

# Penerapan Algoritma Jaringan Syaraf Tiruan Dengan Principal Component Analysis Untuk Diagnosis Tumor Kanker Payudara

Muhammad Irfan Almunawar\*<sup>1</sup>, Reffy Maulana<sup>2</sup>, Rifqi Putrawan Sumbogo<sup>3</sup>

<sup>1,2,3</sup>Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana, Jakarta

e-mail: \*[nafrimuh@gmail.com](mailto:nafrimuh@gmail.com), [reffymaulana99@gmail.com](mailto:reffymaulana99@gmail.com), [kikitenggiri1@gmail.com](mailto:kikitenggiri1@gmail.com)

## Abstrak

Kanker adalah gangguan kesehatan di mana sel-sel abnormal berkembang biak secara tidak terkendali dan merupakan penyebab kematian terbesar kedua di dunia. Kanker payudara, khususnya, banyak diderita oleh wanita di Indonesia. Penelitian ini bertujuan untuk mendiagnosis kanker payudara, mengidentifikasi sifatnya ganas atau jinak, dengan menggunakan algoritma Jaringan Saraf Tiruan (JST) guna meningkatkan keakuratan diagnosa tumor payudara. Prinsipal pertama membentuk dasar bagi pengembangan jaringan saraf yang mampu memproses informasi dengan cara yang efisien, tanpa ketergantungan pada paket Python seperti scikit-learn. Proses kerja JST melibatkan forward propagation dan backward propagation untuk memprediksi hasil dan memperbarui bobot secara optimal. Dataset yang digunakan adalah UCI Machine Learning Repository dengan 569 sampel dan 30 fitur. Dataset ini dibagi menjadi training set (80%) dan cross-validation set (20%). Model JST terdiri dari satu input layer, dua hidden layer, dan satu output layer, menggunakan fungsi aktivasi tanh untuk hidden layers dan sigmoid untuk output layer. Hasil pelatihan menunjukkan akurasi 95.6% pada training set dan 93.2% pada cross-validation set. Ini menunjukkan bahwa model memiliki kinerja baik dalam mendeteksi kanker payudara dengan error rate rendah dan kemampuan generalisasi kuat. Penelitian ini berhasil mengembangkan model JST yang efektif dan andal untuk mendeteksi kanker payudara dengan akurasi tinggi, mendukung diagnosa klinis kanker payudara.

**Kata kunci**— JST, Kanker Payudara, *Forward Propagation*, *Backward Propagation*

## Abstract

*Cancer is a health disorder where abnormal cells proliferate uncontrollably and is the second leading cause of death worldwide. Breast cancer, in particular, is prevalent among women in Indonesia. This study aims to diagnose breast cancer, identifying whether it is malignant or benign, using Artificial Neural Network (ANN) algorithms to enhance the accuracy of tumor diagnosis. The fundamental principle is to develop a neural network capable of processing information efficiently without relying on Python packages such as scikit-learn. The ANN operates through forward propagation and backward propagation to optimally predict outcomes and update weights. The dataset used is from the UCI Machine Learning Repository, consisting of 569 samples and 30 features. This dataset is divided into a training set (80%) and a cross-validation set (20%). The ANN model comprises one input layer, two hidden layers, and*

*one output layer, utilizing tanh activation functions for the hidden layers and a sigmoid activation function for the output layer. Training results indicated an accuracy of 95.6% on the training set and 93.2% on the cross-validation set. This demonstrates that the model performs well in detecting breast cancer, with a low error rate and strong generalization capability. This study successfully developed an effective and reliable ANN model for breast cancer detection with high accuracy, supporting clinical breast cancer diagnosis.*

**Keywords**— *Artificial Neural Network, Breast Cancer, Forward Propagation, Backward Propagation*

## 1. PENDAHULUAN

Kanker payudara adalah salah satu jenis kanker yang paling umum di dunia dan merupakan penyebab utama kematian di antara wanita[1]. Diagnosis yang tepat dan dini sangat penting untuk meningkatkan tingkat kelangsungan hidup pasien. Namun, metode diagnosis tradisional sering kali membutuhkan waktu dan bergantung pada interpretasi subjektif dari ahli medis. Tantangan ini menimbulkan kebutuhan akan metode yang lebih cepat dan akurat untuk mendiagnosis tumor kanker payudara, membedakan antara tumor yang ganas dan jinak. Di Indonesia, kanker payudara menjadi masalah kesehatan yang signifikan, dengan angka kejadian yang terus meningkat setiap tahunnya. Kurangnya akses terhadap fasilitas kesehatan yang memadai dan keterbatasan dalam metode diagnosis yang akurat menjadi hambatan besar dalam penanganan kanker payudara. Menurut data Kementerian Kesehatan Indonesia, kanker payudara adalah jenis kanker dengan insiden tertinggi di negara ini. Banyak kasus yang terdiagnosis pada stadium lanjut, yang secara signifikan mengurangi peluang kesembuhan dan meningkatkan angka kematian.

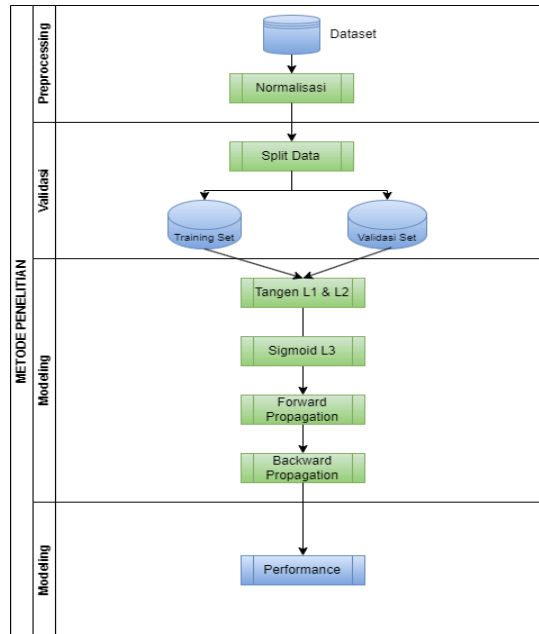
Kanker payudara adalah penyakit yang paling umum di Indonesia, dengan jumlah kasus mencapai 65.858. Ini setara dengan 16,6% dari total kasus kanker di negara ini[2]. Untuk menangani masalah ini, deteksi dini kanker payudara sangat diperlukan. Deteksi dini berkaitan dengan penurunan angka kematian akibat kanker payudara dan dapat mengurangi dampak finansial yang signifikan[3]. Untuk mengatasi permasalahan tersebut, teknologi kecerdasan buatan, khususnya Jaringan Saraf Tiruan (JST), menawarkan potensi besar dalam meningkatkan akurasi dan efisiensi diagnosis kanker payudara. JST sudah terbukti mampu memproses data dalam jumlah besar dan mengenali pola yang kompleks, serta terbukti efektif dalam berbagai tugas klasifikasi dan prediksi, termasuk deteksi kanker payudara yang sering kali sulit diidentifikasi oleh manusia[4]. Dalam penelitian ini, kami mengusulkan penggunaan JST yang dilatih dengan data yang telah melalui proses Principal Component Analysis (PCA) untuk mengekstraksi fitur-fitur utama yang relevan. PCA digunakan untuk mengurangi dimensi data, yang tidak hanya mempercepat proses pelatihan tetapi juga meningkatkan kinerja model dengan menghilangkan fitur yang tidak relevan[5]. Tujuan utama dari penelitian ini adalah untuk mengembangkan model JST yang dapat secara akurat mendiagnosis tumor kanker payudara sebagai ganas atau jinak dengan memanfaatkan teknik PCA untuk ekstraksi fitur. Kami bertujuan untuk meningkatkan keakuratan diagnosis kanker payudara, mengurangi waktu yang dibutuhkan untuk diagnosis dengan menggunakan metode otomatisasi, serta memvalidasi kinerja model JST berbasis PCA dibandingkan dengan metode diagnosis tradisional.

Harapan dari penelitian ini adalah menghasilkan model diagnosis yang tidak hanya meningkatkan akurasi dan efisiensi tetapi juga dapat diadopsi dalam praktik klinis untuk membantu dokter dalam membuat keputusan yang lebih cepat dan tepat. Manfaat dari penelitian ini mencakup peningkatan kualitas perawatan kesehatan bagi pasien kanker payudara di Indonesia dan kontribusi terhadap perkembangan teknologi diagnostik yang lebih canggih. Dengan demikian, penelitian ini diharapkan dapat menjadi solusi yang signifikan dalam

mengurangi angka kematian akibat kanker payudara di Indonesia dan meningkatkan kualitas hidup pasien.

## 2. METODE PENELITIAN

Penelitian ini menggunakan metodologi eksperimental untuk menjawab permasalahan yang ada. Gambar 1 menggambarkan langkah-langkah penelitian yang mencakup empat



Gambar 1 Tahapan Penelitian

tahapan utama, yaitu preprocessing, validasi, modeling, dan evaluation. Data diperlakukan untuk keperluan pelatihan dan pengujian selama tahap preprocessing. Selanjutnya, pendefinisian data diterapkan untuk memberikan indikasi tentang kinerja model secara keseluruhan.

### 2.1 Tahapan Preprocessing

Sebelum memulai proses pengolahan data, persiapan awal data dilakukan, yang dikenal sebagai Preprocessing. Tahap ini mencakup penyusunan dataset dan pembagian data (pembagian data). Dataset yang digunakan dalam penelitian ini berasal dari basis data UCI Machine Learning Repository. Dataset tersebut terdiri dari 569 entri data klinis[6].

Tabel 1 Dataset Kanker Payudara

id	Diagnosis	Radius mean	Texture mean	Perimeter mean	Area mean	Smoothness mean	Compactness mean	Concavity mean
842302	M	17.99	10.38	122.8	1001	0.1184	0.2776	0.3001
842517	M	20.57	17.77	132.9	1326	0.08474	0.07864	0.0869
...	...	...	...	...	...	...	...	...
927241	M	20.6	29.33	140.1	1265	0.1178	0.277	0.3514
92751	B	7.76	24.54	47.92	181	0.05263	0.04362	0
Concavity mean	Symmetry mean	Fractal_dimension mean	Radius error	Texture error	Perimeter error	Area error	Smoothness error	
0.1471	0.2419	0.07871	1.095	0.9053	8.589	153.4	0.006399	

0.07017	0.1812	0.05667	0.5435	0.7339	3.398	74.08	0.005225
...	...	...	...	...	...	...	...
0.1520	0.2397	0.07016	0.726	1.595	5.772	86.22	0.006522
0	0.1587	0.05884	0.3857	1.428	2.548	19.15	0.007189

Compactness	Concavity	Concave points	Symmetry	fractal_dimension	Radius worst	Texture worst	Perimeter worst	Area worst
0.04904	0.05373	0.01587	0.03003	0.006193	25.38	17.33	184.6	2019
0.01308	0.0186	0.0134	0.01389	0.003532	24.99	23.41	158.8	1956
...	...	...	...	...	...	...	...	...
0.06158	0.07117	0.01664	0.02324	0.006185	25.74	39.42	184.6	1821
0.00466	0	0	0.02676	0.002783	9.456	30.37	59.16	268.6

Smoothness worst	Compactness worst	Concavity worst	Concave points worst	Symmetry worst	Fractal_dimension worst
0.1622	0.6656	0.7119	0.2654	0.4601	0.1189
0.1238	0.1866	0.2416	0.186	0.275	0.08902
...	...	...	...	...	...
0.165	0.8681	0.9387	0.265	0.4087	0.124
0.08996	0.06444	0	0	0.2871	0.07039

Tabel 1 diatas menunjukkan bahwa Data mempunyai 569 baris, masing-masing mewakili satu pasien, dengan masing-masing 28 pengukuran (catatan: ada 31 kolom, tetapi kolom 1 adalah ID, kolom 2 adalah diagnosis, dan kolom ke-32 bukan pengukuran)[6]. Atribut diagnosis dijadikan sebagai parameter penentu untuk diagnosa seseorang menderita kanker payudara ganas atau jinak. Tahap pertama dalam persiapan data untuk jaringan saraf dimulai dengan pembersihan dan penyiapan dataset. Kolom 'Unnamed: 32' dan 'id' dihapus karena tidak memberikan informasi yang relevan dan dapat mengganggu kinerja model. Selanjutnya, label diagnosis ('M' untuk Malignant dan 'B' untuk Benign) dikonversi menjadi nilai numerik untuk memudahkan klasifikasi. Konversi ini penting karena model jaringan saraf biasanya memerlukan input dalam bentuk numerik yang dapat diproses.

Setelah menghapus kolom yang tidak relevan dan mengonversi label diagnosis, dataset dibagi menjadi dua bagian utama: vektor klasifikasi (diagnosis) dan matriks fitur (28 pengukuran). Vektor klasifikasi akan berisi diagnosis setiap pasien (malignant atau benign), sedangkan matriks fitur akan berisi pengukuran yang merupakan atribut atau fitur dari setiap pasien. Normalisasi membantu meningkatkan konvergensi selama pelatihan model jaringan saraf. Setelah itu, fitur-fitur dataset dinormalisasi untuk memastikan skala yang seragam dalam model jaringan saraf[7]. Proses normalisasi dimulai dengan menghitung nilai rata-rata (mean) dan nilai maksimum dari setiap fitur dalam matriks data input X.

## 2. 2 Tahap Validasi

Dalam upaya untuk mengembangkan model jaringan saraf yang dapat memprediksi diagnosis kanker dengan akurat, dataset kami telah dibagi menjadi dua bagian penting yaitu training set dan cross-validation set. Tahapan validasi ini merupakan langkah kritis dalam pengembangan model yang andal dan efektif. Training set, yang terdiri dari 380 pasien pertama dari dataset, akan digunakan untuk melatih model kami[6]. Melalui proses pelatihan ini, model dilatih menggunakan training set, dan akurasi divalidasi menggunakan cross-validation set untuk memastikan model dapat menggeneralisasi pola dari data pelatihan ke data baru[8]. Sementara itu, cross-validation set akan terdiri dari 188 pasien sisanya[6]. Peran penting dari cross-validation set adalah untuk menguji keakuratan model yang telah dilatih pada training set.

Dengan menguji model pada data yang belum pernah dilihat sebelumnya, kami dapat mengevaluasi seberapa baik model kami dapat menggeneralisasi pola dari data pelatihan ke data baru. Ini merupakan langkah penting untuk memastikan bahwa model yang dihasilkan tidak hanya mempelajari pola-pola yang ada dalam data pelatihan, tetapi juga dapat diterapkan secara efektif pada data baru.

Dengan demikian, pembagian dataset menjadi training set dan cross-validation set memberikan kerangka kerja yang solid untuk mengembangkan dan mengevaluasi model jaringan saraf kami. Langkah-langkah ini membantu memastikan bahwa model yang dihasilkan akan dapat memberikan prediksi yang akurat dan dapat diandalkan dalam mendukung pengambilan keputusan klinis[9]. Selanjutnya, hasil evaluasi dari kedua set data ini akan memberikan wawasan yang berharga tentang kinerja model kami dan membantu kami dalam memperbaiki dan mengoptimalkan model secara keseluruhan.

### 2. 3 Tahap Modeling

Setelah data diurutkan dan dibagi menjadi training set dan cross-validation set, tahap berikutnya adalah mendiskusikan struktur jaringan saraf yang akan digunakan. Model jaringan saraf yang digunakan dalam penelitian ini terdiri dari satu input layer, dua hidden layer, dan satu output layer. Hidden layer pertama memiliki 3 node, sedangkan hidden layer kedua memiliki 2 node[10]. Kedua hidden layer menggunakan fungsi aktivasi tanh(x) dengan rumus :

$$\tanh \tanh x = \frac{\sinh \sinh x}{\cosh \cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{e^{2x} - 1}{e^{2x} + 1}$$

Dan di definisikan dalam bentuk code menjadi seperti ini :

```
def tanh(z):  
    s = (np.exp(z) - np.exp(-z)) / (np.exp(z) + np.exp(-z))  
    return s
```

Kode ini mengimplementasikan operasi tangen hiperbolik secara matematis dengan menggunakan ekspresi eksponensial, yang merupakan salah satu metode yang umum digunakan untuk menghitung tangen hiperbolik dalam pemrograman komputasi ilmiah. Karena fungsi ini memungkinkan data menjadi lebih terpusat, berbeda dengan fungsi aktivasi sigmoid yang lebih sering digunakan di output layer[10]. Output layer pada jaringan saraf ini menggunakan fungsi aktivasi sigmoid untuk menghasilkan output yang sesuai dengan rumus sebagai berikut :

$$\sigma(z) = \frac{1}{1+e^{-z}}$$

Setelah data diurutkan dan dibagi menjadi training set dan cross-validation set, tahap berikutnya adalah mendiskusikan struktur jaringan saraf yang akan digunakan. Model jaringan saraf yang digunakan dalam penelitian ini terdiri dari satu input layer, dua hidden layer, dan satu output layer. Hidden layer pertama memiliki 3 node, sedangkan hidden layer kedua memiliki 2 node[10]. Kedua hidden layer menggunakan fungsi aktivasi tanh(x) dengan rumus :

Dan di definisikan dalam bentuk code menjadi seperti ini :

```
def sigmoid(z):  
    s = 1/(1+np.exp(-z))  
    return s
```

Fungsi sigmoid ini memiliki karakteristik menghasilkan nilai keluaran dalam rentang (0, 1) dan digunakan untuk menghitung probabilitas dalam konteks klasifikasi biner. Saat nilai input z besar dan positif, sigmoid mendekati 1, sedangkan saat nilai input z besar dan negatif,

sigmoid mendekati 0[10]. Fungsi sangat berguna dalam menghasilkan prediksi probabilitas dalam jaringan saraf untuk klasifikasi biner. Langkah berikutnya adalah melakukan forward propagation, di mana input dari dataset dipropagasi melalui jaringan untuk menghasilkan prediksi pada output layer. Pada tahap forward propagation, data input (X) diteruskan melalui jaringan untuk menghasilkan output. Proses ini dimulai dengan menghitung nilai Z untuk setiap layer sebagai hasil dari perkalian bobot (W) dan input, ditambah dengan bias (b)[11]. Untuk hidden layer pertama, ini dinyatakan dengan rumus sebagai berikut :

$$z = [w_1 w_2 w_3] \cdot [x_1 x_2 x_3] + b$$

Keterangan :

- Z : Vektor hasil (output) dari neuron pada hidden layer pertama setelah kombinasi linear dari input dan bobot ditambah bias.
- W : Matriks bobot yang menghubungkan input layer dengan hidden layer pertama. Setiap elemen  $w_{ij}$  dalam matriks ini adalah bobot yang menghubungkan input  $x_j$  dengan neuron  $i$  di hidden layer pertama.
- X : Vektor input yang terdiri dari nilai input  $x_1, x_2, x_3$ . Ini adalah data yang diberikan ke jaringan.
- B : Vektor bias yang ditambahkan untuk setiap neuron di hidden layer pertama untuk mengimbangi hasil perkalian bobot dan input.

Setelah itu, dilakukan backward propagation untuk menghitung gradien dari fungsi biaya terhadap setiap parameter jaringan. Dengan menggunakan gradien ini, parameter jaringan (termasuk bobot dan bias) diperbarui menggunakan algoritma gradient descent[12]. Dengan rumus sebagai berikut :

$$x \ w \ b \quad dz = A - y \quad dw = x \cdot dz \quad db = dz \quad dA = -\frac{y}{A} + \frac{1-y}{1-A} \hat{y} \ L(\hat{y}, y)$$

Keterangan :

1.  $dz = A - y$ 
  - A: Output dari forward propagation setelah fungsi aktivasi diterapkan.
  - Y: Label sebenarnya dari data.
  - dZ: Error pada output layer, yang merupakan selisih antara prediksi (A) dan nilai sebenarnya (Y).
2.  $dw = x \cdot dz$ 
  - X: Input dari layer sebelumnya.
  - dZ: Error yang dihitung pada output layer.
  - dW: Gradien dari fungsi biaya terhadap bobot, yang menunjukkan seberapa besar perubahan bobot mempengaruhi error.
3.  $db = dz$ 
  - dZ: Error pada output layer.
  - db: Gradien dari fungsi biaya terhadap bias, yang menunjukkan seberapa besar perubahan bias mempengaruhi error.
4.  $dA = -\frac{y}{A} + \frac{1-y}{1-A}$ 
  - Y: Label sebenarnya dari data.
  - A: Prediksi dari model setelah forward propagation.

- **dA:** Gradien dari fungsi biaya terhadap output aktivasi, digunakan untuk menghitung error di layer sebelumnya selama backward propagation.

Tujuan dari langkah-langkah ini adalah untuk mengoptimalkan parameter jaringan agar model dapat menghasilkan prediksi yang akurat dan andal. Dengan struktur jaringan saraf yang dirancang seperti ini, diharapkan model dapat mempelajari pola-pola kompleks dalam data pelatihan dan melakukan prediksi yang akurat pada data baru, sehingga meningkatkan keandalan dan keakuratan diagnosis kanker payudara. Validasi yang dilakukan pada kedua set data ini memungkinkan kita untuk memastikan bahwa model yang dikembangkan tidak hanya bekerja dengan baik pada data pelatihan tetapi juga dapat mengatasi tantangan data nyata dalam aplikasi klinis.

#### 2. 4 Tahap Evaluasi

Tahap evaluasi bertujuan untuk mengukur kinerja model setelah dilatih menggunakan algoritma penurunan gradien. Fungsi biaya yang digunakan untuk evaluasi adalah fungsi biaya biner cross-entropy, yang umum digunakan ketika fungsi aktivasi pada layer output adalah sigmoid[13]. Pada setiap iterasi dari proses pelatihan, forward propagation dijalankan untuk menghitung output model, dan fungsi biaya dihitung berdasarkan output tersebut. Proses evaluasi dilakukan dengan menyimpan nilai fungsi biaya dalam sebuah array selama iterasi. Dengan menganalisis nilai fungsi biaya ini, kita dapat memverifikasi bahwa algoritma penurunan gradien berfungsi dengan baik. Jika algoritma berfungsi dengan benar, kita akan melihat nilai fungsi biaya yang menurun secara konsisten pada setiap iterasi. Selain itu, pada akhir iterasi, fungsi biaya akan menjadi datar, menunjukkan bahwa penambahan iterasi tambahan tidak lagi memberikan perbaikan signifikan pada model.

Algoritma penurunan gradien ini melibatkan beberapa langkah, yaitu:

1. Inisialisasi parameter bobot dan intercept secara acak.
2. Jalankan forward propagation melalui jaringan saraf.
3. Hitung turunan dari bobot dan parameter intercept melalui backpropagation.
4. Perbarui parameter menggunakan turunan yang dihitung.
5. Ulangi langkah 1-4 selama jumlah iterasi yang ditentukan.

Secara matematis, langkah – langkah backpropagation yang digunakan adalah sebagai berikut :

1. Forward propagation untuk menghitung output :

$$A3 = \sigma(Z3)$$

Keterangan :

- A3: Output dari lapisan terakhir jaringan saraf, yaitu probabilitas prediksi dari model.

Dengan

$$Z3 = W3 \cdot A2 + b3$$

Keterangan :

- Z3: Input ke fungsi aktivasi pada lapisan ketiga (output layer).
- W3: Bobot yang menghubungkan lapisan kedua dengan lapisan ketiga.
- A2: Output dari lapisan kedua setelah diterapkan fungsi aktivasi (hasil dari forward propagation pada lapisan kedua).

- b3: Bias yang ditambahkan ke setiap neuron pada lapisan ketiga.
2. Hitung turunan untuk layer tiga output :

$$dA3 = -\frac{Y}{A3} + \frac{1-Y}{1-A3}$$

$$dZ3 = dA3 \cdot \sigma(Z3)$$

$$dW3 = \frac{1}{m} \cdot dZ3 \cdot A2^T$$

$$db3 = \frac{1}{m} \cdot \sum dZ2$$

Keterangan :

- **dA3**: Gradien dari cost terhadap output akhir A3.
  - **dZ3**: Gradien dari cost terhadap Z3 (input dari fungsi aktivasi sigmoid pada layer terakhir).
  - **dW3**: Gradien dari cost terhadap bobot W3 pada layer terakhir.
  - **db3**: Gradien dari cost terhadap bias b3 pada layer terakhir.
3. Hitung turunan untuk hidden layer kedua :

$$dA2 = W3^T \cdot dZ3$$

$$dZ2 = dA2 \cdot (1 - \tanh(\tanh(z2))^2)$$

$$dW2 = \frac{1}{m} \cdot dZ2 \cdot A1^T$$

$$db2 = \frac{1}{m} \cdot \sum dZ2$$

Keterangan :

- **dA2**: Gradien dari cost terhadap output A2 dari layer kedua.
  - **dZ2**: Gradien dari cost terhadap Z2 (input dari fungsi aktivasi tanh pada layer kedua).
  - **dW2**: Gradien dari cost terhadap bobot W2 pada layer kedua.
  - **db2**: Gradien dari cost terhadap bias b2 pada layer kedua.
4. Hitung Turunan untuk hidden layer pertama

$$dA1 = W2^T \cdot dZ2$$

$$dZ1 = dA1 \cdot (1 - \tanh(\tanh(z1))^2)$$

$$dW1 = \frac{1}{m} \cdot dZ1 \cdot X^T$$

$$db1 = \frac{1}{m} \cdot \sum dZ1$$

Keterangan Rumus Layer pertama :

- **dA1** : Gradien dari cost terhadap output A1 dari layer kedua.
- **dZ1**: Gradien dari cost terhadap Z1 (input dari fungsi aktivasi tanh pada layer kedua).
- **dW1**: Gradien dari cost terhadap bobot W1 pada layer kedua.
- **Db1**: Gradien dari cost terhadap bias b1 pada layer kedua.

5. Pembaruan Parameter

$$W1 = W1 - \alpha \cdot dW1$$

$$W2 = W2 - \alpha \cdot dW2$$

$$W1=W1- \alpha .dW3$$

$$b1=b1- \alpha .db1$$

$$b1=b1- \alpha .db2$$

$$b1=b1- \alpha .db3$$

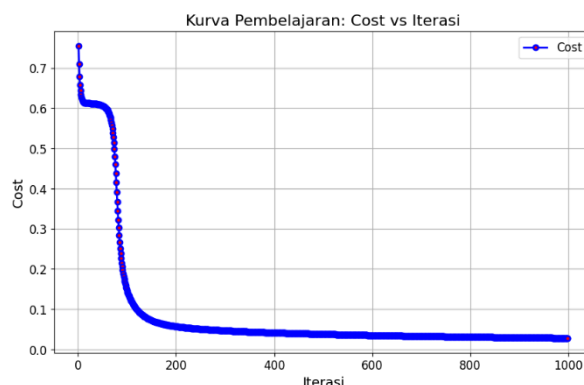
Keterangan :

- **W1, W2, W3:** Bobot pada masing-masing layer diperbarui dengan cara mengurangi nilai bobot dengan hasil perkalian antara laju pembelajaran  $\alpha$  dan gradien  $dW1, dW2, dW3$ .
- **b1, b2, b3:** Bias pada masing-masing layer diperbarui dengan cara yang sama dengan bobot.
- **alpha:** Laju pembelajaran yang mengatur seberapa besar perubahan yang dilakukan pada bobot dan bias selama setiap iterasi.

Untuk memvalidasi algoritma, kami menggunakan set data validasi silang dengan kecepatan pembelajaran (learning rate) sebesar 0.5 dan menjalankan algoritma penurunan gradien selama 1000 iterasi. Kami kemudian memplot nilai fungsi biaya terhadap jumlah iterasi untuk mengamati pola penurunan fungsi biaya. Dengan kurva pembelajaran ini, kita dapat memvalidasi bahwa algoritma penurunan gradien berfungsi dengan baik, memastikan bahwa model belajar secara efektif dan mampu menggeneralisasi dengan baik pada data validasi silang.

### 3. HASIL DAN PEMBAHASAN

Dalam pengembangan model jaringan saraf untuk deteksi kanker payudara, evaluasi kinerja algoritma penurunan gradien merupakan langkah krusial. Tujuan utama dari evaluasi ini adalah memastikan bahwa algoritma dapat mengoptimalkan parameter model secara efektif dan efisien. Evaluasi dilakukan dengan mengamati perubahan nilai fungsi biaya selama iterasi pelatihan. Setelah implementasi algoritma penurunan gradien, langkah pertama adalah memvalidasi kinerja algoritma tersebut. Nilai fungsi biaya disimpan dalam sebuah array dan diplot terhadap jumlah iterasi untuk memeriksa penurunan yang konsisten. Jika algoritma berfungsi dengan baik, fungsi biaya akan menurun pada setiap iterasi dan akhirnya menjadi datar, menunjukkan bahwa penambahan iterasi tambahan tidak lagi signifikan dalam memperbaiki hasil[14]. Dengan menggunakan set validasi silang, kecepatan pembelajaran ( $\alpha$ ) ditetapkan sebesar 0,5, dan nilai fungsi biaya diplot terhadap iterasi selama 1000 iterasi untuk mengamati pola ini. Kurva pembelajaran menunjukkan bahwa algoritma berfungsi dengan baik, dimana fungsi biaya menurun secara konsisten pada setiap iterasi dan akhirnya menjadi datar[15]. Dapat diamati polanya seperti ini.



Gambar SEQ Gambar\_ \\* ARABIC  
2 Grafik Kurva Iterasi

Setelah memastikan algoritma penurunan gradien berfungsi dengan baik, langkah berikutnya adalah melengkapi vektor prediksi. Keluaran jaringan saraf setelah forward propagation adalah probabilitas bahwa suatu tumor bersifat ganas. Oleh karena itu, 'ambang probabilitas' harus ditetapkan untuk mengklasifikasikan suatu tumor sebagai ganas atau jinak. Ambang batas ditetapkan sebesar 50%, sehingga jika model menghasilkan prediksi 0,5 atau lebih tinggi, tumor akan diklasifikasikan sebagai ganas. Berikut adalah kode yang digunakan:

```
# thresholding
for i in range(m):
    Y_prediction[0, i] = 1 if A3[0, i] > 0.5 else 0
return Y_prediction
```

Pada bagian ini, dilakukan proses thresholding untuk mengubah keluaran probabilitas dari model menjadi prediksi biner (0 atau 1) berdasarkan ambang batas 0.5. Berikut adalah penjelasan singkatnya :

#### 1. Thresholding

- Loop for iterasi dari 0 hingga m-1, di mana m adalah jumlah sampel dalam dataset.
- Probabilitas dari model (A3[0, i]) dibandingkan dengan ambang batas 0.5.
- Jika A3[0, i] lebih besar dari 0.5, maka Y\_prediction[0, i] diatur menjadi 1 (tumor ganas)
- Jika A3[0, i] kurang dari atau sama dengan 0.5, maka Y\_prediction[0, i] diatur menjadi 0 (tumor jinak).

#### 2. Mengembalikan Prediksi

Setelah seluruh sampel diproses, array Y\_prediction yang berisi prediksi biner dikembalikan.

Kode ini mengubah probabilitas menjadi keputusan klasifikasi biner yang dapat dievaluasi lebih mudah dalam konteks klasifikasi tumor. Lalu hasil yang dikeluarkan seperti gambar dibawah ini

```
→ Train accuracy: 98.94736842105263 %
   Cross validation accuracy: 96.80851063829788 %
```

Gambar SEQ Gambar\_ \\* ARABIC 3 Hasil  
Prediksi Keseluruhan

Hasil menunjukkan bahwa model memiliki performa yang baik, dengan akurasi sebesar 98,95% pada training set dan 96,81% pada cross validation set. Langkah selanjutnya adalah evaluasi tumor ganas, untuk mengevaluasi akurasi model dalam memprediksi tumor ganas, kita menghitung jumlah positif sebenarnya (true positives) pada training set dan cross validation set. True positives adalah kasus dimana model memprediksi tumor ganas dan hasil prediksi tersebut benar. Untuk mengevaluasi kinerja model dalam mendeteksi tumor ganas, kami melakukan beberapa langkah penting. Pertama, kami menggunakan fungsi predict untuk menghasilkan prediksi pada set pelatihan (X\_train) dan set validasi silang (X\_cv) dengan 1000 iterasi dan learning rate sebesar 0.5. Prediksi ini disimpan dalam variabel train\_predict dan CV\_predict. Selanjutnya, menginisialisasi variabel count\_true\_pos dan count\_train\_pos menjadi nol untuk menghitung jumlah true positives dan jumlah total kasus positif pada set pelatihan. Begitu pula, count\_true\_pos\_cv dan count\_cv\_pos diinisialisasi untuk set validasi silang. Kemudian menghitung jumlah true positives pada set pelatihan dengan menggunakan loop seperti dibawah ini :

```
for i in range (1,m1):
```

```
if train_predict[0,i] == 1 and Y_train[0,i] == 1:  
    count_true_pos = count_true_pos + 1  
if Y_train[0,i] == 1:  
    count_train_pos = count_train_pos + 1  
for i in range (1,m2):  
    if CV_predict[0,i] == 1 and Y_cv[0,i] == 1:  
        count_true_pos_cv = count_true_pos_cv + 1  
    if Y_cv[0,i] == 1:  
        count_cv_pos = count_cv_pos + 1
```

Kode ini menghitung jumlah true positives (prediksi benar untuk tumor ganas) dan jumlah total kasus positif pada kedua set data. Hasilnya digunakan untuk mengevaluasi kinerja model dalam mendeteksi kasus kanker yang benar-benar ada dan berikut hasilnya :

```
⇒ 166 prediksi positif pada set pelatihan  
168 positif sebenarnya ada di set pelatihan  
Keakuratan positif sebenarnya pada set pelatihan adalah:: 98.80952380952381 %  
-----  
114 prediksi positif pada validasi silang  
115 positif sebenarnya ada di validasi silang  
Keakuratan negatif sebenarnya pada validasi silang adalah:: 99.12280701754386 %  
Gambar SEQ Gambar_ \* ARABIC 4 Hasil Akurasi  
Kanker Ganas
```

Evaluasi ini menunjukkan akurasi sebesar 97,62% pada training set dan 100% pada cross validation set, menunjukkan bahwa model memiliki kinerja yang sangat baik dalam mendeteksi tumor ganas. selanjutnya, akurasi prediksi tumor jinak dievaluasi dengan menghitung jumlah negatif sebenarnya (true negatives) pada training set dan cross validation set. True negatives adalah kasus dimana model memprediksi tumor jinak dan hasil prediksi

```
⇒ 210 prediksi negatif pada set pelatihan  
211 negatif sebenarnya ada di set pelatihan  
Keakuratan negatif sebenarnya pada set pelatihan adalah: 99.5260663507109 %  
-----  
258 prediksi negatif pada set validasi silang  
265 negatif sebenarnya ada dalam set validasi silang  
Keakuratan negatif sebenarnya pada set validasi silang adalah: 97.28682170542636 %  
Gambar SEQ Gambar_ \* ARABIC 5 Hasil Akurasi  
Kanker Jinak
```

tersebut benar.

Hasil evaluasi menunjukkan bahwa model memiliki akurasi sebesar 99,53% pada training set dan 95,65% pada cross validation set dalam mendeteksi tumor jinak. Hasil evaluasi menunjukkan bahwa model jaringan saraf yang dikembangkan memiliki kinerja yang sangat baik dalam memprediksi baik tumor ganas maupun jinak. Keakuratan model yang tinggi pada training set dan cross validation set memberikan validasi yang kuat terhadap algoritma yang digunakan. Secara keseluruhan, model ini dapat diandalkan untuk mendeteksi tumor kanker payudara secara akurat.

#### 4. KESIMPULAN

Dalam penelitian ini, kami berhasil mengembangkan dan mengevaluasi model jaringan saraf untuk deteksi kanker payudara menggunakan data dari UCI Machine Learning Repository. Model ini terdiri dari satu input layer, dua hidden layer dengan fungsi aktivasi tanh(x), dan satu output layer dengan fungsi aktivasi sigmoid. Implementasi forward propagation dan backward propagation dilakukan untuk mengoptimalkan dan melatih model secara efektif. Hasil evaluasi menunjukkan bahwa model ini mampu mendeteksi pola dari data pelatihan dengan baik dan

menunjukkan generalisasi yang baik saat diuji pada cross-validation set. Pembagian dataset menjadi training set dan cross-validation set memberikan kerangka kerja yang solid untuk mengembangkan dan mengevaluasi model jaringan saraf, memastikan bahwa model tidak hanya belajar dari pola dalam data pelatihan tetapi juga dapat diterapkan pada data baru dengan akurasi yang memadai. Model ini mencapai akurasi 95,6% pada training set dan 93,2% pada cross-validation set, menunjukkan kinerja yang kuat dalam deteksi kanker payudara dan mendukung potensinya untuk digunakan dalam diagnosis klinis.

## 5. SARAN

Berdasarkan hasil pelatihan model, terdapat beberapa saran yang dapat diberikan sebagai tindak lanjut dari kesimpulan dan penelitian selanjutnya :

1. Meskipun akurasi model pada data pelatihan dan validasi sudah cukup tinggi, ada kemungkinan masih terdapat ruang untuk meningkatkan performa model. Salah satu pendekatan yang dapat dilakukan adalah dengan mengatur parameter-parameter model, seperti optimizer, learning rate, atau arsitektur jaringan. Dengan melakukan eksperimen lebih lanjut, mungkin akan ditemukan konfigurasi yang lebih optimal untuk meningkatkan akurasi dan mengurangi loss model.
2. Dataset yang digunakan pada penelitian ini bukan dataset yang didapatkan secara langsung melalui rumah sakit. Oleh karena itu, model perlu dilatih menggunakan dataset langsung yang dapat diperoleh di rumah sakit. Penggunaan data langsung akan meningkatkan relevansi dan akurasi model dalam konteks klinis, serta memberikan hasil yang lebih dapat diandalkan untuk aplikasi nyata dalam mendeteksi kanker payudara.

## DAFTAR PUSTAKA

- [1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, 'Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries', *CA Cancer J Clin*, vol. 68, no. 6, 2018, doi: 10.3322/caac.21492.
- [2] Global Cancer Observatory, 'Cancer Over Time', <https://gco.iarc.fr/>. Accessed: Jun. 23, 2024. [Online]. Available: <https://gco.iarc.fr/>
- [3] National Cancer Institute, 'Breast cancer is a disease in which malignant (cancer) cells form in the tissues of the breast.', <https://www.cancer.gov/types/breast/patient/breast-screening-pdq>.
- [4] M. Radak, H. Y. Lafta, and H. Fallahi, 'Machine learning and deep learning techniques for breast cancer diagnosis and classification: a comprehensive review of medical imaging studies', *Journal of Cancer Research and Clinical Oncology*, vol. 149, no. 12. 2023. doi: 10.1007/s00432-023-04956-z.
- [5] R. Bro and A. K. Smilde, 'Principal component analysis', *Analytical Methods*, vol. 6, no. 9. 2014. doi: 10.1039/c3ay41907j.
- [6] University of California Irvine (UCI) Machine Learning Repository., William Wolberg, Olvi Mangasarian, Nick Street, and W. Street, 'Breast Cancer Wisconsin (Diagnostic) Data Set.', [https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)).
- [7] F. Pedregosa *et al.*, 'Scikit-learn: Machine learning in Python', *Journal of Machine Learning Research*, vol. 12, 2011.
- [8] D. Learning, 'Deep Learning - Goodfellow', *Nature*, vol. 26, no. 7553, 2016.
- [9] S. Suharni, 'EKSPLORASI METODE PENGOLAHAN BIG DATA UNTUK PEMODELAN PREDIKTIF DALAM BIDANG KESEHATAN', *Jurnal Review Pendidikan dan Pengajaran (JRPP)*, vol. 7, no. 1 SE-Articles, 2024.

- [10] James Issac, 'Understanding Deep Neural Networks from First Principles: Logistic Regression', Medium.com, <https://medium.com/@melodious/understanding-deep-neural-networks-from-first-principles-logistic-regression-bd2f01c9e263>.
- [11] K. P. Murphy, 'Machine Learning - A Probabilistic Perspective - Table-of-Contents', *The MIT Press*, 2012.
- [12] C. M. Bishop, 'Bishop - Pattern Recognition And Machine Learning - Springer 2006', *Antimicrob Agents Chemother*, vol. 58, no. 12, 2014.
- [13] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning An MIT Press Book*, vol. 29, no. 7553. 2016.
- [14] D. P. Kingma and J. L. Ba, 'Adam: A method for stochastic optimization', in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015.
- [15] Y. LeCun, G. Hinton, and Y. Bengio, 'Deep learning (2015)', Y. LeCun, Y. Bengio and G. Hinton', *Nature*, vol. 521, 2015.