

**KLASIFIKASI ASETILASI DAN METILASI PADA DNA HISTON RAGI
MENGUNAKAN METODE *CONVOLUTIONAL NEURAL NETWORK***

(Skripsi)

OLEH

MUHAMAD SEPRYAN ASTRAYESA

1817051028



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023**

**KLASIFIKASI ASETILASI DAN METILASI PADA DNA HISTON RAGI
MENGUNAKAN METODE *CONVOLUTIONAL NEURAL NETWORK***

Oleh

MUHAMAD SEPRYAN ASTRAYESA

Skripsi

**Sebagai Salah Satu Syarat untuk Mendapat Gelar
SARJANA KOMPUTER**

Pada

**Jurusan Ilmu Komputer
Fakultas Matematika dan Ilmu Pengetahuan Alam**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023**

ABSTRAK

KLASIFIKASI ASETILASI DAN METILASI PADA DNA HISTON RAGI MENGUNAKAN METODE *CONVOLUTIONAL NEURAL NETWORK*

Oleh

MUHAMAD SEPRYAN ASTRAYESA

Penelitian ini menyajikan studi eksperimental tentang klasifikasi DNA ragi menggunakan Convolutional Neural Network (CNN) dengan pengkodean *one-hot* pada data DNA. Tujuannya adalah untuk menyelidiki kinerja model CNN dalam mengklasifikasikan urutan DNA ragi secara akurat. Tiga percobaan dilakukan, masing-masing dengan variasi arsitektur model dan teknik pelatihan. Hasilnya dievaluasi berdasarkan metrik akurasi, daya ingat, dan presisi. Di antara ketiga eksperimen tersebut, Eksperimen 3 menunjukkan hasil yang paling menjanjikan. Model yang dilatih mencapai akurasi rata-rata 84,10%, recall 83,27%, dan presisi 86,87% selama fase pelatihan. Namun pada tahap pengujian, model menunjukkan penurunan performa yang sedang, menghasilkan akurasi rata-rata 65,32%, recall 65,13%, dan presisi 69,51%. Meskipun terlihat penurunan performa selama pengujian, Eksperimen 3 mengungguli eksperimen lain dalam hal akurasi klasifikasi. Temuan menunjukkan potensi model CNN untuk secara akurat mengklasifikasikan urutan DNA ragi. Penelitian lebih lanjut diperlukan untuk mengatasi kesenjangan kinerja antara fase pelatihan dan pengujian, dengan fokus pada pengoptimalan arsitektur model dan teknik pelatihan untuk meningkatkan kemampuan generalisasi sistem klasifikasi.

Kata Kunci: Klasifikasi DNA, Jaringan Saraf Konvolusi, pengkodean *one-hot*, metilasi, asetilasi.

ABSTRAK

KLASIFIKASI ASETILASI DAN METILASI PADA DNA HISTON RAGI MENGUNAKAN METODE *CONVOLUTIONAL NEURAL NETWORK*

Oleh

MUHAMAD SEPRYAN ASTRAYESA

This study presents an experimental investigation into the classification of yeast DNA sequences using Convolutional Neural Networks (CNNs) with one-hot encoding applied to the DNA data. The primary objective is to evaluate the performance of CNN models in accurately classifying yeast DNA sequences. Three experiments were conducted, each employing variations in model architecture and training techniques. The outcomes were assessed based on accuracy, recall, and precision metrics. Among the three, Experiment 3 yielded the most promising results. During the training phase, the model achieved an average accuracy of 84.10%, recall of 83.27%, and precision of 86.87%. However, in the testing phase, the model exhibited a moderate performance decline, with an average accuracy of 65.32%, recall of 65.13%, and precision of 69.51%. Despite the observed performance drop during testing, Experiment 3 outperformed the other configurations in terms of classification accuracy. These findings highlight the potential of CNN-based models to effectively classify yeast DNA sequences. Further research is necessary to bridge the performance gap between training and testing phases, focusing on model architecture optimization and training strategies to enhance the generalization capabilities of the classification system.

Keywords: DNA classification, Convolutional Neural Network, one-hot encoding, methylation, acetylation.

Judul Skripsi : **KLASIFIKASI ASETILASI DAN METILASI PADA
DNA HISTON RAGI MENGGUNAKAN METODE
CONVOLUTIONAL NEURAL NETWORK**

Nama Mahasiswa : **Muhamad Sepryan Astrayesa**

NPM : 1817051028

Program Studi : **S1 Ilmu Komputer**

Jurusan : **Ilmu Komputer**

Fakultas : **Matematika dan Ilmu Pengetahuan Alam**



Favorisen R. Lumbanraja, Ph.D.

NIP. 19830110 200812 1 002

2. Ketua Jurusan Ilmu Komputer

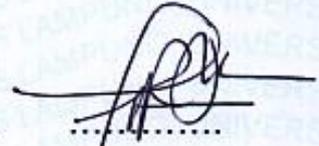
Dwi Sakethi, S.Si., M.Kom.

NIP. 19680611 199801 1 001

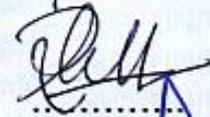
MENGESAHKAN

1. Tim Penguji

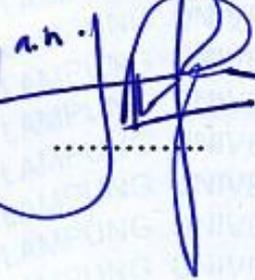
Ketua : Favorisen R. Lumbanraja, Ph.D.



Penguji I
Penguji Pembahas : M. Reza Faisal, S.T., M.T., Ph.D.



Penguji II
Penguji Pembahas : Dr. Ir. Kurnia Muludi, M.S.Sc.



2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam




Dr. Eng. Heri Satria, S.Si., M.Si
NIP 197110012005011002

Tanggal Lulus Ujian Skripsi: 31 Maret 2023

PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Muhamad Sepryan Astrayesa

NPM : 1817051028

Dengan ini menyatakan bahwa skripsi saya yang berjudul "KLASIFIKASI ASETILASI DAN METILASI PADA DNA HISTON RAGI MENGGUNAKAN METODE CONVOLUTIONAL NEURAL NETWORK" merupakan karya saya sendiri dan bukan karya orang lain. Semua tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Apabila di kemudian hari terbukti skripsi saya merupakan hasil penjiplakan atau dibuat orang lain, maka bersedia menerima sanksi sesuai dengan peraturan yang berlaku.

Bandar Lampung, 18 Juni 2025



Muhamad Sepryan Astrayesa

NPM. 1817051028

RIWAYAT HIDUP



Penulis memulai pendidikan formal di SD Al-Azhar 1 Bandar Lampung dan selesai pada tahun 2013. Kemudian melanjutkan pendidikan menengah pertama di MTsN 2 Bandar Lampung yang diselesaikan pada tahun 2016, lalu melanjutkan ke pendidikan menengah atas di SMAN 9 Bandar Lampung yang diselesaikan pada tahun 2018.

Pada tahun 2018 penulis terdaftar sebagai mahasiswa Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur SNMPTN. Selama menjadi mahasiswa, penulis melakukan beberapa kegiatan antara lain.

1. Badan Khusus Himpunan Mahasiswa Jurusan Ilmu Komputer periode 2018/2019.
2. Menjadi Asisten Dosen Jurusan Ilmu Komputer tahun 2019 hingga 2021.
3. Melaksanakan Kerja Pratik pada tahun 2021 di PT. Bank Lampung.
4. Melaksanakan Kuliah Kerja Nyata (KKN) pada tahun ajaran 2021/2022 di Desa Pesawahan, Teluk Betung Selatan, Bandar Lampung, Lampung.
5. Memperoleh sertifikasi Junior Web Developer (JWD) yang oleh Badan Nasional Sertifikasi Profesi (BNSP) pada tahun 2021.
6. Memperoleh sertifikasi Junior Web Developer (JWD) yang diselenggarakan oleh Badan Nasional Sertifikasi Profesi (BNSP) pada tahun 2021.

MOTTO

“Maka nikmat Tuhan kamu yang manakah yang kamu dustakan?”

(QS. Al-Kahfi:13)

“Jika Kamu tidak sanggup menahan lelahnya belajar maka kamu harus sanggup menahan perihnya kebodohan”

(Imam Syafi’i)

“Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya.”

(QS. Al-Baqarah: 286)

PERSEMBAHAN

Alhamdulillahirabbilalamin

Puji syukur kepada Allah SWT yang telah memberikan rahmat dan hidayah-Nya sehingga dapat menyelesaikan penulisan skripsi ini. Shalawat dan salam saya sanjungkan kepada Nabi Muhammad SAW.

Aku persembahkan karya ini kepada.

Ayahanda Samiko dan Ibunda Sri Dwi Arisanti

Sebagai tanda terima kasihku kepada bapak dan ibu yang terkasih dan yang tersayang. Yang senantiasa memberikan yang terbaik, dan melantunkan doa yang selalu menyertaiku. Terima kasih atas semua pengorbanan, perjuangan, dan doa kalian yang tiada hentinya.

Adik Almira Apriyani Astrayesa

Terima kasih telah memberikan semangat, dukungan, dan doa.

Seluruh Keluarga Besar, Sahabat, dan Teman-teman

Terima kasih telah memberikan semangat dan bantuan.

Dosen Pembimbing dan Pembahas

Yang senantiasa membimbing, mengarahkan dan memberi motivasi sejak awal hingga terselesaikannya skripsi ini.

Almamater, Universitas Lampung

SANWACANA

Puji syukur kehadiran Allah SWT, karena telah memberikan rahmat dan hidayah-Nya kepada saya sehingga saya dapat menyelesaikan skripsi dengan judul “Klasifikasi Asetilasi Dan Metilasi Pada DNA Histon Ragi Menggunakan Metode Convolutional Neural Network” Saya berharap skripsi ini dapat menambah pengetahuan bagi pembaca tentang klasifikasi DNA menggunakan metode Deep Learning CNN.

Selama proses penulisan skripsi ini tidak terlepas dari dukungan banyak pihak yang telah membimbing, membantu dan memberi semangat kepada saya, sehingga pada kesempatan ini saya ingin menyampaikan ungkapan terima kasih kepada.

1. Ayah, ibu, dan adik saya yang selalu mendoakan, menyemangati, membiayai serta mendukung saya baik secara moral maupun material. Terima kasih atas doa yang kalian berikan untuk keberhasilan dan kesuksesan saya.
2. Bapak Favorisen R. Lumbanraja, Ph.D. sebagai pembimbing utama serta pembimbing akademik telah membimbing saya, memberikan kritik dan saran, serta membina selama masa studi sampai dengan penulisan skripsi, sehingga studi dan skripsi ini dapat diselesaikan.
3. Ibu Anie Rose Irawati, S.T., M.Cs, Bapak Ardiansyah, S.Kom., M.Kom., dan Didik Kurniawan, S.Si., MT. yang terus mengingatkan, menyemangati, dan memberikan motivasi agar skripsi ini dapat diselesaikan.
4. Bapak Dr. Ir. Kurnia Muludi, M.S.Sc. dan Bapak M. Reza Faisal, S.T., M.T., Ph.D. sebagai pembahas yang telah memberikan masukan, ide, kritik sehingga penulisan skripsi ini dapat diselesaikan.

5. Bapak dan Ibu Dosen Jurusan Ilmu Komputer FMIPA Universitas Lampung yang telah memberikan ilmu selama penulis menempuh pendidikan.
6. Teman–teman satu angkatan Ilmu Komputer 2018 di Universitas Lampung yang telah berjuang bersama selama menempuh pendidikan.
7. Semua pihak yang terlibat selama masa perkuliahan dan penyusunan skripsi yang tidak bisa saya sebutkan satu per satu.

Dalam proses penyusunan skripsi ini tentunya terdapat kekurangan yang disebabkan oleh keterbatasan pengalaman dan pengetahuan penulis. Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna. Semoga skripsi ini membawa manfaat dan keberkahan semua pembaca.

Bandar Lampung, 18 Juni 2025

Muhamad Sepryan Astrayesa
NPM. 1817051028

DAFTAR ISI

	Halaman
DAFTAR ISI.....	xiii
DAFTAR TABEL.....	xv
DAFTAR GAMBAR	xvi
DAFTAR PSEUDOCODE	xvii
I. PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah.....	3
1.4. Tujuan.....	3
1.5. Manfaat.....	3
II. TINJAUAN PUSTAKA	4
2.1. Penelitian Terdahulu	4
2.2. DNA	6
2.3. Histon	7
2.4. One-hot <i>Encoding</i>	8
2.5. Artificial Neural Networks.....	10
2.6. Convolutional Neural Network	11
2.7. Confusion Matrix	12
III. METODOLOGI PENELITIAN.....	14
3.2. Tempat Dan Waktu Penelitian	14

	xiv
3.3. Data Dan Alat.....	16
3.4. Alur Kerja Penelitian.....	20
IV. Hasil dan Pembahasan.....	22
4.1. Pengumpulan Data	22
4.2. Transformasi data menggunakan <i>One-hot Encoding</i>	23
4.3. Pembagian data menjadi <i>training</i> dan <i>testing</i>	24
4.4. Arsitektur Model Convolutional Neural Network (CNN)	24
4.5. Hasil klasifikasi dari model <i>Convolutional Neural Network (CNN)</i>	31
4.6. Pembahasan.....	39
4.7. Perbandingan dengan penelitian sebelumnya	45
V. SIMPULAN DAN SARAN	47
5.1. Simpulan	47
5.2. Saran.....	48
DAFTAR PUSTAKA	49

DAFTAR TABEL

Tabel	Halaman
1. Penelitian terdahulu terkait klasifikasi DNA	4
2. Matrix One-hot <i>Encoding</i>	9
3. Hasil One-hot <i>Encoding</i> dari DNA “CAGGT”	9
4. Confusion matrix pada klasifikasi dua kelas.....	13
5. Jadwal Penelitian.....	15
6. Data DNA ragi (Pokholok et al., 2005).....	17
7. Hasil pelatihan Eksperimen 1.....	31
8. Hasil pelatihan Eksperimen 2.....	32
9. Hasil pelatihan Eksperimen 3.....	34
10. Hasil pengujian Eksperimen 1	35
11. Hasil pengujian <i>Eksperimen 2</i>	36
12. Hasil pengujian Eksperimen 3.	37
13. Perbandingan antara eksperimen 1, 2, dan 3.....	44
14. Perbandingan dengan penelitian sebelumnya	45

DAFTAR GAMBAR

Gambar	Halaman
1. Ilustrasi DNA (U.S. National Library of Medicine).	7
2. Ilustrasi antara histon dan DNA (Elmallah & Micheau, 2019).	8
3. Ilustrasi dari Artificial Neural Network (Nurfikri, 2020).	10
4. Ilustrasi perhitungan pada ANN (Schils, 2020).	11
5. Contoh arsitektur dari <i>Convolutional Neural Network</i> (Matlab).	11
6. Perbandingan total setiap data.	18
7. Perbandingan data positif dan negatif pada setiap <i>dataset</i>	18
8. Alur Penelitian terkait klasifikasi DNA ragi dengan metode CNN.	20
9. Contoh format data yang telah diunduh.	22
10. Arsitektur model pada Eksperimen 1.	25
11. Arsitektur model pada Eksperimen 2.	27
12. Arsitektur model pada Eksperimen 3.	29
13. Hasil pelatihan pada Eksperimen 1.	32
14. Hasil pelatihan pada Eksperimen 2.	33
15. Hasil pelatihan pada Eksperimen 3.	34
16. Hasil pengujian pada Eksperimen 1.	36
17. Hasil pengujian pada Eksperimen 2.	37
18. Hasil pengujian pada Eksperimen 3.	39
19. Perbandingan akurasi latih dan uji pada eksperimen 1.	40
20. Perbandingan akurasi latih dan uji pada eksperimen 2.	41
21. Perbandingan akurasi latih dan uji pada eksperimen 3.	42
22. Perbandingan waktu latih.	43
23. Perbandingan hasil uji Eksperimen 1, 2, dan 3.	44

DAFTAR PSEUDOCODE

Pseudocode	Halaman
1. Pengumpulan data.....	23
2. <i>One-hot encoding</i>	24
3. Pembagian data latih dan data uji.	24
4. Arsitektur model pada eksperimen 1.....	26
5. Arsitektur model pada eksperimen 2.....	28
6. Arsitektur model pada eksperimen 3.....	30

I. PENDAHULUAN

1.1. Latar Belakang

Histon merupakan protein yang bertanggung jawab untuk mengemas DNA menjadi nukleosom, yang merupakan unit dasar kromatin. Histon juga berperan dalam proses regulasi gen. Histon dapat dimodifikasi dengan berbagai senyawa kimia. Beberapa jenis modifikasi yang dapat terjadi pada histon adalah asetilasi dan metilasi. Modifikasi dapat mengubah struktur histon yang dapat mempengaruhi proses transkripsi DNA.

DNA pada nukleus akan tersimpan dalam bentuk kromosom. Kromosom merupakan kumpulan kromatin yang terdiri dari banyak nukleosom. Histon merupakan protein yang memberikan dukungan struktural pada kromosom. Terdapat 5 jenis histon yaitu H1, H2A, H2B, H3 dan H4. Pada nukleus DNA akan tersimpan dalam keadaan terikat dengan *histone octamer*. *Histone octamer* dibentuk dari 8 buah *histone*, terdapat 2 buah untuk setiap jenis H2A, H2B, H3, dan H4.

DNA merupakan struktur dasar pembentuk protein. Pada prosesnya akan terjadi proses transkripsi yang mengubah DNA menjadi RNA kemudian menjadi protein. Dalam proses ini kadang terjadi modifikasi pada asam amino. Contoh dari modifikasi yang dapat terjadi adalah metilasi dan asetilasi. Metilasi merupakan proses penambahan gugus metil ke dalam asam amino. Asetilasi merupakan modifikasi dengan menambahkan gugus asetil pada asam amino.

Pemahaman mengenai gen dan ekspresinya merupakan hal yang penting untuk

memahami bagaimana suatu organisasi hidup (Higashihara et al., 2008). Dengan memahami karakteristik yang dimiliki DNA, kita dapat memahami aktivitas biologi secara lebih menyeluruh. Dengan berkembangnya teknologi pembacaan DNA biaya membaca DNA menjadi lebih murah (Nguyen et al., 2016). Dengan mudahnya pembacaan DNA ini mengakibatkan data dari untai DNA juga terus meningkat, seperti pada GenBank yang diperkirakan telah mencapai 220 juta sekuens pada bulan Juni 2021 (GenBank dan WGS Statistics). Data yang banyak ini akan sangat bermanfaat jika bisa digunakan untuk membantu dalam memahami lebih dalam mengenai DNA.

Sebelumnya telah dilakukan penelitian mengenai klasifikasi DNA. Pada tahun 2008 Higashihara et al. Pada penelitian tersebut diklasifikasikan data DNA dengan algoritme SVM. Sebelum diklasifikasikan fitur pada datanya telah diseleksi dan di-*ranking* dengan algoritme *Random Forest*. Kemudian pada tahun 2016 Nguyen et al., melakukan penelitian pengklasifikasian data DNA dengan menggunakan algoritma seq-CNN. Penelitian terbaru dilakukan oleh Mahmoud dan guo pada tahun 2021 dengan menggunakan *Multilayer Perceptron* dan *Pseudoinverse Learning Algorithm for Autoencoder*. Penelitian tersebut dilakukan dengan tahapan mengubah data DNA menjadi gambar kemudian diklasifikasikan dengan gabungan beberapa model *pretrain deep learning*.

Penelitian ini akan melakukan klasifikasi biner terhadap DNA ragi. DNA ragi akan diklasifikasikan dengan menentukan apakah terdapat histon, terjadi metilasi, atau terjadi asetilasi. Algoritma yang digunakan pada penelitian ini adalah *convolutional neural network*(CNN). DNA akan ditransformasi menggunakan *One-hot encoding* sebelum dijadikan *input* pada model CNN. Penggunaan CNN dan *One-hot encoding* bertujuan untuk melakukan eksperimen seberapa baik pengklasifikasian secara spasial pada data DNA.

1.2. Rumusan Masalah

Berdasarkan pemaparan latar belakang tersebut, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Mengimplementasikan CNN dalam mengklasifikasikan data DNA dan mengevaluasi hasilnya.
2. Membandingkan antara performa algoritme CNN dan algoritme klasifikasi DNA lainnya.

1.3. Batasan Masalah

Batasan masalah dalam penelitian ini yaitu

1. Data yang akan digunakan adalah Peta Genom Nukleus Asetilasi dan Metilasi pada Ragi yang didapat dari Pokholok et al. (2005).
2. Algoritma yang akan digunakan adalah *Convolutional Neural Network* (CNN).

1.4. Tujuan

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mengetahui hasil klasifikasi dari algoritme CNN dalam mengklasifikasi untai DNA.
2. Menentukan *hyperparameter* yang sesuai untuk algoritme CNN dalam memprediksi untai DNA
3. Membandingkan performa algoritme CNN dengan algoritme klasifikasi DNA lainnya.

1.5. Manfaat

Manfaat penelitian ini adalah sebagai berikut:

1. Mendapatkan perbandingan antara klasifikasi dengan CNN dengan penelitian terdahulu.
2. Penelitian ini dapat dijadikan informasi untuk penelitian lanjutan mengenai Klasifikasi Asetilasi dan Metilasi Pada Ragi Dengan menggunakan CNN.

II. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Tabel 1. Penelitian terdahulu terkait klasifikasi DNA

No.	Penelitian	Metode	Data	Hasil
1	<i>Application of a Feature Selection Method to Nucleosome Data: Accuracy Improvement and Comparison with Other Methods (Higashihara et al., 2008)</i>	Pengklasifikasian DNA ragi dengan fitur <i>ranking</i> dan fitur seleksi menggunakan algoritma klasifikasi SVM.	Menggunakan <i>dataset</i> H3, H4, H3K9ac, H3K14ac, H4ac, H3K4me1, H3K4me2, H3K4me3, H3K35me3, H3K79me3 dari penelitian Pokholok et al., (2005)	Menyeleksi fitur penting berhasil meningkatkan akurasi. Rerata akurasi yang didapatkan 75.73%
2	<i>DNA Sequence Classification by Convolutional Neural Network (Nguyen et al., 2016)</i>	Menggunakan algoritme <i>seq-CNN</i> untuk memprediksi terjadinya asetilasi dan metilasi		Rata-rata akurasi yang didapat: 79.4%
3	<i>DNA sequence classification based on MLP with PILAE algorithm (Mahmoud & Guo, 2021)</i>	Mengubah untai DNA menjadi gambar dan diklasifikasikan dengan menggabungkan beberapa <i>pre-train</i> model.	Menggunakan <i>dataset</i> H3, H4, H3K9ac, H3K14ac, dan H4ac dari penelitian Pokholok et al., (2005)	Rata-rata akurasi yang didapat: 98%

Tabel 1 menunjukkan 3 penelitian terdahulu mengenai klasifikasi DNA ragi. Pada penelitian Higashihara et al., (2008) yang bertujuan mengetahui perbedaan

antara penggunaan seleksi fitur dan tidak pada pengklasifikasian DNA. Pada penelitian Pokholok et al., (2005) hanya dilakukan pengklasifikasian tanpa melakukan seleksi fitur dari untai DNA yang ada. Hasil yang didapatkan adalah peningkatan akurasi di hampir semua *dataset* yang digunakan. Hanya pada H3K14ac dan H4ac yang tidak memiliki peningkatan akurasi.

Algoritme fitur seleksi yang digunakan adalah *Random Forest*. Setelah melakukan seleksi fitur dilanjutkan dengan pengklasifikasian dengan metode *Support Vector Machine*. Kernel yang digunakan pada SVM adalah RBF kernel. Setelah mendapatkan hasil klasifikasi, dilakukan perbandingan dengan metode lainnya dan menunjukkan rata-rata hasil yang lebih baik.

Pada penelitian Nguyen et al., (2016) dilakukan pengklasifikasian menggunakan data yang sama pada penelitian. Terdapat tambahan *dataset*, yaitu *dataset Splice and Promotor* dari *UCI machine learning repository*. Untai DNA yang ada akan direpresentasikan sebagai *One-hot vector* agar informasi urutan nukleotida tetap tersimpan. Pada akhirnya hasil klasifikasi yang dilakukan mengalami peningkatan hingga 6%.

Seq-CNN adalah algoritme yang digunakan pada penelitian Nguyen et al., (2016). Algoritme tersebut merupakan algoritme CNN yang dimodifikasi agar dapat mengklasifikasikan untai. Untai DNA ditransformasikan menjadi *One-hot vector* agar memiliki karakteristik yang sama dengan data teks. Setelah ditransformasi, data DNA akan menjadi input untuk model CNN dan dilakukan pengujian agar mendapatkan akurasi yang dimiliki oleh model. Hasil akhirnya model memberikan hingga 6.12% lebih baik pada *dataset* H3K4me3. Peningkatan terendah terjadi pada *dataset* H4 yang hanya meningkat sebesar 0.77%.

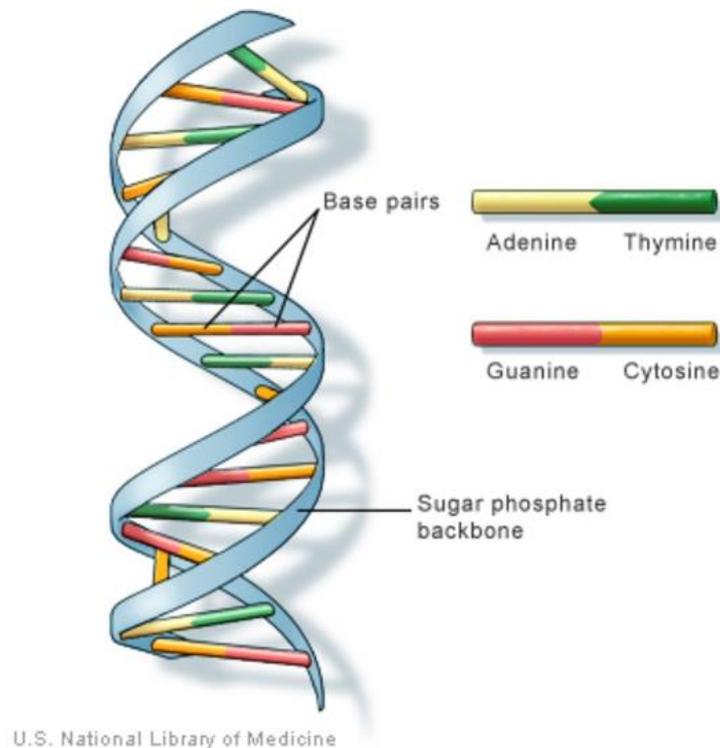
Penelitian Mahmoud & Guo (2021) merupakan penelitian yang tidak merepresentasikan untai DNA dengan *One-hot encoder*. Representasi DNA

yang digunakan adalah dengan menggunakan gambar. Representasi gambar tersebut dihasilkan dari DNA *Data Visualization* yang dikembangkan oleh Neugebauer et al., (2015). Gambar yang dihasilkan akan diklasifikasikan dengan menggunakan *Pseudoinverse Learning Autoencoder* (PILAE). Hasil penelitian menunjukkan peningkatan akurasi yang signifikan. Metode ini juga meningkatkan efisiensi komputasi dengan baik.

Proses pertama yang dilakukan pada penelitian Mahmoud & Guo (2021) adalah mengubah data DNA menjadi gambar dengan DNA *Data Visualization* (Neugebauer et al., 2015). Setelah data diubah menjadi gambar akan dilakukan seleksi fitur dengan *pre-train model* seperti Xception dan VGG. Setelah fitur telah dipilih barulah dilakukan klasifikasi dengan PILAE. Hasil Akurasi yang didapatkan mencapai 98.57% pada *dataset* H3K9ac. Peningkatan akurasi juga mencapai 14.59% pada *dataset* H4ac.

2.2. DNA

DNA merupakan singkatan dari *deoxyribonucleic acid*. DNA adalah rantai polimer yang menyimpan informasi genetik dari beragam makhluk hidup (Mamajanov & Hud, 2014). DNA terdiri dari 4 nukleotida yaitu *adenine* (A), *guanine* (G), *cytosine* (C), and *thymine* (T). Setiap nukleotida akan memiliki pasangan berdasarkan pasangan dasar Watson-Crick(A-T dan G-C). DNA dalam makhluk hidup mayoritas ada dalam bentuk untaian *double helix*. Pada proses tertentu untai tersebut akan diuraikan dan menjadi *single helix*.



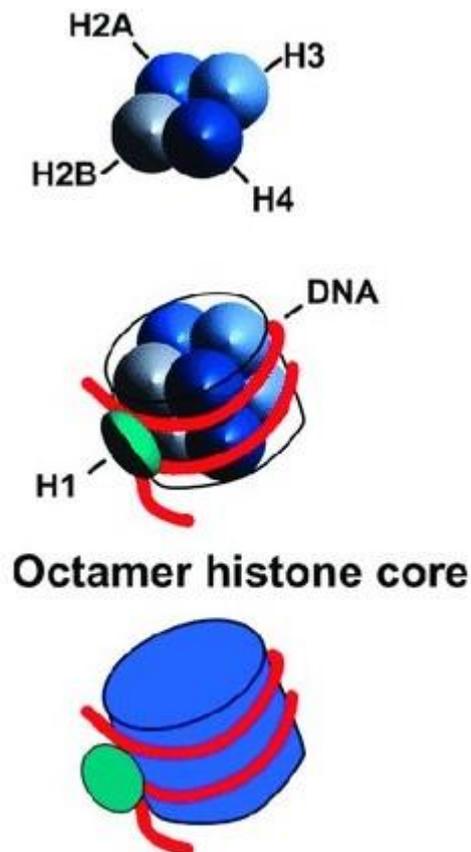
Gambar 1. Ilustrasi DNA (U.S. National Library of Medicine).

Gambar 1 menunjukkan struktur dari untai DNA. DNA pada keadaan normal akan tersimpan dalam bentuk *double helix* namun, pada proses replikasi DNA akan berbentuk *single helix*. Dapat dilihat pada Gambar 1 DNA terdiri dari *base pair* yang menempel pada molekul gula deoksiribosa dan molekul fosfat. *Base pair* pada DNA juga pasti memiliki pasangan dengan aturan adenin dengan timin dan guanin dengan sitosin.

2.3. Histon

Histon adalah protein yang terdapat pada inti sel dan membentuk struktur kromosom. Histon ditemukan pada semua jenis sel eukariotik. Kelompok protein ini terdiri dari lima tipe utama, yaitu H1, H2A, H2B, H3, dan H4, yang memiliki struktur serupa dan berperan dalam pembentukan nukleosom, yaitu struktur kromosom dasar.

Histon H1 berperan sebagai "pengikat" nukleosom dan mengatur interaksi antara nukleosom dan DNA, sementara H2A, H2B, H3, dan H4 membentuk inti nukleosom. Kombinasi dari modifikasi histon seperti asetilasi dan metilasi dapat memengaruhi struktur kromatin dan mengatur aksesibilitas DNA, yang pada akhirnya mempengaruhi aktivitas gen.



Gambar 2. Ilustrasi antara histon dan DNA (Elmallah & Micheau, 2019).

2.4. *One-hot Encoding*

One-hot encoding adalah proses pengodean suatu data kategori menjadi data numerik. Proses ini akan menghasilkan sebuah vektor dengan panjang sebanyak kategori pada data (Brownlee, 2017). Pada data kategori ke i akan diberikan nilai 1 dan yang lainnya 0.

Tabel 2. Matrix One-hot *Encoding*

A	G	C	T
1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	1

Jika diberikan sebuah untai DNA “CAGGT” akan mendapatkan matriks sebagai berikut

Tabel 3. Hasil One-hot *Encoding* dari DNA “CAGGT”

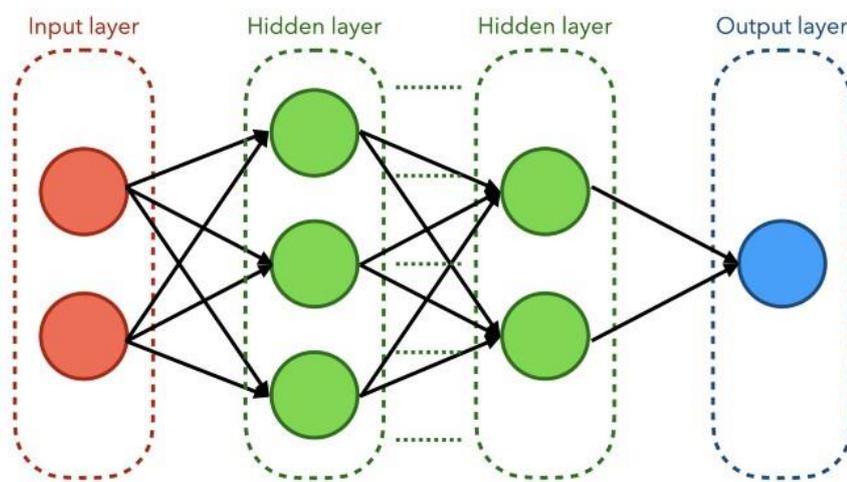
C	A	G	G	T
0	1	0	0	0
0	0	1	1	0
1	0	0	0	0
0	0	0	0	1

Tabel 2 menunjukkan contoh dari One-hot encoding dari 4 jenis basa nukleotida DNA. Berdasarkan Tabel 2, hasil dari encoding DNA “CAGGT” akan membentuk matrix yang diilustrasikan oleh Tabel 3. Oleh karena itu jika terdapat data DNA dengan panjang 500, maka hasil dari One-hot encoding akan berukuran 500x4.

One-hot sudah banyak digunakan pada penelitian klasifikasi DNA (Busia et al., 2018; Fiannaca et al., 2018; Nguyen et al., 2016) Pada penelitian Busia et al., (2018) mengenai klasifikasi DNA dengan untai pendek, *One-hot encoding* digunakan untuk mengkodekan setiap nukleotida dan identitas spesies. Pada penelitian pengklasifikasian taksonomi oleh (Fiannaca et al., 2018) *One-hot encoding* digunakan untuk pengodean nukleotida saja. Sementara pada (Nguyen et al., 2016) *One-hot* digunakan untuk tetap menjaga informasi dari untai selama pengodean nukleotida.

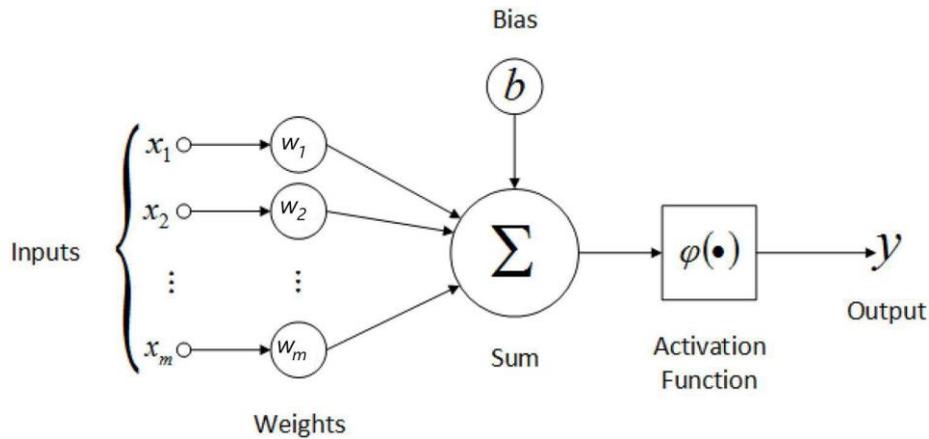
2.5. Artificial Neural Networks

Artificial Neural Networks merupakan sebuah algoritme yang terinspirasi dari otak manusia. Otak manusia merupakan jaringan dari saraf yang terdiri dari reseptor dan efektor. Reseptor atau disebut dendrit bertugas untuk menerima sinyal dari saraf lainnya. Sementara efektor yang disebut akson akan mengirimkan sinyal ke saraf lainnya (Dongare et al., 2012).



Gambar 3. Ilustrasi dari Artificial Neural Network (Nurfikri, 2020).

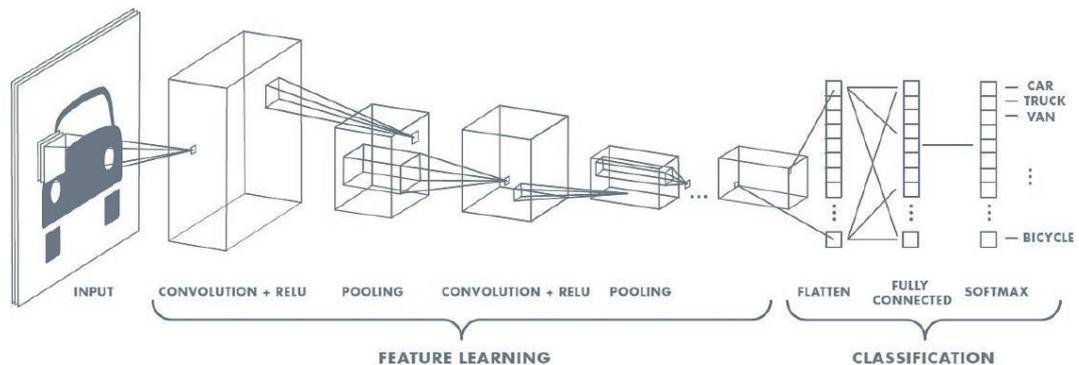
Pada *input layer* setiap neuron akan menerima data masukan kemudian mengalikannya dengan *weight* neuron tersebut dan meneruskannya ke layer selanjutnya. Layer selanjutnya akan menjumlahkan seluruh hasil perkalian data masukan dan *weight* kemudian akan menjadi masukan fungsi aktivasi. Berikut gambaran proses pada ANN ditampilkan pada Gambar 4.



Gambar 4. Ilustrasi perhitungan pada ANN (Schils, 2020).

2.6. Convolutional Neural Network

Convolutional Neural Network (CNN) atau *ConvNet* merupakan bagian dari ANN yang umum digunakan untuk menganalisis sebuah citra digital (Valueva et al., 2020). CNN termasuk dalam jenis *deep neural network* karena kedalaman jaringan yang tinggi dan banyak diaplikasikan pada data citra. Secara garis besar, CNN tidak jauh berbeda dengan arsitektur *neural network*, yaitu memiliki *weight*, *bias*, dan *activation function*. CNN dapat terdiri atas beberapa jenis layer seperti *convolution layer*, *pooling layer*, dan *fully connected layer*.



Gambar 5. Contoh arsitektur dari *Convolutional Neural Network* (Matlab).

Gambar 5 menunjukkan gambaran umum mengenai susunan dari arsitektur CNN. Mulai dari *input layer*, *convolutional layer*, *pooling layer*, *flatten layer*, hingga *fully connected layer*.

3.1.1. Convolutional Layer

Convolutional layer adalah layer pertama yang akan memproses citra yang dimasukkan, sehingga menjadikan layer ini sangat penting pada arsitektur *convolutional neural network* (CNN). Pada layer ini, terjadi operasi konvolusi dan filtrasi untuk mendapatkan ekstraksi fitur dari data atau citra yang dimasukkan. Operasi tersebut dinamakan *feature map* atau *activation map*.

3.1.2. Pooling Layer

Pooling layer adalah lapisan atau *layer* yang mengolah *output* dari *Activation ReLU layer* dengan berbagai macam operasi statistika berdasarkan nilai piksel terdekat. *Pooling layer* dapat mengendalikan *overfitting* dengan cara mengurangi ukuran volume *output* ada *feature map* secara progresif.

3.1.3. Fully Connected Layer

Fully connected layer adalah *layer* yang *neuron-neuron* terhubung dengan semua *neuron* pada lapisan sebelumnya. Lapisan ini akan menerima *input* dari lapisan aktivasi sebelumnya untuk menentukan fitur yang paling berkorelasi dengan kelas tertentu berdasarkan *training dataset* yang ada.

2.7. Confusion Matrix

Confusion matrix adalah salah satu metode yang sering digunakan dalam evaluasi model *machine learning* (Kulkarni et al., 2020). *Confusion matrix* dapat mengevaluasi *binary class* ataupun *multi class classification*. *Confusion matrix* adalah tabel dengan 4 kombinasi berbeda dari nilai prediksi dan nilai aktual. *Confusion matrix* akan menilai dari aspek *accuracy*, *precision*, *recall*, dan *f-1 score*. Berikut contoh tabel *confusion matrix*

Tabel 4. Confusion matrix pada klasifikasi dua kelas

	<i>Predicted Positive</i>	<i>Predicted Negative</i>
<i>Actual Positive</i>	<i>TP</i>	<i>FN</i>
<i>Actual Negative</i>	<i>FP</i>	<i>TN</i>

3.1.4. Accuracy

Akurasi adalah matriks yang paling umum untuk melakukan evaluasi klasifikasi. Akurasi bekerja dengan cara menghitung nilai probabilitas berdasarkan nilai yang benar dari label kelas. Akurasi dirumuskan pada persamaan (1) (Bekkar et al., 2013).

$$Akurasi = \frac{TN+TP}{TN+FP+FN+TP} \dots\dots\dots (1)$$

3.1.5. Precision

Precision adalah nilai kebenaran dari prediksi yang dilakukan oleh *classifier* dengan label kelas yang sudah ada. *Precision* dirumuskan pada persamaan (2) (Bekkar et al., 2013).

$$Precision = \frac{TP}{TP+FP} \dots\dots\dots (2)$$

3.1.6. Sensitivity/Recall

Recall atau *sensitivity* adalah keakuratan dari nilai positif yang ada. *Recall* dirumuskan pada persamaan (3) (Bekkar et al., 2013).

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots (3)$$

III. METODOLOGI PENELITIAN

3.2. Tempat Dan Waktu Penelitian

3.2.1. Tempat Penelitian

Penelitian akan dilakukan di Universitas Lampung, Gedung MIPA Terpadu, pada ruangan laboratorium Rekayasa Perangkat Lunak. Penelitian dilakukan pada laboratorium RPL agar dapat menggunakan fasilitas komputasi yang telah disediakan.

3.2.2. Waktu Penelitian

Berikut merupakan tabel yang menggambarkan jadwal penelitian ini:

Penelitian akan dimulai pada bulan Desember 2021 sampai dengan bulan April 2022. Penelitian akan melalui 3 tahapan yaitu:

1. Persiapan data

Pertama-tama yang data akan dikumpulkan terlebih dahulu. Kemudian data-data yang tidak dapat digunakan akan dibersihkan pada tahap *cleaning*. Data yang sudah dibersihkan akan dikodekan dengan *One-hot encoding*.

2. Transformasi data

Data yang telah dikumpulkan akan dikodekan dengan menggunakan metode *One-hot encoding*. Sehingga data yang awalnya berupa *string* dengan panjang 500 akan ditransformasi menjadi matrix berukuran 500x4.

3. Pemodelan

Selanjutnya akan dibuat model CNN untuk melakukan pengklasifikasian pada data tersebut. Pada tahap awal akan digunakan 50% data. Saat model yang dievaluasi sudah memiliki akurasi yang baik, pelatihan akan dilanjutkan dengan seluruh data yang ada.

4. Evaluasi Penelitian

Tahap terakhir adalah evaluasi penelitian. Evaluasi penelitian akan dilihat seberapa baik model melakukan klasifikasi. Yang terakhir adalah kesimpulan dari hasil penelitian ini.

3.3. Data Dan Alat

3.3.1. Data

Data yang digunakan merupakan data pada penelitian Pokholok et al., (2005). Data berupa kumpulan untai DNA dengan panjang 500 nukleotida. Setiap *dataset* terdapat “H3” dan “H4” menandakan jenis histon. “K” dan angka setelahnya menandakan asam amino yang dimodifikasi (“K14” menandakan

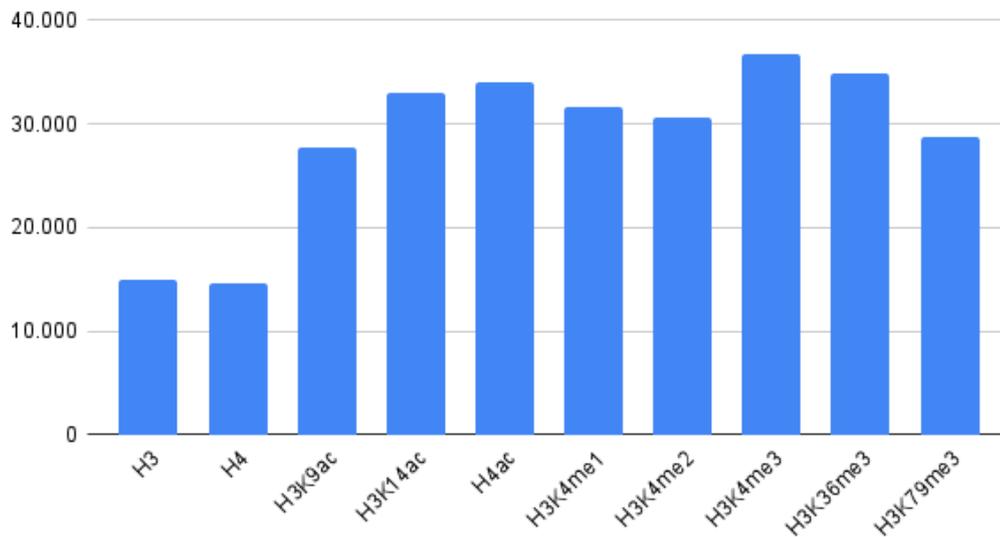
asam amino ke-14, “K”, mengalami modifikasi). “me” dan “ac” dan angka setelahnya menandakan jenis modifikasi yang terjadi (“me2” menandakan terjadinya dimetilasi). Data tersebut juga telah digunakan pada penelitian (Higashihara et al., 2008; Mahmoud & Guo, 2021; Nguyen et al., 2016).

Tabel 6. Data DNA ragi (Pokholok et al., 2005)

<i>Dataset</i>	Deskripsi	Positif	Negatif	Total
H3	Keberadaan H3	7.667	7.298	14.965
H4	Keberadaan H4	6.480	8.121	14.601
H3K9ac	Asetilasi H3K9 terhadap H3	15.415	12.367	27.782
H3K14ac	Asetilasi H3K14 terhadap H3	18.771	14.277	33.048
H4ac	Asetilasi H4	18.410	15.685	34.095
H3K4me1	Monometilasi H3K4 terhadap H3	17.266	14.411	31.677
H3K4me2	Dimetilasi H3K4 terhadap H3	18.143	12.540	30.683
H3K4me3	Trimetilasi H3K4 terhadap H3	19.604	17.195	36.799
H3K36me3	Trimetilasi H3K36 terhadap H3	18.892	15.988	34.880
H3K79me3	Trimetilasi H3K79 terhadap H3	15.337	13.500	28.837

Tabel 6 menunjukkan data yang digunakan pada penelitian ini. Terdapat 10 *dataset* yang akan digunakan pada penelitian ini. Data H3, H4, H3K9ac, H3K14ac, H4ac, H3K4me1, H3K4me2, H3K4me3, H3K36me3, dan H3K79me3 didapatkan dari penelitian Pokholok et al., pada tahun 2005.

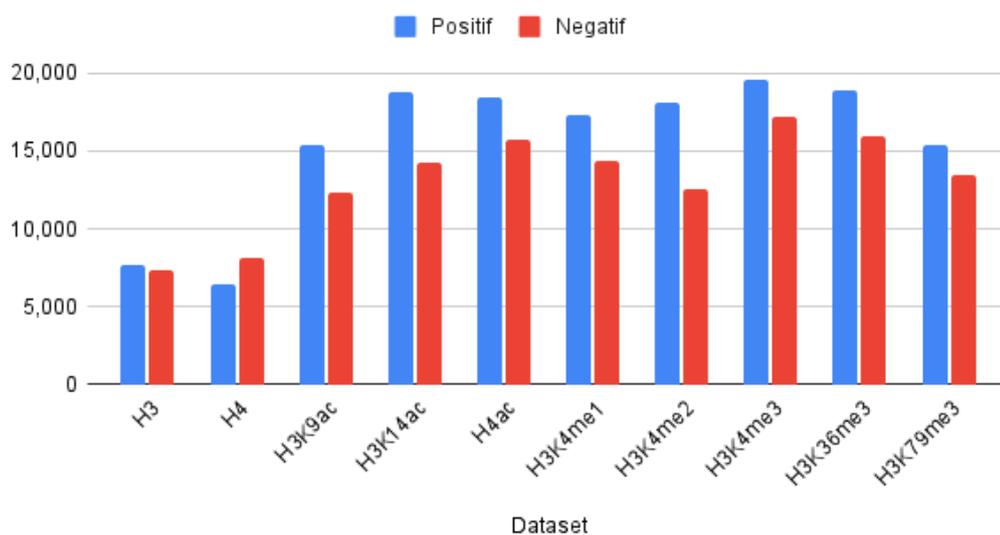
Jumlah DNA pada setiap data



Gambar 6. Pebandingan total setiap data.

Gambar 6 menunjukkan total untai DNA dari setiap *dataset* yang akan digunakan. Dapat dilihat dataset dengan jumlah data terbesar adalah *dataset* H3K4me3 dengan jumlah data sebanyak 34.880. Sementara data dataset dengan jumlah data terkecil adalah *dataset* H4 dengan data DNA sebanyak 14.601.

Positif and Negatif



Gambar 7. Perbandingan data positif dan negatif pada setiap *dataset*.

Gambar 7 menunjukkan jumlah data positif dan negatif DNA dari setiap *dataset* yang akan digunakan. Dapat dilihat bahwa setiap *dataset* memiliki jumlah positif dan negatif yang hampir seimbang, sehingga tidak perlu dilakukan *oversampling* atau *undersampling*. Dapat dilihat juga dataset dengan selisih data terbesar adalah *dataset* H3K4me2 dengan jumlah data sebanyak 5.603. Sementara data dataset dengan selisih data positif dan negatif terkecil adalah *dataset* H4 dengan data DNA sebanyak 369. Data positif terbesar terdapat di data H3K4me3 sebesar 19.604. Data negatif terbesar adalah sebanyak 19.604 di *dataset* H3K4me2.

3.3.2. Alat

1. Hardware

- a. Processor Intel I7 7800K
- b. Random Access Memory 12 GB, DDR4, 2666 Mhz
- c. GPU NVidia Tesla K20 5GB, 208 GB/s CUDA Cores.
- d. SSD SATA Westren Digital 240GB

2. Software

a. Python 3.8.8

Python merupakan bahasa pemrograman yang sering digunakan pada pemrosesan data. Salah satu kelebihan python adalah kemudahan dalam penggunaan dan *syntax* yang menyerupai bahasa alami manusia. Kelebihan lainnya adalah banyaknya dukungan *library* yang mendukung pemrograman pada python.

b. Library Pandas 1.2.4

Pandas merupakan *library* yang digunakan untuk menyimpan data dalam bentuk tabulasi. *Library* ini juga dapat mempermudah dalam pemrosesan data seperti pada saat pembacaan file ke dalam program.

c. Library Numpy 1.19.5

Numpy merupakan sebuah *library* yang mendukung komputasi dengan

kecepatan tinggi untuk *array*. *Library* pandas juga menggunakan numpy untuk menyimpan data *array* pada tabelnya.

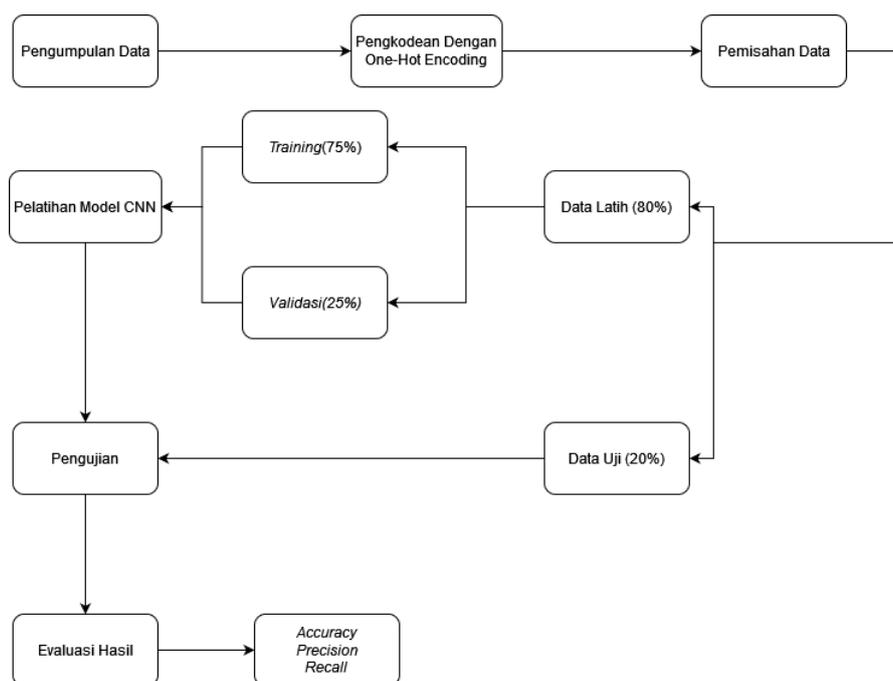
d. *Library Tensorflow 2.5.0*

Tensorflow adalah *library* yang dibuat oleh *google* untuk melakukan pemodelan *deep learning* dengan mudah. *Tensorflow* merupakan *library* yang berjalan pada bahasa pemrograman yang lebih rendah (mendekati bahasa mesin) sehingga memiliki performa komputasi yang baik.

Tensorflow juga telah memuat *Keras* yang sering digunakan juga untuk pemodelan *deep learning*.

3.4. Alur Kerja Penelitian

Penelitian ini akan melalui beberapa tahapan yang sebagai berikut:



Gambar 8. Alur Penelitian terkait klasifikasi DNA ragi dengan metode CNN.

Penjelasan dari setiap tahap penelitian

1. Transformasi fitur

Dataset DNA yang telah dibersihkan akan ditransformasikan fiturnya.

Transformasi fitur ini bertujuan agar data untai DNA dapat menjadi *input* model CNN. Proses ini akan mengubah untai DNA yang merupakan sebuah

string menjadi matriks 2 dimensi. Algoritme yang digunakan adalah algoritme *One-hot encoding*. Algoritme ini akan merepresentasikan setiap nukleotida menjadi sebuah vektor.

2. Pembagian data latih dan data uji

Setelah data DNA diekstrak fiturnya, data akan dibagi menjadi data latih dan data uji. Data latih akan digunakan dalam pembuatan model, sementara data uji akan digunakan setelah selesai pemodelan. Pada data latih akan digunakan 80% dari *dataset* dan data uji akan 20% dari *dataset*. Pembagian data latih dan uji dengan rasio 80-20 digunakan agar eksperiment memiliki rasio yang sama dengan penelitian sebelumnya.

3. Pemodelan pengklasifikasi

Data latih akan digunakan untuk melatih model yang akan dibuat. Pada tahapan ini akan disesuaikan *hyperparameter* pada model CNN. Proses ini akan menggunakan data latih yang akan kembali dibagi menjadi data *training* dan *validation*. Data *training* merupakan 75% dari data latih dan data *validation* merupakan 25% dari data latih. Pembagian data *validation* dari data latih bertujuan untuk memonitor proses pembelajaran model. Proses pelatihan akan berhenti saat akurasi pada data validasi tidak bertambah baik lagi.

4. Prediksi dan evaluasi data uji

Tahap selanjutnya adalah pengujian dengan data uji. Hasil prediksi akan dievaluasi menggunakan *confusion matrix*. Dari *confusion matrix* akan diperoleh *accuracy*, *precision*, *sensitivity*, dan *f-1 score*. Hasil yang didapatkan akan dibandingkan dengan penelitian terdahulu.

V. SIMPULAN DAN SARAN

5.1. Simpulan

Dari penelitian dan pembahasan yang sudah dilakukan mengenai klasifikasi DNA ragi menggunakan metode *convolutional neural network* (CNN) dapat diambil beberapa kesimpulan, antara lain.:

1. Dengan melakukan 3 eksperimen didapatkan hasil sebagai berikut:
 - a. Eksperimen 1 mendapatkan hasil pelatihan dengan rerata *accuracy* sebesar 86.89%, *recall* sebesar 86.27%, dan *precision* sebesar 89.15%. Pengujian mendapatkan hasil dengan rerata *accuracy* 64.83%, *recall* 65.44%, dan *precision* 68.99%. Terdapat perbedaan yang cukup besar antara hasil latih dan hasil uji.
 - b. Eksperimen 2 mendapatkan hasil pelatihan dengan rerata *accuracy* 80.78%, *recall* 85.95%, dan *precision* 82.38%. Pengujian pada eksperimen 2 mendapatkan hasil dengan rerata *accuracy* 64.89%, *recall* 70.59% dan *precision* 67.80%.
 - c. Eksperimen 3 mendapatkan hasil pelatihan dengan rerata *accuracy* 84.10%, *recall* 83.27%, dan *precision* 86.87%. Pengujian pada eksperimen 3 mendapatkan hasil dengan rerata *accuracy* 65.32%, *recall* 65.13% dan *precision* 69.51%.

Dapat disimpulkan pada penelitian ini hasil klasifikasi DNA ragi terbaik didapatkan dari eksperimen 3 yang memiliki rerata akurasi pengujian yang terbaik dari ketiga eksperimen.

2. Dibandingkan dengan penelitian terdahulu hasil eksperimen pada penelitian ini masih belum berhasil meningkatkan akurasi dari proses pengklasifikasian asetilasi dan metilasi pada DNA.

5.2. Saran

1. Penelitian selanjutnya dapat mencoba untuk melakukan encoding dengan jumlah nukelotida yang lebih besar.
2. Penelitian selanjutnya dapat mencoba berbagai teknik untuk meminimalisir terjadinya *overfit* dengan mencoba berbagai kombinasi *layer* yang berbeda
3. Penelitian selanjutnya dapat mencoba melakukan klasifikasi dengan memperhatikan urutan dari untai DNA dengan menggunakan algoritme seperti RNN/LSTM.

DAFTAR PUSTAKA

- Bekkar, M., Djema, H., & Alitouche, T. A. (2013). Evaluation measures for models assessment over imbalanced data sets. *Journal of Information Engineering and Applications*, 3, 27–38.
- Brownlee, J. (2017). *Why One-hot Encode Data in Machine Learning?* Machine Learning Mastery.
- Busia, A., Dahl, G. E., Fannjiang, C., Alexander, D. H., Dorfman, E., Poplin, R., McLean, C. Y., Chang, P. C., & DePristo, M. (2018). A deep learning approach to pattern recognition for short DNA sequences. *BioRxiv*, 0–33.
- Desai, H. P., Parameshwaran, A. P., Sunderraman, R., & Weeks, M. (2020). Comparative Study Using Neural Networks for 16S Ribosomal Gene Classification. *Journal of Computational Biology*, 27(2), 248–258.
- Dongare, A. D., Kharde, R. R., & Kachare, A. D. (2012). Introduction to artificial neural network. *International Journal of Engineering and Innovative Technology (IJEIT)*, 2(1), 189–194.
- Fiannaca, A., la Paglia, L., la Rosa, M., lo Bosco, G., Renda, G., Rizzo, R., Gaglio, S., & Urso, A. (2018). Deep learning models for bacteria taxonomic classification of metagenomic data. *BMC Bioinformatics*
- Elmallah, Mohammed & Micheau, Olivier. (2019). Epigenetic Regulation of TRAIL Signaling: Implication for Cancer Therapy. *Cancers*. 11. 850.
- Higashihara, M., Rebolledo-mendez, J. D., Yamada, Y., & Satou, K. (2008). *Application of a Feature Selection Method to Nucleosome Data: Accuracy Improvement and Comparison with Other Methods*. 5(5).
- Kulkarni, A., Chong, D., & Batarseh, F. A. (2020). Foundations of data imbalance and solutions for a data democracy. In *Data Democracy: At the Nexus of Artificial Intelligence, Software Development, and*

Knowledge Engineering. Elsevier Inc.

- Mahmoud, M. A. B., & Guo, P. (2021). DNA sequence classification based on MLP with PILAE algorithm. *Soft Computing*, 25(5), 4003–4014.
- Mamajanov, I., & Hud, N. v. (2014). *DNA BT - Encyclopedia of Astrobiology* (R. Amils, M. Gargaud, J. Cernicharo Quintanilla, H. J. Cleaves, W. M. Irvine, D. Pinti, & M. Viso, Eds.; pp. 1–7). Springer Berlin Heidelberg.
- Neugebauer, T., Bordeleau, E., Burrus, V., & Brzezinski, R. (2015). DNA data visualization (DDV): Software for generating web-based interfaces supporting navigation and analysis of DNA sequence data of entire genomes. *PLoS ONE*, 10(12), 1–16.
- Nguyen, N. G., Tran, V. A., Ngo, D. L., Phan, D., Lumbanraja, F. R., Faisal, M. R., Abapihi, B., Kubo, M., & Satou, K. (2016). DNA Sequence Classification by Convolutional Neural Network. *Journal of Biomedical Science and Engineering*, 09(05), 280–286.
- Pokholok, D. K., Harbison, C. T., Levine, S., Cole, M., Hannett, N. M., Tong, I. L., Bell, G. W., Walker, K., Rolfe, P. A., Herbolsheimer, E., Zeitlinger, J., Lewitter, F., Gifford, D. K., & Young, R. A. (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell*, 122(4), 517–527.
- Schils, M. (2020). *Audio frame reconstruction from incomplete observations using Deep Learning techniques*.
- Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). DeepFace: Closing the Gap to Human-Level Performance in Face Verification. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 1701–1708.
- Valueva, M. v., Nagornov, N. N., Lyakhov, P. A., Valuev, G. v., & Chervyakov, N. I. (2020). Application of the residue number system to reduce hardware costs of the convolutional neural network implementation. *Mathematics and Computers in Simulation*, 177, 232–243.