

**KLASIFIKASI KANKER BERDASARKAN DATA RNA MENGGUNAKAN  
*DILATED CONVOLUTIONAL NEURAL NETWORK***

**SKRIPSI**



Disusun Oleh  
**MOHAMMAD ANANG SYARIFUDIN**  
**H02217009**

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

**2021**

## PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama : Mohammad Anang Syarifudin

NIM : H02217009

Program Studi : Matematika

Angkatan : 2017

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul "Klasifikasi Kanker Berdasarkan Data RNA Menggunakan *Dilated Convolutional Neural Network*". 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, 23 Januari 2021

Yang menyatakan,



  
Mohammad Anang Syarifudin  
NIM. H02217009

## LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh

Nama : Mohammad Anang Syarifudin

NIM : H02217009

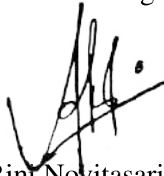
Judul Skripsi : Klasifikasi Kanker Berdasarkan Data RNA Menggunakan

*Dilated Convolutional Neural Network*

telah diperiksa dan disetujui untuk diujikan.

Surabaya, 23 Januari 2021

Pembimbing I



Dian C. Rini Novitasari, M. Kom

NIP. 198511242014032001

Pembimbing II



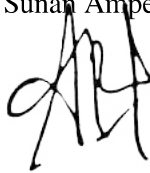
Lutfi Hakim, M.Ag

NIP. 197312252006041001

Mengetahui,

Ketua Program Studi Matematika

UIN Sunan Ampel Surabaya



Ari Fanani, M.Kom

NIP. 198701272014031002

## PENGESAHAN TIM PENGUJI SKRIPSI

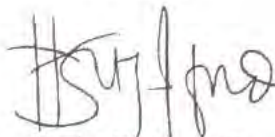
Skripsi oleh

Nama : Mohammad Anang Syarifudin  
NIM : H02217009  
Judul Skripsi : Klasifikasi Penyakit Kanker Berdasarkan Data RNA  
Menggunakan *Dilated Convolutional Neural Network*

Telah dipertahankan di depan Tim Penguji  
pada tanggal 26 Januari 2021

Mengesahkan,  
Tim Penguji

Penguji I



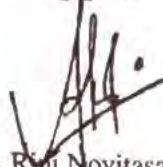
Yuniar Farida, MT  
NIP. 197905272014032002

Penguji II



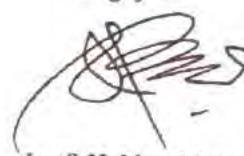
Nurissaidah Elanusa, M. Kom  
NIP. 199011022014032004

Penguji III



Dian C. Rini Novitasari, M. Kom  
NIP. 198511242014032001

Penguji IV



Lutfi Hakim, M. Ag  
NIP. 197312252006041001

Mengetahui,





**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 : MOHAMMAD ANANG SYARIFUDIN  
NIM : H02217009  
Fakultas/Jurusan : SAINTEK / MATEMATIKA  
E-mail address : mohammadanang@gmail.com

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

☒ Skripsi    ☐ Tesis    ☐ Desertasi    ☐ Lain-lain (.....)  
yang berjudul :

KLASIFIKASI KANKER BERDASARKAN DATA RNA MENGGUNAKAN  
DILATED CONVOLUTIONAL NEURAL NETWORK

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, 22 FEBRUARI 2021

Penulis

(MOHAMMAD ANANG SYARIFUDIN)

## ABSTRAK

## Klasifikasi Kanker Berdasarkan Data RNA Menggunakan *Dilated Convolutional Neural Network*

Berdasarkan statistik dari WHO, kanker merupakan salah satu penyakit paling mengancam di dunia. Sebanyak 13% kematian di dunia disebabkan oleh kanker dan terus meningkat setiap tahunnya. Karsinoma merupakan kanker yang paling sering ditemukan pada manusia yaitu 90% dari seluruh kasus kanker di dunia. Salah satu parameter yang dapat digunakan dalam mendeteksi kanker karsinoma yaitu ekspresi RNA. Pada penelitian ini akan dilakukan deteksi kanker berdasarkan ekspresi RNA ke dalam 5 kelas yaitu Breast Carcinoma (BRCA), Kidney Renal clear cell Carcinoma (KIRC), Uterine Corpus Endometrial Carcinoma (UCEC), dan Lung Adenocarcinoma (LUAD) menggunakan metode Dilated Convolutional Neural Network. Beberapa percobaan dilakukan untuk mendapatkan hasil yang terbaik, diantaranya yaitu merubah nilai faktor dilasi dan membandingkan model yang menggunakan augmentasi data atau tanpa menggunakan augmentasi data. Hasil akurasi terbaik yang didapatkan yaitu 99.24% dengan sensitifitas sebesar 98.70%.

**Kata kunci:** Dilated CNN, Kanker, *Deep Learning*, RNA



## DAFTAR ISI

<b>HALAMAN JUDUL</b>	<b>i</b>
<b>HALAMAN PERNYATAAN KEASLIAN</b>	<b>ii</b>
<b>LEMBAR PERSETUJUAN PEMBIMBING</b>	<b>iii</b>
<b>PENGESAHAN TIM PENGUJI SKRIPSI</b>	<b>iv</b>
<b>LEMBAR PERSETUJUAN PUBLIKASI</b>	<b>v</b>
<b>ABSTRAK</b>	<b>vi</b>
<b>ABSTRACT</b>	<b>vii</b>
<b>DAFTAR ISI</b>	<b>1</b>
<b>DAFTAR TABEL</b>	<b>4</b>
<b>DAFTAR GAMBAR</b>	<b>5</b>
<b>I PENDAHULUAN</b>	<b>1</b>
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	4
1.3. Tujuan Penelitian	4
1.4. Manfaat Penelitian	5
1.5. Batasan Masalah	6
1.6. Sistematika Penulisan	6
<b>II TINJAUAN PUSTAKA</b>	<b>8</b>
2.1. Kanker	8
2.1.1. Kanker Paru-Paru	9
2.1.2. Kanker Payudara	10
2.1.3. Kanker Ginjal	10
2.1.4. Kanker Rahim	11
2.2. Ribonucleic Acid (RNA)	12
2.3. 2D Image Creation	13
2.4. Augmentasi Data	13
2.5. k-Fold Cross Validation	16

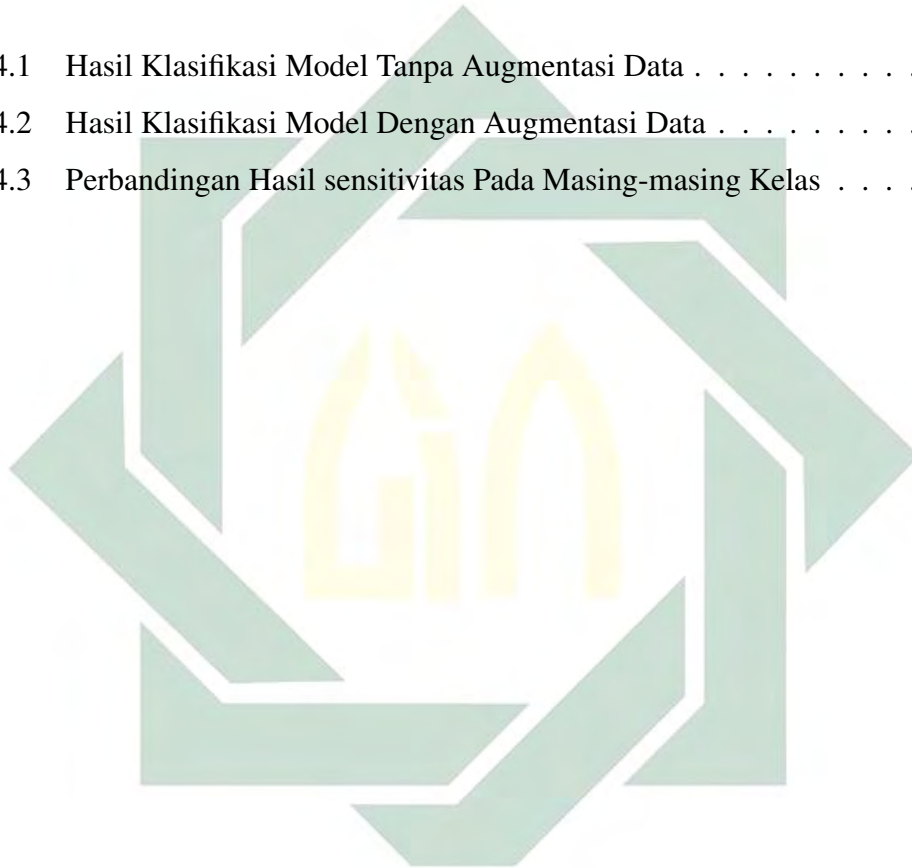






## DAFTAR TABEL

2.1	Contoh Faktor Sekunder dan Persentase Penyebab Kanker . . . . .	8
2.2	Fungsi Aktivasi . . . . .	25
4.1	Hasil Klasifikasi Model Tanpa Augmentasi Data . . . . .	71
4.2	Hasil Klasifikasi Model Dengan Augmentasi Data . . . . .	72
4.3	Perbandingan Hasil sensitivitas Pada Masing-masing Kelas . . . . .	76





4.1	Ilustrasi <i>2D Image Creation</i> . . . . .	40
4.2	Hasil <i>2D Image Creation</i> . . . . .	41
4.3	Contoh Citra <i>Input</i> . . . . .	43
4.4	Contoh Nilai Piksel Citra <i>Input</i> . . . . .	43
4.5	Citra Hasil Refleksi Terhadap Sumbu $x$ . . . . .	45
4.6	Citra Hasil Refleksi Terhadap Sumbu $y$ . . . . .	47
4.7	Citra Hasil Rotasi Sebesar $180^\circ$ . . . . .	49
4.8	Contoh Piksel dengan <i>Padding</i> . . . . .	50
4.9	Hasil <i>Convolution Layer</i> . . . . .	52
4.10	Hasil <i>Activation Layer</i> ReLU . . . . .	53
4.11	Contoh Penerapan <i>Average Pooling</i> . . . . .	54
4.12	Contoh Penerapan <i>Flatten</i> . . . . .	55
4.13	Contoh Proses <i>Dropout</i> . . . . .	56
4.14	Arsitektur <i>Fully Connected Layer</i> . . . . .	56
4.15	(kiri) Model Tanpa Augmentasi Data; (kanan) Model Dengan Augmentasi Data . . . . .	72

# PENDAHULUAN

Kanker merupakan istilah umum yang digunakan untuk menggambarkan penyakit kelainan pertumbuhan pada sel (Khalifa et al., 2020). Berdasarkan statistik yang diberikan oleh *World Health Organization* (WHO), kanker menjadi penyebab 13% kematian di seluruh dunia dengan 8.2 juta jiwa meninggal setiap tahunnya (Zhang et al., 2017). WHO juga telah mengestimasi bahwa pada tahun 2005-2015 sebanyak 84 juta orang meninggal karena kanker (Nugraha et al., 2019). Sedangkan pada tahun 2018, kanker telah menjadi penyebab kematian lebih dari 9 juta orang di seluruh dunia dengan kasus terbesar disebabkan oleh kanker paru-paru ( $>1.7$  juta jiwa) (Bray et al., 2018). Berdasarkan data tersebut, kanker memiliki dampak yang cukup besar pada kehidupan manusia sesuai dengan firman Allah SWT pada Q.S. Al-Baqarah/2:155 mengenai cobaan yang diberikan Allah SWT kepada setiap hamba-Nya.

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

*”Dan sungguh Kami akan menguji kalian dengan sedikit rasa takut, lapar dan kekurangan harta, jiwa dan buah-buahan, dan sampaikan berita gembira pada mereka yang sabar”*

Hingga saat ini, lebih dari 100 jenis kanker yang telah teridentifikasi (Zhang et al., 2017). Namun, secara umum kanker dapat dibagi menjadi 3 kelompok yaitu karsinoma, sarkoma, dan leukemia (limfoma). Dari kasus kanker yang ada di dunia, 90% dari kasus tersebut merupakan jenis kanker karsinoma (Khalifa et al., 2020). Cara umum yang dilakukan untuk mendeteksi kanker yaitu tes laboratorium (urin, darah, dan lain-lain), prosedur pencitraan (*CT scan*, *x-ray*, dan lain-lain), dan biopsi (berdasarkan jaringan pada tubuh). Beberapa cara tersebut membutuhkan biaya yang mahal dan proses yang lama karena dokter akan memeriksa pasien secara manual (Zhang et al., 2017).

[illegible]



sehingga dibutuhkan metode yang cocok untuk melakukan pembelajaran data dalam skala besar (Kurniawan and Ivandari, 2017).

Penelitian mengenai klasifikasi kanker berdasarkan data RNA juga sudah banyak dilakukan, penelitian oleh (Lyu and Haque, 2018) menggunakan salah satu metode *deep learning* yaitu *Convolutional Neural Network* mendapatkan hasil yang baik dengan akurasi sebesar 95.59% dan penelitian lain yang dilakukan oleh (Khalifa et al., 2020) menggunakan metode *Convolutional Neural Network* dan *Binary Particle Swarm Optimization with Decision Tree* (BPSO-DT) sebagai metode seleksi fitur mendapatkan hasil akurasi sebesar 96.90%, akan tetapi kedua penelitian tersebut membutuhkan proses komputasi yang cukup berat. Oleh karena itu, pada penelitian ini digunakan metode *Dilated Convolutional Neural Network* sebagai metode klasifikasi. Metode tersebut dipilih karena pada penelitian sebelumnya yang dilakukan oleh (Lei et al., 2019) (Kudo and Aoki, 2017) mendapatkan hasil bahwa metode *Dilated Convolutional Neural Network* mendapatkan hasil lebih baik dan waktu lebih singkat dalam proses klasifikasi citra dibandingkan dengan metode *Convolutional Neural Network*. Pada penelitian lain yang dilakukan oleh (Zhen et al., 2019), metode *Dilated Convolutional Neural Network* tidak hanya baik digunakan pada data 2 dimensi tetapi juga mendapatkan hasil yang baik ketika digunakan pada data 1 dimensi dibandingkan dengan metode lainnya seperti *Long Short Term Memory* (LSTM), *Gated Recurrent Unit* (GRU), dan lain-lain. Metode *Dilated Convolutional Neural Network* merupakan bagian dari *deep learning*, oleh karena itu proses augmentasi data akan digunakan untuk menghindari *overfitting* (Shorten and Khoshgoftaar, 2019). Selain itu, dari penelitian yang dilakukan oleh (Al-Dhabyani et al., 2019) pada klasifikasi kanker payudara berdasarkan hasil USG menggunakan *deep learning* diperoleh hasil



Berdasarkan latar belakang yang telah dijelaskan, pada penelitian akan dilakukan klasifikasi kanker berdasarkan data RNA menggunakan metode *Dilated Convolutional Neural Network* dengan beberapa percobaan yang diharapkan bisa mendapatkan hasil yang lebih baik.

Berdasarkan latar belakang yang telah dijelaskan, maka peneliti merumuskan permasalahan yang akan diteliti yaitu:

- ### 1.3. Tujuan Penelitian

1. Mengetahui bagaimana proses klasifikasi kanker berdasarkan data RNA menggunakan metode *Dilated Convolutional Neural Network*.

- #### 1.4. Manfaat Penelitian

## 1. Manfaat Teoritis

- [illegible]





## TINJAUAN PUSTAKA

Kanker merupakan istilah yang digunakan untuk berbagai penyakit tumor ganas. Tumor merupakan pembengkakan yang diakibatkan oleh pembelahan sel, baik itu kanker (tumor ganas) atau bukan kanker (tumor jinak). Faktor dari terjadinya kanker pada manusia dibagi menjadi dua yaitu faktor primer (keturunan) dan faktor sekunder seperti yang dijelaskan pada Tabel 2.1 (De Jong, 2005).

Faktor Sekunder	Persentase Penyebab Kanker (%)
Gizi	35
Merokok	30
Infeksi	10
Faktor hormonal	6
Pekerjaan	4
Alkohol	3
Radiasi sinar	3
Pengotoran lingkungan	2
Obat-obatan/tindakan medis	1
Kebiasaan seksual	1
Produk industri	< 1
Makanan	< 1

[illegible]

### 2.1.1. Kanker Paru-Paru

(Sumber: [www.nih.gov](http://www.nih.gov))









(Sumber: [www.researchgate.net](http://www.researchgate.net))

*Ribonucleic Acid* (RNA) atau asam ribonukleat merupakan makromolekul yang berfungsi sebagai penyimpan dan penyalur informasi genetik. Dengan kata lain, RNA dapat mengkonversi informasi genetik dari gen (DNA) ke dalam untai asam amino protein (Sumbono, 2019) seperti pada Gambar 2.5. Satuan yang dapat digunakan untuk menggambarkan kecepatan RNA dalam membaca gen yang ada pada DNA yaitu *Reads Per Kilobase Million* (RPKM). Selain itu, RNA juga dapat berfungsi sebagai onkogen maupun *tumor suppressor gene*. Onkogen merupakan gen yang dapat meningkatkan keganasan sel tumor, sedangkan *tumor suppressor gene* yaitu gen yang bisa mencegah perkembangan tumor. Terjadinya perubahan ekspresi gen pada RNA dapat menjadi indikasi bahwa terdapat kanker pada tubuh manusia (Kresno, 2011). Terdapat beberapa teknik dalam pengambilan RNA dimana Masing-masing teknik memiliki ketentuan dan hasil susunan RNA yang berbeda-beda (Salser, 1974). Pada penelitian ini digunakan data RNA yang diambil menggunakan alat yang bernama *Illumina Hiseq* yang disusun menggunakan metode pada alat tersebut (Clark et al., 2019).

(Sumber: Griffiths et al., 2000)

*2D image creation* merupakan teknik yang dilakukan untuk mengubah data vektor dengan skala  $[a, b]$  menjadi data citra dengan ukuran  $m \times n$  dengan skala  $[0, 255]$ . Apabila  $a$  merupakan nilai minimal dari data vektor,  $b$  merupakan nilai maksimal dari data vektor, dan  $x$  merupakan nilai *input*, maka perubahan skala dapat dilakukan menggunakan Persamaan 2.1 (Khalifa et al., 2020).

Augmentasi data merupakan teknik yang dilakukan untuk memperbanyak set data yang dimiliki. Tujuan dilakukannya augmentasi data adalah untuk mengatasi ketidakseimbangan data sehingga model yang dibangun tidak terlalu condong kearah kelas yang menjadi mayoritas (Shorten and Khoshgoftaar, 2019). Metode yang umum digunakan sebagai teknik augmentasi data adalah modifikasi







### 2.5. *k*-Fold Cross Validation

Diagram illustrating the process of k-fold cross-validation. A horizontal bar represents the dataset, divided into *test* and *train* segments. The *test* segment is highlighted in red. The *train* segment is divided into  $k$  folds. For fold (1), the first segment is red (test) and the rest are white (train). For fold (2), the second segment is red (test) and the rest are white (train). For fold ( $k$ ), the  $k$ -th segment is red (test) and the rest are white (train). A legend at the bottom indicates that a red square represents the *test dataset* and a white square represents the *train dataset*.

(Sumber: Berrar, 2019)

## 2.6. Kecerdasan Buatan

Kecerdasan buatan atau bisa juga disebut *Artificial Intelligence* (AI) merupakan pemodelan suatu sistem yang dapat mempelajari masalah layaknya jaringan saraf otak manusia. Beberapa metode yang sedang dikembangkan dalam kecerdasan buatan adalah *Fuzzy Logic* (Logika Fuzzy), *Evolutionary Computing* (Skema Evolusi), dan *Machine Learning* (Pembelajaran Mesin) (Ahmad, 2017).

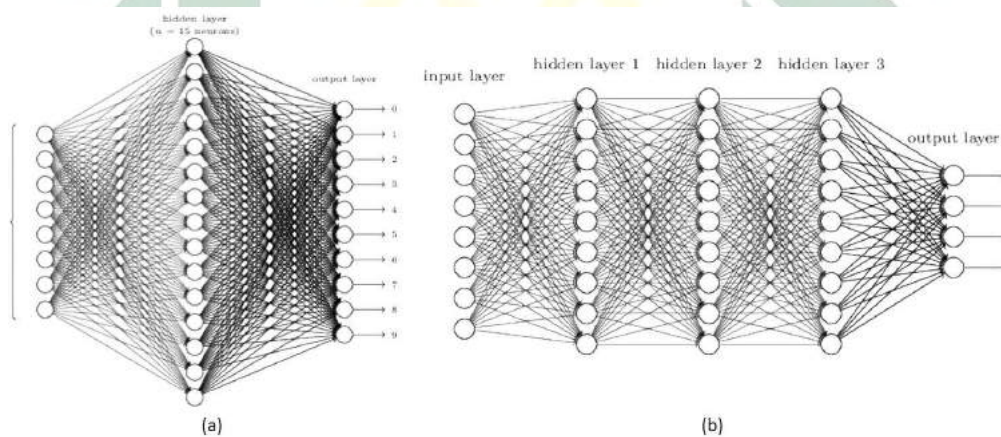
*Machine Learning* atau pembelajaran mesin merupakan model pendekatan suatu sistem agar bisa bekerja semirip mungkin dengan jaringan saraf otak manusia. Metode ini menjadi paling populer karena mampu mempelajari dan menggeneralisasi suatu masalah layaknya otak manusia. Beberapa pengaplikasian metode ini yaitu prediksi dan klasifikasi. Ciri khas dari *machine learning* yaitu adanya proses pembelajaran/pelatihan dan pengujian. Salah satu metode dalam *machine learning* yang sering digunakan adalah *Neural Network* atau Jaringan Saraf Tiruan.

*Neural network* merupakan pemodelan suatu sistem menyerupai karakteristik jaringan saraf biologis yang dimiliki otak manusia agar dapat memproses informasi dengan baik (Wuryandari and Afrianto, 2012). Model ini terdiri dari satu atau beberapa lapisan tersembunyi yang didalamnya terdapat beberapa unit *neuron*. Semakin banyak lapisan tersembunyi maka akan semakin baik kemampuan dari *neural network* tersebut (Ahmad, 2017). Metode dalam *neural network* yang menggunakan banyak lapisan tersembunyi adalah *Deep Learning* atau pembelajaran mendalam.



## 2.7. Deep Learning

*Deep Learning* merupakan bagian dari *machine learning* yang memiliki banyak lapisan sehingga membentuk tumpukan. Lapisan-lapisan tersebut merupakan sebuah algoritma atau metode yang digunakan untuk memproses data sehingga mendapatkan hasil sebaik mungkin (Nurfita and Ariyanto, 2018). Metode ini cocok digunakan dalam mengatasi permasalahan yang kompleks atau permasalahan yang menggunakan data dalam skala besar. Akibatnya, proses pelatihan dalam metode *deep learning* akan membutuhkan waktu yang lama. Perbedaan *deep learning neural network* dengan *simple neural network* dapat dilihat pada Gambar 2.10.



**Gambar 2.10 (a) *Simple Neural Network*; (b) *Deep Learning***

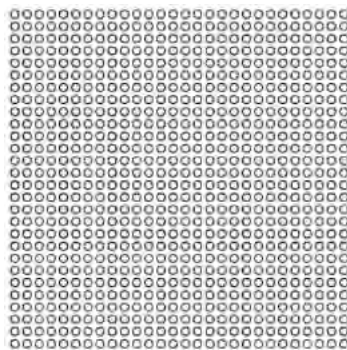
(Sumber: Nielsen, 2015)

## 2.8. Convolutional Neural Network

*Convolutional Neural Network* (CNN) adalah bagian dari *deep learning* yang merupakan pengembangan dari *Multilayer Perceptron* (MLP). *Convolutional neural network* terinspirasi dari jaringan saraf tiruan manusia yang dilakukan oleh Hubel dan Wiesel dalam penelitian *visual cortex* pada indera penglihatan kucing. Pada *convolutional neural network* terdapat suatu lapisan yang memiliki susunan

### 2.8.1. Convolution Layer

*Convolution layer* dapat dibagi menjadi 3 bagian yaitu nilai masukan (*input*), bobot (*filter*), dan nilai keluaran (*feature map*). Misal akan digunakan nilai *input* berupa citra berukuran  $28 \times 28$  seperti pada Gambar 2.11.



(Sumber: Nielsen, 2015)







Dimana  $\mathbf{Z}$  merupakan nilai *output*,  $\mathbf{I}$  merupakan matriks *input*,  $\mathbf{W}$  merupakan matriks *filter*,  $i$  merupakan perhitungan ke- $i$ ,  $f$  merupakan ukuran *filter* yang digunakan, dan  $d$  merupakan faktor dilasi. *Convolution* pada umumnya menggunakan  $d = 1$  sehingga didapatkan Persamaan 2.7.

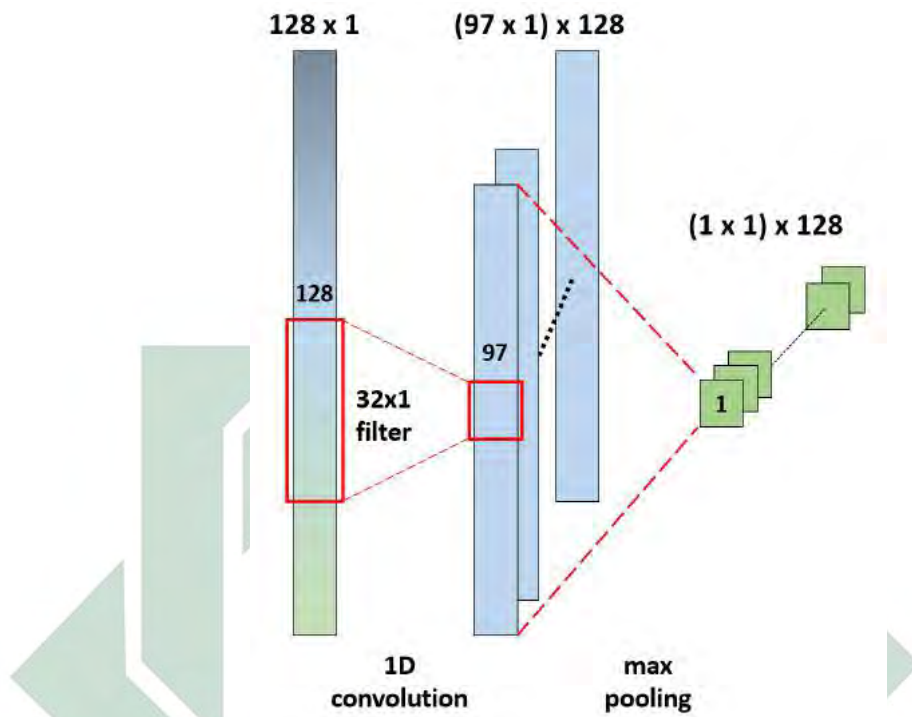
$$z_i = \sum_{a=0}^{f-1} w_a i_{i+a} \quad (2.7)$$

$$z_{(m,n)} = \sum_{a=0}^{f-1} \sum_{b=0}^{f-1} w_{(a,b)} i_{(m+da, n+db)} \quad , d > 0 \quad (2.8)$$

1. *Convolutional 1 Dimension* (CNN 1D)

[illegible]

nilai *output* pada *convolutional 1 dimension* dapat dilakukan menggunakan Persamaan 2.7.

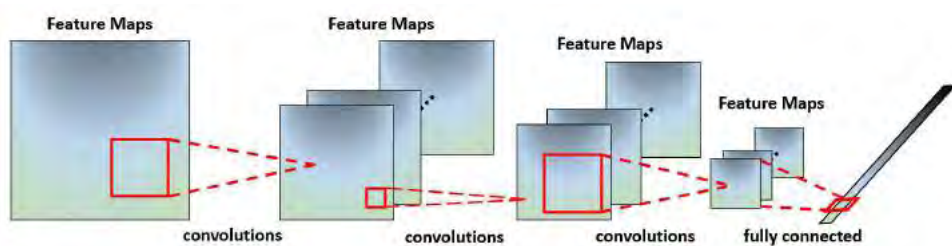


**Gambar 2.15** Arsitektur *Convolutional 1D*

(Sumber: Novitasari et al., 2020)

## 2. Convolutional 2 Dimension (CNN 2D)

*Convolutional 2 dimension* digunakan pada data yang berbentuk dua dimensi (matriks), misalnya citra dan lain-lain. Arsitektur dari *convolutional 2 dimension* dapat dilihat pada Gambar 2.16. Perhitungan nilai *output* pada *convolutional 2 dimension* dapat dilakukan menggunakan Persamaan 2.8.



**Gambar 2.16** *Arsitektur Convolutional 2D*

(Sumber: Novitasari et al., 2020)



*Activation layer* merupakan lapisan yang berisi beberapa fungsi aktivasi yang berfungsi sebagai penentu apakah nilai dari jaringan sebelumnya diaktifkan atau tidak (Suhardi, 2007). Beberapa fungsi aktivasi yang sering digunakan dapat dilihat Tabel 2.2 atau Persamaan 2.10 hingga 2.15 (Munir et al., 2019).

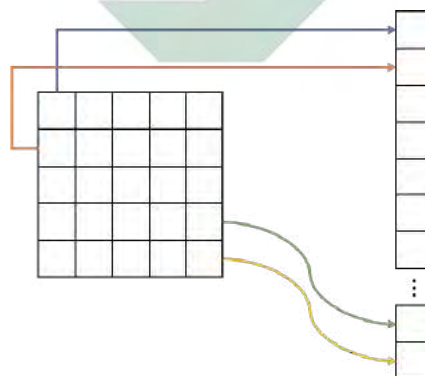
Fungsi Aktivasi	Persamaan
Sigmoid	$y = \frac{1}{1 + e^{-x}} \quad (2.10)$ $\frac{\partial y}{\partial x} = y[1 - y] \quad (2.11)$
Linier	$y = x \quad (2.12)$ $\frac{\partial y}{\partial x} = 1 \quad (2.13)$
ReLU	$y = \max(0, x) \quad (2.14)$ $\frac{\partial y}{\partial x} = \begin{cases} 0 & , x \leq 0 \\ 1 & , x > 0 \end{cases} \quad (2.15)$

Fungsi *softmax* merupakan salah satu jenis fungsi aktivasi lain yang digunakan dalam arsitektur *convolutional neural network*. Fungsi ini digunakan



Dimana  $x$  merupakan nilai dari vektor *input* dan  $c$  merupakan panjang dari vektor *input*.

*Flatten* merupakan teknik yang digunakan untuk merubah hasil keluaran multidimensi dari proses ekstraksi fitur menjadi vektor. Proses tersebut dilakukan karena proses klasifikasi pada *fully connected layer* membutuhkan *input* berupa vektor (Ng et al., 2019). Ilustrasi dari *flatten* dapat dilihat pada Gambar 2.18.



(Sumber: Guillemot et al., 2020)

(Sumber: Labach et al., 2019)

*Fully connected layer* terdiri dari beberapa lapisan yang didalamnya tersusun atas beberapa *node/neuron* yang terkoneksi penuh dengan *layer* sebelumnya. Metode yang digunakan pada lapisan ini yaitu *multilayer perceptron*. Lapisan ini berfungsi mengolah nilai yang didapatkan dari proses sebelumnya sehingga didapatkan hasil sesuai dengan yang diharapkan. Pada intinya, metode yang digunakan pada *fully connected layer* akan melihat kombinasi fitur-fitur yang



$$z_i = b_i + \sum_{j=1}^c x_j w_{ij} \quad (2.18)$$

### 2.8.8. Cross Entropy

[illegible]

$$J = - \sum_{j=1}^h t_j \log y_j \quad (2.19)$$

*Dilated Convolutional Neural Network* merupakan metode yang digunakan untuk mengatasi masalah komputasi yang berat pada *convolutional neural network*. Prinsip dasar dari *dilated convolutional neural network* adalah memberikan lubang/ruang diantara titik satu dengan yang lain pada proses perkalian matriks *input* dengan kernel. Metode ini terbukti pada beberapa penelitian dapat membuat kernel konvolusi memperoleh lebih banyak informasi dari matriks input tanpa meningkatkan proses komputasi (Lei et al., 2019). Proses pada *dilated convolutional neural network* dapat dilakukan menggunakan Persamaan 2.8. Jika metode *convolutional neural network* menggunakan nilai  $d = 1$ , maka metode *dilated convolutional neural network* menggunakan nilai  $d > 1$ . Ilustrasi dari metode *dilated convolutional neural network* dapat dilihat pada Gambar 2.21.











Klasifikasi penyakit kanker merupakan salah satu cara guna melakukan pencegahan pada penderita kanker agar dapat segera ditangani. Hal ini sesuai dengan hadits Rasulullah SAW sebagai berikut.

*"Sesungguhnya Allah telah menurunkan penyakit dan obat, dan menjadikan bagi setiap penyakit terdapat obatnya, maka berobatlah dan jangan berobat dengan sesuatu yang haram". [HR Abu Dawud]*

وَإِذَا مَرِضْتُ فَهُوَ يَشْفِينِ

[illegible]



Kedua ayat di atas menjelaskan bahwa Nabi Ibrahim dan Nabi Ayub memiliki keyakinan bahwa ketika suatu saat tertimpa penyakit, maka hanya Allah SWT yang beliau yakini dapat menyembuhkan. Hal tersebut adalah contoh sikap tawakal dari seorang hamba terhadap Penciptanya yang akan selalu menyayangi hamba-Nya.

Pada proses penyembuhan penyakit kanker, semua cara penyembuhan hanya merupakan usaha dan perantara, yang menentukan kesembuhan hanyalah Allah SWT. Oleh karena itu, klasifikasi penyakit kanker payudara, kanker paru-paru, kanker rahim, dan kanker ginjal merupakan bagian dari usaha manusia untuk mendapatkan kesembuhan.

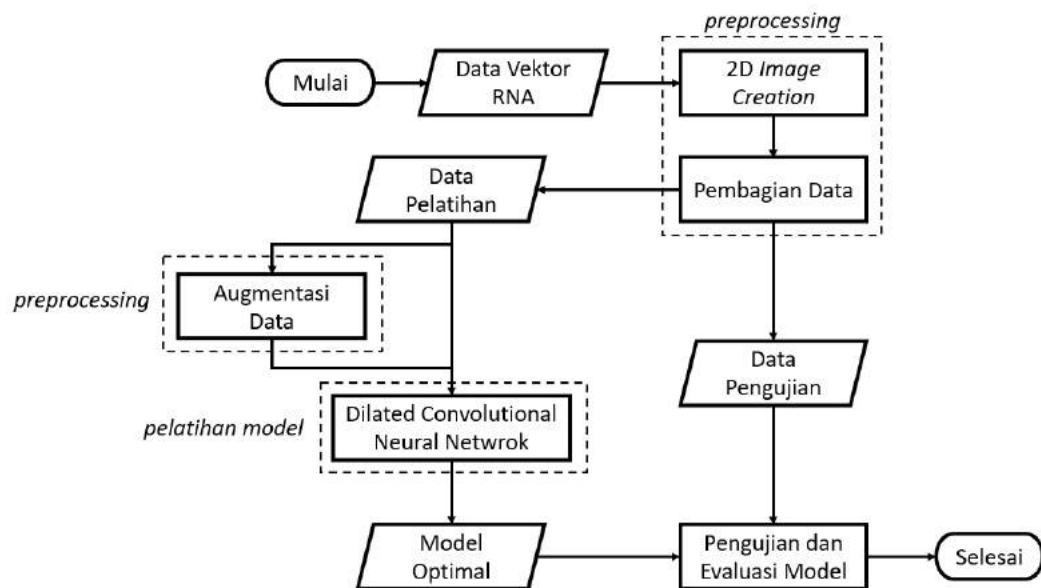
## METODE PENELITIAN

Penelitian ini tergolong ke dalam jenis penelitian kuantitatif karena berdasarkan tujuan yang ingin dicapai, penelitian ini mengandung unsur perhitungan serta data penelitian yang digunakan yaitu berbentuk numerik atau angka. Hasil dari penelitian yang berupa sistem klasifikasi kanker berdasarkan data RNA menggunakan metode *dilated convolutional neural network* diharapkan dapat membantu para ahli medis dalam mendiagnosis kanker pada pasien.

Data yang digunakan dalam penelitian ini merupakan data sekunder RNA yang dapat diakses pada halaman *Mendeley Data* (Ferles et al., 2018). Data tersebut merupakan data vektor berisi susunan gen dari RNA yang diambil menggunakan *Illumina HiSeq Platform* dan terdiri dari 971 kolom yang berisi tentang RPKM (*Read per Kilobase Million*) dari setiap gen. Keseluruhan data berjumlah 1846 yang sudah diklasifikasikan ke dalam 4 kelas yaitu BRCA (kanker payudara), KIRC (kanker ginjal), UCEC (kanker rahim), dan LUAD (kanker paru-paru).

Pada penelitian ini dilakukan beberapa tahapan yaitu *preprocessing*, pelatihan model, pengujian dan evaluasi model. Tahapan pada penelitian dapat

## 1. Preprocessing



**Gambar 3.1 Diagram Alir Sistem Klasifikasi**

## 1. Preprocessing

Langkah pertama pada tahap ini yaitu mengubah vektor RNA yang terdiri dari 971 kolom menjadi matriks dengan jumlah baris sama dengan jumlah kolom. Berdasarkan jumlah kolom yang dimiliki, ukuran yang digunakan pada penelitian ini yakni  $32 \times 32$ . Akan tetapi, matriks dengan ukuran tersebut membutuhkan 1024 elemen, sedangkan jumlah elemen dari vektor RNA yaitu 971. Oleh karena itu akan ditambahkan nilai 0 pada vektor RNA dari kolom 972 – 1024 seperti yang dilakukan oleh (Lyu and Haque, 2018). Setelah semua vektor RNA menjadi matriks, selanjutnya dilakukan *rescaling* dari skala normal  $[0, 12620]$  menjadi skala *image* yaitu  $[0, 255]$ . Langkah kedua yaitu membagi data menjadi set data pelatihan dan set data pengujian. Pembagian data dilakukan menggunakan teknik *k-fold cross validation* dengan  $k = 4$ . Langkah ketiga yang dilakukan yaitu augmentasi data, hal ini dilakukan untuk mendapatkan sampel data yang lebih banyak.

Teknik augmentasi yang dilakukan yaitu refleksi terhadap sumbu  $x$ , refleksi terhadap sumbu  $y$ , dan rotasi sebesar  $180^\circ$ , yang masing-masing dilakukan menggunakan Persamaan 2.2 hingga 2.4. Augmentasi data hanya dilakukan pada set data pelatihan seperti pada penelitian yang dilakukan oleh (Shorten and Khoshgoftaar, 2019). Pada penelitian ini dilakukan percobaan dengan membandingkan hasil klasifikasi menggunakan augmentasi data dan tanpa menggunakan augmentasi data. Hal ini dilakukan untuk mengetahui apakah augmentasi data berpengaruh pada hasil yang diperoleh.

## 2. Pelatihan Model

Pada tahap ini, set data pelatihan hasil dari proses sebelumnya akan dilatih menggunakan *dilated convolutional neural network* dengan arsitektur seperti pada Gambar 3.2. Arsitektur tersebut dipilih karena berdasarkan (Khalifa et al., 2020), arsitektur tersebut mendapatkan hasil yang lebih baik dari arsitektur lainnya. Pada penelitian ini diharapkan metode *dilated convolutional neural network* dapat bekerja lebih baik dengan arsitektur yang sama. Terdapat beberapa *layer* yang digunakan pada arsitektur tersebut yaitu *convolution layer*, *pooling layer*, *activation layer* dengan ReLU sebagai fungsi aktivasi, *dropout layer*, *softmax layer*, *fully connected layer*, dan *cross entropy layer*. Pada tahap ini juga akan dilakukan percobaan dengan mengubah nilai faktor dilasi ( $d$ ) menjadi 2, 4, 6, 8. Hasil yang didapatkan yaitu model dengan bobot dan *filter* terbaik yang nantinya akan digunakan pada proses selanjutnya. Hal ini dilakukan guna mengetahui apakah nilai faktor dilasi akan mempengaruhi hasil klasifikasi.



(Sumber: Khalifa et al., 2020)

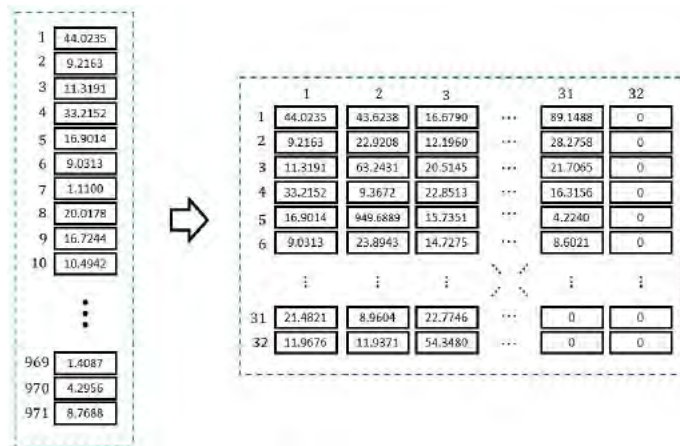
Setelah didapatkan bobot dan *filter* pada proses sebelumnya, langkah selanjutnya yaitu melakukan pengujian terhadap model yang dibangun dengan memberikan set data yang berbeda dengan set data pada proses pelatihan. Set data yang diberikan yaitu set data pengujian hasil dari pembagian menggunakan teknik *k-fold cross validation* pada tahap *preprocessing*. Model akan mengklasifikasikan set data pengujian yang nantinya hasil klasifikasi akan dievaluasi menggunakan konfusi matriks.

## BAB IV

## HASIL DAN PEMBAHASAN

### 4.1. Preprocessing

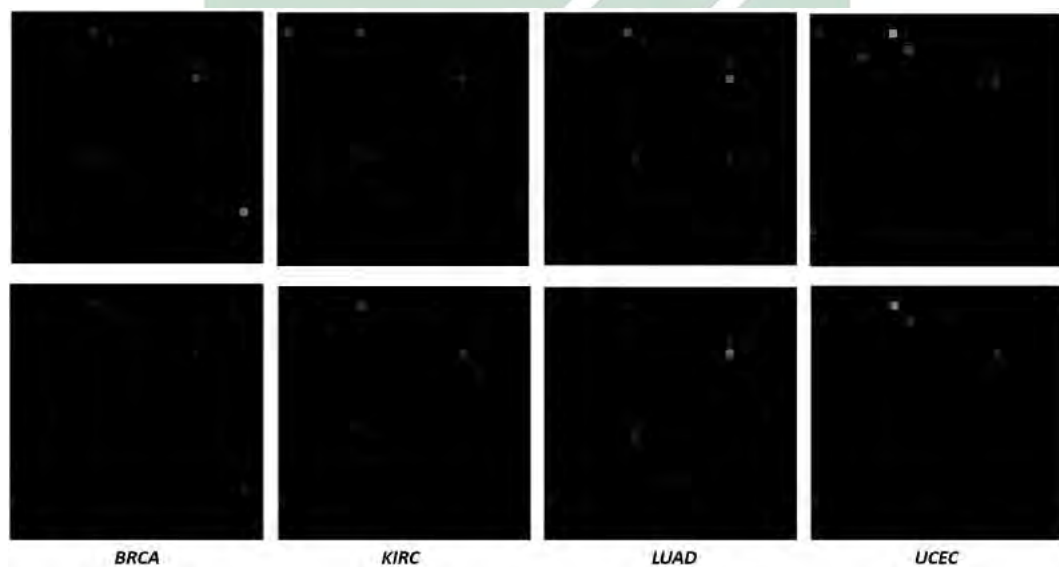
Data yang digunakan pada penelitian ini yaitu data vektor RNA dengan 971 kolom yang berisi tentang RPKM dari masing-masing gen. Arsitektur *dilated convolutional neural network* yang digunakan menggunakan *input layer* dengan masukan berupa matriks, oleh karena itu langkah pertama pada tahap *preprocessing* yaitu mengubah vektor RNA yang terdiri dari 971 kolom menjadi matriks dengan jumlah baris sama dengan jumlah kolom. Berdasarkan jumlah kolom yang dimiliki, ukuran yang digunakan pada penelitian ini yakni  $32 \times 32$ . Akan tetapi, matriks dengan ukuran tersebut membutuhkan 1024 elemen, sedangkan jumlah elemen dari vektor RNA yaitu 971. Oleh karena itu, akan ditambahkan nilai 0 pada vektor RNA dari kolom 972 – 1024. Gambar 4.1 merupakan ilustrasi dari proses *2d image creation* yang dilakukan pada data RNA dari kanker payudara.



### Gambar 4.1 Ilustrasi 2D Image Creation



$$\begin{aligned} PixelValue(0) &= round\left(\frac{(0)(255)}{(12620 - 0)}\right) \\ &= 0 \end{aligned}$$

[illegible]



Proses pelatihan model dilakukan dengan membandingkan beberapa percobaan untuk mendapatkan model yang terbaik. Percobaan yang dilakukan yaitu merubah nilai  $d$  pada *convolution layer*, dan membandingkan hasil model yang menggunakan augmentasi data dan yang tidak menggunakan augmentasi data. Nilai  $d$  yang digunakan yaitu 2, 4, 6, dan 8, sedangkan proses augmentasi data yang dilakukan yaitu refleksi terhadap sumbu  $x$ , refleksi terhadap sumbu  $y$ , serta rotasi sebesar  $180^\circ$ . Arsitektur yang digunakan pada proses pelatihan model dapat dilihat pada Gambar 3.2. Data yang digunakan pada proses pelatihan model sebanyak 1383 – 1386. Berikut merupakan Langkah-langkah yang dilakukan pada proses pelatihan model klasifikasi RNA menggunakan *dilated convolutional neural network*. Sistem yang dibangun akan mengklasifikan kanker menjadi 4 kelas yaitu BRCA dengan kode label [1000], KIRC [0100], LUAD [0010] dan UCEC [0001].

Proses augmentasi dilakukan pada data citra hasil dari proses *2D image creation* yang sudah dipartisi sebanyak  $k$ . Contoh data tersebut dapat dilihat pada Gambar 4.3. Citra tersebut tersusun dari piksel yang memiliki rentang nilai  $[0, 255]$  seperti pada Gambar 4.4. Augmentasi data yang dilakukan yaitu refleksi terhadap sumbu  $x$ , refleksi terhadap sumbu  $y$ , dan rotasi sebesar  $180^\circ$ .





















$$\begin{aligned}
&= \left( (w_{(0,0)} i_{(31,31)}) + (w_{(0,1)} i_{(31,33)}) + (w_{(31,35)} i_{(1,5)}) \right) + \\
&\quad \left( (w_{(1,0)} i_{(33,31)}) + (w_{(1,1)} i_{(33,33)}) + (w_{(33,35)} i_{(3,5)}) \right) + \\
&\quad \left( (w_{(2,0)} i_{(35,31)}) + (w_{(2,1)} i_{(35,33)}) + (w_{(35,35)} i_{(5,5)}) \right) \\
&= \left( (-0.0820 \times 1) + (0.1306 \times 0) + (-0.1955 \times 0) \right) + \\
&\quad \left( (0.0525 \times 0) + (0.0363 \times 0) + (-0.0091 \times 0) \right) + \\
&\quad \left( (0.2203 \times 0) + (0.1783 \times 0) + (-0.0140 \times 0) \right) \\
&= -0.0820
\end{aligned}$$

Perhitungan pada *convolution layer* dilakukan pada masing-masing *filter* sebanyak jumlah *filter* yang digunakan, sehingga setiap *filter* akan mendapatkan hasil yang berbeda-beda.

0.0363	0.2055	0.0525	...	0.2929	0.2203
0.1783	0	0.2021	...	0.8547	0
0.1306	3.3591	-0.0820	...	0.3137	0.0525
0.0363	0	-0.3385	...	0.0421	0
0	1.1629	0	...	0.1383	0.1383
0.1166	0.1643	0.0963	...	-0.0257	0
⋮	⋮	⋮	$\begin{smallmatrix} \times & \times \\ \times & \times \end{smallmatrix}$	⋮	⋮
-0.1955	0	0.1306	...	0	0
-0.2046	0	0.1669	...	0	-0.0820

**Gambar 4.9 Hasil *Convolution Layer***



Pada arsitektur yang digunakan terdapat satu *pooling layer* dengan ukuran  $3 \times 3$  dengan *padding* sama dengan *stride* yaitu sebesar 1. Jenis *pooling layer* yang digunakan adalah *average pooling*. Gambar 4.11 merupakan contoh hasil penerapan dari ReLU *activation layer average pooling* pada Gambar 4.10. Hasil keluaran dari *average pooling* dapat dihitung menggunakan Persamaan 2.9.

Figure 1: A diagram illustrating the iterative process of the proposed algorithm. It shows two stages of a 4x4 grid of boxes. In the first stage, the top-left 3x3 subgrid is highlighted with a blue border. In the second stage, the top-left 3x3 subgrid is again highlighted, but the values in the boxes have changed, indicating an update. Arrows point from the updated values in the second stage to the corresponding values in the first stage, showing the iterative nature of the algorithm. The boxes are arranged in a grid, with some boxes containing numerical values and others containing symbols like '0' or '1'. The diagram is labeled 'Figure 1' and 'Figure 2'.

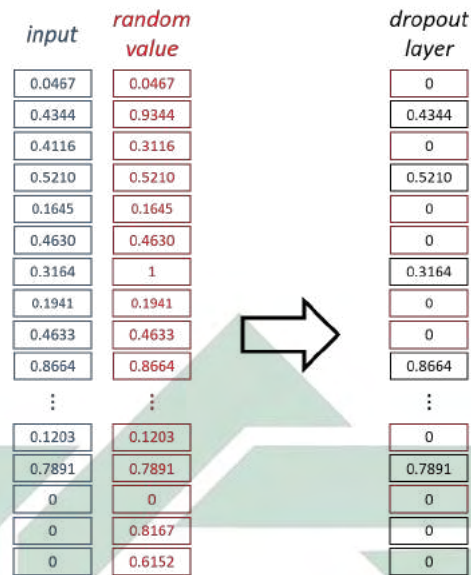
Hasil dari proses *average pooling* akan digunakan pada proses selanjutnya yaitu *dropout layer*. Akan tetapi, sebelum digunakan pada proses selanjutnya, akan

Diagram illustrating the architecture of the proposed model, showing the flow of information from the input layer to the hidden layer and then to the output layer. The input layer consists of 10 nodes, the hidden layer consists of 10 nodes, and the output layer consists of 10 nodes. The input layer is connected to the hidden layer, and the hidden layer is connected to the output layer. The output layer is also connected to the input layer, suggesting a recurrent or residual connection. The diagram is labeled "Figure 1: Architecture of the proposed model".

**Gambar 4.12 Contoh Penerapan *Flatten***

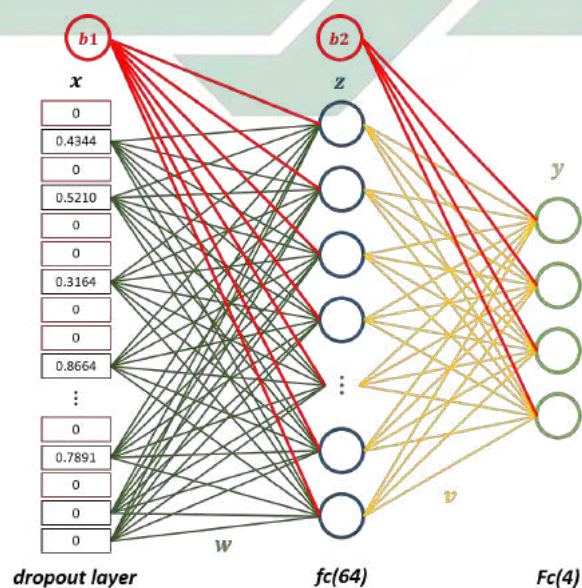
Pada proses ini akan dilakukan seleksi apakah neuron diteruskan ke lapisan selanjutnya atau tidak. Hasil dari proses sebelumnya berupa vektor yang terdiri dari 1024 elemen. Hasil tersebut akan menjadi nilai *input* pada *dropout layer* yang terdiri dari 1024 neuron. Neuron tersebut akan diseleksi dengan probabilitas masing-masing sebesar 0.5. Langkah yang dilakukan yaitu membuat vektor baru terdiri dari 1024 elemen yang berisi nilai acak antara 0 sampai 1. Apabila hasil nilai acak kurang dari nilai probabilitas, maka neuron tersebut tidak digunakan pada proses selanjutnya. Neuron yang tidak digunakan dapat diganti nilai 0 atau dilewati dalam proses perhitungan. Proses pada *dropout layer* dapat dilihat pada Gambar 4.13.



Gambar 4.13 Contoh Proses *Dropout*

#### 4.2.6. Fully Connected Layer

Pada arsitektur yang digunakan terdapat dua *fully connected layer* yang terdiri dari 64 neuron dan 4 neuron. Arsitektur dari *fully connected layer* dapat dilihat pada Gambar 4.14. Berikut merupakan langkah-langkah yang dilakukan pada algoritma *backpropagation*.

Gambar 4.14 Arsitektur *Fully Connected Layer*

- $$\mathbf{W} = \begin{pmatrix} & 1 & 2 & 3 & 4 & \cdots & 1024 \\ 1 & 0.0975 & -0.8003 & 0.9340 & -0.2769 & \cdots & 0.2238 \\ 2 & -0.2785 & -0.1419 & 0.6787 & 0.0462 & \cdots & -0.6797 \\ 3 & 0.5469 & -0.4218 & 0.7577 & 0.0971 & \cdots & 0.6551 \\ 4 & 0.9575 & 0.9157 & 0.7743 & 0.8235 & \cdots & 0.1626 \\ 5 & -0.9649 & 0.7922 & -0.3922 & 0.6948 & \cdots & -0.1190 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 64 & 0.1576 & 0.9595 & 0.6555 & 0.3171 & \cdots & -0.4984 \end{pmatrix}$$

$$\mathbf{V} = \begin{pmatrix} 0.7572 & 0.5308 & 0.0119 & -0.1656 & \dots & 0.5688 \\ -0.7537 & 0.7792 & 0.3371 & 0.6020 & \dots & 0.6892 \\ -0.3804 & -0.9340 & -0.1622 & -0.2630 & \dots & -0.4694 \\ 0.5678 & 0.1299 & 0.7943 & -0.6541 & \dots & 0.7482 \end{pmatrix}$$

$$\mathbf{b1} = \begin{pmatrix} 0.4505 & 0.0838 & -0.2290 & -0.9133 & 0.1524 & \cdots & 0.8258 \end{pmatrix}$$

$$\mathbf{b2} = \begin{pmatrix} -0.9961 & -0.0782 & -0.4427 & 0.1067 \end{pmatrix}$$

[illegible]



$$\begin{aligned} ynet_3 &= b2_3 + (z_1)(v_{3,1}) + (z_2)(v_{3,2}) + \cdots + (z_{64})(v_{3,64}) \\ &= -0.4427 + (3.6585)(-0.3804) + (0)(-0.9340) + \cdots \\ &\quad + (1.4859)(-0.4694) \\ &= 0.4616 \end{aligned}$$

$$\begin{aligned} y_{net_4} &= b_{2_4} + (z_1)(v_{4,1}) + (z_2)(z_{4,2}) + \dots \\ &\quad + (z_{64})(z_{4,64}) \\ &= 0.1067 + (3.6585)(0.5678) + (0)(0.1299) + \dots \\ &\quad + (01.4859)(0.7482) \\ &= 0.0851 \end{aligned}$$

Selanjutnya yaitu menghitung nilai keluaran menggunakan fungsi aktivasi. Pada lapisan ini digunakan fungsi aktivasi *softmax* seperti pada Gambar 3.2. Perhitungan dilakukan menggunakan Persamaan 2.16.

$$y_1 = \frac{e^{-0.8875}}{e^{-0.8875} + e^{0.7761} + e^{0.4616} + e^{0.0851}} = 0.0783$$

$$y_2 = \frac{e^{0.7761}}{e^{-0.8875} + e^{0.7761} + e^{0.4616} + e^{0.0851}} = 0.4131$$

$$y_3 = \frac{e^{0.4616}}{e^{-0.8875} + e^{0.7761} + e^{0.4616} + e^{0.0851}} = 0.3016$$

$$y_4 = \frac{e^{0.0851}}{e^{-0.8875} + e^{0.7761} + e^{0.4616} + e^{0.0851}}$$

$$= 0.2070$$

- Propagasi mundur:

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

[illegible]





$$\begin{aligned}\Delta v_{jk} &= \alpha \frac{\partial J}{\partial z_j} z_k \\ \Delta v_{1,1} &= 0.1 \times -0.9217 \times 3.6585 \\ &= -0.3372 \\ \Delta v_{1,2} &= 0.1 \times -0.9217 \times 0 \\ &= 0 \\ &\vdots \\ \Delta v_{1,64} &= 0.1 \times -0.9217 \times 1.4859 \\ &= -0.1369 \\ \Delta v_{2,1} &= 0.1 \times 0.4131 \times 3.6585 \\ &= 0.1511 \\ &\vdots \\ \Delta v_{4,64} &= 0.1 \times 0.2070 \times 1.4859 \\ &= 0.0308\end{aligned}$$

- [illegible]

$$\begin{aligned}\Delta b_{2_1} &= 0.1 \times -0.9217 \times -0.9961 \\ &= 0.0918\end{aligned}$$

$$\begin{aligned}\Delta b_{2_2} &= 0.1 \times 0.4131 \times -0.0782 \\ &= -0.0032\end{aligned}$$

$$\begin{aligned}\Delta b_{2_3} &= 0.1 \times 0.3016 \times -0.4427 \\ &= -0.0134\end{aligned}$$

$$\begin{aligned}\Delta b_{2_4} &= 0.1 \times 0.2070 \times 0.1067 \\ &= 0.0022\end{aligned}$$

7. Menghitung koreksi nilai bobot dan bias yang nantinya akan digunakan untuk memperbarui bobot **W** dan bias **b1**.

$$\begin{aligned} \text{snet}_k &= \sum_{j=1}^4 \frac{\partial J}{\partial z_j} v_{jk} \\ \text{snet}_1 &= \frac{\partial J}{\partial z_1} v_{1,1} + \frac{\partial J}{\partial z_2} v_{2,1} + \frac{\partial J}{\partial z_3} v_{3,1} + \frac{\partial J}{\partial z_4} v_{4,1} \\ &= -0.9217 \times 0.7572 + 0.4131 \times -0.7537 \\ &\quad 0.3016 \times -0.3804 + 0.2070 \times 0.5678 \\ &= -1.0065 \\ \text{snet}_2 &= \frac{\partial J}{\partial z_1} v_{1,2} + \frac{\partial J}{\partial z_2} v_{2,2} + \frac{\partial J}{\partial z_3} v_{3,2} + \frac{\partial J}{\partial z_4} v_{4,2} \\ &= -0.9217 \times 0.5308 + 0.4131 \times 0.7792 \\ &\quad 0.3016 \times -0.9340 + 0.2070 \times 0.1299 \\ &= -0.4222 \\ &\vdots \end{aligned}$$

$$s_k = \text{snet}_k \frac{\partial z_k}{\partial x}$$

Dimana  $\frac{\partial z_j}{\partial x}$  dapat dilihat pada Persamaan 2.15.

$$\begin{aligned} s_1 &= snet_1 \frac{\partial z_1}{\partial x} \\ &= -1.0065 \times 1 \\ &= -1.0065 \\ s_2 &= snet_2 \frac{\partial z_2}{\partial x} \\ &= -0.4222 \times 0 \\ &= 0 \\ &\vdots \end{aligned}$$

[illegible]

•  
•  
•

$$\begin{aligned}w_{ki} &= w_{ki} + \Delta w_{ki} \\w_{1,1} &= w_{1,1} + \Delta w_{1,1} \\&= 0.0975 + 0 \\&= 0.0975 \\w_{1,2} &= w_{1,2} + \Delta w_{1,2} \\&= -0.8003 + (-0.4372) \\&= -1.2375 \\&\vdots \\w_{1,1024} &= w_{1,1024} + \Delta w_{1,1024} \\&= 0.2238 + 0 \\&= 0.2238 \\w_{2,1} &= w_{2,1} + \Delta w_{2,1} \\&= -0.2785 + 0 \\&= -0.2785 \\&\vdots\end{aligned}$$

- [illegible]

$$\begin{aligned} b_{1_1} &= b_{1_1} + \Delta b_{1_1} \\ &= 0.4505 + (-0.0453) \\ &= 0.4052 \end{aligned}$$

$$\begin{aligned} b_{12} &= b_{12} + \Delta b_{12} \\ &= 0.0838 + 0 \\ &= 0.0838 \end{aligned}$$

$$\begin{aligned} b_{1_{64}} &= b_{1_{64}} + \Delta b_{1_{64}} \\ &= 0.8258 + (-0.0187) \\ &= 0.8071 \end{aligned}$$

$$\begin{aligned} v_{1,1} &= v_{1,1} + \Delta v_{1,1} \\ &= 0.7572 + (-0.3372) \\ &= 0.4200 \end{aligned}$$

[digilib.uinsby.ac.id](#)









Tabel 4.2 Hasil Klasifikasi Model Dengan Augmentasi Data

Faktor Dilasi	k-fold	Data Latih	Data Uji	Akurasi
2	1	1385	461	98.26%
	2	1384	462	98.70%
	3	1383	463	98.92%
	4	1386	460	98.48%
Average				98.59%
4	1	1383	463	98.06%
	2	1385	461	98.48%
	3	1384	462	99.13%
	4	1386	460	98.91%
Average				98.65%
6	1	1385	461	98.70%
	2	1384	462	98.05%
	3	1384	462	98.05%
	4	1385	461	98.92%
Average				98.43%
8	1	1383	463	97.84%
	2	1384	462	98.48%
	3	1385	461	98.92%
	4	1386	460	98.48%
Average				98.43%

Pada tabel tersebut dapat dilihat bahwa hasil akurasi tertinggi dari masing-masing model yang diuji yaitu 99.24% dan 98.65%. Hasil tersebut didapatkan dengan nilai faktor dilasi sebesar 4. Matriks konfusi dari masing-masing model dapat dilihat pada Gambar 4.15.

		Predicted Class					
		BRCA	KIRC	LUAD	UCEC		
True Class	BRCA	220				100%	
	KIRC		134			100%	
	LUAD			41		100%	
	UCEC				67	100%	
		100%	100%	100%	100%	100%	

		Predicted Class					
		BRCA	KIRC	LUAD	UCEC		
True Class	BRCA	219		1		99.5%	0.5%
	KIRC		133	1		99.3%	0.7%
	LUAD	1	1	39		95.1%	4.9%
	UCEC				67	100%	
		99.5%	99.3%	95.1%	100%	99.13%	
		0.5%	0.7%	4.9%			

Gambar 4.15 (kiri) Model Tanpa Augmentasi Data; (kanan) Model Dengan Augmentasi Data



$$\begin{aligned}TFN_3 &= \sum_{j=1, j \neq 1}^4 x_{3,j} \\ &= x_{3,1} + x_{3,2} + x_{3,4} \\ &= 2\end{aligned}$$

$$\begin{aligned}TFN_4 &= \sum_{j=1, j \neq 1}^4 x_{4,j} \\ &= x_{4,1} + x_{4,2} + x_{4,3} \\ &= 0\end{aligned}$$

$$\begin{aligned} sens_1 &= \frac{TTP_1}{TTP_1 + TFN_1} \times 100\% & sens_3 &= \frac{TTP_3}{TTP_3 + TFN_3} \times 100\% \\ &= \frac{219}{219 + 1} 100\% & &= \frac{39}{39 + 2} 100\% \\ &= 99.5\% & &= 95.1\% \end{aligned}$$

$$\begin{aligned} sens_3 &= \frac{TTP_3}{TTP_3 + TFN_3} \times 100\% \\ &= \frac{39}{39 + 2} 100\% \\ &= 95.1\% \end{aligned}$$

$$\begin{aligned} sens_2 &= \frac{TP_2}{TP_2 + TFN_2} \times 100\% & sens_4 &= \frac{TP_4}{TP_4 + TFN_4} \times 100\% \\ &= \frac{133}{133 + 1} 100\% & &= \frac{67}{67 + 0} 100\% \\ &= 99.3\% & &= 100\% \end{aligned}$$

$$\begin{aligned} sens_4 &= \frac{TTP_4}{TTP_4 + TFN_4} \times 100\% \\ &= \frac{67}{67 + 0} 100\% \\ &= 100\% \end{aligned}$$

$$\begin{aligned} \text{sensitivitas} &= \frac{\sum_{i=1}^4 \text{sensi}_i}{4} \\ &= \frac{99.5\% + 99.3\% + 95.1\% + 1000\%}{4} \\ &= 98.48\% \end{aligned}$$

$$\begin{aligned}TFP_1 &= \sum_{j=1, j \neq 1}^4 x_{j,1} \\ &= x_{2,1} + x_{3,1} + x_{4,1} \\ &= 1\end{aligned}$$

$$\begin{aligned}TFP_3 &= \sum_{j=1, j \neq 1}^4 x_{j,3} \\ &= x_{1,3} + x_{1,3} + x_{4,3} \\ &= 2\end{aligned}$$

$$\begin{aligned}TFP_2 &= \sum_{j=1, j \neq 1}^4 x_{j,2} \\ &= x_{1,2} + x_{3,2} + x_{4,2} \\ &= 1\end{aligned}$$

$$\begin{aligned}TFP_4 &= \sum_{j=1, j \neq 1}^4 x_{j,4} \\ &= x_{1,4} + x_{2,4} + x_{3,4} \\ &= 0\end{aligned}$$

$$\begin{aligned}
TTN_1 &= \sum_{j=1, j \neq 1}^4 \sum_{k=1, k \neq 1}^4 x_{j,k} & TTN_3 &= \sum_{j=1, j \neq 3}^4 \sum_{k=1, k \neq 3}^4 x_{j,k} \\
&= x_{2,2} + x_{2,3} + x_{2,4} + \cdots + x_{4,4} + & &= x_{1,1} + x_{1,2} + \cdots + x_{4,4} + \\
&= 241 & &= 419 \\
TTN_2 &= \sum_{j=1, j \neq 2}^4 \sum_{k=1, k \neq 2}^4 x_{j,k} & TTN_4 &= \sum_{j=1, j \neq 4}^4 \sum_{k=1, k \neq 4}^4 x_{j,k} \\
&= x_{1,1} + x_{1,3} + \cdots + x_{4,4} + & &= x_{1,1} + x_{1,2} + \cdots + x_{3,3} + \\
&= 387 & &= 395 \\
spes_1 &= \frac{TTN_1}{TTN_1 + TFP_1} \times 100\% & spes_3 &= \frac{TTN_3}{TTN_3 + TFP_3} \times 100\% \\
&= \frac{241}{241 + 1} 100\% & &= \frac{39}{39 + 2} 100\% \\
&= 99.5\% & &= 95.1\% \\
spes_2 &= \frac{TTN_2}{TTN_2 + TFP_2} \times 100\% & spes_4 &= \frac{TTN_4}{TTN_4 + TFP_4} \times 100\% \\
&= \frac{387}{387 + 1} 100\% & &= \frac{67}{67 + 0} 100\% \\
&= 99.3\% & &= 100\% \\
spesifisitas &= \frac{\sum_{i=1}^4 spes_i}{4} \\
&= \frac{99.5\% + 99.3\% + 95.1\% + 100\%}{4} \\
&= 98.48\%
\end{aligned}$$

Berdasarkan hasil tersebut, model tanpa menggunakan augmentasi data mendapatkan hasil yang lebih baik karena proses augmentasi data akan mempengaruhi susunan data yang dimiliki. Data yang digunakan merupakan data urutan RNA dimana proses penyusunannya menggunakan ketentuan dari metode yang digunakan. Metode yang berbeda akan memberikan urutan yang berbeda dan







وَإِذَا مَرِضْتُ فَهُوَ يَشْفِينِ

وَأَيُّوبَ إِذْ نَادَىٰ رَبَّهُ أَنِّي مَسَّنِيَ الضُّرُّ وَأَنْتَ أَرْحَمُ الرَّاحِمِينَ

Kedua ayat di atas menjelaskan bahwa Nabi Ibrahim dan Nabi Ayub memiliki keyakinan bahwa ketika suatu saat tertimpa penyakit, maka hanya Allah SWT yang beliau yakini dapat menyembuhkan. Hal tersebut adalah contoh sikap tawakal dari seorang hamba terhadap Penciptanya yang akan selalu menyayangi hamba-Nya. Pada proses penyembuhan penyakit kanker, semua cara penyembuhan hanya merupakan usaha dan perantara, yang menentukan kesembuhan hanyalah Allah SWT. Oleh karena itu, klasifikasi penyakit kanker payudara, kanker paru-paru, kanker rahim, dan kanker ginjal merupakan bagian dari usaha manusia untuk mendapatkan kesembuhan.

## BAB V

# PENUTUP

## 5.1. Kesimpulan

Berdasarkan hasil yang didapatkan dari implementasi metode *dilated convolutional neural network* dalam klasifikasi kanker berdasarkan data RNA, kesimpulan yang dapat diambil adalah:

1. Metode yang digunakan merupakan termasuk kedalam metode *deep learning* dimana proses ekstraksi fitur dan klasifikasi terdapat dalam satu arsitektur. Metode tersebut dapat bekerja dengan baik dalam klasifikasi kanker dengan akurasi dan sensitivitas sebesar 99.24% dan 98.70%,.
2. Metode *dilated convolutional neural network* dapat bekerja lebih baik dibandingkan dengan metode *convolutional neural network* dengan arsitektur yang sama. Rata-rata akurasi yang didapatkan yaitu sebesar 99.24% dengan nilai faktor dilasi sebesar 4.
3. Metode augmentasi data dapat mempengaruhi hasil klasifikasi yang didapatkan. Pada Penelitian ini model tanpa augmentasi data dapat memberikan hasil akurasi yang lebih baik dibandingkan dengan model dengan augmentasi data. Hasil yang didapatkan dari masing-masing model yaitu 99.24% untuk model tanpa augmentasi data dan 98.65% untuk model dengan augmentasi data.



October).

W., Fahmy, A., Gomaa, M., and Khaled, H. (2019). Deep learning for data augmentation and classification of bread mold images. *International Journal of Advanced Computer Science*, 10(5):618–627.

N. and Harintaka, H. (2018). Implementasi Metode Convolutional Neural Network Untuk Klasifikasi Tanaman Pada Citra Rona. *Journal of Applied Biology*, 24(2):61.

(2019). Cross-Validation. In *Encyclopedia of Biostatistics*, volume 1, pages 542–545. Elsevier.

- Gomaa, M., and Khale  
mentation and classificat  
*tional Journal of Adva*  
a, H. (2018). Impleme  
asifikasi Tanaman P

- Clark, D. P., Pasdernik, N. J., and McGehee, M. R. (2019). *Molecular Biology, Third Edition*. Academic Press.
- De Jong, W. (2005). *Kanker, Apakah Itu?* EGC.
- Dumoulin, V. and Visin, F. (2016). A guide to convolution arithmetic for deep learning. *arXiv preprint arXiv:1603.07285*, pages 1–31.
- Ferles, C., P, Y., and Naidoo, K. (2018). Cancer Types: RNA Sequencing Values from Tumor Samples/Tissues.
- Griffiths, A. J. F., Miller, J. H. ., Suzuki, D. T., Lewontin, R. C., and Gelbart, W. M. (2000). *An Introduction to Genetic Analysis*. WH Freeman and Company.
- Guillemot, M., Heusele, C., Korichi, R., Schnebert, S., and Chen, L. (2020). Breaking batch normalization for better explainability of deep neural networks through layer-wise relevance propagation. *arXiv*.
- Hamid, A. R. A., Umbas, R., Oka, A. A. G., Mochtar, C. A., Djatisoesanto, W., adi Soedarso, M., Safriadi, F., Danarto, Warli, S. M., Ali, A. Z., Seputra, K. P., Sihombing, A. T., and Hakim, L. (2019). *IAUI Pedoman Tatalaksana Kanker Ginjal 2019*. Ikatan Ahli Urologi Indonesia (IAUI).
- Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*, pages 1–18.
- Ho, Y. and Wookey, S. (2020). The Real-World-Weight Cross-Entropy Loss Function: Modeling the Costs of Mislabeling. *IEEE Access*, 8:4806–4813.
- Hsu, Y. H. and Si, D. (2018). Cancer Type Prediction and Classification Based on RNA-sequencing Data. *Proceedings of the Annual International Conference of*











- Functions: Comparison of trends in Practice and Research for Deep Learning. *arXiv preprint arXiv:1811.03378*, pages 1–20.
- Putra, R. E., Tjandrasa, H., and Suciati, N. (2020). Severity classification of non-proliferative diabetic retinopathy using convolutional support vector machine. *International Journal of Intelligent Engineering and Systems*, 13(4):156–170.
- Rajanbabu, A. (2015). Molecular Pathology and Cytogenetics of Endometrial Carcinoma, Carcinosarcoma, and Uterine Sarcomas. In *Uterine Cancer*, pages 13–23. Springer India, New Delhi.
- Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088):533–536.
- Salser, W. A. (1974). DNA Sequencing Techniques. *Annual Review of Biochemistry*, 43(1):923–965.
- Shorten, C. and Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, 6(1).
- Suhardi, I. (2007). Evaluasi Pagaruh Fungsi Aktivasi Dan Parameter Kemiringannya Terhadap Unjukkerja Pengenalan ( Studi Kasus pada Pengenalan Karakter Angka Tulisan Tangan ). *Jurnal Ilmiah Semesta Teknika*, 10(1):53–68.
- Sumbono, A. (2019). *Biomolekul*. Deepublish.
- Suryo, J. (2010). *Herbal Penyembuh Gangguan Sistem Pernapasan*. Penerbit B First.
- Wuryandari, M. D. and Afrianto, I. (2012). Perbandingan Metode Jaringan Syaraf Tiruan Backpropagation Dan Learning Vector Quantization Pada Pengenalan Wajah. *Komputa*, 1(1):45–51.

- Zhang, Y. H., Huang, T., Chen, L., Xu, Y. C., Hu, Y., Hu, L. D., Cai, Y., and Kong, X. (2017). Identifying and analyzing different cancer subtypes using RNAseq data of blood platelets. *Oncotarget*, 8(50):87494–87511.
- Zhao, J., Mao, X., and Chen, L. (2018). Learning deep features to recognise speech emotion using merged deep CNN. *IET Signal Processing*, 12(6):713–721.
- Zhen, X., Chakraborty, R., Vogt, N., Bendlin, B., and Singh, V. (2019). Dilated Convolutional Neural Networks for Sequential Manifold-Valued Data. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, volume 176, pages 10620–10630. IEEE.
- Zufar, M. and Setiyono, B. (2016). Convolutional Neural Networks Untuk Pengenalan Wajah Secara Real-time. *Jurnal Sains dan Seni ITS*, 5(2):128862.