

**PENDEKATAN MACHINE LEARNING MULTIKELAS
UNTUK KLASIFIKASI SUBTIPE KANKER GINJAL
MENGUNAKAN DATA EKSPRESI MIRNA TCGA**



LAPORAN RESEARCH & TECHNOLOGY PROGRAM

**MITAWATI
00000070345**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2026**

**PENDEKATAN MACHINE LEARNING MULTIKELAS
UNTUK KLASIFIKASI SUBTIPE KANKER GINJAL
MENGUNAKAN DATA EKSPRESI MIRNA TCGA**



LAPORAN RESEARCH & TECHNOLOGY PROGRAM

Diajukan sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana S.Kom

MITAWATI
Mitawati

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2026

HALAMAN PERNYATAAN ORISINALITAS TIDAK PLAGIAT

Dengan ini saya,

Nama : Mitawati
NIM : 00000070345
Program Studi : Informatika

Menyatakan dengan sesungguhnya bahwa Laporan Research & Technology Program saya yang berjudul:

Pendekatan Machine Learning Multikelas untuk Klasifikasi Subtipe Kanker Ginjal Menggunakan Data Ekspresi miRNA TCGA

merupakan hasil karya saya sendiri, bukan merupakan hasil plagiat, dan tidak pula dituliskan oleh orang lain; Semua sumber, baik yang dikutip maupun dirujuk, telah saya cantumkan dan nyatakan dengan benar pada bagian Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan karya ilmiah, saya bersedia menerima konsekuensi untuk dinyatakan TIDAK LULUS. Saya juga bersedia menanggung segala konsekuensi hukum yang berkaitan dengan tindak plagiarisme ini sebagai kesalahan saya pribadi dan bukan tanggung jawab Universitas Multimedia Nusantara.

Tangerang, 30 Desember 2025



Mitawati

HALAMAN PERNYATAAN PENGGUNAAN BANTUAN KECERDASAN ARTIFISIAL (AI)

Saya yang bertanda tangan di bawah ini:

Nama : Mitawati
NIM : 00000070345
Program Studi : Informatika
Judul Laporan : Pendekatan Machine Learning Multikelas untuk Klasifikasi
Subtipe Kanker Ginjal Menggunakan Data Ekspresi miRNA
TCGA

Dengan ini saya menyatakan secara jujur menggunakan bantuan Kecerdasan
Artifisial (AI) dalam pengerjaan Laporan sebagai berikut :

- ☒ Menggunakan AI sebagaimana diizinkan untuk membantu dalam menghasilkan ide-ide utama saja
- ☐ Menggunakan AI sebagaimana diizinkan untuk membantu menghasilkan teks pertama saja
- ☐ Menggunakan AI untuk menyempurnakan sintaksis dan tata bahasa untuk pengumpulan tugas
- ☐ Karena tidak diizinkan: Tidak menggunakan bantuan AI dengan cara apa pun dalam pembuatan tugas

Saya juga menyatakan bahwa:

- (1) Menyerahkan secara lengkap dan jujur penggunaan perangkat AI yang diperlukan dalam tugas melalui Formulir Penggunaan Perangkat Kecerdasan Artifisial (AI)
- (2) Mengakui telah menggunakan bantuan AI dalam tugas saya baik dalam bentuk kata, paraphrase, penyertaan ide atau fakta penting yang disarankan oleh AI dan saya telah menyantumkan dalam sitasi serta referensi
- (3) Terlepas dari pernyataan di atas, tugas ini sepenuhnya merupakan karya saya sendiri

Tangerang, 30 Desember 2025



Mitawati

HALAMAN PERNYATAAN KEABSAHAN PERUSAHAAN

Nama : Mitawati
NIM : 00000070345
Program Studi : Informatika
Fakultas : Teknik dan Informatika

menyatakan bahwa saya melaksanakan kegiatan di:

Nama Perusahaan/Organisasi : LPPM
Alamat : Kampus UMN, Scientia Garden, JL.
Boulevard Gading Serpong, Tangerang
Email Perusahaan/Organisasi : penelitian.lppm@umn.ac.id

1. Perusahaan/Organisasi tempat saya melakukan kegiatan dapat divalidasi keberadaannya.
2. Jika dikemudian hari terbukti ditemukan kecurangan/penyimpangan data yang tidak valid, maka:
 - a. Saya bersedia menerima konsekuensi dinyatakan **TIDAK LULUS**.
 - b. Saya bersedia menerima semua sanksi sesuai peraturan Universitas Multimedia Nusantara.

Pernyataan ini saya buat dengan sebenar-benarnya dan digunakan sebagaimana mestinya.

Tangerang, 30 Desember 2025

Mengetahui

Menyatakan



David Agustriawan, S.Kom., M.Sc.,
Ph.D.



Mitawati

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH MAHASISWA

Yang bertanda tangan di bawah ini:

Nama : Mitawati
NIM : 00000070345
Program Studi : Informatika
Jenjang : S1
Jenis Karya : Laporan Research & Technology Program

Menyatakan dengan sesungguhnya bahwa:

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya di repositori Knowledge Center, sehingga dapat diakses oleh Civitas Akademika/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial dan saya juga tidak akan mencabut kembali izin yang telah saya berikan dengan alasan apapun.
- ☐ Saya tidak bersedia karena dalam proses pengajuan untuk diterbitkan ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*)*.*.

Tangerang, 30 Desember 2025

Yang menyatakan



UNIVERSITAS
MULTIMEDIA
NUSANTARA

Mitawati

** Jika tidak bisa membuktikan LoA jurnal/HKI selama enam bulan ke depan, saya bersedia mengizinkan penuh karya ilmiah saya untuk diunggah ke KC UMN dan menjadi hak institusi UMN.

Halaman Persembahan / Motto

"A good name is to be more desired than great wealth, Favor is better than silver and gold."

Proverbs 22:1 (NASB)



KATA PENGANTAR

(Kata Pengantar dapat dikembangkan dan harus meliputi ucapan rasa syukur, tujuan pembuatan tugas akhir, ucapan terima kasih, dan harapan pada hasil Tugas Akhir ini.)

Mengucapkan terima kasih

1. Bapak Dr. Ir. Andrey Andoko, M.Sc., selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Bapak Arya Wicaksana, S.Kom., M.Eng.Sc., OCA, selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Bapak Januar Wahjudi, S.Kom., M.Sc., sebagai Pembimbing yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya tesis ini..
5. Bapak David Agustriawan, S.Kom., M.Sc., Ph.D., sebagai Pembimbing Lapangan yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya laporan MBKM Penelitian.
6. Orang Tua, keluarga dan teman-teman saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tesis ini.

Semoga karya ilmiah ini dapat memberikan dampak yang positif di dunia penelitian kanker terutama untuk klasifikasi subtype kanker ginjal.

Tangerang, 30 Desember 2025



Mitawati

PENDEKATAN MACHINE LEARNING MULTIKELAS UNTUK KLASIFIKASI SUBTIPE KANKER GINJAL MENGGUNAKAN DATA EKSPRESI MIRNA TCGA

Mitawati

ABSTRAK

kanker ginjal secara akurat menjadi tantangan penting dalam proses diagnosis dan penentuan terapi. Penelitian ini bertujuan untuk mengklasifikasikan subtipen kanker ginjal, yaitu *Kidney Chromophobe* (KICH), *Kidney Renal Clear Cell Carcinoma* (KIRC), dan *Kidney Renal Papillary Cell Carcinoma* (KIRP), menggunakan pendekatan *machine learning* berbasis data ekspresi *microRNA* (miRNA). Dataset yang digunakan berasal dari basis data *The Cancer Genome Atlas* (TCGA) dengan total 1.881 fitur miRNA dan 1.009 sampel. Tahapan penelitian meliputi pra-pemrosesan data, normalisasi, seleksi fitur menggunakan metode *Least Absolute Shrinkage and Selection Operator* (LASSO), serta klasifikasi menggunakan model *Long Short-Term Memory* (LSTM). Metode LASSO diterapkan untuk mengatasi permasalahan *curse of dimensionality* dengan mereduksi fitur miRNA menjadi subset yang lebih informatif. Evaluasi kinerja model dilakukan menggunakan pendekatan *stratified 5-fold cross validation* dengan metrik *accuracy*, *precision*, *recall*, *F1-score*, *Matthews Correlation Coefficient* (MCC), dan *confusion matrix*. Hasil seleksi fitur menunjukkan bahwa LASSO mereduksi 1.881 fitur miRNA menjadi 329 miRNA dengan koefisien non-zero. Pengujian pada tiga skenario jumlah fitur, yaitu 20, 35, dan 50 miRNA, menunjukkan bahwa penggunaan 35 miRNA terpilih memberikan keseimbangan optimal antara kompleksitas model dan kemampuan generalisasi dengan performa klasifikasi yang tinggi dan stabil. Penelitian ini menunjukkan bahwa kombinasi seleksi fitur LASSO dan model LSTM efektif untuk klasifikasi subtipen kanker ginjal berbasis ekspresi miRNA.

Kata kunci: kanker ginjal, klasifikasi multikelas, LASSO, LSTM, *microRNA* (miRNA), seleksi fitur

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

**A MULTICLASS MACHINE LEARNING APPROACH FOR KIDNEY
CANCER SUBTYPE CLASSIFICATION USING TCGA MIRNA
EXPRESSION DATA**

Mitawati

ABSTRACT

Kidney cancer is one of the most heterogeneous malignancies, making accurate identification of kidney cancer subtypes a critical challenge in diagnosis and treatment decision-making. This study aims to classify kidney cancer subtypes, namely Kidney Chromophobe (KICH), Kidney Renal Clear Cell Carcinoma (KIRC), and Kidney Renal Papillary Cell Carcinoma (KIRP), using a machine learning approach based on microRNA (miRNA) expression data. The dataset was obtained from The Cancer Genome Atlas (TCGA) database, consisting of 1,881 miRNA features and 1,009 samples. The research workflow includes data preprocessing, normalization, feature selection using the Least Absolute Shrinkage and Selection Operator (LASSO), and classification using a Long Short-Term Memory (LSTM) model. The LASSO method was employed to address the curse of dimensionality by reducing high-dimensional miRNA features into a more informative subset. Model performance was evaluated using stratified 5-fold cross validation with evaluation metrics including accuracy, precision, recall, F1-score, Matthews Correlation Coefficient (MCC), and confusion matrix. Feature selection results indicate that LASSO reduced the original 1,881 miRNA features to 329 miRNAs with non-zero coefficients. Experiments using three feature scenarios, namely 20, 35, and 50 miRNAs, demonstrate that the selection of 35 miRNAs provides an optimal balance between model complexity and generalization capability, yielding high and stable classification performance. These findings indicate that the combination of LASSO-based feature selection and the LSTM model is effective for classifying kidney cancer subtypes based on miRNA expression data.

Keywords: *feature selection, kidney cancer, LASSO, LSTM, microRNA (miRNA), multiclass classification (in alphabetical order)*

UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERNYATAAN ORISINALITAS	ii
HALAMAN PERNYATAAN PENGGUNAAN BANTUAN KECERDASAN ARTIFISIAL (AI)	iii
HALAMAN PERNYATAAN KEABSAHAN PERUSAHAAN	iv
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	v
HALAMAN PERSEMBAHAN/MOTO	vi
KATA PENGANTAR	vii
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xiv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Urgensi Penelitian	4
1.5 Luaran Penelitian	4
1.6 Manfaat Penelitian	4
BAB 2 TINJAUAN PUSTAKA	6
2.1 Kanker Ginjal	6
2.2 microRNA (miRNA)	7
2.3 Data Ekspresi miRNA dari TCGA	8
2.3.1 Karakteristik Data Ekspresi miRNA	9
2.3.2 Keunggulan Dataset TCGA	9
2.3.3 Tantangan Data Berdimensi Tinggi	10
2.4 Seleksi Fitur pada Data Genomik	11
2.4.1 <i>Least Absolute Shrinkage and Selection Operator</i> (LASSO)	12
2.5 Pembelajaran Mesin (<i>Machine Learning</i>)	12
2.5.1 Klasifikasi Multikelas	12
2.5.2 <i>Long Short-Term Memory</i> (LSTM)	13
2.5.3 Multi-Layer Perceptron	14
2.6 Evaluasi Kinerja Model Klasifikasi	15
2.6.1 <i>Accuracy</i>	16
2.6.2 <i>Confusion Matrix</i>	16
2.6.3 <i>Matthews Correlation Coefficient</i> (MCC)	17
BAB 3 METODE PENELITIAN	18
3.1 Pengumpulan Data	19
3.2 Pra-pemrosesan Data	20
3.3 Struktur Data untuk Klasifikasi	21
3.4 Pembentukan dan Pembagian Dataset	21
3.5 Seleksi Fitur Berbasis LASSO	22
3.6 Pembangunan Model	22
3.6.1 <i>Multi-Layer Perceptron</i> (MLP)	22
3.6.2 <i>Long Short-Term Memory</i> (LSTM)	23
3.7 Evaluasi dan Analisis Model	23

BAB 4	HASIL DAN PEMBAHASAN	25
4.1	Hasil Perancangan	25
4.1.1	Dataset	25
4.2	Pra-pemrosesan Data	27
4.3	Pelabelan Sampel	28
4.4	Hasil Seleksi Fitur	29
4.5	Hasil Klasifikasi	31
4.6	Analisis <i>Confusion Matrix</i>	32
4.7	Analisis <i>Learning Curve</i>	34
4.8	Visualisasi <i>t-SNE</i>	35
4.9	Analisis Hasil Keseluruhan Model	36
4.10	Diskusi Hasil Perancangan	38
BAB 5	SIMPULAN DAN SARAN	39
5.1	Simpulan	39
5.2	Saran	40
DAFTAR PUSTAKA		41



DAFTAR TABEL

Tabel 4.1	Cuplikan Dataset miRNA KICH (Raw)	26
Tabel 4.2	Cuplikan Dataset miRNA KIRC (Raw)	26
Tabel 4.3	Cuplikan Dataset miRNA KIRP (Raw)	26
Tabel 4.4	Cuplikan Dataset Akhir miRNA Kanker Ginjal	26
Tabel 4.5	Ukuran Dataset miRNA Kanker Ginjal	27
Tabel 4.6	Label Kelas Subtipe Kanker Ginjal	28
Tabel 4.7	Hasil Cross-Validation Model dengan 20 miRNA Terpilih .	29
Tabel 4.8	Hasil Cross-Validation Model dengan 35 miRNA Terpilih .	30
Tabel 4.9	Hasil Cross-Validation Model dengan 50 miRNA Terpilih .	30
Tabel 4.10	Top 35 miRNA Terpilih Hasil Seleksi Fitur	32
Tabel 4.11	Perbandingan Performa Model LSTM dengan Berbagai Jumlah Fitur miRNA Hasil Seleksi LASSO	37



DAFTAR GAMBAR

Gambar 2.1	Kerangka <i>Multi-Layer Perceptron</i> (MLP)	15
Gambar 3.1	Alur Penelitian	18
Gambar 4.1	<i>Confusion Matrix</i>	33
Gambar 4.2	<i>Learning Curve</i>	35
Gambar 4.3	Visualisasi <i>t-SNE</i>	36
Gambar 4.4	Perbandingan rata-rata akurasi semua model LSTM dengan Berbagai Jumlah Fitur miRNA Hasil Seleksi LASSO	37



DAFTAR LAMPIRAN

Lampiran 1	PRO-STEP-01 Cover Letter	44
Lampiran 2	PRO-STEP-02 Internship Card	45
Lampiran 3	PRO-STEP-03 Daily Task	46
Lampiran 4	PRO-STEP-04 Verification Form	68
Lampiran 5	PRO-STEP-05 Surat Penerimaan	69
Lampiran 6	Form Bimbingan	70
Lampiran 7	Hasil Pengecekan Similarity Turnitin	71
Lampiran 8	Formulir Penggunaan Perangkat Kecerdasan Artifisial (AI) . .	75
Lampiran 9	Research Paper	76
Lampiran 10	Surat Kerja Sama Penelitian Institusi	87

