

**KLASIFIKASI PENYAKIT KANKER PAYUDARA
MENGUNAKAN ADAM OPTIMIZER UNTUK DEEP
LEARNING**



TESIS

Diajukan sebagai salah satu syarat untuk memperoleh gelar
Magister Ilmu Komputer (M.Kom)

IRMAWATI

14002419

PROGRAM PASCASARJANA MAGISTER ILMU KOMPUTER

NUSA MANDIRI

2022

SURAT PERNYATAAN ORISINALITAS DAN BEBAS PLAGIARISME

Yang bertanda tangan di bawah ini :

Nama : Irmawati
NIM : 14002419
Program Studi : Strata Dua (S2)
Konsentrasi : Data Mining

Dengan ini menyatakan bahwa tesis yang telah saya buat dengan judul: “KLASIFIKASI PENYAKIT KANKER PAYUDARA MENGGUNAKAN ADAM OPTIMIZER UNTUK DEEP LEARNING” adalah hasil karya sendiri, dan semua sumber baik yang kutip maupun yang dirujuk telah saya nyatakan dengan benar dan tesis belum pernah diterbitkan atau dipublikasikan dimanapun dan dalam bentuk apapun.

Demikianlah surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila dikemudian hari ternyata saya memberikan keterangan palsu dan atau ada pihak lain yang mengklaim bahwa tesis yang telah saya buat adalah hasil karya milik seseorang atau badan tertentu, saya bersedia diproses baik secara pidana maupun perdata dan kelulusan saya dari Program Studi Ilmu Komputer (S2) Universitas Nusa Mandiri dicabut/dibatalkan.

Jakarta, 12 Januari 2022
Yang menyatakan,

Irmawati

PERSETUJUAN TESIS

Tesis ini diajukan oleh :


Nama : Irmawati
NIM : 14002419
Program Studi : Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : Data Mining
Judul Tesis : KLASIFIKASI PENYAKIT KANKER PAYUDARA
MENGUNAKAN ADAM OPTIMIZER UNTUK DEEP
LEARNING

Untuk dipertahankan pada periode 2022-2 dihadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Magister Ilmu Komputer (M.Kom) pada Program Studi Ilmu Komputer (S2) Universitas Nusa Mandiri

Jakarta, 12 Januari 2022

PEMBIMBING TESIS

Pembimbing I :
Prof. Dr. Dwiza Riana, S.Si, MM, M.Kom 

Pembimbing II :
Dr. Ferda Ernawan, M.Cs. Ph.D 


DEWAN PENGUJI

Penguji I :

Penguji II :

Penguji III/
Pembimbing :

Program Studi Ilmu Komputer (S2) Universitas Nusa Mandiri

	LEMBAR BIMBINGAN TESIS
	UNIVERSITAS NUSA MANDIRI

NIM : 14002419
 Nama Lengkap : Irmawati
 Dosen Pembimbing I : Prof. Dr. Dwiza Riana, S.Si, MM, M.Kom
 Judul Tesis : Klasifikasi Penyakit Kanker Payudara Menggunakan Adam Optimizer Untuk Deep Learning

No	Tanggal Bimbingan	Pokok Bahasan	Paraf Dosen Pembimbing
1	18 September 2021	Pengajuan tema dan <i>literature review</i>	
2	19 September 2021	Laporan Pengambilan <i>dataset</i>	
3	16 Oktober 2021	Laporan Progress Eksperimen	
4	23 Oktober 2021	Laporan Progress Eksperimen	
5	9 November 2021	Pengajuan BAB III dan BAB IV	
6	12 November 2021	Pengajuan Revisi BAB III dan BAB IV	
7	29 November 2021	Pengajuan BAB I, BAB II dan BAB V	
8	11 Desember 2021	Acc BAB III dan BAB IV	
9	18 Desember 2021	Pengajuan Revisi BAB I, BAB II	
10	5 Januari 2022	Pengajuan Keseluruhan dan Abstrak	
11	6 Januari 2022	Acc Keseluruhan dan Abstrak	
12	21 Januari 2022	Presentasi	

Catatan untuk Dosen Pembimbing.

Bimbingan Tesis

- Dimulai pada tanggal : 18 September 2021
- Diakhiri pada tanggal : 21 Januari 2022
- Jumlah pertemuan bimbingan : 12 (Dua Belas) kali bimbingan
-

Disetujui oleh,
Dosen Pembimbing



Prof. Dr. Dwiza Riana, S.Si, MM, M.Kom

Program Studi Ilmu Komputer (S2) Universitas Nusa Mandiri

	LEMBAR KONSULTASI BIMBINGAN TESIS
	UNIVERSITAS NUSA MANDIRI

NIM : 14002419
 Nama Lengkap : Irmawati
 Dosen Pembimbing II: Dr. Ferda Ernawan, M.Cs. Ph.D
 Judul Tesis : Klasifikasi Penyakit Kanker Payudara Menggunakan Adam
Optimization Untuk Deep Learning

No	Tanggal Bimbingan	Pokok Bahasan	Paraf Dosen Pembimbing
1	18 September 2021	Pengajuan tema dan <i>literature review</i>	
2	19 September 2021	Laporan Pengambilan <i>dataset</i>	
3	16 Oktober 2021	Laporan Progress Eksperimen	
4	23 Oktober 2021	Laporan Progress Eksperimen	
5	9 November 2021	Pengajuan BAB III dan BAB IV	
6	12 November 2021	Pengajuan Revisi BAB III dan BAB IV	
7	29 November 2021	Pengajuan BAB I, BAB II dan BAB V	
8	11 Desember 2021	Acc BAB III dan BAB IV	
9	18 Desember 2021	Pengajuan Revisi BAB I, BAB II	
10	5 Januari 2022	Pengajuan Keseluruhan dan Abstrak	
11	6 Januari 2022	Acc Keseluruhan dan Abstrak	

Bimbingan Tesis:

- Dimulai pada tanggal : 18 September 2021
- Diakhiri pada tanggal : 6 Januari 2022
- Jumlah bimbingan : 11 Kali

Disetujui oleh,

Dosen Pembimbing



(Ferda Ernawan, M.Cs, Ph.D)

KATA PENGANTAR

Puji syukur Alhamdulillah, penulis panjatkan kehadirat Allah SWT, yang telah melimpahkan rahmat dan karunia-Nya sehingga pada akhirnya penulis dapat menyelesaikan tesis ini tepat pada waktunya. Dimana laporan tesis ini penulis sajikan dalam bentuk buku yang sederhana

Penulis menyadari bahwa tanpa bimbingan dan dukungan dari semua pihak dalam pembuatan laporan tesis ini, maka penulis tidak dapat menyelesaikan laporan tesis ini tepat pada waktunya. Untuk itu ijinilah penulis pada kesempatan ini untuk mengucapkan terima kasih kepada:

1. Ketua Universitas Nusa Mandiri, Ibu Prof. Dr. Dwiza Riana, S.Si, MM, M.Kom
2. Ketua Program Studi Ilmu Komputer
3. Pembimbing Tesis, Ibu Prof. Dr. Dwiza Riana, S.Si, MM, M.Kom dan Pak Dr. Ferda Ernawan, M.Cs. Ph.D
4. Orang Tua dan segenap keluarga besar yang selalu mendukung
5. Teman-teman tercinta yang saling menyemangati
6. Rekan-rekan kelas 14.4A.01 Universitas Nusa Mandiri

Serta semua pihak yang tidak dapat penulis sebutkan satu persatu, sehingga atas bantuan semua pihak tersebut penulis dapat menyelesaikan penulisan tesis ini. Penulis menyadari bahwa dalam penulisan tesis ini masih terdapat banyak kekurangan dan kekeliruan, untuk itu penulis mengharapkan kritik dan saran membangun agar penulisan karya ilmiah yang penulis hasilkan di masa yang akan datang menjadi semakin lebih baik.

Akhir kata semoga laporan tesis ini dapat bermanfaat bagi penulis khususnya dan bagi para pembaca yang berminat pada umumnya.

Jakarta, 12 Januari 2022

Penulis,

Irmawati

Program Studi Ilmu Komputer (S2) Universitas Nusa Mandiri

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan di bawah ini, saya :

Nama : Irmawati
NIM : 14002419
Program Studi : Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : Data Mining
Jenis Karya : Tesis

Demi pengembangan ilmu pengetahuan, dengan ini menyetujui untuk memberikan ijin kepada pihak Program Studi Ilmu Komputer (S2) Universitas Nusa Mandiri **Hak Bebas Royalti Non-Eksklusif (*Non-exclusive Royalti-Free Right*)** atas karya ilmiah kami yang berjudul : “KLASIFIKASI PENYAKIT KANKER PAYUDARA MENGGUNAKAN ADAM OPTIMIZER UNTUK DEEP LEARNING” beserta perangkat yang diperlukan (apabila ada).

Dengan **Hak Bebas Royalti Non-Eksklusif** ini pihak Universitas Nusa Mandiri berhak menyimpan, mengalih-media atau *bentuk*-kan, mengelolanya dalam pangkalan data (*database*), mendistribusikannya dan menampilkan atau mempublikasikannya di *internet* atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari kami selama tetap mencantumkan nama kami sebagai penulis/pencipta karya ilmiah tersebut.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Universitas Nusa Mandiri, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 12 Januari 2022
Yang Menyatakan

Irmawati

Program Studi Ilmu Komputer (S2) Universitas Nusa Mandiri

ABSTRAK

Nama : Irmawati
NIM : 14002419
Program Studi : Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : Data Mining
Judul : KLASIFIKASI PENYAKIT KANKER PAYUDARA
MENGUNAKAN ADAM OPTIMIZER UNTUK DEEP
LEARNING

Kanker payudara merupakan penyebab kematian kedua pada pasien wanita di dunia. Kanker payudara telah menyebabkan kematian di lebih dari 100 negara. Diagnosis dini pasien kanker payudara menjadi penting guna mengurangi kemungkinan kematian. Peneliti fokus pada deteksi akurat kanker payudara, metode diagnostik otomatis dan diagnosis kanker payudara. Makalah ini mengusulkan optimasi Adam untuk Deep Learning Algorithm untuk mengklasifikasikan deteksi kanker payudara. Penelitian ini bertujuan untuk mengatasi masalah ketidakstabilan data dan overfitting, serta memperbarui bobot jaringan pada data latih deep learning. Dalam penelitian ini, penulis melakukan eksperimen dengan kombinasi tiga lapisan tersembunyi dan kecepatan pembelajaran untuk meningkatkan akurasi klasifikasi. Eksperimen menggunakan kumpulan data kanker payudara dari Studi UCI: kumpulan data WBCD (Asli), kumpulan data WDBC (Diagnostik) dan kumpulan data Coimbra. Selanjutnya, akurasi skema yang diusulkan dibandingkan dengan tolok ukur deteksi kanker payudara yang ada. Hasil eksperimen menunjukkan bahwa skema yang diusulkan mencapai akurasi 96,3% untuk mengklasifikasikan kanker payudara.

Kata Kunci: Klasifikasi, Kanker Payudara, *Deep Learning*

ABSTRACT

Name : Irmawati
NIM : 14002419
Study of Program : Computer Science
Levels : Strata Dua (S2)
Concentration : Data Mining
Title : **CLASSIFICATION OF BREAST CANCER USING ADAM OPTIMIZER FOR DEEP LEARNING**

Breast cancer is the second leading cause of death among female patient in the worldwide. Breast cancer has caused death in more than 100 countries. Early diagnosis of breast cancer patients become an important in order to reduce the chances of death. Researchers focus on an accurate detection of breast cancer, automatic diagnostic method and diagnose breast cancer. This paper proposes an Adam OPTIMIZER untuk Deep Learning Algorithm for classifying breast cancer detection. This study aims to overcome data instability and overfitting problems, and to update network weights in training data of deep learning. In this study, the authors conducted experiments with a combination of three-hidden layers and learning rates to increase the accuracy of the classification. The experiments used datasets of breast cancer data from the UCI Study: the WBCD dataset (Original), the WDBC dataset (Diagnostics) and the Coimbra dataset. Furthermore, the accuracy of the proposed scheme is compared with the existing benchmark of breast cancer detection. The experimental results show that the proposed scheme achieves an accuracy of 96.3% for classifying breast cancer.

Keywords: Classification, Breast Cancer, Deep Learning, Adam OPTIMIZER

DAFTAR ISI

	Halaman
Lembar Judul	i
Lembar Pernyataan Orisinalitas dan Bebas Plagiarisme	ii
Lembar Persetujuan Tesis	iii
Lembar Bimbingan Thesis	iv
Kata Pengantar	vi
Abstrak	vii
Daftar Isi	x
Daftar Gambar	xi
Daftar Tabel	
BAB I PENDAHULUAN	1
1.1. Latar Belakang Penulisan.....	1
1.2. Identifikasi	4
1.3. Metode Penelitian	5
1.4. Ruang Lingkup Penelitian.....	5
1.5. Kontribusi Tesis	5
1.6. Sistematika Penulisan.....	5
BAB II LANDASAN/KERANGKA PEMIKIRAN	8
2.1. Tinjauan Studi	8
2.2. Teori Tinjauan Pustaka.....	20
BAB III METODOLOGI PENELITIAN	26
3.1. Deskripsi Dataset dan Masalah	26
3.2. Eksperimen Yang Dilakukan	28
BAB IV HASIL PENELITIAN DAN PEMBAHASAN	35
4.1. Hasil Eksplorasi Data	35
4.2. Hasil Data Preprosesing	37
4.3. Hasil Split Data	41
4.4. Normalisasi Data	41
4.5. Hasil Arsitektur Model	41
4.6. Hasil Training dan Validasi Model	46
4.7. Hasil Evaluasi atau Pengujian Model	49
4.8. Model Prediktif yang Diusulkan	52
4.9. Perbandingan	52
BAB IV PENUTUP	56
4.1. Kesimpulan	56
4.2. Saran	56
DAFTAR PUSTAKA	56
DAFTAR RIWAYAT HIDUP	57

Program Studi Ilmu Komputer (S2) Universitas Nusa Mandiri

DAFTAR GAMBAR

	Halaman
Gambar 1.1. Berbagai Kasus Kanker di Indonesia	2
Gambar 2.2. Arsitektur Neural Network.....	13
Gambar 2.2. <i>Artificial Neuron</i>	15
Gambar 3.1. Tahapan Penelitian	28
Gambar 3.2. Distribusi Kelas Target Dataset WBCD	37
Gambar 3.3. Distribusi Kelas Target Dataset WDBC	38
Gambar 3.4. Distribusi Kelas Target Dataset Coimbra	38
Gambar 4.1. Perbandingan Hasil Akurasi.....	54
Gambar 4.2. Perbandingan Hasil Presisi.....	54
Gambar 4.3. Perbandingan Hasil Recall	55
Gambar 4.4. Perbandingan Hasil Akurasi.....	55

DAFTAR TABEL

	Halaman
Tabel 2.1. Tabel Penelitian Terkait	23
Tabel 3.1. Tabel Rincian Dataset Kanker Payudara	27
Tabel 3.2. Tabel Model Arsitektur Deep Learning	32
Tabel 3.3. Tabel Parameter Model	33
Tabel 3.4. Tabel Parameter Optimasi	33
Tabel 3.5. Tabel Spesifikasi Komputer	34
Tabel 4.1. Tabel Dimensi Data	35
Tabel 4.2. Tabel Tipe Data	35
Tabel 4.3. Tabel Missing Values	36
Tabel 4.4. Tabel Distribusi Label Target	36
Tabel 4.5. Tabel Hasil Label Encoder	39
Tabel 4.6. Tabel Dataset WBCD Setelah Feature Encoding	40
Tabel 4.7. Tabel Dataset WDBC Setelah Feature Encoding	40
Tabel 4.8. Tabel Dataset Coimbra Setelah Feature Encoding	41
Tabel 4.9. Tabel Split Data	41
Tabel 4.10. Tabel Model A Dataset WBCD	42
Tabel 4.11. Tabel Model B Dataset WBCD	42
Tabel 4.12. Tabel Model C Dataset WBCD	42
Tabel 4.13. Tabel Model D Dataset WBCD	42
Tabel 4.14. Tabel Model E Dataset WBCD	43
Tabel 4.15. Tabel Model A Dataset WDBC	43
Tabel 4.16. Tabel Model B Dataset WDBC	43
Tabel 4.17. Tabel Model C Dataset WDBC	44
Tabel 4.18. Tabel Model D Dataset WDBC	44
Tabel 4.19. Tabel Model E Dataset WDBC	44
Tabel 4.20. Tabel Model A Dataset Coimbra	45
Tabel 4.21. Tabel Model B Dataset Coimbra	45
Tabel 4.22. Tabel Model C Dataset Coimbra	45
Tabel 4.23. Tabel Model D Dataset Coimbra	45
Tabel 4.24. Tabel Model E Dataset Coimbra.....	46
Tabel 4.25. Tabel Hasil Training dan Validation Model Dataet WBCD.....	47
Tabel 4.26. Tabel Hasil Training dan Validation Model Dataet WDBC.....	48
Tabel 4.27. Tabel Hasil Training dan Validation Model Dataet Coimbra.....	48
Tabel 4.28. Tabel Hasil Akurasi	49
Tabel 4.29. Tabel Hasil Presisi	50
Tabel 4.30. Tabel Hasil Recall.....	51
Tabel 4.31. Tabel Hasil F1-Score	51
Tabel 4.32. Tabel Hasil Rekapitulasi Score.....	52
Tabel 4.33. Tabel Hasil Perbandingan Matrics Performance	52

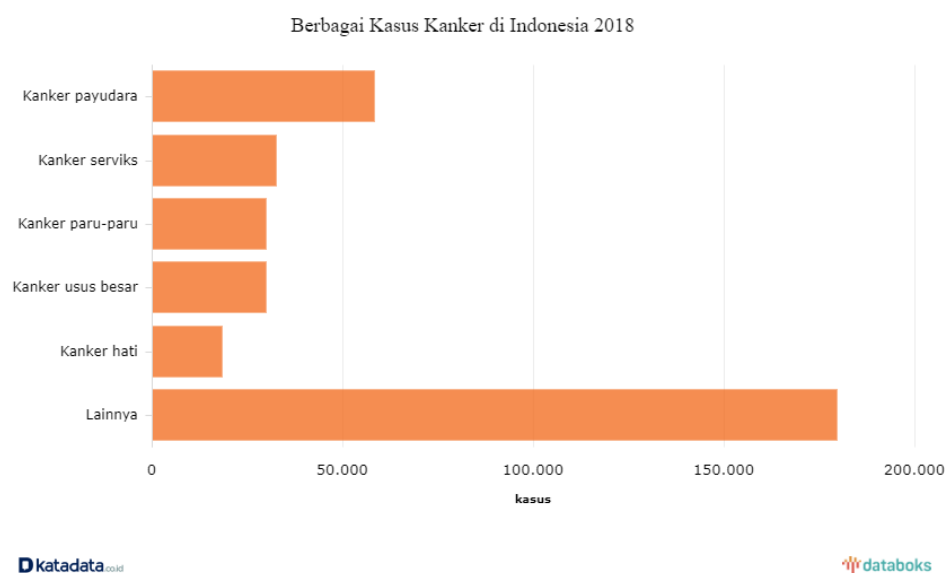
BAB 1

PENDAHULUAN

1.1. Latar Belakang Penulisan

Kanker payudara dapat terjadi pada pria dan wanita, tetapi jauh lebih sering terjadi pada wanita dan merupakan kanker yang paling banyak didiagnosis pada wanita. Kanker payudara adalah jenis kanker paling agresif yang diderita oleh wanita di seluruh dunia dan menjadi penyebab kematian kedua di antara pasien kanker wanita [1]. Pertumbuhan sel payudara yang tidak normal akan menyerang sel-sel di sekitar payudara lebih cepat dan menyebar ke bagian tubuh lainnya [2]. Kanker payudara memiliki dua kelas yang berbeda yaitu jinak dan ganas. Tumor jinak, biasanya dikenal sebagai non-kanker, dibiarkan begitu saja oleh dokter tanpa mengangkatnya karena tidak terlalu agresif terhadap jaringan lain, tetapi kadang-kadang dapat tumbuh menyebabkan rasa sakit atau lainnya. Di sisi lain, tumor ganas atau dikenal sebagai kanker terlalu agresif dan memiliki kemampuan untuk merusak jaringan di sekitarnya dan jika pasien didiagnosis ganas maka dokter melakukan biopsi untuk mengetahui agresivitas atau tingkat keparahan tumor [3]. Penyebab utama penyakit kanker payudara terutama terkait dengan pasien yang mewarisi mutasi genetik pada gen mereka [4]. Tanda-tanda kanker payudara mungkin termasuk benjolan di payudara, perubahan bentuk payudara, lesung pipit pada kulit, cairan yang keluar dari puting susu, dan kemudian menyebar ke area tubuh lainnya.

Data Global Cancer Observatory 2018 dari World Health Organization (WHO) menunjukkan kasus kanker yang paling banyak terjadi di Indonesia adalah kanker payudara, yakni 58.256 kasus atau 16,7% dari total 348.809 kasus kanker. Kementerian Kesehatan (Kemenkes) menyatakan, angka kanker payudara di Indonesia mencapai 42,1 orang per 100 ribu penduduk. Rata-rata kematian akibat kanker ini mencapai 17 orang per 100 ribu penduduk.



Sumber: World Health Organization (WHO), Februari 2019

Gambar 1.1. Berbagai Kasus Kanker di Indonesia

Kanker payudara adalah kanker yang paling sering didiagnosis dan penyebab utama kematian akibat kanker dengan jumlah persentase 11,6%. Pada tahun 2018 kanker payudara di dunia ada sekitar 2,1 juta kasus. Penyakit ini merupakan kanker yang paling sering didiagnosis di sebagian besar negara (154 dari 185) dan juga merupakan penyebab utama kematian akibat kanker di lebih dari 100 negara.[5].

Diagnosis dini pasien kanker payudara sangat penting untuk mencegah perkembangan kanker payudara yang cepat selain dari evolusi prosedur pencegahan [6] dan membantu pemulihan yang cepat dan mengurangi kemungkinan kematian [7]. Diagnosis kanker payudara dapat dilakukan secara manual oleh dokter, namun membutuhkan waktu yang lebih lama dan harus sangat rumit bagi dokter untuk menerapkan klasifikasi tersebut [8]. Ketidaklengkapan data yang relevan juga dapat menyebabkan kesalahan manusia dalam diagnosis [9] Meskipun kemajuan dalam pengobatan kanker payudara menyebabkan penurunan angka kematian akibat kanker payudara di semua kelompok umur, usia muda tetap menjadi faktor risiko tinggi dan memiliki tingkat kelangsungan hidup yang rendah [1]. Karena beratnya penyakit, diperlukan sistem deteksi berbantuan komputer (CAD) menggunakan pendekatan pembelajaran mesin (ML) untuk diagnosis kanker payudara [10] Oleh sebab itu meningkatkan akurasi mengidentifikasi penyakit kanker payudara adalah tugas yang sangat penting [11]. Berbagai metode dapat diterapkan untuk klasifikasi kanker payudara [12] untuk membedakan antara dua jenis tumor payudara yaitu Jinak dan Ganas [3].

Klasifikasi merupakan proses data mining yang bertujuan untuk membagi data ke dalam kelas-kelas untuk memudahkan pengambilan keputusan, karena itu merupakan tugas penting dalam bidang medis. Ada banyak peneliti yang sudah melakukan penelitian untuk memprediksi kanker payudara, seperti yang dilakukan oleh Nur Farahaina Idris dan Mohd Arfian Ismail (2021) tentang perbandingan akurasi antara satu metode dengan metode lain seperti SVM, algoritma C4.5, Naïve Bayes, Random forest, KNN dan Fuzzy-ID3 dengan 3 dataset yang digunakan WBCD asli, dataset WBCD diagnostik dan dataset Coimbra. Dalam penelitian ini

Fuzzy-ID3 menghasilkan performansi yang baik dalam klasifikasi data kanker payudara. Karena metode yang diusulkan dinilai mampu mengatasi kelemahan dari algoritma ID3 tradisional yang tidak mampu menangani data bernilai kontinu. Namun metode yang diusulkan memiliki ketidakstabilan dan masalah *overfitting*.

Berdasarkan uraian diatas, untuk menangani kelemahan yang masih ada pada penelitian sebelumnya, maka penulis mengusulkan Deep Learning menggunakan optimasi Adam. Karena teknik *deep learning* memiliki keunggulan dapat mengevaluasi sejumlah besar data dengan lebih tepat dan lebih cepat, sehingga menghasilkan pengurangan kesalahan selama proses pendeteksian [13].

1.2. Identifikasi Masalah

Berdasarkan latar belakang yang telah diuraikan, terdapat permasalahan yang terjadi pada penelitian Idris dan Ismail [12] Penelitian yang menggunakan metode klasifikasi meliputi SVM, algoritma C4.5, Naïve Bayes, Random forest, KNN, Fuzzy-ID3 dan deep learning algorithm belum dapat menghasilkan akurasi yang optimal di sebabkan hasil yang dihasilkan dari model terdapat ketidakstabilan dan masalah *overfitting* yang terjadi pada data kanker payudara. Ketidakstabilan data dan *overfitting* pada Fuzzy ID3 dapat menyebabkan hasil akurasi belum maksimal. Di sisi lain, *deep learning algorithm* pada umumnya menggunakan *classical stochastic gradient descent* yang menggunakan *single learning rate* untuk semua pembaruan berat pada jaringan. Hal ini memotivasi peneliti untuk menginvestigasi *performence* klasifikasi dalam *deep learning algorithm* untuk meningkatkan akurasi pada deteksi kanker payudara.

1.3. Tujuan Penelitian

Adapun tujuan dari penelitian ini:

1. Penelitian ini mengusulkan menggunakan adam OPTIMIZER untuk deep learning dengan tujuan meningkatkan akurasi pada klasifikasi kanker payudara.
2. Mengevaluasi Adam OPTIMIZER yang di usulkan pada deep learning untuk mempercepat pelatihan, tingkat akselerasi, dan accuracy dalam klasifikasi kanker payudara.

1.4. Ruang Lingkup Penelitian

Penulis membatasi permasalahan pada penelitian dengan menggunakan pengujian algoritma *deep learning* untuk mengklasifikasikan kanker payudara. Data yang digunakan adalah dataset yang di ambil dari UCI *Machine Learning* terdiri dari 3 dataset kanker payudara yaitu WBCD asli, dataset WBCD diagnostik dan dataset Coimbra.

1.5. Kontribusi Thesis

Kontribusi dari tesis ini adalah melakukan usulan model klasifikasi untuk melakukan klasifikasi kanker payudara dengan eksperimen yang dilakukan yaitu pengaplikasian 1 sampai 5 *hidden layer*, 3 *learning rate* dan mengevaluasi model untuk mengukur tingkat keakuratan dengan akurasi, *Presisi*, *Recall* dan *f1-Score*.

1.6. Sistematika Penulisan

Sistematika penulisan dimaksudkan untuk memberikan gambaran yang jelas mengenai masalah yang akan dibahas. Berikut sistematika penulisan pada penelitian ini:

BAB I PENDAHULUAN

Pada bab ini berisi uraian fakta dan data yang berkaitan dengan permasalahan yang akan diteliti, mendeskripsikan mengenai latar belakang masalah, identifikasi masalah, tujuan penelitian, ruang lingkup penelitian, kontribusi tesis dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini berisi tentang landasan teori yang digunakan dalam melakukan penelitian. Pada bab ini membahas tentang landasan-landasan teoritis yang digunakan untuk melaksanakan penelitian yang mencakup tinjauan studi dan tinjauan pustaka serta ringkasan dari karya-karya yang serupa untuk mendukung proses penelitian.

BAB III METODE PENELITIAN

Pada bab ini membahas tentang metode pengumpulan data penelitian yang membahas tentang perencanaan metode penelitian serta konsep dari penelitian yang nantinya akan digunakan sebagai solusi penyelesaian masalah yang diangkat dalam penelitian, eksperimen yang dilakukan, matriks performansi evaluasi model.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini berisi tentang hasil dan pembahasan dari penelitian dimulai dari pengambilan data, pengolahan data dengan menggunakan model yang diusulkan untuk mendapatkan nilai evaluasi dari masing-masing metode yang digunakan serta menjelaskan hasil penelitian dan keterlibatan dalam penelitian.

BAB V PENUTUP

Pada bab ini berisi tentang kesimpulan dari pembahasan pada bab-bab sebelumnya dan saran untuk penelitian di masa depan.

BAB 2

LANDASAN/KERANGKA PEMIKIRAN

2.1. Tinjauan Pustaka

2.2.1. Kanker Payudara

Kanker adalah sekumpulan penyakit di mana sel-sel tubuh tumbuh, berubah, dan berkembang biak tanpa kendali. Biasanya, nama kanker berasal dari bagian tubuh tempat asalnya. Karena itu, kanker payudara mengacu pada pembesaran sel yang tidak terduga dan seringkali cepat yang dimulai di jaringan payudara. Sekelompok sel yang memisahkan dengan cepat dapat membentuk massa jaringan ekstra, yang disebut tumor. Tumor dapat bersifat kanker (ganas) atau non-kanker (jinak). Tumor ganas berjalan melalui jaringan tubuh yang sehat dan menghancurkannya [14].

Faktor risiko pertama yang dapat meningkatkan kemungkinan terjadinya kanker payudara adalah faktor usia, risiko kanker payudara meningkat seiring dengan bertambahnya usia. Faktor lain yang dapat mengintervensi seperti: Faktor keluarga atau genetik. Jenis Kelamin: Wanita adalah yang paling banyak terinfeksi payudara kanker [15].

1. Riwayat wanita: Wanita yang sudah pernah menderita kanker payudara di salah satu payudaranya, memiliki peningkatan risiko terkena kanker di payudara lainnya.
2. Riwayat keluarga: Jika beberapa orang tua dari wanita tersebut telah didiagnosis menderita kanker payudara, terutama pada usia yang lebih muda, risiko terkena kanker payudara meningkat.

Program Studi Ilmu Komputer (S2) Universitas Nusa Mandiri

3. Faktor genetik: Beberapa mutasi genetik meningkatkan risiko kanker payudara.

2.2.2. Data Mining

Data mining merupakan proses penggalian informasi penting dan pola-pola yang tersembunyi dalam data dengan jumlah yang besar [16]

Data mining memiliki 4 fungsi dasar yang biasa kita temui yaitu [17]:

1. Fungsi Klasifikasi, fungsi ini merupakan fungsi untuk menemukan model/fungsi untuk menggambarkan suatu class dari data. Dengan klasifikasi mampu mempelajari data sehingga mampu meramalkan kecenderungan data.
2. Fungsi Prediksi, fungsi ini digunakan untuk menemukan pola data menggunakan variabel dalam memprediksi variabel lain yang belum diketahui jenisnya
3. Fungsi Deskripsi, fungsi ini digunakan untuk menemukan karakteristik data
4. Fungsi Asosiasi, fungsi ini digunakan untuk menemukan hubungan yang ada pada nilai atribut dari banyaknya data

2.2.3. Klasifikasi

Klasifikasi merupakan proses untuk menemukan suatu kelas data suatu objek yang belum diketahui berdasarkan data sebelumnya. Klasifikasi termasuk ke dalam metode pembelajaran atau supervised karena membutuhkan pembelajaran data sebelumnya untuk menentukan hasil dari data baru [17].

Klasifikasi memiliki 4 komponen dasar yaitu:

1. *Class*, merupakan variabel yang menjadi label atau hasil suatu objek;

2. *Predictor*, merupakan variabel yang menjadi atribut dari data yang akan digunakan pada klasifikasi;
3. *Training dataset*, merupakan data yang telah memiliki label sebelumnya;
4. *Testing dataset*, merupakan data baru yang akan dilakukan proses klasifikasi.

2.2.4. Data Preprocessing

Preprocessing merupakan langkah pertama sebelum memodelkan data. Terkadang data yang diperoleh mungkin tidak lengkap, seperti tidak adanya nilai pada baris tertentu. Itu juga bisa berisik, datanya bisa penuh dengan kesalahan dan outlier, dan juga tidak konsisten, dengan perbedaan yang tidak jelas. Jadi, untuk menghilangkan semua hal di atas, kami melakukan praproses data sebelum pemodelan [2].

1. Cleaning Data

Langkah pra-pemrosesan ini pertama-tama melibatkan penghapusan kolom atau entri yang tidak berguna. Pada langkah ini, kolom dihapus dari kumpulan data. Memproses nilai yang hilang juga merupakan tugas penting untuk penambahan data. Melewatkan nilai-nilai yang hilang ini bukanlah solusi yang baik, karena informasi berharga juga dapat hilang. Solusi efektif untuk jenis masalah ini adalah mengganti nilai yang hilang dengan rata-rata entri di kolom itu. Langkah-langkah kecil ini memiliki kemampuan untuk meningkatkan kinerja pengklasifikasi [18].

2. Feature Encoding

Feature encoding merupakan proses mengubah nilai kategorikal pada suatu feature yang berbentuk label ke dalam bentuk numerik. Label encoding

merupakan cara yang sangat sederhana untuk mengonversi nilai kategoris menjadi nilai numerik. Label Encoding hanyalah menetapkan nilai bilangan bulat ke setiap nilai yang mungkin dan berbeda dari variabel kategoris [19]. Label Encoder adalah alat yang efisien untuk mengkodekan level fitur kategoris ke dalam nilai numerik. Label Encoder menyandikan label dengan nilai antara 0 dan kelas -1. Semua fitur kategoris kami dikodekan [8]. Hal ini merupakan langkah penting preprocessing untuk dataset terstruktur dalam supervised learning.

3. Ekstraksi Feature

Pengkodean adalah proses konversi data dari satu formulir ke formulir lainnya. Sebagian besar algoritma mesin pembelajaran tidak dapat menangani nilai kategorikal kecuali kita mengonversinya menjadi nilai numerik [19]

Kami sering memperhatikan fitur-fitur tertentu yang paling berkontribusi pada prediktor atau keluaran. Proses pemilihan variabel keluaran ini disebut dengan metode seleksi fitur. Adanya atribut yang tidak berhubungan dalam kumpulan data dapat mempengaruhi keakuratan kumpulan data. Sebelum pemodelan data, pentingnya ekstraksi fitur mungkin bisa membantu. Yaitu: meningkatkan akurasi, mengurangi over fitting, dan mengurangi waktu latihan [2].

4. Split Data

Membagi dataset dalam tujuan pelatihan dan pengujian dan menghasilkan beberapa grafik untuk data visualisasi [8].

5. Normalisasi

Normalisasi data adalah proses rescaling satu dan lebih atribut ke kisaran 0 sampai 1. Ini berarti nilai terbesar untuk setiap atribut adalah 1 dan nilai terkecil adalah 0. Normalisasi adalah teknik yang baik untuk digunakan ketika tidak mengetahui distribusi data [8].

2.2.5. Arsitektur *Deep Learning*

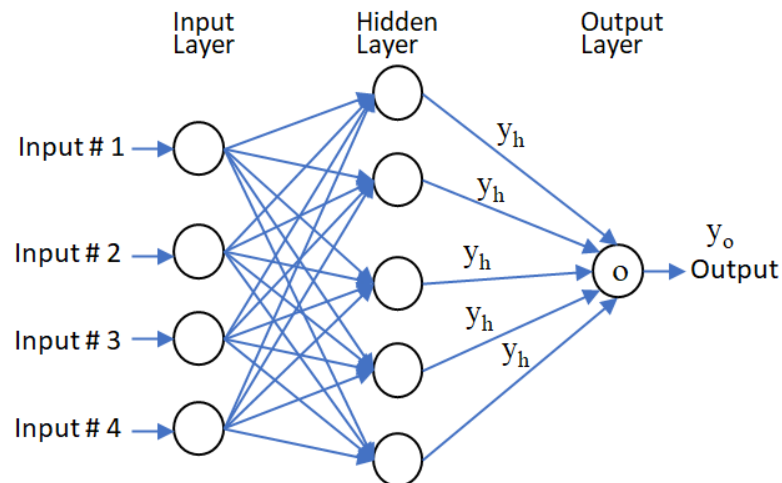
Deep learning merupakan bagian dari *machine learning* yang berbentuk *Artificial Neural Network (ANN)*, algoritma yang digunakan terinspirasi oleh otak manusia, belajar dari sejumlah besar data. *Deep learning* memungkinkan mesin untuk memecahkan masalah yang kompleks, bahkan ketika menggunakan kumpulan data yang sangat beragam, tidak terstruktur, dan saling terkait. Sehingga dapat dikatakan bahwa *deep learning* merupakan pendekatan dalam *Artificial Intelligent (AI)*.

1. *Neural Network*

Neural Network merupakan sebuah sistem pemrosesan informasi yang memiliki karakteristik serupa dengan jaringan neural pada otak manusia. *NN* adalah pemodelan otak manusia dengan definisi paling sederhana dan blok bangunan Neuron [20]. Prinsip kerja *NN* seperti pada otak manusia. *NN* terdiri dari lapisan-lapisan neuron. Arsitektur paling sederhana *NN* setidaknya memiliki lapisan input dan lapisan *output*. Arsitektur *NN* sebagaimana diilustrasikan pada Gambar 2.1.

Pada gambar tersebut terdapat *input*, *input layer*, *hidden layer*, dan *output layer* yang terdiri atas lapisan-lapisan *neuron*. Arsitektur pada gambar tersebut memiliki 3 buah *neuron* pada *input layer*, 5 *neuron* pada *hidden*

layer dan 1 buah *neuron* pada *output layer*. *Neuron* pada *input layer* tidak memiliki *activation function*, sedangkan *neuron* pada *hidden layer* dan *output layer* memiliki *activation function* yang bisa jadi berbeda tergantung sesuai dengan problem yang dihadapi. Pada *hidden layer* dan *output layer* biasanya juga terdapat input lain yang dinamakan bias.



Gambar 2.2. Arsitektur Neural Network

Proses training pada NN secara garis besar terbagi menjadi 2 (dua), yaitu: *Forward Pass* (*Forwardpropagation*) dan *Backward Pass* (*Backpropagation*). *Forwardpropagation* merupakan proses membawa data pada input melewati tiap neuron pada *hidden layer* sampai kepada *output layer*. Setelah sampai pada *output layer* akan dihitung *error*nya. Hasil hitung *error* akan digunakan untuk meng-*update weight* dan bias dengan learning rate tertentu dengan algoritma *backpropagation*. Kedua proses ini akan berulang sampai mendapatkan nilai *weight* dan bias mendapatkan hasil *output* dengan *error* paling kecil.

2. *Backpropagation*

Backpropagation adalah metode inti pembelajaran untuk *deep learning*. Prinsip algoritma *Backpropagation* tidak lain adalah akan menyesuaikan *weight* dan bias berdasarkan hasil hitung *error* pada saat *forwardpropagation*. *Backpropagation of errors* pada dasarnya hanya turunan gradien [21]. Turunan gradien mengacu pada perhitungan gradien pada setiap *weight NN* untuk setiap elemen *training*. Karena *NN* tidak akan mengeluarkan nilai seperti yang diharapkan atau sama persis seperti target, gradien setiap *weight* akan mengindikasikan cara memodifikasi setiap *weight* untuk mencapai *output* yang diharapkan. Jika *NN* mengeluarkan nilai *output* persis apa yang diharapkan atau sama dengan target, maka *gradien* untuk setiap *weight* akan 0. Hal ini menunjukkan bahwa tidak ada perubahan yang diperlukan pada *weight* [22].

3. *Loss Functions*

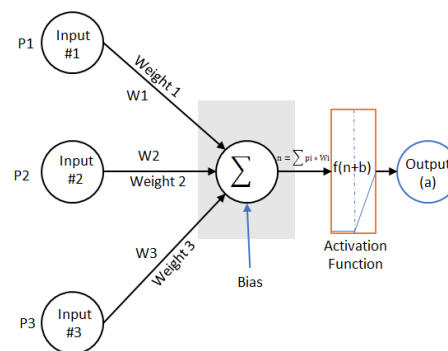
Terjadinya *error* tidak lain adanya perbedaan antara nilai target aktual dan nilai yang diprediksi. Hal ini juga dinamakan *loss*. Untuk memperkirakan nilai kerugian dengan benar diperlukan fungsi kerugian (*loss function*) yang sesuai. Selama proses pelatihan *NN*, bobot dan bias harus diperbarui pada setiap evaluasi untuk mengurangi nilai kerugian dalam evaluasi berikutnya. *Cost function* merupakan kunci untuk menyesuaikan *weight NN* agar model pembelajaran menjadi lebih baik [23]. *Binary Crossentropy* [24] merupakan *loss function* yang mengidentifikasi kerugian berdasarkan nilai probabilitas. Nilai probabilitas terletak antara 0 dan 1. Model sempurna dengan kerugian

nol akan memiliki nilai probabilitas 0. Berikut merupakan rumus yang digunakan untuk klasifikasi biner [23].

2.2.6. *Arsitektur dan Hyperparameter Deep Learning*

1. *Neuron*

Setiap elemen yang memegang input disebut '*neuron*'. *Neuron-neuron* membentuk lapisan yang dinamakan *layer*. Setiap *neuron* dari *input layer* terhubung ke setiap *neuron* dari *hidden layer*, tetapi *neuron* dari layer yang sama tidak saling berhubungan [21]. Proses pembelajaran pada neuron hanyalah modifikasi atau pembaruan *weight* dan bias selama pelatihan dengan *backpropagation*. Proses pembelajaran pada neuron sebagaimana diilustrasikan pada Gambar 2.3. [22].



Sumber: (Jeff Heaton, 2015)

Gambar 2.3. Artificial Neuron

2. *Layer*

Meskipun tidak ada ketentuan secara pasti dalam menentukan jumlah layer pada *NN*, namun jumlah *layer* harus dipilih dengan tepat karena jumlah *layer* yang sangat tinggi dapat menjadi masalah seperti masalah *overfitting* dan jumlah yang lebih rendah dapat menyebabkan model memiliki bias

tinggi dan mengakibatkan kinerja model menjadi tidak baik. Pemilihan jumlah *layer* sangat tergantung pada ukuran data yang digunakan untuk pelatihan.

3. *Activation Function*

Salah satu yang mempengaruhi tingkat akurasi suatu model *NN* adalah tipe fungsi aktivasi (*activation function*) yang digunakan [25]. Fungsi aktivasi digunakan untuk mengenali pola/jaringan *back-propagation*, prediksi/aproksimasi linear, dan pengambilan keputusan. Fungsi aktivasi yang paling sering digunakan saat ini adalah *Rectified Linear Unit (ReLU)* dan *Sigmoid*.

- a. *ReLU* merupakan fungsi aktivasi *non-linear* yang banyak digunakan dalam *NN*. Fungsi *ReLU* tidak memiliki turunan di 0. Namun karena konvensi, gradien 0 disubstitusi ketika x adalah 0. Hal ini berarti akan mengembalikan 0 jika inputnya negatif [25].
- b. *Sigmoid* merupakan fungsi aktivasi termasuk yang paling banyak digunakan karena *non-linear*. *Sigmoid* digunakan ketika melakukan prediksi dalam bentuk biner. Fungsi *sigmoid* mengubah fungsi nilai dalam rentang 0 sampai dengan 1 [25].
- c. Tingkat pembelajaran (*learning rate*) yang terlalu tinggi dapat mengakibatkan gradien yang naik secara ekstrem atau malah turun secara tajam, serta terjebak dalam kondisi yang buruk. *Batch Normalisasi* membantu mengatasi masalah ini [26].

- d. *Batch Normalization* secara signifikan meningkatkan waktu pelatihan, meningkatkan akurasi hanya dengan menurunkan *initial weights* dan sedikit membantu regularisasi [27].
- e. *Epoch* pada pembelajaran *back-propagation* merupakan sekali proses pembaruan bobot (*weight update process*) atau sekali proses iterasi. Untuk setiap *epoch*, algoritma pembelajaran *back-propagation* membangun model jaringan dengan sekumpulan bobot (*weight*) yang berbeda [28]. Jumlah *epoch* merupakan berapa kali seluruh *data training* ditampilkan ke model dan memainkan peran penting dalam menentukan seberapa baik model cocok pada *data training*. Jumlah *epoch* yang tinggi dapat menyebabkan masalah generalisasi pada saat validasi. Jumlah *epoch* yang lebih rendah juga dapat membatasi potensi mendapatkan model yang lebih baik.

2.2.7. Parameter Optimasi

Optimizer merupakan algoritma yang digunakan oleh model untuk memperbarui bobot setiap lapisan setelah setiap iterasi. Di antara variasi pengujian berbagai optimasi *Adam* merupakan optimasi yang paling populer dan paling baik menghasilkan model [29].

1. *Adam* merupakan turunan dari *adaptive moment estimation*, yaitu metode untuk optimasi *stochastic* yang efisien yang hanya memerlukan gradien urutan pertama dengan sedikit persyaratan memori. Metode ini menghitung tingkat pembelajaran adaptif individu untuk parameter yang berbeda dari perkiraan momen pertama dan kedua dari gradien [30].

2. *Learning rate* adalah *hyperparameter* yang mengontrol seberapa banyak perubahan model dalam menanggapi kesalahan yang diperkirakan setiap kali bobot model diperbarui. Memilih *learning rate* merupakan tantangan karena nilai yang terlalu kecil dapat mengakibatkan proses pelatihan yang lama dan dapat macet, sedangkan nilai yang terlalu besar dapat mengakibatkan rangkaian pembelajaran mendapatkan bobot yang kurang optimal atau proses pelatihan terlalu cepat yang tidak stabil [31]. Nilai *learning rate* sangat tergantung pada *optimizer* yang digunakan. Pada Adam umumnya bekerja dengan baik untuk nilai 0.001 atau 0.01, tetapi perlu untuk mencoba semua nilai dari kisaran di atas. Umumnya lebih baik menggunakan algoritma tingkat pembelajaran adaptif seperti Adam daripada menggunakan tingkat pembelajaran yang lebih buruk. Rekomendasi agar menaikkan nilai *batch size* daripada membuat *learning rate* lebih buruk. Ketika kita meningkatkan koefisien momentum, kita meningkatkan akumulasi skala waktu yang diperlukan untuk melupakan gradien lama [32].

2.2.8. Metrics Performance

Metric Performance merupakan suatu nilai yang digunakan untuk mengukur kinerja suatu model. Berikut beberapa *metric performance* yang dapat digunakan untuk mengukur suatu model *machine learning*:

1. Accuracy

Accuracy merupakan salah satu *metric performance* yang paling umum digunakan untuk kinerja klasifikasi, dan didefinisikan sebagai rasio antara sampel yang diklasifikasikan dengan benar ke jumlah total Sampel [33]. *Accuracy* dapat dihitung dengan menggunakan rumus sebagai berikut [34]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2. Precision

Precision adalah rasio sampel berkategori positif yang diklasifikasi dengan benar dibandingkan dengan total sampel yang diklasifikasi sebagai positif.

$$Precision = \frac{TP}{TP + FP}$$

3. Recall

Recall adalah rasio sampel yang diprediksi positif yang diidentifikasi dengan benar.

$$Recall = \frac{TP}{TP + FN}$$

4. F1-score

F1-Score rata-rata harmonic antara *precision* dan *recall*.

$$f1 - score = 2 \frac{Recall * Precision}{Recall + Precision}$$

2.2. Tinjauan Pustaka

Berikut ini merupakan penelitian-penelitian yang telah dilakukan sebelumnya terkait klasifikasi penyakit kanker payudara. Ada banyak teknik pembelajaran *machine learning* yang tersedia untuk mengklasifikasikan kanker payudara [8].

1. Berdasarkan penelitian yang dilakukan oleh [12] yang berjudul “*Breast cancer disease classification using fuzzy-ID3 algorithm with FUZZYDBD method: automatic fuzzy database definition*” Model yang diusulkan adalah metode FUZZYDBD untuk mengklasifikasikan penyakit kanker payudara pada tiga dataset yaitu dataset WBCD (Asli), WDBC (Diagnostik) dan dataset Coimbra dan menghasilkan akurasi yang lebih tinggi, yaitu dengan memperoleh hasil sebesar 94.362, 94.362, 70.690.
2. Berdasarkan penelitian yang dilakukan oleh [35] yang berjudul “Decision support system for breast cancer prediction” Model yang diusulkan adalah Support Vector Machine (SVM) pada dataset WBCD menghasilkan akurasi 86.100.
3. Berdasarkan penelitian yang dilakukan oleh [15] yang berjudul “Using Feature Selection Techniques to Improve the Accuracy of Breast Cancer Classification” Model yang diusulkan adalah C4.5 algorithm untuk mengklasifikasikan penyakit kanker payudara pada dataset WBCD menghasilkan akurasi 92.970.
4. Berdasarkan hasil penelitian yang dilakukan oleh [18] yang berjudul “Effective K-nearest neighbor classifications for Wisconsin breast cancer data sets” Model yang diusulkan *Naïve Bayes* menggunakan dataset WBCD mendapatkan hasil akurasi 91.810

5. Berdasarkan hasil penelitian yang dilakukan oleh [36] yang berjudul “*Performance study of classification algorithms using the breast cancer dataset*” Model yang diusulkan *Random Forest* menggunakan dataset WBCD mendapatkan hasil akurasi 91.660
6. Berdasarkan hasil penelitian yang dilakukan oleh [18] yang berjudul “*Effective K-nearest neighbor classifications for Wisconsin breast cancer data sets*” Model yang diusulkan *K-Nearest Neighbors (KNN)* menggunakan dataset WBCD mendapatkan hasil akurasi 92.570.
7. Berdasarkan penelitian yang dilakukan oleh [2] yang berjudul “*Applications of Machine Learning Techniques to Predict Diagnostic Breast Cancer*” Model yang diusulkan adalah *Support Vector Machine (SVM)* pada dataset WDBC menghasilkan akurasi 61.96.
8. Berdasarkan penelitian yang dilakukan oleh [3] yang berjudul “*Data Mining Algorithms for Classification of Diagnostic Cancer Using Genetic OPTIMIZER Algorithms*” Model yang diusulkan adalah *C4.5 algorithm* pada dataset WDBC menghasilkan akurasi 94.030.
9. Berdasarkan hasil penelitian yang dilakukan oleh [10] yang berjudul “*Machine Learning Classification Techniques for Breast Cancer Diagnosis*” Model yang diusulkan *Naïve Bayes* menggunakan dataset WDBC mendapatkan hasil akurasi 91,18%.
10. Berdasarkan penelitian yang dilakukan oleh [37] yang berjudul “*Classification using probabilistic random forest*” Model yang diusulkan adalah *Random forest* untuk mengklasifikasikan penyakit kanker payudara pada dataset WDBC menghasilkan akurasi 89.370.

11. Berdasarkan penelitian yang dilakukan oleh [2] yang berjudul “Applications of Machine Learning Techniques to Predict Diagnostic Breast Cancer” Model yang diusulkan adalah K-Nearest Neighbors (KNN) pada dataset WDBC menghasilkan akurasi 92.77.
12. Berdasarkan hasil penelitian yang dilakukan oleh [38] yang berjudul “Deep neural network classifier in breast cancer prediction ” Model yang diusulkan Support Vector Machine (SVM) menggunakan dataset Coimbra mendapatkan hasil akurasi 65.960.
13. Berdasarkan hasil penelitian yang dilakukan oleh [39] yang berjudul “Performance Analysis Of Filter Based Feature Selection Methods On Diagnosis Of Breast Cancer And Orthopedics ” Model yang diusulkan C4.5 algorithm menggunakan dataset Coimbra mendapatkan hasil akurasi 68.000.
14. Berdasarkan hasil penelitian yang dilakukan oleh [40] yang berjudul “Reviewing the consistency of the Naïve Bayes Classifier's performance in medical diagnosis and prognosis problems ” Model yang diusulkan Naïve Bayes menggunakan dataset Coimbra mendapatkan hasil akurasi 67.700.
15. Berdasarkan hasil penelitian yang dilakukan oleh [41] yang berjudul “Comparison of machine learning algorithms in breast cancer prediction using the coimbra dataset” Model yang diusulkan Random forest menggunakan dataset Coimbra mendapatkan hasil akurasi 70.310.
16. Berdasarkan hasil penelitian yang dilakukan oleh [42] yang berjudul “Breast cancer–detection system using PCA, multilayer perceptron, transfer learning, and support vector machine ” Model yang diusulkan K-Nearest

Neighbor (KNN) menggunakan dataset Coimbra mendapatkan hasil akurasi 67.120.

Tabel 2.1
Penelitian Terkait

No	Penulis	Judul	Metode	Akurasi
1	(Idris Da Ismail, 2021) [12]	Breast Cancer Disease Classification Using Fuzzy-ID3 Algorithm With FUZZYDBD Method: Automatic Fuzzy Database Definition	FUZZYDBD	94.362
2	(Kumari, Singh & Ahlawat, 2019) [35]	Decision Support System For Breast Cancer Prediction. International Journal For Research In Applied Science And Engineering Technology	Support Vector Machine (SVM)	86.100
3	(Saoud Et Al., 2019) [15]	Using Feature Selection Techniques To Improve The Accuracy Of Breast Cancer Classification	C4.5 Algorithm	92.970
4	(Assiri, Nazir & Velastin, 2020) [43]	<i>Machine Learning Classification Techniques For Breast Cancer Diagnosis</i>	<i>Naïve Bayes</i>	91.810
5	(Pyingkodi Et Al., 2020) [36]	Breast Tumor Classification Using An Ensemble Machine Learning Method	Random Forest	91.660.
6	(Mushtaq Et Al., 2020) [18]	Effective K-Nearest Neighbor Classifications For Wisconsin Breast Cancer Data Sets	<i>K-Nearest Neighbors (KNN)</i>	92.570.
7	(Chaurasia & Pal, 2020) [2]	Applications Of Machine Learning	Support Vector Machine (SVM)	61.9614

		Techniques To Predict Diagnostic Breast Cancer		
8	(Khan Et Al., 2017) [3]	Data Mining Algorithms For Classification Of Diagnostic Cancer Using Genetic OPTIMIZER Algorithms	C4.5 Algorithm	94.030
9	(Omondiagbe, Veeramani & Sidhu, 2019) [10]	Machine Learning Classification Techniques For Breast Cancer Diagnosis	Naïve Bayes	91.18
10	(Gondane & Susheela Devi, 2015)[37]	Classification Using Probabilistic Random Forest	Random Forest	89.370.
11	(Chaurasia & Pal, 2020) [2]	Applications Of Machine Learning Techniques To Predict Diagnostic Breast Cancer	K-Nearest Neighbors (KNN)	92.77
12	(Poorani & Balasubramanie, 2019) [38]	Deep Neural Network Classifier In Breast Cancer Prediction	Support Vector Machine (SVM)	65.960.
13	(Kayaalp & Basarslan, 2019)[39]	Performance Analysis Of Filter Based Feature Selection Methods On Diagnosis Of Breast Cancer And Orthopedics	C4.5 Algorithm	68.000.
14	(Fauziyyah, Abdullah & Nurrohmah, 2020) [40]	Reviewing The Consistency Of The Naïve Bayes Classifier's Performance In Medical Diagnosis And Prognosis Problems	Naïve Bayes	67.700.
15	(Austria Et Al., 2019) KNN [41]	Comparison Of Machine Learning Algorithms In Breast Cancer Prediction Using The Coimbra Dataset	Random Forest	70.310

16	<i>(Chiu, Li & Kuo, 2020) [42]</i>	Breast Detection Using Multilayer Perceptron, Support Machine	Cancer-System PCA, Transfer And Vector	K-Nearest Neighbor (KNN)	67.120.
----	--	---	--	--------------------------	---------

BAB 3

METODOLOGI PENELITIAN

3.1. Deskripsi Dataset dan Masalah

Kanker payudara menjadi penyebab utama kematian kedua di antara pasien kanker wanita di seluruh dunia. Berdasarkan penelitian yang dilakukan pada tahun 2019, ada sekitar 250.000 wanita di seluruh Amerika Serikat yang didiagnosis menderita kanker payudara invasif setiap tahun. Pencegahan kanker payudara tetap menjadi tantangan di dunia saat ini karena pertumbuhan sel kanker payudara merupakan proses multistep yang melibatkan banyak jenis sel. Diagnosis dini dan deteksi kanker payudara adalah salah satu pendekatan terbesar untuk mencegah penyebaran kanker dan meningkatkan tingkat kelangsungan hidup.

Dataset yang digunakan untuk penelitian ini merupakan data sekunder atau data publik yang diambil dari *UCI Machine Learning*. Eksperimen ini dilakukan dengan menggunakan tiga set data kanker payudara, yaitu: dataset WBCD (Asli), dataset WDBC (Diagnostik) dan dataset Coimbra.

Database Kanker Payudara Wisconsin, yang dikenal sebagai kumpulan data WBCD (Asli), terdiri dari satu atribut target diantaranya yaitu kelas jinak dan ganas. Dan berisi sembilan atribut prediktif, ketebalan rumpun, keseragaman ukuran sel, keseragaman bentuk sel, adhesi marginal, ukuran sel epitel tunggal, kromatin hambar, inti telanjang, nukleolus normal dan mitosis. Serta satu atribut non-prediktif dalam kumpulan data ini adalah nomor ID. Dataset terdiri dari 699 kasus, 458 sampel jinak dan 241 sampel ganas. Ini memiliki 16 nilai yang hilang, dan

penghapusan daftar telah diterapkan untuk menangani data yang hilang, yang mengarah ke 683 *instances*.

Dataset Kanker Payudara Diagnostik Wisconsin, juga dikenal sebagai WDBC (Diagnostik). Sumber dataset ini adalah dari University of Wisconsin. Kumpulan data ini terdiri dari 569 *instances* tanpa nilai yang hilang dan memiliki satu atribut kelas target yaitu jinak (B) atau ganas (M). Atribut prediktif dari kumpulan data terdiri dari sepuluh fitur bernilai nyata yang dihitung untuk setiap inti, seperti radius, tekstur, keliling, luas, kehalusan, kekompakan, kecekungan, kecekungan, simetri, dan dimensi fraktal. Rata-rata, kesalahan standar dan radius (rata-rata dari tiga pembacaan nilai terbesar) dihitung untuk setiap inti yang memimpin kumpulan data memiliki lebih dari 32 atribut, termasuk atribut non-prediktif yaitu ID pasien.

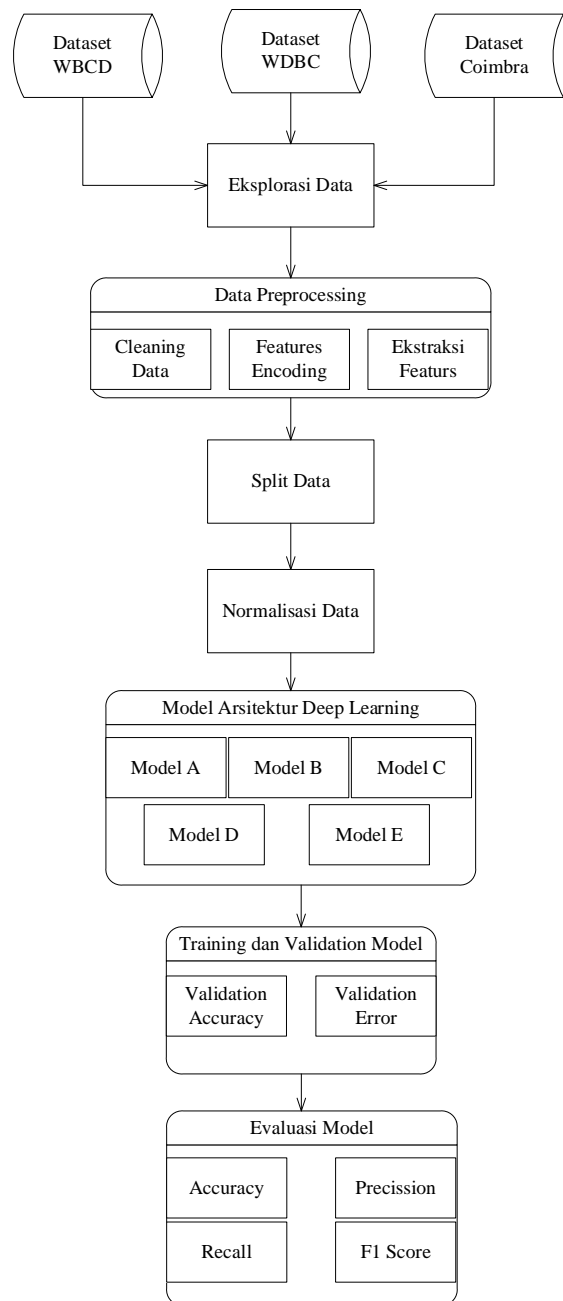
Sedangkan Dataset Kanker Payudara Coimbra yang digunakan dalam makalah ini berasal dari peneliti dari Fakultas Kedokteran Universitas Coimbra dan Pusat Rumah Sakit Universitas Coimbra. Dataset ini berisi 116 *instances* yang terdiri dari satu atribut target yang memiliki dua kelas diantaranya *healthy control* dan *patient*. Dataset ini juga memiliki sembilan atribut prediktif seperti usia (tahun), BMI (kg/m²), glukosa (mg/dL), insulin (μ U/mL), HOMA, leptin (ng/mL), adiponektin (μ g/mL), resistin (ng/mL) dan MCP-1 (pg/dL).

Tabel 3.1. merupakan Tabel kumpulan data kanker payudara yang digunakan pada penelitian ini, berdasarkan jumlah atribut, jumlah *instances* dan jumlah kelas.

Tabel 3.1. Rincian Dataset Kanker Payudara

Dataset	Jumlah Atribut	Jumlah Instance	Jumlah Kelas
WBCD (Original)	11	683	2
WDBC (Diagnostic)	32	569	2
Coimbra	10	116	2

3.2. Eksperimen Yang Dilakukan



Gambar 3.1. Tahapan Penelitian

Eksperimen penelitian dilakukan sesuai dengan tahapan-tahapan sebagaimana disajikan pada Gambar 3.1. Penelitian dimulai dari melakukan eksplorasi dataset, data *preprocessing*, split data, normalisasi data, membangun arsitektur model *deep learning*, *training* dan *validation* model, serta *evaluation model*. Penjelasan dari setiap tahapan diatas dijelaskan sebagai berikut:

3.2.1. Eksplorasi Data

Pentingnya pemahaman terhadap data diperlukan eksplorasi data untuk membantu dalam menentukan teknik preprosesing yang tepat. Berikut tahapan pada saat eksplorasi data:

1. Mengecek dimensi data

Mengecek dimensi data bertujuan untuk mengetahui jumlah baris dan kolom.

2. Mengecek type data

Mengecek tipe data bertujuan untuk memastikan tipe data sudah sesuai dengan isi data.

3. Mengecek data yang kosong

Dari pengecekan tersebut dapat diketahui data yang kosong sehingga jika ada data kosong yang masuk ke model kemungkinan model yang dihasilkan tidak berkualitas.

4. Statistik distribusi label target

Tujuan dari statistik distribusi label target ini untuk mengetahui, ada berapa banyak data, dan tipe data.

3.2.2. Data Preprocessing

Tahapan persiapan data dilakukan untuk memastikan bahwa dataset yang digunakan untuk *training* dan *testing* merupakan data yang berkualitas. Jika dataset yang digunakan masih memiliki *noise* maka model yang dihasilkan juga tidak akan berkualitas dan memiliki *bias*.

1. *Cleaning Data*

Pra-pemrosesan data dimulai dengan cleaning data yang terdiri menghapus kolom “*id*” pada dataset asli (WBCD) dan dataset diagnostik (WDBC). Kemudian mengubah nama *class* fitur (sebagai kelas target) menjadi fitur dengan nama “*diagnosis*” untuk ketiga dataset, hal tersebut dilakukan karena jika tidak mengubah nama kelas fitur maka akan dibaca sebagai fungsi.

2. *Features Encoding*

Untuk melakukan konversi nilai pada variable dalam bentuk kategorikal ke dalam bentuk numerik menggunakan teknik *Label Encoding* pada atribut target dari ketiga dataset tersebut. *Label Encoder* dari library *Scikit Learn*. Kemudian merubah label target klasifikasi dari “*category*” menjadi “*numeric*”, pada dataset WBCD (asli) dan dataset WDBC (diagnostik) berubah “*malignant*” menjadi “1” dan “*benign*” menjadi “0”. Sedangkan pada dataset Coimbra mengubah “*healthy*” menjadi “0” dan “*patient*” menjadi “1”.

3. *Ekstraksi Features*

Ekstraksi fitur (*dummy*) pada penelitian ini bertujuan untuk mengubah data input menjadi kolom biner untuk setiap kategori semua fitur input yang sebelumnya telah mengalami proses *fuzzifikasi* dimana setiap atribut yang telah dihasilkan berada dalam nilai himpunan fuzzy (rendah, sedang, tinggi). Merujuk pada penelitian yang dilakukan oleh [12].

3.2.3. *Split Data*

Split data: 80% data *training*, 20% data *testing*, data *training* digunakan untuk membangun dan melatih model, data *testing* yang akan digunakan untuk evaluasi model dan pengujian model.

3.2.4. *Normalisasi Data*

Minimisasi data dapat digunakan untuk menskalakan data dalam rentang yang sama, sehingga setiap atribut input memiliki bias yang diminimalkan dalam Jaringan Saraf Tiruan. Ini juga dapat mempercepat waktu pelatihan dengan memulai proses pelatihan untuk semua atribut dalam rentang yang sama. Dalam penelitian ini digunakan *Min-Max Scaler* dari *scikit learn* untuk merubah range dengan skala input menjadi 0-1.

3.2.5. Arsitektur Deep Learning

Model Arsitektur Deep Learning dibuat terdiri dari input layer, hidden layer dan output layer. Tabel 3.2 merupakan model yang dibangun dengan skenario lima model yang terdiri Model-A memiliki satu hidden layer, Model-B memiliki dua hidden layer, Model-C memiliki tiga hidden layer, Model-D memiliki empat hidden layer, dan model-E memiliki lima hidden layer. Semua model yang dibuat menerapkan prinsip kerja *learning back-propagation algorithm* dan semua model diterapkan pada ketiga jenis dataset yang digunakan pada penelitian ini.

Tabel 3.2. Model Arsitektur Deep Learning

Arsitektur	layer
Model-A	Satu input layer (6), satu hidden layer (15), satu output layer (1)
Model-B	Satu input layer (6), dua hidden layer (15), satu output layer (1)
Model-C	Satu input layer (6), tiga hidden layer (15), satu output layer (1)
Model-D	Satu input layer (6), empat hidden layer (15), satu output layer (1)
Model-E	Satu input layer (6), lima hidden layer (15), satu output layer (1)

Eksperimen dilakukan dengan beberapa variasi *setting hyperparameter*. Evaluasi *performance model* dilakukan dengan membandingkan hasil skor *metric performance* yang didapatkan berdasarkan:

- a. optimasi dan besarnya *learning rate* yang digunakan
- b. jumlah *epoch* yang digunakan pada setiap variasi model
- c. jumlah *hidden layer* pada model *neural network*.

Fungsi aktivasi pada input layer dan *hidden layer* menggunakan *ReLU* sedangkan untuk output layer menggunakan *Sigmoid*. Eksperimen masing-masing

akan dilakukan dengan jumlah epoch 20 dengan *batch size* 5. Variasi Model Deep Learning dengan parameter dan nilainya sebagaimana pada Tabel 3.3.

Tabel 3.3. Parameter Model

Parameter	Nilai Parameter
<i>Aktivasi Input Layer</i>	<i>ReLU (Rectified Linear Unit)</i>
<i>Aktivasi Hidden Layer</i>	<i>ReLU (Rectified Linear Unit)</i>
<i>Aktivasi Output</i>	<i>Sigmoid</i>
<i>Neuron</i>	<i>6, 15, 1</i>
<i>Epoch</i>	<i>20</i>
<i>Batch Size</i>	<i>5</i>
<i>Loss Function</i>	<i>Binary Crossentropy</i>
<i>Jenis Layer</i>	<i>Dense</i>

Berdasarkan komposisi pada Tabel tersebut, eksperimen akan dilakukan terhadap model dengan optimasi Adam. Optimasi menggunakan nilai *learning rate* 0.1, 0.01 dan 0.001 disajikan pada Tabel 3.4.

Tabel 3.4. Parameter Optimasi

Optimasi	Learning Rate
<i>Adams</i> <i>(Adaptive Moments Estimation)</i>	<i>0.1</i>
	<i>0.01</i>
	<i>0.001</i>

Teknik validasi model yang diimplementasikan pada penelitian ini menggunakan data testing untuk memvalidasi model. Data testing digunakan karena sedikitnya jumlah data pada dataset sehingga data tidak dibagi menjadi data validation.

3.2.6. Training dan Validation Model

Eksperimen dilakukan dengan bantuan *google colab*, bahasa pemrograman yang digunakan *python*. Spesifikasi komputer yang digunakan untuk *running* program sebagaimana disajikan dalam tabel 3.5.

Tabel 3.5. Spesifikasi Komputer

<i>Operating System</i>	<i>Windows 10 Pro 64-bit</i>
<i>Processor</i>	<i>Intel ® Core™ i5-2430M CPU @2.40GHZ 4CPUs, ~2.4GHz</i>
<i>Memory</i>	<i>8192MB RAM</i>
<i>Hardisk</i>	<i>1 TB</i>

3.2.7. Evaluasi Model

Setelah dilakukan eksperimen dan validasi atas model selama proses *training*, selanjutnya akan dilakukan evaluasi atas setiap model. Evaluasi model menggunakan *data testing*. Proses evaluasi menggunakan *scikit-learn libraries*. *Metrics performance* yang digunakan untuk mengevaluasi model yaitu *accuracy*, *f1-score*, *precision*, dan *recall*.

BAB 4

HASIL PENELITIAN DAN PEMBAHASAN

4.1. Hasil Eksplorasi Data

Eksplorasi data dimulai dengan mengecek dimensi data, mengecek tipe data, mengecek data yang kosong dan mengecek distribusi data label target dengan tujuan memastikan kembali terkait jumlah data dan tipe data. Adapun hasil dari eksplorasi data sebagai berikut:

1. Dimensi Data

Tabel 4.1. Dimensi Data

Dataset	Jumlah Instance	Jumlah Atribut
WBCD (Original)	683	11
WDBC (Diagnostic)	569	32
Coimbra	116	10

Berdasarkan tabel 4.1. jumlah *instance* dataset WBCD (original) sebanyak 683 dengan jumlah atribut 11. Dataset WDBC jumlah *instance* 569 dengan jumlah atribut 32 dan dataset coimbra jumlah *instance* 116 dengan jumlah atribut 10, dan dapat dinyatakan jumlah data dan atribut sudah sesuai.

2. Tipe Data

Tabel 4.2. Tipe Data

WBCD	WDBC	Coimbra	Tipe Data
ID	ID	-	int
<i>clumpthickness</i>	<i>radiusnuca</i>	<i>age</i>	<i>object</i>
<i>uniformitycellsize</i>	<i>texturenuca</i>	<i>bmi</i>	<i>object</i>
<i>uniformitycellshape</i>	<i>perimeternuca</i>	<i>glucose</i>	<i>object</i>
<i>marginaladhesion</i>	<i>areanuca</i>	<i>insulin</i>	<i>object</i>
<i>singleepithelialsize</i>	<i>smoothnessnuca</i>	<i>homa</i>	<i>object</i>
<i>barenuclai</i>	<i>compactnessnuca</i>	<i>leptin</i>	<i>object</i>
<i>chromatin</i>	<i>concavitynuca</i>	<i>adiponectin</i>	<i>object</i>
<i>nucleoli</i>	<i>concavepointsnuca</i>	<i>resistin</i>	<i>object</i>
<i>mitoses</i>	<i>symmetrynuca</i>	<i>mcp.1</i>	<i>object</i>

<i>class</i>	<i>fractaldimensionnuca</i>	<i>class</i>	<i>object</i>
...

Pada tabel 4.2 diatas, dataset WBCD memiliki atribut dengan tipe data sebanyak satu int dan sepuluh object, dataset WDBC memiliki atribut dengan tipe data sebanyak satu int dan tiga puluh satu object, dataset coimbra memiliki atribut dengan tipe data sebanyak sepuluh object.

3. Mengecek data yang kosong

Tabel 4.3. Missing Values

WBCD	WDBC	Coimbra	Missing Values
<i>clumpthickness</i>	<i>radiusnuca</i>	<i>age</i>	0
<i>uniformitycellsize</i>	<i>texturenuca</i>	<i>bmi</i>	0
<i>uniformitycellshape</i>	<i>perimeternuca</i>	<i>glucose</i>	0
<i>marginaladhesion</i>	<i>areanuca</i>	<i>insulin</i>	0
<i>singleepithelialsize</i>	<i>smoothnessnuca</i>	<i>homa</i>	0
<i>barenuclei</i>	<i>compactnessnuca</i>	<i>leptin</i>	0
<i>chromatin</i>	<i>concavitynuca</i>	<i>adiponectin</i>	0
<i>nucleoli</i>	<i>concavepointsnuca</i>	<i>resistin</i>	0
<i>mitoses</i>	<i>symmetrynuca</i>	<i>mcp.1</i>	0
<i>class</i>	<i>fractaldimensionnuca</i>	<i>class</i>	0
...

Berdasarkan hasil tabel 4.3. setelah dilakukan pengecekan data apakah masih terdapat data yang kosong. Dari hasil yang disajikan tidak terdapat data yang kosong baik di dataset WBCD, WDBC ataupun dataset Coimbra.

4. Statistik distribusi label target

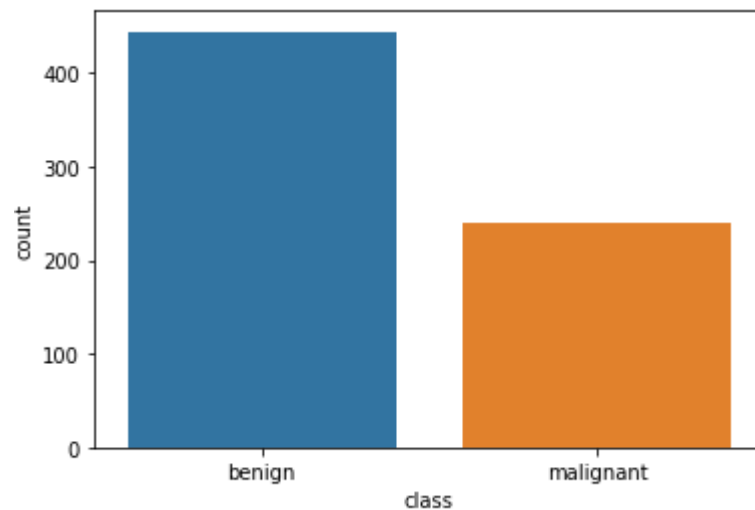
Tabel 4.4. Distribusi Label Target

Dataset	Target Class	Jumlah	Total Instance
WBCD	<i>benign</i>	444	683
	<i>malignant</i>	239	
WDBC	<i>benign</i>	357	569
	<i>malignant</i>	212	
Coimbra	<i>patients</i>	68	116

	<i>healthy</i>	48	
--	----------------	----	--

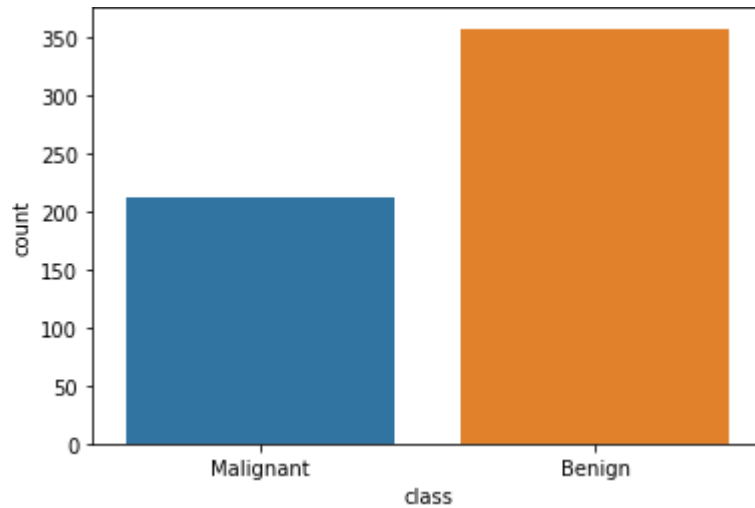
Berdasarkan tabel 4.4. Setiap dataset memiliki dua *target class*. Dataset WBCD dan WDBC memiliki dua target kelas yang sama yaitu "*malignant*" dan "*benign*". Sedangkan dataset Coimbra memiliki dua kelas target yaitu "*patients*" dan "*healthy*".

Dataset WBCD dengan target kelas "*benign*" memiliki jumlah 444 data dan target kelas "*malignant*" memiliki jumlah 239 data dengan total *instance* 683. Berikut visualisasi distribusi data kelas target dataset WBCD.



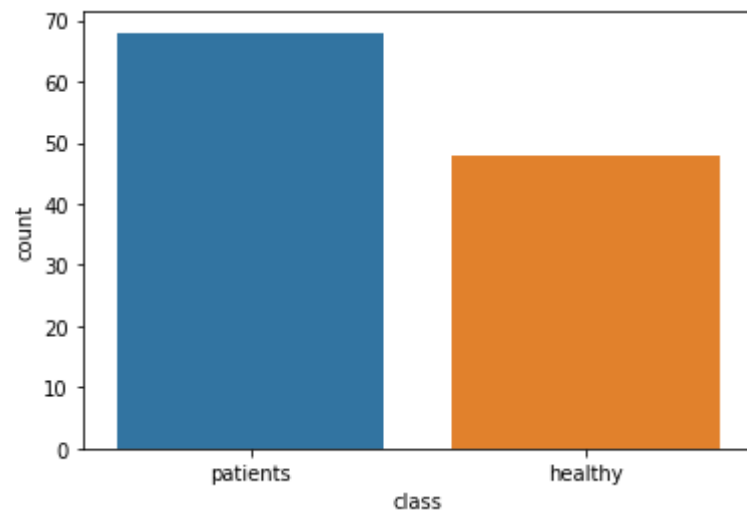
Gambar 3.2. Distribusi Kelas Target Dataset WBCD.

Dataset WDBC dengan target kelas "*benign*" memiliki jumlah 357 data dan target kelas "*malignant*" memiliki jumlah 212 data dengan total *instance* 569. Berikut visualisasi distribusi data kelas target dataset WDBC.



Gambar 3.3 Distribusi Kelas Target Dataset WDBC

Dataset Coimbra dengan target kelas “*patients*” memiliki jumlah 68 data dan target kelas “*healthy*” memiliki jumlah 48 data dengan total *instance* 116. Berikut visualisasi distribusi data kelas target dataset Coimbra.



Gambar 3.4. Distribusi Kelas Target Dataset Coimbra.

4.2. Hasil Data Preprocessing

Preprocessing dimulai dengan *cleaning data*, menghapus kolom *id* pada dataset WBCD dan WDBC. Sehingga jumlah atribut setelah dilakukan

penghapusan untuk dataset WBCD menjadi 10 atribut dan dataset WDBC menjadi 31 atribut.

Kemudian mengubah *feature class* dengan nama “*diagnosis*” pada ketiga jenis dataset, jika tidak dirubah nama *features class* akan terbaca sebagai fungsi. Kemudian melakukan *feature encoding* yang dilakukan dengan menggunakan *library label encoder* dari *scikit learn* untuk mengubah label target klasifikasi menjadi numerik. Adapun hasil dari proses *feature encoding* di sajikan pada tabel 4.5. Pada dataset WBCD dan WDBC label “*benign*” menjadi “0” dan label “*malignant*” menjadi “1”. Sedangkan dataset Coimbra label “*patients*” menjadi “1” dan label “*healthy*” menjadi “0”.

Ekstraksi fitur yang dilakukan hanya untuk atribut data yang bertipe object dan menghasilkan jumlah kolom sebanyak 28 pada dataset WBCD, jumlah kolom sebanyak 94 pada dataset WDBC, dan jumlah kolom sebanyak 28 pada dataset Coimbra. Hal ini karena ekstraksi fitur membentuk isian data pada setiap kolom atribut menjadi kolom baru yang mewakili setiap satu isi data dan atribut menjadi satu kolom dengan nilai biner yaitu 1 atau 0.

Tabel 4.5. Hasil Label Encoder

Dataset	Kategori	Label Encoder
WBCD	<i>benign</i>	0
	<i>malignant</i>	1
WDBC	<i>benign</i>	0
	<i>malignant</i>	1
Coimbra	<i>patients</i>	1
	<i>healthy</i>	0

Sepuluh *instances* data yang sudah dilakukan proses *feature encoding* dan ekstraksi fitur disajikan pada Tabel 4.6-4.8 Dibawah ini.

Tabel 4.6. Dataset WBCD setelah *Feature Encoding*

Clumpthickness_High	Clumpthickness_Low	Clumpthickness_Medium	diagnosis
0	0	1	0
0	0	1	0
0	1	0	0
0	0	1	0
0	0	1	0
1	0	0	1
0	1	0	0
0	1	0	0
0	1	0	0
0	0	0	0

Tabel 4.6. Dataset WDBC setelah *Feature Encoding*

radiusnucA_High	radiusnucA_Low	radiusnucA_Medium	diagnosis
0	0	1	1
0	0	1	1
0	0	1	1
0	1	0	1
0	0	1	1
0	0	1	1
0	0	1	1
0	0	1	1
0	0	1	1
0	0	1	1

Tabel 4.8. Dataset Coimbra setelah *Feature Encoding*

Age_High	Age_Low	Age_Medium	diagnosis
0	0	1	1
0	0	1	1
0	0	1	1
0	0	1	1
1	0	0	1
1	0	0	0
0	0	1	1
0	1	0	1
0	0	1	1
0	1	0	0

4.3. Hasil Split Data

Pembagian split data pada penelitian ini *data training* 80%, data *testing* 20%, untuk hasil split dataset WBCD pada data train dihasilkan sebanyak 546 data dan 137 data test, untuk dataset WDBC dihasilkan data train sebanyak 455 data dan 114 data test, sedangkan untuk data Coimbra dihasilkan data train sebanyak 92 dan data test sebanyak 24 data.

Tabel 4.9. Split Data

Diagnosis	WBCD		WDBC		COIMBRA	
	<i>train</i>	<i>test</i>	<i>train</i>	<i>test</i>	<i>train</i>	<i>test</i>
0	357	87	290	67	36	12
1	189	50	165	47	56	12
Total	546	137	455	114	92	24

4.4. Normalisasi Data

Tujuan dari normalisasi data untuk merubah range input data menjadi 0-1, berikut hasil dari normalisasi data pada input data WBCD.

```
array([[1., 0., 0., ..., 0., 1., 0.],
       [0., 1., 0., ..., 0., 1., 0.],
       [0., 0., 1., ..., 0., 1., 0.],
       ...,
       [0., 1., 0., ..., 0., 1., 0.],
       [0., 1., 0., ..., 0., 1., 0.],
       [1., 0., 0., ..., 0., 1., 0.]])
```

4.5. Hasil Arsitektur Model

1. Dataset WBCD

Hasil arsitektur model pada dataset WBCD untuk kelima skenario model deep learning yaitu model A pada tabel 4.10, model B pada tabel 4.11, model C pada tabel 4.12, model D pada tabel 4.13, dan model E pada tabel 4.14 beserta jumlah parameternya.

Tabel 4.10. Model A Dataset WBCD

Layer (Type)	Output Shape	Param
input (Dense)	(None, 6)	168
dense (Dense)	(None, 15)	105
output (Dense)	(None, 1)	16
Total params: 289		
Trainable params: 289		
Non-trainable params: 0		

Tabel 4.11. Model B Dataset WBCD

Layer (Type)	Output Shape	Param
input (Dense)	(None, 6)	168
dense_3 (Dense)	(None, 15)	105
dense_4 (Dense)	(None, 15)	240
output (Dense)	(None, 1)	16
Total params: 529		
Trainable params: 529		
Non-trainable params: 0		

Tabel 4.12. Model C Dataset WBCD

Layer (Type)	Output Shape	Param
input (Dense)	(None, 6)	168
dense_9 (Dense)	(None, 15)	105
dense_10 (Dense)	(None, 15)	240
dense_11 (Dense)	(None, 15)	240
output (Dense)	(None, 1)	16
Total params: 769		
Trainable params: 769		
Non-trainable params: 0		

Tabel 4.13. Model D Dataset WBCD

Layer (Type)	Output Shape	Param
input (Dense)	(None, 6)	168
dense_18 (Dense)	(None, 15)	105
dense_19 (Dense)	(None, 15)	240
dense_20 (Dense)	(None, 15)	240
dense_21 (Dense)	(None, 15)	240
output (Dense)	(None, 1)	16
Total params: 1,009		
Trainable params: 1,009		
Non-trainable params: 0		

Tabel 4.14. Model E Dataset WBCD

Layer (Type)	Output Shape	Param
input (Dense)	(None, 6)	168
dense_30 (Dense)	(None, 15)	105
dense_31 (Dense)	(None, 15)	240
dense_32 (Dense)	(None, 15)	240
dense_33 (Dense)	(None, 15)	240
dense_34 (Dense)	(None, 15)	240
output (Dense)	(None, 1)	16
Total params: 1,249		
Trainable params: 1,249		
Non-trainable params: 0		

2. Dataset WDBC

Hasil arsitektur model pada dataset WDBC untuk kelima skenario model deep learning yaitu model A pada tabel 4.15, model B pada tabel 4.16, model C pada tabel 4.17, model D pada tabel 4.18, dan model E pada tabel 4.19 beserta jumlah parameternya.

Tabel 4.15. Model A Dataset WDBC

Layer (Type)	Output Shape	Param
input (Dense)	(None, 6)	564
dense (Dense)	(None, 15)	105
output (Dense)	(None, 1)	16
Total params: 685		
Trainable params: 685		
Non-trainable params: 0		

Tabel 4.16. Model B Dataset WDBC

Layer (Type)	Output Shape	Param
input (Dense)	(None, 6)	564
dense_3 (Dense)	(None, 15)	105
dense_4 (Dense)	(None, 15)	240
output (Dense)	(None, 1)	16
Total params: 925		
Trainable params: 925		
Non-trainable params: 0		

Tabel 4.17. Model C Dataset WDBC

Layer (Type)	Output Shape	Param
input (Dense)	(None, 6)	564
dense_9 (Dense)	(None, 15)	105
dense_10 (Dense)	(None, 15)	240
dense_11 (Dense)	(None, 15)	240
output (Dense)	(None, 1)	16
Total params: 1,165		
Trainable params: 1,165		
Non-trainable params: 0		

Tabel 4.18. Model D Dataset WDBC

Layer (Type)	Output Shape	Param
input (Dense)	(None, 6)	564
dense_18 (Dense)	(None, 15)	105
dense_19 (Dense)	(None, 15)	240
dense_20 (Dense)	(None, 15)	240
dense_21 (Dense)	(None, 15)	240
output (Dense)	(None, 1)	16
Total params: 1,405		
Trainable params: 1,405		
Non-trainable params: 0		

Tabel 4.19. Model E Dataset WDBC

Layer (Type)	Output Shape	Param
input (Dense)	(None, 6)	564
dense_30 (Dense)	(None, 15)	105
dense_31 (Dense)	(None, 15)	240
dense_32 (Dense)	(None, 15)	240
dense_33 (Dense)	(None, 15)	240
dense_34 (Dense)	(None, 15)	240
output (Dense)	(None, 1)	16
Total params: 1,645		
Trainable params: 1,645		
Non-trainable params: 0		

3. Dataset Coimbra

Hasil arsitektur model pada dataset Coimbra untuk kelima skenario model deep learning yaitu model A pada tabel 4.20, model B pada tabel 4.21,

model C pada tabel 4.22, model D pada tabel 4.23, dan model E pada tabel 4.24 beserta jumlah parameternya.

Tabel 4.20. Model A Dataset Coimbra

Layer (Type)	Output Shape	Param
input (Dense)	(None, 6)	168
dense (Dense)	(None, 15)	105
output (Dense)	(None, 1)	16
Total params: 289		
Trainable params: 289		
Non-trainable params: 0		

Tabel 4.21. Model B Dataset Coimbra

Layer (Type)	Output Shape	Param
input (Dense)	(None, 6)	168
dense_3 (Dense)	(None, 15)	105
dense_4 (Dense)	(None, 15)	240
output (Dense)	(None, 1)	16
Total params: 529		
Trainable params: 529		
Non-trainable params: 0		

Tabel 4.22. Model C Dataset Coimbra

Layer (Type)	Output Shape	Param
input (Dense)	(None, 6)	168
dense_9 (Dense)	(None, 15)	105
dense_10 (Dense)	(None, 15)	240
dense_11 (Dense)	(None, 15)	240
output (Dense)	(None, 1)	16
Total params: 769		
Trainable params: 769		
Non-trainable params: 0		

Tabel 4.23. Model D Dataset Coimbra

Layer (Type)	Output Shape	Param
input (Dense)	(None, 6)	168
dense_18 (Dense)	(None, 15)	105
dense_19 (Dense)	(None, 15)	240
dense_20 (Dense)	(None, 15)	240
dense_21 (Dense)	(None, 15)	240
output (Dense)	(None, 1)	16

Total params: 1,009
Trainable params: 1,009
Non-trainable params: 0

Tabel 4.24. Model E Dataset Coimbra

Layer (Type)	Output Shape	Param
input (Dense)	(None, 6)	168
dense_30 (Dense)	(None, 15)	105
dense_31 (Dense)	(None, 15)	240
dense_32 (Dense)	(None, 15)	240
dense_33 (Dense)	(None, 15)	240
dense_34 (Dense)	(None, 15)	240
output (Dense)	(None, 1)	16
Total params: 1,249		
Trainable params: 1,249		
Non-trainable params: 0		

4.6. Hasil Training dan Validasi Model

Hasil *training* dan *validation* model untuk dataset kanker payudara yaitu dataset WBCD disajikan pada Tabel 4.25, untuk nilai *accuracy* (Acc) *training* model terbaik dihasilkan pada Model-D algoritma *deep learning* dengan *learning rate* 0,01 dan *hidden layer* empat sebesar 0.9817 dengan nilai *loss* (L) paling minimumnya sebesar 0.0637 pada *learning rate* 0.01 dan *hidden layer* empat. Kemudian nilai *validation accuracy* (VAcc) terbaik pada Model-B dengan *hidden layer* dua dan Model-E dengan *hidden layer* lima algoritma *Deep Learning* pada *learning rate* 0.01 sebesar 0.9635, dan nilai *validation loss* (VL) paling minimum terdapat pada Model-C algoritma *Deep Learning* dengan *learning rate* 0.01 dan *hidden layer* tiga sebesar 0.1106.

Tabel 4.25. Hasil Training dan Validation Model Dataset WBCD

Dataset	Arsitektur Model	Learnng Rate	Batch Size	Training		Validation	
				Acc	Loss	VAcc	VLoss
WBCD	A	0.1	20	0.9780	0.1265	0,9549	0.1375
		0.01	20	0.9780	0.0670	0.9562	0.1279

		0.001	20	0.9707	0.0796	0.9489	0.1212
	B	0.1	20	0.6538	0.6527	0.6566	0.6350
		0.01	20	0.9799	0.0781	0.9635	0.1269
		0.001	20	0.9799	0.0770	0.9562	0.1142
	C	0.1	20	0.9762	0.1024	0.9489	0.1785
		0.01	20	0.9780	0.0932	0.9562	0.1106
		0.001	20	0.9799	0.0771	0.9489	0.1122
	D	0.1	20	0.6538	0.6513	0.6350	0.6594
		0.01	20	0.9817	0.0637	0.9489	0.1860
		0.001	20	0.9780	0.0748	0.9562	0.1111
	E	0.1	20	0.6538	0.6516	0.6350	0.6630
		0.01	20	0.9707	0.0932	0.9635	0.9707
		0.001	20	0.6538	0.6457	0.6350	0.6562

Hasil *training* dan *validation* model untuk dataset kanker payudara yaitu dataset WDBC disajikan pada Tabel 4.26, untuk nilai *accuracy* (Acc) *training* model terbaik dihasilkan pada Model-A algoritma *deep learning* dengan *learning rate* 0,01 dan *hidden layer* satu sebesar 0.9890 dengan nilai *loss* (L) paling minimumnya sebesar 0.0370 pada *learning rate* 0.01 dan *hidden layer* satu. Kemudian nilai *validation accuracy* (VAcc) terbaik pada Model-A dengan *hidden layer* satu algoritma *Deep Learning* pada *learning rate* 0.01 sebesar 0.9561, dan nilai *validation loss* (VL) paling minimum terdapat pada Model-B algoritma *Deep Learning* dengan *learning rate* 0.001 dan *hidden layer* dua sebesar 0.1336.

Tabel 4.26. Hasil Training dan Validation Model Dataset WDBC

Dataset	Arsitektur Model	Learnng Rate	Batch Size	Training		Validation	
				Acc	Loss	VAcc	VLoss
WDBC	A	0.1	20	0.9758	0.0840	0.9386	0.4909
		0.01	20	0.9890	0.0370	0.9561	0.2361
		0.001	20	0.9714	0.0633	0.9474	0.9474
	B	0.1	20	0.9758	0.0904	0.9386	0.6942
		0.01	20	0.9846	0.0499	0.9474	0.2701
		0.001	20	0.9802	0.0596	0.9386	0.1336
	C	0.1	20	0.9407	0.2494	0.9211	0.9276
		0.01	20	0.9802	0.0424	0.9474	0.1716
		0.001	20	0.9868	0.0537	0.9386	0.1799
	D	0.1	20	0.6374	0.6626	0.5877	0.6824

		0.01	20	0.9824	0.0562	0.9298	0.2289
		0.001	20	0.9824	0.0578	0.9474	0.1473
	E	0.1	20	0.6374	0.6611	0.5877	0.6783
		0.01	20	0.9780	0.0614	0.9561	0.1968
		0.001	20	0.9802	0.0548	0.9474	0.1779

Hasil *training* dan *validation* model untuk dataset kanker payudara yaitu dataset Coimbra disajikan pada Tabel 4.27, untuk nilai *accuracy* (Acc) *training* model terbaik dihasilkan pada Model-C algoritma *deep learning* dengan *learning rate* 0,01 dan *hidden layer* tiga sebesar 0.8478 dengan nilai *loss* (L) paling minimumnya sebesar 0.3550 pada *learning rate* 0.01 dan *hidden layer* tiga. Kemudian nilai *validation accuracy* (VAcc) terbaik pada Model-A dengan *hidden layer* satu, Model-B dengan *hidden layer* dua, Model-C dengan *hidden layer* tiga, Model-D dengan *hidden layer* empat dari algoritma *Deep Learning* pada *learning rate* 0.01 dan 0.001 sebesar 0.7500, juga dihasilkan sebesar 0.7500 pada Model-E dengan *hidden layer* lima dan *learning rate* 0.001. dan nilai *validation loss* (VL) paling minimum terdapat pada Model-A algoritma *Deep Learning* dengan *learning rate* 0.001 dan *hidden layer* satu sebesar 0.6481.

Tabel 4.27. Hasil Training dan Validation Model Dataset Coimbra

Dataset	Arsitektur Model	Learnng Rate	Batch Size	Training		Validation	
				Acc	Loss	VAcc	VLoss
Coimbra	A	0.1	20	0.6087	0.6708	0.5000	0.7196
		0.01	20	0.8261	0.3757	0.7500	1.4904
		0.001	20	0.7609	0.4887	0.7500	0.6481
	B	0.1	20	0.6087	0.6735	0.5000	0.7304
		0.01	20	0.8152	0.3710	0.7500	1.1456
		0.001	20	0.7609	0.5093	0.7500	0.6866
	C	0.1	20	0.6087	0.6723	0.5000	0.7219
		0.01	20	0.8478	0.3550	0.7500	1.6751
		0.001	20	0.7717	0.4938	0.7500	0.6786
	D	0.1	20	0.6087	0.6699	0.5000	0.7205
		0.01	20	0.8370	0.3560	0.7500	1.2391
		0.001	20	0.7826	0.4978	0.7500	0.7226

	E	0.1	20	0.6087	0.6724	0.5000	0.7165
		0.01	20	0.8152	0.4099	0.7500	1.8653
		0.001	20	0.6087	0.5165	0.5000	0.7453

4.7. Hasil Evaluasi atau Pengujian Model

1. Hasil Akurasi

Tabel 4.28 menjelaskan hasil akurasi yang di dapat dari penelitian yang telah dilakukan dengan tiga dataset yang diujikan yaitu dataset WBCD, WDBC, dan Coimbra. dihasilkan akurasi tertinggi pada dataset WBCD adalah 96,3% pada *learning rate* 0,01 dengan Model-B *hidden layer* 2 dan Model-E *hidden layer* 5. Sedangkan pada dataset WDBC dihasilkan akurasi sebesar 95,6% berada pada *learning rate* 0,01 dengan Model-E *hidden layer* lima. Dan dari dataset Coimbra hasil akurasi tertinggi yaitu 75% pada Model-A, Model-B, Model-C dan Model-D dengan *learning rate* 0.01 dan 0.001. Pada Model-E sebesar 75% dihasilkan dari *learning rate* 0.01.

Tabel 4.28. Hasil Accuracy

Dataset	Learning Rate	Hasil Model Deep Learning (%)				
		Model-A	Model-B	Model-C	Model-D	Model-E
WBCD	0.1	94.8	63.5	94.8	63.5	63.5
	0.01	95.6	96.3	95.6	94.8	96.3
	0.001	94.8	95.6	94.8	95.6	63.5
WDBC	0.1	93.8	93.8	92.1	58.7	58.7
	0.01	95.6	94.7	94.7	92.9	95.6
	0.001	94.7	93.8	93.8	94.7	94.7
COIMBRA	0.1	50	50	50	50	50
	0.01	75	75	75	75	75
	0.001	75	75	75	75	50

2. Hasil Presisi

Tabel 4.29 menjelaskan hasil presisi yang di dapat dari penelitian yang telah dilakukan dengan tiga dataset yang diujikan yaitu dataset WBCD, WDBC, dan

Coimbra. dihasilkan presisi tertinggi pada dataset WBCD adalah 95,4% pada *learning rate* 0,01 dengan Model-B *hidden layer* 2 dan Model-E *hidden layer* 5. Sedangkan pada dataset WDBC dihasilkan presisi sebesar 95,6% berada pada *learning rate* 0,01 dengan Model-E *hidden layer* lima dan Model-A *hidden layer* satu. Dan dari dataset Coimbra hasil presisi tertinggi yaitu 83.3% pada Model-A, Model-B, dan Model-D dengan *learning rate* 0.001. Pada Model-C sebesar 83.3% dihasilkan dari *learning rate* 0.01 dan 0.001.

Tabel 4.29. Hasil Presisi

Dataset	Learning Rate	Hasil Model Deep Learning (%)				
		Model-A	Model-B	Model-C	Model-D	Model-E
WBCD	0.1	94.0	31.7	94.0	31.7	31.7
	0.01	94.6	95.4	94.7	94.0	95.4
	0.001	94.9	94.7	94.0	94.7	31.7
WDBC	0.1	93.5	94.1	91.6	29.3	29.3
	0.01	95.6	94.3	94.3	92.5	95.6
	0.001	94.5	93.5	93.5	94.5	94.5
COIMBRA	0.1	25	25	25	25	25
	0.01	78.1	78.1	83.3	78.1	78.1
	0.001	83.3	83.3	83.3	83.3	25

3. Hasil Recall

Tabel 4.28 menjelaskan hasil recall yang di dapat dari penelitian yang telah dilakukan dengan tiga dataset yang diujikan yaitu dataset WBCD, WDBC, dan Coimbra. dihasilkan recall tertinggi pada dataset WBCD adalah 97,1% pada *learning rate* 0,01 dengan Model-B *hidden layer* 2 dan Model-E *hidden layer* 5. Sedangkan pada dataset WDBC dihasilkan recall sebesar 95,3% berada pada *learning rate* 0,01 dengan Model-E *hidden layer* lima dan Model-A *hidden layer* satu. Dan dari dataset Coimbra hasil recall tertinggi yaitu 75% pada Model-A, Model-B, Model-C dan Model-D dengan *learning rate* 0.01 dan 0.001. Pada Model-E sebesar 75% dihasilkan dari *learning rate* 0.01.

Tabel 4.30. Hasil Recall

Dataset	Learning Rate	Hasil Model Deep Learning (%)				
		Model-A	Model-B	Model-C	Model-D	Model-E
WBCD	0.1	95.1	50	95.1	50	50
	0.01	96.5	97.1	96.1	95.1	97.1
	0.001	95.1	96.1	95.1	96.1	50
WDBC	0.1	93.8	93.1	92.6	50	50
	0.01	95.3	94.8	94.8	93.3	95.3
	0.001	94.5	93.8	93.8	94.5	94.5
COIMBRA	0.1	50	50	50	50	50
	0.01	75	75	75	75	75
	0.001	75	75	75	75	50

4. Hasil *f1* Score

Tabel 4.28 menjelaskan hasil *f1-score* yang di dapat dari penelitian yang telah dilakukan dengan tiga dataset yang diujikan yaitu dataset WBCD, WDBC, dan Coimbra. dihasilkan *f1-score* tertinggi pada dataset WBCD adalah 96,1% pada *learning rate* 0,01 dengan Model-B *hidden layer* 2 dan Model-E *hidden layer* 5. Sedangkan pada dataset WDBC dihasilkan *f1-score* sebesar 95,4% berada pada *learning rate* 0,01 dengan Model-E *hidden layer* lima dan Model-A *hidden layer* satu. Dan dari dataset Coimbra hasil *f1-score* tertinggi yaitu 74.2% pada Model-A, Model-B, Model-D dan Model-E dengan *learning rate* 0.01.

Table 4.31. Hasil *f1* score

Dataset	Learning Rate	Hasil Model Deep Learning (%)				
		Model-A	Model-B	Model-C	Model-D	Model-E
WBCD	0.1	94.5	39	94.5	39	38.8
	0.01	95.3	96.1	95.3	94.5	96.1
	0.001	94.5	95.3	94.5	95.3	38.8
WDBC	0.1	93.6	93.5	91.9	37	37
	0.01	95.4	94.6	94.6	92.8	95.4
	0.001	94.5	93.6	93.6	94.5	94.5
COIMBRA	0.1	33	33	33	33	33
	0.01	74.2	74.2	73.3	74.2	74.2
	0.001	73.3	73.3	73.3	73.3	33

4.8. Model Prediktif yang Diusulkan

Berikut ini rekapitulasi hasil eksperimen akurasi, presisi, recall, *f1 score* yang sudah diujikan ke tiga dataset kanker payudara WBCD, WDBC, dan Coimbra.

Table 4.32. Rekapitulasi Score

<i>Metric Performance</i>				
Dataset	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>f1 Score</i>
WBCD	96.3	95.4	97.1	96.1
WDBC	95.6	95.6	95.3	95.4
Coimbra	75	83.3	75	74.2

Hasil pengujian menunjukkan *metric Performance* tiga dataset kanker payudara, Hasil terbaik dataset WBCD memiliki *accuracy* terbaik 96.3%, *precision* 95.4%, *recall* 97.1% dan *f1 score* 96.1%. Sedangkan hasil terbaik dataset WDBC memiliki *accuracy* terbaik 95.6%, *precision* 95.6%, *recall* 95.3% dan *f1 score* 95.4%. Terakhir hasil terbaik dataset Coimbra memiliki *accuracy* terbaik 75%, *precision* 83.3%, *recall* 75% dan *f1 score* 74.2%.

4.9. Perbandingan

Perbandingan *metrics performance* model *deep learning* yang diusulkan dengan penelitian kanker payudara yang telah dilakukan sebelumnya oleh Idris dan Ismail [12] dengan algoritma *FUZZYDBD* sebagaimana disajikan pada Tabel 4.30 seperti dibawah ini:

Table 4.33. Perbandingan *Matrics Performance*

Dataset	<i>Accuracy</i>		<i>Precision</i>		<i>Recall</i>		<i>f1 Score</i>	
	Idris and Ismail (2021)	Skema yang Diusulkan	Idris and Ismail (2021)	Skema yang Diusulkan	Idris and Ismail (2021)	Skema yang Diusulkan	Idris and Ismail (2021)	Skema yang Diusulkan
WBCD	94.3	96.3	94.1	95.4	93.3	97.1	93.7	96.1

WDBC	94.5	95.6	94.2	95.6	94.0	95.3	94.1	95.4
Coimbra	70.6	75	69.7	83.3	69.0	75	69.2	74.2

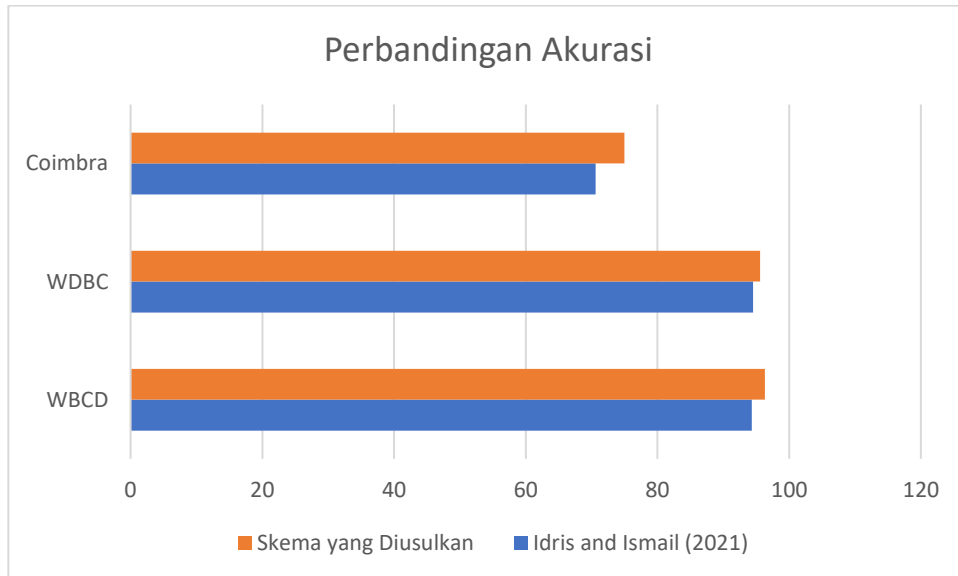
Darti tabel 4.30 menunjukkan *accuracy* pada penelitian sebelumnya pada dataset kanker payudara WBCD memiliki *accuracy* 94.3%, *precision* 94.1%, *recall* 93.3% dan *f1 score* 93.7%. Sedangkan hasil terbaik dataset WBCD pada model yang diusulkan memiliki *accuracy* terbaik 96.3%, *precision* 95.4%, *recall* 97.1% dan *f1 score* 96.1%.

Dataset kanker payudara WDBC pada penelitian sebelumnya memiliki *accuracy* 94.5%, *precision* 94.2%, *recall* 94.0% dan *f1 score* 94.1%. Sedangkan hasil terbaik dataset WDBC pada model yang diusulkan memiliki *accuracy* terbaik 95.6%, *precision* 95.6%, *recall* 95.3% dan *f1 score* 95.4%.

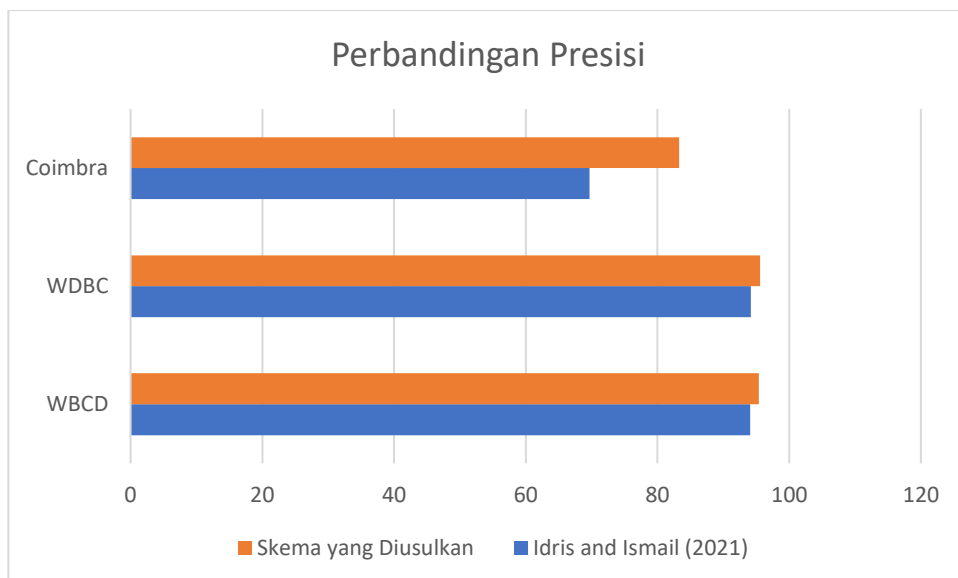
Dataset kanker payudara Coimbra pada penelitian sebelumnya memiliki *accuracy* 70.6%, *precision* 69.7%, *recall* 69% dan *f1 score* 69.2%. Sedangkan hasil terbaik dataset Coimbra memiliki *accuracy* terbaik 75%, *precision* 83.3%, *recall* 75% dan *f1 score* 74.2%.

Dari penjelasan diatas dapat disimpulkan bahwa *score* yang didapatkan oleh Model *Deep Learning* terhadap penelitian yang dilakukan sebelumnya oleh Idris dan Ismail [12] baik *accuracy*, *precision*, *recall* dan *f1-score* mendapatkan hasil yang lebih tinggi.

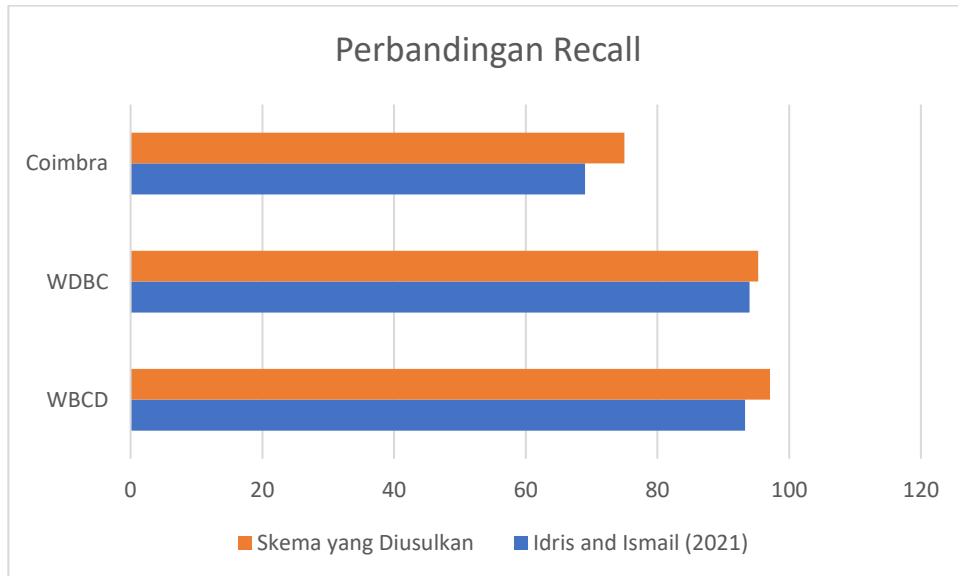
Berikut ini visualisasi perbandingan hasil *accuracy*, *precision*, *recall* dan *f1-score* model yang disulkan yaitu *Deep Learning* dengan penelitian Idris dan Ismail [12] dengan algoritma *FUZZYDBD*.



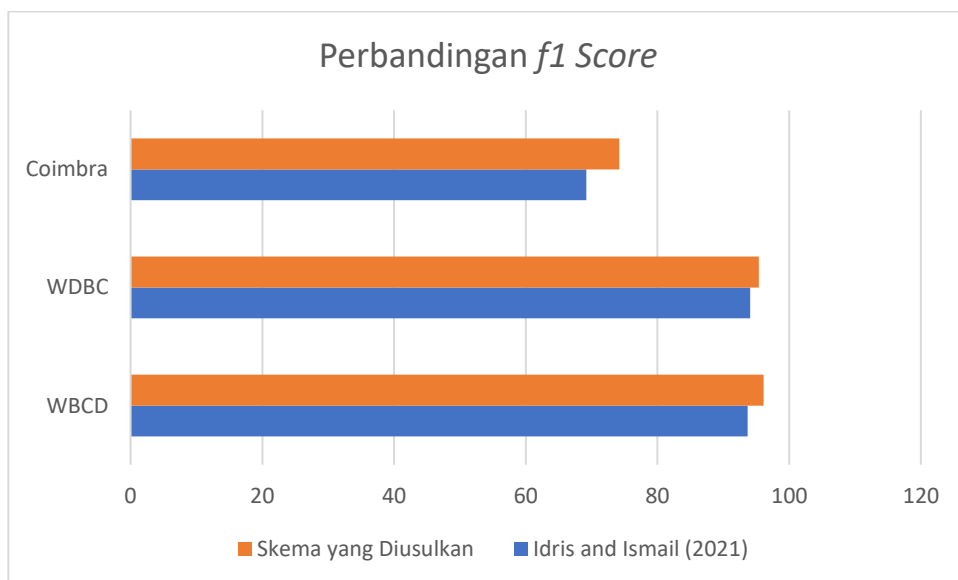
Gambar 4.1. Perbandingan hasil akurasi antara model yang diusulkan dan penelitian yang dilakukan oleh idris dan ismail [12]



Gambar 4.2. Perbandingan hasil presisi antara model yang diusulkan dan penelitian yang dilakukan oleh idris dan ismail [12]



Gambar 4.3. Perbandingan hasil *recall* antara model yang diusulkan dan penelitian yang dilakukan oleh idris dan ismail [12]



Gambar 4.4. Perbandingan hasil *f1 score* antara model yang diusulkan dan penelitian yang dilakukan oleh idris dan ismail [12]

BAB 5

PENUTUP

5.1. Kesimpulan

Berdasarkan analisis dan pembahasan hasil penelitian, dapat disimpulkan:

1. Arsitektur Model *Deep Learning* dapat digunakan untuk mengklasifikasikan penyakit kanker payudara.
2. Penggunaan optimasi Adam dan variasi *hidden layer* mendapatkan *score metrics performance* yang berbeda-beda.
3. Model *Deep Learning* yang diusulkan mendapatkan *score metric performance* terbaik dibandingkan dengan *score metric performance* penelitian sebelumnya yang dilakukan oleh Idris dan Ismail (2021) [12] menggunakan Algoritma *FUZZYDBD*.

5.2. Saran

Peneliti menyadari bahwa dalam kegiatan penelitian selalu ada ruang untuk *improvement* baik dalam dataset maupun metode yang digunakan, oleh karena itu, untuk penelitian selanjutnya, peneliti memberikan saran:

1. Penelitian klasifikasi kanker payudara yang lebih baik dan akurat, sebaiknya menambahkan jumlah data dan atribut pada *dataset*, karena pada dunia medis memiliki data yang sangat kompleks.
2. Supaya dapat diperbandingkan Model *Deep Learning* yang sudah diusulkan, sebaiknya dibuat lebih banyak lagi variasi kombinasi parameter model maupun parameter optimasi.

DAFTAR REFERENSI

- [1] H. B. Lee And W. Han, “Unique Features Of Young Age Breast Cancer And Its Management,” *J. Breast Cancer*, Vol. 17, No. 4, Pp. 301–307, 2014.
- [2] V. Chaurasia And S. Pal, “Applications Of Machine Learning Techniques To Predict Diagnostic Breast Cancer,” *SN Comput. Sci.*, Vol. 1, No. 5, Pp. 1–11, 2020.
- [3] R. A. Khan, T. Suleman, M. S. Farooq, M. H. Rafiq, And M. A. Tariq, “Data Mining Algorithms For Classification Of Diagnostic Cancer Using Genetic OPTIMIZER Algorithms,” *IJCSNS Int. J. Comput. Sci. Netw. Secur.*, Vol. 12, No. March, Pp. 207–211, 2017.
- [4] W. Majeed *Et Al.*, “Breast Cancer: Major Risk Factors And Recent Developments In Treatment,” *Asian Pacific J. Cancer Prev.*, Vol. 15, No. 8, Pp. 3353–3358, 2014.
- [5] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, And A. Jemal, “Global Cancer Statistics 2018: GLOBOCAN Estimates Of Incidence And Mortality Worldwide For 36 Cancers In 185 Countries,” *CA. Cancer J. Clin.*, Vol. 68, No. 6, Pp. 394–424, 2018.
- [6] Y. S. Sun *Et Al.*, “Risk Factors And Preventions Of Breast Cancer,” *Int. J. Biol. Sci.*, Vol. 13, No. 11, Pp. 1387–1397, 2017.
- [7] E. M. Badr, M. A. Salam, And H. Ahmed, “Optimizing Support Vector Machine Using Gray Wolf Optimizer Algorithm For Breast Cancer Detection,” Vol. 3, No. November, 2019.
- [8] N. Khuriwal And N. Mishra, “Breast Cancer Diagnosis Using Deep Learning Algorithm,” *Proc. - IEEE 2018 Int. Conf. Adv. Comput. Commun. Control Networking, ICACCCN 2018*, Pp. 98–103, 2018.
- [9] E. Zaitseva, V. Levashenko, J. Rabcan, And E. Krsak, “Application Of The Structure Function In The Evaluation Of The Human Factor In Healthcare,” *Symmetry (Basel)*, Vol. 12, No. 1, P. 93, 2020.
- [10] D. A. Omondiagbe, S. Veeramani, And A. S. Sidhu, “Machine Learning Classification Techniques For Breast Cancer Diagnosis,” *IOP Conf. Ser. Mater. Sci. Eng.*, Vol. 495, No. 1, 2019.
- [11] E. Badr, M. Abdulsalam, And H. Ahmed, “The Impact Of Scaling On Support Vector Machine In Breast Cancer Diagnosis,” *Int. J. Comput. Appl.*, Vol. 175, No. 19, Pp. 15–19, 2020.
- [12] N. F. Idris And M. A. Ismail, “Breast Cancer Disease Classification Using Fuzzy-ID3 Algorithm With FUZZYDBD Method: Automatic Fuzzy Database Definition,” *Peerj Comput. Sci.*, Vol. 7, Pp. 1–22, 2021.
- [13] A. E. Minarno And E. A. Hakim, “KLASIFIKASI CITRA THORAX X - RAY MENGGUNAKAN TRANSFER LEARNING INCEPTION V3,” Pp. 53–62, 2021.

- [14] E. Alickovic And A. Subasi, *Normalized Neural Networks For Breast Cancer Classification*, Vol. 73. Springer International Publishing, 2020.
- [15] H. Saoud, A. Ghadi, M. Ghailani, And B. A. Abdelhakim, *Using Feature Selection Techniques To Improve The Accuracy Of Breast Cancer Classification*, Vol. 1. Springer International Publishing, 2019.
- [16] B. Aisyah And Y. Sulisty, “Klasifikasi Kanker Payudara Menggunakan Algoritma Gain Ratio,” *J. Tek. Elektro*, Vol. 8, No. 2, Pp. 43–46, 2016.
- [17] A. P. Ayudhitama And U. Pujianto, “Analisa 4 Algoritma Dalam Klasifikasi Liver Menggunakan Rapidminer,” *J. Inform. Polinema*, Vol. 6, No. 2, Pp. 1–9, 2020.
- [18] Z. Mushtaq, A. Yaqub, S. Sani, And A. Khalid, “Effective K-Nearest Neighbor Classifications For Wisconsin Breast Cancer Data Sets,” *J. Chinese Inst. Eng. Trans. Chinese Inst. Eng. A*, Vol. 43, No. 1, Pp. 80–92, 2020.
- [19] J. T. Hancock And T. M. Khoshgoftaar, “Survey On Categorical Data For Neural Networks,” *J. Big Data*, Vol. 7, No. 1, 2020.
- [20] M. M. Mijwil, “Artificial Neural Networks Advantages And Disadvantages,” *Linkedin*, No. March, Pp. 1–2, 2018.
- [21] I. Mackie., *Introduction To Deep Learning*. Springer, 2018.
- [22] J. Heaton, *Artificial Intelligent For Humans*. .
- [23] Y. Ho And S. Wookey, “The Real-World-Weight Cross-Entropy Loss Function: Modeling The Costs Of Mislabeling,” *IEEE Access*, Vol. 8, Pp. 4806–4813, 2020.
- [24] L. Ghani, L. K. Mihardja, And P. Sumber Daya Dan Pelayanan Kesehatan Jl Percetakan, “Faktor Risiko Dominan Penderita Stroke Di Indonesia Faktor Risiko Dominan Penderita Stroke Di Indonesia DOMINANT RISK FACTORS OF STROKE IN INDONESIA,” *Bul. Penelit. Kesehat.*, Pp. 49–58, 2016.
- [25] S. Sharma, S. Sharma, And A. Athaiya, “Activation Functions In Neural Networks,” *Int. J. Eng. Appl. Sci. Technol.*, Vol. 04, No. 12, Pp. 310–316, 2020.
- [26] S. Loffe And Christian Sz, “Australian Literary Journalism And ‘Missing Voices’: How Helen Garner Finally Resolves This Recurring Ethical Tension,” *Journal. Pract.*, Vol. 10, No. 6, Pp. 730–743, 2016.
- [27] C. Garbin, X. Zhu, And O. Marques, “Dropout Vs. Batch Normalization: An Empirical Study Of Their Impact To Deep Learning,” *Multimed. Tools Appl.*, Vol. 79, No. 19–20, Pp. 12777–12815, 2020.
- [28] J. G. Carney And P. Cunningham, “The Epoch Interpretation Of Learning,” P. 5, 1998.

- [29] A. Sharma, "Guided Stochastic Gradient Descent Algorithm For Inconsistent Datasets," *Appl. Soft Comput. J.*, Vol. 73, Pp. 1068–1080, 2018.
- [30] D. P. Kingma And J. L. Ba, "Adam: A Method For Stochastic OPTIMIZER," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, Pp. 1–15, 2015.
- [31] K. C. Yuan, L. W. Tsai, K. S. Lai, S. T. Teng, Y. S. Lo, And S. J. Peng, "Using Transfer Learning Method To Develop An Artificial Intelligence Assisted Triaging For Endotracheal Tube Position On Chest X-Ray," *Diagnostics*, Vol. 11, No. 10, 2021.
- [32] S. L. Smith, P. J. Kindermans, C. Ying, And Q. V. Le, "Don't Decay The Learning Rate, Increase The Batch Size," *6th Int. Conf. Learn. Represent. ICLR 2018 - Conf. Track Proc.*, No. 2017, Pp. 1–11, 2018.
- [33] A. Tharwat, "Classification Assessment Methods," *Appl. Comput. Informatics*, Vol. 17, No. 1, Pp. 168–192, 2018.
- [34] T. Djatna, M. K. D. Hardhienata, And A. F. N. Masruriyah, "An Intuitionistic Fuzzy Diagnosis Analytics For Stroke Disease," *J. Big Data*, Vol. 5, No. 1, 2018.
- [35] M. M. Kumari, "The Impact Of News Information On The Stock Recommendation System : A Survey," Pp. 64–68, 2018.
- [36] M. Pyingkodi, M. M., Shanthi, S., Saravanan, T. M., Thenmozhi, K., Nanthini, K., Hemalatha, D., ... & Dhivya, "Performance Study Of Classification Algorithms Using The Breast Cancer Dataset," *Int. J. Futur. Gener. Commun. Netw.*, 2020.
- [37] R. Gondane And V. Susheela Devi, "Classification Using Probabilistic Random Forest," *Proc. - 2015 IEEE Symp. Ser. Comput. Intell. SSCI 2015*, Pp. 174–179, 2015.
- [38] S. Poorani And P. Balasubramanie, "Deep Neural Network Classifier In Breast Cancer Prediction," *Int. J. Eng. Adv. Technol.*, Vol. 9, No. 1, Pp. 2106–2109, 2019.
- [39] F. Kayaalp And M. S. Basarslan, "Performance Analysis Of Filter Based Feature Selection Methods On Diagnosis Of Breast Cancer And Orthopedics," *6th Int. Congr. Fundam. Appl. Sci. 2019*, Vol. 2019, No. July, Pp. 1–11, 2019.
- [40] N. A. Fauziyyah, S. Abdullah, And S. Nurrohmah, "Reviewing The Consistency Of The Naïve Bayes Classifier's Performance In Medical Diagnosis And Prognosis Problems," *AIP Conf. Proc.*, Vol. 2242, No. June, 2020.
- [41] V. H. Austria YD, Goh ML, Maria LS, Lalata JA, Goh JE, "Comparison Of Machine Learning Algorithms In Breast Cancer Prediction Using The Coimbra Dataset.," *Int. J. Simul. Syst. Sci. Technol.*, 2019.
- [42] H. J. Chiu, T. H. S. Li, And P. H. Kuo, "Breast Cancer–Detection System

Using PCA, Multilayer Perceptron, Transfer Learning, And Support Vector Machine,” *IEEE Access*, Vol. 8, Pp. 204309–204324, 2020.

- [43] A. S. Assiri, S. Nazir, And S. A. Velastin, “Breast Tumor Classification Using An Ensemble Machine Learning Method,” *J. Imaging*, Vol. 6, No. 6, 2020.