## KLASIFIKASI KANKER BERDASARKAN DATA RNA MENGGUNAKAN ELMAN RECURRENT NEURAL NETWORK

## **SKRIPSI**



# Disusun Oleh **EKA ALIFIA KUSNANTI H92218044**

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

2022

#### PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama

EKA ALIFIA KUSNANTI

NIM

H92218044

Program Studi :

Matematika

Angkatan

2018

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul "KLASIFIKASI KANKER BERDASARKAN DATA RNA MENGGUNAKAN ELMAN RECURRENT 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, 12 Januari 2022

Yang menyatakan,

EKA ALIFIA KUSNANTI

NIM. H92218044

### LEMBAR PERSETUJUAN PEMBIMBING

#### Skripsi oleh

Nama

EKA ALIFIA KUSNANTI

NIM

H92218044

Judul Skripsi

KLASIFIKASI KANKER BERDASARKAN DATA RNA

MENGGUNAKAN ELMAN RECURRENT NEURAL

**NETWORK** 

telah diperiksa dan disetujui untuk diujikan.

Pembimbing I

Dian Candra Rini Novitasari, M. Kom

NIP. 198511242014032001

Pembimbing I

Dr. Abdullol Hamid, M.Pd

NIP. 198508282014031003

Mengetahui, Ketua Program Studi Matematika UIN Sunan Ampel Surabaya

Aris Fanani, M.Kom

NIP. 198701272014031002

## PENGESAHAN TIM PENGUJI SKRIPSI

#### Skripsi oleh

Nama

: EKA ALIFIA KUSNANTI

NIM

H92218044

Judul Skripsi : KLASIFIKASI KANKER BERDASARKAN DATA RNA

MENGGUNAKAN ELMAN RECURRENT NEURAL

**NETWORK** 

Telah dipertahankan di depan Tim Penguji pada tanggal 12 Januari 2022

> Mengesahkan, Tim Penguji

Penguji I

Nurissaidah Ulinhuha, M. Kom

NIP. 199011022014032004

Penguji II

Putroue Keumala Intan, M.Si

NIP. 198805282018012001

Penguji III

Dian Candra Rini Novitasari, M. Kom

NIP. 198511242014032001

Dr. Abdull Hamid, M.Pd

NIP. 1985/8282014031003

Mengetahui,

kan Fakultas Sains dan Teknologi

Siman Ampel Surabaya

matur Rusydiyah, M.Ag

197312272005012003



## 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 akade	emika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:
Nama	EKA ALIFIA KUSNANTI
	H92218099
	SAINTER /MATEMATIKA
E-mail address	eraalifia z@gmail.com
UIN Sunan Ampel ✓ Sekripsi   yang berjudul :	an ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :  Tesis   Desertasi   Lain-lain ()
	HA MENGGUHAKAN ELMAN RECURPENT
	MEURAL HETWORK
Perpustakaan UIN mengelolanya dala menampilkan/memakademis tanpa perpenulis/pencipta da Saya bersedia untu Sunan Ampel Surah	yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Ekslusif ini Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, am bentuk pangkalan data (database), mendistribusikannya, dan publikasikannya di Internet atau media lain secara <i>fulltext</i> untuk kepentingan elu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai n atau penerbit yang bersangkutan.  k menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN baya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta
dalam karya ilmiah s Demikian pernyataa	n ini yang saya buat dengan sebenarnya.
2	Surabaya, 12 Januari 2022
	Penulis
	( sustain

Eta Alista busnonti )
nama terang dan tanda tangan

#### **ABSTRAK**

## KLASIFIKASI KANKER BERDASARKAN DATA RNA MENGGUNAKAN ELMAN RECURRENT NEURAL NETWORK

Kanker termasuk salah satu penyakit mematikan yang jumlah penderitanya terus bertambah setiap tahunnya. Perkembangan dari sel kanker dapat dengan cepat menyebar ke seluruh bagian tubuh melalui aliran darah atau dari sistem limfatik sehingga dapat menyebabkan kematian. Hal tersebut dapat terjadi karena terdapat kekacauan yang ada pada gen. Hal dasar dalam genetika yaitu pemantauan dari ekspresi gen itu sendiri, yaitu dengan mengukur dari mRNA bukan dari proteinnya dikarenakan urutan dari mRNA yang akan berhibridasi dengan komplementer dari DNA dan juga RNA. Tujuan dari penelitian ini yaitu klasifikasi kanker berdasarkan data RNA menggunakan metode Elman Recurrent Neural Network. Dalam jaringan recurrent terdapat sebanyak dua buah input, yaitu input sebenarnya dan contextual input. Proses jalannya iterasi jauh lebuh cepat karena adanya *feedback*, sehingga pembaruan parameter serta konvergensi juga menjadi lebih cepat. Data yang digunakan adalah data RNA dengan empat kelas yaitu BRCA atau breast adenocarcinoma (kanker payudara), KIRC atau kidney renal clear cell carcinoma (kanker ginjal), UCEC atau uterine corpus endometrial carcinoma (kanker rahim), dan LUAD atau lung adenocarcinoma (kanker Data akan dilakukan preprocessing menggunakan minmax scaler kemudian diklasifikasi menggunakan ERNN dengan uji coba pembagian data, learning rate, dan jumlah hidden layer. Kombinasi parameter paling baik didapatkan pada 20 node hidden layer I, 50 node hidden layer II, dan learning rate Pada model tersebut didapatkan akurasi mencapai 99.19 %, sensitivitas sebesar 99.03 % dan Spesifisitas sebesar 99.72 %. Waktu yang dibutuhkan untuk model tersebut yaitu 19 detik.

Kata kunci: RNA, Kanker, ERNN, Klasifikasi, Ekspresi gen.

#### **ABSTRACT**

## CLASSIFICATION OF CANCER BASED ON RNA DATA USING *ELMAN*RECURRENT NEURAL NETWORK

Cancer is one of the deadliest diseases whose number of sufferers continues to increase every year. The development of cancer cells can quickly spread to all parts of the body through the bloodstream or from the lymphatic system so that it can cause death. This can happen because there is a disorder that exists in the gene. The basic thing in genetics is the monitoring of gene expression itself, namely by measuring from mRNA not from protein because 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 Elman Recurrent Neural Network method. In the recurrent network there are two inputs, namely the actual input and the contextual input. The iteration process is much faster due to feedback, so parameter updates and convergence are also faster. The data used are RNA data with four classes, namely BRCA or breast adenocarcinoma (breast cancer), KIRC or kidney renal clear cell carcinoma (kidney cancer), UCEC or uterine corpus endometrial carcinoma (uterine cancer), and LUAD or lung adenocarcinoma (lung cancer). The data will be preprocessed using a minmax scaler then classified using ERNN with trials of data sharing, learning rate, and the number of hidden layers. The best combination of parameters was obtained at 20 nodes hidden layer I, 50 nodes hidden layer II, and learning rate 0.1. In this model, the accuracy reached 99.19 %, sensitivity of 99.03 % and specificity of 99.72 %. The time required for the model is 19 seconds.

**Keywords**: RNA, Cancer, ERNN, Classification, Gene expression.

## **DAFTAR ISI**

HA	ALAN	IAN JUDUL	j	
HALAMAN PERNYATAAN KEASLIAN			ii	
LE	LEMBAR PERSETUJUAN PEMBIMBING ii			
PE	NGE	SAHAN TIM PENGUJI SKRIPSI	iv	
LE	EMBA	AR PERSETUJUAN PUBLIKASI	V	
AF	BSTR	AK	vi	
AF	BSTR	ACT	vii	
DA	(FTA	R ISI	1	
DA	(FTA	R TABEL	3	
DA	(FTA	R GAMBAR <mark>.</mark>	4	
I	PEN	DAHULUAN <mark></mark>	6	
	1.1.	Latar Belakang Masalah	6	
	1.2.	Rumusan Masalah	13	
	1.3.		13	
	1.4.	Manfaat Penelitian	13	
	1.5.	Batasan Masalah	14	
	1.6.	Sistematika Penulisan	15	
II	TINJ	JAUAN PUSTAKA	17	
	2.1.	Ribonucleic Acid (RNA)	17	
	2.2.	Kanker Paru-paru	18	
	2.3.	Kanker Payudara	20	
	2.4.	Kanker Ginjal	21	
	2.5.	Kanker Rahim	22	
	2.6.	Minmax Scaler	23	
	2.7.	Jaringan Saraf Tiruan	24	
	2.8.	Komponen Jaringan Saraf Tiruan	26	
	2.9.	Arsitektur Jaringan Saraf Tiruan	26	

	2.10. Recurrent Neural Network (RNN)	28
	2.11. Elman Recurrent Neural Network (ERNN)	29
	2.12. Confusion Matrix	34
	2.13. Penyakit dan kesehatan dalam prespektif islam	37
Ш	I METODE PENELITIAN	40
	3.1. Jenis Penelitian	40
	3.2. Jenis dan Sumber Data	40
	3.3. Kerangka Penelitian	41
IV	HASIL DAN PEMBAHASAN	44
	4.1. <i>Preprocessing</i>	44
	4.2. Pelatihan dan Pengujian Model	49
	4.3. Evaluasi Model	62
	4.4. Aplikasi Klasifikasi Kanker	69
	4.5. Integrasi Keislaman	73
$\mathbf{V}$	PENUTUP	77
	5.1. Kesimpulan	77
	5.2. Saran	77
DA	AFTAR PUSTAKA	78

## **DAFTAR TABEL**

3.1	Sampel data RNA	41
3.2	Tabel evaluasi	43
4.1	Sampel data RNA sebelum normalisasi	49
4.2	Sampel data RNA sesudah normalisasi	49
4.3	Bobot dan bias input layer menuju hidden layer I	50
4.4	Bobot dan bias input layer menuju hidden layer I	51
4.5	Bobot dan bias hidden layer II menuju output layer	51
4.6	Data ke-1 proses training	52
4.7	Hasil net input layer menuju hidden layer	53
4.8	Hasil koreksi net j	58
4.9	Hasil $\delta_j$	59
4.10	Perbaikan bobot v	59
	Koreksi bias hidden layer I menuju hidden layer II	
4.12	Perbaikan bias hidden layer I menuju hidden layer II	61
4.13	Perbaikan bobot v	61
4.14	Hasil klasifikasi pada pembagian data 90%:10%	62
4.15	Hasil klasifikasi pada pembagian data 80%:20%	63
4.16	Hasil klasifikasi pada pembagian data 70%:30%	64

## DAFTAR GAMBAR

2.1	Struktur letak RNA	17
2.2	(a) citra screening thoracic computed tomography (CT) scan dari	
	kanker paru-paru malignant (b) Hasil rontgen kanker paru-paru	20
2.3	(a) Hasil mamografi kanker payudara (b) Hasil tangkapan oleh	
	kamera gambar termal	21
2.4	Hasil CT Scan kanker ginjal	22
2.5	Hasil biopsi kanker rahim	23
2.6	Fungsi Aktivasi	25
2.7	Single Layer Network	27
2.8	Multi Layer Network	28
2.9	Competitive Layer	28
2.10	Arsitektur RNN sed <mark>er</mark> hana	29
2.11	Arsitektur ERNN	30
2.12	Confusion Matrix	35
3.1	Diagram alir	42
4.1	Visualisasi data kanker payudara	45
4.2	Visualisasi data kanker ginjal	46
4.3	Visualisasi data kanker paru-paru	47
4.4	Visualisasi data kanker rahim	48
4.5	Performance model terbaik	65
4.6	Hasil confusion matrix model terbaik	66
4.7	Data RNA kanker payudara terklasifikasi salah	68
4.8	Data RNA kanker paru-paru terklasifikasi salah	68
4.9	Kemiripan data	69
4.10	Halaman utama aplikasi	70

4.11	11 Halaman pemilihan file dan halaman utama dengan data yang sudah		
	di masukkan	71	
4.12	Halaman klasifikasi single data	72	
4.13	Halaman klasifikasi multiple data	73	



#### **BABI**

#### **PENDAHULUAN**

#### 1.1. Latar Belakang Masalah

Kanker termasuk salah satu penyakit mematikan yang jumlah penderitanya terus bertambah setiap tahunnya. Pada 30 tahun belakangan ini klasifikasi penyakit kanker semakin berkembang, akan tetapi masih belum terdapat identifikasi macam kanker dengan pendekatan secara umum (Putri et al., 2020). Kanker merupakan penyakit yang menyebabkan kematian manusia paling utama pada berbagai negara di dunia. WHO menyatakan ketika tahun 2019 telah ada 18.1 juta kasus terbaru serta 9.6 juta kasus kematian yang diakibatkan oleh penyakit kanker, serta jumlahnya diprediksi akan terus meningkat sebesar 70% dalam dua dekade yang akan datang (Adiwijaya, 2018). Para peneliti melakukan berbagai penelitian dengan segala metode guna mengetahui pendekatan yang paling baik untuk mengklasifikasikan kanker. Diantaranya yaitu mendiagnosa kanker dengan menggunakan ekspresi gen (Putri et al., 2020).

Perkembangan dari sel kanker dapat dengan cepat menyebar ke seluruh bagian tubuh mana saja dengan masuk melalui aliran darah atau dari sistem limfatik dan hal tersebut dapat menyebabkan kematian. Hal tersebut dapat terjadi karena terdapat kekacauan yang ada pada gen. Hal dasar dalam genetika yaitu pemantauan dari ekspresi gen itu sendiri, yaitu dengan mengukur dari mRNA bukan dari proteinnya dikarenakan urutan dari mRNA yang akan berhibridasi dengan komplementer dari DNA dan juga RNA (Rusamsi et al., 2018). Ketika

proses membuat protein, sebuah instruksi akan ditranskripsi oleh *messenger RNA* (mRNA) yang berasal dari DNA dan berada pada gen sisa didalam nukleus serta proses membuat ribosom sel. Hasilnya yaitu keadaan sel dan perubahan level dalam mRNA akan berkolerasi. Pengukuran level pada mRNA dilakukan agar mendapat informasi apa yang sudah terjadi dalam keadaan sel tersebut sehingga bisa memahami kejadian proses biologi didalamnya (Pradana et al., 2018).

Pada kehidupan sehari-hari, kesehatan termasuk aspek yang sangat diperhatikan bersamaan dengan aspek penting lainnya seperti perekonomian dan juga pendidikan. Pada pengertian *Ad-Dhoruriyyat* atau juga bisa disebut *maqashid syari'ah* yang disampaikan oleh al-Ghazali, ternyata suatu bidang kesehatan termasuk ke dalam *Hifdz An-Nafs* (memelihara jiwa). Hal tersebut memperlihatkan jika dalam kehidupan sehari-harinya sangat penting bagi manusia untuk memelihara jiwa mereka dari berbagai jenis gangguan yang ada seperti datangnya sebuah penyakit atau adanya musuh. Namun, ketika manusia memelihara dan menjaga jiwanya dalam aspek kesehatan juga harus sesuai dengan aturan atau arahan yang sudah ada.

Kesehatan sendiri memiliki dua arti, yakni kesehatan jasmani atau biasa juga disebut menjadi kata *as-shihah* serta kesehatan rohani yang juga biasa disebut *afiat*. Pada KBBI, kata *afiat* sendiri dipersamakan dengan kata *as-shihah*. *As-shihah* memiliki pengertian seluruh badan dalam keadaan baik dan semua bagian-bagiannya terbebas dari adanya penyakit, sedangkan *afiat* memiliki pengertian kuat dan sehat. Kata *al-afiyah* serta *as-shihah* sendiri sering muncul baik di ayat al-Qur'an, hadis, maupun doa-doa, diantaranya dalam al-Qur'an pada surah Al-Isra' ayat 82 sebagai berikut:

Artinya: "Kami turunkan dari al-Qur'an sesuatu yang menjadi penawar dan rahmat bagi orang-orang mukmin, sedangkan bagi orang-orang zalim (al-Qur'an itu) hanya akan menambah kerugian".

Ayat al-Qur'an pada surah Al-Isra' di atas memiliki kandungan hikmah jika al-Qur'an merupakan penyembuh untuk orang-orang yang beriman. Seorang mukmin ketika dibacakan ayat-ayat al-Qur'an akan bergetar dan menambah keimanannya baik jasmani maupun rohani serta rahmat Allah Swt. begitu sangat luas bagi orang yang beriman (Akbar and Budiyanto, 2020).

Sedangkan hadis tentang kesehatan yang diriwayatkan oleh al-Bukhari yaitu:

حَدَّثَنَا الْمَكِيُّ بْنُ إِبْرَاهِيمَ أَخْبَرَنَا عَبْدُ اللَّهِ بْنُ سَعِيدٍ هُوَ ابْنُ أَبِي هِنْدٍ عَنْ أَبِيهِ عَنْ ابْنِ عَبَّاسٍ رَضِيَ اللَّهُ عَنْهُمَا قَالَ,قَالَ النَّبِيُّ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ عُمَتَانِ مَغْبُونٌ فِيهِمَا كَثِيرٌ مِنْ النَّاسِ الصِّحَّةُ وَالْفَرَاغُ قَالَ عَبَّاسٌ الْعَنْبَرِيُّ حَدَّثَنَا صَفْوَانُ بْنُ عِيسَى عَنْ عَبْدِ اللَّهِ بْنِ سَعِيدِ بْنِ أَبِي هِنْدٍ عَنْ أَبِيهِ مِعْتُسَ ابْنَ عَبَّاسٍ عَنْ النَّبِيِّ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ مِثْلَهُ

Artinya: "Telah menceritakan kepada kami Al Makki bin Ibrahim telah mengabarkan kepada kami Abdullah bin Sa'id yaitu Ibnu Abu Hind dari Ayahnya dari Ibnu Abbas radliallahu 'anhuma dia berkata; Nabi Muhammad Saw.: "Dua kenikmatan yang sering dilupakan oleh kebanyakan manusia adalah kesehatan dan waktu luang." 'Abbas Al 'Anbari mengatakan; telah menceritakan kepada kami Shufwan bin Isa dari Abdullah bin Sa'id bin Abu Hind dari Ayahnya saya mendengar Ibnu Abbas dari Nabi Muhammad Saw. seperti hadis di atas".

Berdasarkan bunyi dari hadis tentang kesehatan yang diriwayatkan oleh al-Bukhari tersebut, Allah Swt. telah memberikan dua kenikmatan kepada hamba-Nya. Namun, manusia sering kali melupakan hal itu, yaitu nikmat

kesehatan dan nikmat waktu luang. Sesungguhnya suatu kerugian bagi seorang manusia ketika dirinya tidak memiliki rasa syukur terhadap segala hal yang telah diberikan oleh Allah Swt. Oleh karena itu, harusnya manusia patut senantiasa bersyukur kepada Allah swt., sehingga kita masih diberi kenikmatan berupa kesehatan dan waktu luang. Dari hadis tersebut kita bisa menjadikannya pelajaran untuk selalu menjaga kesehatan kita, agar senantiasa bisa menjalankan perintah Allah Swt. dalam menggunakan waktu luang sebaik-baiknya serta menghindari perbuatan yang diharamkan oleh Allah Swt., seperti yang sudah tercantum dalam al-Qur'an (Chairi).

Menjaga kesehatan menjadi sangat penting karena hal tersebut juga berpengaruh terhadap kegiatan ibadah dan perbuatan sesama manusia. Jika kita memiliki tubuh yang sehat dan bugar maka kita dapat beribadah secara maksimal serta melakukan berbagai hal baik terhadap sesama manusia. Maka dari itu lebih baik menjaga kesehatan dari pada mengobati suatu penyakit. Hal ini sesuai dengan salah satu kaidah fikih yang berkaitan mengenai pencegahan penyakit.

Artinya: "Menolak kerusakan lebih utama dari menerima kemaslahatan".

Menurut pendapat Azam, jika bertentangan kerusakan dan kemaslahatan, maka lebih utama mendahulukan menolak kerusakan. Hal tersebut dikarenakan *syara*' terhadap meninggalkan suatu yang terlarang itu lebih besar daripada melakukan suatu yang diperintahkan. Selain itu di dalam sesuatu yang terlarang akan ada hikmah yanng terkandung di dalamnya. Hal ini sama halnya berlaku pada kita yang senantiasa harus melakukan pencegahan akan datangnya suatu penyakit (Shodiq, 2020).

Berkaitan dengan pencegahan akan datangnya penyakit, salah satu

penyebab adanya suatu penyakit yaitu karena terdapat kekacauan yang ada pada gen. Hal dasar dalam genetika yaitu pemantauan dari ekspresi gen itu sendiri, yaitu dengan mengukur dari mRNA. Data genom makhluk hidup seperti RNA ini, didapatkan melalui teknik Whole Genome Sequencing (WGS) menggunakan komputasi dan algoritme kompleks yang disebut bioinformatika. Pengurutan asam ribonukleat (RNA-Seq) mengukur tingkat ekspresi beberapa transkrip secara bersamaan. Pembacaan dapat berupa gen, ekson, atau wilayah minat lainnya. Berbagai alat komputasi telah dikembangkan untuk mempelajari patogen atau virus dari data RNA-Seq dengan mengklasifikasikannya menurut atribut dalam beberapa kelas yang telah ditentukan. Pengembangan model klasifikasi sangat dianjurkan untuk diagnosis dan klasifikasi penyakit, pemantauan penyakit pada tingkat molekuler dan penelitian biomarker penyakit potensial. Kemajuan dalam bioinformatika ini, bersama dengan perkembangan dalam klasifikasi berbasis pembelajaran mesin, akan menyediakan suatu alat yang kuat untuk klasifikasi informasi transkriptom yang tersedia melalui data RNA-Seq (Jabeen et al., 2018).

Dalam dunia komputasi terdapat istilah *Artificial Intelligence* atau kecerdasan buatan yaitu berbagai macam metode dan algoritma yang bekerja berdasarkan kecerdasan mesin dan memiliki banyak aplikasi di berbagai bidang ilmu, termasuk teori dan aplikasi logika fuzzy, buatan teknik kecerdasan dan sosiologi, penilaian risiko dan identifikasi bahaya, pembelajaran mesin, dan algoritma meta-heuristik dan teknik pengelompokan (Pirouz et al., 2020). Dengan kemajuan kecerdasan buatan dalam diagnostik medis, konsep yang digunakan dalam pembelajaran mesin yaitu di mana apa yang disebut pembelajar ansambel menggunakan seperangkat pengklasifikasi yang berbeda beda guna menghasilkan klasifikasi yang lebih tepat daripada pengklasifikasi tunggal (Hekler et al., 2019).

Salah satu metode pada kecerdasan buatan yaitu metode jaringan saraf tiruan. Arsitektur yang terdapat pada jaringan saraf tiruan sangat beragam yang digunakan sesuai dengan kebutuhan rancangan dalam suatu aplikasi. Jaringan pada metode Elman Recurrent Neural Network (ERNN) sering disebut dengan jaringan recurrent yang bersifat parsial karena pada jaringan ini umumnya hanya memiliki koneksi yang hanya berbentuk feed forward. Jaringan Simple recurrent network mempunyai bentuk arsitektur yang hampir sama dengan jaringan recurrent spacemodel dengan hidden layer yang bersifat non linear tanpa adanya unit delay. Dalam jaringan recurrent terdapat sebanyak dua buah input, yaitu input sebenarnya dan contextual input. Proses jalannya iterasi jauh lebuh cepat karena adanya feedback, sehingga pembaruan parameter serta konvergensi juga menjadi lebih cepat (Singh et al., 2018). Pada penelitian sebelumnya yang pernah mengklasifikasikan kanker berdasarkan data RNA menggunakan preprocessing normalisasi menghasilkan akurasi yang bagus dengan menggunakan 11 kelas menghasilkan akurasi 96.70% pada test dataset (Zhao et al., 2019).

Klasifikasi kanker menggunakan data RNA telah dilakukan oleh beberapa peneliti, diantaranya yaitu Ho Sun Shon dkk menggunakan metode *Convolutional Neural Network* (CNN) dengan permasalahan kanker lambung yang menghasilkan yang menghasilkan nilai akurasi sebesar 95.96% (Shon et al., 2019). Peneliti lain yaitu Joseph M. de Guia dkk dengan penelitiannya yang mengklasifikasikan kanker dengan metode *DeepGx* mendapatkan nilai akurasi 95.65% untuk data 33 kelas (Joseph et al., 2019). Peneliti lain yang melakukan klasifikasi kanker adalah Murtada K. Elbashir dkk tentang implementasi algoritma CNN untuk klasifikasi kanker payudara menggunakan data RNA dengan hasil akurasi 98.76% (Elbashir et al., 2019). Klasifikasi kanker menggunakan data RNA menghasilkan akurasi

yang cukup baik dengan metode Neural Network.

Dalam permasalahan lain terdapat peneliti yang menggunakan ERNN untuk klasifikasi yaitu S. Ramkumar, J dkk mengklasifikasikan Eog Signal menghasilkan rata-rata akurasi 91% dan 93.57% (Thilagaraj et al., 2021). Sedangkan peneliti Wildan Kurniadi dan Irfan Maliki menggunakan metode ERNN untuk klasifikasi *Speech Recognition* menghasilkan akurasi terbaik 83.88% (Kurniadi and Maliki, 2018). Peneliti lain dengan permasalahan Financial menggunakan data berpola mengaplikasikan Elman Recurrent Random Neural Network serta forecasting model BPNN, STNN, ERNN, dan ST-ERNN. Pada ERNN menghasilkan MAPE terkecil sebesar 0.5191 (Wang et al., 2016). Selanjutnya peneliti Prisca Deviani Pakan dengan permasalahan kanker menggunakan data mikro array dengan 10 kelas menghasilkan MSE mencapai 0.15 (Pakan, 2019). Pada penelitian terdahulu yang telah disebutkan menghasilkan nilai akurasi yang cukup bagus menggunakan metode ERNN pada permasalahan klasifikasi sehingga metode ini cocok digunakan untuk kasus klasifikasi kanker.

Berdasarkan ulasan berbagai penelitian mengenai permasalahan klasifikasi kanker menggunakan data RNA dan kelebihan metode ERNN untuk klasifikasi. Maka dalam penelitian ini akan dilakukan klasifikasi kanker berdasarkan data RNA menggunakan metode ERNN agar dapat membantu mengklasifikasi kanker dengan baik dan mendapatkan nilai akurasi yang bisa diperhitungkan sehingga dapat digunakan untuk diagnosa dini penyakit kanker yang selanjutnya dapat di evaluasi lebih lanjut dengan pemeriksaan yang lainnya.

#### 1.2. Rumusan Masalah

Berdasarkan isi penjelasan dari latar belakang, terdapat rumusan masalah yang sudah dibuat sebagai berikut:

- Bagaimana hasil dan evaluasi model dari klasifikasi kanker berdasarkan data RNA dengan metode ERNN?
- 2. Bagaimana model arsitektur ERNN yang paling optimal untuk klasifikasi kanker berdasarkan data RNA menggunakan parameter pembagian data, hidden layer dan learning rate?

#### 1.3. Tujuan Penelitian

Berdasarkan dari poin-poin rumusan masalah,, maka tujuan pada penelitian ini sebagai berikut:

- Dapat mengetahui hasil klasifikasi kanker berdasarkan data RNA menggunakan metode ERNN dan evaluasi model menggunakan confusion matrix.
- 2. Dapat mengetahui model arsitektur yang paling optimal dari klasifikasi kanker berdasarkan data RNA menggunakan ERNN dengan parameter pembagian data, *hidden layer* dan *learning rate*.

#### 1.4. Manfaat Penelitian

Berbagai manfaat dari penelitian ini disebutkan sebagai berikut:

#### 1. Manfaat Teoritis

Dari segi teoritis, penelitian ini memberikan tambahan wawasan dan ilmu

pengetahuan dalam bidang sains dan teknologi serta dapat digunakan untuk referensi dalam mengklasifikasikan kanker berdasarkan data RNA menggunakan metode ERNN.

#### 2. Manfaat Praktis

#### (a) Bagi Peneliti

Dapat dijadikan suatu sistem untuk mendeteksi penyakit kanker menggunakan data RNA dengan metode ERNN.

#### (b) Bagi Pembaca

Penelitian ini bisa dijadikan sebuah wawasan serta pengetahuan tentang klasifikasi penyakit kanker menggunakan metode ERNN.

#### (c) Bagi medis

Penelitian ini dapat menjadi bagian prototype alat kesehatan untuk membantu ahli medis dalam mengklasifikasikan kanker berdasarkan data RNA dengan akurasi yang maksimal.

#### 1.5. Batasan Masalah

Penelitian ini menetapkan batasan masalah sebagai berikut:

- Data yang akan diklasifikasikan pada penelitian ini yaitu dataset RNA dari Mendeley dengan empat kelas yaitu BRCA atau breast adenocarcinoma (kanker payudara), KIRC atau kidney renal clear cell carcinoma (kanker ginjal), UCEC atau uterine corpus endometrial carcinoma (kanker rahim), dan LUAD atau lung adenocarcinoma (kanker paru-paru).
- 2. Parameter yang digunakan untuk klasifikasi penyakit kanker menggunakan metode ERNN yaitu *learning rate*, *hidden layer*, dan pembagian data.

#### 1.6. Sistematika Penulisan

Cara penulisan pada penelitian terbagi manjadi lima bab yang berisikan sebagai berikut:

#### 1. BAB I PENDAHULUAN

Bagian dari bab ini berisikan mengenai penjelasan dari latar belakang masalah tentang penyakit kanker dan konsep terdeteksinya sel kanker melalui RNA menggunakan perhituangan algoritma kecerdasan buatan, rumusan dan batasan masalah yang digunakan, serta tujuan, manfaat, dan sistematika penulisan penelitian.

#### 2. BAB II TINJAUAN PUSTAKA

Bagian dari bab ini berisikan tentang literatur yang digunakan tentang RNA, Penjelasan mengenai berbagai penyakit kanker serta cara mendeteksinya yaitu kanker paru-paru, kanker payudara, kanker ginjal, kanker rahim, proses *preprocessing* dan penjelasan mengenai algoritma metode ERNN, serta metode evaluasinya yang menggunakan. *confusion matrix*.

#### 3. BAB III METODE PENELITIAN

Bagian dari bab ini berisikan mengenai sumber data yang digunakan serta jenis dan alur penelitian dari mulai pengumpulan data RNA manusia yang memiliki penyakit kanker paru-paru, kanker payudara, kanker ginjal, kanker rahim, tahapan proses penelitian mulai dari input data hingga hasil, tahapan uji coba serta analisis hasil model terbaik menggunakan tabel berisikan hasil *confusion matrix*.

#### 4. BAB IV HASIL DAN PEMBAHASAN

Bagian dari bab ini berisikan mengenai hasil yang didapatkan dari proses

hasil uji coba klasifikasi kanker menggunakan metode ERNN beserta pembahasannya terkait perhitungan metode ERNN 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

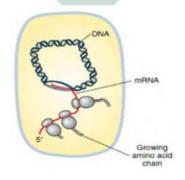
Bagian dari bab ini berisikan mengenai kesimpulan akhir dan saran.

#### **BAB II**

#### TINJAUAN PUSTAKA

#### 2.1. Ribonucleic Acid (RNA)

Pada umumnya, protein-coding messenger ribonucleic acid (mRNAs) adalah pengontrol utama sel, yang menjalankan fungsi yang diperlukan untuk kehidupan (Wang et al., 2021). Satuan dari kecepatan RNA ketika membaca gen terhadap DNA disebut Reads Per Kilobase Million (RPKM). Di lain sisi, RNA memiliki fungsi lain yaitu menjadi onkogen atau tumor suppressor gene. Onkogen adalah gen untuk meningkatkan keganasan dari sel tumor, sedangkan tumor suppressor gene merupakan gen untuk mencegah berkembangnya tumor. Adanya perubahan pada nilai suatu ekspresi gen terhadap RNA dijadikan indikator jika telah ada penyakit kanker dalam tubuh manusia. Teknik dalam pengambilan RNA memiliki berbagai macam cara serta tiap teknik juga mempunyai aturan dan susunan dari RNA yang dihasilkan berbeda-beda (Syarifudin, 2021).



Gambar 2.1 Struktur letak RNA

Sumber: (Syarifudin, 2021)

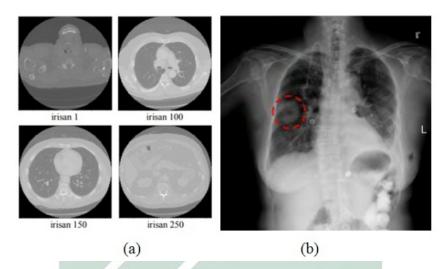
#### 2.2. Kanker Paru-paru

Kanker paru-paru termasuk penyakit ganas yang banyak ditemukan hampir di seluruh belahan dunia serta merupakan penyebab utama alasan kematian akibat penyakit keganasan. Ketika tahun 2015, American Cancer Society menyatakan jika kanker paru menjadi kasus nomer 2 paling banyak untuk penyakit dengan keganasan di mana nomer satu yaitu kanker prostat yang terjadi terhadap pria serta kanker payudara terhadap wanita (Husen et al., 2016). Dengan kasus kematian karena keganasan, kanker paru-paru ini berada pada urutan pertama baik terjadi terhadap pria ataupun wanita. Kanker paru-paru juga menduduki urutan nomer 6 dalam penderita sejumlah 4.9%. Faktor penyebab penyakit kanker paru-paru yaitu seorang perokok aktif maupun pasif, terdapat faktor lain juga seperti polusi udara yang kotor dan lingkungan kerja yang buruk (Prasetio and Susanti, 2019).

Penyakit kanker paru-paru ini juga sering kali ditemukan pada pasien yang tidak merokok, hal tersebut terjadi pada orang yang terpapar bahan kimia yang berbahaya atau bisa jadi seseorang yang menghirup asap rokok dari orang di sekitarnya, kondisi seperti itu biasa disebut perokok pasif. Upaya mengobati penyakit kanker paru-paru ini jika masih dalam tahap stadium awal maka keberhasilannya akan tinggi, namun apabila kanker paru-paru masih dalam tingkat stadium awal , tidak ada gejala yang ditimbulkan. Beberapa gejala umum pada penderita kanker paru-paru antara lain batuk yang kronis atau bahkan hingga mengeluarkan darah, berat badan penderita akan turun secara drastis, terasa nyeri pada bagian dada dan tulang juga disertai rasa sesak nafas. Beberapa faktor yang bisa menjadi penyebab adanya kanker paru-paru antara lain terdapat salah satu keturunan keluarga yang menderita kanker paru, daerah tempat tinggal atau lingkungan tempat bekerja yang tidak sehat dengan udara yang tercemar zat

berbahaya atau bisa jadi adanya polusi udara yang kotor, dan pernah melakukan radioterapi udara, serta pernah menjalani radioterapi udara (Priyanti, 2021).

Penyakit kanker dapat dideteksi dengan berbagai cara, salah satunya yaitu menggunakan citra digital, data patologi, dan data vektor RNA. Pengolahan citra digital saat ini digunakan dalam berbagai bidang riset. Teknologi CT Scan dikembangkan melalui proses pemindaian setiap lapisan-lapisan jaringan tubuh dari berbagai arah dengan menggunakan radiasi sinar X. Hasil dari pemindaian dengan menggunakan teknologi ini disebut sebagai citra X-ray Maulida et al. (2013). Sedangkan data RNA-Seq dengan mengklasifikasikannya menurut atribut dalam beberapa kelas yang telah ditentukan. Namun, alat komputasi dan pendekatan untuk menganalisis kumpulan data yang kompleks masih kurang. Pengembangan model klasifikasi sangat dianjurkan untuk diagnosis dan klasifikasi penyakit, pemantauan penyakit pada tingkat molekuler dan penelitian biomarker penyakit potensial. Dalam bab ini, kami membahas berbagai pendekatan pembelajaran mesin untuk klasifikasi data RNA-Seq dan implementasinya. Kemajuan dalam bioinformatika ini, bersama dengan perkembangan dalam klasifikasi berbasis pembelajaran mesin, akan menyediakan kotak peralatan yang kuat untuk klasifikasi informasi transkriptom yang tersedia melalui data RNA-Seq (Jabeen et al., 2018).



Gambar 2.2 (a) citra screening thoracic computed tomography (CT) scan dari kanker paru-paru malignant (b) Hasil rontgen kanker paru-paru.

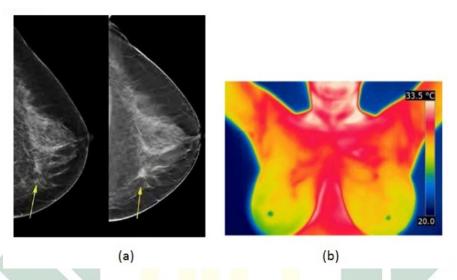
Sumber: (Anugerah, 2018)

#### 2.3. Kanker Payudara

Kanker payudara adalah penyakit di mana terdapat sel-sel ganas pada jaringan payudara. Penyakit ini termasuk salah satu kanker yang pada umumnya diderita oleh wanita pada negara-negara maju, namun di Indonesia menurut hasil patologi menunjukkan bahwa penyakit ini memiliki peringkat nomer dua dengan penderita paling banyak setelah kanker leher rahim (Sirait, 2009). Faktor risiko adanya penyakit ini yaitu jenis kelamin di mana perbandingan pria dan wanita 1:100. Faktor resiko lain dari kanker payudara ini yaitu riwayat penyakit keluarga dengan penyakit kanker payudara dengan kemungkinan 15,79%, *menarche* dini sebesar 8,77%, *nullipara* sebesar 7,02% serta penggunaan pil di mana kandungannya terdapat esterogen dengan konsumsi jangka panjang dengan resiko sebesar 42,11%. Selain itu, faktor resiko lainnya yaitu menopause yang terlambat, riwayat dari pemberian ASI, serta badan yang obesitas (Anggorowati, 2013).

Ada beberapa cara untuk mendeteksi kanker payudara secara dini, salah

satunya adalah dengan mengidentifikasi dan mendeteksinya melalui pemeriksaan rontgen dan kamera thermal. Citra yang dihasilkan dari sinar x tersebut disebut dengan mammogram, sedangkan image hasil kamera thermal disebut dengan termografi (RASMI, 2020).



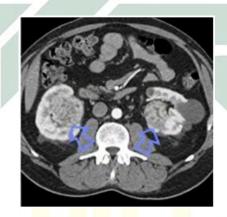
Gambar 2.3 (a) Hasil mamografi kanker payudara (b) Hasil tangkapan oleh kamera gambar termal

Sumber: (a) (Syarifudin, 2021) (b) (Ekici and Jawzal, 2020).

#### 2.4. Kanker Ginjal

Penyakit ginjal termasuk penyakit yang penderitanya tidak memandang jenis kelamin. Faktor adanya penyakit ini yaitu kebiasaan buruk seperti jarang meminum air putih, sering minum minuman yang mengandung alkohol, serta jarang olahraga. Penyakit ini akan sangat mudah terjadi ketika daya tahan tubuh mulai berkurang dan menurun, sehingga memungkinkan masuknya bakteri atau mikroorganisme ke dalam tubuh (Hendry et al., 2021).

Kanker ginjal merupakan penyebab kematian dengan urutan nomor 16. Tumor ginjal menduduki nomor tiga dengan jumlah kematian paling banyak dibawah urutan penyakit tumor prostat serta penyakit tumor kandung kemih. Asal dari tumor ginjal yaitu tumor primer yang terdapat pada ginjal maupun tumor sekunder yang muasalnya dari mestastasis keganasan pada tempat yang lain. Hasil dari statistik menyatakan jika insiden terjadinya kanker ginjal 2 kali lipat kemungkinannya menimpa pada jenis kelamin laki-laki jika dibandingkan dengan perempuan serta 59% terjadinya penyakit kanker ginjal terdapat pada negara berkembang dimana urutan paling atas yaitu Amerika Utara serta Eropa dan paling bawah yaitu Asia serta Afrika (Melisa et al., 2016).

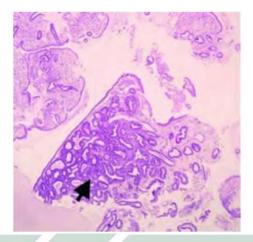


Gambar 2.4 Hasil CT Scan kanker ginjal

Sumber: (Syarifudin, 2021)

#### 2.5. Kanker Rahim

Kanker rahim yaitu tumor yang ganas dan tumbuh pada kantong selaput yang berada dalam perut atau rahim. Penyakit ini pada umumnya terjadi pada wanita dan sering kali dihubungkan dengan sperma. Penyebab dari penyakit ini yaitu dimana hormon proogresteron dan esterogen tidak seimbang (Saputra et al., 2020). Penyakit kanker rahim yaitu penyakit dengan keganasan ginekologi dimana kasus yang terjadi umumnya pada wanita (Christiyawati et al., 2016).



Gambar 2.5 Hasil biopsi kanker rahim

Sumber: (Syarifudin, 2021)

#### 2.6. Minmax Scaler

Normalisasi merupakan langkah *preprocessing* yang biasanya dilakukan agar data memiliki rentan yang sama (Anindya and Kantarcioglu, 2018). Salah satunya yaitu menggunakan Minmax Scaler dimana pada persamaannya menggunakan nilai Min serta Max pada proses pengubahan range data awal menjadi skala 0 sampai 1 (Zhang et al., 2019). Persamaan 2.1 menyatakan rumus perhitungannya (Harlianto et al., 2020):

$$x_i' = \frac{x_i - min(x)}{max(x) - min(x)}$$
(2.1)

 $x_i$ = Data ke-i

min(x)= Data terkecil

max(x)= Data terbesar

#### 2.7. Jaringan Saraf Tiruan

Jaringan Saraf Tiruan (JST) yaitu salah satu metode dalam kecerdasan buatan yang diaplikasikan untuk mengolah suatu informasi, prinsip kerja dari metode ini mirip dengan jaringan saraf yang ada pada manusia. Manusia memiliki suatu jaringan saraf yang terdiri dari sel dengan jumlah yang banyak dan dinamakan dengan neutron. Jaringan saaraf tiruan juga sering disebut dengan neural network (NN) yaitu suatu sistem yang memproses suatu informasi dengan karakteristik yang hampir sama dengan jaringan saraf biologis. NN dibentuk sebagai model generalisasi matematis jaringan saraf biologis dengan perumpamaan berikut:

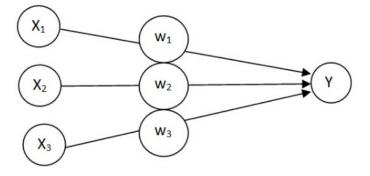
- 1. Proses pengolahan suatu informasi akan dilakukan dengan elemen sederhana yang dinamakan neuron dengan jumlah yang banyak.
- 2. Sinyal yang dikirim oleh neuron satu dengan neuron lainnya dihubungkan melalui sebuah tautan.
- 3. Tiap link antar neuron mempunyai nilai bobot yang fungsinya untuk mengukur kuat atau lemahnya suatu sinyal (bobot positif dapat memperkuat sinyal, sedangkan bobot negatif dapat melemahkan sinyal).
- 4. Tiap neuron akan memanfaatkan fungsi aktivasi (biasanya yang digunakan bukan fungsi linier) ke input (besar bobot sinyal input) untuk menentukan sinyal output (Walid, 2019).

Di dalam neutron memiliki 3 elemen utama yang bisa disamakan dengan jaringan saraf tiruan adalah soma, akson dan dendrit. Dalam jaringan saraf tiruan istilah neutron juga disebut sebagai node atau unit. Tiap node akan diasosiasikan

terhadap node lain melalui sebuah layer dengan nilai bobot yang besarnya sudah ditentukan. Bobot ini berfungsi untuk membawa informasi yang akan digunakan dalam proses memecahkan masalah. Pada jaringan saraf manusia juga terdapat proses serupa yang terjadi pada celah sinaptik dalam suatru proses kimia. Pada setiap lapisan berisi node-node yang memiliki sifat atau karakteristik yang sama. Setiap node akan memiliki fungsi aktivasi yang berasal dari input (Sudarsono, 2016). Jaringan saraf tiruan yaitu generalisasi bentuk matematika dari pemahaman manusia yang dikembangkan untuk memperoleh informasi yang diinginkan. Jaringan saraf memiliki ciri-ciri sebagai berikut:

- 1. Pola hubungan antar neutron disebut arsitektur jaringan.
- 2. Metode yang digunakan untuk menentukan nilai suatu bobot sebagai penghubung dinamakan pelatihan atau pembelajaran atau juga sebagai algoritma.
- 3. Fungsi aktivasi dijalankan oleh neutron dari *input* jaringan untuk mendapatkan *output* (Andriyani and Sitohang, 2018).

Perhatikan fungsi aktivasi pada 2.6 menunjukkan X berupa input, w berupa bobot, dan Y merupakan hasil output.



Gambar 2.6 Fungsi Aktivasi

#### 2.8. Komponen Jaringan Saraf Tiruan

Pada dasarnya *output*. Sejalan dengan perkembangan ilmu dan teknologi di masa sekarang, jaringan saraf tiruan memiliki tambahan satu lapisan, yaitu lapisan tersembunyi yang terletak di antara lapisan *input* dan lapisan *output* (Lesnussa et al., 2015)):

#### 1. Input Layer

Lapisan ini terdiri dari node-node *input*. Unit input akan menerima pola data untuk menggambarkan suatu masalah.

#### 2. Hidden Layer

Node-node di lapisan ini disebut unit tersembunyi. Selurus proses akan dilakukan pada lapisan ini, tetapi *output*-nya tidak dapat diamati secara langsung.

#### 3. Output Layer

Lapisan ini berisi node keluaran dari jaringan syaraf tiruan untuk suatu masalah.

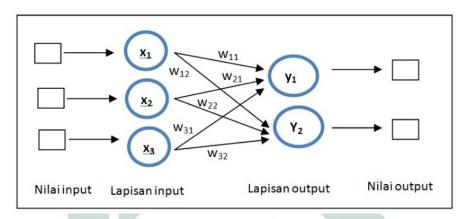
#### 2.9. Arsitektur Jaringan Saraf Tiruan

Pada penerapannya, jaringan saraf tiruan mempunyai berbagai macam arsitektur jaringan dalam aplikasi penggunaannya arsitektur dipilih berdasarkan dari desainnya. Berikut ini berbagai macam arsitektur jaringan saraf tiruan (Andriyani and Sitohang, 2018):

#### 1. Jaringan Lapisan Tunggal (Single Layer Network)

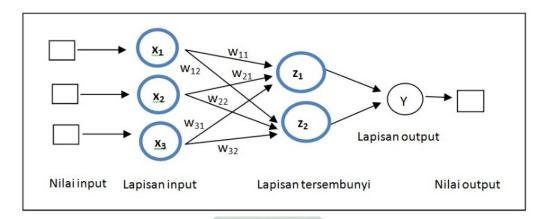
Arsitektur terdiri dari terdiri dari lapisan *input* dan lapisan *output* yang jumlahnya masing-masing satu. Neuron-neuron pada lapisan *input* akan

terhubung langsung dengan neuron-neuron pada lapisan *output*. Jadi pada jaringan ini hanya menerima *input* kemudian langsung diproses menjadi *output* tanpa harus melalui *hidden layer*.



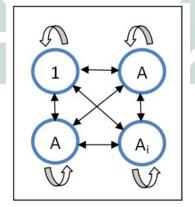
Gambar 2.7 Single Layer Network

2. Jaringan Banyak Lapisan (*Multi Layer Network*) Jaringan ini terdiri dari banyak lapisan yang memiliki fungsi berbeda, yang terdiri dari tiga lapisan, yaitu lapisan *input*, lapisan tersembunyi, dan lapisan *output*. Bentuk arsitektur yang seperti ini dapat memecahkan masalah yang lebih kompleks dari pada jaringan dengan lapisan tunggal, tetapi proses pelatihannya seringkali memakan waktu lama. jaringan dengan n input  $(x_1, x_2, \ldots)$ , lapisan tersembunyi yang terdiri dari p neuron  $(z_1, z_2, \ldots)$ , dan m neuron output  $(Y_1, Y_2, \ldots)$ .



Gambar 2.8 Multi Layer Network

3. Jaringan apisan Kompetitif (*Competitive Layer*) Jaringan dengan lapisan kompetitif memiliki model jaringan yang sama dengan jaringan tunggal atau ganda. Bedanya ada loop umpan balik, di mana sinyal mengalir dua arah, baik maju atau mundur. Jaringan *Competitive Layer* ditunjukkan pada Gambar 2.9.

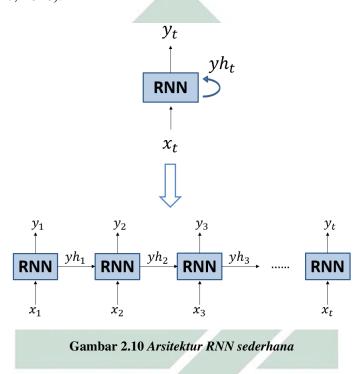


Gambar 2.9 Competitive Layer

#### 2.10. Recurrent Neural Network (RNN)

Recurrent Neural Network adalah salah satu bagian dari model artificial neural network yang mempunyai feedback dari keluaran hidden layer ke masukan

input layer. Recurrent Neural Network mempunyai struktur dan algoritma pelatihan yang lebih kompleks dibandingkan feed-forward neural network. Pada arsitektur model ini, output dari network digunakan kembali sebagai input network. Sifat dasar Recurrent Neural Network tersebut membuktikan bahwa metode ini tepat diterapkan untuk data berbentuk sequence atau list (Radjabaycolle, 2020).

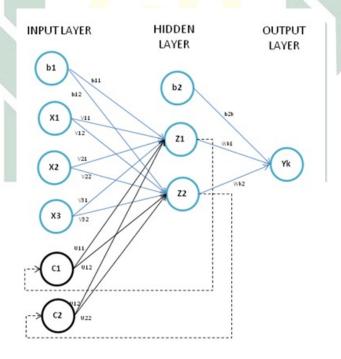


#### 2.11. Elman Recurrent Neural Network (ERNN)

ERNN adalah salah satu metode yang dikembangkan dari dari RNN (Radjabaycolle, 2020). Metode ini digunakan pada berbagai macam bidang seperti, *language modeling*, dan *machine translation* (Walid, 2019). Arsitektur ERNN juga biasa disebut sebagai jaringan repetitif parsial karena bentuknya yang memiliki koneksi berupa *feed forward* (Dhahri, 2020). Adannya *hidden layer* pada jaringan ini mengakibatkan ERNN dapat mempelajari hubungan antara *input layer* dengan *output layer*. Hal tersebut terjadi karena banyaknya jumlah neuron pada

hidden layer (Wang et al., 2021).

ERNN juga merupakan salah satu variasi dari Multi Layer Perceptron (MLP). Perbedaannya pada ERNN terdapat lapisan tambahan yang berdekatan dengan *input layer* dan *hidden layer*. Lapisan ini berisikan informasi yang berasal dari waktu sebelumnya. Pada dasarnya metode ini memiliki prinsip kerja dimana *input* akan disalurkan secara *feed forward* kemudian akan diberikan aturan pembelajaran. Jaringan bisa mengenali urutan dan melakukan pekerjaan yang bersamaan, misalkan dalam kasus prediksi urutan yang terletak di luar MPL. Tidak hanya menggunakan nilai dari luar jaringan akan tetapi juga akan ditambahkan nilai dari propagasi sebelumnya yang merupakan output dari hidden layer. Berikut arsitektur ERNN ditunjukkan pada Gambar 2.11.



Gambar 2.11 Arsitektur ERNN

Secara umum, jaringan Elman menggunakan fungsi aktivasi sigmoid bipolar pada lapisan tersembunyi (*recurrent*) dan linier untuk lapisan keluaran.

Berbeda dengan *Backpropagation Neural Network* (BPNN), pada jaringan Elman ini, fungsi aktivasinya bisa berupa fungsi apa saja, baik kontinu maupun diskontinu. Jumlah neuron pada setiap jaringan yang diatur cukup banyak, begitu juga dengan jumlah *hidden layer* yang disesuaikan dengan kompleksitas permasalahan. Penundaan yang terjadi pada waktu sebelumnya dapat digunakan untuk waktu sekarang. Seperti halnya BPNN, proses pelatihan pada jaringan Elman dapat dilakukan dengan dua cara yaitu dengan melatih dan mengadaptasi fungsi (Walid, 2019). Setiap kali sebuah pola disajikan, unit menghitung aktivasinya seperti halnya dalam jaringan *feed forward*. Adapun tahapan dalam algoritma JST *Recurrent Elman* adalah sebagai berikut:

## Langkah 1:

Pelatihan dilakukan pada pola *vector* tiap waktu(t) atau masukan menuju *hidden* layer menggunakan jaringan recurrent

$$net_j(t) = y_{hj} + \sum_{h=0}^{m} u_{hj} y_h(t-1) + \theta_j$$
 (2.2)

$$net_{j}(t) = \sum_{i}^{n} v_{ij} x_{i}(t) + \sum_{h}^{m} u_{hj} y_{h}(t-1) + \theta_{j}$$
(2.3)

Pada perhitungan pertama  $context\ layer$  masih belum memiliki nilai dari hidden layer pada perhitungan sebelumnya sehingga nilai  $y_h$  ini di copy dari  $hidden\ layer$  menuju ke  $context\ layer$  menjadi  $y_h(t-1)$ . Selanjutnya nilai  $y_h(t-1)$  dihitung dengan bobot awal dari  $context\ layer$  menuju  $hidden\ layer$  menggunakan nilai bobot v awal dari  $input\ layer$  ke  $hidden\ layer$ . Sehingga pada perhitungan pertama menggunakan Persamaan berikut.

$$net_j(t) = \sum_{i=1}^{n} v_{ij} x_i(t) + \sum_{i=1}^{n} v_{ij} y_h(t-1) + \theta_j$$
 (2.4)

Pada perhitungan kedua,  $context\ layer$  menggunakan nilai hasil copy dari keluaran  $hidden\ layer$  pada perhitungan sebelumnya (perhitungan pertama) dan menggunakan bobot  $u_{hj}$  yang berukuran sesuai dengan banyak  $node\ hidden\ layer$ . Hal ini juga berlaku pada perhitungan seterusnya hingga selesai.

 $x_i = \text{nilai } input \text{ ke-i, i=1,2,3...dst.}$ 

 $v_{ij}$  = nilai bobot *input layer node* ke i menuju *hidden layer node* ke j

 $u_{hj}=$ nilai bobot $context\ layer\ node$ ke h<br/> menuju  $hidden\ layer\ node$ ke j

 $y_h$  = hasil copy nilai keluaran *hidden layer* ke t-1

 $\theta_j$  = bias *node* ke-j

n = jumlah node input

h = indeks node context

 $m = \text{jumlah } node \ hidden \ layer$ 

Hitung keluaran hidden laye $r(y_i)$  dengan menggunakan fungsi aktivasi.

$$f(net_j(t)) = \frac{1}{1 + \epsilon^{-net_j}}$$
 (2.5)

Langkah 2:

Hasil keluaran akan ditentukan dengan lapisan/layer context-nya serta output yang memiliki nilai bobot (w)

$$y_k(t) = g(net_k(t)) (2.6)$$

Selanjutnya akan lebih lanjut terjabarkan seperti berikut

$$net_k(t) = \sum_{j=0}^{m} w_{jk} f(net_j(t)) + \theta_j$$
(2.7)

 $y_k$  = hasil keluaran dari fungsi aktivasi  $y_{net_k}$ 

 $w_{jk}$  = nilai bobot *hidden layer* menuju *output layer* 

## Langkah 3:

Nilai bobot pada jaringan ini akan selalu diperbarui, yang mana dihitung dengan koreksi bobot kemudian akan digunakan untuk perhitungan menentukan nilai bobot selanjutnya yang akan dipakai.

## Langkah 4:

Menghitung kesalahan setiap layer untuk menghitung koreksi bobot.

## Langkah 5:

Perhitungan dilakukan dengan aturan turunan rantai diferensiasi, nilai bobot yang berkaitan dengan *error* akan dihitung dengan rumus berikut

$$\delta_k = (d_k - y_k)g'(net_k(t)) \tag{2.8}$$

Sedangkan, untuk error di lapisan sembunyi dihitung menggunakan rumus dibawah

$$\delta_j = \sum_{k=1}^m \delta_k w_{jk} f'(net_j) \tag{2.9}$$

 $\delta_k$  = kesalahan pada *output layer* 

 $\delta_j$  = kesalahan pada *hidden layer* 

 $d_k = \text{target}$ 

 $y_k = \text{hasil } output$ 

#### Langka 6:

Koreksi nilai bobot pada keluaran akan menggunakan rumus seperti dibawah

$$\Delta w_{ik} = \eta \delta_k f(net_i) \tag{2.10}$$

Sedangkan untuk koreksi bobot pada masukan

$$\Delta v_{ij} = \eta \delta_i x_i \tag{2.11}$$

 $\Delta w_{jk}$  = koreksi bobot w

 $\Delta v_{ij}$  = koreksi bobot v

 $\eta = lerning rate$ 

Selanjutnya menghitung koreksi bias.

$$\Delta\theta_k = \eta \delta_k \tag{2.12}$$

 $\Delta \theta_k$  = koreksi bias

Langkah 7:

Berdasarkan pada nilai setiap waktunya(t), nilai bobot yang baru menggunakan penjumlahan koreksi bobot dan bobot lama. Lakukan hal tersebut pada setiap bobot.

$$w_{kj}baru = w_{kj}lama + \Delta w_{kj} \tag{2.13}$$

 $\Delta u_{hj}$  =perubahan bobot u

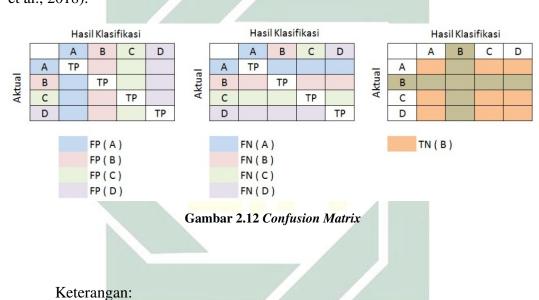
Langkah 8:

Pelatihan/testing hanya berhenti jika nilai error < target error (Radjabaycolle, 2020).

## 2.12. Confusion Matrix

Confusion Matrix yaitu alat evaluasi visual yang digunakan dalam pembelajaran mesin. Kolom dari Confusion Matrix mewakili hasil kelas prediksi, dan baris mewakili hasil kelas yang sebenarnya. Alat evaluasi ni menghitung

semua kemungkinan kasus klasifikasi pada permasalah. Mengambil masalah klasifikasi biner sebagai contoh, dimensi *Confusion Matrix* adalah 2 × 2. Serangkaian ukuran kinerja algoritma dapat ditentukan menggunakan *Confusion Matrix*, seperti tingkat pemeriksaan area hasil kelas positif dan tingkat penarikan kelas negatif. Langkah-langkah ini umumnya berlaku di semua algoritma klasifikasi (Xu et al., 2020). *Confusion Matrix* berisikan informasi hasil dari klasifikasi yang telah diprediksi dengan benar oleh sistem klasifikasi (Gunawan et al., 2018).



- 1. *True Positive* (TP) yaitu nilai yang mempresentasikan banyak kelas positif yang terklasifikasi secara benar di kelas positif.
- 2. *True Negative* (TN) yaitu nilai yang mempresentasikan banyak kelas negatif yang terklasifikasi secara benar di kelas negatif.
- 3. *False Positive* (FP) yaitu nilai yang mempresentasikan banyak kelas positif yang terklasifikasi salah di kelas negatif.
- 4. *False Negative* (FN) yaitu nilai yang mempresentasikan banyak kelas negatif yang terklasifikasi salah di kelas positif.

Selanjutnya akan dihitung spesifisitas, sensitivitas, dan akurasi. Berikut penjelasan dan rumusnya:

#### 1. Spesifisitas

Spesifisitas adalah hasil yang mewakili banyak data yang diklasifikasikan dengan benar dalam kelas negatif. Dengan demikian, semakin besar nilai spesifisitas, semakin baik suatu sistem dapat mengklasifikasikan kelas negatif dengan benar. Spesifisitas dapat dihitung menggunakan Persamaan 2.14 berikut:

$$Spesifisitas = \frac{\sum \frac{TN}{TN + FP}}{n}x100 \tag{2.14}$$

### 2. Sensitivitas

Sensitivitas adalah kebalikan dari spesifisitas, yaitu hasil mewakili jumlah data yang diklasifikasikan dengan benar dalam kelas positif. Dengan demikian, semakin besar sensitivitasnya, maka semakin naik sistem bisa mengklasifikasikan kelas positif dengan benar. Sensitivitas dapat dihitung menggunakan Persamaan 2.15 berikut:

$$Sensitivitas = \frac{\sum \frac{TP}{TP + FN}}{n} x 100 \tag{2.15}$$

#### 3. Akurasi

Akurasi adalah hasil yang mewakili banyak data yang diklasifikasikan menurut data sebenarnya. Besar akurasi berbanding lurus dengan kinerja sistem. Sehingga suatu sistem dapat disebut baik apabila memiliki nilai akurasi yang besar. Dengan kata lain hasil klasifikasi harus menunjukkan banyak data yang terklasifikasi sesuai dengan kelasnya. Akurasi dapat

dihitung menggunakan Persamaan 2.16 berikut:

$$Akurasi = \frac{TP_{all}}{n_{all}}x100 (2.16)$$

### 2.13. Penyakit dan kesehatan dalam prespektif islam

Berbagai pendapat dari oarang-orang barat menganggap bahwa ibadah merupakan suatu hal yang irasional. Menurut pandangan merka ibadah bukan hal yang dapat dikaji secara ilmiah, yang diapat ditunjukkan bukti kebenarannya menggunakan data serta fakta. Akan tetapi, banyak penelitian yang telah dilakukan membuktikan bahwa ibadah mempunyai dampak signifikan pada kesehatan dan pengobatan seseorang.

Padahal hal ini sudah dapat kita temui sejak zaman Nabi Muhammad Saw. Menurut Amin Syukur dalam bukunya yang berjudul "Zikir Menyembuhkan Kankerku", beliau menuliskan jika dalam kehidupan ini terdapat dua macam penyakit, yaitu fisik dan psikis. Kedua sumber penyakit tersebut berasal dari pikiran seseorang atau perasaan dan juga bisa berasal dari perut. Hal tersebut sesuai dengan perkataan dari salah satu seorang tabib dari Arab yaitu Al-Harith bin Kildah, beliau mengatakan bahwa perut merupakan sumber dari suatu penyakit serta makanan merupakan oabat dari berbagai macam penyakit. Baik itu penyakit fisik maupun psikis keduanya saling berpengaruh satu sama lain. Sakit fisik yang terasa nyeri akan berpengaruh pada psikologis, dan sakit psikologis juga mempengaruhi sakit fisik. Maka dari itu pengobatan suatu penyakit tidak hanya dilakukan secara medis, akan tetapi juga dilakukan melalui doa, zikir, serta ibadah. Menurut Amin Syukur, pengobatan yang dilakukan dengan perpaduan medis serta zikir dinamakan pengobatan "min-plus" (Warto, 2019).

Suatu penyakit datang dari Allah Swt. pasti karena suatu alasan sehingga kita harus senantiasa bersabar. Hal ini juga termasuk salah satu cara seorang hamba berikhtiar kepada Allah Swt. Cobaan berupa penyakit juga dialami oleh Nabi Ayub A.s. sebagaimana disebutkan pada surah Al-Anbiya ayat 83, Allah Swt berfirman,

Artinya:"(Ingatlah) Ayyub ketika dia berdoa kepada Tuhannya, "(Ya Tuhanku,) sesungguhnya aku telah ditimpa penyakit, padahal Engkau Tuhan Yang Maha Penyayang dari semua yang penyayang"

Selain itu, dalam agama Islam diajarkan pula agar manusia segera menyembuhkan dirinya, seperti mencari obat atau pergi ke tabib apabila sedang sakit. Karena setiap penyakit yang datang pasti ada obatnya. Hal tersebut menunjukkan bahwa jika itu berusaha untuk menyembuhkan diri maka pasti ada jalannya. Hal tersebut sesuai dengan salah satu sabda Nabi Muhammad Saw., yang berbunyi:

حَدَّثَنَا حَفَصُ بُنُ عُمَرَ النَّمَرِيُّ حَدَّثَنَا شُعْبَةُ عَنْ زِيَادِبْنِ عِلَاقَةَ عَنْ أَسَامَةَ بْنِ شَرِيكٍ قَالَ أَتَيْتُ النَّبِيَّ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ وَأَصْحَابَهُ كَأَنَّمَا عَلَى رُءُوسِهِمُ الطَّيْرُ فَسَلَّمْتُ ثُمَّ قَعَدْتُ فَجَاءَ النَّهِ أَنتَدَاوَى فَقَالَ تَدَاوَوَا قَعَدْتُ فَجَاءَ الْأَعْرَابُ مِنْ هَا هُنَا وَهَا هُنَا فَقَالُوا يَا رَسُولَ اللَّهِ أَنتَدَاوَى فَقَالَ تَدَاوَوَا فَعَدْتُ فَجَاءَ الْأَعْرَابُ مِنْ هَا هُنَا وَهَا هُنَا فَقَالُوا يَا رَسُولَ اللَّهِ أَنتَدَاوَى فَقَالَ تَدَاوَوا فَإِنَّ اللَّهَ عَزَ وَجَلَّ لَمْ يَضَعُ دَاءً إِلَّا وَضَعَ لَهُ دَوَاءً غَيْرَ دَاءٍ وَاحِدٍ الْهَرَمُ (رواه أبو داوود)

Artinya: "(ABUDAUD - 3357): Telah menceritakan kepada kami Hafsh bin Umar An Namari telah menceritakan kepada kami Syu'bah dari Ziyad bin 'Ilaqah dari Usamah bin Syarik ia berkata, "Aku pernah mendatangi Nabi Muhammad Saw. dan para sahabatnya, dan seolah-olah di atas kepala mereka terdapat burung. Aku

kemudian mengucapkan salam dan duduk, lalu ada seorang Arab badui datang dari arah ini dan ini, mereka lalu berkata, "Wahai Rasulullah, apakah boleh kami berobat?" Beliau menjawab: "Berobatlah, sesungguhnya Allah Swt. tidak menciptakan penyakit melainkan menciptakan juga obatnya, kecuali satu penyakit, yaitu pikun."." (HR. Abu Dawud dari Abud Darda' radhiallahu 'anhu).

Penyakit sangat penting dihindari karena dapat membahayakan diri sendiri, terutama penyakit ganas seperti kanker yang dapat menyebabkan kematian. Oleh karena itu alangkah baiknya jika kita senantiasa menjaga diri dari hal-hal yang membahayakan diri sendiri. Hal ini sesuai dengan salah satu kaidah fikih terkait kemudaratan:

Artinya: "Kemudaratan (ker<mark>ug</mark>ian, kesulitan, atau bahaya) harus dihilangkan"

Kemudaratan yang dimaksud yaitu salah satunya hal-hal yang berbahaya seperti penyakit. Sehingga penting bagi kita untuk menjaga diri dan kesehatan agar terhindar dari suatu penyakit(Shodiq, 2020).

Menjaga diri dapat dilakukan dengan mencegah datangnya suatu penyakit yang menimpa tubuh terutama penyakit berbahaya seperti kanker. Salah satu cara mendeteksi kanker yaitu dengan melihat susunan RNA pada manusia. Hal tersebut dapat dilakukan menggunakan bantuan metode kecerdasan buatan. Adanya deteksi kanker tersebut dapat membantu kita untuk menjaga diri sehingga. jika terdeteksi maka dapat segera diobati. Dengan demikian kita dapat senantiasa manjaga diri dan kesehatan agar tidak ada yang membahayakan diri sendiri. Hal tersebut penting dilakukan karena kesehatan itu sendiri pada *maqashid syari'ah* yang disampaikan oleh Al-Ghazali termasuk ke dalam *Hifdz An-Nafs* (memelihara jiwa) (Akbar and Budiyanto, 2020).

### **BAB III**

## METODE PENELITIAN

### 3.1. Jenis Penelitian

Penelitian ini mengklasifikasikan penyakit kanker menggunakan metode ERNN berdasarkan data RNA. Penelitian yang dilakukan ini termasuk pada jenis penelitian kuantitatif karena jenis dari data yang digunakan pada penelitian ini bentuk angka atau numerik serta terdapat proses perhitungan dan analisis pada hasil yang didapatkan. Hasil pada penelitian ini yaitu sistem klasifikasi penyakit kanker menggunakan metode ERNN berdasarkan data RNA yang nantinya diharapkan bisa berguna untuk membantu para ahli di bidang medis untuk mengklasifikasikan penyakit kanker dengan cepat dan akurasi yang dapat diperhitungkan.

## 3.2. Jenis dan Sumber Data

Data sekunder yang digunakan pada penelitian ini berisikan data vektor yang isinya merupakan susunan gen dari RNA pengambilannya menggunakan *Illumina HiSeq Platform*. Data tersebut diakses dari halaman website dari Mendeley Data (Ferles et al., 2018). Jumlah data sebanyak 1846 data terdiri dari 971 kolom yang isinya RPKM (*Read per Kilobase Million* dari setiap gen) dengan 4 kelas yaitu BRCA atau *breast adenocarcinoma* (kanker payudara) berjumlah 878 data, KIRC atau *kidney renal clear cell carcinoma* (kanker ginjal) berjumlah 537 data, UCEC atau *uterine corpus endometrial carcinoma* (kanker rahim) berjumlah 269 data, dan LUAD atau *lung adenocarcinoma* (kanker paru-paru) berjumlah 162

data. Data tersebut akan dibagi menjadi data latih dan data uji untuk proses melakukan klasifikasi. Sampel data RNA ditunjukkan pada Tabel 3.1

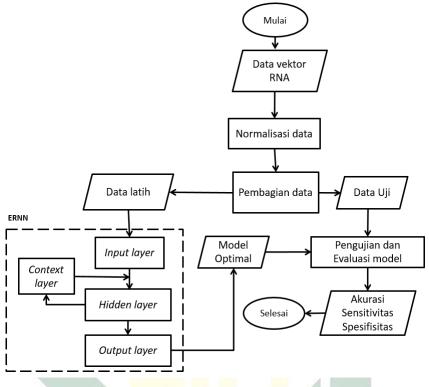
Tabel 3.1 Sampel data RNA

	1	2	3		971	Kelas
	1	<u> </u>	3		9/1	Keias
1	44.0235	9.2163	11.3191		8.7688	BRCA
2	29.7462	9.7656	40.5401		5.7585	BRCA
3	35.7993	9.8848	3.8860		5.3062	BRCA
4	26.4904	7.0858	10.8040		7.5005	BRCA
1	34.3220	16.5141	4.4705		5.1566	KIRC
2	42.4248	3.2142	29.0936		4.0475	KIRC
3	33.9790	15. <mark>192</mark> 0	13.3896		4.2263	KIRC
4	20.0473	9 <mark>.32</mark> 91	6.3995		5.2631	KIRC
1	47.5236	3 <mark>7.5</mark> 110	6.7251		2.4951	LUAD
2	64.0034	1 <mark>0.7555</mark>	18.7964		8.4382	LUAD
3	38.3356	6.8341	5.7191		5.4018	LUAD
4	23.7498	8.3384	7.2932	<b></b>	4.8625	LUAD
1	13.6022	8.8507	10.3864	,/.	1.6655	UCEC
2	41.3168	10.8444	18.3367		7.0020	UCEC
3	34.5631	29.8378	33.7547		3.0592	UCEC
4	38.0035	9.4083	7.1472		5.7118	UCEC

# 3.3. Kerangka Penelitian

Kerangka penelitian ini ditunjukkan pada diagram alir yang dapat dilihat pada Gambar 3.1. Pada diagram tersebut menunjukkan bahwa tahapan dari penelitian ini terdiri dari *preprocessing* menggunakan normalisasi, pelatihan serta

pengujian model dengan metode ERNN, dan evaluasi model menggunakan textitconfusion matrix.



Gambar 3.1 Diagram alir

## 1. Preprocessing

Pada tahap ini data vektor RNA yang berukuran 971x1 diubah menjadi skala 0 hingga 1 menggunakan metode normalisasi minmax scaler. Berikutnya data akan dibagi menjadi data latih atau data training dan data uji atau data testing dengan perbandingan banyak data 90:10, 80:20, dan 70:30.

#### 2. Pelatihan model

Pada tahap ini dilakukan pelatihan model menggunakan algoritma ERNN. Sebelumnya akan dilakukan inisialisasi nilai *learning rate*, banyak *epoch*, dan toleransi *error*. Data yang sudah dibagi akan masuk ke *input layer*, *hidden layer*, *context layer*, *dan output layer*. Pada tiap *layer* akan dihitung bobotnya

dan selalu diperbarui setiap iterasi untyuk menemukan bobot terbaik. Bobot tersebut yang nantinya digunakan untuk mengklasifikasikan di layer akhir.

# 3. Pengujian model

Pada tahap ini dilakukan uji coba pembagian data, *learning rate*, jumlah *hidden layer* pertama, dan jumlah *hidden layer* kedua. Nilai *hidden layer* yang akan digunakan yaitu 20 dan 50, 50 dan 100, serta 80 dan 150. Nilai *learning rate* yang akan digunakan yaitu 0.1, 0.2, 0.3, 0.4, dan 0.5. Uji coba dilakukan dengan berbagai parameter agar dapat mengetahui model terbaik dan paling optimal.

#### 4. Evauasi model

Pada tahap ini akan dilakukan evaluasi terhadap model yang sudah dibuat.

Perhitungan menggunakan confusion matrix dengan menghitung sensitivitas, spesifisitas, dan akurasi.

Tabel 3.2 Tabel evaluasi

Hidden	Hidden	Learning	Akurasi	Sensitivitas	Spesifisitas	Time
layer I	layer II	rate	(%)	(%)	(%)	
20	50	0,1				
50	100	0,2				
80	150	0,3				
		0,4				
		0,5				

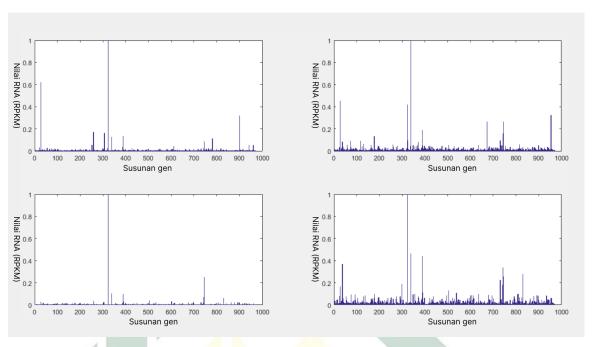
## **BAB IV**

### HASIL DAN PEMBAHASAN

## 4.1. Preprocessing

Pada penelitian ini menggunakan data vektor berisi susunan gen dari RNA yang diambil menggunakan Illumina HiSeq Platform. Satuan dari kecepatan RNA ketika membaca gen terhadap DNA disebut Reads Per Kilobase Million (RPKM). Satu data terdiri dari 971 kolom yang berisi tentang RPKM (Read per Kilobase Million dari setiap gen). Perubahan pada ekspresi gen terhadap RNA dijadikan indikasi jika telah ada kanker dalam tubuh manusia tersebut. Perbedaan tinggi rendahnya nilai suatu ekspresi gen setiap penderita dengan jenis kanker yang sama dipengaruhi oleh faktor yang mempengaruhi penderita tersebut. Hal itu terjadi sebagaimana suatu penyakit kanker dapat terjadi karena berbagai faktor. Berikut visualisasi data setiap kelasnya.

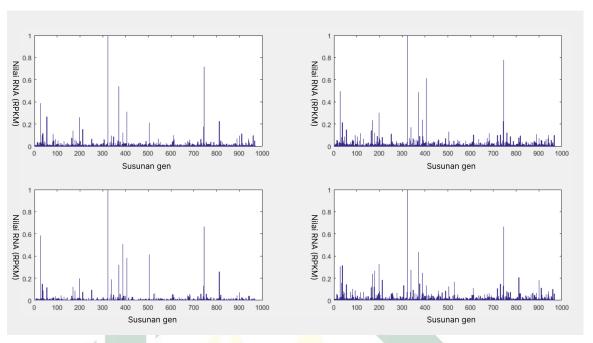
## 1. BRCA (breast adenocarcinoma)



Gambar 4.1 Visualisasi data kanker payudara

Susunan gen pada penderita payudara memiliki nilai maksimum pada kisaran gen dengan urutan 300 – 400. Beberapa gen dengan nilai yang lebih tinggi dari area lain juga terdapat pada kisaran tersebut. Beberapa penderita memiliki nilai gen yang cukup tinggi di awal susunan gen serta pada susunan kisaran 700 hingga 800an. Ekspresi gen pada urutan yang tidak disebutkan memiliki nilai yang tidak jauh beda dan cukup stabil.

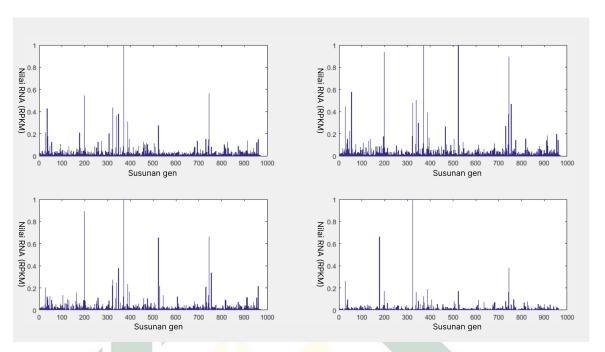
# 2. KIRC (kidney renal clear cell carcinoma)



Gambar 4.2 Visualisasi data kanker ginjal

Penderita kanker ginjal memiliki nilai ekspresi gen tertinggi pada kisaran urutan gen 300 – 400 dan nilai yang sangat tinggi pada 700 – 800. Pada urutan awal penderita penyakit ini juga memiliki nilai ekspresi gen yang cukup tinggi. Pada kisaran 200, 500, dan 900 terdapat beberapa ekspresi gen yang lebih menonjol dari gen lainnya yang rendah dan cukup stabil.

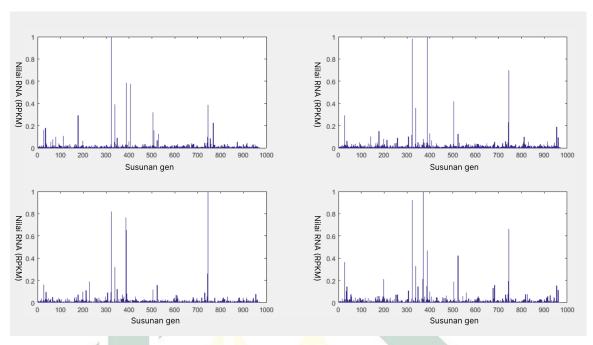
# 3. LUAD (lung adenocarcinoma)



Gambar 4.3 Visualisasi data kanker paru-paru

Penderita penyakit kanker paru-paru memiliki 3 hingga 5 nilai ekspresi gen yang sangat tinggi. Mayoritas terdapat pada urutan 300 – 400. Beberapa nilai tertinggi lainnya terdapat pada urutan 200an serta700 – 800. Ekspresi gen yang lainnya memiliki nilai yang kurang stabil, beberapa banyak yang menonjol. Beberapa lainnya juga memiliki nilai yang cukup tinggi.

# 4. UCEC (uterine corpus endometrial carcinoma)



Gambar 4.4 Visualisasi data kanker rahim

Penderita kanker rahim memiliki dua nilai ekspresi gen yang tertinggi pada urutan gen 300 – 400. Terdapat penderita yang juga memiliki nilai tertinggi pada uruta gen 700 – 800. Beberapa gen yang menonjol terdapat pada urutan awal serta beberapa lainnya juga terdapat pada urutan akhir. Pada urutan 500 an terdapat ekspresi gen yang cukup tinggi. Ekspresi gen yang lainnya cukup stabil meskipun ada beberapa yang naik turun.

Data setiap kelas cukup sulit jika hanya dibedakan dengan cara pengamatan melalui visualisasi data. Hal tersebut karena jumlah parameter yang sangat banyak sehingga diperlukan metode lain untuk klasifikasi kanker berdasarkan data RNA. Selain itu, data RNA memiliki rentan data yang besar sehingga perlu dilakukan *preprocessing* normalisasi data menggunakan Persamaan 2.1.

Tabel 4.1 Sampel data RNA sebelum normalisasi

$x_1$	$x_2$	$x_3$	$x_4$	 $x_{971}$
44.0235	9.2163	11.3191	33.2152	 8.7688

$$x_1 = \frac{44.0235 - 0}{3.7564 \times 10^3 - 0}$$
$$= 0.0117$$

Tabel 4.2 Sampel data RNA sesudah normalisasi

$x_1$	$x_2$	$x_3$	$x_4$	 $x_{971}$
0.0117	0.0025	0.0030	0.0088	 0.0023

Perhitungan normalisasi dilakukan pada seluruh data RNA. Seanjutnya data akan diklasifikasi menggunakan metode ERNN.

## 4.2. Pelatihan dan Pengujian Model

Pelatihan model dilakukan menggunakan metode ERNN dengan uji coba parameter pembagian data, *learning rate*, dan jumlah node pada *hidden layer*. Data yang digunakan yaitu data RNA yang sudah dinormalisasi. Jumlah data sebanyak 1846, di mana terdiri dari BRCA atau *breast adenocarcinoma* (kanker payudara) berjumlah 878 data, KIRC atau *kidney renal clear cell carcinoma* (kanker ginjal) berjumlah 537 data, UCEC atau *uterine corpus endometrial carcinoma* (kanker rahim) berjumlah 269 data, dan LUAD atau *lung adenocarcinoma* (kanker paru-paru) berjumlah 162 data. Data tersebut akan dibagi menjadi data latuh dan

data uji untuk proses melakukan klasifikasi. Berikut contoh perhitungan ERNN dengan jumlah node *hidden layer* I 20 dan node *hidden layer* II 50.

Langkah awal dari metode ERNN yaitu melakukan inisialisasi, seperti learning rate, maksimal epoch, toleransi error, bobot input layer menuju hidden layer I, bobot hidden layer I menuju hidden layer II, bobot h=hidden layer II menuju Output Layer, dan bias.

*Learning rate* = 0.1

Maksimal epoch = 1000

Toleransi error = 0.0001

Inisialisasi bobot dan bias input layer menuju hidden layer I

Tabel 4.3 Bobot dan bias input layer menuju hidden layer I

	$v_1$	$v_2$	$v_3$	$v_4$		$v_{20}$
b	0.2217	0.4954	0.0886	0.0302		0.7441
$x_1$	0.5523	0.3812	0.7953	0.0686	4	0.1603
$x_2$	0.7347	0.7477	0.4604	0.6134		0.6561
$x_3$	0.5783	0.8043	0.2444	0.4180	<i></i>	0.4967
$x_4$	0.2514	0.7883	0.0423	0.0916		0.4129
:	:	: `		:	٠٠.	÷
$x_{971}$	0.4104	0.9755	0.9616	0.6071		0.7780

Inisialisasi bobot dan bias hidden layer I menuju hidden layer II

Tabel 4.4 Bobot dan bias input layer menuju hidden layer I

	$v_1$	$v_2$	$v_3$	$v_4$		$v_{50}$
b	0.6825	0.0417	0.5568	0.5111		0.8608
1	0.5111	0.4510	0.0075	0.8151		0.2829
2	0.7904	0.4651	0.1651	0.4875		0.4795
3	0.0648	0.6645	0.1033	0.1897		0.3148
4	0.9911	0.3331	0.2036	0.4030		0.8794
:	://	<b>/</b> : , ,	<b>/</b> :	÷	٠	÷
20	0.0071	0.0052	0.0014	0.3796		0.4932

Inisialisasi bobot dan bias hidden layer II menuju output layer

Tabel 4.5 Bobot dan bias hidden layer II menuju output layer

	w
b	0.4421
1	0.8194
2	0.3966
3	0.5705
4	0.0041
:	:
50	0.4300

Selanjutnya yaitu menghitung proses ERNN epoch ke-1 menggunakan data ke-1. Data yang digunakan ditunjukkan pada Tabel 4.6.

Tabel 4.6 Data ke-1 proses training

$x_1$	$x_2$	$x_3$	$x_4$	 $x_{971}$
0.0117	0.0025	0.0030	0.0088	 0.0023

Menghitung nilai net dari hidden layer I menggunakan Persamaan 2.4.

$$y_{h1} = \sum_{i}^{n} x_{i}(t)v_{ij}$$

$$= (x_{1}v_{11}) + (x_{2}v_{12}) + (x_{3}v_{13}) + \dots + (x_{971}v_{1971})$$

$$= (0.0117 \times 0.5523) + (0.0025 \times 0.7347) + (0.0030 \times 0.5783)$$

$$+ \dots + (0.0023 \times 0.4104)$$

$$= 0.0065 + 0.0018 + 0.0017 + \dots + (9.5802 \times 10^{-4})$$

$$= 4.1092$$

Nilai dari  $y_h$  ini di copy dari hidden layer menuju ke context layer menjadi  $y_h(t-1)$  karena pada perhitungan pertama context layer masih belum memiliki nilai dari hidden layer pada perhitungan sebelumnya. Selanjutnya nilai  $y_h(t-1)$  dihitung dengan bobot awal dari context layer menuju hidden layer menggunakan nilai bobot v awal dari input layer ke hidden layer, sehingga pada perhitungan pertama ini ukurannya menyesuaikan bobot yang digunakan dan menggunakan  $y_h$  node ke-1 pada setiap nilai  $y_h(t-1)$ . Hal tersebut juga dilakukan pada node lain pada perhitungan awal. Pada perhitungan selanjutnya context layer sudah memiliki nilai yang di copy dari hidden layer sehingga ukuran  $y_h(t-1)$  sama dengan banyak node yang digunakan pada hidden layer dengan ukuran bobot yang menyesuaikan banyak node nya. Pada perhitungan pertama didapatkan hasil net

sebagai berikut.

$$net_{j}(t) = y_{hj} + \sum_{i}^{n} v_{ij}y_{h}(t-1) + \theta_{j}$$

$$net_{1}(t) = y_{h1} + \sum_{i}^{n} v_{ij}y_{h}(t-1) + \theta_{1}$$

$$= 4.1092 + ((0.5523 \times 4.1092) + (0.7347 \times 4.1092) + (0.5783 \times 4.1092)$$

$$+ \dots + (0.4104 \times 4.1092)) + 0.2217$$

$$= 4.1092 + (1.9916 \times 10^{3}) + 0.2217$$

$$= 1.9959 \times 10^{3}$$

Hitung dengan cara serupa terhadap setiap node pada *hidden layer* I. Hasil dari net setiap node ditunjukkan pada Tabel

Tabel 4.7 Hasil net input layer menuju hidden layer

net	Hasil
$net_1$	$1.9959 \times 10^3$
$net_2$	$1.9835 \times 10^3$
$net_3$	$2.3151 \times 10^3$
$net_4$	$2.4270 \times 10^3$
$net_5$	$2.1457 \times 10^3$
<b>:</b>	:
$net_{20}$	$1.9135 \times 10^3$

Hasil dari nilai net dihitung menggunakan fungsi aktivasi seperti pada

Persamaan 2.5 dan akan menjadi keluaran hidden layer I.

$$f(netj(t)) = \frac{1}{1 + \epsilon^{-net_j}}$$

$$= \frac{1}{1 + \epsilon^{-net_1}}$$

$$= \frac{1}{1 + \epsilon^{-1.9959 \times 10^3}}$$

$$= 1$$

Pada perhitungan saat ini keluaran setiap node pada hidden layer I atau  $(y_j)$  bernilai 1. Selanjutnya akan dihitung keluaran hidden layer II dengan cara yang sama. Menghitung nilai keluaran dari hidden layer II.

$$y_{h1}2 = \sum_{i}^{n} y_{j}(t)v1_{ij}$$

$$= (y_{1}v1_{11}) + (y_{2}v1_{12}) + (y_{3}v1_{13}) + \dots + (y_{20}v1_{120})$$

$$= (1 \times 0.5111) + (1 \times 0.7904) + (1 \times 0.0648) + \dots + (1 \times 0.0071)$$

$$= 0.5111 + 0.7904 + 0.0648 + \dots + 0.0071$$

$$= 7.8719$$

Nilai dari  $y_{h1}$ 2 ini selanjutnya di copy dari  $hidden\ layer\ menuju\ ke\ context\ layer$ 

menjadi  $y_h(t-1)$ , sehingga didapatkan hasil net hidden layer II sebagai berikut.

$$net2_{j}(t) = y_{hj} + \sum_{i}^{n} v_{ij}y_{h2}(t-1) + \theta_{j}$$

$$net2_{1}(t) = y_{h1}2 + \sum_{i}^{n} v_{ij}y_{h}(t-1) + \theta_{1}$$

$$= 7.8719 + ((0.5111 \times 7.8719) + (0.7904 \times 7.8719) + (0.0648 \times 7.8719)$$

$$+ \dots + (0.0071 \times 7.8719)) + 0.6825$$

$$= 7.8719 + 61.9663 + 0.6825$$

$$= 70.5207$$

Hasil dari nilai net dihitung menggunakan fungsi aktivasi dan akan menjadi keluaran *hidden layer* II.

$$f(net2j(t)) = \frac{1}{1 + \epsilon^{-net_j}}$$

$$= \frac{1}{1 + \epsilon^{-net_1}}$$

$$= \frac{1}{1 + \epsilon^{-70.5207}}$$

Pada perhitungan saat ini keluaran setiap node pada *hidden layer* II bernilai 1. Selanjutnya akan dihitung keluaran *output layer* dengan cara yang sama.

Menghitung nilai keluaran dari *output layer* menggunakan Persamaan 2.7.

$$net_k(t) = \sum_{j}^{m} w_{jk} y_j(t) + \theta_j$$

$$= (w_{11} f(net_1)) + (w_{21} f(net_2)) + (w_{31} f(net_3)) + \dots +$$

$$(w_{501} f(net_5 0)) + \theta_1$$

$$= (0.8194 \times 1) + (0.3966 \times 1) + (0.5705 \times 1) + \dots + 0.4300 \times 1)$$

$$+0.4421$$

$$= 26.6079$$

Hasil  $net_k(t)$  akan diaktifkan menggunakan fungsi aktivasi dan menjadi keluaran dari  $output\ layer$  menggunakan Persamaan 2.6.

$$f(net_k(t)) = \frac{1}{1 + \epsilon^{-net_k}}$$
$$= \frac{1}{1 + \epsilon^{-26.6079}}$$
$$= 1$$

Selanjutnya akan dihitung nilai *error* dari setiap *layer* untuk perbaikan bobot menggunakan persamaan 2.8.

$$\delta_k = (d_k - y_k)g'(net_k(t))$$

$$= (d_k - y_k)(y_k(t)(net_k(t) - y_k(t)))$$

$$= 7.1241 \times 10^{-11}$$

Menghitung koreksi bobot w dengan menggunakan nilai learning rate seperti pada

Persamaan 2.10.

$$\Delta w_{jk} = \eta \delta_k f(net 2_j)$$
  
=  $0.1 \times 7.1241 \times 10^{-11} \times 1$   
=  $7.1241 \times 10^{-12}$ 

Hasil dari  $f(net2_j)$  sama sehingga nilai koreksi bobot w juga bernilai sama. Selanjutnya hitung koreksi bobot v dengan cara yang sama dan hitung koreksi bias menggunakan Persamaan 2.12.

$$\Delta \theta_k = \eta \delta_k$$

$$= 0.1 \times 7.1241 \times 10^{-11}$$

$$= 7.1241 \times 10^{-12}$$

Hitung koreksi bobot output layer dan hidden layer II

$$\delta net_1 = \sum_{k=1}^{m} \delta_k \times w_{kj}$$

$$= 3.5621 \times 10^{-9} \times 0.8194$$

$$= 2.9188 \times 10^{-9}$$

Hasil lengkapnya ditunjukkan pada Tabel 4.9

Tabel 4.8 Hasil koreksi net j

net	Hasil
$net_1$	$2.9188 \times 10^{-9}$
$net_2$	$1.4129 \times 10^{-9}$
$net_3$	$2.0329 \times 10^{-9}$
$net_4$	$1.4481 \times 10^{-11}$
$net_5$	$3.0086 \times 10^{-9}$
1:,/	:
$net_{50}$	$1.5317 \times 10^{-9}$

Selanjutnya hitung kesalahan pada fungsi sigmoid dengan menggunakan

# Persamaan 2.9

$$\delta_j = \delta net_1 f'(net_j)$$

$$= 3.5621 \times 10^{-9} \times 1$$

$$= 3.5621 \times 10^{-9}$$

Hasil lengkapnya ditunjukkan pada Tabel 4.9

Tabel 4.9 Hasil  $\delta_j$ 

$\delta_j$	Hasil
1	$2.9188 \times 10^{-9}$
2	$1.4129 \times 10^{-9}$
3	$2.0329 \times 10^{-9}$
4	$1.4481 \times 10^{-11}$
5	$3.0086 \times 10^{-9}$
ij	:
50	$1.5317 \times 10^{-9}$

Setelah di dapatkan hasil tersebut, hitung koreksi bobot v, yaitu bobot antara hidden layer I dan hidden layer menggunakan learning rate.

$$\Delta v_{11} = 0.1 \times 2.9188 \times 10^{-9}$$

$$= 2.9188 \times 10^{-10}$$

Sehingga didapatkan koreksi bobot yang ditunjukkan pada Tabel 4.10

Tabel 4.10 Perbaikan bobot v

	$v_1$	$v_2$	$v_3$	$v_4$		$v_{50}$
1	$2.9188 \times 10^{-10}$	$1.4129 \times 10^{-10}$	$2.0321 \times 10^{-10}$	$1.4481 \times 10^{-12}$		$1.5317 \times 10^{-10}$
2	$2.9188 \times 10^{-10}$	$1.4129 \times 10^{-10}$	$2.0321 \times 10^{-10}$	$1.4481 \times 10^{-12}$		$1.5317 \times 10^{-10}$
3	$2.9188 \times 10^{-10}$	$1.4129 \times 10^{-10}$	$2.0321 \times 10^{-10}$	$1.4481 \times 10^{-12}$		$1.5317 \times 10^{-10}$
4	$2.9188 \times 10^{-10}$	$1.4129 \times 10^{-10}$	$2.0321 \times 10^{-10}$	$1.4481 \times 10^{-12}$		$1.5317 \times 10^{-10}$
:	<b>:</b>	÷	i i	÷	٠	:
20	$2.9188 \times 10^{-10}$	$1.4129 \times 10^{-10}$	$2.0321 \times 10^{-10}$	$1.4481 \times 10^{-12}$		$1.5317 \times 10^{-10}$

Hitung koreksi bobot v antara *input layer* dan *hidden layer* I dengan cara yang sama. Selanjutnya hitung koreksi bias.

$$\Delta\theta_1 = \eta\delta_1$$

$$= 0.1 \times 2.9188 \times 10^{-9}$$

$$= 2.9188 \times 10^{-10}$$

Nilai koreksi bias ditunjukkan pada Tabel

Tabel 4.11 Koreksi bias hidden layer I menuju hidden layer II

Bias	Hasil
1	$2.9188 \times 10^{-10}$
2	$1.4129 \times 10^{-10}$
3	$2.0321 \times 10^{-10}$
4	$1.4481 \times 10^{-12}$
:	: /
50	$1.5317 \times 10^{-12}$

Langkah akhir yaitu perbarui bobot dengan cara manjumlahkan bobot lama dengan koreksi bobot.

$$w_{11}baru = w_{11}lama + \Delta w_{11}$$
  
=  $0.8194 + 7.1241 \times 10^{-12}$   
=  $0.8194$ 

Hasil pembaruan bobot ditunjukkan pada Tabel

Tabel 4.12 Perbaikan bias hidden layer I menuju hidden layer II

	w baru
1	0.8194
2	0.3966
3	0.5705
4	0.0041
:	
50	0.4300

Hasil pembaruan bobot v, yaitu bobot antara *hidden layer* I dan *hidden layer* I ditunjukkan pada Tabel 4.13

Tabel 4.13 Perbaikan bobot v

	$v_1$	$v_2$	$v_3$	$v_4$		$v_{50}$
1	1.0222	0.4431	0.0635	1.5969	4.	0.3393
2	1.5808	0.9021	0.0149	1.6301		0.5659
3	0.1297	0.9302	0.3301	0.9751		0.4795
4	1.9822	1.3290	0.2065	0.3793		0.9589
:	:	: \		:	٠٠.	:
20	0.0142	0.0104	0.0027	0.7592		0.9864

Nilai bobot akan terus diperbarui hingga mendapatkan bobot yang optimal atau sudah mencapai maksimum epoch. Proses perhitungan ERNN ini akan diaplikasikan untuk klasifikasi kanker berdasarkan RNA dengan sebanyak 1846 data. Setelah dilakukan pelatihan, akan dilakukan tahap pengujian. ERNN diuji coba dengan parameter *learning rate*, pembagian data, dan jumlah node pada *hidden layer*.

#### 4.3. Evaluasi Model

Pada penelitian ini melakukan evaluasi model dengan confusion matrix. Hasil dari confusion matrix berupa akurasi, sensitivitas, dan spesifisitas yang di hitung menggunakan Persamaan 2.14, 2.15, dan 2.16. Nilai tersebut mempresentasikan ukuran kinerja algoritma ERNN dalam melakukan klasifikasi kanker berdasarkan data RNA. Hasil klasifikasi pada pembagian data 90%:10% ditunjukkan pada Tabel 4.14

Tabel 4.14 Hasil klasifikasi pada pembagian data 90%:10%

Hidden	Hidden	Learning	Akurasi	Sensitivitas	Spesifisitas	Time
layer I	layer II	rate	<u>(%)</u>	(%)	(%)	
20	50	0,1	9 <mark>7.30</mark>	94.63	98.76	00:00:18
		0,2	96. <mark>76</mark>	94.44	98.61	00:00:16
		0,3	97.30	93.75	98.90	00:00:16
		0,4	97.30	95.58	98.86	00:00:16
		0,5	97.84	95.84	99.05	00:00:17
50	100	0,1	98.38	97.92	99.28	00:00:34
		0,2	97.84	96.78	99.09	00:00:32
		0,3	96.22	93.24	98.42	00:00:32
		0,4	96.22	90.53	98.52	00:00:32
		0,5	97.84	95.84	99.05	00:00:33
80	150	0,1	97.30	94.70	98.80	00:00:59
		0,2	97.84	94.89	99.05	00:00:47
		0,3	97.30	93.75	98.76	00:00:48
		0,4	95.68	88.67	98.51	00:00:48
		0,5	96.22	92.36	98.57	00:00:47

Berdasarkan Tabel 4.14, Uji coba pertama dilakukan dengan pembagian data *training* 90% dan data *testing* 10%. Hasil terbaik pada uji coba ini didapatkan pada kombinasi 50 node *hidden layer* I, 100 node *hidden layer* II, dan *learning rate* 0.1. Pada model tersebut didapatkan akurasi mencapai 98.38%, sensitivitas sebesar 97.92% dan Spesifisitas sebesar 99.28%. Waktu yang dibutuhkan untuk model tersebut yaitu 34 detik.

Tabel 4.15 Hasil klasifikasi pada pembagian data 80%:20%

Hidden	Hidden	Learning	Akurasi	Sensitivitas	Spesifisitas	Time
layer I	layer II	rate	(%)	(%)	(%)	
20	50	0,1	99.19	99.03	99.72	00:00:19
	4	0,2	9 <mark>9.</mark> 19	97 <mark>.</mark> 97	99.73	00:00:18
		0,3	98 <mark>.10</mark>	95.80	99.38	00:00:20
		0,4	<mark>98.64</mark>	98. <mark>21</mark>	99.55	00:00:22
		0,5	98.92	98.33	99.63	00:00:18
50	100	0,1	98.64	98.21	99.55	00:00:37
		0,2	97.29	94.56	99.09	00:00:41
		0,3	98.92	98.79	99.64	00:00:38
		0,4	98.92	97.83	99.61	00:00:39
		0,5	97.83	96.13	99.29	00:00:37
80	150	0,1	97.83	96.97	99.29	00:00:59
		0,2	98.37	97.51	99.45	00:00:59
		0,3	98.92	98.36	99.64	00:00:57
		0,4	97.83	96.16	99.30	00:00:57
		0,5	98.92	98.08	99.67	00:00:54

Berdasarkan Tabel 4.15, Uji coba kedua di lakukan dengan pembagian data

training 80% dan data testing 20 %. Hasil terbaik pada uji coba ini didapatkan pada kombinasi 20 node hidden layer I, 50 node hidden layer II, dan learning rate 0.1. Pada model tersebut didapatkan akurasi mencapai 99.19 %, sensitivitas sebesar 99.03 % dan Spesifisitas sebesar 99.72 %. Waktu yang dibutuhkan untuk model tersebut yaitu 19 detik.

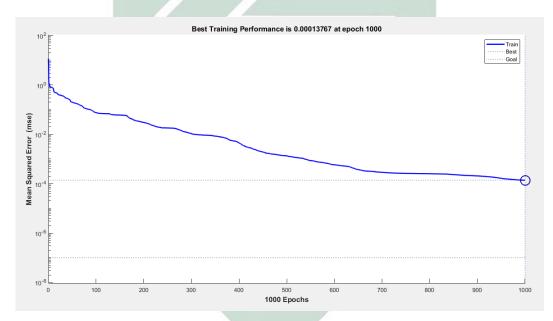
Tabel 4.16 Hasil klasifikasi pada pembagian data 70%:30%

Hidden	Hidden	Learning	Akurasi	Sensitivitas	Spesifisitas	Time
layer I	layer II	rate	(%)	(%)	(%)	
20	50	0,1	97.65	94.69	99.21	00:00:16
		0,2	97.47	94.66	99.16	00:00:18
	4	0,3	9 <mark>7.</mark> 83	95 <mark>.3</mark> 9	99.20	00:00:16
		0,4	96. <mark>93</mark>	93.00	99.99	00:00:16
		0,5	97.29	95. <mark>61</mark>	99.07	00:00:19
50	100	0,1	97.83	95. <mark>96</mark>	99.29	00:00:30
		0,2	96.57	92.23	98.89	00:00:32
		0,3	96.75	93.83	98.98	00:00:42
		0,4	95.67	90.87	98.57	00:00:30
		0,5	97.65	97.12	99.24	00:00:31
80	150	0,1	96.21	92.74	98.75	00:00:54
		0,2	96.93	94.53	99.00	00:00:53
		0,3	94.36	94.36	99.07	00:00:48
		0,4	95.85	93.31	98.65	00:00:44
		0,5	95.31	90.22	98.49	00:00:44

Berdasarkan Tabel 4.16, Uji coba ketiga dilakukan dengan pembagian data *training* 70% dan data *testing* 30 %. Hasil terbaik pada uji coba ini didapatkan

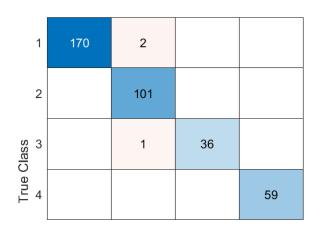
pada kombinasi 50 node *hidden layer* I, 100 node *hidden layer* II, dan *learning rate* 0.1. Pada model tersebut didapatkan akurasi mencapai 97.83 %, sensitivitas sebesar 95.96 % dan Spesifisitas sebesar 99.29 %. Waktu yang dibutuhkan untuk model tersebut yaitu 16 detik. Bedasarkan keseluruhan uji coba, hasil terbaik didapatkan pada uji coba kedua pada kombinasi 20 node *hidden layer* I, 50 node *hidden layer* II, dan *learning rate* 0.1. Hal itu menunjukkan model tersebut dapat mempelajari pola data dengan baik. Berikut *performance* dari model terbaik di tunjukkan pada

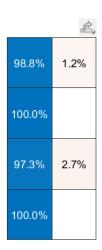
#### Gambar 4.5



Gambar 4.5 Performance model terbaik

Model tersebut menghasilkan akurasi, sensitivitas dan spesifisitas yang tinggi dalam waktu 19 detik. Hasil *Confusion matrix* model terbaik ditunjukkan pada 4.6.





100.0%	97.1%	100.0%	100.0%
	2.9%		
1	2	3	4
		<b>Predicted Class</b>	

Gambar 4.6 Hasil confusion matrix model terbaik

# Ket:

1 = Kanker payudara

2 = Kanker ginjal

3 = Kanker paru-paru

4 = Kanker rahim

$$Spesifisitas = \frac{\sum \frac{TN}{TN+FP}}{n} x 100\%$$

$$= \frac{\frac{197}{197+0} + \frac{265}{265+3} + \frac{332}{332+0} + \frac{310}{310+0}}{4} x 100\%$$

$$= 99.72\%$$

$$Sensitivitas = \frac{\sum \frac{TP}{TP+FN}}{n} x 100\%$$

$$= \frac{\frac{170}{170+2} + \frac{101}{101+0} + \frac{36}{36+1} + \frac{59}{59+0}}{4} x 100\%$$

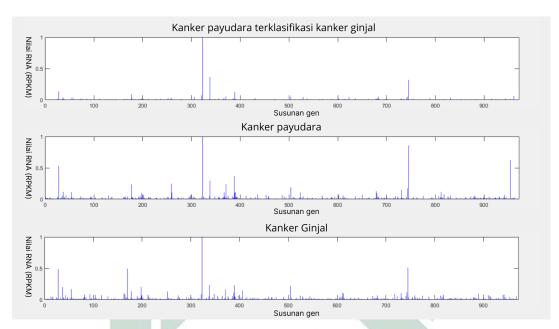
$$= 99.03\%$$

$$Akurasi = \frac{TP_{all}}{n_{all}} x 100\%$$

$$= \frac{170 + 101 + 36 + 59}{369} x 100\%$$

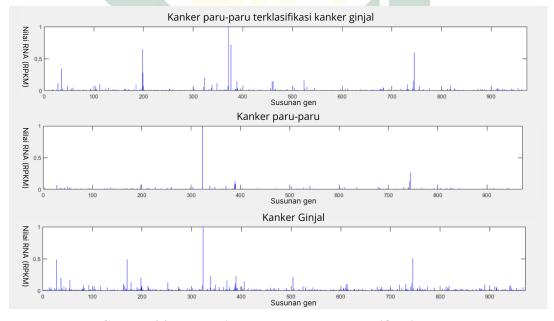
$$= 99.19\%$$

Berdasarkan hasil spesifisitas, sensitifitas, dan akurasi tersebut menunjukkan metode ERNN dapat mengklasifikasikan kanker dengan baik menggunakan data RNA. Jika di lihat hasil *confusion matrix* hanya terdapat tiga data yang salah kasifikasi yaitu dua data kanker payudara dan satu data kanker paru-paru. Ketiga data tersebut terklasifikasi ke dalam kelas kanker ginjal. Hal tersebut menunjukkan bahwa Hasil klasifikasi cenderung ke kelas kanker ginjal. Berikut visualisasi data RNA kelas kanker payudara yang terklasifikasi kanker ginjal ditunjukkan pada Gambar 4.7.



Gambar 4.7 Data RNA kanker payudara terklasifikasi salah

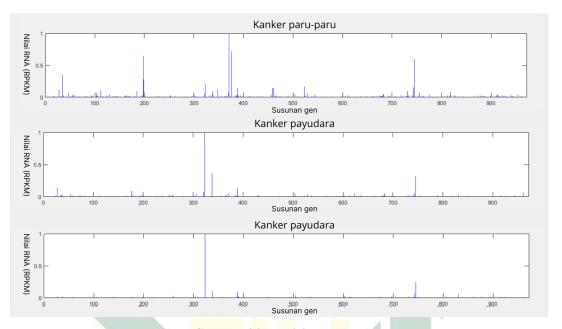
Sedangkan visualisasi data RNA kelas kanker paru-paru yang terklasifikasi kanker ginjal ditunjukkan pada Gambar 4.7.



Gambar 4.8 Data RNA kanker paru-paru terklasifikasi salah

Kesalahan dari klasifikasi itu dapat terjadi karena data kanker payudara dan

kanker paru-paru yang terklasifikasi salah tersebut memiliki kemiripan atau pola data RNA yang hampir sama dengan kanker ginjal. Berikut contoh data kanker payudara yang terklasifikasi kanker ginjal karena pola nilai RNA nya yang memiliki kemiripan ditunjukkan pada Gambar 4.9



## Gambar 4.9 Kemiripan data

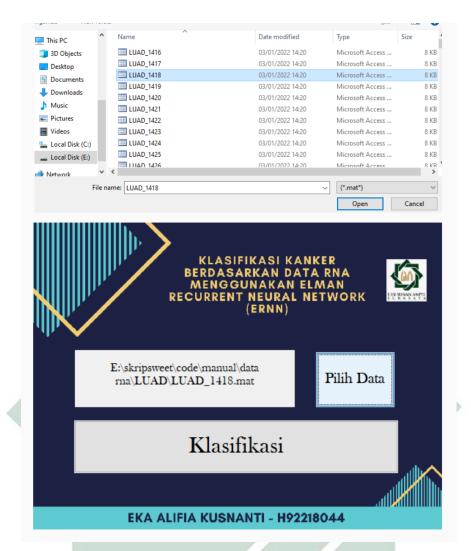
## 4.4. Aplikasi Klasifikasi Kanker

Model dari hasil uji coba terbaik digunakan untuk membuat aplikasi klasifikasi kanker. Aplikasi ini diharapkan dapat membantu ahli medis dalam melakukan klasifikasi. Penggunaan aplikasinya yaitu dengan cara memasukkan data RNA dan akan keluar hasil klasifikasi kanker berdasarkan data tersebut. Hasil klasifikasi jenis kanker yang dikeluarkan yaitu kanker payudara, kanker ginjal, kanker paru-paru, dan kanker rahim. Berikut Halaman utama dari aplikasi ditunjukkan pada Gambar 4.10.



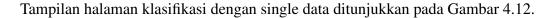
Gam<mark>ba</mark>r 4.10 Halaman utama aplikasi

Pada halaman utama pengguna diminta untuk memasukkan data yang ingin diklasifikasi. Klik "Pilih Data" untuk mencari data dan akan muncul kotak pemilihan file data RNA. Pada aplikasi ini, pengguna dapat menggunakan single data maupun multiple data. Pilih data dan klik "Open" untuk memasukkan ke Selanjutnya akan muncul kembali halaman utama dengan tampilan aplikasi. tulisan letak file yang sudak kita pilih, contohnya seperti "E:/skripsweet/code/manual/data rna/LUAD/LUAD\_1418.mat". Pastikan lokasi dan nama data sudah sesuai dengan nama file yang sudah di pilih. Jika sudah benar maka klik "Klasifikasi" untuk melakukan klasifikasi. Berikut tampilan halaman pemilihan file dan halaman utama yang sudah di masukkan data ditunjukkan pada Gambar 4.11



Gambar 4.11 Halaman pemilihan file dan halaman utama dengan data yang sudah di masukkan

Setelah di klik "Klaifiksi" akan ditujukan ke halaman klasifikasi. Pada halaman tersebut ditampilkan tabel data RNA berukuran 1 × 971 sebanyak data yang dimasukkan dan terdapat kotak hasil klasifikasi seperti "Kanker Paru-paru". Jika pengguna memasukkan data multiple, maka pengguna harus memasukkan urutan data yang ingin diketahui hasil klasifikasi nya dan klik "CEK". Selain itu juga terdapat tombol "Quit" dan "BACK". Klik "BACK" jika ingin memasukkan data baru dan melakukan klasifikasi lagi. Klik "Quit" jika ingin menutup aplikasi.





Gamba<mark>r 4.12 Halaman kla</mark>sifika<mark>si s</mark>ingle data

Sedangkan Tampilan halaman klasifikasi dengan multiple data ditunjukkan pada Gambar 4.13.



Aplikasi klasifikasi kanker ini diharapkan dapat bermaanfaat sebagaimana mestinya terutama bagi ahli medis dan penderita penyakit kanker.

## 4.5. Integrasi Keislaman

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

langit dan bumi, perbedaan bahasamu dan warna kulitmu. Sungguh, pada yang demikian itu benar-benar terdapat tanda-tanda bagi orang-orang yang mengetahui".

Maknah dari ayat al-Qur'an tersebut adalah Allah Swt., menciptakan manusia dengan ciri yang berbeda. Sebagai contoh, meskipun satu keluarga memiliki satu garis keturunan yang sama pasti terdapat perbedaan di antara mereka. Hal itu sam seperti DNA setiap manusia pasti berbeda antara manusia sat dengan yang lainnya. Jika terjadi perubahan pada susunan suatu DNA, maka arti dan maknah dari DNA nya juga menjadi berubah. Begitu pun dengan RNA, jika usunan RNA maka maknanya juga berubah.

Adanya perubahan pada nilai suatu ekspresi gen terhadap RNA dijadikan indikator jika telah ada penyakit kanker dalam tubuh manusia. Hal tersebut digunakan dalam penelitian ini untuk mengklasifikasikan penyakit kanker. Seseorang yang berusaha sembuh dari penyakitnya merupakan suatu bentuk ikhtiar terhadap Allah Swt. Karena suatu penyakit yang datang dari Allah Swt, pasti memiliki maksud tertentu. Hal ini di jelaskan dalam salah satu hadis Nabi Muhamad Saw., Hadis tersebut diriwayatkan al-Bukhari no 5210 yang berbunyi: حَدَّ ثَنِي عَبْدُ اللَّهِ بُنُ مُحَمَّدٍ حَدَّ ثَنَا عَبْدُ الْمَلِكِ بُنُ عَمْرٍ و حَدَّ ثَنَا وُعَنْ أَبِي هُرَيْرَةً عَنْ عَطَاءِ بُنِ يَسَارٍ عَنْ أَبِي سَعِيدٍ الْخُدْرِيِ وَعَنْ أَبِي هُرَيْرَةً عَنْ النَّبِي صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ قَالَ مَا يُصِيبُ الْمُسْلِمَ مِنْ نَصَبٍ وَلَا وَصَبٍ وَلَا هَمٍ وَلَا هَمْ وَلَا أَذًى وَلَا غَمْ حَتَى الشَّوْكَةِ يُشَاكُهَا إِلَّا كَفَّرَ اللَّهُ بِهَا مِنْ خَطَايَاهُ حُزْنِ وَلَا أَذًى وَلَا غَمْ حَتَى الشَّوْكَةِ يُشَاكُهَا إِلَّا كَفَّرَ اللَّهُ بِهَا مِنْ خَطَايَاهُ

Artinya: "Telah menceritakan kepadaku Abdullah bin Muhammad telah menceritakan kepada kami Abdul Malik bin 'Amru telah menceritakan kepada

kami Zuhair bin Muhammad dari Muhammad bin 'Amru bin Halhalah dari 'Atha' bin Yasar dari Abu Sa'id Al Khudri dan dari Abu Hurairah dari Nabi shallallahu 'alaihi wasallam beliau bersabda: "Tidaklah seorang muslim tertimpa suatu penyakit dan keletihan, kehawatiran dan kesedihan, dan tidak juga gangguan dan kesusahan bahkan duri yang melukainya melainkan Allah akan menghapus kesalahan-kesalahannya".

Menjaga diri dari penyakit memang tidak mudah, seperti menjaga pola makan dan cara hidup yang baik dan benar. Tetapi hal tersebut harus tetap senantiasa dilakukan agar tubuh tetap sehat jasmani dan rohani karena mencegah diri dari penyakit lebih baik daripada mengobati penyakit tersebut. Hal ini sesuai dengan salah satu kaidah fikih sebagai berikut.

Artinya: "Menolak itu lebih kuat dari pada menghilangkan".

Mencegah memang lebih baik daripada mengobati. Sebagaimana hasil dari penelitian ini diharapkan dapat memudahkan ahli medis dalam melakukan klasifikasi kanker. Hal tersebut penting karena kanker sangat mudah menyebar dalam tubuh manusia dan merupakan penyakit berbahaya yang dapat menyebabkan kematian.

Suatu musibah seperti penyakit memang tidak ada yang tau kapan datangnya, akan tetapi kita harus senantiasa menjaga diri dari hal-hal berbahaya seperti mencegah dan mengobati jika terkena penyakit. Maka dari itu dengan adanya penelitian ini diharapkan para penderita kanker dapat diberi kemudahan untuk mengetahui penyakit kanker yang di deritanya.

Hasil dalam penelitian ini dapat membantu penderita penyakit kanker

payudara, kanker ginjal, kanker paru-paru, dan kanker rahim. Penyakit tersebut diklasifikasikan berdasarkan data RNA penderita. Hasil dari klasifikasi juga memiliki akurasi yang baik sehingga diharapkan bisa dijadikan acuan untuk mengetahui jenis kanker pada tubuh seseorang agar dapat segera ditangani dengan baik.



### **BAB V**

### **PENUTUP**

## 5.1. Kesimpulan

Berdasarkan hasil penelitian ini tentang klasifikasi kanker berdasarkan dara RNA menggunakan metode ERNN didapatkan kesimpulan sebagai berikut.

- 1. Metode ERNN dapat diaplikasikan dengan baik untuk klasifikasi kanker berdasarkan data RNA karena dapat mengenali 4 kelas kanker Hal tersebut di tunjukkan dengan hasil *confusion matrix* yang rata-rata nya bernilai tinggi, pada akurasi 97.45% sensitifitas 95.19%, dan spesifisitas 99.12%.
- 2. Klasifikasi kanker berdasarkan data RNA menggunakan metode ERNN mndapatkan model terbaik pada *hidden layer* yang lebih kecil. Kombinasi parameter paling baik di dapatkan pada 20 node *hidden layer* I, 50 node *hidden layer* II, dan *learning rate* 0.1. Pada model tersebut di dapatkan akurasi mencapai 99.19 %, sensitivitas sebesar 99.03 % dan Spesifisitas sebesar 99.72 %. Waktu yang dibutuhkan untuk model tersebut yaitu 19 detik.

#### 5.2. Saran

Pada penelitian tentang klasifikasi kanker berdasarkan data RNA menggunakan metode Elman Recurrent Neural Network ini masih memiliki banyak kekurangan sehingga di harapkan kedepannya dapat diperbaiki lagi. Berikut beberapa saran yang dapat di perbaiki:

- 1. Menambahkan data RNA jenis kanker yang lain sehingga dapat di buat aplikasi yang bisa melakukan klasifikasi berbagai jenis kanker.
- Melakukan klasifikasi kanker berdasarkan data RNA menurut tingkat stadiumnya sehingga dapat diketahui tingkat keganasan kanker yang di klasifikasi.
- 3. Menambahkan data RNA kelas normal sehingga dapat digunakan untuk deteksi dini penyakit kanker.

## **DAFTAR PUSTAKA**

- Adiwijaya, A. (2018). Deteksi Kanker Berdasarkan Klasifikasi Microarray Data. *Jurnal Media Informatika BUDIDARMA*, 2(4):181–186.
- Akbar, D. L. and Budiyanto, B. (2020). Konsep kesehatan dalam al-qur'an dan hadis. *Al-Bayan: Jurnal Ilmu al-Qur'an dan Hadist*, 3(2):157–173.
- Andriyani, S. and Sitohang, N. (2018). Implementasi Metode Backpropagation Untuk Prediksi Harga Juanl Kelapa Sawit Berdasarkan Kualitas Buah. *JURTEKSI (Jurnal Teknol. dan Sist. Informasi*), 4(2):155–164.
- Anggorowati, L. (2013). Faktor risiko kanker payudara wanita. *KEMAS: Jurnal Kesehatan Masyarakat*, 8(2).
- Anindya, I. C. and Kantarcioglu, M. (2018). Adversarial anomaly detection using centroid-based clustering. In 2018 IEEE International Conference on Information Reuse and Integration (IRI), pages 1–8. IEEE.
- Anugerah, A. G. (2018). Klasifikasi Tingkat Keganasan Kanker Paru-paru pada Citra Computed Tomography (CT) Scan Menggunakan Metode Convolutional Neural Network= Classification Level Of Cancer Rate Rations In Computed Tomography (Ct) Scan Using Convolutional Neural Network Method.
- Chairi, A. N. Perspektif Agama Islam terhadap Keluarga Sehat di Kota Muara Bulian Kabupaten Batang Hari Jambi Tahun 2017. *Scientia Journal*, 6(1):120–128.

- Christiyawati, M. D., Yatmihatun, S., and Sumanto, S. (2016). Metode Terapi Hipnopunktur (Hipnoterapi Dan Terapi Akupunktur) Untuk Peningkatan Kualitas Hidup Pasien Kanker Rahim Pasca Kemoterapi Di Kota Surakarta. *Jurnal Keterapian Fisik*, 1(1):9–14.
- Dhahri, H. (2020). Biogeography-Based Optimization for Weight Optimization in Elman Neural Network Compared with Meta-Heuristics Methods. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, 11(2):82–103.
- Ekici, S. and Jawzal, H. (2020). Breast cancer diagnosis using thermography and convolutional neural networks. *Medical hypotheses*, 137:109542.
- Elbashir, M. K., Ezz, M., Mohammed, M., and Saloum, S. S. (2019). Lightweight convolutional neural network for breast cancer classification using RNA-seq gene expression data. *IEEE Access*, 7:185338–185348.
- Gunawan, B., Sastypratiwi, H., and Pratama, E. E. (2018). Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. *JEPIN* (*Jurnal Edukasi dan Penelitian Informatika*), 4(2):113–118.
- Harlianto, D., Mardiyati, S., Lestari, D., Zili, A. H., and Devila, S. (2020).
  Indonesia tuberculosis morbidity rate forecasting using recurrent neural network.
  In AIP Conference Proceedings, volume 2242, page 30006. AIP Publishing LLC.
- Hekler, A., Utikal, J. S., Enk, A. H., Hauschild, A., Weichenthal, M., Maron, R. C.,
  Berking, C., Haferkamp, S., Klode, J., and Schadendorf, D. (2019). Superior
  skin cancer classification by the combination of human and artificial intelligence.
  European Journal of Cancer, 120:114–121.
- Hendry, M. A., Gumanof, M. I., Mulya, F. R., and Meidelfi, D. (2021). Sistem

- Pakar Penyakit Ginjal Berbasis Web. *Inspiration: Jurnal Teknologi Informasi dan Komunikasi*, 11(1):80–88.
- Husen, A., Suharti, C., and Hardian, H. (2016). Hubungan Antara Derajat Nyeri Dengan Tingkat Kualitas Hidup Pasien Kanker Paru Yang Menjalani Kemoterapi. Diponegoro Medical Journal (Jurnal Kedokteran Diponegoro), 5(4):545–557.
- Jabeen, A., Ahmad, N., and Raza, K. (2018). Machine learning-based state-of-theart methods for the classification of rna-seq data. In *Classification in BioApps*, pages 133–172. Springer.
- Joseph, M., Devaraj, M., and Leung, C. K. (2019). DeepGx: deep learning using gene expression for cancer classification. In 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pages 913–920. IEEE.
- Kurniadi, W. and Maliki, I. (2018). Speech Recognition Menggunakan Elman Recurrent Neural Network Untuk Kata Dalam Bahasa Indonesia.
- Lesnussa, Y. A., Latuconsina, S., and Persulessy, E. R. (2015). Aplikasi Jaringan Saraf Tiruan Backpropagation untuk Memprediksi Prestasi Siswa SMA (Studi kasus: Prediksi Prestasi Siswa SMAN 4 Ambon). *Jurnal Matematika Integratif, ISSN*, pages 1412–6184.
- Maulida, N., Paramitha, D. F., Sukarno, E. A., and Arifin, A. Z. (2013). Klasifikasi Kanker Paru-Paru Menggunakan Pengolahan Citra. *J. Tek. Pomits*, 2(1).
- Melisa, J., Monoarfa, A., and Tjandra, F. (2016). Profil penderita karsinoma sel ginjal (renal cell carcinoma) di RSUP Prof. Dr. RD Kandou Manado periode 2013-2015. *e-CliniC*, 4(2).

- Pakan, P. D. (2019). Klasifikasi Data Kanker Berdasrkan Mikro Array Gen Menggunakan Jaringan Syaraf Tiruan. *Jurnal Ilmiah Flash*, 5(1).
- Pirouz, B., Shaffiee Haghshenas, S., Shaffiee Haghshenas, S., and Piro, P. (2020). Investigating a serious challenge in the sustainable development process: analysis of confirmed cases of COVID-19 (new type of coronavirus) through a binary classification using artificial intelligence and regression analysis. *Sustainability*, 12(6):2427.
- Pradana, A. C., Adiwijaya, A., and Aditsania, A. (2018). Implementasi Algoritma
  Binary Particle Swarm Optimization (BPSO) dan C4. 5 Decision Tree untuk
  Deteksi Kanker Berdasarkan Klasifikasi Microarray Data. *eProceedings of Engineering*, 5(3).
- Prasetio, R. T. and Susanti, S. (2019). Prediksi Harapan Hidup Pasien Kanker Paru Pasca Operasi Bedah Toraks Menggunakan Boosted k-Nearest Neighbor. *Jurnal Responsif: Riset Sains & Informatika*, 1(1):64–69.
- Priyanti, E. (2021). Penerapan Algoritma Neural Network untuk Klasifikasi Kanker Paru. *Bianglala Informatika*, 9(1):56–60.
- Putri, D. R., Adiwijaya, A., and Sibaroni, Y. (2020). Deteksi Kanker Berdasarkan Klasifikasi Data Microarray Menggunakan Least Absolute Shrinkage and Selection Operator dan Functional Link Neural Network. *Jurnal Media Informatika BUDIDARMA*, 4(4):947–954.
- Radjabaycolle, J. (2020). Prediksi Indeks Harga Konsumen (IHK) Kota Ambon Menggunakan Elman Recurrent Neural Network (ERNN). *Tensor: Pure and Applied Mathematics Journal*, 1(2):65–75.

- RASMI, R. P. (2020). Peningkatan Hasil Diagnosis Kanker Payudara dari Hasil Citra Mammogram Menggunakan Metode Ekstraksi Ciri dan Klasifikasi.
- Rusamsi, S. R., Adiwijaya, A., and Rohmawati, A. A. (2018). Deteksi Kanker Berdasarkan Klasifikasi Microarray Data Menggunakan Wavelet Trasnsform Dan Backpropagation Termodifikasi Dengan Conjugate Gradient Flechter Reeves. *eProceedings of Engineering*, 5(1).
- Saputra, P. A., Rafdhi, F., and Bernadisman, D. (2020). Sistem Pakar Diagnosa Penyakit Kanker Rahim dengan Metode Forward Chainning. *Jurnal Visualika*, 6(1):64–74.
- Shodiq, S. (2020). Penanganan Covid-19 Dalam Pendekatan Kaidah Fikih dan Ushul Fikih (Analisis Kebijakan Pembatasan Sosial Berskala Besar Dibidang Keagamaan). *Al-Adalah: Jurnal Hukum dan Politik Islam*, 5(2):113–134.
- Shon, H. S., Yi, Y., Kim, K. O., Cha, E.-J., and Kim, K.-A. (2019). Classification of stomach cancer gene expression data using CNN algorithm of deep learning. *Journal of Biomedical and Translational Research (JBTR)*, 20(1):15–20.
- Singh, S., Pandey, S. K., Pawar, U., and Janghel, R. R. (2018). Classification of ECG arrhythmia using recurrent neural networks. *Procedia computer science*, 132:1290–1297.
- Sirait, A. M. (2009). Hubungan kontrasepsi pil dengan tumor/kanker payudara di Indonesia.
- Sudarsono, A. (2016). Jaringan Syaraf Tiruan Untuk Memprediksi Laju Pertumbuhan Penduduk Menggunakan Metode Bacpropagation (Studi Kasus Di Kota Bengkulu). *Jurnal Media Infotama*, 12(1).

- Syarifudin, M. A. (2021). Klasifikasi kanker berdasarkan data RNA menggunakan Dilated Convolutional Neural Network.
- Thilagaraj, M., Arunkumar, N., Ramkumar, S., and Hariharasitaraman, S. (2021).
  Electrooculogram signal identification for elderly disabled using Elman network.
  Microprocessors and Microsystems, 82:103811.
- Walid, W. (2019). Peramalan Penjualan Harga Saham PT Bank Rakyat (Persero) Tbk BBRI Indonesia dengan Menggunakan Recurren Neural Nerwork (RNN). In *PRISMA, Prosiding Seminar Nasional Matematika*, volume 2, pages 139–147.
- Wang, J., Wang, J., Fang, W., and Niu, H. (2016). Financial time series prediction using elman recurrent random neural networks. *Computational intelligence and neuroscience*, 2016.
- Wang, Y., Wang, L., Yang, F., Di, W., and Chang, Q. (2021). Advantages of direct input-to-output connections in neural networks: The Elman network for stock index forecasting. *Information Sciences*, 547:1066–1079.
- Warto, W. (2019). Ibadah dan Kesehatan dalam Perspektif Islam dan Sains. Rausyan Fikr: Jurnal Pemikiran dan Pencerahan, 15(2).
- Xu, J., Zhang, Y., and Miao, D. (2020). Three-way confusion matrix for classification: A measure driven view. *Information sciences*, 507:772–794.
- Zhang, T., Song, S., Li, S., Ma, L., Pan, S., and Han, L. (2019). Research on gas concentration prediction models based on LSTM multidimensional time series. *Energies*, 12(1):161.
- Zhao, J., Mao, X., and Chen, L. (2019). Speech emotion recognition using deep

1D & 2D CNN LSTM networks. *Biomedical Signal Processing and Control*, 47:312–323.

