

**KLASIFIKASI TUMOR PAYUDARA DENGAN CITRA MAMMOGRAM
MENGUNAKAN METODE *CONVOLUTIONAL NEURAL NETWORK*
(CNN)**

(Skripsi)

Oleh

MUHAMMAD FAJRU RAMADHAN

1857051003



**JURUSAN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
2022**

**KLASIFIKASI TUMOR PAYUDARA DENGAN CITRA MAMMOGRAM
MENGUNAKAN METODE *CONVOLUTIONAL NEURAL NETWORK*
(CNN)**

Oleh

MUHAMMAD FAJRU RAMADHAN

Skripsi

**Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA ILMU KOMPUTER**

Pada

**Jurusan Ilmu Komputer
Fakultas Matematika dan Ilmu Pengetahuan Alam**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2022**

ABSTRAK

KLASIFIKASI TUMOR PAYUDARA DENGAN CITRA MAMMOGRAM MENGUNAKAN METODE *CONVOLUTIONAL NEURAL NETWORK* (CNN)

Oleh

MUHAMMAD FAJRU RAMADHAN

Tumor adalah massa jaringan abnormal. Tumor terjadi karena adanya perubahan genetik (mutasi) pada sel. Sel-sel abnormal yang terus tumbuh dan tidak dapat dihancurkan oleh mekanisme tubuh secara perlahan akan memakan sel-sel sehat di sekitarnya. Kondisi ini akan membuat tumor menjadi ganas dan mulai menjadi kanker yang berbahaya. Tumor dapat tumbuh di beberapa bagian tubuh, termasuk payudara. Umumnya tumor payudara diderita oleh wanita, namun tidak menutup kemungkinan bagi pria juga. Penelitian ini berfokus pada klasifikasi tumor payudara untuk deteksi dini kanker payudara berdasarkan citra mammogram payudara wanita. Algoritma *deep learning* digunakan dalam penelitian ini untuk mendeteksi tumor payudara. Klasifikasi tersebut menggunakan algoritma *deep learning* berupa *Convolutional Neural Network* (CNN). Penelitian ini menggunakan arsitektur AlexNet dan GoogleNet. Tujuan dari penelitian ini adalah untuk menganalisis *accuracy*, *precision*, *recall*, dan *F1 score* arsitektur AlexNet dan GoogleNet untuk klasifikasi tumor payudara. Data yang digunakan adalah citra mammogram payudara dari MIAS (*Mammographic Image Analysis Society*) *Database*. Setelah dilakukan beberapa percobaan, didapatkan hasil tertinggi, yaitu *accuracy* 93,24%, *precision* 93,30%, *recall* 93,24%, dan *F1 score* 93,23% yang diperoleh dari arsitektur GoogLeNet dengan skenario pemisahan data 80% *training* 20% *testing* dan dimensi input 224×224 piksel. Penelitian ini memperlihatkan bahwa CNN dengan arsitektur AlexNet dan GoogLeNet mampu mendeteksi tumor payudara dari citra mammogram.

Kata Kunci: *Deep learning*, Tumor payudara, Citra mammogram, AlexNet, GoogLeNet, CNN.

ABSTRACT

CLASSIFICATION OF BREAST TUMOR WITH MAMMOGRAM IMAGE USING CONVOLUTIONAL NEURAL NETWORK (CNN)

By

MUHAMMAD FAJRU RAMADHAN

Tumor is a mass of abnormal tissue. Tumor occurs due to genetic changes (mutations) in cells. Abnormal cells that continue to grow and cannot be destroyed by the body's mechanisms will slowly eat the healthy cells around them. This condition will make tumors become malignant and start being a dangerous cancer. Tumors can grow in several part of the body, including breast. Commonly, breast tumors suffered by women, but it is possible for men too. This research focuses on classifying breast tumor for early detection of breast cancer based on mammogram images of women's breasts. Deep learning algorithm is used in this research to detect breast tumor. The classification uses a deep learning algorithm in the form of Convolutional Neural Network (CNN). This research use AlexNet and GoogleNet architectures. Purpose of this research is to analyze the accuracy, precision, recall, and F1 score of AlexNet and GoogleNet architectures for classifying breast tumor. The data are mammogram images of breast from MIAS (Mammographic Image Analysis Society) Database. After many trials, the highest results are 93.24% accuracy, 93.30% precision, 93.24% recall, and 93.23% F1 score that acquired from GoogLeNet architecture with data separation scenario of 80% training 20% testing and input dimensions of 224×224 pixels. This research shows that CNN with AlexNet and GoogLeNet architectures is able to detect breast tumor from mammogram images.

Keywords: Deep learning, Breast tumor, Mammogram images, AlexNet, GoogLeNet, CNN.

Judul Skripsi : **KLASIFIKASI TUMOR PAYUDARA
DENGAN CITRA MAMMOGRAM
MENGGUNAKAN METODE
CONVOLUTIONAL NEURAL NETWORK
(CNN)**

Nama Mahasiswa : **Muhammad Fajru Ramadhan**

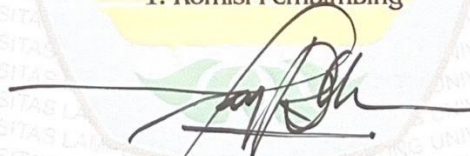
Nomor Pokok Mahasiswa : **1857051003**

Jurusan : **Ilmu Komputer**

Fakultas : **Matematika dan Ilmu Pengetahuan Alam**

MENYETUJUI

1. Komisi Pembimbing



Favorisen R. Lumbanraja, Ph.D.
NIP 19830110 200812 1 002

2. Ketua Jurusan Ilmu Komputer



Didik Kurniawan, S.Si., M.T.
NIP 19800419 200501 1 004

MENGESAHKAN

1. Tim Penguji

Ketua

: **Favorisen R. Lumbanraja, Ph.D.**



Penguji Pembahas

: **M. Reza Faisal, S.T., M.T., Ph.D.**



Penguji Pembahas

: **Dr. rer. nat. Akmal Junaidi, M.Sc.**



2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Dr. Eng. Supto Dwi Yuwono, S.Si., M.T.

NIP. 19740705 200003 1 001

Tanggal Lulus Ujian Skripsi : 25 Juli 2022

PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Muhammad Fajru Ramadhan

NPM : 1857051003

Dengan ini menyatakan bahwa skripsi saya yang berjudul “KLASIFIKASI TUMOR PAYUDARA DENGAN CITRA MAMMOGRAM DENGAN METODE *CONVOLUTIONAL NEURAL NETWORK (CNN)*” adalah benar hasil karya sendiri dan bukan orang lain. Seluruh tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Jika di kemudian hari terbukti skripsi saya adalah hasil penjiplakan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan peraturan yang berlaku.

Bandar Lampung, 04 Agustus 2022

Penulis,



Muhammad Fajru Ramadhan
NPM. 1857051003

RIWAYAT HIDUP



Penulis dilahirkan di Bandar Lampung pada tanggal 27 November 2000 sebagai anak kedua dari dua bersaudara dari pasangan Bapak Bambang Nardianto dan Ibu Maidewi. Penulis menyelesaikan pendidikan Sekolah Dasar (SD) di SD Kartika II-5 Bandar Lampung pada tahun 2012, pendidikan Sekolah Menengah Pertama (SMP) di SMP Negeri 1 Bandar Lampung pada tahun 2015, dan pendidikan Sekolah Menengah Atas (SMA) di SMA Negeri 9 Bandar Lampung pada tahun 2018.

Penulis terdaftar sebagai mahasiswa Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung pada tahun 2018 melalui jalur Seleksi Mandiri Masuk PTN-Barat (SMM PTN-Barat). Selama menjadi mahasiswa, penulis mengemban beberapa tanggung jawab dan berpartisipasi dalam berbagai kegiatan, antara lain.

1. Menjadi Ketua Angkatan Jurusan Ilmu Komputer 2018.
2. Menjadi anggota Adapter Himpunan Mahasiswa Jurusan Ilmu Komputer pada periode 2018/2019.
3. Menjadi anggota pengurus di Bidang Kaderisasi Himpunan Mahasiswa Jurusan Ilmu Komputer (Himakom) pada periode 2018/2019 dan 2019/2020.
4. Menjadi Ketua Pelaksana Pekan Raya Jurusan VIII yang diselenggarakan oleh Himpunan Mahasiswa Jurusan Ilmu Komputer (Himakom) pada tahun 2020.
5. Menjadi Asisten Dosen Jurusan Ilmu Komputer untuk mata kuliah Aljabar Linier pada periode semester ganjil tahun ajaran 2019/2020.

6. Menjadi Asisten Dosen Jurusan Ilmu Komputer untuk mata kuliah Pemrograman Terstruktur pada periode semester genap tahun ajaran 2019/2020.
7. Menjadi Asisten Dosen Jurusan Ilmu Komputer untuk mata kuliah Basis Data pada periode semester ganjil tahun ajaran 2020/2021.
8. Menjadi Asisten Dosen Jurusan Ilmu Komputer untuk mata kuliah Pemrograman Web pada periode semester genap tahun ajaran 2020/2021.
9. Melaksanakan Kerja Praktik di PT Poseidon Aquakultura Sejahtera pada tahun 2021.
10. Melaksanakan Kuliah Kerja Nyata (KKN) pada periode II tahun 2021 di Kelurahan Bumi Waras, Kecamatan Bumi Waras, Kota Bandar Lampung.
11. Mengikuti ujian sertifikasi dan mendapatkan sertifikat *Junior Web Developer* (JWD) yang diselenggarakan oleh Badan Nasional Sertifikasi Profesi (BNSP) pada tahun 2021.

MOTTO

“Maka sesungguhnya bersama kesulitan itu ada kemudahan.”

(QS Al-Insyirah: 5)

“Life is like riding a bicycle. To keep your balance, you must keep moving.”

(Albert Einstein)

“There is only one thing that makes a dream impossible to achieve: the fear of failure.”

(Paulo Coelho)

“Jangan pernah menyerah jika kamu masih ingin mencoba. Jangan biarkan penyesalan datang karena kamu selangkah lagi untuk menang”

(R.A. Kartini)

PERSEMBAHAN

Alhamdulillahilabbilamin

Puji syukur kepada Allah SWT yang telah memberikan rahmat dan hidayah-Nya sehingga dapat menyelesaikan penulisan skripsi ini. Shalawat dan salam saya sanjungkan kepada Nabi Muhammad SAW.

Aku persembahkan karya ini kepada.

Ayahanda Bambang Nardianto dan Ibunda Maidewi

Sebagai tanda terima kasihku kepada ayah dan ibu yang tercinta dan yang tersayang. Terima kasih telah mendidik dan membesarkanku dengan kasih sayang kalian. Terima kasih selalu mendukungku dan mendoakanku dalam segala pilihanku. Terima kasih atas semua pengorbanan, perjuangan, dan doa kalian yang tiada hentinya. Terima kasih ayah dan ibu.

Kakakku Dita Ananda

Terima kasih telah memberikan semangat, dukungan, dan doa.

Seluruh Keluarga Besar, Sahabat, dan Teman-teman

Terima kasih telah memberikan semangat dan bantuan.

Dosen Pembimbing dan Pembahas

Yang senantiasa membimbing, mengarahkan dan memberi motivasi sejak awal hingga terselesaikannya skripsi ini.

Almamater Tercinta, Universitas Lampung

SANWACANA

Puji syukur kehadirat Allah SWT, karena telah memberikan rahmat dan hidayah-Nya kepada saya sehingga saya dapat menyelesaikan skripsi dengan judul “Klasifikasi Tumor Payudara dengan Citra Mammogram menggunakan Metode *Convolutional Neural Network* (CNN).” Saya berharap skripsi ini dapat menambah pengetahuan bagi pembaca tentang klasifikasi data gambar menggunakan metode *deep learning* CNN.

Selama proses penulisan skripsi ini tidak terlepas dari dukungan banyak pihak yang telah membimbing, membantu dan memberi semangat kepada saya, sehingga pada kesempatan ini saya ingin menyampaikan ungkapan terima kasih kepada.

1. Ayah, ibu, kakak, dan keluarga saya yang selalu mendoakan, menyemangati, membiayai serta mendukung saya baik secara moral maupun material. Terima kasih atas doa yang kalian berikan untuk keberhasilan dan kesuksesan saya.
2. Bapak Favorisen R. Lumbanraja, Ph.D. sebagai pembimbing utama yang telah membimbing saya, memberikan kritik dan saran, serta membina dalam menyelesaikan skripsi ini, sehingga skripsi ini dapat diselesaikan dengan baik dan tepat waktu.
3. Bapak M. Reza Faisal, S.T., M.T., Ph.D. sebagai pembahas pertama yang telah memberikan masukan yang sangat bermanfaat dalam penulisan dan perbaikan skripsi ini.
4. Bapak Dr. rer. nat. Akmal Junaidi, M. Sc. sebagai pembahas kedua dan selaku sekretaris Jurusan Ilmu Komputer yang telah memberikan masukan yang sangat bermanfaat dalam penulisan dan perbaikan skripsi ini, serta telah membantu proses administrasi untuk pelaksanaan seminar dan sidang skripsi.

5. Bapak Ardiansyah, S.Kom., M.Kom. selaku Pembimbing Akademik yang telah mengarahkan dalam pengambilan mata kuliah pilihan dan memberikan motivasi selama masa perkuliahan.
6. Bapak Dr. Suropto Dwi Yuwono, M.T. selaku Dekan FMIPA Universitas Lampung.
7. Bapak Didik Kurniawan, S.Si., M.T. selaku Ketua Jurusan Ilmu Komputer Universitas Lampung.
8. Bapak dan Ibu Dosen Jurusan Ilmu Komputer FMIPA Universitas Lampung yang telah memberikan ilmu dan pengalaman dalam hidup untuk menjadi lebih baik.
9. Ibu Ade Nora Maela, Bang Zainuddin, dan Mas Ardi Nofalian yang telah membantu segala urusan administrasi di Jurusan Ilmu Komputer.
10. Annisa Arum Adiza, Agnes Aura Ainisha, Safaana Salwa Salsabilla, Isna Mutiara, Ajeng Prihandini, Dinda Putri Yana, dan Irhab Nicky Febio yang selalu memberikan semangat, dukungan moral, dan canda tawa kepada penulis selama kehidupan perkuliahan.
11. Arbi Hidayatullah, Ridho Alrafi, dan Rika Ningtias Azhari yang telah membantu dan menemani dalam mengurus administrasi seminar dan wisuda.
12. Teman-teman Grup Whatsapp yang namanya selalu berubah dan Heemahooraa yang telah menemani, membantu, dan berbagi canda tawa selama masa perkuliahan.
13. Teman-teman seperbimbingan skripsi yang telah membantu dan memberikan ide selama penyusunan skripsi.
14. Teman-teman Discord yang telah menemani untuk menghilangkan penat dan berbagi cerita.
15. Teman-teman Ilmu Komputer 2018 yang telah menjadi keluarga dan berbagi berbagai macam memori yang seru dan indah selama menjalankan masa studi di Jurusan Ilmu Komputer Universitas Lampung.
16. Semua pihak yang terlibat selama masa perkuliahan dan penyusunan skripsi yang tidak bisa saya sebutkan satu per satu.

Dalam proses penyusunan skripsi ini tentunya terdapat kekurangan yang disebabkan oleh keterbatasan pengalaman dan pengetahuan saya. Kritik dan saran

dari para pembaca sangat saya harapkan sebagai pembelajaran untuk saya. Semoga skripsi ini dapat bermanfaat bagi para pembaca.

Bandar Lampung, 04 Agustus 2022

Penulis,

Muhammad Fajru Ramadhan
NPM. 1857051003

DAFTAR ISI

	Halaman
DAFTAR ISI.....	xiv
DAFTAR TABEL.....	xvi
DAFTAR GAMBAR	xvii
DAFTAR <i>PSEUDOCODE</i>	xix
I. PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah.....	3
1.4. Tujuan.....	4
1.5. Manfaat.....	4
II. TINJAUAN PUSTAKA	5
2.1 Penelitian Terdahulu.....	5
2.2 Tumor dan Kanker Payudara.....	10
2.3 Mammogram	12
2.4 <i>Preprocessing</i>	14
2.5 <i>Convolutional Neural Network (CNN)</i>	15
2.5.1. AlexNet	15
2.5.2. GoogLeNet.....	17
2.5.3. <i>Convolutional Layer</i>	18
2.5.4. <i>Activation Rectified Linear Units (ReLU) Layer</i>	19
2.5.5. <i>Pooling Layer</i>	19
2.5.6. <i>Fully Connected Layer</i>	19
2.5.7. <i>Softmax</i>	20
2.6 <i>Confusion Matrix</i>	20
2.6.1. <i>Accuracy</i>	21
2.6.2. <i>Recall</i>	21

2.6.3.	<i>Precision</i>	21
2.6.4.	<i>F1 Score</i>	22
III.	METODOLOGI PENELITIAN	23
3.1.	Tempat dan Waktu Penelitian	23
3.1.1.	Tempat Penelitian.....	23
3.1.2.	Waktu Penelitian	23
3.2.	Data dan Alat.....	25
3.2.1.	Data	25
3.2.2.	Alat.....	29
3.3.	Alur Kerja Penelitian.....	31
3.3.1.	Pengumpulan Data	32
3.3.2.	<i>Data Preprocessing</i>	32
3.3.3.	Augmentasi Data.....	32
3.3.4.	Pemisahan Data.....	33
3.3.5.	Model	34
3.3.6.	Evaluasi.....	37
IV.	HASIL DAN PEMBAHASAN	38
4.1.	<i>Data Preprocessing</i>	38
4.1.1.	<i>Import Data</i>	38
4.1.2.	<i>Data Cropping</i> dan <i>Resizing</i>	39
4.2.	Augmentasi Data	39
4.3.	<i>Import Data</i> Hasil Augmentasi dan Pemberian Label.....	41
4.4.	Pembagian Data.....	42
4.5.	Klasifikasi menggunakan Model CNN tanpa Augmentasi Data.....	43
4.6.	Pelatihan Model CNN dengan Augmentasi Data.....	44
4.7.	Hasil Klasifikasi menggunakan Model CNN dengan Augmentasi Data	59
4.6.1.	Klasifikasi dengan Model AlexNet.....	59
4.6.2.	Klasifikasi dengan Model GoogLeNet.....	61
4.7.	Pembahasan	62
4.8.	Perbandingan Dengan Penelitian Sebelumnya.....	65
V.	PENUTUP	69
5.1.	Kesimpulan.....	69
5.2.	Saran.....	71
	DAFTAR PUSTAKA	72

DAFTAR TABEL

Tabel	Halaman
1. Penelitian terdahulu terkait klasifikasi penyakit menggunakan CNN	5
2. Pengelompokan stadium berdasarkan sistem klasifikasi TNM (Panigroro et al., 2019).....	11
3. Komposisi arsitektur AlexNet (Sabilla, 2020)	16
4. Komposisi arsitektur GoogLeNet (Sabilla, 2020).....	17
5. <i>Confusion matrix</i> (Han & Kamber, 2011).....	20
6. Rencana waktu penelitian	24
7. Detail informasi dari beberapa data (J Suckling, 1994).....	26
8. Komposisi arsitektur AlexNet yang digunakan pada penelitian ini.....	34
9. Komposisi arsitektur GoogLeNet yang digunakan pada penelitian ini	35
10. Hasil klasifikasi model dari empat arsitektur, tanpa augmentasi dengan dimensi gambar 224×224 piksel.....	43
11. Hasil klasifikasi model dari empat arsitektur, tanpa augmentasi dengan dimensi gambar 256×256 piksel.....	43
12. Rata-rata hasil pelatihan model dari empat arsitektur CNN dengan dimensi gambar 224×224 piksel	57
13. Rata-rata hasil pelatihan model dari empat arsitektur CNN dengan dimensi gambar 256×256 piksel	57
14. Rata-rata hasil validasi model dari empat arsitektur CNN dengan dimensi gambar 224×224 piksel	57
15. Rata-rata hasil validasi model dari empat arsitektur CNN dengan dimensi gambar 256×256 piksel	58
16. Hasil perbandingan klasifikasi	63
17. Perbandingan penelitian sebelumnya dengan AlexNet.....	66
18. Perbandingan penelitian sebelumnya dengan GoogLeNet	67

DAFTAR GAMBAR

Gambar	Halaman
1. Tingkatan stadium pada kanker payudara secara umum (Alodokter, 2021).....	12
2. Pesawat mammografi (Sidabutar, 2012).....	13
3. Hasil mammogram (J Suckling, 1994).	13
4. Arsitektur <i>Convolutional Neural Network</i> (CNN) (Matlab).....	15
5. Ilustrasi arsitektur AlexNet lima <i>layers</i> (Sabilla, 2020).....	15
6. Ilustrasi arsitektur GoogLeNet 13 <i>layers</i> (Sabilla, 2020).	17
7. Ukuran dimensi data mammogram payudara dari MIAS.	26
8. Grafik penyebaran jumlah data.	28
9. Grafik penyebaran data berdasarkan CLASS dan SEVERITY.	28
10. Diagram alur kerja penelitian klasifikasi tumor payudara dengan citra mammogram menggunakan metode <i>Convolutional Neural Network</i> (CNN)..	31
11. Beberapa contoh augmentasi pada data abnormal dengan cara rotasi.	33
12. Beberapa contoh augmentasi pada data normal dengan cara rotasi.	33
13. Ilustrasi arsitektur AlexNet yang digunakan pada penelitian ini.	34
14. Ilustrasi arsitektur GoogLeNet yang digunakan pada penelitian ini.....	35
15. Ilustrasi dan komposisi <i>inception block</i> pada GoogLeNet yang dipakai pada penelitian ini (Szegedy et al., 2015).	36
16. Grafik training model AlexNet 70% <i>training</i> dan 30% <i>testing</i> dimensi 224×224 piksel.	48
17. Grafik training model AlexNet 80% <i>training</i> dan 20% <i>testing</i> dimensi 224×224 piksel.	49
18. Grafik training model GoogLeNet 70% <i>training</i> dan 30% <i>testing</i> dimensi 224×224 piksel.	50

19. Grafik training model GoogLeNet 80% <i>training</i> dan 20% <i>testing</i> dimensi 224×224 piksel.	51
20. Grafik training model AlexNet 70% <i>training</i> dan 30% <i>testing</i> dimensi 256×256 piksel.	52
21. Grafik training model AlexNet 80% <i>training</i> dan 20% <i>testing</i> dimensi 256×256 piksel.	53
22. Grafik training model GoogLeNet 70% <i>training</i> dan 30% <i>testing</i> dimensi 256×256 piksel.	54
23. Grafik training model GoogLeNet 80% <i>training</i> dan 20% <i>testing</i> dimensi 256×256 piksel.	55
24. Plot <i>confusion matrix</i> model AlexNet.	60
25. Plot <i>confusion matrix</i> model GoogLeNet.	61
26. Perbandingan <i>accuracy</i> antar skenario dan arsitektur.	63
27. Perbandingan <i>precision</i> antar skenario dan arsitektur.	64
28. Perbandingan <i>recall</i> antar skenario dan arsitektur.	64
29. Perbandingan <i>F1 score</i> antar skenario dan arsitektur.	65
30. Perbandingan <i>accuracy</i> AlexNet dengan penelitian sebelumnya.	66
31. Perbandingan <i>accuracy</i> GoogLeNet dengan penelitian sebelumnya.	67

DAFTAR PSEUDOCODE

Pseudocode	Halaman
1. Kode program <i>import</i> data.	38
2. Kode program <i>cropping</i> dan <i>resizing</i> data.	39
3. Kode program augmentasi (rotasi) data abnormal.	40
4. Kode program augmentasi (rotasi) data normal.	41
5. Kode program <i>import</i> data hasil augmentasi.	41
6. Kode program pemberian label.	42
7. Kode program pembagian data 70% <i>training</i> dan 30% <i>testing</i>	42
8. Kode program pembagian data 80% <i>training</i> dan 20% <i>testing</i>	43
9. Kode program model CNN arsitektur AlexNet.	45
10. Kode program model CNN arsitektur GoogLeNet.	46
11. Kode program <i>compile</i> dan pelatihan model 80% <i>training</i> dan 20% <i>testing</i>	47
12. Kode program <i>compile</i> dan pelatihan model 70% <i>training</i> dan 30% <i>testing</i>	47
13. Kode program evaluasi model.	59

I. PENDAHULUAN

1.1. Latar Belakang

Tumor merupakan sel-sel yang tumbuh secara abnormal di tubuh. Terdapat dua jenis tumor, yaitu tumor jinak (*benign*) dan tumor ganas (*malignant*). Tumor jinak tidak bersifat kanker dan kebanyakan tidak membahayakan penderita. Tumor ganas bersifat kanker dan dapat menyebabkan kematian. Tumor ganas yang terus tumbuh secara tidak terkontrol dan memakan sel-sel sehat di sekitarnya dan menyebabkan penyakit kanker. Kanker merupakan salah satu penyakit yang mematikan di dunia. Kanker terjadi akibat adanya perubahan (mutasi) genetik pada sel. Mutasi genetik tersebut akan membuat sel menjadi abnormal. Sel abnormal yang terus tumbuh dan tidak dapat dihancurkan oleh mekanisme tubuh, secara perlahan akan menjadi kanker yang berbahaya (Willy, 2019). Di dunia, berdasarkan data dari Globocan 2020, angka kematian akibat kanker tercatat mencapai 10 juta jiwa dari 19,3 juta kasus pada tahun 2020. Angka kematian tersebut meningkat dibandingkan pada tahun 2018, yaitu 9,6 juta jiwa dari 18,1 juta kasus. Di Indonesia sendiri, kasus kematian akibat kanker pada tahun 2020 mencapai 235 ribu jiwa dari 397 ribu kasus (Globocan, 2020).

Tumor dapat tumbuh di beberapa bagian tubuh, salah satunya adalah payudara. Tumor payudara banyak diderita oleh wanita, namun tidak menutup kemungkinan diderita oleh pria. Tumor payudara yang bersifat ganas akan menjadi kanker payudara. Menurut data dari Globocan 2020, kanker payudara menyumbang angka 2,3 juta (11,7%) kasus kanker di dunia dengan tingkat kematian sebanyak 685 ribu jiwa. Di Indonesia, kanker payudara merupakan penyebab kematian kedua pada wanita akibat kanker.

Menurut data dari Globocan 2020, kanker payudara merupakan kanker yang paling banyak diderita oleh masyarakat Indonesia dengan jumlah 66 ribu jiwa (16,6%) dari total kasus kanker yang ada di Indonesia. Dari 66 ribu jiwa tersebut, semuanya adalah wanita. Angka kematian kanker payudara di Indonesia mencapai 23 ribu. Angka tersebut menempatkan kanker payudara sebagai kanker yang paling mematikan kedua di Indonesia (Globocan, 2020).

Tumor payudara yang sudah tumbuh cukup lama dapat dengan mudah dideteksi karena adanya benjolan pada payudara yang terlihat dan teraba, namun tumor payudara yang baru tumbuh akan sulit dideteksi karena tidak teraba dan terlihat. Jika tumor payudara tersebut jinak, maka tidak ada masalah. Namun, jika ternyata tumor payudara tersebut ganas, maka keterlambatan pendeteksian tersebut dapat membahayakan nyawa penderita. Pendeteksian dini tumor payudara dapat dilakukan dengan melakukan CT dan PET Scans pada bagian payudara, kemudian hasilnya akan dilihat oleh manusia. Namun pada usia tumor payudara yang masih muda, pendeteksian dengan cara tersebut rentan terjadi human error. Komputer mengambil peran untuk mengurangi resiko tersebut. Algoritma-algoritma *machine learning* (ML) dan *deep learning* (DL) mulai digunakan untuk mendeteksi kanker. Algoritme-algoritme tersebut dinilai dapat membantu manusia dalam melakukan klasifikasi karena memiliki tingkat akurasi yang tinggi.

Hal-hal tersebut didukung dengan banyaknya penelitian tentang klasifikasi penyakit dengan menggunakan algoritme-algoritme *machine learning* dan *deep learning*. Contohnya, penelitian yang dilakukan oleh Puspitasari et al. menggunakan algoritma *machine learning* berupa *support vector machine* (SVM) untuk melakukan klasifikasi penyakit gigi dengan hasil tingkat akurasi sebesar 94% (Puspitasari et al., 2018). Penelitian yang dilakukan oleh Hanin et. al. untuk mendeteksi penyakit kulit manusia menggunakan algoritma *deep learning* CNN dengan data citra kulit manusia menghasilkan tingkat akurasi sebesar 96% (Hanin et al., 2021). Penelitian oleh Adiwijaya menghasilkan tingkat akurasi deteksi kanker sebesar 98% dengan metode

Principal Component Analysis (PCA) dan *Evolving Neural Network (ENN)* (Adiwijaya, 2018).

Berhubungan dengan penelitian sebelumnya, penelitian ini bertujuan untuk mendeteksi atau mengklasifikasi tumor payudara sebagai pendeteksian dini kanker payudara. Deteksi atau klasifikasi akan dilakukan menggunakan metode *Convolutional Neural Network (CNN)* dengan data berupa gambar citra mammogram payudara wanita. Penelitian ini diharapkan dapat membantu manusia, khususnya praktisi di bidang kesehatan, untuk melakukan prediksi tumor payudara secara cepat dan akurat.

1.2. Rumusan Masalah

Rumusan masalah pada penelitian ini, antara lain.

1. Mendeteksi tumor payudara melalui citra *mammogram* payudara *right craniocaudal (RCC)* dan *left craniocaudal (LCC)* dengan metode *convolutional neural network (CNN)*.
2. Membandingkan hasil klasifikasi antar arsitektur *convolutional neural network (CNN)* yang digunakan pada penelitian ini, dalam mendeteksi tumor payudara melalui citra mammogram payudara *right craniocaudal (RCC)* dan *left craniocaudal (LCC)*.

1.3. Batasan Masalah

Batasan masalah pada penelitian ini, antara lain.

1. Penelitian ini akan dilakukan pemodelan dengan metode *convolutional neural network (CNN)* untuk klasifikasi atau deteksi tumor payudara.
2. Data yang digunakan adalah citra mammogram payudara wanita berjumlah 322 citra yang akan dideteksi kelainannya. Data ini didapatkan dari database MIAS (*Mammographic Image Analysis Society*) yang dapat diakses melalui link <http://peipa.essex.ac.uk/info/mias.html>.
3. Hasil pengujian prediksi menggunakan model yang dibuat akan ditunjukkan melalui beberapa parameter, antara lain *accuracy*, *precision*, *recall*, dan *F1 score*.

4. Tidak dikembangkannya sistem pada penelitian ini, melainkan hanya berupa model.

1.4. Tujuan

Tujuan dari penelitian ini, antara lain.

1. Mengevaluasi kinerja metode *convolutional neural network* (CNN) dalam klasifikasi atau deteksi tumor payudara.
2. Mengetahui hyperparameter yang optimal pada metode *convolutional neural network* (CNN), sehingga menghasilkan model dengan kinerja yang maksimal.

1.5. Manfaat

Manfaat dari penelitian ini, antara lain.

1. Dapat mendeteksi tumor payudara penyebab kanker payudara lebih dini dan akurat.
2. Dapat menentukan *hyperparameter* yang optimal untuk metode *convolutional neural network* (CNN) pada kasus klasifikasi atau deteksi tumor payudara.

II. TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Terdapat beberapa penelitian terdahulu yang berhubungan dengan penelitian ini, dapat dilihat pada Tabel 1.

Tabel 1. Penelitian terdahulu terkait klasifikasi penyakit menggunakan CNN

No	Penelitian	Metode	Data	Hasil
1	Klasifikasi Citra Mammogram Menggunakan Metode K-Means, GLCM, dan <i>Support Vector Machine</i> (SVM) (Setiawan & Putra, 2018).	<ul style="list-style-type: none"> Ekstraksi fitur: <i>Gray level co-occurrence matrix</i> (GLCM) Klasifikasi: <i>Support vector machine</i> (SVM). 	40 data citra mammogram payudara dari MIAS (http://peipa.essex.ac.uk/info/mias.html).	<i>Accuracy</i> : 80%.
2	<i>Mammograms Classification Using Features Vector and Naïve Bayes Classifier</i> (Santika & Angries, 2013).	<ul style="list-style-type: none"> Ekstraksi Fitur: GLCM, GLRLM, dan <i>Chip Histogram</i>. Klasifikasi: <i>Naïve Bayes</i>. 	292 data citra mammogram payudara dari MIAS (http://peipa.essex.ac.uk/info/mias.html).	<ul style="list-style-type: none"> Klasifikasi normal atau abnormal <i>Accuracy</i>: 78%. Klasifikasi <i>benign</i> atau <i>malignant</i> (abnormal) <i>Accuracy</i>: 68%.
3	<i>Accuracy Enhancement in Early Detection</i>	Klasifikasi: <i>Convolutional neural network</i>	322 data citra mammogram payudara dari	<i>Accuracy</i> model <ul style="list-style-type: none"> VGG-16: 64%. VGG-19: 61%.

<p><i>of Breast Cancer on Mammogram Images with Convolutional Neural Network (CNN)</i> (Susilo & Sugiharti, 2021).</p>	<p>(CNN) model VGG-16, VGG-19, ResNet50.</p>	<p>MIAS (http://peipa.ssex.ac.uk/info/mias.html).</p>	<p>•ResNet50: 71%.</p>
<p>4 <i>Sistem Klasifikasi Penyakit Kulit Menggunakan Convolutional Neural Network (CNN)</i> (Hanin et al., 2021).</p>	<p>Klasifikasi: <i>Convolutional neural network</i> (CNN).</p>	<p>1500 data citra kulit berpenyakit (cacar air, campak, jerawat, dan skabies).</p>	<p>•Accuracy: 96,53%. •Precision: 95%. •Recall: 95%. •F1-Score: 95%.</p>

Tabel 1 memperlihatkan beberapa penelitian terdahulu yang berkaitan dengan penelitian ini. Penelitian yang pertama dilakukan oleh Setiawan dan Putra pada tahun 2018. Penelitian ini mengklasifikasikan atau mendeteksi kanker payudara menggunakan metode segmentasi *k-means*, ekstraksi fitur *gray level co-occurrence matrix* (GLCM), dan metode klasifikasi *support vector machine* (SVM) dengan data berupa citra mammogram kedua payudara wanita yang diambil dari posisi kiri dan kanan (RCC dan LCC) . Data citra mammogram ini diambil dari *database* MIAS (*Mammographic Image Analysis Society*). Data citra yang digunakan pada penelitian ini berjumlah 40 citra yang terdiri dari 30 data *training* (15 citra normal dan 15 citra abnormal) dan 10 data *testing* (5 citra normal dan 5 citra abnormal). Sebelum dilakukan klasifikasi, data melalui 3 tahap *processing*. Yang pertama, dilakukan segmentasi menggunakan *k-means clustering* dengan jumlah *cluster* sebanyak 4. Kemudian, citra hasil segmentasi dilakukan proses perbaikan, seperti *filling* objek, morfologi erosi, dilasi, *bwareaopen*, dan *masking*. Yang terakhir, dilakukan ekstraksi ciri pada data menggunakan metode GLCM. Setelah dilakukan data *processing*, data diklasifikasikan menggunakan *support vector machine* (SVM) dan menghasilkan nilai akurasi sebesar 80% (Setiawan & Putra, 2018).

Hasil penelitian yang dilakukan oleh Setiawan dan Putra memperlihatkan bahwa, pendeteksian kanker payudara melalui citra mammogram payudara dengan metode *support vector machine* (SVM) belum menghasilkan akurasi prediksi yang diharapkan. Hal ini terjadi karena kurangnya data yang digunakan pada proses *training*, yaitu hanya 40 citra dan ekstraksi fitur yang kurang maksimal. Oleh karena itu, untuk meningkatkan akurasi prediksi yang maksimal, data yang digunakan dapat diperbanyak lagi.

Penelitian kedua dilakukan oleh Angries dan Santika pada tahun 2013. Penelitian ini bertujuan untuk mengklasifikasikan kanker payudara menggunakan metode *naïve bayes* dengan data citra mammogram kedua payudara wanita (RCC dan LCC) yang diambil dari *database* MIAS. Data yang digunakan sebanyak 292 citra gambar yang terdiri dari 198 gambar payudara normal, 52 gambar payudara abnormal-*benign*, dan 42 gambar payudara abnormal-*malignant*. Data dibagi menjadi 50% data *training*, 30% data *validation*, dan 20% data *testing*. Sebelum dilakukan klasifikasi, data gambar dilakukan ekstraksi fitur menggunakan *gray level co-occurrence matrices* (GLCM), *gray level run length matrices* (GLRLM), dan *chip histogram*. *Features Vector* yang digunakan pada penelitian ini, antara lain *entropy*, *energy*, *contrast*, *sum average*, *variance*, *correlation*, *maximum probability*, *inverse difference moment*, *cluster shade*, dan SREL. Pengklasifikasian pada penelitian ini dilakukan dua kali, yaitu untuk mengklasifikasikan payudara termasuk normal atau abnormal dan untuk mengklasifikasikan payudara yang termasuk abnormal ke dalam tipe *benign* atau *malignant*. Pengujian masing-masing klasifikasi dilakukan dengan menerapkan beberapa skenario untuk mencari akurasi yang optimal. Setelah dilakukan beberapa pengujian, didapatkan akurasi tertinggi untuk klasifikasi payudara normal atau abnormal sebesar 78% dan akurasi tertinggi untuk klasifikasi payudara abnormal termasuk ke dalam tipe *benign* atau *malignant* sebesar 68% (Santika & Angries, 2013).

Penelitian yang dilakukan oleh Angries dan Santika menghasilkan tingkat akurasi yang rendah. Hal tersebut dikarenakan metode klasifikasi *naïve bayes*

yang tidak dapat menghasilkan model yang cukup baik untuk mendeteksi kanker payudara melalui citra mammogram. Model yang kurang baik dapat terjadi karena kurangnya fitur-fitur yang dihasilkan dari *feature extraction* untuk mendeteksi kanker payudara melalui citra mammogram.

Penelitian selanjutnya dilakukan oleh Susilo dan Sugiharti pada tahun 2021. Penelitian ini bertujuan untuk mengetahui kinerja augmentasi data dan *transfer learning* pada CNN dalam mendeteksi kanker payudara. Data yang digunakan dalam penelitian ini adalah data citra mammogram kedua payudara wanita (RCC dan LCC) yang diambil dari *database* MIAS sebanyak 322 data. Sebelum dilakukan klasifikasi, data melewati tahap *preprocessing*, berupa *cropping* dan *resizing*. Data dipotong sedemikian rupa, sehingga didapatkan piksel x dan y diantara 250 sampai 950. *Cropping* atau pemotongan data difokuskan pada area yang terdapat kanker atau tumor. Kemudian, data diubah ukurannya (*resizing*) menjadi 224×224 piksel. Setelah dilakukan *preprocessing*, dilakukan augmentasi data menggunakan *ImageDataGenerator* dengan 9 parameter, yaitu *zoom*, *rotation*, *width shift*, *height shift*, *shear*, *horizontal flip*, dan *vertical flip*. Data dibagi menjadi 70% data *training* dan 30% data *testing*, kemudian dilakukan klasifikasi menggunakan CNN model VGG-16, VGG-19, dan ResNet50 dengan tujuan untuk mengklasifikasikan data menjadi tujuh kelas (CARC, CIRC, SPIC, MISC, ARCH, ASYM, dan NORM). Klasifikasi menghasilkan *accuracy* pada model VGG-16 sebesar 64%, VGG-19 sebesar 61%, dan ResNet50 sebesar 71% (Susilo & Sugiharti, 2021).

Penelitian yang dilakukan oleh Susilo dan Sugiharti menghasilkan akurasi klasifikasi kanker payudara yang lebih tinggi dibandingkan dengan penelitian sejenis dengan metode CNN. Namun, akurasi yang dihasilkan masih tergolong rendah dibandingkan penelitian dengan data yang sama menggunakan algoritme *machine learning* lainnya. Penggunaan model *transfer learning* dan augmentasi data yang kurang maksimal menjadi salah dua penyebab rendahnya tingkat akurasi klasifikasi tersebut. Selain itu, data

yang sedikit juga dapat menyebabkan rendahnya hasil akurasi untuk klasifikasi kanker payudara.

Penelitian keempat dilakukan oleh Hanin, Patmasari, dan Fu'adah pada tahun 2021. Penelitian ini bertujuan untuk mengklasifikasikan atau mendeteksi penyakit kulit menggunakan *Convolutional Neural Network* (CNN). Data diperoleh dari penelitian sebelumnya yang berjudul *Identifikasi Penyakit Kulit Berdasarkan Kombinasi Segmentasi Warna dan Analisis Tekstur Dengan Deteksi Binary Large Object (BLOB) Menggunakan Jaringan Syaraf Tiruan Learning Vector Quantiza* yang dilakukan oleh Setiyowibowo Aditya P. pada tahun 2012. Data berupa gambar berformat jpg sebanyak 1500 citra data penyakit kulit yang terdiri dari cacar air, campak, skabies, dan jerawat dengan pembagian data *training* sebesar 75% atau 1125 citra dan data *testing* sebesar 25% atau 375 citra. Sebelum dilakukan klasifikasi, dilakukan *preprocessing* data berupa *resize* (mengubah resolusi) citra gambar menjadi 64×64 pixel.

Setelah data di-*resize*, data masuk ke dalam tahap *feature extraction learning* yang akan dilakukan proses konvolusi, aktivasi ReLU, dan *pooling*. Setiap konvolusi akan digunakan ukuran kernel yang sama, yaitu 3×3 . Pada setiap *hidden layer*, jumlah *output channel* masing-masing sebesar 8, 16, 32, 64, 128. Setiap *layer* pada proses konvolusi menggunakan *activation ReLU* dan akan dilakukan proses *max pooling* yang akan menghasilkan citra yang telah tereduksi. Setelah proses *max pooling* selesai, tahap selanjutnya adalah *classification*. Tahap *classification* dimulai dengan proses *flattening*, kemudian aktivasi *softmax* yang akan mengklasifikasikan data citra menjadi empat kelas (cacar air, campak, skabies, dan jerawat) *output* penyakit kulit manusia.

Dari model CNN tersebut, dilakukan beberapa percobaan dengan beberapa skenario yang berbeda. Skenario pertama akan menguji performa model berdasarkan jumlah *hidden layer* (3, 4, 5, dan 6). Skenario kedua akan menguji performa model berdasarkan *optimizer* yang digunakan (Adam,

Nadam, SGD, dan RMSprop). Skenario ketiga akan menguji performa model berdasarkan nilai *learning rate* (0,001; 0,01; dan 0,1). Dari pengujian ketiga skenario tersebut, menghasilkan performa model yang paling optimal dengan parameter *hidden layer* berjumlah 5 (8, 16, 32, 64, 128), *optimizer* menggunakan *adam optimizer*, dan nilai *learning rate* sebesar 0,001 yang menghasilkan nilai akurasi sebesar 96,53%, *precision* 95%, *recall* 95%, *f1-score* 95%, dan *loss* 0,2486 (Hanin et al., 2021).

Penelitian oleh Hanin, Patmasari, dan Fu'adah menghasilkan tingkat akurasi yang tinggi dan memuaskan. Penggunaan *convolutional neural network* (CNN) dalam klasifikasi penyakit kulit melalui gambar atau citra kulit sangatlah cocok. CNN, sebagai salah satu metode *deep learning*, mampu untuk melakukan ekstraksi fitur yang baik, sehingga dapat menghasilkan model klasifikasi dan akurasi yang bagus. Selain itu, data gambar atau citra kulit yang digunakan dapat memperlihatkan penyakit kulit dengan jelas.

Dari penelitian yang telah dilakukan sebelumnya, pendeteksian kanker payudara menggunakan citra mammogram payudara dengan metode SVM, *naïve bayes*, dan CNN masih belum menghasilkan tingkat akurasi yang memuaskan. Tingkat akurasi berada pada rentang 60% - 80%. Penelitian ini bertujuan untuk mendeteksi tumor payudara sebagai pendeteksian dini kanker payudara melalui citra mammogram payudara menggunakan metode *convolutional neural network* (CNN) agar mendapatkan tingkat akurasi yang diharapkan.

2.2 Tumor dan Kanker Payudara

Tumor adalah sel-sel yang tumbuh secara abnormal pada tubuh. Tumor atau sel-sel abnormal ini terus tumbuh, sehingga membentuk benjolan di payudara (Ramli et al., 2005). Tumor terbagi menjadi dua jenis, yaitu tumor jinak (*benign*) dan tumor ganas (*malignant*). Tumor jinak tidak bersifat kanker dan kebanyakan tidak membahayakan penderita. Tumor ganas bersifat kanker dan dapat menyebabkan kematian. Tumor dapat tumbuh di berbagai bagian tubuh manusia, tak terkecuali pada payudara. Pertumbuhan tumor yang agresif dan

tidak terkendali pada payudara menjadikan tumor tersebut bersifat ganas dan dapat menyebabkan kanker payudara. Kanker payudara merupakan tumor pada payudara yang bersifat ganas dan merusak sel sehat di sekitarnya. Kanker payudara banyak diderita oleh wanita, tetapi tidak menutup kemungkinan dapat diderita oleh pria (Tapan, 2005). Kanker payudara merupakan salah satu jenis kanker terbanyak yang ada di Indonesia. Menurut *pathological based registration* di Indonesia, kanker payudara menempati urutan pertama dengan frekuensi relatif sebesar 18,6%. Diperkirakan 12 dari 100.000 wanita di Indonesia mengidap kanker payudara (Panigroro et al., 2019). Sampai saat ini, penyebab pasti kanker payudara masih belum ditemukan, tetapi para pakar kesehatan menduga bahwa penyebab kanker payudara adalah multifaktorial (Tambunan, 1991). Multifaktorial atau polifaktorial adalah pewarisan penyakit genetik yang disebabkan oleh beberapa pasang gen bersama dengan lingkungan (Hartono, 2010).

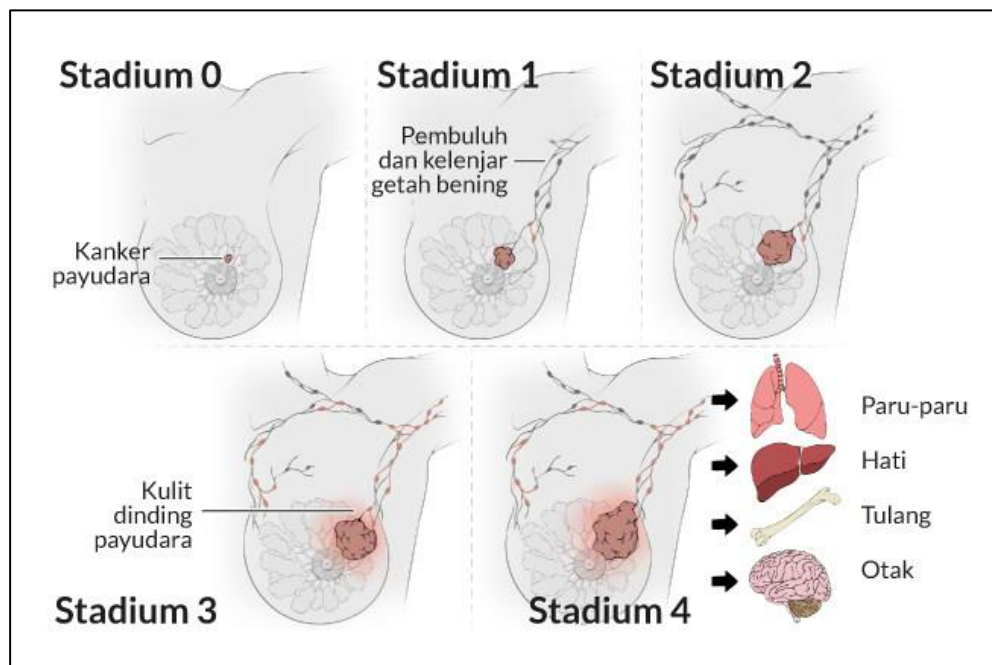
Salah satu sistem pengkategorian kanker payudara adalah sistem klasifikasi TNM. Sistem klasifikasi TNM (Tumor, *nodul*, dan *metastasis*) adalah sistem stadium kanker yang menggambarkan tingkatan kanker di tubuh pasien. Berdasarkan sistem klasifikasi TNM yang dikeluarkan oleh *American Joint Committee on Cancer* (AJCC), kanker payudara dikategorikan berdasarkan beberapa tingkatan atau stadium, yang dapat dilihat pada Tabel 2.

Tabel 2. Pengelompokan stadium berdasarkan sistem klasifikasi TNM (Panigroro et al., 2019)

Stadium	Tumor (T)	Regional (N)	Metastatis (M)
Stadium 0	Tis	N0	M0
Stadium IA	T1	N0	M0
Stadium IB	T0	N1mic	M0
	T1	N1mic	M0
Stadium IIA	T0	N1	M0
	T1	N1	M0
Stadium IIB	T2	N0	M0
	T2	N1	M0
	T3	N0	M0
Stadium IIIA	T0	N2	M0
	T1	N2	M0

Stadium	Tumor (T)	Regional (N)	Metastatis (M)
	T2	N2	M0
	T3	N1-N2	M0
Stadium IIIB	T4	N1-N2	M0
Stadium IIIC	Semua T	N3	M0
Stadium IV	Semua T	Semua N	M1

Secara garis besar, perkembangan atau peningkatan stadium kanker payudara dipengaruhi oleh ukuran tumor yang ada di payudara. Semakin besar tumor yang ada, semakin tinggi tingkat stadiumnya, seperti yang dapat dilihat pada Gambar 1.



Gambar 1. Tingkatan stadium pada kanker payudara secara umum (Alodokter, 2021).

2.3 Mammogram

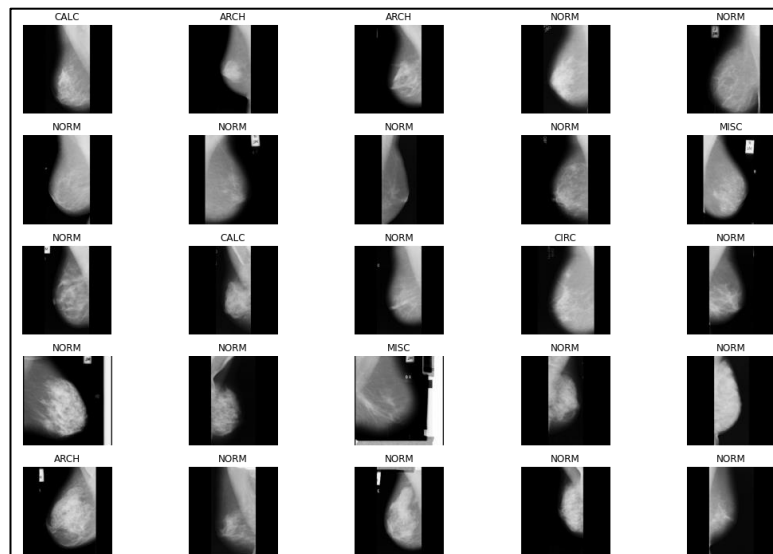
Mammogram atau mammografi merupakan salah satu metode pemeriksaan payudara menggunakan sinar-X atau *X-ray* dosis rendah yang memperlihatkan kelainan pada payudara yang berbentuk sangat kecil (mikrokalsifikasi) dan pertumbuhannya tidak dapat teraba. Mikrokalsifikasi adalah bintik kecil, seperti kapur, pada payudara yang terbentuk dari kalsium.

Pemeriksaan dengan metode mammogram menggunakan alat yang dinamakan dengan pesawat mammografi, seperti yang dapat dilihat pada Gambar 2 (Helja et al., 2013).



Gambar 2. Pesawat mammografi (Sidabutar, 2012).

Sinar-X dari mammografi dapat menembus jaringan payudara yang mengalami kelainan secara jelas (Novianti & Purnami, 2012). Hasil dari pemeriksaan mammogram atau mammografi berupa foto atau citra *X-ray* dari bagian tubuh yang dipindai, seperti pada Gambar 3.



Gambar 3. Hasil mammogram (J Suckling, 1994).

Terdapat beberapa faktor yang dilihat pada saat dilakukannya pemeriksaan mammografi di payudara, antara lain (Novianti & Purnami, 2012).

1. *Intermediate Findings*, yaitu variabel yang menjelaskan keadaan sel atau jaringan yang ada di payudara. Variabel ini terdiri dari lima indikator, yaitu *well defined*, *developing*, *architectural*, *skin thickening*, dan *asymetry*.
2. *Suspicious for Malignancy*, yaitu variabel yang menjelaskan bentuk tumor yang ada di dalam payudara. Variabel ini terdiri dari tiga indikator, yaitu *mass*, *calcification*, dan *speculated sign*.
3. *Breast Imaging Reporting and Data System (BIRADS) Category*, yaitu variabel yang digunakan untuk memprediksi tingkat keganasan benjolan pada pasien kanker payudara.
4. Letak abnormal, yaitu variabel untuk melihat di mana letak kelainan pada payudara.

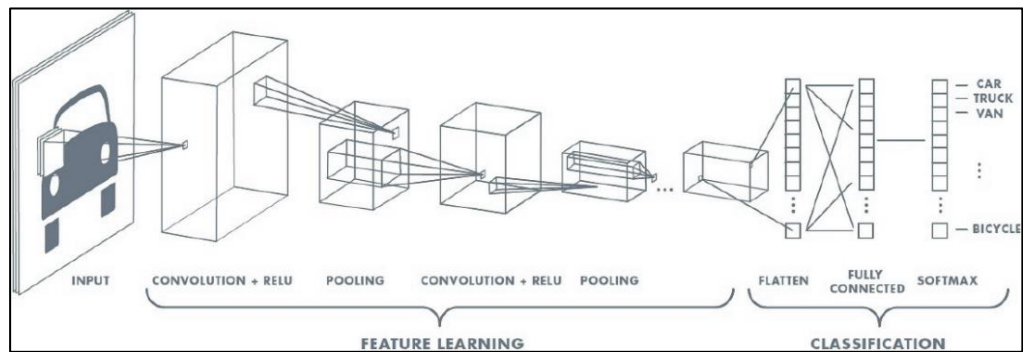
2.4 *Preprocessing*

Data preprocessing adalah teknik yang digunakan untuk mengubah data mentah ke dalam bentuk atau format yang lebih berguna dan efisien. Proses ini diperlukan karena data mentah seringkali memiliki format atau bentuk yang sulit untuk dilakukan *training* pada suatu model. Proses *preprocessing* dapat membantu model dalam melatih data, sehingga menghasilkan kinerja model menjadi lebih baik (Oliver, 2021).

Pada penelitian ini, dilakukan augmentasi pada data yang ada menggunakan modul OpenCV. Augmentasi data dilakukan untuk menghasilkan sampel data *training* yang baru dari data asli yang ada. Augmentasi yang dilakukan berupa rotasi gambar dari 1 sampai 180 derajat dengan beberapa ketentuan. Selain itu, dilakukan perubahan format data, *cropping*, dan perubahan dimensi yang semula nya 1024×1024 piksel, menjadi 224×224 piksel dan 256×256 piksel dengan menggunakan teknik *bilinear interpolation* yang sudah tersedia di *library* OpenCV. *Bilinear interpolation* atau interpolasi bilinear menentukan nilai piksel yang baru berdasarkan rata-rata nilai atau bobot empat piksel (2×2) terdekat dari piksel yang ada pada gambar aslinya (Singh et al., 2009).

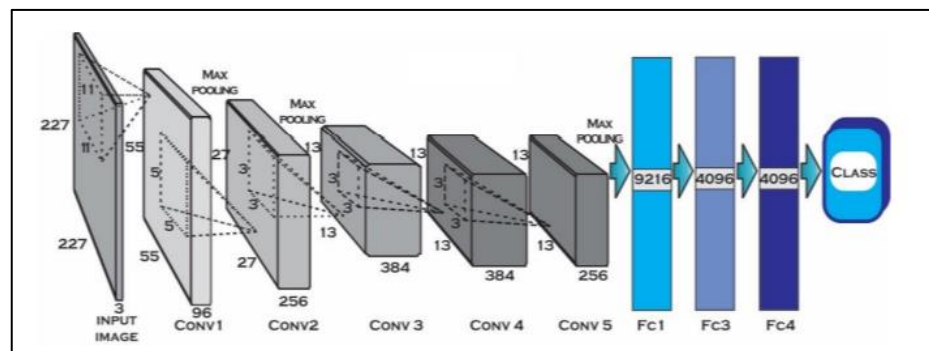
2.5 Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) adalah suatu algoritma *deep learning* yang dikembangkan dari *multilayer perceptron* (MLP) yang didesain untuk mengelola data dua dimensi dan tiga dimensi. CNN termasuk dalam jenis *deep neural network* karena kedalaman jaringan yang tinggi dan banyak diaplikasikan pada data citra (Kim, 2017). CNN tidak jauh berbeda dengan arsitektur *neural network*, yaitu memiliki *weight*, *bias*, dan *activation function*. Secara garis besar, *layer* pada CNN terdiri atas *convolution layer*, *activation ReLU layer*, *pooling layer*, dan *fully connected layer* (Hanin et al., 2021). Secara umum arsitektur *convolutional neural network* (CNN) dapat dilihat pada Gambar 4. Pada penelitian ini, ada dua arsitektur CNN yang akan digunakan, yaitu arsitektur GoogLeNet dan AlexNet.



Gambar 4. Arsitektur *Convolutional Neural Network* (CNN) (Matlab).

2.5.1. AlexNet



Gambar 5. Ilustrasi arsitektur AlexNet lima *layers* (Sabilla, 2020).

AlexNet merupakan sebuah arsitektur CNN yang terdiri dari model-model *convolution* (Sabilla, 2020). AlexNet memiliki *layers* yang tidak terlalu banyak. Secara ringkas, AlexNet dapat terdiri dari lima *convolution layers*, tiga *pooling layers*, dan dua *fully-connected layers*. Ilustrasi arsitektur AlexNet dapat dilihat pada Gambar 5 dan komposisi arsitektur AlexNet dapat dilihat pada Tabel 3.

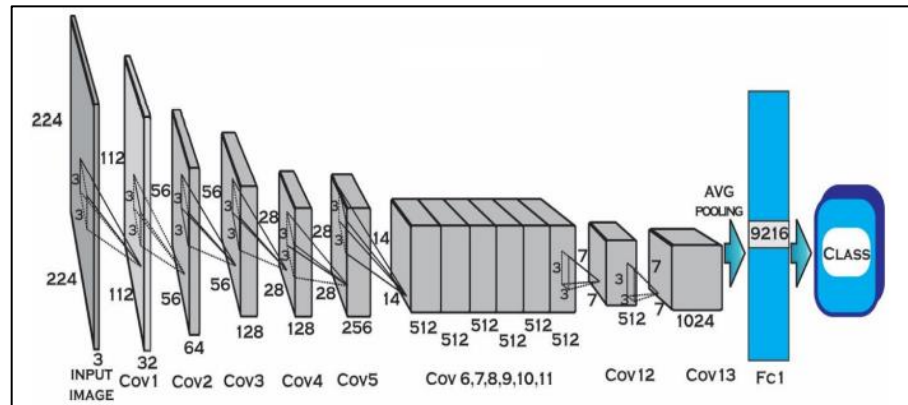
Tabel 3. Komposisi arsitektur AlexNet (Sabilla, 2020)

	<i>Layer</i>	<i>Output Size</i>	<i>Kernel Size</i>	<i>Stride</i>	<i>Activation</i>
<i>Input</i>	<i>Image</i>	227×227×3	-	-	-
1	<i>Convolution</i>	55×55×96	11×11	4	<i>Relu</i>
	<i>Max Pooling</i>	27×27×96	3×3	2	<i>Relu</i>
2	<i>Convolution</i>	27×27×256	5×5	1	<i>Relu</i>
	<i>Max Pooling</i>	13×13×256	3×3	2	<i>Relu</i>
3	<i>Convolution</i>	13×13×256	3×3	1	<i>Relu</i>
4	<i>Convolution</i>	13×13×384	3×3	1	<i>Relu</i>
5	<i>Convolution</i>	13×13×256	3×3	1	<i>Relu</i>
	<i>Max Pooling</i>	6×6×256	3×3	2	<i>Relu</i>
6	<i>FullyConnected</i>	9216	-	-	-
7	<i>FullyConnected</i>	4096	-	-	-
8	<i>FullyConnected</i>	4096	-	-	-
<i>Output</i>	<i>FullyConnected</i>	1000	-	-	<i>Softmax</i>

Tabel 3 memperlihatkan komposisi AlexNet secara ringkas dan umum. Pada penelitian ini, komposisi AlexNet diubah untuk menyesuaikan skenario penelitian yang ada. *Output size* pada *layer input* akan diubah menjadi 224×224 piksel menyesuaikan skenario dimensi *input* yang ada. Selain itu, *activation function* diubah menjadi *sigmoid* karena penelitian ini merupakan *binary classification* yang mana hanya terdapat dua *class* (abnormal dan

normal). *Sigmoid* digunakan untuk *binary classification* dan *softmax* digunakan untuk *multi-classification* (Basta, 2020).

2.5.2. GoogLeNet



Gambar 6. Ilustrasi arsitektur GoogLeNet 13 layers (Sabilla, 2020).

GoogLeNet merupakan sebuah arsitektur model CNN yang dikembangkan oleh orang-orang Google. Arsitektur ini dikembangkan dengan melakukan modifikasi pada arsitektur AlexNet (Guo et al., 2017). GoogLeNet terdiri dari 22 layers yang memiliki parameter (Jika *pooling* dihitung, maka berjumlah 27 layers) (Szegedy et al., 2015). Secara sederhana, GoogLeNet dapat terdiri dari lima hingga 13 layers. Waktu komputasi GoogLeNet akan lebih lama dibandingkan dengan AlexNet karena memiliki fitur *convolution* dan jumlah *layer* yang lebih banyak (Sabilla, 2020). Ilustrasi arsitektur GoogLeNet dapat dilihat pada Gambar 6 dan komposisi arsitektur GoogLeNet dapat dilihat pada Tabel 4.

Tabel 4. Komposisi arsitektur GoogLeNet (Sabilla, 2020)

	<i>Layer</i>	<i>Output Size</i>	<i>Kernel Size</i>	<i>Stride</i>
<i>Input</i>	<i>Image</i>	224×224×3	-	-
1	<i>Convolution</i>	112×112×64	3×3	2
	<i>Max Pooling</i>	56×56×64	3×3	2
2	<i>Convolution</i>	56×56×192	3×3	1

	<i>Max Pooling</i>	28×28×192	3×3	2
3	<i>Inception (3a)</i>	28×28×256	-	-
4	<i>Inception (3b)</i>	28×28×480	-	-
	<i>Max Pooling</i>	14×14×512	3×3	2
5	<i>Inception (4a)</i>	14×14×512	-	-
6	<i>Inception (4b)</i>	14×14×512	-	-
7	<i>Inception (4c)</i>	14×14×512	-	-
8	<i>Inception (4d)</i>	14×14×528	-	-
9	<i>Inception (4e)</i>	14×14×832	-	-
	<i>Max Pooling</i>	7×7×832	3×3	2
10	<i>Inception (5a)</i>	7×7×832	-	-
11	<i>Inception (5b)</i>	7×7×1024	-	-
	<i>Average Pool</i>	1×1×1024	7×7	1
	<i>Dropout 0,4</i>	1×1×1024	-	-
	<i>Linear</i>	1×1×1000	-	-
	<i>Softmax</i>	1×1×1000	-	-

Tabel 4 memperlihatkan komposisi GoogLeNet secara umum. Sama seperti AlexNet, komposisi GoogLeNet pada penelitian ini diubah untuk menyesuaikan skenario penelitian. *Output size* pada *layer input* akan diubah menjadi 256×256 piksel menyesuaikan skenario dimensi *input* yang ada. Selain itu, *activation function* juga diubah menjadi *sigmoid* seperti halnya AlexNet.

2.5.3. Convolutional Layer

Convolutional layer adalah *layer* pertama yang akan memproses citra yang dimasukkan, sehingga menjadikan *layer* ini sangat penting pada arsitektur *convolutional neural network* (CNN). *Layer* ini akan mendeteksi fitur-fitur yang dimiliki data gambar atau citra, seperti dua kuping, dua mata, dan lain-lain. Fitur-fitur pada data gambar tersebut didapatkan dengan cara melakukan *scanning* fitur pada gambar sesuai dengan *kernel* atau *filter size* yang telah ditentukan. Ukuran *kernel* atau *filter* sangat penting dalam CNN karena memengaruhi fitur-fitur

apa saja yang bisa didapatkan. Semakin besar ukuran *kernel* atau *filter*, semakin sulit fitur-fitur kecil bisa didapatkan, begitu juga sebaliknya. Oleh karena itu, pada CNN umumnya terdapat lebih dari satu *layer* konvolusi yang memiliki ukuran *kernel* yang berbeda-beda agar mengurangi jumlah fitur yang tidak terdeteksi. Hasil dari *convolutional layer* dinamakan *feature map* atau *activation map* (Hanin et al., 2021).

2.5.4. Activation Rectified Linear Units (ReLU) Layer

Activation ReLU layer merupakan *layer* yang berfungsi sebagai pengubah nilai piksel pada citra menjadi nol, jika terdapat nilai piksel yang kurang dari nol. Perubahan tersebut bertujuan untuk meningkatkan kualitas citra untuk meminimalisir kesalahan (Hanin et al., 2021).

2.5.5. Pooling Layer

Pooling layer adalah lapisan atau *layer* yang mengolah output dari *Activation ReLU layer* dengan berbagai macam operasi statistika berdasarkan nilai piksel terdekat. *Pooling layer* dapat mengendalikan *overfitting* dengan cara mengurangi ukuran *volume output* pada *feature map* secara progresif (Kim, 2017).

2.5.6. Fully Connected Layer

Fully connected layer adalah lapisan yang berisikan kumpulan dari proses *convolution* atau konvolusi (Hijazi et al., 2015). *Neuron-neuron* pada lapisan ini terhubung dengan semua *neuron* pada lapisan sebelumnya (Albelwi & Mahmood, 2017). Lapisan ini akan menerima *input* dari lapisan aktivasi sebelumnya untuk menentukan fitur yang paling berkorelasi dengan kelas tertentu berdasarkan *training dataset* yang ada (Hanin et al., 2021).

2.5.7. Softmax

Softmax activation berfungsi untuk mengubah *output* dari *layer* terakhir menjadi distribusi probabilitas dasarnya. Persamaan softmax activation dapat dituliskan seperti pada Persamaan (1).

$$f_i(\vec{x}) = \frac{e^{x_i}}{\sum_{j=1}^k e^{x_j}}, \text{ di mana } i = 1, 2, 3, \dots, k \dots \dots \dots (1)$$

Notasi f_i merupakan hasil fungsi untuk setiap elemen ke- i pada keluaran vektor kelas dan x adalah vektor yang berisi nilai dari *fully connected layer* terakhir. *Softmax* dapat menghitung probabilitas dari semua kelas, kemudian akan diambil nilai vektor yang bernilai real. Nilai vektor tersebut akan diubah menjadi nilai dengan rentang nol sampai satu yang jika semuanya dijumlahkan akan bernilai satu (Hanin et al., 2021).

2.6 Confusion Matrix

Confusion Matrix adalah suatu fungsi atau alat yang dapat menganalisis sebuah *classifier* termasuk ke dalam kategori baik atau buruk, dalam mengenali *tuple* dari kelas yang berbeda (Han & Kamber, 2011). Terdapat beberapa istilah yang ada pada *confusion matrix*, seperti yang dapat dilihat pada Tabel 5.

Tabel 5. *Confusion matrix* (Han & Kamber, 2011)

<i>Actual Class</i>	<i>Predicted Class</i>		Total
	<i>Yes</i>	<i>No</i>	
<i>Yes</i>	<i>TP</i>	<i>FN</i>	<i>P</i>
<i>No</i>	<i>FP</i>	<i>TN</i>	<i>N</i>
Total	<i>P'</i>	<i>N'</i>	<i>P + N</i>

Adapun istilah yang terdapat pada Tabel 3, sebagai berikut (Han & Kamber, 2011).

- TP (True Positive)*, yaitu jumlah data dengan nilai sebenarnya positif dan nilai prediksi positif.

- b. *FP (False Positive)*, yaitu jumlah data dengan nilai sebenarnya negatif dan nilai prediksi positif.
- c. *FN (False Negative)*, yaitu jumlah data dengan nilai sebenarnya positif dan nilai prediksi negatif.
- d. *TN (True Negative)*, yaitu jumlah data dengan nilai sebenarnya negatif dan nilai prediksi negatif.

Nilai dari *TP* dan *TN* memberikan informasi ketika *classifier* benar dalam melakukan klasifikasi data, sedangkan nilai *FP* dan *FN* memberikan informasi ketika *classifier* salah dalam melakukan klasifikasi data (Han & Kamber, 2011).

Berdasarkan penelitian sebelumnya, parameter *confusion matrix* yang digunakan untuk mengevaluasi kinerja model yang dibuat adalah *accuracy*, *recall*, *precision*, dan *f1 score* (Hanin et al., 2021).

2.6.1. Accuracy

Accuracy adalah matrik yang paling umum untuk melakukan evaluasi klasifikasi. *Accuracy* bekerja dengan cara menghitung nilai probabilitas berdasarkan nilai yang benar dari label kelas. *Accuracy* dinyatakan dengan Persamaan (2) (Bekkar et al., 2013).

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \dots\dots\dots(2)$$

2.6.2. Recall

Recall atau *sensitivity* adalah keakuratan dari nilai positif yang ada. *Recall* dirumuskan pada Persamaan (3) (Bekkar et al., 2013).

$$recall = \frac{TP}{TP+FN} \times 100\% \dots\dots\dots(3)$$

2.6.3. Precision

Precision adalah nilai kebenaran dari prediksi yang dilakukan oleh *classifier* dengan label kelas yang sudah ada. *Precision* dapat dituliskan seperti Persamaan (4) (Bekkar et al., 2013).

$$precision = \frac{TP}{TP+FP} \times 100\% \dots\dots\dots(4)$$

2.6.4. *F1 Score*

F1 score adalah nilai rata-rata harmonik dari *recall* dan *precision* (Powers, 2019). *F1 score* dinyatakan dengan Persamaan (5).

$$f1\ score = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \times 100\% \dots\dots\dots(5)$$

III. METODOLOGI PENELITIAN

3.1. Tempat dan Waktu Penelitian

Penelitian ini dilakukan pada tempat dan waktu yang dijelaskan sebagai berikut.

3.1.1. Tempat Penelitian

Penelitian dilaksanakan di Lab Rekayasa Perangkat Lunak FMIPA Universitas Lampung. Digunakan salah satu *personal computer* (PC) yang ada di Lab Rekayasa Perangkat Lunak (RPL), untuk menjalankan python dengan *graphics processing unit* (GPU) Nvidia Tesla K20.

3.1.2. Waktu Penelitian

Waktu atau rencana penelitian dapat dilihat pada Tabel 6.

Tabel 6. Rencana waktu penelitian

Kegiatan	2021																2022																			
	Desember				Januari				Februari				Maret				April				Mei				Juni				Juli				Agustus			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Pemahaman dan Pengumpulan Data	■	■																																		
<i>Praprocessing</i> Data			■	■																																
Pemodelan Data, Prediksi CNN, dan Evaluasi Kinerja					■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
a. Arsitektur GoogLeNet					■	■	■	■	■	■	■	■	■	■	■	■																				
• 80% <i>training</i> , 20% <i>testing</i>					■	■	■	■	■	■	■	■	■	■	■	■																				
• 70% <i>training</i> , 30% <i>testing</i>									■	■	■	■	■	■	■	■																				
b. Arsitektur AlexNet																	■	■	■	■	■	■	■	■												
• 80% <i>training</i> , 20% <i>testing</i>																	■	■	■	■	■	■	■	■												
• 70% <i>training</i> , 30% <i>testing</i>																					■	■	■	■	■	■	■	■								

Rencana penelitian yang ditunjukkan pada Tabel 6 dimulai pada Desember 2021 s.d. Juni 2022. Secara garis besar, rencana penelitian dibagi menjadi tiga tahapan, antara lain.

1. Pelaksanaan Penelitian

a. Pengumpulan atau Pengambilan Data

Data citra mammogram payudara wanita telah dikumpulkan oleh J Suckling pada tahun 1994 yang kemudian dipublikasikan.

Data penelitian dapat diakses melalui *link* <http://peipa.essex.ac.uk/info/mias.html>.

b. *Preprocessing* Data

Dilakukan *praprocessing* data berupa *cropping*, *resizing*, dan augmentasi data.

c. Pemodelan Data

Pada penelitian ini, dibuat model *convolutional neural network* (CNN) dengan arsitektur GoogLeNet dan AlexNet. Masing-masing arsitektur dibagi menjadi dua skenario berdasarkan pembagian data *training* dan *testing*, yaitu 80% *training*, 20% *testing* dan 70% *training* dan 30% *testing*.

2. Evaluasi Penelitian

Pada tahapan ini, dilakukan evaluasi kinerja model menggunakan *confusion matrix* dengan parameter *accuracy*, *recall*, *precision*, dan *f1 score*.

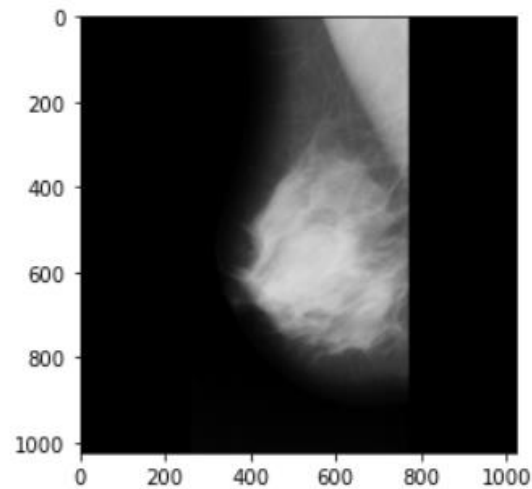
3.2. Data dan Alat

3.2.1. Data

Data yang digunakan bersumber dari alamat <http://peipa.essex.ac.uk/info/mias.html>. Data tersebut merupakan data citra mammogram payudara wanita yang diambil berdasarkan *right craniocaudal* (RCC) dan *left craniocaudal* (LCC). Data merupakan gambar tiga dimensi (3D) berukuran 1024 piksel \times 1024 piksel \times 3 (RGB) dan berjumlah sebanyak 322 data citra mammogram. Data dikumpulkan oleh J Suckling dan dipublikasikan pada tahun 1994 dengan judul *The Mammographic Image Analysis Society Digital*

Mammogram Database. Pada hakikatnya, data ini dibagi menjadi tujuh kelas. Namun pada penelitian ini, data hanya akan dibagi menjadi dua kelas, yaitu kelas normal dan abnormal dengan masing-masing sebanyak 207 data normal dan 115 data abnormal. Detail dari beberapa data yang digunakan, dapat dilihat pada Tabel 7.

(1024, 1024, 3)



Gambar 7. Ukuran dimensi data mammogram payudara dari MIAS.

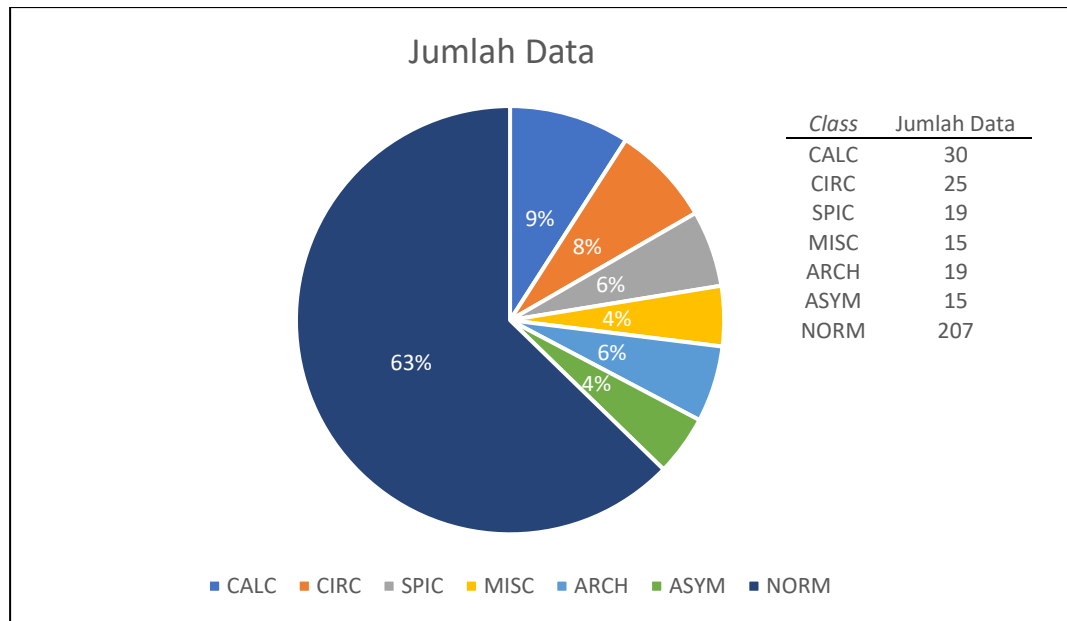
Tabel 7. Detail informasi dari beberapa data (J Suckling, 1994)

REFNUM	BG	CLASS	SEVERITY	X	Y	RADIUS
mdb001	G	CIRC	B	535	425	197
mdb002	G	CIRC	B	522	280	69
mdb003	D	NORM				
mdb004	D	NORM				
mdb005	F	CIRC	B	477	133	30
mdb005	F	CIRC	B	500	168	26
mdb006	F	NORM				
mdb007	G	NORM				
mdb008	G	NORM				
mdb009	F	NORM				
mdb010	F	CIRC	B	525	425	33

Keterangan dari nama kolom pada Tabel 7 adalah sebagai berikut.

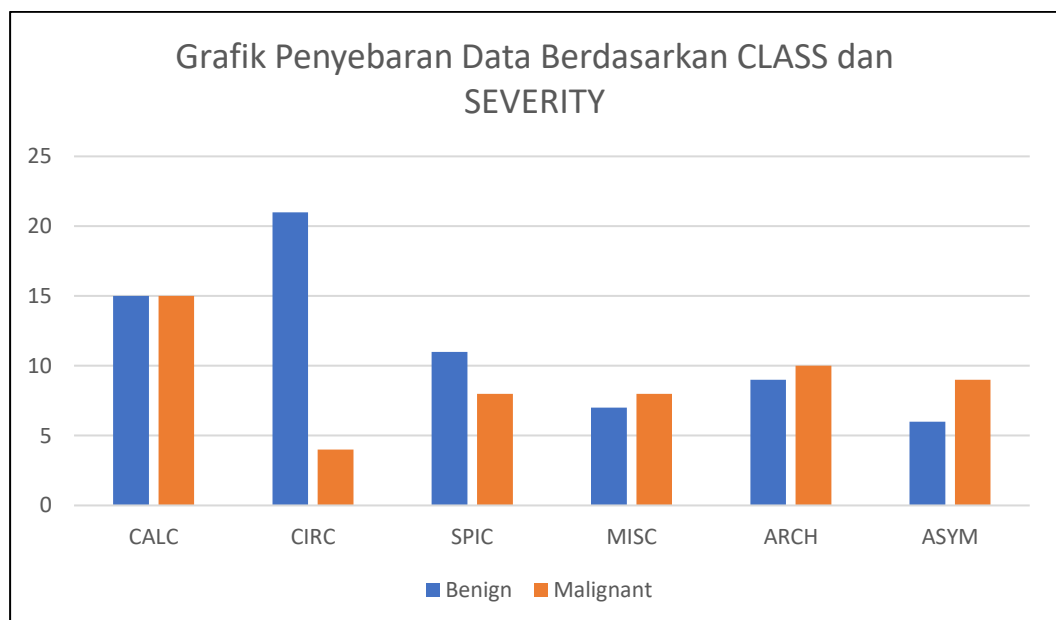
- a. REFNUM merupakan nomor referensi dari MIAS *database*.
- b. BG merupakan karakter dari jaringan dasar yang dibagi menjadi beberapa kategori, yaitu.
 - F : *Fatty*
 - G : *Fatty-glandular*
 - D : *Dense-glandular*
- c. CLASS merupakan keabnormalan yang ada. Keabnormalan dibagi menjadi beberapa kategori, yaitu.
 - CALC : *Calcification*
 - CIRC : *Well-defined / circumscribed masses*
 - SPIC : *Spiculated masses*
 - MISC : *Other, ill-defined masses*
 - ARCH : *Architectural distortion*
 - ASYM : *Asymmetry*
 - NORM : *Normal*
- d. SEVERITY merupakan tingkat keganasan dari keabnormalan. Tingkat keganasan dibagi menjadi beberapa kategori, yaitu.
 - B : *Benign*
 - M : *Malignant*
- e. X dan Y merupakan koordinat gambar dari pusat keabnormalan.
- f. RADIUS merupakan radius perkiraan (dalam piksel) dari lingkaran yang melingkupi keabnormalan.

Grafik penyebaran jumlah data dapat dilihat pada Gambar 8.



Gambar 8. Grafik penyebaran jumlah data.

Grafik penyebaran data berdasarkan CLASS dan SEVERITY-nya dapat dilihat pada Gambar 9.



Gambar 9. Grafik penyebaran data berdasarkan CLASS dan SEVERITY.

3.2.2. Alat

Alat yang digunakan pada penelitian ini, terdiri dari dua jenis, yaitu *hardware* dan *software*.

3.2.2.1. *Hardware*

Hardware yang digunakan, antara lain.

- a. *Processor*: Intel Core i5-9400F, 2.90 Ghz, *Cache* 9M, 6 *Cores*, 6 *Threads*.
- b. *Random Access Memory (RAM)*: 16 GB, DDR4, 2666 Mhz.
- c. *Storage*: SSD M.2 NVMe 256 GB.
- d. *Network Interface*: TP-Link Archer T2U Plus, 200 Mbps, 2.4 Ghz.
- e. *Video Graphics Array (VGA)*: Nvidia GTX 1660 Ti, 1536 Nvidia CUDA *Cores*, 1500 Mhz *Base Clock*, 6 GB GDDR6 VRAM *Memory*.

3.2.2.2. *Software*

Software yang digunakan, antara lain.

- a. *Sistem Operasi*: Windows 10 Pro, 64-bit, versi 21H1, OS *Build* 19043.1288.
- b. *Tools*:
 1. Python 3.8.8, 64-bit dan Jupyter Notebook 6.3.0, 64-bit.
 2. *Library*:
 - A. *Library* Opencv 4.5.3
OpenCV (*Open Source Computer Vision*) adalah sebuah *library* pada Python yang digunakan untuk mengolah gambar dan video untuk mendapatkan informasi yang ada di dalamnya (Anggri, 2017). OpenCV 4.5.3 dikembangkan oleh banyak orang, beberapa diantaranya adalah Alexander Alekhin, Giles Payne, dan Anatoliy Talamanov.

B. *Library* Numpy 1.19.5

NumPy (*Numerical Python*) adalah sebuah *library* pada Python yang biasa digunakan untuk membentuk objek n-dimensional *array* (Rohman, 2019). NumPy 1.19.5 dikembangkan oleh delapan orang, antara lain Charles Harris, Christoph Gohlke, Matti Picus, Reghuveer Devulapalli, Sebastian Berg, Simon Graham, Veniamin Petrenko, dan Bernie Gray.

C. *Library* Pandas 1.2.4

Pandas adalah sebuah *library* di Python yang biasa digunakan untuk membentuk struktur dan menganalisis data. Struktur data dasar pada pandas dinamakan *dataframe* (Rezkie, 2020). Pandas 1.2.4 dikembangkan oleh sembilan orang, yaitu Fangchen Li, Joris Van den Bossche, Matthew Zeitlin, MeeseeksMachine, Pandas Development Team, Simon Hawkins, jbrockmendel, mzeitlin11, dan patrick.

D. *Library* Tensorflow 2.5.0

Tensorflow adalah *library* pada Python yang berisikan API atau *tools* yang biasa digunakan untuk *machine learning* dan *artificial intelligence* (AI). Salah satu contoh API atau *tools* yang ada di tensorflow dan sering digunakan adalah keras. Tensorflow 2.5.0 dikembangkan oleh banyak orang, beberapa diantaranya adalah 8bitmp3, Aaron S. Mondal, dan Abhilash Mahendrakar.

E. *Library* Scikit-learn 0.24.1

Scikit-learn atau sklearn adalah *library* di Python yang berfungsi untuk membantu dalam melakukan *data processing* atau *training data* pada *machine*

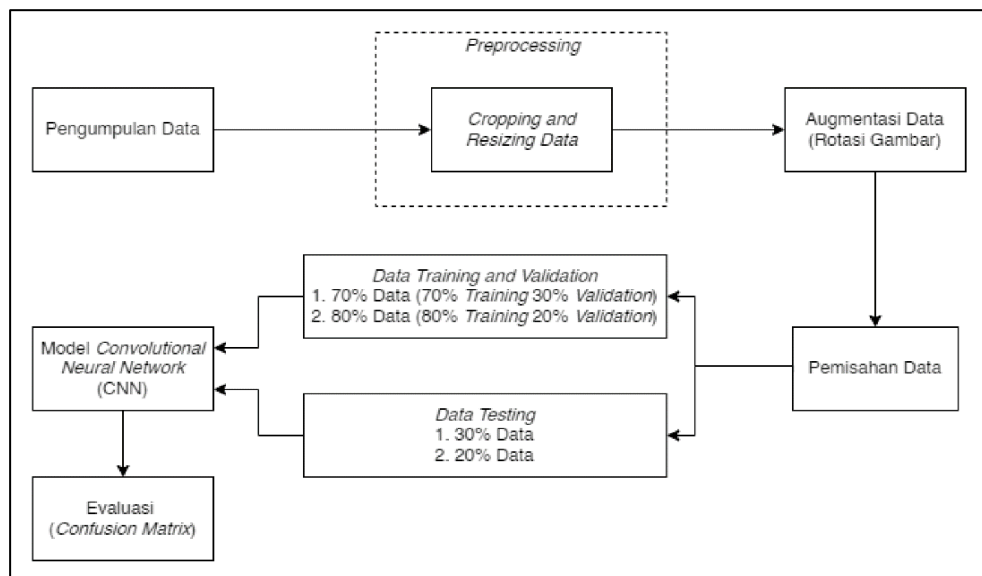
learning dan *deep learning* (Andhika, 2019). Scikit-learn 0.24.1 dikembangkan oleh banyak orang, beberapa diantaranya adalah Abo7atm, Adam Spannbauer, dan Adrin Jalali.

F. *Library* Matplotlib 3.3.4

Matplotlib adalah *library* pada Python yang biasa digunakan untuk memvisualisasikan data dalam bentuk plot atau grafik dua dimensi (2D) atau tiga dimensi (3D). Matplotlib 3.3.4 dikembangkan oleh banyak orang, beberapa diantaranya adalah Thomas A. Caswell, Michael Droettboom, dan Antony Lee.

3.3. Alur Kerja Penelitian

Alur penelitian ini didasari oleh penelitian yang dilakukan Susilo dan Sugiharti pada tahun 2021. Diagram alur penelitian dapat dilihat pada Gambar 10.



Gambar 10. Diagram alur kerja penelitian klasifikasi tumor payudara dengan citra mammogram menggunakan metode *Convolutional Neural Network* (CNN).

Secara rinci, diagram alur kerja penelitian pada Gambar 9 dapat dijelaskan sebagai berikut.

3.3.1. Pengumpulan Data

Pada tahap ini, dilakukan pengumpulan dan pengambilan data melalui link <http://peipa.essex.ac.uk/pix/mias/>. Data yang dikumpulkan merupakan data gambar *grayscale* citra mammogram payudara dengan format *pgm*. Data diunduh dan disimpan dalam satu folder.

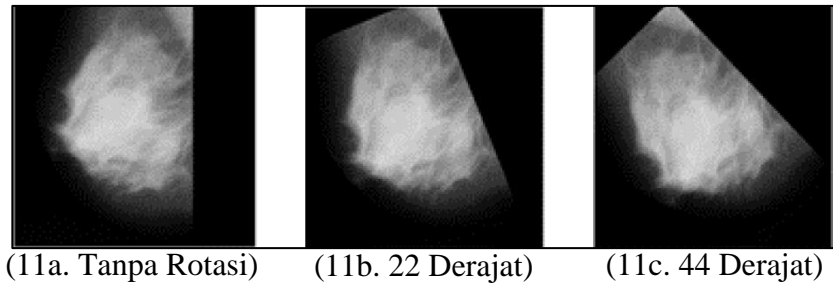
3.3.2. Data Preprocessing

Pada tahap *preprocessing*, ada dua hal utama yang akan dilakukan, yaitu *cropping* dan *resizing*. Sebelum dilakukan *cropping* dan *resizing*, data diubah formatnya menjadi *bmp (bitmap)*. *Cropping* gambar difokuskan pada bagian payudara yang terdapat tumornya. Hal ini bertujuan untuk lebih memfokuskan pelatihan model pada gambar payudara yang terdapat tumor untuk menghasilkan model yang lebih baik. *Resizing* gambar bertujuan untuk mengurangi waktu latih model. Data gambar yang semulanya berukuran 1024×1024 piksel akan diubah menjadi ukuran 224×224 piksel dan 256×256 piksel. Teknik interpolasi yang digunakan untuk *resizing* adalah *bilinear interpolation*.

3.3.3. Augmentasi Data

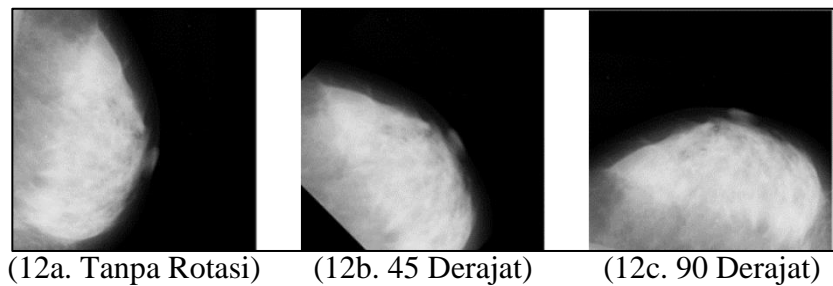
Augmentasi data dilakukan untuk memperbanyak data agar terhindar dari *overfitting* dan menghasilkan model yang lebih baik. Augmentasi data yang dilakukan, yaitu berupa rotasi gambar dengan menggunakan *library* OpenCV. Rotasi dilakukan pada data gambar dari nilai 1 sampai 180 derajat dengan beberapa ketentuan. Ketentuan rotasinya, antara lain.

- a. Data gambar dengan label abnormal dirotasi setiap kelipatan 22 derajat, sehingga jumlah data abnormal bertambah dari 115 data menjadi 1035 data. Contoh rotasi gambar data abnormal dapat dilihat pada Gambar 11.



Gambar 11. Beberapa contoh augmentasi pada data abnormal dengan cara rotasi.

- b. Data gambar dengan label normal dirotasi setiap kelipatan 45 derajat, sehingga jumlah data normal bertambah dari 207 data menjadi 1035 data. Contoh rotasi gambar data normal dapat dilihat pada Gambar 12.



Gambar 12. Beberapa contoh augmentasi pada data normal dengan cara rotasi.

Dengan augmentasi data, jumlah data yang semula berjumlah 322 data menjadi 2070 data.

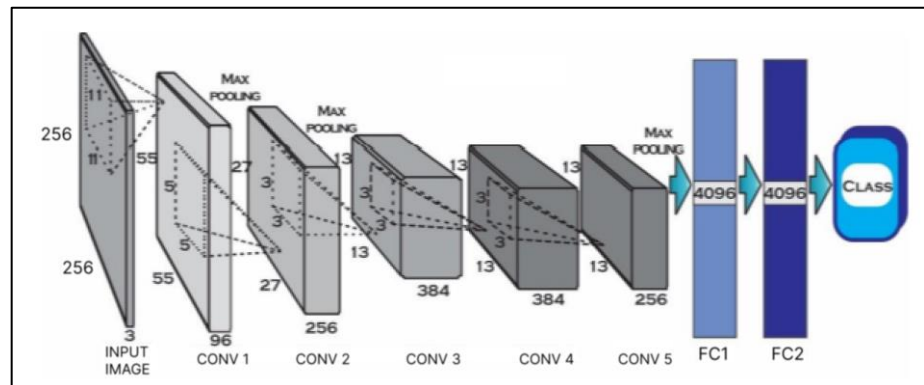
3.3.4. Pemisahan Data

Pemisahan data akan dilakukan menggunakan fungsi *train_test_split* dari *library* scikit-learn. Fungsi tersebut akan memilih data secara acak sesuai dengan persentase data *training* atau *testing* yang ditentukan. Terdapat dua skenario pembagian data yang akan dilakukan. Skenario pertama, data akan dibagi menjadi 70% *training* (dengan 30% digunakan sebagai *validation*) dan 30% *testing*. Maka pada skenario pertama, sebanyak 1449 data akan digunakan untuk *training* model dan 621 data akan digunakan untuk *testing* atau

klasifikasi menggunakan model yang sudah dilatih sebelumnya. Skenario kedua, data akan dibagi menjadi 80% *training* (dengan 20% digunakan sebagai *validation*) dan 20% *testing*. Maka pada skenario kedua, sebanyak 1656 data akan digunakan untuk *training* model dan 414 data akan digunakan untuk *testing* atau klasifikasi menggunakan model yang sudah dilatih sebelumnya.

3.3.5. Model

Pada penelitian ini akan dibuat model *Convolutional Neural Network* (CNN) dengan dua jenis arsitektur, yaitu arsitektur GoogLeNet dan AlexNet. Masing-masing model tersebut akan dicoba ke dalam masing-masing skenario pemisahan data dan dimensi input untuk mendapatkan model dengan akurasi yang paling baik. Ilustrasi dan komposisi AlexNet yang digunakan pada penelitian ini, dapat dilihat pada Gambar 13 dan Tabel 8.



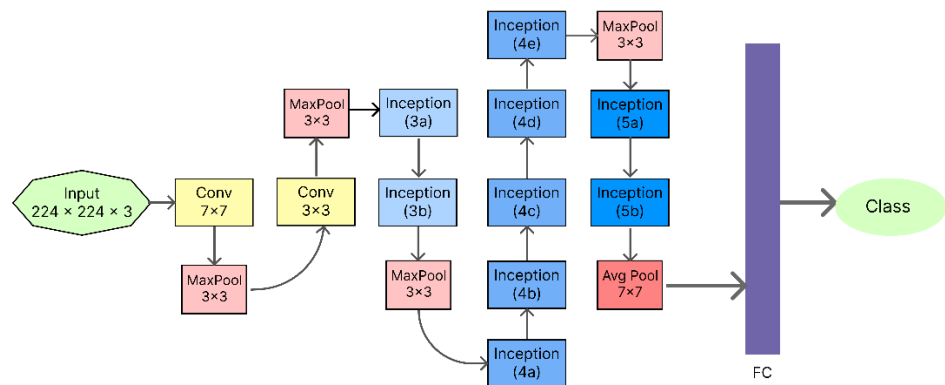
Gambar 13. Ilustrasi arsitektur AlexNet yang digunakan pada penelitian ini.

Tabel 8. Komposisi arsitektur AlexNet yang digunakan pada penelitian ini

	<i>Layer</i>	<i>Output Size</i>	<i>Kernel Size</i>	<i>Stride</i>	<i>Activation</i>
<i>Input</i>	<i>Image</i>	256×256×3	-	-	-
1	<i>Convolution</i>	55×55×96	11×11	4	<i>Relu</i>
	<i>Max Pooling</i>	27×27×96	3×3	2	<i>Relu</i>

2	<i>Convolution</i>	27×27×256	5×5	1	<i>Relu</i>
	<i>Max Pooling</i>	13×13×256	3×3	2	<i>Relu</i>
3	<i>Convolution</i>	13×13×256	3×3	1	<i>Relu</i>
4	<i>Convolution</i>	13×13×384	3×3	1	<i>Relu</i>
5	<i>Convolution</i>	13×13×256	3×3	1	<i>Relu</i>
	<i>Max Pooling</i>	6×6×256	3×3	2	<i>Relu</i>
6	<i>FullyConnected</i>	4096	-	-	-
7	<i>FullyConnected</i>	4096	-	-	-
	<i>Output FullyConnected</i>	1000	-	-	<i>Sigmoid</i>

Ilustrasi dan komposisi GoogLeNet yang digunakan pada penelitian ini, dapat dilihat pada Gambar 14 dan Tabel 9.



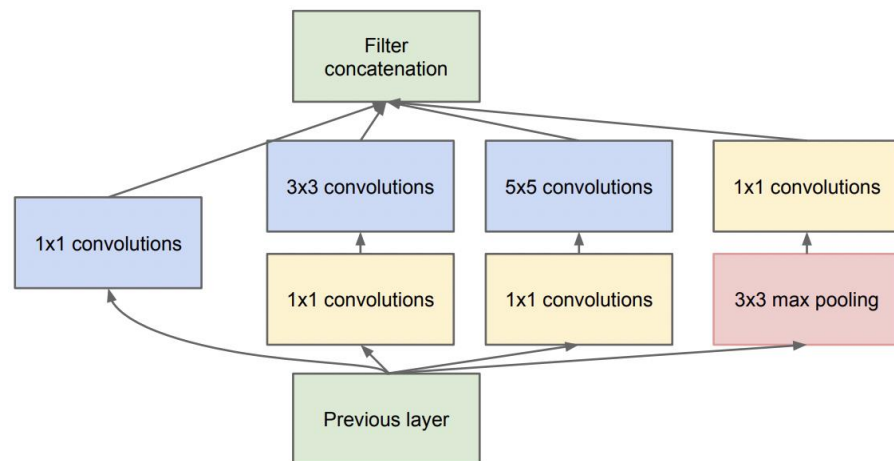
Gambar 14. Ilustrasi arsitektur GoogLeNet yang digunakan pada penelitian ini.

Tabel 9. Komposisi arsitektur GoogLeNet yang digunakan pada penelitian ini

	<i>Layer</i>	<i>Output Size</i>	<i>Kernel Size</i>	<i>Stride</i>
<i>Input</i>	<i>Image</i>	224×224×3	-	-
1	<i>Convolution</i>	112×112×64	3×3	2
	<i>Max Pooling</i>	56×56×64	3×3	2
2	<i>Convolution</i>	56×56×192	3×3	1
	<i>Max Pooling</i>	28×28×192	3×3	2
3	<i>Inception (3a)</i>	28×28×256	-	-

4	<i>Inception</i> (3b)	28×28×480	-	-
	<i>Max Pooling</i>	14×14×512	3×3	2
5	<i>Inception</i> (4a)	14×14×512	-	-
6	<i>Inception</i> (4b)	14×14×512	-	-
7	<i>Inception</i> (4c)	14×14×512	-	-
8	<i>Inception</i> (4d)	14×14×528	-	-
9	<i>Inception</i> (4e)	14×14×832	-	-
	<i>Max Pooling</i>	7×7×832	3×3	2
10	<i>Inception</i> (5a)	7×7×832	-	-
11	<i>Inception</i> (5b)	7×7×1024	-	-
	<i>Average Pool</i>	1×1×1024	7×7	1
	<i>Dropout</i> 0,4	1×1×1024	-	-
	<i>Linear</i>	1×1×1000	-	-
	<i>Sigmoid</i>	1×1×1000	-	-

Pada GoogLeNet terdapat *inception layer* atau *inception block* yang berisikan *convolution layer* yang bekerja secara bersamaan dan paralel. Ilustrasi dan komposisi dari *inception block* yang digunakan oleh GoogLeNet dapat dilihat pada Gambar 15.



Gambar 15. Ilustrasi dan komposisi *inception block* pada GoogLeNet yang dipakai pada penelitian ini (Szegedy et al., 2015).

Dimensi *input* pada AlexNet dan GoogLeNet yang aslinya adalah 256×256 piksel dan 224×224 piksel diubah untuk menyesuaikan skenario dimensi *input* yang ada, yaitu AlexNet dengan *input* 224×224 piksel dan GoogLeNet dengan *input* 256×256 piksel. Selain itu, *activation function* kedua model juga diubah menjadi *sigmoid*.

3.3.6. Evaluasi

Pada tahap ini, dilakukan evaluasi untuk mengetahui kinerja model menggunakan *confusion matrix*. Dalam melakukan evaluasi, terdapat beberapa parameter yang akan digunakan. Parameter yang digunakan, antara lain *accuracy*, *recall*, *precision*, dan *f1 score*.

V. PENUTUP

5.1. Kesimpulan

Dari penelitian dan pembahasan yang sudah dilakukan mengenai klasifikasi tumor payudara dengan citra *mammogram* menggunakan metode *convolutional neural network* (CNN) dapat diambil beberapa kesimpulan, antara lain.

1. *Convolutional neural network* (CNN) arsitektur AlexNet dan GoogLeNet dapat mendeteksi atau mengklasifikasi tumor payudara sebagai pendeteksian dini kanker payudara melalui gambar atau citra *mammogram*. Di antara kedua arsitektur tersebut, GoogLeNet memiliki akurasi yang lebih tinggi dibandingkan AlexNet pada semua skenario. Arsitektur GoogLeNet dengan pembagian data 80% *training* 20% *testing* dan dimensi input gambar 224×224 piksel memberikan hasil yang terbaik berupa *accuracy* 93,24%, *precision* 93,30%, *recall* 93,24%, dan *f1 score* 93,23%. Dalam mendapatkan hasil tersebut, terdapat beberapa tahapan dan skenario percobaan yang dilakukan, antara lain.
 - a. Data gambar citra *mammogram* payudara normal dan abnormal (terdapat tumor) yang dikumpulkan atau diambil dari MIAS (*Mammographic Image Analysis Society*) berjumlah 322 data gambar berdimensi 1024×1024 piksel yang terdiri dari 207 data citra payudara normal dan 115 data citra payudara abnormal.
 - b. Data yang diambil akan melalui tiga tahap *preprocessing* yaitu *cropping*, *resizing* dan augmentasi. *Cropping* atau pemotongan gambar difokuskan pada gambar payudara yang terdapat tumor, sehingga menghasilkan gambar berdimensi 700×700 . Kemudian, data di-*resize* dimensinya masing-masing menjadi 224×224 piksel dan

256×256 piksel. Yang terakhir, augmentasi data berupa rotasi, untuk memperbanyak data. Sehingga, jumlah data setelah diaugmentasi adalah 2070 data citra yang terdiri dari 1035 data citra payudara normal dan 1035 data citra payudara abnormal.

- c. Data diklasifikasi menggunakan dua arsitektur CNN, dua skenario pembagian data, dan dua dimensi input gambar. Arsitektur yang digunakan adalah AlexNet dan GoogLeNet. Data dibagi menjadi *training* dan *testing* dengan dua skenario pembagian yang berbeda, yaitu 70% *training* 30% *testing* dan 80% *training* 20% *testing*. Terdapat dua input dimensi gambar yang digunakan, yaitu 224×224 piksel dan 256×256 piksel.
2. Pada penelitian ini, terdapat beberapa percobaan klasifikasi yang dilakukan berdasarkan skenario pembagian dan dimensi *input* arsitektur. Hasil dari klasifikasi-klasifikasi tersebut, antara lain.
 - a. Arsitektur AlexNet dengan skenario 70% *training* 30% *testing* dan dimensi input 224×224 piksel menghasilkan *accuracy* 84,86%, *precision* 85,08%, *recall* 84,86%, dan *f1 score* 84,82%. Sedangkan dengan arsitektur dan skenario pembagian data yang sama, tetapi dengan dimensi input 256×256 piksel, menghasilkan *accuracy* 85,99%, *precision* 87,02%, *recall* 85,99%, dan *f1 score* 85,86%.
 - b. Arsitektur AlexNet dengan skenario 80% *training* 20% *testing* dan dimensi input 224×224 piksel menghasilkan *accuracy* 86,96%, *precision* 87,01%, *recall* 86,96%, dan *f1 score* 86,95%. Sedangkan dengan arsitektur dan skenario pembagian data yang sama, tetapi dengan dimensi input 256×256 piksel, menghasilkan *accuracy* 82,85%, *precision* 83,07%, *recall* 82,85%, dan *f1 score* 82,82%.
 - c. Arsitektur GoogLeNet dengan skenario 70% *training* 30% *testing* dan dimensi input 224×224 piksel menghasilkan *accuracy* 91,14%, *precision* 91,23%, *recall* 91,14%, dan *f1 score* 91,13%. Sedangkan dengan arsitektur dan skenario pembagian data yang sama, tetapi dengan dimensi input 256×256 piksel, menghasilkan *accuracy* 91,30%, *precision* 91,73%, *recall* 91,30%, dan *f1 score* 91,27%.

- d. Arsitektur GoogLeNet dengan skenario 80% *training* 20% *testing* dan dimensi input 224×224 piksel menghasilkan *accuracy* 93,24%, *precision* 93,30%, *recall* 93,24%, dan *f1 score* 93,23%. Sedangkan dengan arsitektur dan skenario pembagian data yang sama, tetapi dengan dimensi input 256×256 piksel, menghasilkan *accuracy* 90,82%, *precision* 91,38%, *recall* 90,82%, dan *f1 score* 90,79%.

5.2. Saran

Saran pada penelitian ini, antara lain.

1. Penelitian ini dapat dilakukan menggunakan arsitektur atau model CNN lain untuk mendapatkan hasil klasifikasi pembandingan dengan arsitektur yang digunakan pada penelitian ini. Beberapa arsitektur atau model lain yang dapat dicoba, antara lain MobileNet yang didesain untuk memaksimalkan akurasi secara efektif dengan sumber daya perangkat keras yang terbatas, LeNet yang memiliki arsitektur sederhana dan parameter yang sedikit, tetapi dapat menghasilkan hasil akurasi yang baik, dan arsitektur lainnya.
2. Penelitian ini dapat dilakukan menggunakan metode klasifikasi yang baru, selain CNN, untuk mendapatkan hasil klasifikasi pembandingan dengan metode CNN yang digunakan pada penelitian ini.
3. Penelitian ini dapat dilakukan dengan jenis *preprocessing*, augmentasi data, dan atau teknik penyeimbangan data lainnya agar mendapatkan hasil klasifikasi pembandingan dengan teknik yang digunakan pada penelitian ini. Selain itu, dapat ditambahkan *hyperparameter tuning* untuk membandingkan hasil klasifikasinya dengan penelitian ini.
4. Dapat dilakukan klasifikasi lanjutan untuk mengetahui citra *mammogram* payudara abnormal masuk ke dalam kategori atau *severity benign* (jinak) atau *malignant* (ganas).

DAFTAR PUSTAKA

- Adiwijaya, A. 2018. Deteksi Kanker Berdasarkan Klasifikasi Microarray Data. *Jurnal Media Informatika Budidarma*, 2(4), 181.
- Albelwi, S., & Mahmood, A. 2017. A framework for designing the architectures of deep Convolutional Neural Networks. *Entropy*, 19(6).
- Andhika, W. 2019. Belajar machine-learning, basic of scikit-learn. Medium. <https://medium.com/@wahyuandhika/belajar-machine-learning-basic-of-scikit-learn-a1685db819a8>. Diakses pada 25 Oktober 2021.
- Basta, N. 2020. The Differences between Sigmoid and Softmax Activation Functions. Medium. <https://medium.com/arteos-ai/the-differences-between-sigmoid-and-softmax-activation-function-12adee8cf322>. Diakses pada 25 Juni 2022.
- Bekkar, M., Djemaa, H. K., & Alitouche, T. A. 2013. Evaluation Measures for Models Assessment over Imbalanced Data Sets. *Journal of Information Engineering and Applications*, 3(10), 27–38.
- Globocan. 2020. The Global Cancer Observatory - All cancers. *International Agency for Research on Cancer - WHO*, 419, 199–200.
- Globocan. 2020. The Global Cancer Observatory - Breast cancers. *International Agency for Research on Cancer - WHO*, 631, 50–51.
- Globocan. 2020. The Global Cancer Observatory - Cancer Incident in Indonesia. *International Agency for Research on Cancer - WHO*, 858, 1–2.
- Guo, Z., Chen, Q., Wu, G., Xu, Y., Shibasaki, R., & Shao, X. 2017. Village building identification based on Ensemble Convolutional Neural Networks. *Sensors*

(Switzerland), 17(11), 1–22.

- Han, J., & M. Kamber. 2011. *Data Mining: Concepts and Techniques Tutorial*. San Fransisco: Morgan Kaufman Publisher.
- Hanin, M. A., Patmasari, R., & Nur, R. Y. 2021. Sistem Klasifikasi Penyakit Kulit Menggunakan Convolutional Neural Network (Cnn) Skin Disease Classification System Using Convolutional Neural Network (Cnn). *E-Proceeding of Engineering*, 8(1), 273–281.
- Hartono. 2010. Aspek Genetika Glaukoma Sudut Terbuka Kronis. *Indonesia Ophthalmology Meeting on World Glaucoma Day 2010*, 11-16.
- Helja, M., Nurhasanah, & Sampurno, J. 2013. Analisis Fraktal Citra Mammogram Berbasis Tekstur Sebagai Pendukung Diagnosis Kanker Payudara. *Positron*, 3(2), 35–39.
- Hijazi, S., Kumar, R., & Rowen, C. 2015. *Using Convolutional Neural Networks for Image Recognition*. California: Cadence Design Systems Inc.
- Kumar, S. 2020. Understanding 8 types of Cross-Validation. Towards Data Science. <https://towardsdatascience.com/understanding-8-types-of-cross-validation-80c935a4976d>. Diakses pada 05 Maret 2022.
- Nareza, M. 2021. Diagnosis Kanker Payudara. Alodokter. <https://www.alodokter.com/kanker-payudara/diagnosis>. Diakses pada 11 Oktober 2021.
- Setiawan, K. N., & Putra, I. M. S. 2018. Klasifikasi Citra Mammogram Menggunakan Metode K-Means, GLCM, dan Support Vector Machine (SVM). *Jurnal Ilmiah Merpati (Menara Penelitian Akademika Teknologi Informasi)*, 6(1), 13.
- Novianti, F. A., & Purnami, S. W. 2012. Analisis Diagnosis Pasien Kanker Payudara Menggunakan Regresi Logistik dan Support Vector Machine (SVM) Berdasarkan Hasil Mamografi. *Jurnal SAINS Dan Seni ITS*, 1(1), D147–D152.
- Nugraha, F. S., Shidiq, M. J., & Rahayu, S. 2019. Analisis Algoritma Klasifikasi

- Neural Network Untuk Diagnosis Penyakit Kanker Payudara. *Jurnal Pilar Nusa Mandiri*, 15(2), 149–156.
- Oliver, A. 2021. Bikin Data Lebih Mudah Dibaca, yuk, Kenalan dengan Data Preprocessing. Glints. <https://glints.com/id/lowongan/data-preprocessing-adalah/>. Diakses pada 17 Januari 2022.
- P, Anggri Yulio. 2017. Mengenal OpenCV (Open Source Computer Vision Library). DEVTRIK. <https://devtrik.com/opencv/mengenal-opencv-open-source-computer-vision-library/>. Diakses pada 25 Oktober 2021.
- P, Kim. 2017. *MATLAB Deep Learning: With Machine Learning, Neural Networks, and Artificial Intelligence*. New York: Apress.
- Panigoro, S., Hernowo, B. S., & Purwanto, H. 2019. Panduan Penatalaksanaan Kanker Payudara (Breast Cancer Treatment Guideline). *Jurnal Kesehatan Masyarakat*, 4(4), 1–50.
- Powers, D. M. W. 2019. What the F-measure doesn't measure *ArXiv*.
- Puspitasari, A. M., Ratnawati, D. E., & Widodo, A. W. 2018. Klasifikasi Penyakit Gigi Dan Mulut Menggunakan Metode Support Vector Machine. *J-Ptiik*, 2(2), 802–810.
- Ramli, H. M, Umbas, R., & Panigoro, S. S. 2005. *Deteksi Dini Kanker*. Jakarta: Balai Penerbit Fakultas Kedokteran Indonesia.
- Rezkia, S. M. 2020. Belajar Pandas untuk Tingkatkan Kompetensi Bahasa Pemrograman Pythonmu!. DQLab. Annissa Widya Davita. <https://www.dqlab.id/belajar-pandas-untuk-tingkatkan-kompetensi-python>. Diakses pada 25 Oktober 2021.
- Rohman, Y. A. 2019. Pengenalan NumPy, Pandas, Matplotlib. Medium. <https://medium.com/@yasirabd/pengenalan-numpy-pandas-matplotlib-b90bafd36c0>. Diakses pada 25 Oktober 2021.
- Sabilla, I. A. 2020. Arsitektur Convolutional Neural Network (Cnn) Untuk Klasifikasi Jenis Dan Kesegaran Buah Pada Neraca Buah. *Tesis*,

201510370311144, 1–119.

- Santika, D. D., & Angries, N. 2013. MAMMOGRAMS CLASSIFICATION USING FEATURES VECTOR AND NAIVE BAYES CLASSIFIER. *Jurnal Ilmiah Ilmu Komputer FIK - UPH*, 9, 1–9.
- Sidabutar, S. 2012. *Kontrol Kualitas Sistem Mammografi Digital: Direct Radiography (DR)*. (Skripsi). Universitas Indonesia. Depok.
- Singh, T. R., Singh, O. I., Singh, K. M., Singh, T. R., & Sinam, T. 2009. *Image Magnification based on Directed Linear Interpolation*. 367–371.
- Suckling, J. 1994. The Mammographic Image Analysis Society Digital Mammogram Database. *Experta Medica International Congress*, 1069, 375-378.
- Susilo, A. B., & Sugiharti, E. 2021. Accuracy Enhancement in Early Detection of Breast Cancer on Mammogram Images with Convolutional Neural Network (CNN) Methods using Data Augmentation and Transfer Learning. *Journal of Advances in Information Systems and Technology*. 3(April), 9–16.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. 2015. Going deeper with convolutions. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 07-12-June-2015*, 1–9.
- Tambunan, G. W. 1991. *Diagnosis dan Tatalaksana Sepuluh Jenis Kanker Terbanyak di Indonesia*. Jakarta: EGC.
- Tapan, E. 2005. *Kanker, Antioksidan, & Terapi Komplementer*. Jakarta: PT Elex Media Komputindo.
- Willy, T. 2019. Kanker. Alodokter. <https://www.alodokter.com/penyakit-kanker>. Diakses pada 24 September 2021.