

**KLASIFIKASI KANKER BERDASARKAN DATA RNA MENGGUNAKAN
1D-CNN**

SKRIPSI



**UIN SUNAN AMPEL
S U R A B A Y A**

Disusun Oleh

BILQIS SINNA FAJRIN PRASETYA PUTRI

H92219043

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL
SURABAYA**

2023

PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama : BILQIS SINNA FAJRIN PRASETYA PUTRI

NIM : H92219043

Program Studi : Matematika

Angkatan : 2019

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul "KLASIFIKASI KANKER BERDASARKAN DATA RNA MENGGUNAKAN 1D-CNN". Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 12 Juli 2023

Yang menyatakan,



BILQIS SINNA FAJRIN PRASETYA PUTRI
NIM. H92219043

LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh

Nama : BILQIS SINNA FAJRIN PRASETYA PUTRI
NIM : H92219043
Judul skripsi : KLASIFIKASI KANKER BERDASARKAN DATA RNA
MENGGUNAKAN 1D-CNN

telah diperiksa dan disetujui untuk diujikan.

Pembimbing I



Ahmad Hanif Asyhar, M.Si
NIP. 198601232014031001

Pembimbing II



Lutfi Hakim, M.Ag
NIP. 197312252006041001

Mengetahui,
Ketua Program Studi Matematika
UIN Sunan Ampel Surabaya



Yuniar Farida, M.T
NIP. 197905272014032002

PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi oleh

Nama : BILQIS SINNA FAJRIN PRASETYA PUTRI
NIM : H92219043
Judul Skripsi : KLASIFIKASI KANKER BERDASARKAN DATA RNA
MENGGUNAKAN 1D-CNN

Telah dipertahankan di depan Tim Penguji
pada tanggal 12 Juli 2023

Mengesahkan,
Tim Penguji

Penguji I



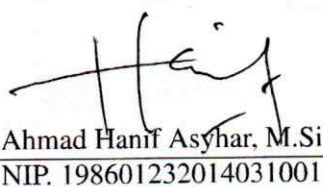
Dr. Dian Candra Rini Novitasari, M.Kom
NIP. 198511242014032001

Penguji II



Wika Dianita Utami, M.Sc
NIP. 199206102018012003

Penguji III



Ahmad Hanif Asyhar, M.Si
NIP. 198601232014031001

Penguji IV



Dr. Lutfi Hakim, M.Ag
NIP. 197312252006041001

Mengetahui,

Dekan Fakultas Sains dan Teknologi
UIN Sunan Ampel Surabaya



Dr. A. Saepul Hamdani, M.Pd
NIP. 196507312000031002



UIN SUNAN AMPEL
SURABAYA

KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
PERPUSTAKAAN

Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300
E-Mail: perpus@uinsby.ac.id

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : BILQIS SINNA FAJRIN PRASETYA PUTRI
NIM : H92219043
Fakultas/Jurusan : SAINTEK / MATEMATIKA
E-mail address : bilqissinna321@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :

Sekripsi Tesis Desertasi Lain-lain (.....)

yang berjudul :

KLASIFIKASI KANKER BERDASARKAN DATA RNA
MENGGUNAKAN 1D-CNN

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 12 JULI 2023

Penulis

(BILQIS SINNA FAJRIN P.P.)
nama terang dan tanda tangan

ABSTRAK

KLASIFIKASI KANKER BERDASARKAN DATA RNA MENGGUNAKAN 1D-CNN

Kanker termasuk penyakit yang mematikan dengan jumlah penderitanya terus meningkat tiap tahun. Sel kanker mengalami perkembangan dengan cepat menyebar ke bagian tubuh mana pun melalui aliran darah atau kelenjar getah bening akibat kelainan genetik. Ekspresi gen dipantau dengan mengukur mRNA. Ini karena urutan mRNA berhibridisasi menjadi komplemen DNA dan RNA. Penelitian ini bertujuan untuk mengklasifikasikan kanker berdasarkan data RNA menggunakan metode 1D Convolutional Neural Network. Data yang digunakan adalah data RNA dengan empat kelas yaitu BRCA (*breast adenocarcinoma*) atau kanker payudara, KIRC (*kidney renal clear cell carcinoma*) atau kanker ginjal, LUAD (*lung adenocarcinoma*) atau kanker paru-paru, dan UCEC (*uterine corpus endometrial carcinoma*) atau kanker rahim. Selanjutnya, data akan mengalami pembagian menggunakan *k-Fold Cross Validation* selanjutnya diklasifikasi menggunakan 1D-CNN dengan uji coba pembagian data, *batch size*, dan *learning rate*. Selanjutnya model optimal akan dievaluasi menggunakan matriks konfusi. Model terbaik didapatkan dengan nilai *learning rate* yaitu 0.01, nilai *k-fold* yaitu 2 dan nilai *batch size* yaitu 64. Pada model tersebut mendapatkan akurasi sebesar 99.18%, sensitivitas sebesar 98.29%, dan spesifisitas sebesar 99.48%.

Kata kunci: Ekspresi gen, Kanker, Klasifikasi, Matriks konfusi, 1D-CNN

ABSTRACT

CLASSIFICATION OF CANCER BASED ON RNA DATA USING 1D-CNN

Cancer is a deadly disease whose number of sufferers continues to increase every year. Moreover, the development of cancer cells can spread quickly to all parts of the body through the bloodstream or from the lymphatic system due to a genetic disorder. There is monitoring of gene expression by measuring from mRNA. This is because of the sequence of mRNA will hybridize with complementary DNA and RNA. The purpose of this study is to classify cancer based on RNA data using the 1D Convolutional Neural Network method. Furthermore, the data used are RNA data with four classes, namely BRCA (breast adenocarcinoma) or breast cancer, KIRC (kidney renal clear cell carcinoma) or kidney cancer, LUAD (lung adenocarcinoma) or cancer lung, and UCEC (uterine corpus endometrial carcinoma) or uterine cancer. In addition, the data will be divided using k-Fold Cross Validation and then classified by using 1D-CNN with trials of data sharing, batch size, and learning rate. Next, the optimal model will be evaluated using a confusion matrix. The best combination of parameters is obtained at the number of learning rate is 0.01, the number of k-fold is 2 and the number of batch size is 64. In this model, the accuracy reaches 99.18%, sensitivity of 98.29% and specificity of 99.48%.

Keywords: Gene expression, Cancer, Classification, Confusion matrix, 1D-CNN

UIN SUNAN AMPEL
S U R A B A Y A

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PERSETUJUAN PEMBIMBING	ii
PENGESAHAN TIM PENGUJI SKRIPSI	iii
HALAMAN PERNYATAAN KEASLIAN	iv
MOTTO	v
HALAMAN PERSEMBAHAN	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
ABSTRAK	xiii
ABSTRACT	xiv
I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	4
1.3. Tujuan Penelitian	5
1.4. Manfaat Penelitian	5
1.5. Batasan Masalah	5
1.6. Sistematika Penulisan	6
II TINJAUAN PUSTAKA	8
2.1. <i>Ribonucleic Acid (RNA)</i>	8
2.2. Kanker Payudara	9
2.3. Kanker Ginjal	10
2.4. Kanker Paru-paru	10
2.5. Kanker Rahim	11
2.6. <i>k-Fold Cross Validation</i>	11
2.7. <i>Deep Learning</i>	12

2.8. Convolutional Neural Network	13
2.8.1. Convolution Layer	13
2.8.2. Activation Layer	15
2.8.3. Fully Connected Layer	15
2.8.4. Softmax Layer	18
2.8.5. Cross Entropy	18
2.8.6. 1D Convolutional Neural Network	19
2.9. Confusion Matrix	20
2.10. Penyakit dalam Perspektif Islam	21
III METODE PENELITIAN	24
3.1. Jenis Penelitian	24
3.2. Sumber Data	24
3.3. Tahapan Penelitian	25
IV HASIL DAN PEMBAHASAN	29
4.1. Preprocessing	29
4.2. Pelatihan Model	33
4.2.1. Convolution Layer	33
4.2.2. Activation Layer	36
4.2.3. Fully Connected Layer	36
4.3. Pengujian dan Evaluasi Model	52
4.4. Penyakit dalam Perspektif Islam	59
V PENUTUP	62
5.1. Simpulan	62
5.2. Saran	62
DAFTAR PUSTAKA	63

DAFTAR TABEL

3.1 Sampel Data RNA	25
4.1 Hasil Klasifikasi Model dengan <i>Learning Rate</i> = 0.1	53
4.2 Hasil Klasifikasi Model dengan <i>Learning Rate</i> = 0.01	54
4.3 Hasil Klasifikasi Model dengan <i>Learning Rate</i> = 0.001	55
4.4 Rata-rata Akurasi berdasarkan Nilai <i>Learning Rate</i>	57



UIN SUNAN AMPEL
S U R A B A Y A

DAFTAR GAMBAR

2.1 Struktur Letak RNA	9
2.2 <i>k-Fold Cross Validation</i>	12
2.3 <i>Fully Connected Layer</i>	15
2.4 Arsitektur 1D-CNN	19
2.5 <i>Confusion Matrix</i>	20
3.1 <i>Flowchart Penelitian</i>	26
3.2 Arsitektur 1D-CNN	27
4.1 Visualisasi data RNA kanker payudara	29
4.2 Visualisasi data RNA kanker ginjal	30
4.3 Visualisasi data RNA kanker paru-paru	31
4.4 Visualisasi data RNA kanker rahim	32
4.5 Perhitungan <i>convolution layer</i> $z_{(1)}$ dengan <i>padding</i> = 1	34
4.6 Perhitungan <i>convolution layer</i> $z_{(2)}$ dengan <i>stride</i> = 1	35
4.7 Hasil <i>Convolution Layer</i>	35
4.8 Hasil Fungsi Aktivasi ReLU	36
4.9 Rata-rata Akurasi berdasarkan Nilai <i>learning Rate</i>	57
4.10 <i>Confusion Matrix</i> model terbaik	58

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Penyakit mematikan dengan penderita terus bertambah tiap tahun salah satunya adalah kanker (Ramroach et al., 2020). Dalam 30 tahun terakhir telah berkembang klasifikasi untuk mengidentifikasi jenis kanker karena penyakit kanker penyebab kematian utama manusia di berbagai negara. Di tahun 2019 tercatat terdapat 18.1 juta kasus baru kanker ditambah 9.6 juta kasus kematian karena kanker pada data World Health Organization (WHO), dan jumlah itu diperkirakan selama 20 tahun mendatang akan meningkat 70% (Adiwijaya, 2018).

Di Indonesia sendiri, pada tahun 2020 tercatat sebanyak hampir 400.000 kasus baru kanker (Prasetio and Susanti, 2019). Berdasarkan data di atas, kasus kanker di dunia dan Indonesia tergolong sangat tinggi dan akan terus meningkat. Faktor pemicu kanker tumbuh pada tubuh antara lain yaitu usia, faktor keturunan, gaya hidup, dan juga lingkungan. Gaya hidup yang tidak sehat akan menyebabkan seseorang berpotensi lebih tinggi terkena berbagai penyakit, termasuk kanker. Sehingga sangat penting untuk seseorang menerapkan pola hidup sehat.

Dalam Islam kesehatan termasuk aspek penting, karena kesehatan merupakan nikmat yang besar dari Allah sehingga setiap hamba harus mensyukuri nikmat tersebut. Terkait pentingnya kesehatan ini, terdapat dalam hadis Rasulullah yaitu:

نِعْمَتَانِ مَغْبُونٌ فِيهِمَا كَثِيرٌ مِنَ النَّاسِ ، الصَّحَّةُ وَالْفَرَاغُ

artinya: “Dua kenikmatan yang sering dilupakan oleh kebanyakan manusia adalah kesehatan dan waktu luang.” [HR. Al-Bukhari: 6412, at-Tirmidzi: 2304, Ibnu Majah: 4170]. Setiap nikmat yang diterima hamba dari Allah wajib untuk disyukuri termasuk nikmat sehat dan waktu luang. Menjaga nikmat sehat dengan cara menerapkan pola hidup yang sehat juga lingkungan yang bersih merupakan salah satu cara mensyukuri nikmat sehat. Pada zaman sekarang, banyak manusia abai akan kesehatannya, contohnya yaitu budaya malas gerak dan semua serba instan. Keduanya termasuk pola hidup yang tidak sehat. Budaya malas gerak ini sangat tidak baik untuk kesehatan. Selain itu, sering mengkonsumsi makanan instan juga berdampak pada kesehatan manusia.

Cara umum dalam pendeteksian kanker yaitu dengan tes laboratorium (darah atau urin), prosedur pencitraan seperti *CT Scan*, *X-Ray* dan lainnya, serta biopsi pada jaringan tubuh. Cara-cara tersebut memerlukan biaya yang mahal serta proses yang lama karena pemeriksaan tersebut secara manual oleh dokter.

Penelitian terkait kanker yang telah dilakukan guna pendeteksian ataupun pengklasifikasi kanker. Pada proyek penelitian *The Cancer Genome Atlas* (TCGA) yang dilakukan oleh *National Cancer Institute* (NCI) bersama *National Human Genome Research Institute* (NHGRI) menghasilkan data berupa ekspresi gen, DNA metilasi, ekspresi protein, serta RNA (Khalifa et al., 2020). Berbagai penelitian menggunakan berbagai macam metode untuk menemukan pendekatan terbaik klasifikasi kanker yaitu termasuk ekspresi gen untuk diagnosis kanker (Stricker et al., 2021). Asam deoksiribonukleat (DNA) dan asam ribonukleat (RNA) merupakan pemegang kunci informasi genetik serta peranan struktural tubuh. Sel kanker dengan cepat berkembang dan menyebar ke tubuh dengan cepat

melalui aliran darah dan sistem limfatik yang berpotensi pada kematian. Hal ini terjadi karena adanya kelainan pada gen. Dalam pembuatan protein, instruksi ditranskripsi *messenger* RNA (mRNA) (Olbrecht et al., 2021).

Dalam tubuh makhluk hidup termasuk manusia, seluruh informasi genetik disimpan pada DNA. mRNA berasal dari DNA dan ditemukan dalam inti sel dan sitoplasma. mRNA akan membawa informasi dari DNA menuju ke ribosom dan menghasilkan perubahan keadaan sel dan tingkat mRNA yang berkorelasi. Untuk mendapatkan informasi mengenai kejadian didalam keadaan sel perlu mengukur tingkat mRNA kemudian didalamnya dapat dipahami kejadian proses biologi (Cheng et al., 2018). Bioinformatika memanfaatkan komputasi dan algoritma kompleks dengan teknik *Whole Genome Sequencing* (WGS) untuk mendapatkan data genom makhluk hidup seperti RNA. Tingkat ekspresi beberapa transkrip secara bersamaan dapat diukur dengan pengurutan asam ribonukleat (RNA-Seq) yang berupa pembacaan gen, pembacaan ekson, atau pembacaan wilayah minat lainnya. Dalam komputasi yang mempelajari terkait patogen atau virus yang diklasifikasi pada beberapa kelas yang ditentukan dari data RNA-Seq dengan atribut telah banyak dikembangkan. Pemantauan penyakit tingkat molekuler, diagnosis dan klasifikasi penyakit, juga penelitian biomarker dari penyakit potensial memanfaatkan pengembangan model klasifikasi (Kabir and Ludwig, 2019). Perkembangan dalam klasifikasi berbasis pembelajaran mesin dan kemajuan dalam bioinformatika akan memberikan alat yang kuat guna untuk proses klasifikasi data RNA-Seq dengan informasi transkriptom yang tersedia (Chantsalnyam and Tayara, 2020).

Pengembangan klasifikasi dalam komputasi sangat beragam, contohnya dalam *deep learning* yang memiliki algoritma seperti jaringan saraf *neuron* tiruan.

Pada penelitian tentang klasifikasi kanker menggunakan RNA dengan berbagai metode *deep neural network*, *convolutional neural network* (CNN) memiliki akurasi tertinggi sebesar 97.81% (Mamun and Mondal, 2019). Rukhsar, dkk juga melakukan penelitian mengenai RNA-Seq untuk klasifikasi kanker menggunakan *deep learning*. Pada penelitian tersebut dilakukan perbandingan antara metode CNN, AlexNet, GoogleNet, VGG16, VGG19, ResNet50, ResNet101, dan ResNet152. Hasil akurasi tertinggi diperoleh CNN sebesar 97% (Rukhsar et al., 2022). Metode 1D-CNN juga digunakan pada penelitian klasifikasi kaki normal, kaki cavus, dan kaki planus. Penelitian tersebut menggunakan 80 data untuk di klasifikasi menjadi 3 kelas dengan metode 1D-CNN didapatkan akurasi tertinggi sebesar 95% (Mei et al., 2020). Pada penelitian yang dilakukan oleh Dan Li, dkk mengenai sinyal EKG untuk klasifikasi penyakit jantung menjadi 5 kelas menggunakan metode 1D-CNN. Penelitian tersebut mendapatkan akurasi tertinggi sebesar 97.5% (Li et al., 2017).

Penggunaan data RNA (Ribonucleic Acid) untuk klasifikasi dengan 1D CNN (Convolutional Neural Network) telah menjadi pilihan yang populer dalam analisis data biologis dan bioinformatika. Alasan utama untuk menggunakan data RNA dengan 1D CNN adalah karena data RNA berbentuk numerik sehingga bentuk data tersebut cocok digunakan untuk klasifikasi dengan 1D-CNN. Berdasarkan penelitian pentingnya klasifikasi kanker, keefektifan penggunaan data RNA dalam pengklasifikasian kanker, dan keunggulan metode 1D-CNN untuk klasifikasi sehingga akan dilakukan klasifikasi kanker dengan metode 1D-CNN yang berdasarkan pada data RNA kanker dengan tujuan membantu klasifikasi kanker yang dimana nilai akurasinya dapat dihitung sehingga dapat dimanfaatkan guna pendiagnosaan dini terhadap kanker yang kemudian dapat evaluasi dengan

pemeriksaan lanjutan.

1.2. Rumusan Masalah

Berdasarkan paparan dari latar belakang diatas, didapat beberapa rumusan masalah yaitu:

1. Bagaimana proses klasifikasi kanker menggunakan metode 1D-CNN berdasarkan data RNA?
2. Bagaimana model optimal dari 1D-CNN dalam klasifikasi kanker menggunakan data RNA?

1.3. Tujuan Penelitian

Dilihat pada rumusan masalah yang telah didapatkan, tujuan dari peneliti dalam melakukan penelitian ini yaitu sebagai berikut:

1. Untuk mengetahui proses klasifikasi kanker menggunakan metode 1D-CNN berdasarkan data RNA.
2. Untuk mengetahui model optimal dari 1D-CNN dalam klasifikasi kanker menggunakan data RNA.

1.4. Manfaat Penelitian

Manfaat dari penelitian ini adalah untuk memberikan pengetahuan baru mengenai pengklasifikasian kanker menggunakan 1D-CNN berdasarkan data RNA. Adanya penelitian ini, diharapkan semakin banyak informasi yang tersedia, juga semakin mudah dalam mempelajari ilmu terkait klasifikasi. Selain itu, penelitian ini bermanfaat dalam pengembangan penelitian di dunia kesehatan dan

membantu pengembangan klasifikasi kanker sehingga dapat lebih mempercepat pencegahan dan pengobatan serta mengurangi kasus kanker di tahun berikutnya.

1.5. Batasan Masalah

Batasan masalah dalam penelitian ini yaitu:

1. Data yang digunakan dalam proses klasifikasi yaitu dataset RNA kanker yang didapat dari Mendeley dengan atribut 4 kelas yaitu BRCA atau kanker payudara, KIRC atau kanker ginjal, LUAD atau kanker paru-paru, dan UCEC atau kanker rahim.
2. Parameter dalam proses klasifikasi yang akan digunakan yaitu pembagian data, *bacth size*, dan *learning rate*.

1.6. Sistematika Penulisan

Sistematika penulisan pada penelitian ini yaitu terdapat tiga bagian yaitu:

1. BAB I PENDAHULUAN

Pada bab ini terdiri dari latar belakang masalah terkait penyakit kanker dan pendeteksiannya melalui RNA yang dikuatkan dengan penelitian terdahulu, rumusan masalah, tujuan penelitian serta manfaat penelitian, kemudian batasan masalah, serta sistematika penulisan.

2. BAB II TINJAUAN PUSTAKA

Pada bab ini terdiri dari literatur terkait RNA, penjelasan singkat mengenai 4 kanker yaitu payudara, ginjal, paru-paru, dan rahim, kemudian proses normalisasi data dan algoritma metode 1D-CNN, serta metode evaluasi model.

3. BAB III METODE PENELITIAN

Pada bab ini berisi terkait sumber data, jenis serta alur penelitian. Alur penelitian dimulai dari pengumpulan data RNA manusia yang memiliki penyakit kanker, kemudian masuk pada proses penelitian dengan input data, selanjutnya tahap pelatihan dan pengujian serta analisis model terbaik yang didapatkan.

4. BAB IV HASIL DAN PEMBAHASAN

Bagian dari bab ini berisikan mengenai hasil yang didapatkan dari proses hasil uji coba klasifikasi kanker menggunakan metode 1D-CNN beserta pembahasannya terkait perhitungan metode 1D-CNN pada rumusan masalah yang sudah disebutkan, uraian tahapan pada penelitian serta analisis model terbaik dengan mengamati hasil akurasi, sensitivitas dan spesifisitas yang dihasilkan. Pembahasan integrasi keislaman mengenai penyakit kanker dan kesimpulan akhir penelitian.

5. BAB V PENUTUP

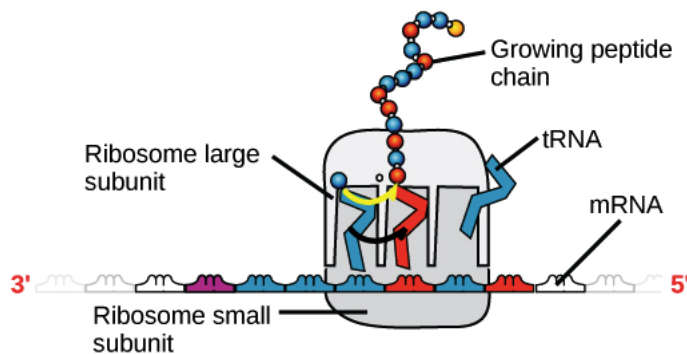
Bab ini berisi kesimpulan dari hasil penelitian yang telah dilakukan dan saran untuk penelitian selanjutnya.

BAB II

TINJAUAN PUSTAKA

2.1. *Ribonucleic Acid (RNA)*

Ribonucleic Acid atau asam ribonukleat biasa disingkat RNA adalah makromolekul dengan fungsi untuk menyimpan dan menyalurkan informasi genetik (Stricker et al., 2021). RNA akan melakukan konversi informasi genetik dari gen yaitu DNA ke dalam untaian asam amino protein yang merupakan hasil dari proses translasi. Proses translasi adalah proses dimana urutan nukleotida menjadi urutan asam amino pada protein yang merupakan hasil dari transkripsi DNA menjadi mRNA. Setiap asam amino dikodekan melalui urutan tiga nukleotida yang disebut kodon. Proses menentukan urutan awal dari protein ditandai dengan adanya kodon start (AUG) dan diakhiri dengan kodon stop (UAA, UAG, atau UGA). *Reads Per Kilobase Million (RPKM)* merupakan satuan saat pembacaan gen pada DNA yang merupakan kecepatan RNA. Indikasi bahwa terdapat kanker pada tubuh manusia yaitu terjadinya perubahan ekspresi gen pada RNA. RNA mempunyai fungsi onkogen dan *tumor suppressor gene*. 2 hal ini adalah sesuatu yang berlawanan. Gen yang meningkatkan keganasan sel tumor disebut onkogen, kebalikannya yaitu gen yang dapat mencegah perkembangan tumor disebut *tumor suppressor gene* (Elbashir and Ezz, 2019). Hasil dari susunan RNA akan berbeda-beda bergantung pada teknik dan aturan tiap teknik yang digunakan pada pengambilan RNA.



Gambar 2.1 Struktur Letak RNA

(Sumber: (Hou et al., 2021))

Setelah mRNA diproduksi, mRNA akan berasosiasi dengan ribosom. Ribosom akan merakit protein dari asam amino dengan menggunakan informasi dalam mRNA untuk membuat protein dengan urutan tertentu, membaca nukleotida mRNA dalam kodon dan menambahkan asam amino tertentu bagi setiap kodon (Alam et al., 2020).

2.2. Kanker Payudara

Sel-sel ganas yang tumbuh pada jaringan payudara disebut kanker payudara. Kanker dapat tumbuh pada saluran atau kelenjar payudara, jaringan ikat payudara, dan jaringan lemak (Liu et al., 2018). Secara umum, akan ada benjolan pada payudara yang menimbulkan nyeri. Kanker payudara menempati peringkat dua penderita terbanyak di Indonesia setelah kanker rahim menurut hasil patologi. Berbagai cara pengobatan kanker payudara antara lain yaitu operasi pengangkatan sel tumor, kemoterapi, dan radioterapi, serta terapi *hormone* (dengan *tamoxifen* atau *aromatase inhibitor*) (Yu et al., 2020).

Jenis kanker payudara ada 2 yaitu bersifat invasif dan bersifat non invasif. Kanker dengan penyebaran melalui kelenjar getah bening atau aliran darah yang menyebabkan organ lain juga terkena kanker disebut kanker payudara invasif.

Sementara kanker payudara non invasif tergolong jinak yang jarang muncul benjolan dan dapat terdeteksi dengan mamografi. Berbagai cara mendeteksi kanker payudara, antara lain dengan SADARI (Periksa Payudara Sendiri) kemudian melalui pemeriksaan rontgen yang menghasilkan citra mammograf dan kamera thermal yang menghasilkan image termograf (Normalisa, 2018).

2.3. Kanker Ginjal

Sel-sel ganas yang tumbuh pada ginjal disebut kanker ginjal. Kanker ginjal dapat berupa tumor primer dimana tumor tersebut tumbuh di ginjal atau tumor sekunder yaitu tumor dari organ lainnya di tubuh. Banyak faktor yang dapat menyebabkan pemicu dari kanker ginjal yaitu kebiasaan buruk seperti sering mengkonsumsi alkohol, jarang minum air putih dan tidak berolahraga. Saat sistem kekebalan tubuh manusia menurun, menjadi lebih rentan bagi bakteri dan mikroorganisme penyebab penyakit untuk dengan mudah memasuki tubuh (Pakan, 2019). Pria lebih besar kemungkinan terjangkit penyakit kanker ginjal 2 kali lipat dibanding wanita menurut hasil statistik dan sering terjadi pada negara berkembang (Wulan et al., 2022).

2.4. Kanker Paru-paru

Salah satu penyakit ganas yaitu kanker paru-paru. Berbagai faktor yang mempengaruhinya, yaitu seperti perokok aktif maupun pasif, polusi udara yang tercemar zat berbahaya, dan lingkungan yang tidak sehat, keturunan keluarga yang menderita kanker paru, dan pernah melakukan radioterapi udara (Fajri and Anifah, 2018). Pada pasien perokok pasif, kanker ini tumbuh setelah terpapar bahan kimia berbahaya atau asap rokok di lingkungannya. Proses pendeteksian kanker dapat dilakukan dengan citra digital ataupun data patologi, dan juga data RNA. Pada

klasifikasi data RNA-Seq akan mengklasifikasikan berdasarkan atribut dalam beberapa kelas (Prasetio and Susanti, 2019).

2.5. Kanker Rahim

Kanker rahim adalah jenis tumor ganas di bidang ginekologi yang berkembang di rongga perut atau dinding rahim (Yuningsih et al., 2019). Faktor penyebabnya antara lain obesitas, menopause, dan diabetes. Gejala dari kanker rahim seperti nyeri panggul juga nyeri pada saat buang air kecil, bahkan terjadinya pendarahan yang abnormal. Penyakit ini disebabkan oleh ketidakseimbangan antara hormon progesteron dan hormon estrerogen.

2.6. *k-Fold Cross Validation*

k-Fold Cross Validation merupakan metode untuk mengevaluasi efektivitas model yang telah dibangun. Evaluasi dilakukan dengan menggunakan sampel acak yang berperan sebagai data uji, dengan tujuan untuk mengukur performa model tersebut. Data yang dimiliki akan di partisi sebanyak k dengan ukuran yang kira-kira sama. Model kemudian dilatih pada partisi $k - 1$ dan diuji pada partisi yang tersisa (Lyu et al., 2022). Ilustrasi dari teknik ini dapat dilihat pada Gambar 2.2



Gambar 2.2 k -Fold Cross Validation

(Sumber: (Lyu et al., 2022))

Pada Gambar 2.2 terlihat bagaimana gambaran pada pembagian data dengan k -fold. Pada percobaan pertama, subset D_1 dijadikan sebagai data pengujian dan subset lainnya dijadikan sebagai data pelatihan, pada percobaan kedua subset D_1, D_3, \dots, D_k menjadi data pelatihan dan D_2 menjadi data pengujian, dan seterusnya.

2.7. Deep Learning

Deep Learning merupakan bagian dari *machine learning* dengan bentuk berupa tumpukan karena terdiri dari beberapa lapisan (Mamun and Mondal, 2019). Lapisan ini merupakan suatu algoritma atau metode yang memproses data untuk hasil terbaik. Dalam *deep learning*, proses klasifikasi diproses pada komputer secara langsung dari Gambar atau suara. Permasalahan dengan data berskala besar dan kompleks sangat cocok menggunakan metode *deep learning*. Namun berakibat pada waktu yang cenderung lama pada proses pelatihan.

2.8. Convolutional Neural Network

Convolutional Neural Network atau biasa disebut CNN merupakan pengembangan metode *Multilayer Perceptron* (MLP) yang merupakan bagian dari *deep learning* (Tahir et al., 2019). CNN (*Convolutional Neural Network*) dapat digunakan dalam klasifikasi data terlabel menggunakan metode (*supervised learning*). CNN bekerja dengan cara melatih data yang memiliki variabel target, sehingga dapat mengelompokkan data ke dalam kelas yang ada (Tasdelen and Sen, 2021). Lapisan pada *convolutional neural network* terdiri dari susunan 3D yaitu lebar yang merupakan ukuran lapisan, tinggi juga merupakan ukuran lapisan serta kedalaman yang merupakan jumlah lapisan. Komponen-komponen utama pada *convolutional neural network* yaitu *Convolution Layer*, *Pooling Layer*, *Fully Connected Layer* dan *Dropout*.

2.8.1. Convolution Layer

Konvolusi dapat diartikan sebagai operasi dua fungsi yaitu *input* dan *filter* (kernel atau bobot) yang menghasilkan matriks baru yang disebut *feature map* (Li et al., 2017). Selanjutnya, terjadi proses pergeseran *filter* sebesar s yang biasa disebut *stride* menuju posisi selanjutnya. Perhitungan akan terus berjalan sampai *filter* menutup seluruh nilai *input*. Terdapat 3 bagian pada *convolution layer* yaitu *input* atau nilai masukan, kemudian bobot atau *filter*, yang terakhir yaitu nilai keluaran berupa *feature map*. Terdapat beberapa istilah dalam CNN yaitu, faktor dilasi (f) yang merupakan parameter yang mempengaruhi jarak antar elemen filter saat konvolusi, kemudian istilah *stride* (s) yang merupakan besarnya pergeseran filter, dan istilah *padding* juga dipakai dalam 1D-CNN dimana *padding* adalah proses penambahan nilai 0 di setiap tepi data *input* dengan tujuan untuk menjaga

ukuran *output* dari *convolution layer* agar sama dengan ukuran input. Ukuran *output* dapat dihitung dengan Persamaan 2.1 (Widadi et al., 2020).

$$u_z = \frac{m - (d + f) + 2p}{s} + 1 \quad (2.1)$$

Dengan u adalah ukuran *output*, m adalah ukuran *input*, d adalah faktor dilasi, f adalah ukuran *filter*, p adalah ukuran *padding*, dan s merupakan besar *stride*.

Lalu nilai matriks z didapatkan dengan Persamaan sebagai berikut:

$$z_{i,j} = f_{i,j} \times i_{i,j} \quad (2.2)$$

Nilai *output* pada proses *convolution* berasal dari perkalian antara nilai *filter* dengan nilai *input*. Perkalian ini bergeser secara bertahap sesuai *stride* dan hasilnya dijumlahkan untuk menghasilkan satu nilai matriks *output* (Widadi et al., 2020). Persamaan 2.3 menyatakan operasi pada proses *convolution*.

$$z_{i,j} = \sum_{k=0}^{f-1} w_{i,j} \times i_{i,j} \quad (2.3)$$

Dengan z adalah nilai *outputnya*, f merupakan ukuran *filter* yang digunakan, w adalah matriks *filter*, i adalah matriks *input*. Asumsi untuk Persamaan 2.3 adalah untuk konvolusi satu dimensi. Sedangkan untuk konvolusi pada dua dimensi pada Persamaan 2.4 dengan *filter* $w_{(a,b)}$ dan data *input* $i_{(m,n)}$

$$z_{(m,n)} = \sum_{a=0}^{f-1} \sum_{b=0}^{f-1} w_{(a,b)} i_{(m+da,n+db)}, d > 0 \quad (2.4)$$

2.8.2. Activation Layer

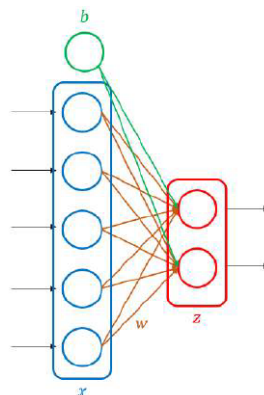
Activation layer adalah lapisan yang berisi beberapa fungsi aktivasi yang menentukan apakah nilai dari jaringan sebelumnya diaktifkan atau tidak. Pada penelitian ini menggunakan fungsi aktivasi ReLU pada Persamaan 2.5 dan 2.6.

$$y = \max(0, x) \quad (2.5)$$

$$\frac{\partial y}{\partial x} = \begin{cases} 0, & x \leq 0 \\ 1, & x > 0 \end{cases} \quad (2.6)$$

2.8.3. Fully Connected Layer

Dalam *fully connected layer* terdapat lapisan-lapisan. Lapisan dalam *fully connected layer* terdiri dari node/neuron dimana node/neuron tersebut terkoneksi penuh dengan layer sebelumnya. Hasil dari proses sebelumnya akan diolah menggunakan *multilayer perceptron* untuk mencapai hasil yang sesuai (Liu et al., 2018). Hasil *output activation layer* akan menjadi *input* pada *fully connected layer* berupa *feature map*. Pada Gambar 2.3 merupakan ilustrasi *fully connected layer*.



Gambar 2.3 Fully Connected Layer

Metode *fully connected layer* awalnya melihat kemudian mempelajari kombinasi dari fitur lapisan sebelumnya, kemudian akan menghasilkan rangkaian fitur-fitur dimana fitur tersebut termasuk ke dalam kelas tertentu. Penggunaan *multilayer perceptron* umumnya dengan metode *backpropagation* dimana terdapat propagasi maju dan propagasi mundur. Langkah-langkah pada *layer* ini sebagai berikut:

1. Langkah 1: Inisialisasi nilai bobot awal dengan rentang 0 sampai 1.
2. Propagasi maju (*feedforward*)

Pada propagasi maju, rumus untuk mencari keluaran pada *layer* ini ditunjukkan pada Persamaan 2.7 dengan x merupakan hasil *output* dari *layer* sebelumnya serta nilai bias b_1 dan bobot W yang telah diinisialisasi sebelumnya.

$$znet_j = b_{1j} + \sum_{i=1}^n x_i w_{i,j} \quad (2.7)$$

Dengan b_{1j} merupakan bobot bias ($j = 1, 2, 3, \dots, p$), x_i merupakan nilai *input* ke- i ($i = 1, 2, 3, \dots, n$), $w_{i,j}$ merupakan bobot. Setelah mendapatkan nilai $znet_j$, akan dilakukan fungsi aktivasi menggunakan ReLU dengan rumus pada Persamaan 2.8.

$$z_j = \max(0, znet_j) \quad (2.8)$$

Kemudian menghitung *layer* selanjutnya menggunakan Persamaan 2.9 dengan x merupakan hasil keluaran dari *layer* z serta nilai bias b_2 dan bobot V yang telah diinisialisasi sebelumnya..

$$ynet_j = b_{2j} + \sum_{i=1}^n x_i v_{i,j} \quad (2.9)$$

Dengan b_{2j} merupakan bobot bias ($j = 1, 2, 3, \dots, p$), x_i merupakan nilai *input* ke- i ($i = 1, 2, 3, \dots, n$), $v_{i,j}$ merupakan bobot.

3. Propagasi mundur

Menghitung kesalahan pada layer sebelumnya menggunakan Persamaan [2.10](#).

$$\left(\frac{\partial J}{\partial z_i}\right) = y_i - t_i \quad (2.10)$$

Kemudian menghitung perubahan bobot dan bias dengan Persamaan [2.11](#) dan [2.12](#).

$$\Delta v_{jk} = \alpha \frac{\partial J}{\partial z_j} z_k \quad (2.11)$$

$$\Delta b_{2j} = \alpha \frac{\partial J}{\partial z_j} b_j \quad (2.12)$$

Menghitung kesalahan pada layer sebelumnya menggunakan Persamaan [2.13](#) dan [2.14](#).

$$snet_k = \sum_{j=1}^c \frac{\partial J}{\partial z_j} v_{jk} \quad (2.13)$$

$$s_k = snet_k \frac{\partial z_k}{\partial x} \quad (2.14)$$

Kemudian menghitung perubahan bobot dan bias dengan Persamaan [2.15](#) dan [2.16](#).

$$\Delta w_{ki} = \alpha s_k x_i \quad (2.15)$$

$$\Delta b_{1k} = \alpha s_k b_{1k} \quad (2.16)$$

4. Perubahan bobot

Menghitung semua perubahan nilai bobot pada masing-masing bagian

menggunakan Persamaan [2.17](#), [2.18](#), [2.19](#), dan [2.20](#).

$$w_{ki} = w_{ki} + \Delta w_{ki} \quad (2.17)$$

$$b_{1k} = b_{1k} + \Delta b_{1k} \quad (2.18)$$

$$v_{jk} = v_{jk} + \Delta v_{jk} \quad (2.19)$$

$$b_{2j} = b_{2j} + \Delta b_{2j} \quad (2.20)$$

2.8.4. Softmax Layer

Fungsi *softmax* adalah salah satu fungsi aktivasi yang digunakan dalam arsitektur CNN untuk menghitung probabilitas dari suatu vektor yang menghasilkan keluaran berupa nilai dalam rentang antara 0 dan 1, dengan jumlah probabilitas sama dengan 1. Fungsi *softmax* dapat dihitung menggunakan Persamaan [2.21](#) dan [2.22](#) dengan c merupakan jumlah kelas.

$$y_i = \frac{e^{z_i}}{\sum_{j=1}^c e^{z_j}} \quad (2.21)$$

$$\frac{\partial y_i}{\partial z_j} = \begin{cases} y_i(1 - y_i), & i = j \\ -(y_i)(y_j), & i \neq j \end{cases} \quad (2.22)$$

2.8.5. Cross Entropy

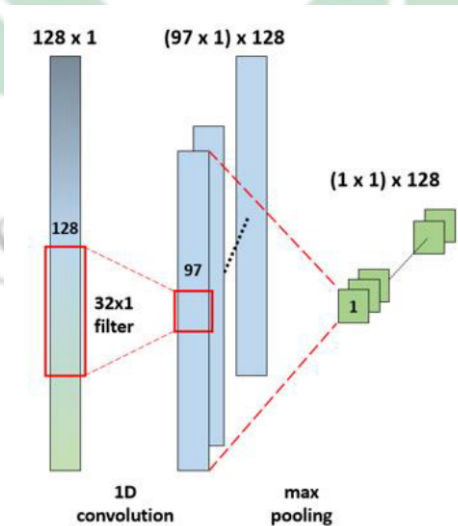
Cross entropy adalah fungsi yang digunakan untuk mengatur bobot dalam *neural network* untuk membuat model yang lebih baik. Fungsi ini menghitung rasio error antara hasil *neural network* dan objek yang diinginkan. Hasil ini kemudian digunakan sebagai indikasi seberapa besar bobot jaringan yang perlu

diperbarui. Jaringan memperbarui bobot hingga mendapatkan keluaran dengan tingkat kesalahan terendah (He et al., 2021). Apabila hasil keluaran sebanyak h dari jaringan *neural network* disimbolkan sebagai y dan target disimbolkan dengan t , maka fungsi *cross entropy* J dapat ditulis seperti pada Persamaan 2.23.

$$J = - \sum_{j=1}^h t_j \log y_j \quad (2.23)$$

2.8.6. 1D Convolutional Neural Network

Convolutional 1 dimension biasa diterapkan pada data dengan bentuk numerik atau vektor seperti sinyal, *time series*, dan lain-lain (Abdoli et al., 2019). Nilai *output* dari 1 dimensi *convolutional neural network* dapat dihitung dengan Persamaan ???. Gambar 2.4 merupakan arsitektur dari 1 dimensi *convolutional neural network*.

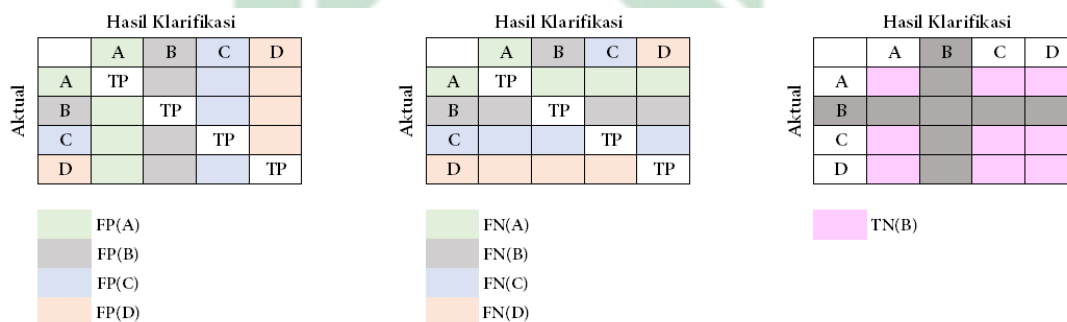


Gambar 2.4 Arsitektur 1D-CNN

(Sumber: (Novitasari et al., 2020))

2.9. Confusion Matrix

Setelah mendapat hasil klasifikasi akan dilakukan proses pengujian menggunakan *confusion matrix* dengan 3 parameter yaitu akurasi, sensitivitas, dan spesifisitas untuk mengetahui seberapa baik kerja sistem dalam melakukan proses klasifikasi (Novitasari et al., 2020). Kolom dan baris pada *confusion matrix* mewakili hasil klasifikasi dimana hasil kelas prediksi diwakilkan oleh kolom dan hasil kelas sebenarnya diwakilkan oleh baris. *Confusion matrix* untuk klasifikasi 4 kelas dengan ukuran 4×4 ditampilkan pada Gambar 2.5



Gambar 2.5 Confusion Matrix

Untuk memperoleh nilai akurasi, sensitivitas, dan spesifisitas perlu menemukan nilai dari TP (*True Positive*), TN (*True Negative*), FP (*False Positive*), dan FN (*False Negative*). Selanjutnya menghitung nilai akurasi, sensitivitas, dan spesifisitas dengan l merupakan jumlah kelas dan i merupakan identitas kelas untuk kelas ke- i .

1. Akurasi

Akurasi adalah suatu nilai yang digunakan untuk mengukur tingkat keakuratan dalam suatu sistem deteksi atau klasifikasi (Yu et al., 2020). Jika sistem pendeteksian akurat maka sistem dapat mendeteksi dengan baik. Sebaliknya jika sistem pendeteksian tidak akurat maka sistem tersebut dikatakan tidak dapat mendeteksi dengan baik. Untuk menghitung nilai

akurasi menggunakan Persamaan [2.24](#).

$$Akurasi = \frac{\sum_{i=1}^l TP_i}{n_{all}} \times 100\% \quad (2.24)$$

2. Sensitivitas

Sensitivitas adalah nilai dari jumlah data yang bernilai positif dan terdeteksi secara benar sehingga masuk dalam kelas positif ([Mei et al., 2020](#)). Apabila sensitivitas bernilai tinggi maka sistem tersebut dapat mengklasifikasi kanker dengan baik. Untuk menghitung nilai sensitivitas menggunakan Persamaan [2.25](#).

$$Sensitivitas = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l (TP_i + FN_i)} \times 100\% \quad (2.25)$$

3. Spesifisitas

Spesifisitas adalah nilai dari jumlah data yang bernilai negatif dan terdeteksi masuk dalam kelas negatif ([Mei et al., 2020](#)). Apabila spesifisitas bernilai tinggi maka sistem tersebut dapat mengklasifikasi kelas negatif dengan baik. Untuk menghitung nilai spesifisitas menggunakan Persamaan [2.26](#).

$$Spesifisitas = \frac{\sum_{i=1}^l TN_i}{\sum_{i=1}^l (TN_i + FP_i)} \times 100\% \quad (2.26)$$

2.10. Penyakit dalam Perspektif Islam

Allah Swt senantiasa menguji hambaNya dengan ujian-ujian sesuai kemampuannya. Ujian-ujian tersebut akan berbeda untuk setiap hamba. Ada yang diuji dengan harta, kedudukan, kesehatan dan lain sebagainya. Dalam menghadapi ujian Allah, baik bagi seorang hamba untuk senantiasa bersabar. Sakit juga dapat

diartikan sebagai bentuk peringatan Allah kepada hambaNya agar senantiasa mengingat segala dosa-dosa selama hidup. Kondisi sehat dan kondisi sakit adalah dua kondisi yang senantiasa dialami oleh setiap manusia. Terkait dengan ujian dari Allah, disebutkan dalam surah Al-Baqarah ayat 155, Allah Swt berfirman :

وَلَنَبْلُوَنَّكُمْ بِشَيْءٍ مِّنَ الْخَوْفِ وَالْجُوعِ وَنَقْصٍ مِّنَ الْأَمْوَالِ وَالْأَنْفُسِ وَالثَّمَرَاتِ ۗ
وَبَشِّرِ الصَّابِرِينَ

Artinya: "Dan sungguh akan Kami berikan cobaan kepadamu, dengan sedikit ketakutan, kelaparan, kekurangan harta, jiwa dan buah-buahan. Dan berikanlah berita gembira kepada orang-orang yang sabar." [QS Al-Baqarah:155].
Bermacam-macam cobaan yang akan diberikan Allah kepada hambaNya. Cobaan tersebut antara lain dicabutnya nikmat tenang sehingga selalu merasa khawatir, kemudian cobaan berupa kelaparan, kekurangan harta atau kebangkrutan, dan cobaan pada jiwa yang dapat diartikan sebagai penyakit. Penyakit merupakan salah satu bentuk ujian Allah terhadap hambaNya. Kanker merupakan salah satu penyakit yang sangat ditakuti oleh banyak orang. Karena kanker memiliki sifat yang dapat menyebar ke bagian tubuh lainnya sehingga kanker terkenal akan keganasannya. Meskipun demikian, kita harus selalu berprasangka baik terhadap Allah SWT. Seorang hamba yang diuji penyakit apabila bersabar dan berikhtiar sembuh, maka akan menggugurkan dosa-dosanya. Karena Allah telah menjamin bahwa setiap penyakit itu ada obatnya. Hal ini sesuai dengan hadits Rasulullah SAW sebagai berikut.

إِنَّ اللَّهَ أَنْزَلَ الدَّاءَ وَالِدَّوَاءَ وَجَعَلَ لِكُلِّ دَاءٍ دَوَاءً فَتَدَاوَوْا وَلَا تَدَاوَوْا بِحَرَامٍ

Artinya: "Allah telah menurunkan penyakit dan juga obatnya. Allah menjadikan setiap penyakit ada obatnya. Maka berobatlah, namun jangan berobat dengan yang

haram.” [HR Abu Dawud]. Pada hadist tersebut terdapat anjuran untuk orang yang sakit agar berobat untuk mengetahui penyakitnya, karena tiap penyakit akan ada obatnya. Jika pengobatan yang dilakukan tepat terhadap suatu penyakit, dengan izin Allah maka penyakit yang diderita akan hilang dan menjadi sehat kembali.

Berobat untuk mengetahui penyakit secara dini sangat dianjurkan, agar mendapatkan penanganan yang benar. Jika telah menerima penanganan yang benar, maka indikasi dari komplikasi satu penyakit dengan penyakit lainnya dapat dicegah. Sesuai dengan kaidah fikih menyebutkan:

الضَّرَرُ يُدْفَعُ عَلَى قَدْرِ الْإِمْكَانِ

Artinya: ”Bahaya itu harus dihilangkan sebisa mungkin”. Kaidah tersebut menegaskan bahwa bahaya harus bisa dihilangkan ataupun dicegah semaksimal mungkin, sehingga sangat penting bagi seseorang yang sakit untuk segera memeriksakan penyakitnya. Semakin cepat ditangani akan semakin tinggi kemungkinan untuk disembuhkan. Salah satu cara yang dapat dilakukan sebagai bentuk pencegahan penyakit kanker yaitu dengan klasifikasi penyakit kanker yang apabila terdeteksi agar dapat mengambil langkah untuk segera ditangani.

Dalam proses penyembuhan kanker, semua metode penyembuhan hanyalah upaya dan mediasi, dan hanya Allah SWT yang menentukan obatnya. Dengan demikian, klasifikasi kanker payudara, ginjal, paru-paru, dan rahim menggunakan data RNA dengan metode 1D-CNN merupakan bagian dari upaya manusia untuk mencapai kesembuhan.

BAB III

METODE PENELITIAN

3.1. Jenis Penelitian

Penelitian ini akan melakukan proses klasifikasi berdasarkan data RNA penyakit kanker dengan metode 1D-CNN. Data RNA kanker merupakan data angka atau numerik sehingga penelitian ini termasuk jenis penelitian kuantitatif, kemudian akan masuk kedalam hitungan dan hasilnya akan dianalisis. Penelitian ini akan menghasilkan suatu sistem klasifikasi berdasarkan data RNA penyakit kanker dengan metode 1D-CNN.

3.2. Sumber Data

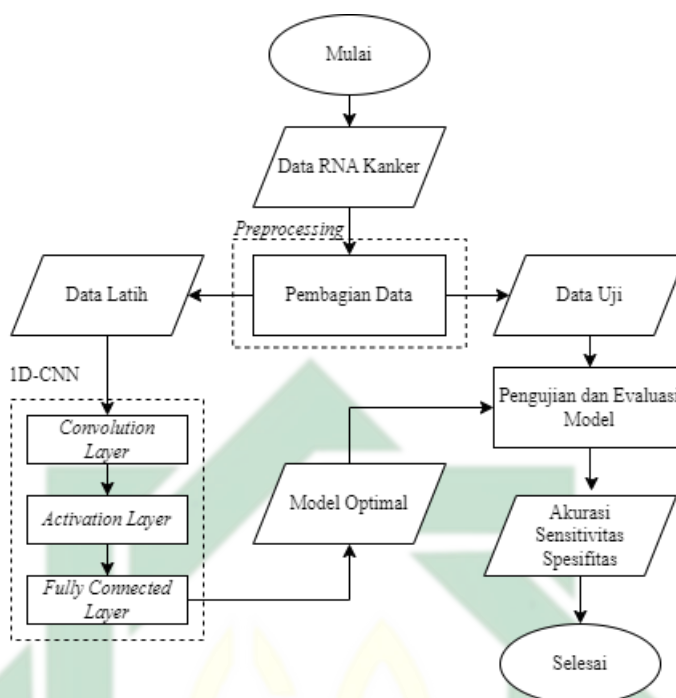
Data yang digunakan pada penelitian ini adalah data berisi susunan gen dari RNA. Pengambilan susunan gen ini menggunakan *Illumina HiSeq Platform*. Data tersebut diakses dari halaman website dari Mendeley Data (Ferles et al., 2018). Data tersebut terdiri dari 1846 data yang berisi 971 kolom dengan isi RPKM (Read per Kilobase Million dari setiap gen). Terdapat empat kelas didalamnya, yaitu yang pertama BRCA (*breast adenocarcinoma*) atau kanker payudara sebanyak 878 data, kemudian KIRC (*kidney renal clear cell carcinoma*) atau kanker ginjal sebanyak 537 data, dan LUAD (*lung adenocarcinoma*) atau kanker paru-paru sebanyak 162 data, serta UCEC (*uterine corpus endometrial carcinoma*) atau kanker rahim sebanyak 269 data. Pada tabel 3.1 terdapat beberapa sampel data dari masing-masing kelas kanker.

Tabel 3.1 Sampel Data RNA

	AARS	ABCB6	ABCC5	ABCF1	ABCF3	...	ZW10	Kelas
1	44.0253	9.2163	11.3191	33.2152	16.9014	...	8.7688	BRCA
2	29.7461	9.7656	40.5401	30.1691	20.0473	...	5.7585	BRCA
3	35.7993	9.8847	3.8860	29.9842	17.1359	...	5.3062	BRCA
1	15.7318	19.5532	5.4462	9.3998	12.0172	...	5.1802	KIRC
2	27.2393	13.0371	8.6423	22.4536	12.9371	...	5.3371	KIRC
3	37.1648	17.2921	10.7791	23.8232	15.5651	...	4.5325	KIRC
1	37.1429	8.9555	8.7000	25.0928	15.0067	...	2.2503	LUAD
2	38.3631	6.5501	9.2564	36.5378	14.7896	...	6.7724	LUAD
3	48.4049	6.3276	4.4723	27.3103	15.8212	...	6.8177	LUAD
1	35.5389	11.7667	4.2605	33.7091	23.0639	...	4.4601	UCEC
2	65.2397	16.6786	9.1418	41.0158	19.3605	...	7.5601	UCEC
3	30.5216	37.4309	13.0927	44.3704	38.5401	...	2.0903	UCEC

3.3. Tahapan Penelitian

Tahapan pada penelitian ini dimulai dari *preprocessing* yaitu data akan mengalami pembagian, kemudian masuk pada proses pelatihan serta pengujian model dengan metode 1D-CNN, dan terakhir masuk pada proses pengevaluasian model dengan *confusion matrix*. Pada Gambar [3.1](#) merupakan *flowchart* dari penelitian ini.



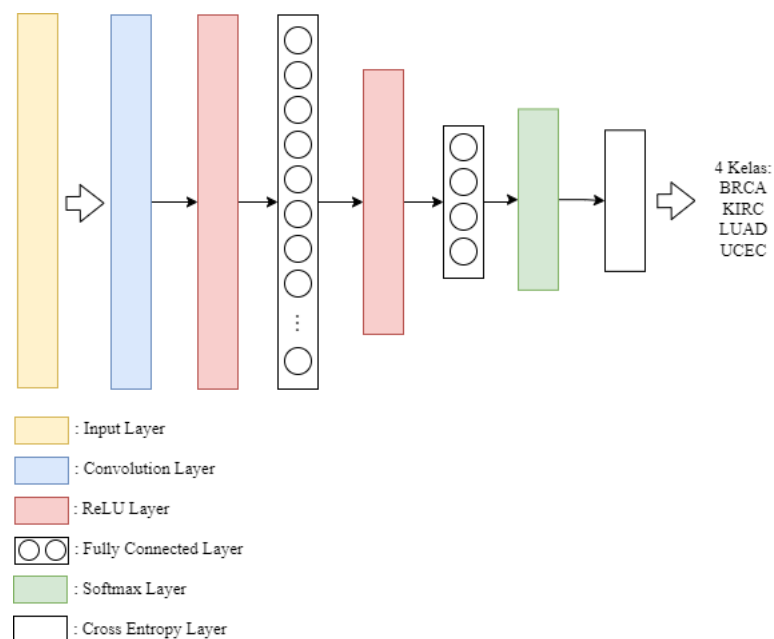
Gambar 3.1 Flowchart Penelitian

1. *Preprocessing*

Tahap *preprocessing* berisi pembagian data. Data RNA kanker dengan ukuran 971x1 akan dibagi menjadi 2 bagian, yaitu data pelatihan (*data training*) dan data pengujian (*data testing*) dengan *k-fold* sebanyak 5.

UIN SUNAN AMPEL
S U R A B A Y A

2. Pelatihan Model



Gambar 3.2 Arsitektur 1D-CNN

Tahap pelatihan model ini dengan algoritma 1D-CNN. Data yang sebelumnya telah terbagi akan masuk ke *convolution layer*, kemudian masuk ke *activation layer* dan *fully connected layer*. Pada tiap layer akan dihitung bobotnya dan selalu diperbarui setiap iterasi dengan tujuan penemuan bobot terbaik. Klasifikasi di layer akhir akan menggunakan bobot terbaik.

3. Pengujian Model

Tahap pengujian model berisi uji coba pada pembagian data, *batch size*, dan *learning rate*. Nilai *learning rate* akan menggunakan 3 macam nilai yaitu 0.1, 0.01, dan 0.001. Nilai *batch size* akan menggunakan 4 macam nilai yaitu 8, 16, 32, dan 64. Dengan melakukan berbagai parameter untuk uji coba akan mendapat model terbaik dan paling optimal.

4. Evaluasi Model

Tahap ini akan melakukan evaluasi terhadap model terbaik yang sudah dibuat. Nilai akurasi, sensitivitas, dan spesifisitas akan dihitung dengan *confusion matrix*.



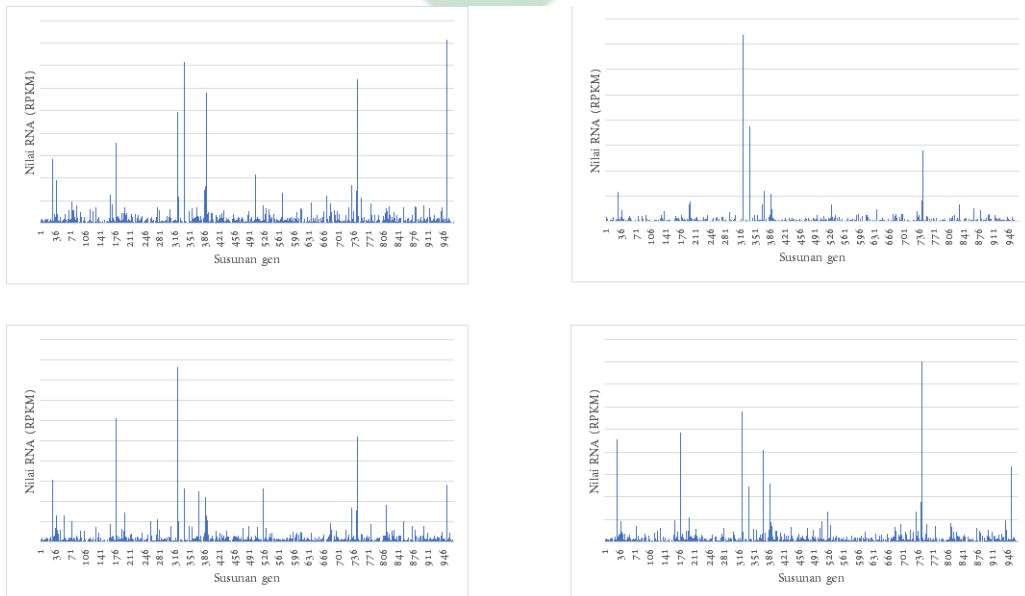
BAB IV

HASIL DAN PEMBAHASAN

4.1. Preprocessing

Penelitian ini menggunakan data vektor yang mengandung genom RNA yang diperoleh pada platform Illumina HiSeq. Satuan kecepatan di mana RNA membaca gen terhadap DNA disebut Reads Per Kilobase Million (RPKM). Satu data terdiri dari 971 kolom RPKM (Read per Kilobase Million dari setiap gen). Perubahan ekspresi gen pada RNA dianggap sebagai indikator ada atau tidaknya kanker dalam tubuh manusia. Perbedaan tinggi rendahnya ekspresi gen pada setiap pasien dengan jenis kanker yang sama dipengaruhi oleh faktor spesifik pasien. Hal tersebut dikarenakan kanker bisa muncul karena banyak faktor. Berikut adalah visualisasi data untuk masing-masing kelas.

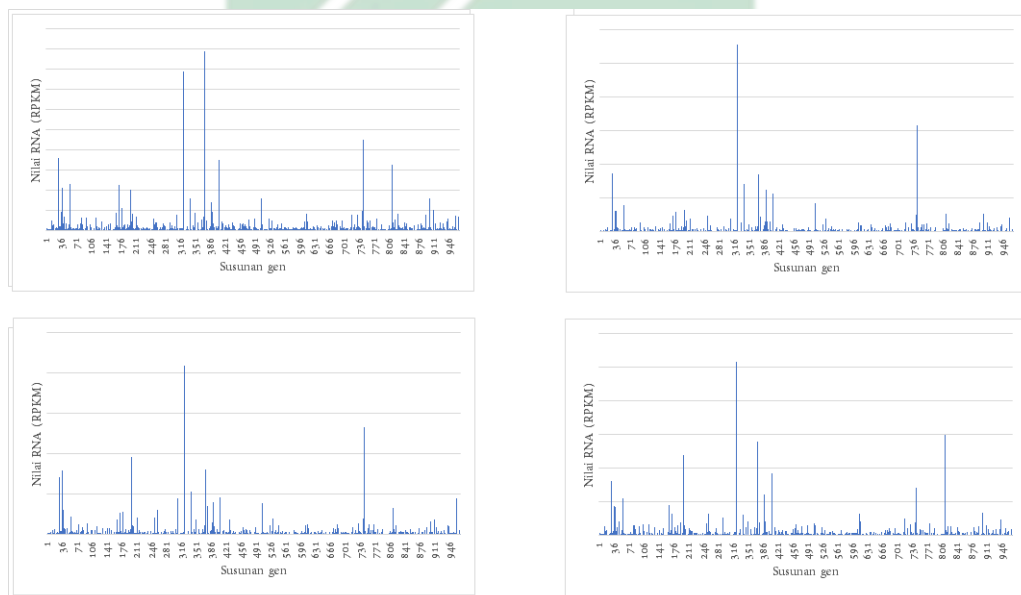
1. BRCA atau kanker payudara



Gambar 4.1 Visualisasi data RNA kanker payudara

Pada pasien payudara tingkat ekspresi gen tertinggi berada pada kisaran 300-400 kelas gen. Beberapa pasien memiliki skor gen yang relatif tinggi antara 700 dan 800 pada awal urutan gen. Ekspresi gen dalam urutan yang tidak disebutkan menunjukkan nilai yang tidak jauh berbeda satu sama lain dan cukup stabil.

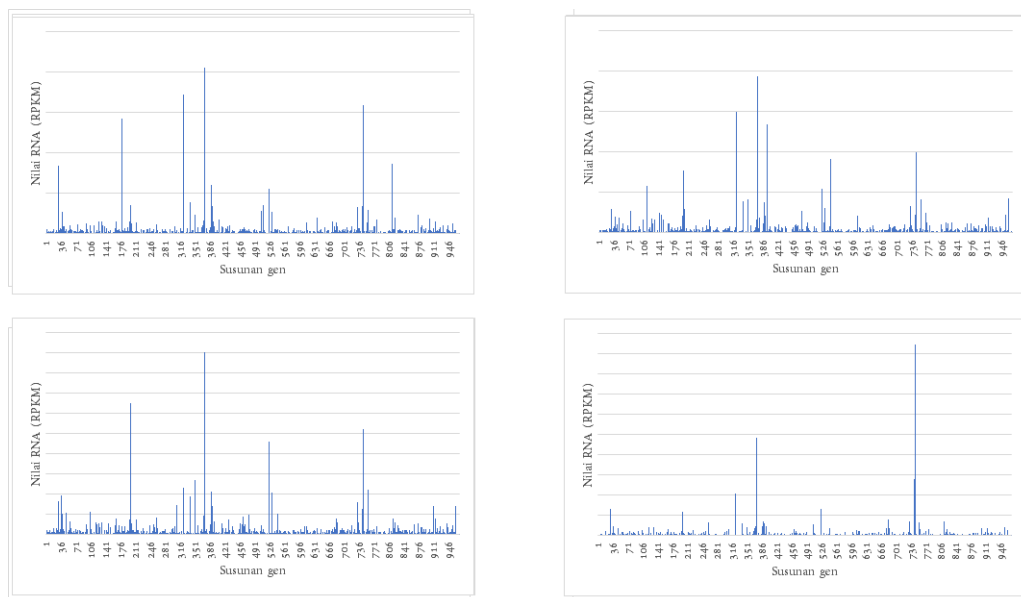
2. KIRC atau kanker ginjal



Gambar 4.2 Visualisasi data RNA kanker ginjal

Pada penderita kanker ginjal, tingkat ekspresi gen tertinggi berada pada kisaran urutan gen 300-400 dan 700-800. Pasien dengan penyakit ini juga memiliki tingkat ekspresi gen yang relatif tinggi pada kisaran 200, 500 dan 900 dibandingkan gen lain yang rendah dan cukup stabil.

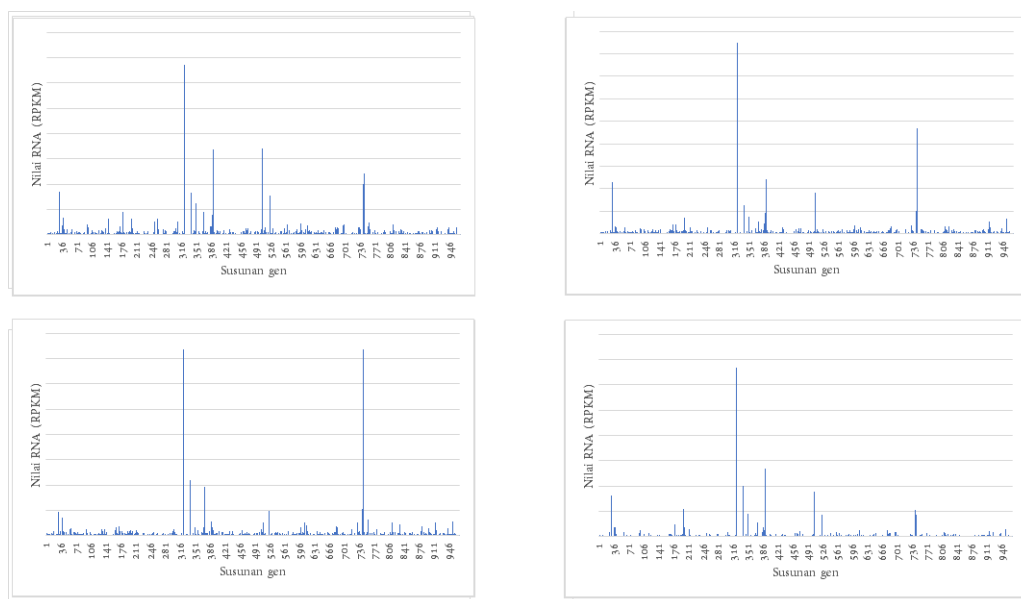
3. LUAD atau kanker paru-paru



Gambar 4.3 Visualisasi data RNA kanker paru-paru

Pada kanker paru-paru susunan gen memiliki nilai tinggi pada urutan 300-400, kemudian pada urutan 200 dan 700-800. Pada urutan lainnya, nilai gen cenderung sama dan cukup stabil.

4. UCEC atau kanker rahim



Gambar 4.4 Visualisasi data RNA kanker rahim

Pada kanker rahim nilai ekspresi gen yang tertinggi pada urutan gen 300-400 dan urutan gen 700-800. Sehingga gen yang menonjol terdapat pada urutan awal dan urutan akhir. Pada urutan lainnya, nilai gen cenderung sama dan cukup stabil.

Pembedaan setiap kelas dalam data menjadi rumit jika hanya mengandalkan pengamatan visual terhadap representasi data. Hal ini disebabkan oleh jumlah variabel yang sangat besar, sehingga diperlukan pendekatan lain untuk mengelompokkan kanker berdasarkan data RNA. Data tersebut kemudian dibagi menjadi set data pelatihan dan data pengujian menggunakan *k-fold cross validation* dengan nilai $k = 5$. Pembagian data ditetapkan sebanyak 80% data pelatihan dan 20% data pengujian. Sehingga didapatkan data pelatihan dan pengujian sebanyak 1476 data dan 370 data. Selanjutnya data akan diklasifikasi menggunakan metode 1D-CNN.

4.2. Pelatihan Model

Pada penelitian ini dilakukan beberapa percobaan menggunakan metode 1D-CNN dengan uji coba parameter pembagian data, *batch size*, dan *learning rate*. Data yang digunakan pada penelitian ini terdiri dari 1846 data yang berisi 971 kolom dengan empat kelas didalamnya, yaitu yang pertama BRCA (*breast adenocarcinoma*) atau kanker payudara sebanyak 878 data, kemudian KIRC (*kidney renal clear cell carcinoma*) atau kanker ginjal sebanyak 537 data, dan LUAD (*lung adenocarcinoma*) atau kanker paru-paru sebanyak 162 data, serta UCEC (*uterine corpus endometrial carcinoma*) atau kanker rahim sebanyak 269 data. Data tersebut dibagi menjadi data latih dan data uji menggunakan *K-fold* untuk proses klasifikasi. Sistem yang dibangun akan mengklasifikasikan kanker menjadi 4 kelas yaitu BRCA dengan label [1000], kemudian KIRC dengan label [0100], kemudian LUAD dengan label [0010], dan UCEC dengan label [0001]. Langkah awal dari metode 1D-CNN yaitu melakukan inisialisasi toleransi error yaitu 0.0001. Selanjutnya, data latih akan masuk kedalam *convolution layer*, kemudian *activation layer*, dan *fully connected layer*.

4.2.1. Convolution Layer

Proses pada *convolution layer* dilakukan menggunakan *filter* berdimensi 3×1 , *stride* (*s*) sebesar 1, *padding* (*p*) sama dengan faktor dilasi (*d*) yaitu 1. Nilai-nilai tersebut disesuaikan dengan arsitektur pada 1D-CNN. Berikut ini merupakan contoh perhitungan pada *convolution layer*. Dimensi hasil yang didapatkan dapat

dihitung menggunakan Persamaan [2.1](#).

$$u = \frac{m - (d + f) + 2p}{s} + 1$$

$$u = \frac{971 - (1 + 3) + 2(1)}{1} + 1$$

$$u = 969$$

Perhitungan pada *convolution layer* dilakukan menggunakan filter (w) berukuran 3×1 . Nilai filter didapat secara acak dalam rentang nilai $[-1, 1]$. Nilai filter (w) pada percobaan perhitungan yaitu:

$$w = \begin{bmatrix} -0.1349 \\ 0.7733 \\ -0.5188 \end{bmatrix}$$

Matriks *input* (i) di kalikan dengan matriks filter w menggunakan Persamaan [2.3](#) seperti berikut:

-0.1349
0.7733
-0.5188

 \times

0	44.0235	9.2162	11.319	33.2151	...	4.2529	4.3371	9.3331	0
---	---------	--------	--------	---------	-----	--------	--------	--------	---

Gambar 4.5 Perhitungan *convolution layer* $z_{(1)}$ dengan *padding* = 1

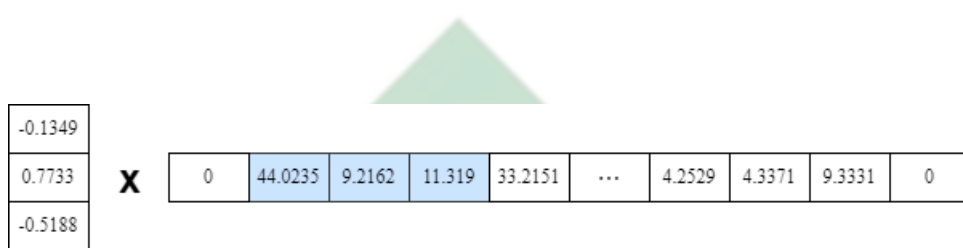
$$z_{(1)} = \sum_{a=0}^{3-1} w_a i_0$$

$$= ((-0,1349 \times 0) (0.7733 \times 44.0235) + (-0.5188 \times 9.2162))$$

$$= -4.6841$$

Kemudian perkalian bergeser sebesar $s = 1$ sehingga untuk nilai $z_{(2)}$ didapatkan:

$$\begin{aligned} z_{(2)} &= \sum_{a=0}^{3-1} w_a i_0 \\ &= ((-0,1349 \times 44.0235) + (0.7733 \times 9.2162) + (-0.5188 \times 11.319)) \\ &= -9.7222 \end{aligned}$$



-0.1349	X	0	44.0235	9.2162	11.319	33.2151	...	4.2529	4.3371	9.3331	0
0.7733											
-0.5188											

Gambar 4.6 Perhitungan *convolution layer* $z_{(2)}$ dengan *stride* = 1

Perhitungan akan berlanjut sampai seluruh matriks *input* telah dikalikan dengan matriks filter. Sehingga didapatkan hasil sebagai berikut:

$$\begin{aligned} z_{(3)} &= \sum_{a=0}^{3-1} w_a i_1 \\ &= ((-0,1349 \times 9.2162) + (0.7733 \times 11.3190) + (-0.5188 \times 33,2151)) \\ &= 15.3898 \end{aligned}$$

$$\begin{aligned} z_{(969)} &= \sum_{a=0}^{3-1} w_a i_{(969)} \\ &= ((-0,1349 \times 4.3371) + (0.7733 \times 9.3331) + (-0.5188 \times 0)) \\ &= -1.4174 \end{aligned}$$

-4.6841	-9.7222	15.3898	...	-2.2591	-1.4174
---------	---------	---------	-----	---------	---------

Gambar 4.7 Hasil *Convolution Layer*

4.2.2. Activation Layer

Setelah mendapatkan hasil dari *convolution layer*, penelitian ini menggunakan fungsi aktivasi *Rectified Linear Unit* (ReLU) dengan tujuan untuk mengantisipasi nilai negatif pada proses sebelumnya. Hasil dari proses *convolution layer* pada Gambar 4.7 masih terdapat nilai negatif. Menghitung fungsi aktivasi dengan Persamaan 2.5 sehingga didapatkan hasil pada Gambar 4.8.

$$\begin{aligned}
 y(1) &= \max(0, -4.6841) \\
 &= 0 \\
 y(2) &= \max(0, -9.7222) \\
 &= 0 \\
 y(3) &= \max(0, 15.3898) \\
 &= 15.3898 \\
 &\vdots \\
 y(969) &= \max(0, -1.4174) \\
 &= 0
 \end{aligned}$$

0	0	15.3898	...	0	0
---	---	---------	-----	---	---

Gambar 4.8 Hasil Fungsi Aktivasi ReLU

4.2.3. Fully Connected Layer

Pada arsitektur yang digunakan terdapat dua *fully connected layer* dengan 384 neuron dan 4 neuron. Langkah-langkah yang dilakukan pada algoritma *backpropagation* sebagai berikut:

1. Inisialisasi bobot awal W dan V serta bias b_1 dan b_2 secara acak dalam rentang -1 sampai 1. Nilai toleransi kesalahan (ϵ) yang digunakan pada penelitian ini yaitu 0.00001 dan laju pemahaman (α) yaitu 0.1. Ukuran matriks W , V , b_1 dan b_2 tergantung dari banyaknya V pada lapisan x , z , dan y . Penelitian ini menggunakan 969 neuron pada lapisan x , 384 neuron pada lapisan z , serta 4 neuron pada lapisan y . Didapatkan ukuran matriks W sebesar 969×384 , matriks V berukuran 384×4 , matriks b_1 berukuran 1×384 , dan matriks b_2 berukuran 1×4 .

$$W = \begin{bmatrix} -0.01647 & 0.02324 & -0.01623 & -0.02199 & \cdots & -0.00457 \\ -0.01975 & -0.01206 & -0.03626 & -0.02838 & \cdots & 0.03606 \\ -0.02352 & 0.01097 & -0.00358 & -0.00805 & \cdots & -0.00419 \\ -0.03575 & -0.0186 & 0.03782 & 0.00793 & \cdots & 0.01520 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0.04261 & 0.02008 & 0.01932 & -0.00154 & \cdots & -0.02794 \\ -0.03158 & -0.02753 & -0.01439 & -0.00138 & \cdots & -0.0364 \end{bmatrix} \quad (969 \times 384)$$

$$V = \begin{bmatrix} 0.11903 & -0.11535 & -0.06742 & -0.06965 \\ 0.03663 & 0.07418 & -0.10273 & 0.08844 \\ 0.11129 & -0.01269 & 0.07583 & -0.11341 \\ \vdots & \vdots & \vdots & \vdots \\ 0.08762 & -0.09837 & -0.02266 & -0.12314 \end{bmatrix} \quad (384 \times 4)$$

$$b_1 = \begin{bmatrix} 0.00033 & -0.00177 & -0.00099 & 0.00257 & 0.00065 & \cdots & -0.00157 \end{bmatrix}$$

$$b_2 = \begin{bmatrix} 0.000165 & -0.00077 & 0.000349 & 0.00085 \end{bmatrix}$$

Propagasi maju :

2. Menghitung nilai keluaran pada lapisan z menggunakan Persamaan 2.7 dengan nilai x inputan merupakan hasil keluaran dari *convolution layer*, nilai w adalah matriks bobot W dan nilai b adalah nilai b_1 yang didapatkan dari inisialisasi pada langkah pertama.

$$\begin{aligned} znet_1 &= b_{1_1} + (x_1)(w_{1,1}) + (x_2)(w_{1,2}) + \dots + (x_{969})(w_{1,969}) \\ &= 0.00033 + (0)(-0.01647) + (0)(0.02324) + \dots + \\ &\quad (0)(-0.00457) \\ &= -0.294737 \end{aligned}$$

$$\begin{aligned} znet_2 &= b_{1_2} + (x_1)(w_{2,1}) + (x_2)(w_{2,2}) + \dots + (x_{969})(w_{2,969}) \\ &= -0.00177 + (12.4674)(-0.01975) + (0)(-0.01206) + \\ &\quad \dots + (2.34544)(0.03606) \\ &= 7.305873 \end{aligned}$$

$$\begin{aligned} znet_3 &= b_{1_3} + (x_1)(w_{3,1}) + (x_2)(w_{3,2}) + \dots + (x_{969})(w_{3,969}) \\ &= -0.000993 + (0)(-0.02352) + (17.5343)(0.01097) + \\ &\quad \dots + (1.38247)(-0.00358) \\ &= 0.0026073 \end{aligned}$$

⋮

$$\begin{aligned} znet_{384} &= b_{1_{384}} + (x_1)(w_{384,1}) + (x_2)(w_{384,2}) + \dots \\ &\quad + (x_{969})(w_{384,969}) \\ &= -0.00157 + (0)(-0.03158) + (0)(-0.02753) + \dots + (0)(-0.0364) \\ &= -0.186681 \end{aligned}$$

Hasil dari z_{net_i} kemudian dihitung dengan fungsi aktivasi ReLU menggunakan Persamaan 2.8 dengan hasil z_i .

$$\begin{aligned} z_1 &= \max(0, -2.94737) \\ &= 0 \end{aligned}$$

$$\begin{aligned} z_2 &= \max(0, 7.305873) \\ &= 73.05873 \end{aligned}$$

$$\begin{aligned} z_3 &= \max(0, 0.026073) \\ &= 0.026073 \end{aligned}$$

$$\vdots$$

$$\begin{aligned} z_{384} &= \max(0, -1.86681) \\ &= 0 \end{aligned}$$

3. Menghitung nilai keluaran pada lapisan y menggunakan Persamaan 2.9 dengan nilai x inputan merupakan hasil keluaran dari lapisan z , nilai w adalah matriks bobot V dan nilai b adalah nilai b_2 yang didapatkan dari inialisasi pada langkah pertama.

$$\begin{aligned} y_{net_1} &= b_{2_1} + (z_1)(v_{1,1}) + (z_2)(v_{1,2}) + \dots + (z_{384})(v_{1,384}) \\ &= 0.000165 + (0)(0.11903) + (73.0569)(0.03663) + \dots \\ &\quad + (0)(0.08762) \\ &= 1.17325 \end{aligned}$$

$$\begin{aligned}
 y_{net_2} &= b_{2_2} + (z_1)(v_{2,1}) + (z_2)(v_{2,2}) + \dots + (z_{384})(v_{2,384}) \\
 &= -0.00077 + (0)(-0.11535) + (73.0569)(0.07418) + \dots \\
 &\quad + (2.34544)(-0.09837) \\
 &= 2.50932
 \end{aligned}$$

$$\begin{aligned}
 y_{net_3} &= b_{2_3} + (z_1)(v_{3,1}) + (z_2)(v_{3,2}) + \dots + (z_{384})(v_{3,384}) \\
 &= 0.000349 + (0)(-0.06742) + (73.0569)(-0.10273) + \dots \\
 &\quad + (1.38247)(-0.02266) \\
 &= -21.3589
 \end{aligned}$$

$$\begin{aligned}
 y_{net_4} &= b_{2_4} + (z_1)(v_{4,1}) + (z_2)(v_{4,2}) + \dots + (z_{384})(v_{4,384}) \\
 &= 0.00085 + (0)(-0.06965) + (73.0569)(0.08844) + \dots \\
 &\quad + (0)(-0.12314) \\
 &= -9.25393
 \end{aligned}$$

Langkah berikutnya yaitu menghitung nilai keluaran dengan fungsi aktivasi *softmax* menggunakan Persamaan 2.21.

$$\begin{aligned}
 y_1 &= \frac{e^{0.11725}}{e^{0.11725} + e^{0.250385} + e^{-2.13532} + e^{-0.92876}} \\
 &= 0.38478
 \end{aligned}$$

$$\begin{aligned}
 y_2 &= \frac{e^{0.250385}}{e^{0.11725} + e^{0.250385} + e^{-2.13532} + e^{-0.92876}} \\
 &= 0.43967
 \end{aligned}$$

$$\begin{aligned}
 y_3 &= \frac{e^{-2.13532}}{e^{0.11725} + e^{0.250385} + e^{-2.13532} + e^{-0.92876}} \\
 &= 0.04045
 \end{aligned}$$

$$\begin{aligned}
 y_4 &= \frac{e^{-0.92876}}{e^{0.11725} + e^{0.250385} + e^{-2.13532} + e^{-0.92876}} \\
 &= 0.13518
 \end{aligned}$$

4. Menghitung nilai eror guna mengetahui kinerja sistem yang dibangun apakah baik. Pada penelitian ini menggunakan rumus *cross entropy* untuk perhitungan nilai eror dengan Persamaan 2.23. Variabel t merupakan label asli dari data yang digunakan, sedangkan y merupakan hasil perhitungan pada lapisan y . Pada perhitungan ini digunakan data dari kanker payudara, oleh karena itu nilai dari variabel t yaitu [1000]. Apabila nilai error yang didapat lebih besar dari nilai toleransi error yang diinisialisasi pada langkah pertama, maka dilakukan propagasi mundur hingga memperoleh nilai error yang diinginkan.

$$\begin{aligned}
 J &= - \sum_{j=1}^h t_j \log y_j \\
 &= - (t_1 \log y_1 + t_2 \log y_2 + t_3 \log y_3 + t_4 \log y_4) \\
 &= - (1 \log 0.38478 + 0 \log 0.43967 + 0 \log 0.04045 + 0 \log 0.13518) \\
 &= - (-0.41478) \\
 &= 0.41478
 \end{aligned}$$

Propagasi mundur :

5. Menghitung kesalahan pada lapisan sebelumnya $\left(\frac{\partial J}{\partial z_i}\right)$ dengan persamaan berikut:

$$\frac{\partial J}{\partial z_i} = \frac{\partial J}{\partial y_j} \cdot \frac{\partial y_j}{\partial z_i}$$

Dimana $\frac{\partial y_j}{\partial z_i}$ merujuk pada Persamaan 2.22, sehingga $\frac{\partial J}{\partial y_j}$ sebagai berikut:

$$\begin{aligned}\frac{\partial J}{\partial y_j} &= -\sum_{j=1} \frac{\partial t_j \log y_j}{\partial y_j} \\ &= -\sum_{j=1} \frac{t_j}{y_j}\end{aligned}$$

Berdasarkan Persamaan 2.22, kemudian didapatkan:

$$\begin{aligned}\frac{\partial J}{\partial z_i} &= \frac{\partial J}{\partial y_j} \cdot \frac{\partial y_j}{\partial z_i} \\ &= -\frac{t_i}{y_i} y_i (1 - y_i) + \sum_{j \neq 1} \frac{t_j}{y_j} (y_i) (y_j) \\ &= t_i y_i - t_i + y_i \sum_{j \neq 1} t_j \\ &= y_i \left(t_i + \sum_{j \neq 1} t_j \right) - t_i \\ &= y_i - t_i\end{aligned}$$

Sehingga nilai $\frac{\partial J}{\partial z_i}$ untuk $i = (1, 2, 3, 4)$ didapatkan:

$$\frac{\partial J}{\partial z_1} = y_1 - t_1$$

$$= 0.38478 - 1$$

$$= -0.61522$$

$$\frac{\partial J}{\partial z_2} = y_2 - t_2$$

$$= 0.43967 - 0$$

$$= 0.43967$$

$$\frac{\partial J}{\partial z_3} = y_3 - t_3$$

$$= 0.04045 - 0$$

$$= 0.04045$$

$$\begin{aligned}\frac{\partial J}{\partial z_4} &= y_4 - t_4 \\ &= 0.13518 - 0 \\ &= 0.13518\end{aligned}$$

6. Menghitung koreksi nilai bobot serta bias yang selanjutnya akan digunakan guna memperbarui bobot pada V dan b_2 .

Nilai koreksi bobot V:

$$\Delta v_{jk} = \alpha \frac{\partial J}{\partial z_j} z_k$$

$$\begin{aligned}\Delta v_{1,1} &= 0.1 \times -0.61522 \times 0 \\ &= 0\end{aligned}$$

$$\begin{aligned}\Delta v_{1,2} &= 0.1 \times -0.61522 \times 7.3041 \\ &= -0.4493\end{aligned}$$

⋮

$$\begin{aligned}\Delta v_{1,384} &= 0.1 \times -0.61522 \times 0 \\ &= 0\end{aligned}$$

$$\begin{aligned}\Delta v_{2,1} &= 0.1 \times 0.43967 \times 0 \\ &= 0\end{aligned}$$

$$\begin{aligned}\Delta v_{2,2} &= 0.1 \times 0.43967 \times 7.3041 \\ &= 0.3210\end{aligned}$$

⋮

$$\begin{aligned}\Delta v_{2,384} &= 0.1 \times 0.43967 \times 0 \\ &= 0\end{aligned}$$

$$\Delta v_{3,1} = 0.1 \times 0.04045 \times 0$$

$$= 0$$

$$\Delta v_{3,2} = 0.1 \times 0.04045 \times 7.3041$$

$$= 0.0295$$

⋮

$$\Delta v_{3,384} = 0.1 \times 0.04045 \times 0$$

$$= 0$$

$$\Delta v_{4,1} = 0.1 \times 0.13518 \times 0$$

$$= 0$$

$$\Delta v_{4,2} = 0.1 \times 0.13518 \times 7.3041$$

$$= 0.0987$$

⋮

$$\Delta v_{4,384} = 0.1 \times 0.13518 \times 0$$

$$= 0$$

Nilai koreksi bias b_2 :

$$\Delta b_{2_j} = \alpha \frac{\partial J}{\partial z_j} b_j$$

$$\Delta b_{2_1} = 0.1 \times -0.61522 \times 0.00016$$

$$= -0.0000101$$

$$\Delta b_{2_2} = 0.1 \times 0.43967 \times -0.00077$$

$$= -0.0000337$$

$$\Delta b_{2_1} = 0.1 \times 0.04045 \times 0.00034$$

$$= 0.0000014$$

$$\Delta b_{2_1} = 0.1 \times 0.13518 \times 0.00085$$

$$= 0.0000114$$

7. Menghitung koreksi nilai bobot serta bias yang selanjutnya akan digunakan guna memperbarui bobot pada W dan b_1 .

$$snet_k = \sum_4^{j=1} \frac{\partial J}{\partial z_j} v_{jk}$$

$$snet_1 = \frac{\partial J}{\partial z_1} v_{1,1} + \frac{\partial J}{\partial z_2} v_{2,1} + \frac{\partial J}{\partial z_3} v_{3,1} + \frac{\partial J}{\partial z_4} v_{4,1}$$

$$= -0.61522 \times 0.11903 + 0.43967 \times -0.11535 +$$

$$0.04045 \times -0.06742 + 0.13518 \times -0.06965$$

$$= -0.13608$$

$$snet_2 = \frac{\partial J}{\partial z_1} v_{1,2} + \frac{\partial J}{\partial z_2} v_{2,2} + \frac{\partial J}{\partial z_3} v_{3,2} + \frac{\partial J}{\partial z_4} v_{4,2}$$

$$= -0.61522 \times 0.03663 + 0.43967 \times 0.07418 +$$

$$0.04045 \times -0.10273 + 0.13518 \times 0.08844$$

$$= 0.01787$$

$$snet_3 = \frac{\partial J}{\partial z_1} v_{1,3} + \frac{\partial J}{\partial z_2} v_{2,3} + \frac{\partial J}{\partial z_3} v_{3,3} + \frac{\partial J}{\partial z_4} v_{4,3}$$

$$= -0.61522 \times 0.11129 + 0.43967 \times -0.01269 +$$

$$0.04045 \times 0.07583 + 0.13518 \times -0.11341$$

$$= -0.08631$$

⋮

$$\begin{aligned}
 s_{net_{384}} &= \frac{\partial J}{\partial z_1} v_{1,384} + \frac{\partial J}{\partial z_2} v_{2,384} + \frac{\partial J}{\partial z_3} v_{3,384} + \frac{\partial J}{\partial z_4} v_{4,384} \\
 &= -0.61522 \times 0.08762 + 0.43967 \times -0.09837 + \\
 &\quad 0.04045 \times -0.02266 + 0.13518 \times -0.12314 \\
 &= -0.11472
 \end{aligned}$$

Kemudian untuk nilai s_k yaitu $s_k = s_{net_k} \frac{\partial z_k}{\partial x}$ dimana untuk nilai $\frac{\partial z_k}{\partial x}$ dapat dilihat di Persamaan 2.6. Sehingga didapatkan:

$$\begin{aligned}
 s_1 &= s_{net_1} \frac{\partial z_1}{\partial x} \\
 &= -0.13608 \times 0 \\
 &= 0
 \end{aligned}$$

$$\begin{aligned}
 s_2 &= s_{net_2} \frac{\partial z_2}{\partial x} \\
 &= 0.01787 \times 1 \\
 &= 0.01787
 \end{aligned}$$

$$\begin{aligned}
 s_3 &= s_{net_3} \frac{\partial z_3}{\partial x} \\
 &= -0.08631 \times 1 \\
 &= -0.08631
 \end{aligned}$$

⋮

$$\begin{aligned}
 s_{384} &= s_{net_{384}} \frac{\partial z_{384}}{\partial x} \\
 &= -0.11472 \times 0 \\
 &= 0
 \end{aligned}$$

Nilai koreksi bobot W:

$$\Delta w_{ki} = \alpha s_k x_i$$

$$\begin{aligned} \Delta w_{1,1} &= 0.1 \times 0 \times 0 \\ &= 0 \end{aligned}$$

$$\begin{aligned} \Delta w_{1,2} &= 0.1 \times 0 \times 0 \\ &= 0 \end{aligned}$$

$$\begin{aligned} \Delta w_{1,3} &= 0.1 \times 0 \times 0.00153 \\ &= 0 \end{aligned}$$

⋮

$$\begin{aligned} \Delta w_{1,969} &= 0.1 \times 0 \times 0 \\ &= 0 \end{aligned}$$

$$\begin{aligned} \Delta w_{2,1} &= 0.1 \times 0.01787 \times 0 \\ &= 0 \end{aligned}$$

⋮

$$\begin{aligned} \Delta w_{384,969} &= 0.1 \times 0 \times 0 \\ &= 0 \end{aligned}$$

UIN SUNAN AMPEL
S U R A B A Y A

Nilai koreksi bias b_1 :

$$\Delta b_{1_k} = \alpha s_k b_{1_k}$$

$$\begin{aligned} \Delta b_{1_1} &= \alpha s_1 b_{1_1} \\ &= 0.1 \times 0 \times 0.00033 \\ &= 0 \end{aligned}$$

$$\begin{aligned}\Delta b_{1_2} &= \alpha s_2 b_{1_2} \\ &= 0.1 \times 0.01787 \times -0.00177 \\ &= -0.00000316\end{aligned}$$

$$\begin{aligned}\Delta b_{1_3} &= \alpha s_3 b_{1_3} \\ &= 0.1 \times -0.08631 \times -0.00099 \\ &= 0.00008544\end{aligned}$$

⋮

$$\begin{aligned}\Delta b_{1_{384}} &= \alpha s_{384} b_{1_{384}} \\ &= 0.1 \times 0 \times -0.00157 \\ &= 0\end{aligned}$$

8. Memperbarui bobot dan bias dengan menggunakan nilai pada proses sebelumnya pada tiap lapisan.

Nilai bobot W:

$$w_{ki} = w_{ki} + \Delta w_{ki}$$

$$\begin{aligned}w_{1,1} &= w_{1,1} + \Delta w_{1,1} \\ &= -0.01647 + 0 \\ &= -0.01647\end{aligned}$$

$$\begin{aligned}w_{1,2} &= w_{1,2} + \Delta w_{1,2} \\ &= 0.02324 + 0 \\ &= 0.02324\end{aligned}$$

$$\begin{aligned}
 w_{1,3} &= w_{1,3} + \Delta w_{1,3} \\
 &= -0.01623 + 0 \\
 &= -0.01623
 \end{aligned}$$

$$\vdots$$

$$\begin{aligned}
 w_{1,969} &= w_{1,1} + \Delta w_{1,1} \\
 &= -0.00457 + 0 \\
 &= -0.00457
 \end{aligned}$$

$$\begin{aligned}
 w_{2,1} &= w_{2,1} + \Delta w_{2,1} \\
 &= -0.01975 + 0 \\
 &= -0.01975
 \end{aligned}$$

$$\vdots$$

$$\begin{aligned}
 w_{384,969} &= w_{384,969} + \Delta w_{384,969} \\
 &= -0.0364 + 0 \\
 &= -0.0364
 \end{aligned}$$

Nilai bias b_1 :

$$\begin{aligned}
 b_{1_1} &= b_{1_1} + \Delta b_{1_1} \\
 &= 0.00033 + 0
 \end{aligned}$$

$$= 0.00033$$

$$\begin{aligned}
 b_{1_2} &= b_{1,2} + \Delta b_{1_2} \\
 &= -0.00177 + (-0.00000316) \\
 &= -0.001773
 \end{aligned}$$

$$\begin{aligned}
 b_{13} &= b_{1,3} + \Delta b_{13} \\
 &= -0.00099 + 0.00008544 \\
 &= -0.000904
 \end{aligned}$$

$$\vdots$$

$$\begin{aligned}
 b_{1384} &= b_{1,384} + \Delta b_{1384} \\
 &= -0.00157 + 0 \\
 &= -0.00157
 \end{aligned}$$

Nilai bobot V:

$$v_{jk} = v_{jk} + \Delta_{jk}$$

$$\begin{aligned}
 v_{1,1} &= v_{1,1} + \Delta_{1,1} \\
 &= 0.11903 + 0 \\
 &= 0.11903
 \end{aligned}$$

$$\begin{aligned}
 v_{1,2} &= v_{1,2} + \Delta_{1,2} \\
 &= 0.03663 + (-0.4493)
 \end{aligned}$$

$$= -0.41267$$

$$\vdots$$

$$\begin{aligned}
 v_{1,384} &= v_{1,384} + \Delta_{1,384} \\
 &= 0.08762 + 0 \\
 &= 0.08762
 \end{aligned}$$

UIN SUNAN AMPEL
S U R A B A Y A

$$\begin{aligned}
 v_{2,1} &= v_{2,1} + \Delta_{2,1} \\
 &= -0.11535 + 0 \\
 &= -0.11535
 \end{aligned}$$

⋮

$$\begin{aligned}
 v_{4,384} &= v_{4,384} + \Delta_{4,384} \\
 &= -0.12314 + 0 \\
 &= -0.12314
 \end{aligned}$$

Nilai bias b_2 :

$$\begin{aligned}
 b_{2_1} &= b_{2_1} + \Delta b_{2_1} \\
 &= 0.000165 + (-0.0000101) \\
 &= 0.0001549
 \end{aligned}$$

$$\begin{aligned}
 b_{2_2} &= b_{2,2} + \Delta b_{2_2} \\
 &= -0.00077 + (-0.0000337) \\
 &= -0.0008037
 \end{aligned}$$

$$\begin{aligned}
 b_{2_3} &= b_{2,3} + \Delta b_{2_3} \\
 &= 0.000349 + 0.0000014 \\
 &= 0.0003504
 \end{aligned}$$

$$\begin{aligned}
 b_{2_4} &= b_{2_4} + \Delta b_{2_4} \\
 &= 0.00085 + 0.0000114 \\
 &= 0.0008614
 \end{aligned}$$

UIN SUNAN AMPEL
S U R A B A Y A

Sehingga didapatkan nilai baru untuk bobot W dan V serta bias b_1 dan b_2 yaitu:

$$W = \begin{bmatrix} -0.01647 & 0.02324 & -0.01623 & \cdots & -0.00457 \\ -0.01975 & -0.01206 & -0.03626 & \cdots & -0.01377 \\ -0.02352 & 0.01095 & -0.00359 & \cdots & -0.00893 \\ -0.03575 & -0.0186 & 0.03782 & \cdots & 0.05042 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ -0.02409 & -0.03088 & -0.02812 & \cdots & -0.0364 \end{bmatrix}$$

$$V = \begin{bmatrix} 0.11903 & -0.11535 & -0.06742 & -0.06965 \\ -0.41267 & 0.39518 & 0.02954 & 0.09874 \\ \vdots & \vdots & \vdots & \vdots \\ 0.08762 & -0.09837 & -0.02266 & -0.12314 \end{bmatrix}$$

$$b_1 = \begin{bmatrix} 0.00033 & -0.001773 & -0.000904 & \cdots & -0.00157 \end{bmatrix}$$

$$b_2 = \begin{bmatrix} 0.0001549 & -0.0008037 & 0.0003504 & 0.0008614 \end{bmatrix}$$

Perhitungan diulang kembali dari langkah 2 menggunakan bobot dan bias baru yang sudah didapatkan. Perhitungan dilakukan hingga nilai error pada langkah 4 lebih kecil dari nilai error yang diinisialisasi pada langkah pertama.

4.3. Pengujian dan Evaluasi Model

Evaluasi model dalam penelitian ini dengan *confusion matrix*. Hasil dari *confusion matrix* mencakup tingkat akurasi, sensitivitas, dan spesifisitas yang di hitung menggunakan Persamaan [2.24](#), [2.25](#), dan [2.26](#). Nilai tersebut mempresentasikan ukuran kinerja algoritma 1D-CNN dalam melakukan klasifikasi

kanker berdasarkan data RNA. Hasil klasifikasi dengan *Learning Rate* = 0.1 ditunjukkan pada Tabel 4.1

Tabel 4.1 Hasil Klasifikasi Model dengan *Learning Rate* = 0.1

<i>k-fold</i>	<i>Batch Size</i>	Akurasi	Sensitivitas	Spesifisitas
1	8	98.37	98.49	97.48
	16	97.56	96.80	95.83
	32	96.75	95.22	95.48
	64	98.64	98.22	98.39
2	8	98.64	96.53	97.94
	16	98.10	97.11	95.89
	32	98.37	97.91	97.49
	64	97.01	96.07	95.91
3	8	98.91	97.81	99.37
	16	98.10	96.86	97.98
	32	98.64	97.94	98.04
	64	96.74	97.01	94.86
4	8	96.47	92.93	94.00
	16	98.91	98.65	99.04
	32	96.47	94.31	95.39
	64	97.01	96.12	94.39
5	8	97.01	94.72	95.70
	16	98.10	97.66	96.84
	32	97.28	96.85	96.68
	64	98.37	97.03	97.29

Berdasarkan Tabel 4.1, uji coba dilakukan dengan nilai *learning rate* = 0.1

dengan jumlah $k\text{-fold} = 5$ dan 5 macam $batch\ size$. Hasil terbaik pada uji coba ini didapatkan pada $k\text{-fold} = 4$ dan nilai $batch\ size = 16$ dengan akurasi mencapai 98.91%, sensitivitas sebesar 98.65%, dan spesifisitas sebesar 99.04%.

Tabel 4.2 Hasil Klasifikasi Model dengan $Learning\ Rate = 0.01$

$k\text{-fold}$	$Batch\ Size$	Akurasi	Sensitivitas	Spesifisitas
1	8	95.40	94.43	93.75
	16	95.40	92.27	94.27
	32	96.48	94.29	93.21
	64	97.56	96.66	97.65
2	8	98.64	97.05	97.64
	16	96.74	94.02	96.43
	32	97.28	96.22	96.97
	64	99.18	98.29	99.48
3	8	96.20	94.25	93.62
	16	97.28	96.73	95.33
	32	97.56	96.87	94.88
	64	97.56	96.59	96.59
4	8	96.47	96.28	93.66
	16	98.10	95.88	97.63
	32	97.28	95.90	94.35
	64	95.66	93.07	94.19

<i>k-fold</i>	<i>Batch Size</i>	Akurasi	Sensitivitas	Spesifisitas
5	8	96.47	93.26	96.11
	16	95.39	93.48	94.51
	32	96.20	94.08	93.51
	64	96.74	96.59	95.38

Berdasarkan Tabel 4.2, uji coba selanjutnya dilakukan dengan nilai *learning rate* = 0.01 dengan jumlah *k-fold* = 5 dan 5 macam *batch size*. Hasil terbaik pada uji coba ini didapatkan pada *k-fold* = 2 dan nilai *batch size* = 64 dengan akurasi mencapai 99.18%, sensitivitas sebesar 98.29%, dan spesifisitas sebesar 99.48%.

Tabel 4.3 Hasil Klasifikasi Model dengan *Learning Rate* = 0.001

<i>k-fold</i>	<i>Batch Size</i>	Akurasi	Sensitivitas	Spesifisitas
1	8	95.13	93.94	91.06
	16	95.40	93.19	95.69
	32	96.48	93.60	95.27
	64	94.59	92.03	91.41
2	8	96.74	94.83	92.77
	16	98.37	96.09	97.70
	32	95.39	94.63	91.65
	64	95.93	95.20	92.90
3	8	95.93	94.40	94.07
	16	96.20	94.23	94.24
	32	95.12	93.98	93.30
	64	95.66	92.82	94.50

<i>k-fold</i>	<i>Batch Size</i>	Akurasi	Sensitivitas	Spesifisitas
4	8	95.12	92.21	91.37
	16	98.37	97.36	95.73
	32	95.66	93.75	92.83
	64	96.20	94.50	95.17
5	8	96.20	95.07	93.98
	16	97.49	96.95	97.09
	32	95.93	94.64	97.00
	64	95.93	95.06	93.19

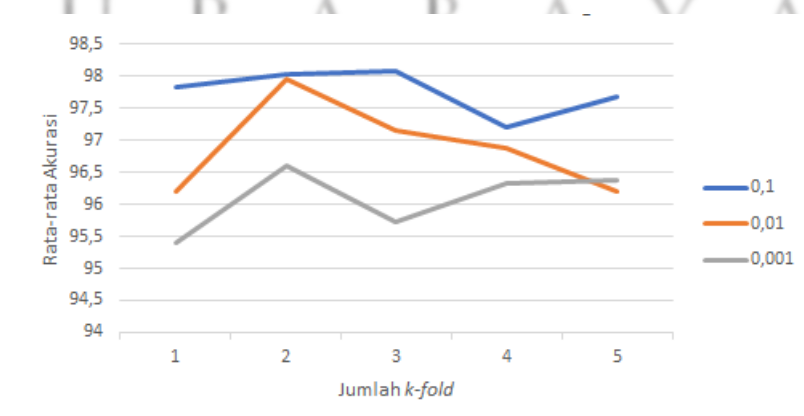
Berdasarkan Tabel 4.3, uji coba terakhir dilakukan dengan nilai *learning rate* = 0.001 dengan jumlah *k-fold* = 5 dan 5 macam *batch size*. Hasil terbaik pada uji coba ini didapatkan pada *k-fold* = 4 dan nilai *batch size* = 16 dengan akurasi mencapai 98.37%, sensitivitas sebesar 97.36%, dan spesifisitas sebesar 95.73%.

Pada beberapa uji coba, nilai *k-fold* dengan hasil akurasi tertinggi didapatkan oleh nilai $k = 4$ dan $k = 2$. Nilai *batch size* dengan akurasi tertinggi pada *batch size* = 16 dan 64.

Selanjutnya untuk mengetahui pengaruh besaran *learning rate*, dilakukan analisis terkait hasil berdasarkan nilai *learning rate*. Hasil akurasi pada tiap *learning rate* akan di rata-rata pada Tabel 4.4.

Tabel 4.4 Rata-rata Akurasi berdasarkan Nilai *Learning Rate*

<i>Learning rate</i>	<i>k-fold</i>	Rata-rata Akurasi
0.1	1	97.83
	2	98.03
	3	98.09
	4	97.21
	5	97.69
0.01	1	96.21
	2	97.96
	3	97.15
	4	96.87
	5	96.20
0.001	1	95.40
	2	96.60
	3	95.72
	4	96.33
	5	96.38

Gambar 4.9 Rata-rata Akurasi berdasarkan Nilai *learning Rate*

Dari Tabel 4.4 dilakukan *plotting* dengan hasil pada Gambar 4.9. Pada grafik tersebut, hasil rata-rata akurasi tertinggi diperoleh oleh nilai *learning rate* = 0.1 yang ditunjukkan oleh garis biru. Hal ini terjadi karena nilai *learning rate* = 0.1 dapat membantu model mencapai konvergensi dengan lebih cepat dan stabil. Sementara rata-rata akurasi pada nilai *learning rate* = 0.001 cenderung lebih rendah dari uji coba lainnya.

Hasil terbaik dari beberapa uji coba yang telah dilakukan didapatkan pada model dengan nilai *learning rate* = 0.01, *k-fold* = 2 dan nilai *batch size* = 64 yaitu akurasi mencapai 99.18%, sensitivitas sebesar 98.29%, dan spesifisitas sebesar 99.48%. Hasil *confusion matrix* model terbaik ditunjukkan pada 4.10

Actual Values	BRCA	171	1	0	0
	KIRC	0	110	0	0
	LUAD	2	0	30	0
	UCEC	0	0	0	56
	Predicted Values	BRCA	KIRC	LUAD	UCEC

Gambar 4.10 *Confusion Matrix* model terbaik

Gambar 4.10 merupakan hasil *confusion matrix* dari model terbaik yang didapatkan. Pada hasil tersebut, terdapat 3 data yang salah terklasifikasi yaitu satu data kanker payudara (BRCA) yang terklasifikasi kanker ginjal (KIRC) dan dua data kanker paru-paru (LUAD) yang terklasifikasi kanker payudara (BRCA). Sedangkan untuk kanker payudara, kanker ginjal, kanker paru-paru, dan kanker

rahim yang terklasifikasi benar sebanyak 171, 110, 30, dan 56.

$$\begin{aligned} Akurasi &= \frac{\sum_{i=1}^l TP_i}{n_{all}} \times 100\% \\ &= \frac{171 + 110 + 30 + 56}{370} \\ &= 99.18\% \end{aligned}$$

$$\begin{aligned} Sensitivitas &= \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l (TP_i + FN_i)} \times 100\% \\ &= \frac{\frac{171}{171+1} + \frac{110}{110+0} + \frac{30}{30+2} + \frac{56}{56+0}}{4} \times 100\% \\ &= 98.29\% \end{aligned}$$

$$\begin{aligned} Spesifisitas &= \frac{\sum_{i=1}^l TN_i}{\sum_{i=1}^l (TN_i + FP_i)} \times 100\% \\ &= \frac{\frac{171}{171+2} + \frac{110}{110+1} + \frac{30}{30+0} + \frac{56}{56+0}}{4} \times 100\% \\ &= 99.48\% \end{aligned}$$

4.4. Penyakit dalam Perspektif Islam

Pada penelitian ini didapatkan hasil bahwa metode 1D CNN dapat bekerja dengan baik dalam klasifikasi kanker berdasarkan data RNA dengan akurasi mencapai 99.18%, sensitivitas sebesar 98.29%, dan spesifisitas sebesar 99.48%. Hal ini dipengaruhi oleh data RNA yang merupakan ekspresi gen sehingga menjadi salah satu indikator terdeteksinya suatu penyakit. Sesuai dengan firman Allah SWT dalam QS Ar-Rum ayat 22 yang berbunyi:

وَمِنْ آيَاتِهِ خَلْقُ السَّمَوَاتِ وَالْأَرْضِ وَاخْتِلَافُ أَلْسِنَتِكُمْ وَأَلْوَانِكُمْ
إِنَّ فِي ذَلِكَ لَآيَاتٍ لِّلْعَالَمِينَ

Artinya: "Dan di antara tanda-tanda (kebesaran)-Nya ialah penciptaan

langit dan bumi, perbedaan bahasamu dan warna kulitmu. Sungguh, pada yang demikian itu benar-benar terdapat tanda-tanda bagi orang-orang yang mengetahui” [QS. Ar-Rum:22]. Ayat tersebut menerangkan bahwa manusia diciptakan oleh Allah SWT dengan ciri yang berbeda-beda. Setiap manusia akan selalu berbeda dengan manusia lainnya meskipun memiliki garis keturunan yang sama. Hal tersebut berlaku pada RNA manusia dimana setiap makhluk hidup pasti memiliki susunan RNA yang berbeda-beda. Apabila dalam suatu susunan RNA dilakukan perubahan, maka representasi dari susunan tersebut juga bisa berubah.

Hasil dari penelitian juga dapat digunakan sebagai sarana untuk mengetahui tipe kanker yang dialami oleh pasien sehingga dapat memperoleh obat dan pengobatan yang sesuai. Ikhtiar periksa ke dokter merupakan salah satu usaha untuk mendapatkan kesembuhan. Namun kesembuhan datangnya hanya dari Allah SWT. Karenanya kita juga harus berdoa memohon kesembuhan kepada Allah. Pada masa Nabi Muhammad SAW, Nabi sendiri mengamalkan atau membaca doa secara teratur untuk memohon perlindungan kepada Allah agar keluarganya selalu diberi kesehatan. Doa ini juga dibacakan oleh Nabi saat mengunjungi orang yang sedang sakit untuk meminta kesembuhan yang cepat. Sesuai dengan hadist berikut:

اللَّهُمَّ رَبَّ النَّاسِ أَذْهِبِ الْبَأْسَ وَأَشْفِهِ وَأَنْتَ الشَّافِي لَا شِفَاءَ إِلَّا شِفَاؤُكَ
شِفَاءَ لَا يُغَادِرُ سَقَمًا

Artinya: ”Ya Allah, Rabb manusia, hilangkanlah kesusahan dan berilah dia kesembuhan, Engkau Zat Yang Maha Menyembuhkan. Tidak ada kesembuhan kecuali kesembuhan dari-Mu, kesembuhan yang tidak meninggalkan penyakit lain.” [HR Bukhari dan Muslim]. Pada hadist tersebut merupakan doa yang diajarkan oleh Nabi Muhammad untuk memohon kesembuhan karena hanya Allah

yang dapat memberikan kesembuhan kepada hamba-Nya. Selain doa tersebut, ada juga doa yang diamalkan oleh Nabi Ayub AS. Nabi Ayub diberi cobaan oleh Allah berupa penyakit kulit sehingga banyak orang menghindarinya karena penyakit tersebut. Doa tersebut terdapat dalam Al-Quran Surat Al-Anbiyaa ayat 83 yang berbunyi:

وَأَيُّوبَ إِذْ نَادَى رَبَّهُ أَنِّي مَسَّنِيَ الضُّرُّ وَأَنْتَ أَرْحَمُ الرَّاحِمِينَ

Artinya: "Dan (ingatlah kisah) Ayub, ketika dia berdoa kepada Tuhannya, "(Ya Tuhanku), sungguh, aku telah ditimpa penyakit, padahal Engkau Tuhan Yang Maha Penyayang dari semua yang penyayang." [QS. Al-Anbiya:83]. Ayat tersebut menjelaskan bahwa Nabi Ayub meyakini bahwa jika nabi terkena penyakit, hanya Allah SWT yang nabi percaya mampu menyembuhkan penyakit tersebut. Ini adalah contoh tindakan tawakal dari seorang hamba terhadap Penciptanya yang akan selalu mencintai hamba-Nya. Dalam proses pemulihan penyakit kanker, semua metode pemulihan hanya merupakan upaya dan perantara, yang menentukan kesembuhan hanya Allah SWT. Oleh karena itu, kategorisasi penyakit kanker payudara, kanker paru-paru, kanker rahim, dan kanker ginjal merupakan bagian dari usaha manusia untuk mencapai kesembuhan.

BAB V

PENUTUP

5.1. Simpulan

Berdasarkan hasil yang diperoleh dari implementasi metode 1D *convolutional neural network* pada klasifikasi kanker berdasarkan data RNA, dapat ditarik kesimpulan sebagai berikut:

1. Metode yang digunakan termasuk kedalam metode *deep learning* sehingga proses ekstraksi fitur dan klasifikasi terdapat dalam satu arsitektur. Metode tersebut dapat bekerja dengan baik dalam klasifikasi kanker karena dapat mengenali 4 kelas kanker dengan akurasi sebesar 99.18%.
2. Klasifikasi kanker berdasarkan data RNA menggunakan metode 1D-CNN mendapatkan model terbaik pada kombinasi parameter dengan nilai *learning rate* = 0.01, *k-fold* = 2 dan nilai *batch size* = 64. Pada model tersebut mendapatkan akurasi sebesar 99.18%, sensitivitas sebesar 98.29%, dan spesifisitas sebesar 99.48%.

5.2. Saran

Pada penelitian mengenai klasifikasi kanker berdasarkan RNA menggunakan 1D-CNN masih memiliki beberapa kekurangan sehingga perlu adanya perbaikan. Beberapa hal yang disarankan untuk penelitian mendatang yaitu:

1. Menambahkan data RNA kelas normal sehingga dapat digunakan untuk deteksi dini penyakit kanker.
2. Melakukan klasifikasi kanker berdasarkan data RNA menurut tingkat stadiumnya sehingga dapat diketahui tingkat keganasan kanker yang di klasifikasi.



UIN SUNAN AMPEL
S U R A B A Y A

DAFTAR PUSTAKA

- Abdoli, S., Cardinal, P., and Lameiras Koerich, A. (2019). End-to-end environmental sound classification using a 1D convolutional neural network. *Expert Systems with Applications*, 136:252–263.
- Adiwijaya (2018). Deteksi Kanker Berdasarkan Klasifikasi Microarray Data. *Jurnal Media Informatika Budidarma*, 2(4):181.
- Alam, W., Ali, S. D., Tayara, H., and Chong, K. T. (2020). A CNN-Based RNA N6-Methyladenosine site predictor for multiple species using heterogeneous features representation. *IEEE Access*, 8:138203–138209.
- Chantsalnyam, T. and Tayara, H. (2020). ncRDeep : Non-coding RNA classification with convolutional neural network. *Computational Biology and Chemistry*, 88:107364.
- Cheng, J., Zhuo, H., Xu, M., Wang, L., Xu, H., Peng, J., and Hou, J. (2018). Regulatory network of circRNA – miRNA – mRNA contributes to the histological classification and disease progression in gastric cancer. *Journal of Translational Medicine*, 16(1):1–14.
- Elbashir, M. K. and Ezz, M. (2019). Lightweight Convolutional Neural Network for Breast Cancer Classification Using RNA-Seq Gene Expression Data. *IEEE Access*, 7:185338–185348.
- Fajri, M. I. and Anifah, L. (2018). Deteksi Status Kanker Paru-Paru Pada Citra Ct Scan Menggunakan Metode Fuzzy Logic. *Teknik Elektro*, 7 no. 3:121–126.

- Ferles, C., Papanikolaou, Y., and Naidoo, K. (2018). Cancer Types: RNA Sequencing Values from Tumor Samples/Tissues.
- He, J., Li, X., Chen, Y., Chen, D., Guo, J., and Zhou, Y. (2021). Deep Transfer Learning Method Based on 1D-CNN for Bearing Fault Diagnosis. *Shock and Vibration*, 2021(1).
- Hou, X., Zaks, T., Langer, R., and Dong, Y. (2021). Lipid nanoparticles for mRNA delivery. *Nature Reviews Materials*, 6(12):1078–1094.
- Kabir, M. F. and Ludwig, S. A. (2019). Classification Models and Survival Analysis for Prostate Cancer Using RNA Sequencing and Clinical Data. *IEEE International Conference on Big Data (Big Data) Classification*, pages 2736–2745.
- Khalifa, N. E. M., Taha, M. H. N., Ezzat Ali, D., Slowik, A., and Hassanien, A. E. (2020). Artificial intelligence technique for gene expression by tumor RNA-Seq Data: A novel optimized deep learning approach. *IEEE Access*, 8:22874–22883.
- Li, D., Zhang, J., Zhang, Q., and Wei, X. (2017). Classification of ECG signals based on 1D convolution neural network. *2017 IEEE 19th International Conference on e-Health Networking, Applications and Services, Healthcom 2017*, 2017-Decem:1–6.
- Liu, K., Kang, G., Zhang, N., and Hou, B. (2018). Breast Cancer Classification Based on Fully-Connected Layer First Convolutional Neural Networks. *IEEE Access*, 6(c):23722–23732.
- Lyu, Z., Yu, Y., Samali, B., Rashidi, M., Mohammadi, M., Nguyen, T. N., and Nguyen, A. (2022). Back-Propagation Neural Network Optimized by K-Fold

- Cross-Validation for Prediction of Torsional Strength of Reinforced Concrete Beam. *Materials*, 15(4).
- Ma, W. and Lu, J. (2017). An Equivalence of Fully Connected Layer and Convolutional Layer. *arXiv preprint arXiv:1712.01252*, 1(3):1–9.
- Mamun, A. A. and Mondal, A. M. (2019). Long Non-coding RNA Based Cancer Classification using Deep Neural Networks. *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, 18(1):541.
- Mei, Z., Ivanov, K., Zhao, G., Wu, Y., Liu, M., and Wang, L. (2020). Foot type classification using sensor-enabled footwear and 1D-CNN. *Measurement: Journal of the International Measurement Confederation*, 165:108184.
- Normalisa (2018). Fuzzy Inference System dengan Metode Mamdani untuk Prediksi Kanker Payudara. *Jurnal Teknologi Sistem Informasi dan Aplikasi*, 1(1):37–44.
- Novitasari, D. C. R., Hendradi, R., Caraka, R. E., Rachmawati, Y., Fanani, N. Z., Syarifudin, A., Toharudin, T., and Chen, R. C. (2020). Detection of COVID-19 chest x-ray using support vector machine and convolutional neural network. *Communications in Mathematical Biology and Neuroscience*, 2020:1–19.
- Olbrecht, S., Busschaert, P., Qian, J., Vanderstichele, A., Loverix, L., Gorp, T. V., Nieuwenhuysen, E. V., Han, S., Broeck, A. V. D., and Coosemans, A. (2021). High-grade serous tubo-ovarian cancer refined with single-cell RNA sequencing : specific cell subtypes influence survival and determine molecular subtype classification. *Biomedcentral*, 13:1–30.

- Pakan, P. D. (2019). Klasifikasi Data Kanker Berdasarkan Mikro Array Gen Menggunakan Jaringan Syaraf Tiruan. *Jurnal Ilmiah Flash*, 5(1):1–4.
- Pourafshin, F. (2021). Big Data Mining in Internet of Things Using Fusion of Deep Features. *International Journal of Scientific Research & Engineering Trends*, 7(2):1089–1093.
- Prasetio, R. T. and Susanti, S. (2019). Prediksi Harapan Hidup Pasien Kanker Paru Pasca Operasi Bedah Toraks Menggunakan Boosted k-Nearest Neighbor. *Jurnal Responsif*, 1(1):64–69.
- Ramroach, S., Joshi, A., and John, M. (2020). Optimisation of cancer classification by machine learning generates an enriched list of candidate drug targets and biomarkers. *Molecular Omics*, 16(2):113–125.
- Rukhsar, L., Bangyal, W. H., Khan, M. S. A., Ibrahim, A. A. A., Nisar, K., and Rawat, D. B. (2022). Analyzing RNA-Seq Gene Expression Data Using Deep Learning Approaches for Cancer Classification. *Applied Sciences (Switzerland)*, 12(4):1–17.
- Stricker, M., Asim, M. N., and Stricker, M. (2021). CircNet : an encoder – decoder-based convolution neural network (CNN) for circular RNA identification. *Neural Computing and Applications*, 34(14):11441–11452.
- Tahir, M., Tayara, H., and Chong, K. T. (2019). iPseU-CNN : Identifying RNA Pseudouridine Sites Using Convolutional Neural Networks. *Molecular Therapy: Nucleic Acid*, 16:463–470.
- Tasdelen, A. and Sen, B. (2021). A hybrid CNN-LSTM model for pre-miRNA classification. *Scientific Reports*, (0123456789):1–9.

- Widadi, R., Widodo, B. A., and Zulherman, D. (2020). Klasifikasi Sinyal EEG pada Sistem BCI Pergerakan Jari Manusia Menggunakan Convolutional Neural Network. *Techno.Com*, 19(4):459–467.
- Wulan, T. D., Salim, H. M., Kurniastuti, I., and Bintarti, T. W. (2022). Klasifikasi Citra Histologi Ginjal Berdasarkan Ekstraksi Fitur Statistik Menggunakan Perceptron. *Remik: Riset dan E-Jurnal Manajemen Informatika Komputer*, 6(3):418–425.
- Yu, Z., Wang, Z., Yu, X., and Zhang, Z. (2020). RNA-Seq-Based Breast Cancer Subtypes Classification Using Machine Learning Approaches. *Computational Intelligence and Neuroscience*, 2020:1–13.
- Yuningsih, A., Prayudha, J., and Syaifuddin, M. (2019). Implementasi Sistem Pakar Mendiagnosa Penyakit Kanker Rahim dalam Kandungan dengan Menggunakan Metode Teorema Bayes. *Jurnal Cyber Tech*, 2(9):1–10.



UIN SUNAN AMPEL
S U R A B A Y A