SKRIPSI

PEMODELAN *MULTILAYER PERCEPTRON* UNTUK KLASIFIKASI STATUS KEKAMBUHAN KANKER PARU BERDASARKAN PROFIL METILASI DNA

(STUDI KASUS : PASIEN *LUNG ADENOCARCINOMA* STADIUM I PASCAOPERASI)

NAFAL MAULANA ABDILLAH
22106010055

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
YOGYAKARTA

2026

PEMODELAN *MULTILAYER PERCEPTRON* UNTUK KLASIFIKASI STATUS KEKAMBUHAN KANKER PARU BERDASARKAN PROFIL METILASI DNA

(STUDI KASUS : PASIEN *LUNG ADENOCARCINOMA* STADIUM I PASCAOPERASI)

Skripsi

Untuk memenuhi sebagian persyaratan

mencapai derajat Sarjana S-1

Program Studi Matematika



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

diajukan oleh

NAFAL MAULANA ABDILLAH

22106010055

Kepada

PROGRAM STUDI MATEMATIKA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA

YOGYAKARTA

2026

SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR

Hal : Persetujuan Skripsi / Tugas Akhir
Lamp :

Kepada
Yth. Dekan Fakultas Sains dan Teknologi
UIN Sunan Kalijaga Yogyakarta
di Yogyakarta

Assalamu'alaikum wr. wb.

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka kami selaku pembimbing berpendapat bahwa skripsi Saudara:

Nama	: Nafal Maulana Abdillah
NIM	: 22106010055
Judul Skripsi	: Pemodelan *Multilayer Perceptron* Untuk Klasifikasi Status Kekambuhan Kanker Paru Berdasarkan Profil Metilasi DNA (Studi Kasus : Pasien *Lung Adenocarcinoma* Stadium I Pascaoperasi)

sudah dapat diajukan kembali kepada Program Studi Matematika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu dalam Program Studi Matematika.

Dengan ini kami mengharap agar skripsi/tugas akhir Saudara tersebut di atas dapat segera dimunaqasyahkan. Atas perhatiannya kami ucapkan terima kasih.

Wassalamu'alaikum wr. wb.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

Yogyakarta, 3 Agustus 2026
Pembimbing

Dr. Epha Diana Supandi, S.Si., M.Sc.
NIP: 197509122008012015

KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

PENGESAHAN TUGAS AKHIR

Nomor : B-2006/Un.02/DST/PP.00.9/08/2026

Tugas Akhir dengan judul : Pemodelan Multilayer Perceptron Untuk Klasifikasi Status Kekambuhan Kanker Paru Berdasarkan Profil Metilasi DNA (Studi Kasus : Pasien Lung Adenocarcinoma Stadium I Pascaoperasi)

yang dipersiapkan dan disusun oleh:

Nama : NAFAL MAULANA ABDILLAH
Nomor Induk Mahasiswa : 22106010055
Telah diujikan pada : Kamis, 13 Agustus 2026
Nilai ujian Tugas Akhir : A

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR

Ketua Sidang

Dr. Epha Diana Supandi, S.Si., M.Sc.
SIGNED

Valid ID: 6a87fe3627a0a

Penguji I

Muhamad Rashif Hilmi, S.Si., M.Sc.
SIGNED

Valid ID: 6a87d70e65442

Penguji II

Dr. Sugiyanto, S.Si., ST., M.Si.
SIGNED

Valid ID: 6a868ee7a384f

Yogyakarta, 13 Agustus 2026
UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi

Prof. Dr. Dra. Hj. Khurul Wardati, M.Si.
SIGNED

Valid ID: 6a880a4891f40

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

SURAT PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini:

Nama	: Nafal Maulana Abdillah
NIM	: 22106010055
Program Studi	: Matematika
Fakultas	: Sains dan Teknologi

Dengan ini menyatakan bahwa isi skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar sarjana di suatu Perguruan Tinggi dan sesungguhnya skripsi ini merupakan hasil pekerjaan penulis sendiri sepanjang pengetahuan penulis, bukan duplikasi atau saduran dari karya orang lain kecuali bagian tertentu yang penulis ambil sebagai bahan acuan. Apabila terbukti pernyataan ini tidak benar, sepenuhnya menjadi tanggung jawab penulis.

Yogyakarta, 4 Agustus 2026



Nafal Maulana Abdillah

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

HALAMAN PERSEMBAHAN

Karya sederhana ini penulis persembahkan untuk:

Kedua orang tua penulis yang telah memberikan do'a, kasih dan cinta nya yang tak terbatas.

Keluarga dekat, sahabat dan semua orang–orang yang pernah saya kenal.

Almamater kampus tercinta yang banyak memberikan warna dalam hidup selama 4 tahun Universitas Islam Negeri Sunan Kalijaga Yogyakarta.

"Matematika angkatan 2022 tercinta"

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

MOTTO

*"**Karena sesungguhnya sesudah kesulitan itu ada kemudahan. Sesung-guhnya sesudah kesulitan itu ada kemudahan.**"*

(QS. Al Insyirah:5-6)

*"**Tidak ada yang instan didunia ini. Nikmatilah prosesnya. Tidak peduli prosesnya lambat atau cepat dan yang terpenting jangan selalu bandingkan diri kita dengan orang lain karena disetiap kesuksesan seseorang dibaliknya selalu ada darah perjuangan yang diteteskan.**"*

*"**Percayalah akan potensi diri sendiri. Maksimalkan apa yang sudah Allah SWT. titipkan kepada diri kita sebagai bentuk rasa Syukur kepada-Nya**"*

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

PRAKATA

Puji syukur kehadirat Allah SWT yang telah melimpahkan rahmat dan karunianya yang tak ternilai harganya berupa keimanan, kesabaran, kekuatan dan kelancaran. Shalawat serta salam semoga selalu tercurahkan kepada Nabi Muhammad SAW sehingga penulis dapat menyelesaikan skripsi ini yang berjudul "Pemodelan *Multilayer Perceptron* untuk Klasifikasi Status Kekambuhan Kanker Paru Berdasarkan Profil Metilasi DNA (Studi kasus : Pasien *Lung Adenocarcinoma* stadium I pascaoperasi"

Mengenai skripsi ini, penulis ingin mengucapkan terima kasih kepada pembimbing skripsi penulis, Ibu Dr. Epha Diana Supandi, S.Si., M.Sc. yang sudah membimbing penulis hingga menyelesaikan tugas akhir dan memberikan saran dan dukungan yang membangun. Penulis ingin berterima kasih juga kepada Bapak Deddy Rahmadi, M.Sc., yang sudah membimbing penulis diawal dengan segala keterbatasan penulis dalam menyusun tugas akhir. Tanpa mengurangi rasa hormat, penulis juga ingin menyampaikan ucapan terima kasih dan apresiasi yang sebesar-besarnya kepada:

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

1. Bapak Prof. Noorhaidi, S.Ag., M.A., M.Phil., Ph.D, selaku Rektor UIN Sunan Kalijaga Yogyakarta.
2. Ibu Dr. Khurul Wardati, M.Si., selaku Dekan Fakultas Sains dan Teknologi, Universitas Islam Negeri Sunan Kalijaga Yogyakarta.
3. Ibu Dr. Epha Diana Supandi, S.Si., M.Sc, selaku Ketua Program Studi Matematika

4. Seluruh dosen program studi Matematika dan staf fakultas Sains dan Teknologi yang senantiasa memberikan ilmu dan layanan terbaik kepada penulis dari awal hingga akhir perkuliahan.

5. Kedua orang tua penulis dan keluarga besar yang telah memberikan banyak nasehat, kasih sayang, doa dan dukungannya sehingga penulis sampai ke titik ini.

6. Keluarga Matematika angkatan 2022 yang telah membersamai selama 4 tahun dibangku perkuliahan, saling mendukung, berbagi ilmu dan banyak kenangan manis yang selalu dikenang.

7. Teman-teman seperbimbingan dan seperjuangan

8. Terakhir semua pihak yang penulis tidak bisa sebutkan satu-persatu, semoga dimudahkan segala urusannya.

Penulis sangat menyadari tidak ada yang sempurna di dunia ini. Oleh karena itu segala bentuk saran, kritik atas skripsi ini penulis menerimanya dengan senang hati. Penulis berharap skripsi ini dapat memberikan manfaat khususnya bagi penulis dan semua kalangan yang membutuhkan untuk dijadikan referensi dalam menulis. Akhir kata semoga Allah SWT melimpahkan segala rahmat dan hidayah-Nya kepada kita semua, *Aamiin*.

Yogyakarta, 24 Agustus 2026

Nafal Maulana Abdillah

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR ISI

HALAMAN JUDUL . . . . . **i**

HALAMAN PERSETUJUAN . . . . . **ii**

HALAMAN PENGESAHAN . . . . . **iii**

HALAMAN PERNYATAAN . . . . . **iv**

HALAMAN PERSEMBAHAN . . . . . **v**

HALAMAN MOTTO . . . . . **vi**

PRAKATA . . . . . **vii**

DAFTAR ISI . . . . . **ix**

DAFTAR TABEL . . . . . **xiii**

DAFTAR GAMBAR . . . . . **xiv**

DAFTAR LAMBANG . . . . . **xv**

INTISARI . . . . . **xvii**

ABSTRACT . . . . . **xviii**

I PENDAHULUAN **1**

1.1. Latar Belakang Masalah 1

1.2. Rumusan Masalah 6

1.3. Batasan Masalah 6

1.4. Tujuan Penelitian 7

1.5. Manfaat Penelitian 8

1.6. Tinjauan Pustaka 9

1.7. Sistematika Penulisan 15

II LANDASAN TEORI **16**

2.1. Vektor dan Matriks 16

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

2.1.1. Vektor 16
2.1.2. Penjumlahan dan Pengurangan Vektor 17
2.1.3. Perkalian Skalar dengan Vektor 18
2.1.4. Perkalian Titik Vektor 18
2.1.5. Matriks 19
2.1.6. Penjumlahan dan Pengurangan Matriks 19
2.1.7. Perkalian Matriks dengan Skalar 20
2.1.8. Perkalian Matriks dengan Matriks 21
2.1.9. Transpos Matriks 22
2.2. Turunan 22
2.2.1. Turunan Fungsi Satu Peubah 22
2.2.2. Turunan Parsial 23
2.2.3. Aturan Rantai (*Chain Rule*) 23
2.3. *High-dimensional data* 24
2.4. *Feature Selection* 25
2.4.1. *Variance Threshold* 26
2.4.2. *Lasso regression* 27
2.5. *Machine Learning* 28
2.6. Klasifikasi 31
2.6.1. Evaluasi Model Klasifikasi 32
2.7. *Artificial Neural Network* 35
2.8. *Multilayer Perceptron* (MLP) 39
2.8.1. *Forward Propagation* 40
2.8.2. *Activation Function* 41
2.8.3. *Loss Function* 43
2.8.4. *Backpropagation* 45

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

2.9. *SHapley Additive exPlanation* (SHAP) 52
2.10. Kanker paru 54
2.11. Metilasi DNA 54
2.12. Representasi Matematis Data Metilasi DNA (*Beta-Value*) 56
III METODE PENELITIAN 58
3.1. Jenis dan Metode Penelitian 58
3.2. Jenis dan Sumber Data Penelitian 60
3.3. Objek dan Variabel Penelitian 61
3.4. Metode Pengumpulan Data 62
3.5. Metode Analisis Data 63
3.6. Alur Penelitian 67
IV PEMBAHASAN 68
4.1. Pemodelan Matematis dan Implementasi Komputasi *Multilayer Perceptron* (MLP) 68
4.1.1. Transformasi Linear pada Propagasi Maju (*Forward Propagation*) 69
4.1.2. Penerapan Fungsi Aktivasi Non-Linear: ReLU dan Softmax 76
4.1.3. Formulasi Fungsi Kerugian *Categorical Cross-Entropy* . . . 81
4.1.4. Penerapan Aturan Rantai (*Chain Rule*) pada Propagasi Balik (*Backpropagation*) 84
4.1.5. Optimisasi Parameter Menggunakan *Stochastic Gradient Descent* (SGD) 93
4.2. Contoh Perhitungan Manual Berdasarkan Hasil Pemodelan MLP . . 94
4.2.1. Inisiasi Awal 96
4.2.2. Perhitungan *Forward Propagation* 98
4.2.3. Perhitungan *Backpropagation* 106

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

4.2.4. Perhitungan *Update* Bobot Menggunakan *Stochastic Gradient Descent* (SGD) 114

V STUDI KASUS 118

5.1. Deskripsi dan Pra-pemrosesan data 118

5.2. Skenario Eksperimen 120

5.2.1. Desain *Pipeline* Pemodelan 121

5.2.2. Pengaturan Parameter 122

5.2.3. Optimasi Parameter 123

5.2.4. Skema Validasi Model 124

5.3. Evaluasi Performa Model 125

5.3.1. Evaluasi Hasil Optimasi Parameter 125

5.3.2. Evaluasi Akhir Model Terbaik 127

5.4. Interpretasi Model 130

5.5. Anotasi Genomik 131

VI PENUTUP 134

6.1. Kesimpulan 134

6.2. Saran 135

DAFTAR PUSTAKA 137

LAMPIRAN 142

A Data Penelitian 142

B *Source Code* 143

C Biodata Penulis 144

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR TABEL

1.1 Ringkasan Penelitian Terdahulu 12

2.1 Analogi antara Jaringan Saraf Biologis dan Jaringan Saraf Tiruan . . 36

4.1 Sampel Data Asli Status Kekambuhan Pasien 96

5.1 Data Metilasi DNA dan Klinis Pasien (Data X) 119

5.2 Data Status Kekambuhan (Data Y) 119

5.3 Ruang Pencarian *Hyperparameter* Pemodelan 123

5.4 Konfigurasi Parameter Model 126

5.5 Hasil Evaluasi Performa Model (Diurutkan Berdasarkan *Recall Mean*) 126

5.6 Hasil Evaluasi Performa Model 128

5.7 Top 20 Faktor Prediksi Kekambuhan Kanker Paru Stadium I Berdasarkan Skor SHAP 131

5.8 Hasil Anotasi Biomarker Prediktif ke dalam Gen Referensi Modern (hg38) 133

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR GAMBAR

2.1 Perbandingan *Classical Programming* dengan *Machine Learning* . . 29

2.2 Tabel *Confusion Matrix* 33

2.3 Perbandingan *Biological Neuron* dengan *Artificial Neuron*. 36

2.4 Arsitektur *Multilayer Perceptron* 39

2.5 Contoh Arsitektur *Multilayer Perceptron* 47

3.1 Diagram alur penelitian 67

4.1 Arsitektur *Multilayer Perceptron* yang digunakan pada penelitian ini 69

4.2 Contoh Arsitektur *Multilayer Perceptron* Untuk Perhitungan Manual 95

1.1 Data Metilasi DNA 142

2.1 *Source Code* Pemodelan *Multilayer Perceptron* 143

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR LAMBANG

$\mathbf{X}$	= matriks masukan (*input features*) berdimensi $N \times N_{in}$
$\mathbf{Y}$	= matriks target sesungguhnya (*true labels*) dalam *one-hot encoding*
$y_j, \mathbf{y}$	= nilai target sesungguhnya untuk kelas/neuron ke-j
N	= jumlah sampel data dalam satu *batch* populasi
N_L	= jumlah neuron/kelas pada lapisan keluaran (*output layer*)
l	= indeks lapisan (*layer*) pada jaringan saraf tiruan
j	= indeks neuron target pada lapisan saat ini (l)
i	= indeks neuron masukan pada lapisan sebelumnya ($l-1$)
m	= indeks perulangan (*iterator*) deret penjumlahan kelas/neuron
$z_j^{(l)}, \mathbf{Z}^{(l)}$	= nilai masukan mentah (*raw logit*) pada lapisan ke-l
$a_j^{(l)}, \mathbf{A}^{(l)}$	= nilai keluaran fungsi aktivasi (*activation output*) pada lapisan ke-l
$w_{ij}^{(l)}, \mathbf{W}^{(l)}$	= bobot (*weight*) penghubung neuron i lapisan $(l-1)$ ke neuron j lapisan l
$b_j^{(l)}, \mathbf{b}^{(l)}$	= bias pada lapisan ke-l

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

$\mathcal{L}$ = nilai fungsi kerugian (*Categorical Cross-Entropy Loss*) skalar

$\frac{\partial \mathcal{L}}{\partial \mathbf{Z}^{(l)}}$ = turunan parsial / gradien kerugian terhadap matriks *logit* lapisan ke-l

$\frac{\partial \mathcal{L}}{\partial \mathbf{W}^{(l)}}$ = turunan parsial / gradien kerugian terhadap matriks bobot lapisan ke-l

$\frac{\partial \mathcal{L}}{\partial \mathbf{b}^{(l)}}$ = turunan parsial / gradien kerugian terhadap vektor bias lapisan ke-l

$\frac{\partial \mathcal{L}}{\partial \mathbf{A}^{(l)}}$ = turunan parsial / gradien kerugian terhadap matriks aktivasi lapisan ke-l

α = laju pembelajaran (*learning rate*)

$\odot$ = operator produk Hadamard (perkalian matriks *element-wise*)

$\mathbb{I}(\cdot)$ = fungsi indikator (*binary masking matrix*)

$\mathbf{M}^T$ = transpos dari matriks $\mathbf{M}$

$\ln$ = fungsi logaritma natural

$\sum$ = operator penjumlahan deret (*sigma*)

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

INTISARI

Pemodelan *Multilayer Perceptron* untuk Klasifikasi Status Kekambuhan Kanker Paru Berdasarkan Profil Metilasi DNA

(Studi kasus : Pasien *Lung Adenocarcinoma* stadium I pascaoperasi)

Nafal Maulana Abdillah

22106010055

Meskipun tingkat keberhasilan operasi kanker paru *Lung Adenocarcinoma* (LUAD) Stadium I cukup tinggi, sekitar 27% pasien tetap mengalami kekambuhan fatal pasca-operasi. Profil metilasi DNA adalah prediktor epigenetik potensial untuk mendeteksi risiko ini, namun analisisnya sering terkendala oleh dimensi fitur yang sangat tinggi. Penelitian ini bertujuan mengklasifikasikan status kekambuhan dengan membangun algoritma *Multilayer Perceptron* (MLP) *from scratch* menggunakan NumPy guna memberikan transparansi matematis dan optimasi perhitungan yang mengefisiensikan komputasi model. Model terbaik didapat dengan parameter *Variance Threshold* (0,001) dan *Lasso* (maksimal 5.000 fitur), dilanjutkan pelatihan MLP dengan arsitektur 256 dan 64 *hidden neurons*, *learning rate* 0,1, serta 5.000 *epochs*. Pengujian model terbaik pada *data test* menghasilkan tingkat *Accuracy* 65% dan *Recall* 43% serta *Cross-Validation Accuracy* 63,04% dan *Recall* 56,67% di mana performa prediksi dipengaruhi oleh fenomena ketidakseimbangan kelas. Meskipun demikian, evaluasi menggunakan *SHapley Additive exPlanations* (SHAP) melalui anotasi genomik referensi hg38 berhasil mengidentifikasi gen-gen biomarker paling dominan dan berpengaruh terhadap kekambuhan tumor, yaitu TTC6, C5orf66, TDRD10, dan BHLHA15. Secara keseluruhan, pemodelan ini tidak hanya mampu melakukan klasifikasi status kekambuhan, tetapi juga sukses menyajikan transparansi matematis dan relevansi biologis di balik kekambuhan tumor.

Kata Kunci: Kanker Paru-Paru, Metilasi DNA, *High-Dimensional Data*, *Multilayer Perceptron*, NumPy, SHAP.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

ABSTRACT

Multilayer Perceptron Modeling for Lung Cancer Recurrence Status Classification Based on DNA Methylation Profiles

(Case Study: Postoperative Stage I Lung Adenocarcinoma Patients)

Nafal Maulana Abdillah

22106010055

Although the surgical success rate for Stage I Lung Adenocarcinoma (LUAD) is relatively high, approximately 27% of patients still experience fatal post-operative recurrence. DNA methylation profiles serve as potential epigenetic predictors for detecting this risk; however, their analysis is often constrained by high feature dimensionality. This study aims to classify recurrence status by constructing a Multilayer Perceptron (MLP) algorithm from scratch using NumPy to provide mathematical transparency and optimization that enhances computational efficiency. The optimal model was achieved using Variance Threshold (0.001) and Lasso (maximum of $5{,}000$ features), followed by MLP training with an architecture of 256 and 64 hidden neurons, a learning rate of 0.1, and $5{,}000$ epochs. Evaluation of the best model on the test dataset yielded an Accuracy of 65% and a Recall of 43%, alongside a Cross-Validation Accuracy of 63.04% and Recall of 56.67%, where predictive performance was influenced by class imbalance. Nevertheless, evaluation using SHapley Additive exPlanations (SHAP) paired with the hg38 reference genomic annotation successfully identified the most dominant biomarker genes influencing tumor recurrence: TTC6, C5orf66, TDRD10, and BHLHA15. Overall, this modeling approach not only classifies recurrence status but also successfully delivers mathematical transparency and biological relevance underlying tumor recurrence.

Keywords: Lung Cancer, DNA Methylation, High-Dimensional Data, Multilayer Perceptron, NumPy, SHAP.

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Di era digital, perkembangan dan persebaran informasi terjadi sangat masif. Era ini menuntun kita menuju ke era *big data*, dimana ketersediaan data berskala besar ini memicu adanya perubahan paradigma dalam penanganan data dan tentunya juga mempengaruhi dalam proses pemodelan matematika. Salah satu tantangan paling mendasar yang muncul dari fenomena ini adalah data berdimensi tinggi atau yang biasa kita kenal dengan *high-dimensional data*. Kondisi ini terjadi ketika jumlah fitur atau variabel (p) jauh melebihi jumlah sampel atau observasi (n) yang secara matematis dapat direpresentasikan sebagai $p \gg n$. Jika ditinjau dari dalam ruang dimensi tinggi, volume ruang data berekspansi secara eksponensial sehingga titik-titik antar data menjadi sangat tersebar atau renggang. Fenomena yang secara matematis dikenal sebagai *Curse of Dimensionality* ini menyebabkan jarak antar observasi menjadi kurang bermakna, sehingga menyulitkan algoritma pembelajaran mesin untuk menemukan pola klasifikasi yang valid untuk digeneralisasi (Hastie et al., 2015).

Metode konvensional seperti regresi logistik linier sering kali gagal dalam menangani masalah klasifikasi dengan struktur matriks $p \gg n$ (Negnevitsky, 2011). Secara aljabar, ketika jumlah variabel lebih besar daripada jumlah observasi, maka sistem persamaan linier yang terbentuk bersifat *underdetermined*. Jika jumlah variabel sebanyak p variabel, maka sistem persamaan linear perlu sebanyak n persama-

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

an linear dengan $n \geq p$ untuk bisa menemukan satu jawaban pasti. Kasus ini tentu mengarah pada masalah singularitas. Matriks kovarians dari data berdimensi tinggi tidak dapat di invers, sehingga estimasi parameter menggunakan metode kuadrat terkecil atau *Maximum Likelihood Estimation* (MLE) tidak memiliki solusi tunggal yang unik. Selain itu, metode linier konvensional berasumsi bahwa batas antara p dan n itu berupa garis lurus atau bidang datar, sehingga tidak memiliki fleksibilitas untuk menangkap pola interaksi non-linear yang sangat kompleks antar variabel.

Sebagai solusi atas keterbatasan aljabar pada pemodelan linier tersebut, *Multilayer Perceptron* (MLP) menawarkan kerangka komputasi non-linear yang sangat tangguh. Secara konseptual, MLP merupakan bentuk arsitektur dasar dari jaringan saraf tiruan yang memproses data masukan melalui komposisi fungsi bersarang. Dalam formulasi matematisnya, komputasi pada setiap lapisan tersembunyi direpresentasikan sebagai transformasi affine, yaitu komputasi aljabar berupa perkalian matriks bobot ($w_{ji}^{(l)}$) dengan vektor fitur input ($a_i^{(l-1)}$) yang kemudian dijumlahkan dengan vektor bias ($b_j^{(l)}$). Hasil transformasi linier ini tidak berdiri sendiri, melainkan dilewatkan pada sebuah fungsi aktivasi nonlinear ($\sigma^{(l)}$), sehingga luaran dari suatu lapisan ke-l dapat secara matematis dinotasikan sebagai $a_j^{(l)} = \sigma^{(l)}\left(\sum_{i=1}^{N_{l-1}} w_{ji}^{(l)} a_i^{(l-1)} + b_j^{(l)}\right)$ (Goodfellow et al., 2016). Proses propagasi maju atau *forward propagation* ini secara berurutan memproyeksikan data spasial berdimensi tinggi ke dalam ruang fitur baru yang lebih representatif untuk klasifikasi (Bishop & Nasrabadi, 2006).

Keunggulan dari MLP tidak hanya terletak pada struktur matriksnya, melainkan pada keberadaan fungsi aktivasi nonlinear nya. Tanpa adanya fungsi aktivasi nonlinear, serangkaian perkalian matriks pada MLP secara aljabar akan tereduksi menjadi satu pemetaan linier tunggal, yang pada akhirnya akan mengulangi kele-

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

mahan dan singularitas regresi konvensional. Melalui penambahan nonlinearitas inilah MLP dilandasi oleh Teorema Aproksimasi Universal (Heryadi & Wahyono, 2021). Teorema tersebut memberikan jaminan matematis bahwa sebuah jaringan *feedforward* dengan setidaknya satu lapisan tersembunyi dan jumlah neuron yang memadai, berkapasitas untuk mengaproksimasi fungsi analitik kontinyu sekompleks apa pun dengan tingkat kepresisian yang diinginkan. Dalam konteks klasifikasi medis, kapasitas ini menjadikan MLP sangat ideal untuk memodelkan pola interaksi genetik tingkat tinggi antar CpG sites metilasi DNA, yang hubungannya di dalam sel secara biologis bersifat sangat rumit, sinergis, dan tidak linier (Moore et al., 2010).

Selain fleksibilitas pemetaannya yang dijamin oleh teorema, ketangguhan MLP juga ditopang oleh kerangka optimisasi numerik yang solid. Untuk meminimalkan kesalahan prediksi, MLP menggunakan fungsi kerugian (*loss function*) seperti *Categorical Cross Entropy*. Fungsi kerugian ini bertindak sebagai fungsi objektif dalam proses pembelajarannya yang menjadi step awal proses optimasi parameter. Proses pencarian parameter optimal dilakukan melalui algoritma propagasi balik atau *backpropagation*, yang merupakan implementasi komputasional dari Aturan Rantai (*Chain Rule*) dalam kalkulus multivariabel (Rumelhart et al., 1986). Algoritma ini menghitung secara eksak nilai gradien dari fungsi kerugian terhadap jutaan parameter bobot dan bias yang ada di dalam jaringan. Melalui metode optimisasi berbasis gradien, parameter matriks jaringan diperbarui secara iteratif untuk bergerak menuruni lereng fungsi kerugian menuju titik konvergensi minimum, memungkinkan model untuk mengadaptasikan parameter secara optimal (LeCun et al., 2015).

Salah satu bidang yang sangat membutuhkan keandalan integrasi formulasi

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

matematis tersebut untuk menangani data berdimensi tinggi adalah bidang *precision medicine*, spesifiknya pada ranah analisis genomik. Dalam bioinformatika, data molekuler seorang pasien direpresentasikan dalam bentuk matriks yang memiliki ratusan ribu titik fitur genetika, namun sering kali hanya segelintir fitur spesifik yang benar-benar memiliki relevansi klinis terhadap perkembangan suatu penyakit.

Salah satu patologi dengan urgensi tertinggi secara global adalah kanker paru-paru(Bray et al., 2024). Penyakit ini secara konsisten menempati peringkat teratas sebagai penyebab mortalitas akibat kanker, baik di dunia maupun di Indonesia(Bray et al., 2024). Dari berbagai subtipe histologis yang ada, *Lung Adenocarcinoma* (LUAD) merupakan varian yang paling dominan, mencakup sekitar 40 % dari total insiden kanker paru-paru (Denisenko et al., 2018). Penanganan klinis kanker paru-paru pada varian LUAD ini tentu sangat bergantung pada fase dimana tumor tersebut terdeteksi, sehingga menjadikan diagnosis pada tahap yang paling awal sebagai kunci utama penyelamatan pasien.

Fakta medis menunjukkan bahwa secara umum, membasmi kanker paru tidak berhenti pada meja operasi saja. Sekitar 30% hingga 55% pasien *Non-small cell lung cancer* (NSCLC) yang telah menjalani operasi reseksi kuratif pada akhirnya tetap mengalami kekambuhan dan meninggal akibat penyakit tersebut (Uramoto & Tanaka, 2014). Tantangan medis yang krusial juga masih terlihat pada pasien di tahap awal. Meskipun secara makroskopis tumor primer telah diangkat sepenuhnya, data klasifikasi TNM dari *International Association for the Study of Lung Cancer* (IASLC) menunjukkan bahwa tingkat kelangsungan hidup 5 tahun (*5-year survival rate*) pada pasien patologis Stadium I tidaklah sempurna, yakni berkisar antara 73% hingga 90% tergantung pada ukuran sub-stadium tumor. Fakta ini secara matematis mengonfirmasi bahwa hingga 27% pasien NSCLC Stadium I pada akhirnya tetap

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

mengalami kegagalan pengobatan yang berakibat pada kematian (Goldstraw et al., 2016). Fenomena ini mengindikasikan bahwa terdapat perubahan genetik tingkat molekuler yang tidak kasat mata, yang membutuhkan pendekatan komputasi seperti *Machine Learning* untuk mendeteksi potensi kekambuhan sejak dini.

Sebagai respons terhadap kegagalan diagnosis makroskopis tersebut, pemantauan pada tingkat molekuler, khususnya melalui modifikasi epigenetik seperti metilasi DNA, menawarkan solusi *early warning system* yang jauh lebih sensitif. Metilasi DNA merupakan proses biologis di mana gugus metil ditambahkan pada area *CpG sites* di untai DNA. Profil metilasi ini sering kali mengalami hipermetilasi atau hipometilasi jauh sebelum mutasi genetik primer atau manifestasi fisik tumor muncul. Pengukuran metilasi DNA, seperti yang disediakan oleh basis data *The Cancer Genome Atlas* (TCGA), menghasilkan data dengan ratusan ribu kolom *CpG sites* yang menjadikannya representasi sempurna dari permasalahan data berdimensi tinggi ($p \gg n$) yang membutuhkan penyelesaian matematis dan komputasi yang kokoh.

Berdasarkan pemaparan permasalahan komputasi dan urgensi medis di atas, penelitian ini dilakukan untuk menyelesaikan persoalan klasifikasi status kekambuhan pada pasien kanker paru tipe LUAD berbasis profil metilasi DNA. Penelitian ini difokuskan pada perumusan persamaan matematis dan implementasi algoritma *Multilayer Perceptron* dari dasar dengan memanfaatkan pustaka NumPy yang ada di bahasa pemrograman python. Untuk mengatasi data berdimensi tinggi yang melekat pada data epigenetik, arsitektur model akan didahului oleh tahapan seleksi fitur menggunakan kombinasi *Variance Threshold* dan regresi LASSO, guna memastikan bahwa model yang dihasilkan tidak hanya akurat secara prediksi, tetapi juga transparan, terhindar dari *overfitting*, dan efisien secara komputasional.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

1.2. Rumusan Masalah

Berdasarkan uraian pada latar belakang, diperoleh rumusan masalah dari penelitian ini yaitu sebagai berikut:

1. Bagaimana Pemodelan algoritma *Multilayer Perceptron* (MLP) untuk klasifikasi status kekambuhan pada pasien kanker paru *Lung Adenocarcinoma* (LUAD) stadium I pascaoperasi berdasarkan data metilasi DNA?
2. Bagaimana implementasi algoritma berdasarkan pemodelan MLP menggunakan pustaka dasar NumPy untuk klasifikasi status kekambuhan pasien LUAD stadium I berdasarkan data metilasi DNA?
3. Bagaimana performa model MLP yang diimplementasikan dalam mengklasifikasikan status kekambuhan pasien LUAD stadium I berdasarkan data metilasi DNA?
4. Bagaimana menentukan tingkat kepentingan variabel pada data metilasi DNA berdasarkan model MLP terbaik?

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

1.3. Batasan Masalah

Agar penelitian tetap fokus dan terukur, maka Ruang lingkup permasalahan yang dibahas dalam penelitian ini dibatasi pada hal-hal berikut:

1. Data yang digunakan dibatasi pada data metilasi DNA dari penderita kanker paru-paru tipe *Lung Adenocarcinoma* (LUAD) khusus pada Stadium I pascaoperasi. Data jenis kanker paru lainnya seperti *Squamous Cell Carcinoma* atau *Small Cell Lung Cancer* maupun stadium lanjut tidak termasuk dalam cakupan penelitian ini.

2. Klasifikasi yang dilakukan bersifat biner, yaitu membedakan antara pasien yang pada akhirnya mengalami kekambuhan dan pasien yang terbebas dari tumor pascaoperasi reseksi. Penelitian ini tidak memprediksi waktu terjadinya kekambuhan atau *Survival Analysis*.

3. Model komputasi yang dibangun dibatasi pada arsitektur *Multilayer Perceptron* (MLP) sederhana. Penelitian ini tidak melakukan perbandingan performa dengan arsitektur *Deep Learning* tingkat lanjut lainnya.

4. Pembangunan model MLP dibentuk dari dasar dengan mengimplementasikan langsung hasil formulasi matematikanya menggunakan pustaka NumPy pada bahasa pemrograman Python. *Framework Machine Learning* tingkat tinggi seperti TensorFlow, Keras, atau PyTorch tidak digunakan untuk membangun arsitektur inti dari *Neural Network* tersebut.

5. Performa model diukur berdasarkan *Confusion Matrix* dengan metrik standar klasifikasi, seperti Akurasi (*Accuracy*), Presisi (*Precision*), Sensitivitas (*Recall*), dan *F1-Score*.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

1.4. Tujuan Penelitian

1. Memodelan algoritma *Multilayer Perceptron* (MLP) untuk klasifikasi status kekambuhan pada pasien kanker paru *Lung Adenocarcinoma* (LUAD) stadium I pascaoperasi berdasarkan data metilasi DNA?

2. Mengimplementasi algoritma berdasarkan pemodelan MLP menggunakan pustaka dasar NumPy untuk klasifikasi status kekambuhan pasien LUAD stadium I berdasarkan data metilasi DNA?

3. Mengevaluasi performa model MLP yang diimplementasikan dalam meng-

klasifikasikan status kekambuhan pasien LUAD stadium I berdasarkan data metilasi DNA?

4. Menentukan tingkat kepentingan variabel pada data metilasi DNA berdasarkan model MLP, serta mengetahui implikasi klinisnya terhadap status kekambuhan pasien LUAD Stadium I?

1.5. Manfaat Penelitian

Secara teoritis, penelitian ini diharapkan dapat memberikan kontribusi literatur terhadap bidang Matematika dan komputasi biomedis, khususnya dalam memetakan interaksi nonlinear pada data berdimensi tinggi ($p \gg n$) seperti profil metilasi DNA. Melalui implementasi algoritma *Multilayer Perceptron* (MLP) secara fundamental menggunakan pustaka NumPy, penelitian ini membuka proses (*black-box*) dari mekanisme *Neural Network*. Hal ini akan memberikan pemahaman matematis yang komprehensif mengenai proses komputasi pada *forward propagation* dan *backpropagation* sehingga dapat menjadi rujukan ilmiah bagi penelitian selanjutnya yang berfokus pada fondasi arsitektur dasar pembelajaran mesin.

Secara praktis, model komputasi yang dihasilkan berpotensi menjadi landasan awal bagi pengembangan sistem pendukung keputusan klinis di bidang onkologi. Model ini dapat membantu tenaga medis dalam mengidentifikasi secara dini pasien kanker paru tipe *Lung Adenocarcinoma* (LUAD) stadium I yang memiliki risiko tinggi mengalami kekambuhan, sehingga langkah intervensi dan pemantauan pasca-operasi dapat diprioritaskan. Selain itu, basis kode program yang dibangun dari dasar dalam penelitian ini dapat dimanfaatkan oleh para pengembang dan peneliti lain sebagai *baseline* yang transparan, mudah dimodifikasi, dan dikembangkan lebih lanjut untuk kebutuhan klasifikasi genomik di masa mendatang.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

1.6. Tinjauan Pustaka

Penelitian ini mengkaji literatur ilmiah dari berbagai riset terdahulu, yang mencakup artikel jurnal, tesis, skripsi, buku teks, serta sumber akademik terpercaya lainnya yang relevan dengan topik bahasan. Berikut adalah beberapa penelitian terdahulu yang dijadikan acuan utama:

1. Sandoval et al. (2013) melakukan penelitian dengan menerapkan metode analisis kelangsungan hidup (*survival analysis*) melalui pendekatan Cox Proportional Hazards Regression untuk mengidentifikasi biomarker metilasi DNA dalam memprediksi risiko kekambuhan pasca-operasi pada pasien kanker paru *Non-Small-Cell Lung Cancer* (NSCLC) stadium I. Hasil penelitian menunjukkan bahwa hipermetilasi pada lima gen spesifik (HIST1H4F, PCDHGB6, NPBWR1, ALX1, dan HOXA9) merupakan prediktor independen yang signifikan terhadap waktu kelangsungan hidup bebas kambuh (*Relapse-Free Survival*) yang lebih singkat, di mana kelompok pasien dengan tingkat metilasi tinggi terbukti memiliki probabilitas risiko kekambuhan yang jauh lebih mematikan dibandingkan dengan kelompok pasien bermetilasi rendah.
2. Astuti & Adiwijaya (2019) mengkaji solusi untuk mengatasi masalah kutukan dimensi (*curse of dimensionality*) pada data genetik microarray melalui metode ekstraksi fitur. Penelitian tersebut memanfaatkan teknik statistika multivariat *Principal Component Analysis* (PCA) untuk mereduksi dimensi sebelum dilakukan klasifikasi menggunakan algoritma *Linear Discriminant Analysis* (LDA). Hasil pengujian menyimpulkan bahwa penerapan PCA secara komputasional mampu menyusutkan ribuan fitur data microarray tanpa kehilangan informasi varians yang vital. Kombinasi pendekatan ekstraksi fitur dan klasifikasi ini sukses meningkatkan efisiensi waktu pemrosesan sekaligus

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

menghasilkan tingkat akurasi deteksi kanker yang optimal.

3. Penelitian oleh Mangalik et al. (2023) menyoroti urgensi tahap seleksi fitur dalam memproses dataset microarray kanker yang memiliki fitur berlebih. Studi ini mengimplementasikan algoritma optimasi *Binary Particle Swarm Optimization* (BPSO) pada tahap pra-pemrosesan guna menyaring kumpulan fitur genetik yang paling relevan dan mengeliminasi fitur pendeteksi yang redundan. Subset data hasil seleksi BPSO tersebut kemudian diklasifikasikan menggunakan algoritma *Distance Weight K-Nearest Neighbor* (DWKNN). Hasil penelitian membuktikan bahwa algoritma BPSO sangat efektif dalam mereduksi noise data, yang ditandai dengan adanya lonjakan metrik performa akurasi klasifikasi yang signifikan pada model DWKNN jika dibandingkan dengan pemodelan yang menggunakan seluruh fitur data mentah.

4. Almufareh et al. (2026) mengeksplorasi penggunaan model komputasi berbasis Deep Learning untuk mengklasifikasikan subtipe utama kanker paru, yakni Adenokarsinoma dan Karsinoma Sel Skuamosa, menggunakan profil metilasi DNA. Melalui validasi multi-kohort, arsitektur Jaringan Saraf Tiruan pada studi ini terbukti mampu mencapai performa klasifikasi yang sangat baik. Selain performa prediksi, keunggulan utama dari studi ini terletak pada penerapan algoritma *SHapley Additive exPlanations* (SHAP) untuk interpretabilitas model. Analisis SHAP tersebut berhasil memetakan dan mengurutkan situs-situs CpG spesifik yang memberikan kontribusi terbesar terhadap keputusan klasifikasi, sehingga mampu memberikan penjelasan secara transparan dan bermakna biologis terhadap model komputasi yang bersifat *black-box*.

5. (Penelitian ini) melakukan penelitian dengan menerapkan metode Machine Learning melalui arsitektur *Multi-Layer Perceptron* (MLP) yang dikombina-

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

sikan dengan teknik reduksi dimensi Regresi LASSO dan *variance threshold* untuk mengklasifikasikan status kekambuhan pasien kanker paru stadium I pascaoperasi berdasarkan data profil metilasi DNA, serta mengidentifikasi fitur situs CpG teratas yang paling signifikan pengaruhnya terhadap probabilitas kekambuhan melalui algoritma interpretasi *SHapley Additive exPlanations* (SHAP).

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

Tabel 1.1 Ringkasan Penelitian Terdahulu

No	Judul Penelitian	Tujuan	Metode	Variabel	Hasil
1	*A Prognostic DNA Methylation Signature for Stage I Non–Small-Cell Lung Cancer* (Sandoval et al., 2013)	Identifikasi biomarker metilasi DNA untuk memprediksi pasien kanker paru Non-Small-Cell Lung Cancer (NSCLC) stadium I yang berisiko tinggi mengalami kekambuhan setelah operasi	Seleksi fitur dilakukan menggunakan Univariate dan Multivariate Cox Proportional Hazards Regression dengan teknik step-wise selection. Evaluasi statistik dilakukan menggunakan metode Kaplan-Meier, Log-rank test, dan kurva time-dependent ROC.	Variabel Independen (X): Skor risiko klinis yang dihitung berdasarkan status metilasi DNA pada lima gen spesifik hasil seleksi (HIST1H4F, PCDHGB6, NPBWR1, ALX1, dan HOXA9). Variabel Dependen (Y): Relapse-Free Survival (RFS), yakni variabel waktu kontinu yang dihitung sejak tanggal operasi hingga terjadinya peristiwa kekambuhan atau kematian.	Penelitian menemukan bahwa hipermetilasi pada kelima gen tersebut merupakan prediktor independen yang signifikan terhadap waktu bebas kambuh (RFS) yang lebih singkat (p-value < 0.05). Secara klinis, pasien dengan metilasi tinggi (pada dua gen atau lebih) memiliki Hazard Ratio (HR) yang tinggi, menandakan risiko kekambuhan pasca-operasi yang jauh lebih mematikan dibandingkan pasien bermetilasi rendah.
2	*Principal Component Analysis* Sebagai Ekstraksi Fitur Data Microarray Untuk Deteksi Kanker Berbasis *Linear Discriminant Analysis* (Astuti & Adiwijaya, 2019)	Mereduksi masalah kutukan dimensi (*curse of dimensionality*) pada data genetik microarray yang sangat besar agar proses komputasi deteksi kanker menjadi lebih cepat, efisien, dan akurat.	Ekstraksi fitur dan reduksi dimensi menggunakan statistika multivariat *Principal Component Analysis* (PCA). Fitur hasil ekstraksi tersebut kemudian diklasifikasikan menggunakan algoritma *Linear Discriminant Analysis* (LDA).	Variabel Independen (X): Komponen utama (Principal Components) hasil transformasi dan ekstraksi dari ribuan fitur data ekspresi gen (microarray). Variabel Dependen (Y): Label kelas status pasien (Normal vs Kanker).	Penggunaan PCA secara signifikan mampu mereduksi ribuan dimensi fitur microarray tanpa kehilangan banyak varians informasi penting. Kombinasi PCA dan metode klasifikasi LDA terbukti mampu meningkatkan efisiensi waktu komputasi sekaligus menghasilkan akurasi deteksi kanker yang sangat baik.

Bersambung ke halaman berikutnya...

Lanjutan Tabel 1.1

No	Judul Penelitian	Tujuan	Metode	Variabel	Hasil
3	Analisis Seleksi Fitur Binary PSO Pada Klasifikasi Kanker Berdasarkan Data Microarray Menggunakan DWKNN (Mangalik et al., 2023)	Mengoptimalkan kinerja algoritma klasifikasi kanker dengan menyaring dan menyeleksi fitur genetik (dari data microarray berdimensi tinggi) yang paling relevan dan membuang fitur yang redundan (tidak penting).	Seleksi fitur dengan Binary Particle Swarm Optimization (BPSO). klasifikasinya menggunakan metode Distance Weight K-Nearest	Variabel Independen (X): Kumpulan subset fitur (gen) yang telah terseleksi secara optimal dari dataset microarray asli. Variabel Dependen (Y): Kategori atau jenis kanker yang diuji.	Penerapan seleksi fitur BPSO terbukti sangat efektif membuang noise dan fitur yang tidak relevan. Klasifikasi DWKNN yang menggunakan data hasil seleksi BPSO mengalami lonjakan metrik performa yang signifikan dibandingkan model DWKNN yang menggunakan keseluruhan data microarray mentah.
4	*Deep-Learning-Based Classification of Lung Adenocarcinoma and Squamous Cell Carcinoma Using DNA Methylation Profiles: A Multi-Cohort Validation Study* (Almufareh et al., 2026)	Mengklasifikasikan subtipe kanker paru antara Adenokarsinoma dan Karsinoma Sel Skuamosa menggunakan model Deep Learning, serta mengidentifikasi biomarker metilasi DNA yang paling berpengaruh menggunakan algoritma SHAP	klasifikasi menggunakan arsitektur *Deep Learning* dengan validasi *multi-cohort*. intepretasi model menggunakan algoritma SHAP.	Variabel Independen (X): data metilasi DNA pada Adenokarsinoma dan Karsinoma Sel Skuamosa. Variabel Dependen (Y): Kelas subtipe kanker paru Adenokarsinoma atau Karsinoma Sel Skuamosa.	Model Deep Learning terbukti mencapai performa klasifikasi yang sangat tinggi. Selain itu, analisis SHAP berhasil memetakan dan mengurutkan situs CpG spesifik yang memiliki kontribusi terbesar terhadap hasil prediksi, memberikan nilai penjelasan biologis yang transparan pada model.

Bersambung ke halaman berikutnya...

Lanjutan Tabel 1.1

No	Judul Penelitian	Tujuan	Metode	Variabel	Hasil
5	Pemodelan *Multilayer Perceptron* untuk Klasifikasi Status Kekambuhan Kanker Paru Berdasarkan Profil Metilasi DNA (Studi Kasus : Pasien *Lung Adenocarcinoma* stadium I pascaoperasi) (Penelitian ini)	Membangun dan mengoptimasi model *MultiLayer Perceptron* untuk klasifikasi status kekambuhan pasien kanker paru pascaoperasi berdasarkan profil metilasi DNA, serta mengidentifikasi fitur metilasi DNA paling krusial yang berkontribusi terhadap prediksi model MLP.	Seleksi fitur dengan *variance threshold* dan *lasso regression*. Kemudian fitur tersebut di klasifikasikan menggunakan MLP. fitur penting MLP di intepretasikan menggunakan SHAP	Variabel Independen (X): data metilasi DNA pada *adenocarcinoma* Variabel Dependen (Y) : Status Kekambuhan, yaitu 1 (Kambuh) dan 0 (Tidak kambuh).	Diperolehnya arsitektur model MLP yang paling optimal beserta hasil evaluasi performa klasifikasinya melalui berbagai metrik ukur. Selain itu, dihasilkan pula daftar situs CpG teratas yang paling signifikan pengaruhnya terhadap probabilitas kekambuhan kanker berdasarkan nilai SHAP value, yang kemudian dianotasikan secara biologis.

1.7. Sistematika Penulisan

Sistematika pada penelitian ini yaitu:

BAB I PENDAHULUAN

Pada bab ini berisi tentang latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, tinjauan pustaka dan sistematika penulisan.

BAB II DASAR TEORI

Pada bab ini menjelaskan terkait landasan teori yang menjadi acuan, berisi tentang konsep, hipotesis dan rancangan penelitian.

BAB III METODE PENELITIAN

Pada bab ini membahas terkait metode penelitian yang akan digunakan dalam pemecahan masalah yang berisi jenis dan metode penelitian, jenis dan sumber data penelitian, objek dan variabel penelitian, metode pengumpulan data, metode analisis data, serta alur penelitian.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini akan dilakukan perancangan arsitektur model

BAB V STUDI KASUS

Pada bab ini akan dilakukan proses analisis data, implementasi model, dan optimalisasi model yang kemudian hasilnya diinterpretasikan secara terperinci.

BAB VI PENUTUP

Pada bab ini berisi kesimpulan dan saran dari penelitian yang dilakukan oleh penulis.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

BAB VI

PENUTUP

6.1. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, pemodelan matematis algoritma *Multilayer Perceptron* (MLP) untuk mengklasifikasikan status kekambuhan pasien kanker paru *Lung Adenocarcinoma* (LUAD) stadium I berbasis data metilasi DNA berhasil dirumuskan secara terperinci mencakup alur *forward propagation*, perhitungan *loss*, *backpropagation*, hingga *weight update*. Pemodelan ini menerapkan optimasi perhitungan matematis melalui penyederhanaan perhitungan matematika sehingga mampu meningkatkan efisiensi komputasi model secara signifikan. Seluruh formulasi matriks tersebut terbukti fungsional dan presisi saat diimplementasikan secara langsung menggunakan pustaka dasar NumPy tanpa bergantung pada *framework deep learning* eksternal.

Model *Multilayer Perceptron* (MLP) dengan performa terbaik dicapai melalui kombinasi hiperparameter *Variance Threshold* sebesar 0,001, seleksi fitur *Lasso* maksimal 5.000 fitur, *learning rate* 0,1, arsitektur *hidden layer* berjumlah 256 neuron pada *hidden layer* pertama dan 64 neuron pada *hiddden layer* kedua, serta pelatihan selama 5.000 *epochs*. Pada tahap evaluasi *Cross-Validation*, konfigurasi ini menghasilkan metrik *Accuracy* sebesar 63,04%, *Recall* 56,67%, *Precision* 41,0%, dan *F1-Score* 46,1%, sementara pada pengujian *data test* diperoleh *Accuracy* sebesar 65%, *Recall* 43%, *Precision* 43%, dan *F1-Score* 43%. Selanjutnya, interpretasi tingkat kepentingan variabel menggunakan metode SHAP yang dianotasikan ke ka-

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

mus genomik hg38 berhasil mengidentifikasi gen **TTC6**, **C5orf66**, **TDRD10**, dan **BHLHA15** sebagai fitur metilasi DNA yang paling dominan dan berpengaruh dalam menentukan risiko kekambuhan pasien.

6.2. Saran

Berdasarkan hasil evaluasi dan sejumlah keterbatasan yang diidentifikasi selama penelitian ini berlangsung, terdapat beberapa saran yang direkomendasikan untuk penyempurnaan dan pengembangan penelitian di masa mendatang:

1. **Penanganan Ketidakseimbangan Kelas (*Imbalanced Data*):** Mengingat tingginya angka *False Positive* dan performa prediksi yang masih bias terhadap kelas mayoritas (sembuh), penelitian selanjutnya sangat disarankan untuk mengimplementasikan teknik penanganan ketidakseimbangan data. Pendekatan pada tingkat data seperti *Synthetic Minority Over-sampling Technique* (SMOTE) atau pendekatan pada tingkat algoritma seperti penerapan bobot kelas (*Class Weighting*) dan *Focal Loss* dapat dieksplorasi untuk memaksa model lebih memperhatikan pola dari kelas minoritas (kambuh).

2. **Penerapan Metode Optimasi *Hyperparameter* Lanjutan:** Walaupun *Random Search* terbukti secara statistik lebih efisien daripada *Grid Search*, pencarian parameter pada penelitian ke depan dapat ditingkatkan dengan menggunakan metode optimasi terpandu seperti *Bayesian Optimization* (misalnya memanfaatkan kerangka kerja *Optuna*). Metode ini mampu belajar dari riwayat iterasi sebelumnya untuk mengarahkan pencarian ke area parameter yang paling menjanjikan, sehingga berpotensi menghasilkan arsitektur model yang lebih optimal dengan waktu komputasi yang lebih efisien.

3. **Integrasi Pemodelan *Multi-Omics*:** Kekambuhan kanker paru merupakan

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

peristiwa dengan mekanisme biologis yang sangat kompleks. Untuk meningkatkan ketajaman prediksi model, penelitian selanjutnya dapat mengintegrasikan data metilasi DNA dengan modalitas data molekuler lainnya (seperti data ekspresi gen RNA-Seq atau data mutasi somatik). Penggabungan pemodelan secara *multi-omics* akan memberikan konteks patofisiologis yang lebih utuh dibandingkan profil epigenomik tunggal.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR PUSTAKA

Almufareh, M. F., Tehsin, S., Humayun, M., Kausar, S., & Farooq, A. (2026). Deep-learning-based classification of lung adenocarcinoma and squamous cell carcinoma using dna methylation profiles: A multi-cohort validation study. *Cancers*, 18(4).

Anton, H., Bivens, I. C., & Davis, S. (2012). *Calculus: Early Transcendentals*. John Wiley & Sons, 10th edition.

Anton, H. & Rorres, C. (2014). *Elementary Linear Algebra: Applications Version*. John Wiley & Sons, Hoboken, New Jersey, 11th edition.

Astuti, W. & Adiwijaya, K. (2019). Principal component analysis sebagai ekstraksi fitur data microarray untuk deteksi kanker berbasis linear discriminant analysis. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 3.

Baylin, S. B. & Jones, P. A. (2011). A decade of exploring the cancer epigenome—biological and translational implications. *Nature Reviews Cancer*, 11(10):726–734.

Bergstra, J. & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(2):281–305.

Bibikova, M., Barnes, B., Huo, T., Sabaliauskas, V., Kelly, E., Kristen, C., et al. (2011). High density dna methylation expression profiling on the infinium platform. *Genomics*, 98(4):288–295.

Bird, A. (2007). Perceptions of epigenetics. *Nature*, 447(7143):396–398.

STATE ISLAMIC UNIVERSITY SUNAN KALIJAGA YOGYAKARTA

Bishop, C. M. & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning*. Springer.

Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & Jemal, A. (2024). Global cancer statistics 2022: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 74(3):229–263, https://acsjournals.onlinelibrary.wiley.com/doi/abs/10.3322/caac.21834.

Chollet, F. & Chollet, F. (2021). *Deep learning with Python*. simon and schuster, 2nd edition.

Cohen, P. R. (1995). *Empirical methods for artificial intelligence*, volume 139. MIT press Cambridge, MA.

Creswell, J. W. & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

Dedeurwaerder, S., Defrance, M., Calonne, E., Denis, H., Sotiriou, C., & Fuks, F. (2011). Evaluation of the infinium humanmethylation450 beadchip for the significance of dna methylation analysis in human cancer. *Epigenomics*, 3(6):771–784.

Denisenko, T. V., Budkevich, I. N., & Zhivotovsky, B. (2018). Cell death-based treatment of lung adenocarcinoma. *Cell death & disease*, 9(2):117.

Du, P., Zhang, X., Huang, C. C., Jafari, M., Kibbe, W. A., Hou, L., & Lin, S. M. (2010). Comparison of beta-value and m-value methods for quantifying methylation levels by illumina infinium platform. *BMC Bioinformatics*, 11:587.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

Esteller, M. (2008). Epigenetics in cancer. *New England Journal of Medicine*, 358(11):1148–1159.

Fawcett, T. (2006). Introduction to roc analysis. *Pattern Recognition Letters*, 27:861–874.

Goldstraw, P., Chansky, K., Crowley, J., Rami-Porta, R., Asamura, H., Eberhardt, W. E., Nicholson, A. G., Groome, P., Mitchell, A., Bolejack, V., et al. (2016). The iaslc lung cancer staging project: proposals for revision of the tnm stage groupings in the forthcoming (eighth) edition of the tnm classification for lung cancer. *Journal of Thoracic oncology*, 11(1):39–51.

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning*. MIT press Cambridge.

Guyon, I. & Elisseeff, A. (2003). An introduction to variable and feature selection. *J. Mach. Learn. Res.*, 3(null):1157–1182.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics)*. Springer.

Hastie, T., Tibshirani, R., & Wainwright, M. (2015). Statistical learning with sparsity. *Monographs on statistics and applied probability*, 143(143):8.

He, H. & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering*, 21(9):1263–1284.

Heryadi, Y. & Wahyono, T. (2020). *Machine Learning: Konsep dan Implementasi*. GAVA MEDIA.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

Heryadi, Y. & Wahyono, T. (2021). *Dasar-Dasar Deep Learning dan Implementasinya*. Gava Media, Yogyakarta.

Jones, P. A. (2012). Functions of dna methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics*, 13(7):484–492.

Kuhn, M. & Johnson, K. (2013). *Applied Predictive Modeling*. Mathematics and Statistics. Springer New York.

Laird, P. W. (2010). Principles and challenges of genome-wide dna methylation analysis. *Nature Reviews Genetics*, 11(3):191–203.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553):436–444.

Lundberg, S. M. & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.

Mangalik, Y., Saragih, T. H., Nugrahadi, D., Muliadi, M., & Mazdadi, M. (2023). Analisis seleksi fitur binary pso pada klasifikasi kanker berdasarkan data microarray menggunakan dwknn. *Jurnal Informatika Polinema*, 9:143–152.

Moore, M. J., Soltis, P. S., Bell, C. D., Burleigh, J. G., & Soltis, D. E. (2010). Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proceedings of the National Academy of Sciences*, 107(10):4623–4628.

Negnevitsky, M. (2011). *Artificial Intelligence: A Guide to Intelligent Systems*. Pearson Education Limited.

Nielsen, M. A. (2015). *Neural Networks and Deep Learning*. Determination Press, San Francisco, CA.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

Oates, B. J., McLean, R., & Griffiths, M. (2022). *Researching information systems and computing*. SAGE publications Ltd.

Purcell, E. J., Varberg, D., & Rigdon, S. E. (2007). *Calculus*. Pearson Prentice Hall, 9th edition.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *nature*, 323(6088):533–536.

Sandoval, J., Mendez-Gonzalez, J., Nadal, E., Chen, G., Carmona, F. J., Sayols, S., Moran, S., Heyn, H., Vizoso, M., Gomez, A., et al. (2013). A prognostic dna methylation signature for stage i non–small-cell lung cancer. *Journal of clinical oncology*, 31(32):4140–4147.

Sokolova, M. & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45:427–437.

Stewart, J. (2015). *Calculus*. Cengage Learning, 8th edition.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288.

Uramoto, H. & Tanaka, F. (2014). Recurrence after surgery in patients with nsclc. *Translational lung cancer research*, 3(4):242.

Zhou, W., Laird, P. W., & Shen, H. (2017). Comprehensive characterization, annotation and innovative use of infinium dna methylation beadchip probes. *Nucleic Acids Research*, 45(4):e22–e22, https://doi.org/10.1093/nar/gkw967.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA