	7	526	511	29	14	94,77	97,33	96,01
	10	526	511	29	14	94,77	97,33	96,01

Berdasarkan tabel 4.9 hasil pengujian klasifikasi dengan menggabungkan dari keempat fitur yaitu GLCM, RGB, HSV, dan Histogram. Dari hasil pengujian, tingkat sensitivitas dan akurasi tertinggi terdapat pada metode *backpropagation* untuk *fold* 7 dengan nilai 96,51% dan 96,94%. Sedangkan untuk tingkat spesifisitas tertinggi diperoleh dengan metode *backpropagation* untuk nilai *fold* 10 dengan nilai 97,55% yang berarti memiliki kemungkinan untuk dapat terdiagnosa dengan benar pada citra IDC positif terdeteksi IDC positif cukup tinggi. Dari penggabungan keempat fitur tersebut membuktikan bahwa benar adanya dapat meningkatkan nilai akurasi. Setelah itu, dengan nilai sensitivitas 96,51% kesalahan untuk kasus sebagai terdiagnosa IDC positif terdeteksi IDC negatif rendah yaitu 3,49%. Dengan ini dapat dinyatakan hasil pengujian dari penggabungan keempat fitur telah berhasil diimplementasikan dengan menghasilkan data yang lebih akurat.

Dari keempat fitur yang digunakan dalam pengujian menghasilkan 15 ciri yang berbeda. Dalam pengujian terhadap citra histopatologi pada IDC ini, banyaknya ciri-ciri yang digunakan memiliki pengaruh dalam meningkatkan nilai akurasi. Akan tetapi penggabungan keempat fitur tersebut memerlukan waktu eksekusi sedikit lebih lama jika dibandingkan ketika hanya menggunakan satu atau dua fitur. Penyebab lamanya waktu eksekusi yang digunakan karena banyaknya ciri serta penggunaan *classifier* yang lebih dari satu dalam melakukan pengujian.

4.3 Perbandingan Paper Terdahulu

Pada tahap ini, peneliti membandingkan metode yang telah digunakan sebelumnya dengan metode yang diusulkan dari data yang sama yaitu IDC *breast histopatology images* yang didapat pada situs Kaggle. Parameter yang dibandingkan selain metode yang digunakan adalah akurasi dan jumlah *dataset*. Tabel 4.10 merupakan perbandingan dari performa pada metode terdahulu dengan metode yang diusulkan.

Tabel 4.10 Perbandingan performa dengan metode terdahulu

No	Penulis (Tahun)	Metode	Akurasi	Jumlah data Histopatologi yang digunakan
1	Aziza M. Romano, Alexander A. Hernandez (2019) [4].	Menggunakan ekstraksi fitur CNN dengan arsitektur <i>convolutional layer</i> , <i>pooling layer</i> baru yaitu <i>accept-recept pooling</i> , <i>dropout</i> , dan <i>fully connected layer</i> .	85,41%	90.000
2	Fransisco Perdigon Romero, An Tang, Samuel Kadoury (2019) [5].	Menggunakan CNN dengan model arsitektur dari Inception yang mengintegrasikan <i>batch normalization</i> (BN) pada tiap langkah <i>convolution</i> , <i>max pooling</i> , <i>dropout</i> , dan <i>fully connected</i> .	89%	227.522

3	Fina Afifana Rahman (2020) [6].	Menggunakan CNN, MobileNet, dan VGG-16.	87,9%	50.840
4	Metode yang diusulkan.	Gabungan Ekstraksi Fitur dari GLCM, RGB, HSV, dan Histogram menggunakan <i>classifier backpropagation</i> dan SVM.	96,94%	1080

Berdasarkan tabel 4.10 diatas, penggunaan metode yang diusulkan diketahui menghasilkan akurasi yang lebih tinggi dibandingkan dengan metode yang digunakan pada penelitian sebelumnya. Nilai akurasi yang dicapai yaitu 96,94% dengan jumlah data yang digunakan 1080 citra. Jika dilihat, metode yang digunakan pada penelitian terdahulu hanya menggunakan metode CNN (*Convolutional Neural Network*) dengan penambahan beberapa model, sedangkan metode yang diusulkan memakai empat fitur dan memiliki 15 ciri yang digunakan. Untuk jumlah data, metode yang diusulkan menggunakan jumlah data paling sedikit diantara penelitian sebelumnya. Semakin banyak jumlah data yang digunakan, maka semakin besar juga kemungkinan jumlah data yang salah.

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

1. Berdasarkan hasil penelitian, menggunakan empat fitur citra dengan jumlah 15 ciri. Masing-masing fitur tersebut adalah GLCM yang terdiri dari 4 ciri yaitu *contrast*, *correlation*, *energy*, dan *homogeneity*. Setelah itu ada fitur RGB yang terdiri dari 3 ciri yaitu *red*, *green*, dan *blue*. Dilanjutkan dengan fitur HSV yang terdiri dari 3 fitur yaitu *hue*, *saturation*, dan *value*. Kemudian yang terakhir Histogram memiliki 5 ciri yaitu *mean*, *variance*, *standard deviation*, *skewness*, dan *entropy*.
2. Penggabungan dari keempat ciri fitur mendapatkan hasil akurasi tertinggi dengan menggunakan *classifier backpropagation*. Nilai akurasi yang didapatkan yaitu 96,94%, dengan nilai sensitivitas 96,51%. Dari hasil penelitian dapat dinyatakan hasil pengujian dari penggabungan keempat fitur telah berhasil diimplementasikan dengan menghasilkan data yang lebih akurat.

5.2 Saran

Terdapat beberapa saran sebagai pertimbangan untuk dapat digunakan pada penelitian selanjutnya, yaitu:

- a. Melakukan pemisahan *dataset* bagus dan rusak, sehingga dapat menggunakan *dataset* dengan skala besar tanpa mengkhawatirkan penggunaan *dataset* yang rusak.
- b. Menambah ekstraksi fitur serta penggunaan *classifier* yang berbeda.
- c. Melakukan seleksi fitur untuk mendapatkan hasil yang lebih akurat dalam deteksi dan klasifikasi kanker payudara.

DAFTAR PUSTAKA

- [1] X. Zhou *et al.*, “A Comprehensive Review for Breast Histopathology Image Analysis Using Classical and Deep Neural Networks,” *Prep. Pap. IEEE Trans. JOURNALS*, vol. 4, 2016.
- [2] C. DeSantis, J. Ma, L. Bryan, and A. Jemal, “Breast cancer statistics, 2013,” *CA. Cancer J. Clin.*, vol. 64, no. 1, pp. 52–62, 2014.
- [3] R. Erwandi and S. Suyanto, “Klasifikasi Kanker Payudara Menggunakan Residual Neural Network,” *Indones. J. Comput.*, vol. 5, no. 1, pp. 45–52, 2020.
- [4] A. M. Romano and A. A. Hernandez, “Enhanced Deep Learning Approach for Predicting Invasive Ductal Carcinoma from Histopathology Images,” *Int. Conf. Artif. Intell. Big Data*, pp. 142–148, 2019.
- [5] F. P. Romero, “Multi-Level Batch Normalization in Deep Networks for Invasive Ductal Carcinoma Cell Discrimination in Histopathology Images,” *Int. Symp. Biomed. Imaging (ISBI 2019)*, pp. 1092–1095, Apr. 2019.
- [6] F. A. Rahman, “Klasifikasi Invasive Ductal Carcinoma Menggunakan Convolutional Neural Network,” Skripsi S1, Universitas Muhammadiyah Malang, Malang, 2020.
- [7] E. F. Himmah, M. Widyaningsih, and Maysaroh, “Identifikasi Kematangan Buah Kelapa Sawit Berdasarkan Warna RGB Dan HSV Menggunakan Metode K-Means Clustering,” *Sains dan Inform.*, vol. 6, no. 2, pp. 193–202, 2020.
- [8] A. S. Hyperastuty and R. Riries, “Artificial Neural Network Dalam Menentukan Grading Histopatologi Kanker Payudara,” *Biosains Pascasarj.*, vol. 19, no. 2, pp. 176–188, 2017.
- [9] T. A. C. S. medical and editorial content Team, “Understanding a Breast Cancer Diagnosis Types of Breast Cancer,” *American Cancer Society*, 2019. [Online]. Available: <https://www.cancer.org/content/dam/CRC/PDF/Public/8580.00.pdf>.
- [10] S. I. dan T. M. Zaini, *Pengolahan Citra Digital*, Maret. Bandar Lampung: Anugrah Utama Raharja (AURA), 2014.
- [11] A. H. H. Slamet, B. H. Purnomo, and D. W. Soediby, “Model Jaringan Syaraf Tiruan untuk Prakiraan Harga Komponen Bahan Baku Pakan Unggas di PT XYZ,” *Ind. J. Teknol. dan Manaj. Agroindustri*, vol. 9, no. 2, pp. 151–161, 2020.
- [12] J. Han, J. Pei, and M. Kamber, *Data Mining Concepts and Techniques*, Third. Waltham: Morgan Kaufmann, 2012.
- [13] P. Mooney, “Breast Histopathology Images,” 2017. [Online]. Available: <https://www.kaggle.com/paultimothymooney/breast-histopathology-images>.
- [14] S. Mutrofin, Y. Agustiawan, and E. Kurniawan, “Ekstraksi Fitur Warna Menggunakan Hue Saturation Value Untuk Klasifikasi Tumbuhan Berdasarkan Citra Daun,” Jombang, 2015.
- [15] A. Ramdhan, “Klasifikasi Citra Rontgen Paru-Paru dengan Ekstraksi Fitur Histogram dan Metode Naive Bayes Classifier,” Skripsi S1, Institut Teknologi Sepuluh Nopember,

Surabaya, 2016.



LAMPIRAN

Lampiran 1 – Hasil ekstraksi pada fitur GLCM

No	<i>Contrast</i>	<i>Correlation</i>	<i>Energy</i>	<i>Homogeneity</i>	Kelas
1	0.737736	0.664265726	0.287688	0.808707309	0
2	1.060823	0.627639433	0.143539	0.754442246	0
3	1.29536	0.596912946	0.289491	0.775945454	0
4	0.945814	0.791031119	0.098319	0.735440045	0
5	0.73353	0.753459547	0.131014	0.778986603	0
6	0.882436	0.697461985	0.107111	0.745001562	0
7	1.044577	0.59318825	0.191612	0.751470603	0
8	1.022407	0.637678388	0.121395	0.729472442	0
9	1.053975	0.490102373	0.11572	0.716588817	0
10	0.869463	0.529215227	0.158959	0.746435062	0
11	0.684611	0.170842543	0.398667	0.821977995	0
12	0.607212	0.541506871	0.330382	0.822082153	0
13	1.08005	0.368999479	0.141384	0.697546841	0
14	1.869954	0.423100532	0.082726	0.629277528	0
15	0.621556	0.610466668	0.178353	0.781990143	0
16	0.725812	0.680471474	0.165345	0.782981258	0
17	1.499317	0.566515894	0.072096	0.667925968	0
18	0.997666	0.351518822	0.137375	0.702663647	0
19	1.078222	0.590634462	0.122553	0.715208073	0
20	1.453265	0.522869601	0.08121	0.672604089	0
21	0.611793	0.83221248	0.185758	0.806047862	0
22	0.891533	0.71490522	0.136895	0.763796404	0
23	0.817724	0.694819563	0.135829	0.76058552	0
24	0.756868	0.57351256	0.146728	0.746410523	0
25	0.899863	0.612249141	0.113648	0.735440441	0
26	0.831493	0.724233692	0.173534	0.778193079	0
27	0.726768	0.716341214	0.149244	0.786114883	0
28	0.561541	0.516636559	0.212547	0.779979141	0
29	1.356526	0.674661281	0.10349	0.706499187	0
30	1.030902	0.679231842	0.16249	0.745741774	0
31	0.883932	0.432006005	0.141546	0.726402679	0
32	0.744742	0.546711513	0.128747	0.746508087	0
33	1.003053	0.396739132	0.105223	0.693703422	0
34	0.986195	0.464375985	0.098748	0.696289254	0
35	1.032682	0.510733823	0.091585	0.700219388	0
36	0.762661	0.536277913	0.126981	0.734206164	0
37	0.790812	0.411085038	0.178374	0.74702735	0
38	0.704296	0.461764055	0.190739	0.759831945	0
39	0.624246	0.378608658	0.250307	0.78081886	0

No	<i>Contrast</i>	<i>Correlation</i>	<i>Energy</i>	<i>Homogeneity</i>	Kelas
40	0.746787	0.690202699	0.138446	0.751791476	0
41	0.431406	0.43811059	0.285582	0.804337082	0
42	0.503236	0.605856551	0.382913	0.838998855	0
43	1.089673	0.604594102	0.090072	0.710659795	0
44	0.795473	0.261412947	0.17393	0.727312057	0
45	0.874302	0.645775653	0.227619	0.779061363	0
46	0.856241	0.546353204	0.13232	0.747391504	0
47	0.962736	0.471895477	0.172045	0.742516347	0
48	0.868684	0.429178828	0.199172	0.75429394	0
49	1.371064	0.607527223	0.080511	0.684187179	0
50	1.189204	0.539406685	0.094426	0.699528773	0
51	0.938955	0.35014357	0.122017	0.707124427	1
52	1.261083	0.472875853	0.094122	0.677864674	1
53	1.361633	0.419553902	0.099587	0.670047237	1
54	1.161137	0.461973417	0.105684	0.694356102	1
55	0.964796	0.225059029	0.142899	0.696184472	1
56	1.5809	0.333849233	0.079064	0.642216889	1
57	1.876828	0.370808159	0.075015	0.632720731	1
58	1.081087	0.544572568	0.149947	0.759519159	1
59	1.833886	0.567671259	0.082555	0.644299399	1
60	0.945562	0.477540791	0.147419	0.74500774	1
61	0.725535	0.465686708	0.191381	0.760797758	1
62	1.335239	0.263211366	0.095104	0.662841906	1
63	0.673751	0.506192151	0.139867	0.760257497	1
64	0.84783	0.470397083	0.255006	0.77480935	1
65	0.675229	0.343250803	0.22649	0.766125746	1
66	1.585298	0.556474127	0.076548	0.661631979	1
67	2.142099	0.253708224	0.061433	0.591369742	1
68	1.678794	0.637619815	0.108081	0.68555347	1
69	1.504411	0.423272686	0.088777	0.653062196	1
70	1.219769	0.658648138	0.092056	0.699016451	1
71	1.500975	0.697359841	0.059457	0.669603538	1
72	0.571443	0.618960514	0.201302	0.80550892	1
73	1.010437	0.574389875	0.120737	0.735768083	1
74	1.050137	0.432162079	0.131846	0.711609017	1
75	0.822291	0.270973589	0.24812	0.760803172	1
76	1.205337	0.527003553	0.107422	0.718869776	1
77	1.469671	0.630594062	0.069911	0.691010759	1
78	1.723805	0.600162296	0.066918	0.648053033	1
79	1.620519	0.612228451	0.060387	0.669189713	1
80	1.694511	0.383949406	0.083057	0.650996876	1
81	0.921808	0.368371637	0.164223	0.738069971	1
82	1.31918	0.614441524	0.075452	0.68355522	1
83	1.335868	0.410851988	0.094742	0.680966477	1

No	Contrast	Correlation	Energy	Homogeneity	Kelas
84	1.295852	0.357727577	0.119251	0.69507011	1
85	1.393957	0.475237575	0.092554	0.679828231	1
86	0.970571	0.596570812	0.092853	0.712013154	1
87	1.225504	0.636071675	0.091707	0.707176142	1
88	0.685312	0.39093843	0.201877	0.777929404	1
89	0.829069	0.414815756	0.169708	0.738378265	1
90	2.210714	0.576248184	0.05353	0.621299989	1
91	0.682209	0.598732552	0.144159	0.747840414	1
92	0.672539	0.603353784	0.12195	0.759634701	1
93	0.858834	0.62155273	0.102489	0.730447869	1
94	0.986183	0.448642649	0.114938	0.71443711	1
95	1.019059	0.412859907	0.13707	0.722182236	1
96	1.569577	0.527330166	0.08252	0.677231972	1
97	1.048355	0.56496323	0.10379	0.715497085	1
98	1.006139	0.647212336	0.135861	0.734919339	1
99	0.912445	0.463063347	0.24573	0.782171387	1
100	1.237726	0.546172632	0.111428	0.700211717	1



Lampiran 2 – Hasil ekstraksi pada fitur RGB

No	Red	Green	Blue	Kelas
1	235.1568	211.3772	228.8872	0
2	228.0308	190.9264	216.262	0
3	224.7616	208.0456	220.734	0
4	205.7152	154.2152	187.3052	0
5	222.7724	175.5996	201.1376	0
6	219.5752	165.7608	194.968	0
7	229.4264	199.742	217.4316	0
8	226.5352	185.0228	207.6072	0
9	225.838	193.1944	212.8516	0
10	228.764	201.8272	218.542	0
11	201.854	110.6304	157.4848	0
12	205.1104	102.3336	145.8048	0
13	152.69	92.028	145.6816	0
14	157.8456	97.7708	146.6172	0
15	222.5352	141.7288	181.2252	0
16	220.5144	141.9048	181.014	0
17	206.7804	149.7736	183.736	0
18	205.0176	130.5884	171.888	0
19	208.418	137.9612	171.9212	0
20	222.4456	175.7128	197.9012	0
21	219.3136	169.6104	199.5728	0
22	219.37	166.584	198.0896	0
23	226.7424	186.5916	208.8964	0
24	219.2264	161.1192	193.1596	0
25	225.0468	167.7924	197.8356	0
26	230.7732	186.2908	209.2204	0
27	220.2964	150.2808	187.0208	0
28	222.4812	126.8848	170.6732	0
29	220.6756	175.7468	201.0328	0
30	226.606	194.1928	213.842	0
31	217.4636	164.7912	197.6332	0
32	220.5156	169.7808	200.6392	0
33	215.8584	176.7524	205.034	0
34	214.998	173.4232	202.48	0
35	215.8936	160.554	190.9464	0
36	215.38	149.1728	183.0548	0
37	222.9508	155.9128	189.2376	0
38	223.1628	153.5428	187.7276	0
39	214.8432	119.074	163.4132	0
40	215.5828	144.7348	182.368	0
41	215.4712	127.32	171.8828	0
42	214.5668	128.4044	169.8408	0

No	Red	Green	Blue	Kelas
43	214.4272	170.2312	198.482	0
44	204.9636	133.1048	176.8084	0
45	233.1344	203.088	219.8712	0
46	230.3628	194.5948	215.8336	0
47	227.6568	206.5584	219.1092	0
48	229.3468	210.5984	221.6836	0
49	216.998	164.7744	195.0288	0
50	148.6588	94.8868	147.9384	0
51	148.0092	103.5648	153.1464	1
52	149.1052	104.848	155.34	1
53	205.5088	146.456	179.5748	1
54	181.2412	130.6532	170.9756	1
55	139.706	99.5916	150.4156	1
56	139.684	98.8364	148.8952	1
57	143.2368	89.9032	146.4324	1
58	129.0112	91.844	144.0312	1
59	193.4052	150.0196	182.7496	1
60	188.584	141.6072	180.9904	1
61	140.3912	98.0008	149.5632	1
62	150.1964	93.9844	148.18	1
63	170.6252	98.6956	147.9376	1
64	175.456	96.3028	146.2652	1
65	182.0924	136.1852	179.9268	1
66	162.6796	117.6764	165.2552	1
67	144.6564	98.2172	147.54	1
68	164.0328	109.3416	157.0704	1
69	173.146	117.6832	161.6616	1
70	171.8412	123.128	164.138	1
71	230.0464	204.9596	217.0188	1
72	220.5952	190.7976	207.2536	1
73	163.6832	118.7484	155.624	1
74	175.518	115.936	153.0556	1
75	176.8388	129.1516	165.2236	1
76	166.434	122.5592	160.5968	1
77	161.206	113.56	147.1716	1
78	187.34	133.2536	160.2708	1
79	167.2532	93.462	133.6748	1
80	169.3136	85.4912	127.2752	1
81	210.6448	159.8916	183.3912	1
82	203.0736	139.18	168.6504	1
83	187.9784	136.9228	171.0832	1
84	179.3564	139.8524	173.778	1
85	203.6852	127.4932	158.8456	1

No	<i>Red</i>	<i>Green</i>	<i>Blue</i>	Kelas
86	206.0124	162.9728	186.1748	1
87	168.888	100.9324	142.3872	1
88	164.4956	90.224	129.8812	1
89	175.2676	113.4764	147.2124	1
90	202.8228	153.9476	181.9468	1
91	189.6924	141.02	173.4172	1
92	182.4688	129.638	164.7776	1
93	168.2452	106.354	149.4136	1
94	176.546	111.2252	151.9684	1
95	173.3964	120.0268	157.6784	1
96	178.1236	119.3992	158.2332	1
97	175.2528	124.6132	161.424	1
98	178.8072	117.6272	156.822	1
99	178.6832	127.4448	163.9972	1
100	187.372	130.274	165.2768	1



Lampiran 3 – Hasil ekstraksi pada fitur HSV

No	<i>Hue</i>	<i>Saturation</i>	<i>Value</i>	Kelas
1	0.777543	0.114177	0.924784	0
2	0.823602	0.177605	0.896218	0
3	0.712666	0.094564	0.885274	0
4	0.825036	0.274811	0.807705	0
5	0.835949	0.223465	0.874184	0
6	0.851494	0.256923	0.861644	0
7	0.802161	0.140873	0.900725	0
8	0.836438	0.194046	0.888918	0
9	0.836125	0.152964	0.886279	0
10	0.812722	0.125489	0.897956	0
11	0.908715	0.451961	0.791798	0
12	0.922843	0.505641	0.804406	0
13	0.846731	0.418239	0.609663	0
14	0.858826	0.40823	0.623242	0
15	0.913499	0.367328	0.87274	0
16	0.902023	0.363518	0.864915	0
17	0.881827	0.293016	0.811424	0
18	0.903868	0.370131	0.804056	0
19	0.910879	0.348556	0.81747	0
20	0.879048	0.222064	0.872522	0
21	0.824798	0.241168	0.860685	0
22	0.859626	0.252363	0.860773	0
23	0.854959	0.187807	0.889512	0
24	0.897141	0.2714	0.859815	0
25	0.862584	0.262718	0.882791	0
26	0.798168	0.2044	0.905427	0
27	0.893187	0.323576	0.864071	0
28	0.921266	0.432219	0.872522	0
29	0.850647	0.222032	0.866251	0
30	0.823881	0.158002	0.889981	0
31	0.881062	0.248109	0.853567	0
32	0.889448	0.23608	0.865609	0
33	0.864765	0.188784	0.847653	0
34	0.869558	0.201706	0.843967	0
35	0.892062	0.263612	0.847315	0
36	0.9084	0.313083	0.844692	0
37	0.907705	0.303728	0.874489	0
38	0.908048	0.314993	0.875332	0
39	0.920188	0.447546	0.842629	0
40	0.894924	0.337241	0.845614	0
41	0.912934	0.410609	0.845005	0
42	0.915097	0.404989	0.841518	0

No	Hue	Saturation	Value	Kelas
43	0.857757	0.218877	0.842221	0
44	0.895313	0.353321	0.804141	0
45	0.803621	0.1389	0.915741	0
46	0.840719	0.163108	0.904946	0
47	0.798636	0.101466	0.893754	0
48	0.762692	0.089797	0.900627	0
49	0.874013	0.255211	0.851771	0
50	0.855512	0.187904	0.876088	0
51	0.814674	0.355419	0.613586	1
52	0.81034	0.356917	0.623903	1
53	0.895894	0.294205	0.806505	1
54	0.862388	0.285651	0.712433	1
55	0.800423	0.360766	0.593633	1
56	0.804887	0.365294	0.588819	1
57	0.817939	0.431258	0.604728	1
58	0.787039	0.40246	0.567799	1
59	0.854703	0.239863	0.77189	1
60	0.857616	0.260321	0.745285	1
61	0.80028	0.37159	0.600662	1
62	0.82781	0.411472	0.623133	1
63	0.876172	0.433991	0.677242	1
64	0.887906	0.45437	0.69051	1
65	0.828964	0.291037	0.734279	1
66	0.82339	0.322083	0.663139	1
67	0.821242	0.386497	0.593038	1
68	0.849138	0.359496	0.653475	1
69	0.85205	0.347613	0.687245	1
70	0.84219	0.321948	0.681831	1
71	0.846326	0.114117	0.902409	1
72	0.826676	0.145384	0.865485	1
73	0.850949	0.299089	0.657329	1
74	0.881927	0.345119	0.693955	1
75	0.859787	0.292179	0.707816	1
76	0.841033	0.305097	0.673774	1
77	0.877356	0.326958	0.632204	1
78	0.891366	0.309571	0.734698	1
79	0.901204	0.456964	0.656168	1
80	0.908716	0.49901	0.664091	1
81	0.908217	0.255418	0.826071	1
82	0.908928	0.320384	0.796856	1
83	0.877016	0.281843	0.741548	1
84	0.847317	0.240958	0.710809	1
85	0.930022	0.38407	0.798765	1
86	0.889315	0.223409	0.807912	1

No	Hue	Saturation	Value	Kelas
87	0.892979	0.406689	0.663462	1
88	0.900822	0.455446	0.64678	1
89	0.90585	0.389241	0.687336	1
90	0.899186	0.246695	0.795661	1
91	0.883539	0.263787	0.744267	1
92	0.883737	0.301246	0.715882	1
93	0.876757	0.378696	0.663882	1
94	0.888188	0.377944	0.694827	1
95	0.865547	0.327601	0.681598	1
96	0.875245	0.340901	0.699324	1
97	0.864029	0.306369	0.690445	1
98	0.884708	0.351833	0.705413	1
99	0.87038	0.301025	0.701374	1
100	0.870608	0.316046	0.735586	1



Lampiran 4 – Hasil ekstraksi pada fitur Histogram

No	Mean	Variance	Standard Deviation	Skewness	Entropy	Kelas
1	220.4668	1224.976	34.99966	1.235474	6.02477	0
2	204.906	1541.608	39.26331	0.729955	6.643103	0
3	214.4844	1638.684	40.48066	1.493444	5.491263	0
4	173.3736	2269.821	47.64264	0.24111	6.708932	0
5	192.6136	1457.872	38.18209	0.014727	6.432178	0
6	185.1796	1426.557	37.76979	0.164705	6.617162	0
7	210.632	1252.151	35.38575	0.984264	6.101282	0
8	200.0044	1352.344	36.77424	0.448347	6.587043	0
9	205.1828	948.7918	30.80246	0.978562	6.485828	0
10	211.7792	846.2576	29.09051	0.951676	6.139998	0
11	143.2412	388.143	19.70135	1.282944	6.022262	0
12	138.018	626.5209	25.0304	1.778028	6.039943	0
13	116.2836	785.4896	28.02659	1.358921	6.596293	0
14	121.2948	1528.929	39.10152	1.01731	7.08671	0
15	170.3876	695.023	26.36329	1.163832	6.289114	0
16	169.8636	1037.011	32.20265	0.97823	6.275236	0
17	170.6896	1676.868	40.94957	0.239151	7.062376	0
18	157.5496	741.4691	27.22993	0.592661	6.688841	0
19	162.8932	1233.631	35.12309	0.493073	6.816414	0
20	192.2144	1453.9	38.13005	0.310859	6.712587	0
21	187.8804	1718.949	41.46021	0.245556	5.844219	0
22	185.9532	1456.496	38.16406	0.362555	6.150711	0
23	201.1212	1330.631	36.47781	0.430724	6.435317	0
24	182.154	833.1295	28.86398	0.644651	6.555082	0
25	188.3316	1125.491	33.54834	0.289415	6.553125	0
26	202.2072	1560.203	39.4994	0.446137	6.02781	0
27	175.4012	1131.071	33.63139	0.815558	6.239727	0
28	160.454	477.5527	21.85298	1.370781	6.112429	0
29	192.0512	2063.761	45.42864	0.501741	6.753507	0
30	206.1272	1579.109	39.73801	1.021059	6.446605	0
31	184.29	720.1611	26.83582	0.199599	6.601291	0
32	188.4676	743.0898	27.25967	0.091838	6.671076	0

No	Mean	Variance	Standard Deviation	Skewness	Entropy	Kelas
33	191.6524	762.758	27.61807	0.220013	6.695026	0
34	189.168	851.5646	29.18158	0.008327	6.713265	0
35	180.552	949.1825	30.80881	0.140588	6.805746	0
36	172.8308	736.8558	27.14509	0.441971	6.683769	0
37	179.756	598.8461	24.47133	0.449938	6.485893	0
38	178.2536	573.0637	23.93875	0.727627	6.375548	0
39	152.7656	437.9027	20.92612	0.423685	6.238991	0
40	170.2064	1164.631	34.12669	0.826976	6.282707	0
41	158.7524	265.4999	16.29417	2.026642	5.686391	0
42	158.8876	524.0214	22.89151	2.155967	5.690829	0
43	186.644	1336.049	36.55201	0.054943	6.659326	0
44	159.5832	433.5383	20.82158	0.259429	6.263482	0
45	213.984	1240.995	35.22776	1.09223	6.168654	0
46	207.7088	916.4592	30.27308	0.571526	6.44721	0
47	214.2912	845.1688	29.07179	1.481916	6.166694	0
48	217.4676	703.5874	26.52522	1.272162	5.946269	0
49	183.8272	1683.387	41.0291	0.05857	6.872527	0
50	198.1608	1215.89	34.86962	0.55616	6.70271	0
51	122.502	1130.825	33.62774	0.923657	6.90558	1
52	123.8404	1105.875	33.2547	0.981133	6.864417	1
53	167.884	986.1009	31.40224	0.149201	6.900916	1
54	150.3772	539.2389	23.22152	0.593781	6.482378	1
55	117.3768	1142.255	33.79726	0.601277	6.982821	1
56	116.772	1405.247	37.48663	0.809563	7.102524	1
57	112.284	1174.616	34.27267	0.2338	6.797605	1
58	108.896	2046.498	45.23824	1.054743	7.130838	1
59	166.7208	871.4116	29.51968	0.097356	6.711257	1
60	160.152	638.3961	25.2665	0.617702	6.493494	1
61	116.5436	831.2233	28.83094	0.461567	6.793313	1
62	116.9792	587.4396	24.23715	0.113978	6.490674	1
63	125.8132	708.7503	26.62236	1.159068	6.31236	1
64	125.6552	412.9595	20.32141	1.292086	6.080803	1
65	154.8952	1764.789	42.00939	0.551304	6.997233	1

No	Mean	Variance	Standard Deviation	Skewness	Entropy	Kelas
66	136.5512	1379.185	37.13738	0.412508	7.150725	1
67	117.7304	2200.391	46.90832	1.24771	6.941834	1
68	131.138	1225.901	35.01287	0.93316	6.938142	1
69	139.274	1690.418	41.11469	0.738553	7.018724	1
70	142.3756	2339.606	48.36947	0.450223	7.219961	1
71	213.8296	666.8622	25.82367	1.235618	6.299053	1
72	201.596	1181.506	34.37305	0.790656	6.559667	1
73	136.3876	864.9454	29.40995	0.46144	6.760155	1
74	137.9812	529.0968	23.00211	0.328528	6.399627	1
75	147.5124	1189.523	34.48946	0.267588	6.911577	1
76	140.026	1822.902	42.69546	0.648968	7.088316	1
77	131.638	1997.101	44.68894	0.750595	7.045444	1
78	152.512	1902.787	43.62095	0.076555	7.175519	1
79	120.1036	1314.766	36.2597	0.847952	7.040419	1
80	115.322	647.1575	25.43929	0.827598	6.557443	1
81	177.7536	1654.705	40.67806	0.511634	7.121724	1
82	161.6436	1072.161	32.74388	0.342739	6.90123	1
83	156.086	908.7178	30.14495	0.077849	6.707879	1
84	155.5292	1230.329	35.07605	0.370405	6.889437	1
85	153.8456	1101.319	33.18613	0.456159	6.891099	1
86	178.4916	1465.75	38.28511	0.016594	6.645159	1
87	125.9656	467.8356	21.62951	0.035371	6.243328	1
88	116.9492	642.2714	25.34307	0.609272	6.424597	1
89	135.792	2547.997	50.47769	0.435785	7.376754	1
90	171.7568	750.9585	27.40362	0.4576	6.642203	1
91	159.2764	762.8672	27.62005	0.104338	6.721197	1
92	149.4484	1032.329	32.12988	0.381359	6.896103	1
93	129.7664	811.5678	28.48803	0.211058	6.762313	1
94	135.4004	816.0993	28.56745	0.072357	6.731232	1
95	140.2752	1582.103	39.77566	0.844076	7.013201	1
96	141.3724	1116.385	33.41235	0.906219	6.822638	1
97	143.958	1383.768	37.19904	1.158039	6.67433	1
98	140.3972	793.0314	28.16081	0.214616	6.425239	1

No	Mean	Variance	Standard Deviation	Skewness	Entropy	Kelas
99	146.926	1231.674	35.09522	1.054305	6.779383	1
100	151.3308	1260.513	35.5037	1.113507	6.607795	1

