



**EKSTRAKSI FITUR 1D DAN 2D BERBASIS PROTEIN DESKRIPTOR
AMINO ACID COMPOSITION (AAC) UNTUK KLASIFIKASI
ASETILASI PADA PROTEIN LISIN DENGAN CONVOLUTIONAL
NEURAL NETWORK**

Skripsi

**Untuk Memenuhi Persyaratan Dalam
Menyelesaikan Sarjana Strata S1-Ilmu Komputer**

Oleh

LAILA ADAWIYAH

NIM. 1811016120011

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
JULI 2023**



**EKSTRAKSI FITUR 1D DAN 2D BERBASIS PROTEIN DESKRIPTOR
AMINO ACID COMPOSITION (AAC) UNTUK KLASIFIKASI
ASETILASI PADA PROTEIN LISIN DENGAN CONVOLUTIONAL
NEURAL NETWORK**

Skripsi

**Untuk Memenuhi Persyaratan Dalam
Menyelesaikan Sarjana Strata S1-Ilmu Komputer**

**Oleh
LAILA ADAWIYAH
NIM. 1811016120011**

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
JULI 2023**

SKRIPSI

EKSTRAKSI FITUR 1D DAN 2D BERBASIS PROTEIN DESKRIPTOR AMINO ACID COMPOSITION (AAC) UNTUK KLASIFIKASI ASETILASI PADA PROTEIN LISIN DENGAN *CONVOLUTIONAL NEURAL NETWORK*

Oleh :

LAILA ADAWIYAH
1811016120011

Telah dipertahankan di depan Dosen Penguji pada tanggal 21 Juli 2023:

Susunan Dosen Penguji :

Pembimbing I



Mohammad Reza Faisal, S.T M.T PhD
NIP. 197612202008121001

Dosen Penguji I



Dwi Kartini, S.Kom., M.Kom
NIP. 198704212012122003

Pembimbing II



Triando H. Saragih, S.Kom., M.Kom
NIP. 199308242019031012

Dosen Penguji II



Rudy Herteno, S.Kom M.Kom
NIP. 198809252022031003



Banjarbaru, 15 Juli 2023

Kerua Program Studi Ilmu Komputer

Iwan Budiman, S.T., M.Kom
NIP. 197703252008121001

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan disuatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Banjarbaru, 21 Juli 2023



Laila Adawiyah

NIM. 1811016120011

ABSTRAK

EKSTRAKSI FITUR 1D DAN 2D BERBASIS PROTEIN DESKRIPTOR AMINO ACID COMPOSITION (AAC) UNTUK KLASIFIKASI ASETILASI PADA PROTEIN LISIN DENGAN *CONVOLUTIONAL NEURAL NETWORK*

(Oleh: Laila Adawiyah; Pembimbing: Mohammad Reza Faisal, S.T M.T PhD dan Triando H. Saragih, S.Kom ; 2023; 61 halaman)

Post-Translational Modification (PTM) atau yang disebut dengan modifikasi pasca-translasi merupakan mekanisme penting yang dibutuhkan dalam komponen protein, Modifikasi diperlukan untuk pertumbuhan sel, regulasi transkripsi, serta metabolisme yang dibutuhkan untuk kehidupan sehari-hari, Oleh karena itu perlu adanya metode komputasi (*in silico*) karena dapat mempercepat proses pengidentifikasian. Penelitian yang telah ada mengidentifikasi asetilasi lisin dengan menggunakan metode komputasi, salah satunya yaitu klasifikasi. Pada penelitian ini klasifikasi protein dimulai dengan mengekstraksi sekuens protein menjadi fitur numerik dengan deskriptor protein yang dalam penelitian ini menggunakan *Amino Acid Composition* selanjutnya dilakukan segmentasi untuk menambah jumlah fitur sehingga kemudian data dirubah menjadi 1D dan 2D. Pada akhirnya klasifikasi protein dilakukan dengan metode convolutional neural network yaitu 1D CNN dan 2D CNN. Penelitian ini memberikan hasil kinerja yang terbaik pada penggunaan metode 2D CNN dibandingkan 1D CNN serta peningkatan akurasi yang didapatkan dari penggunaan segmentasi protein.

Kata Kunci : *Post-Translational Modification*, asetilasi lisin, segmentasi protein, *amino acid composition*, 1D CNN, 2D CNN, convolutional neural network

ABSTRACT

1D AND 2D FEATURE EXTRACTION BASED ON PROTEIN DESCRIPTOR AMINO ACID COMPOSITION (AAC) FOR ACETYLATION CLASSIFICATION OF PROTEIN LYSINE USING CONVOLUTIONAL NEURAL NETWORK

(By: Laila Adawiyah; Supervisor: Mohammad Reza Faisal, S.T M.T PhD and Triando H. Saragih, S.Kom ; 2023; 61 pages)

Post-Translational Modification (PTM) or what is called post-translational modification is an important mechanism needed in protein components. Modification is needed for cell growth, regulation of transcription, and metabolism needed for everyday life. Therefore, there is a need for computational methods (in silico) because it can speed up the identification process. Existing research has identified lysine acetylation using computational methods, one of which is classification. In this study, protein classification was started by extracting protein sequences into numerical features with protein descriptors which in this study used Amino Acid Composition which was then segmented to increase the number of features so as to improve the classification performance of the protein descriptors, then the data was used as a matrix. Furthermore, protein classification was carried out using the convolutional neural network method, namely 1D CNN and 2D CNN. This study provides the best performance results using the 2D CNN method compared to 1D CNN and the increased accuracy obtained from the use of protein segmentation.

Keywords: *Post-Translational Modification, lysine acetylation, protein segmentation, amino acid composition, 1D CNN, 2D CNN, convolutional neural network*

PRAKATA

Assalamualaikum Warahmatullahi Wabarakatuh

Puji syukur penulis panjatkan ke hadirat Allah SWT karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul “Ekstraksi fitur 1D dan 2D berbasis protein deskriptor *Amino Acid Composition* (AAC) untuk klasifikasi asetilasi pada protein lisin dengan *Convolutional Neural Network*” untuk memenuhi syarat dalam menyelesaikan Pendidikan S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Tak lupa penulis panjatkan sholawat dan salam ke hadirat Rasulullah Muhammad SAW beserta para sahabat, keluarga, dan pengikut beliau hingga yaumul qiamah.

Pada lembar ini penulis ingin menyampaikan ucapan terimakasih kepada pihak- pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan skripsi ini, adapun yang dimaksud adalah sebagai berikut:

1. Keluarga terutama orang tua yaitu Ibu Wahidah , Bapak Adriansyah dan Ibu Norliani dan yang selalu memberikan bantuan, semangat, doa, dan dukungan dalam proses menyelesaikan skripsi ini.
2. Bapak M. Reza Faisal., S.T., M.T., Ph.D selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
3. Bapak Triando H. Saragih, S.Kom., M.Kom selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
4. Bapak Irwan Budiman S.T., M.Kom selaku Ketua Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
5. Bapak Rudy Herteno, S.Kom M.Kom selaku dosen pembimbing akademik dan seluruh Dosen dan staf Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
6. Untuk teman-teman seperjuangan nelisa, munawarrah, sora, fachlia, fenny, fathmah, sadih dan erny dan teman-teman lainnya yang selalu memberikan dorongan agar terselesaikannya penulisan skripsi ini

7. Teman-teman dan kerabat khususnya teman-teman Ilmu Komputer 2018, yang sudah ikut serta dalam memberikan semangat, memberikan doa serta membantu mengerjakan penelitian ini.
8. Semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian skripsi ini.

Penulis berharap skripsi ini dapat bermanfaat memberikan sumbangan pikiran dan ilmu pengetahuan baik bagi penulis sendiri maupun bagi pembaca. Penulis menyadari bahwa skripsi ini jauh dari kata sempurna, oleh karena itu penulis mengharapkan kritik dan saran yang membangun guna kesempurnaan skripsi ini untuk perbaikan dikemudian hari.

Banjarbaru, 21 Juli 2023



Laila Adawiyah

NIM. 1811016120011

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	iii
PERNYATAAN.....	iv
ABSTRAK.....	v
ABSTRACT.....	vi
PRAKATA.....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xii
DAFTAR LAMPIRAN.....	xiv

BAB I PENDAHULUAN

1.1 Latar Belakang	1
1.2 Perumusan Masalah.....	2
1.3 Batasan Masalah.....	3
1.4 Tujuan.....	3
1.5 Manfaat Penelitian.....	3

BAB II TINJAUAN PUSTAKA

2.1 Kajian Terdahulu	4
2.2 Keaslian Penelitian	5
2.3 Landasan Teori	8
2.3.1. Protein	8
2.3.2. Post-Translational Modification.....	13
2.3.3. Asetilasi.....	14
2.3.4. Dataset Protein	15
2.3.5. <i>Sequence Segmentation</i>	15
2.3.6. Feature Extraction	17
2.3.7. <i>Deep learning</i>	19
2.3.8. <i>Convolution Neural Network</i>	20
2.3.9. Evaluasi Matrik.....	29

BAB III METODE PENELITIAN

3.1	Alat Penelitian	32
3.2	Bahan Penelitian.....	32
3.3	Variabel Penelitian	33
3.3	Prosedur Penelitian.....	33

BAB IV HASIL DAN PEMBAHASAN

4.1	Hasil.....	35
4.1.1.	Pengumpulan Data	35
4.1.2.	Ekstraksi Fitur.....	36
4.1.3.	<i>Amino Acid Composition</i>	43
4.1.4.	Pembagian Data	44
4.1.5.	Klasifikasi	45
4.2	Evaluasi	50
4.3	Pembahasan	53

BAB V PENUTUP

5.1	Kesimpulan.....	60
5.1	Saran	60

DAFTAR PUSTAKA

RIWAYAT HIDUP

DAFTAR TABEL

Tabel	Halaman
Tabel 1 Keaslian penelitian.....	5
Tabel 2 Rancangan penelitian	8
Tabel 3 Daftar descriptor protein ProtR.....	19
Tabel 4 Confusion Matrix	30
Tabel 5 Data sekuens protein lisin	32
Tabel 6 Jumlah data protein lisin	35
Tabel 7 data protein lisin dari kelas positif dan negatif	35
Tabel 8 Pembagian panjang sekuen protein k=3	41
Tabel 9 Pembagian panjang sekuen protein k=4	41
Tabel 10 Pembagian panjang sekuen protein k=5	42
Tabel 11 Segmentasi AAC.....	43
Tabel 12 Contoh penerapan AAC pada dengan fitur 120.....	44
Tabel 13 Contoh penerapan AAC pada dengan fitur 160	44
Tabel 14 Contoh penerapan AAC pada dengan fitur 200.....	44
Tabel 15 pembagian data train dan test.....	45
Tabel 16 Model 1D CNN.....	46
Tabel 17 Model 2D CNN.....	46
Tabel 18 Dimensi setelah direshape kedalam bentuk 1D dan 2D.....	47
Tabel 19. Parameter 1D CNN.	48
Tabel 20. Parameter 2D CNN	49
Tabel 21 Hasil pengujian klasifikasi protein lisin.....	50
Tabel 22 Confusion Matrix 1D CNN fitur 120.....	54
Tabel 23 Confusion Matrix 2D CNN fitur 120.....	55
Tabel 24 Confusion Matrix 1D CNN fitur 160.....	55
Tabel 25 Confusion Matrix 2D CNN fitur 160.....	56
Tabel 26 Confusion Matrix 1D CNN fitur 200.....	56
Tabel 27 Confusion Matrix 2D CNN fitur 200.....	56

DAFTAR GAMBAR

Gambar	Halaman
Gambar 1 Ilustrasi terjadinya protein.....	8
Gambar 2 Modifikasi protein pasca-translasi	14
Gambar 3 Penambahan gugus asetil pada protein lisin.....	15
Gambar 4 Ilustrasi dari adjacent	16
Gambar 5 Ilustrasi overlapped segment	16
Gambar 6 (a) Simple Neural Network; (b) Deep Learning.....	20
Gambar 7 Contoh Input pada Convolution Layer.....	21
Gambar 8 Ilustrasi Convolution Layer.....	22
Gambar 9 Contoh Padding	22
gambar 10 Ilustrasi convolution layer dengan stride	23
Gambar 11 Arsitektur convolutional 1D.....	25
Gambar 12 Arsitektur Convolutional 2D.....	25
Gambar 13 Max Pooling dan Average Pooling.	26
Gambar 14 Grafik fungsi aktivasi pada ReLU layer.....	27
Gambar 15 Ilustrasi dari flatten.....	28
Gambar 16 <i>Fully Connected Layer</i>	29
Gambar 17 Alur penelitian.....	33
Gambar 18 Segmentasi dengan parameter k=3.....	37
Gambar 19 Segmentasi dengan parameter k=4.....	38
Gambar 20 Segmentasi dengan k=5.....	39
Gambar 21 pembagian data.....	45
Gambar 22 Model plot 1D CNN	50
Gambar 23 Model plot 2D CNN	50
Gambar 24 Grafik akurasi.....	51
Gambar 25 Grafik Sensitivity.	52
Gambar 26 Grafik Specificity	52
Gambar 27 Grafik MCC.....	53
Gambar 28 hasil klasifikasi protein lisin 1D CNN	57
Gambar 29 Hasil klasifikasi 2D CNN.....	58

Gambar 30 Grafik akurasi 1D CNN.....	58
Gambar 31 Grafik akurasi 2D CNN.....	59

DAFTAR LAMPIRAN

Lampiran

Lampiran 1 *Source code dataset asetilasi negatif*

Lampiran 2 *Source code dataset asetilasi positif.*

Lampiran 3 *Source code ekstraksi fitur AAC*

Lampiran 4 *Source code adjacent segment*

Lampiran 5 *Source code overlapped segment*

Lampiran 6 Source code pembagian data

Lampiran 7 Source code model 1D CNN

Lampiran 8 Model 1D CNN.

Lampiran 9 Train model 1D CNN

Lampiran 10 Source code Model 2D CNN

Lampiran 11 Model 2D CNN

Lampiran 12 Train model 2D CNN

Lampiran 13 Evaluasi model 1D CNN

Lampiran 14 Evaluasi model 2D CNN