



**PENERAPAN MODEL HIBRIDA *CONVOLUTIONAL NEURAL NETWORK* DAN *VISION TRANSFORMER* PADA KLASIFIKASI CITRA DENGAN VISUAL HETEROGEN UNTUK IDENTIFIKASI JENIS SAMPAH**

**SKRIPSI**

**untuk memenuhi persyaratan  
dalam menyelesaikan program sarjana Strata-1 Statistika**

**Oleh  
MUHAMMAD FAUZAN ADZIM  
NIM. 2211017110001**

**PROGRAM STUDI S-1 STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMBUNG MANGKURAT  
BANJARBARU  
JANUARI 2026**



**PENERAPAN MODEL HIBRIDA *CONVOLUTIONAL NEURAL NETWORK* DAN *VISION TRANSFORMER* PADA KLASIFIKASI CITRA DENGAN VISUAL HETEROGEN UNTUK IDENTIFIKASI JENIS SAMPAH**

**SKRIPSI**

**untuk memenuhi persyaratan  
dalam menyelesaikan program sarjana Strata-1 Statistika**

**Oleh  
MUHAMMAD FAUZAN ADZIM  
NIM. 2211017110001**

**PROGRAM STUDI S-1 STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMBUNG MANGKURAT  
BANJARBARU  
JANUARI 2026**

# SKRIPSI

## PENERAPAN MODEL HIBRIDA *CONVOLUTIONAL NEURAL NETWORK* DAN *VISION TRANSFORMER* PADA KLASIFIKASI CITRA DENGAN VISUAL HETEROGEN UNTUK IDENTIFIKASI JENIS SAMPAH

Oleh  
**Muhammad Fauzan Adzim**  
NIM. 2211017110001

Telah dipertahankan pada hari Senin, tanggal 22-12-2025 dan disetujui oleh dosen pembimbing dan dosen penguji sebagai berikut:

### Pembimbing I



Dewi Sri Susanti, S.Si., M.Si.  
NIP. 19730516 199903 2 002

### Penguji I



Selvi Annisa, S.Si., M.Si.  
NIP. 19921226 202203 2 016

### Pembimbing II



Sigit Dwi Prabowo, S.Mat., M.Stat.  
NIP. 19960528 202406 1 003

### Penguji II



Oni Soesanto, S.Si., M.Si.  
NIP. 19730126 200501 1 003

Banjarbaru, 6 Januari 2026  
Mengetahui,  
Koordinator Program Studi Statistika  
FMIPA ULM

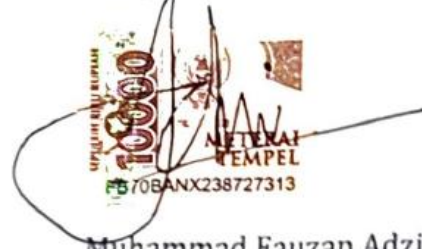


Dewi Anggraini, S.Si., M.App.Sci., Ph.D.  
NIP. 19830328 200501 2 001

## PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu 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, 5 Januari 2026



The image shows a handwritten signature in black ink over a rectangular stamp. The stamp contains the text 'KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN' at the top, 'PRODI STATISTIKA' in large letters in the middle, and 'METERAI TEMPEL' and '708ANX238727313' at the bottom.

Muhammad Fauzan Adzim  
NIM. 2211017110001

PRODI STATISTIKA

## ABSTRAK

**Penerapan Model Hibrida *Convolutional Neural Network* dan *Vision Transformer* pada Klasifikasi Citra dengan Visual Heterogen untuk Identifikasi Jenis Sampah** (Oleh: Muhammad Fauzan Adzim; Pembimbing: Dewi Sri Susanti dan Sigit Dwi Prabowo, 2025; 68 halaman)

Dalam klasifikasi citra, *convolutional neural network* (CNN) berfokus pada pola lokal, sedangkan *vision Transformer* (ViT) menekankan konteks global. Integrasi keduanya dalam model hibrida berpotensi menghasilkan representasi citra yang lebih utuh. Klasifikasi citra sampah dengan karakteristik visual yang heterogen dapat dimanfaatkan untuk mengevaluasi hibrida CNN-ViT. Selain itu, klasifikasi tersebut mendukung kebutuhan identifikasi jenis sampah yang akurat bagi pengelolaan sampah yang efektif. Penelitian ini mengkaji penerapan hibrida CNN-ViT untuk klasifikasi citra sampah dengan visual heterogen. Data yang digunakan terdiri dari 24.705 berformat citra RGB pada dua kelas, yaitu *organic* dan *recyclable*. Tahapan penelitian meliputi eksplorasi data, *resize*, pembagian data 80:10:10 (*training*, *validation*, dan *testing*), augmentasi pada data *training*, serta standardisasi. Model dilatih selama 50 *epoch* menggunakan *binary cross entropy* dan *optimizer* Adam. Evaluasi dilakukan pada *epoch* terbaik (<50) yang ditentukan berdasarkan akurasi validasi tertinggi. Sebagai pembandingan, model CNN dan ViT juga dilatih serta dievaluasi secara terpisah. Pada data *testing*, diperoleh akurasi hibrida CNN-ViT sebesar 91.54%, CNN sebesar 91.78%, dan ViT sebesar 87.17%. Temuan ini menunjukkan bahwa CNN dapat menjadi *baseline* yang efektif dan efisien, sementara hibrida CNN-ViT memberikan kinerja yang kompetitif dengan CNN dan layak dipertimbangkan sebagai alternatif untuk klasifikasi sampah berbasis citra.

Kata Kunci: Hibrida, *Convolutional Neural Network*, *Vision Transformer*, Sampah, Klasifikasi Citra

## ABSTRACT

**Implementation of a Hybrid Convolutional Neural Network and Vision Transformer Model to the Classification of Visually Heterogeneous Images for Waste Type Identification** (By: Muhammad Fauzan Adzim; Supervisors: Dewi Sri Susanti and Sigit Dwi Prabowo, 2025; 68 pages)

In image classification, convolutional neural networks (CNNs) focus on local patterns, whereas vision Transformers (ViTs) emphasize global context. Integrating the two in a hybrid model has the potential to produce a more complete image representation. Waste image classification with heterogeneous visual characteristics can be utilized to evaluate the CNN-ViT hybrid. Furthermore, such classification supports the need for accurate waste-type identification for effective waste management. This study examines the application of a CNN-ViT hybrid for classifying waste images with heterogeneous visuals. The data used consists of 24,705 images in RGB format across two classes: organic and recyclable. The research stages include data exploration, resizing, an 80:10:10 data split (training, validation, and testing), augmentation of the training data, and standardization. The model is trained for 50 epochs using binary cross-entropy and the Adam optimizer. Evaluation is performed at the best epoch, determined by the highest validation accuracy. For comparison, CNN and ViT models are also trained and evaluated separately. On the testing data, the CNN-ViT hybrid achieved an accuracy of 91.54%, the CNN 91.78%, and the ViT 87.17%. These findings indicate that the CNN can serve as an effective and efficient baseline, while the CNN-ViT hybrid provides competitive performance with the CNN and is worth considering as an alternative for image-based waste classification.

Keywords: Hybrid, Convolutional Neural Network, Vision Transformer, Waste, Image Classification

## PRAKATA

Puji syukur ke hadirat Allah Swt. atas segala rahmat, taufik, hidayah, dan inayah-Nya sehingga penulis dapat menyelesaikan Tugas Akhir berjudul “Penerapan Model Hibrida *Convolutional Neural Network* dan *Vision Transformer* pada Klasifikasi Citra dengan Visual Heterogen untuk Identifikasi Jenis Sampah”. Penulisan ini bertujuan untuk memenuhi salah satu persyaratan dalam menyelesaikan program sarjana pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lambung Mangkurat. Penulis menyadari bahwa dalam proses penyusunan masih terdapat berbagai keterbatasan, baik dari segi penulisan maupun perumusan kalimat. Penulis bersyukur atas terselesaikannya Tugas Akhir ini. Atas bimbingan dan dukungan dari berbagai pihak, Tugas Akhir ini dapat diselesaikan. Dengan tulus, penulis mengucapkan terima kasih kepada pihak-pihak yang telah membantu, khususnya kepada:

1. Orang tua dan keluarga yang selalu memberikan dukungan serta mendoakan untuk penyelesaian Tugas Akhir;
2. Ibu Dewi Sri Susanti, S.Si., M.Si. dan Bapak Sigit Dwi Prabowo, S.Mat., M.Stat. selaku dosen pembimbing yang telah bersedia meluangkan waktu, tenaga, dan pikiran dalam pelaksanaan penelitian serta penyelesaian Tugas Akhir;
3. Ibu Selvi Annisa, S.Si., M.Si. dan Bapak Oni Soesanto, S.Si., M.Si. selaku dosen penguji yang telah memberikan masukan, saran, dan bantuan dalam perbaikan penulisan Tugas Akhir;
4. Ketua Jurusan Statistika FMIPA ULM beserta seluruh dosen dan staf yang telah memberikan ilmu, motivasi, nasihat, dan dukungan selama masa perkuliahan;
5. Diri penulis sendiri yang telah berusaha keras dan ikhlas selama proses penelitian maupun penyusunan Tugas Akhir;
6. Teman-teman beserta pihak lain yang telah memberikan dukungan dan semangat kepada penulis selama penelitian maupun penyusunan penulisan Tugas Akhir.

Penulis menyadari bahwa Tugas Akhir ini masih memiliki berbagai keterbatasan. Penulis sangat mengharapkan kritik dan saran yang konstruktif dari berbagai pihak sebagai bahan perbaikan dan penyempurnaan di masa yang akan datang. Akhir kata, penulis mengucapkan terima kasih dan memanjatkan doa agar segala kebaikan yang diberikan oleh pihak-pihak yang telah membantu memperoleh balasan yang sebaik-baiknya. Semoga karya ini dapat memberikan manfaat, memperluas wawasan, dan menjadi referensi yang berguna bagi pembaca.

Banjarbaru, 5 Januari 2026

Muhammad Fauzan Adzim

## DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
PERNYATAAN.....	iii
ABSTRAK.....	iv
ABSTRACT .....	v
PRAKATA.....	vi
DAFTAR ISI .....	vii
DAFTAR GAMBAR.....	ix
DAFTAR TABEL .....	x
DAFTAR LAMPIRAN.....	xi
DAFTAR ISTILAH, LAMBANG, DAN SINGKATAN.....	xii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah .....	3
1.3 Tujuan Penelitian .....	4
1.4 Manfaat Penelitian .....	4
BAB II TINJAUAN PUSTAKA .....	5
2.1 Kajian Penelitian Terdahulu .....	5
2.2 Kajian Teori.....	7
2.2.1 Klasifikasi Citra .....	7
2.2.2 Eksplorasi Data.....	8
2.2.3 <i>Preprocessing</i> .....	9
2.2.4 <i>Neural Network</i> .....	12
2.2.5 <i>Convolutional Neural Network (CNN)</i> .....	16
2.2.6 <i>Vision Transformer (ViT)</i> .....	21
2.2.7 Hibrida CNN-ViT.....	29
2.2.8 Evaluasi Model .....	31
2.2.9 Sampah dan Jenisnya.....	32
BAB III METODE PENELITIAN .....	34
3.1 Sumber Data.....	34
3.2 Variabel Penelitian.....	34
3.3 Prosedur Penelitian.....	34
BAB IV HASIL DAN PEMBAHASAN.....	39
4.1 Eksplorasi Data.....	39
4.2 <i>Preprocessing</i> .....	40
4.2.1 <i>Resize</i> .....	40
4.2.2 <i>Data Splitting</i> .....	41
4.2.3 Augmentasi Citra.....	42
4.2.4 Pemetaan Citra ke Representasi Numerik.....	43
4.2.5 Standardisasi.....	44
4.3 Pembentukan <i>Batch</i> .....	47
4.4 Pembentukan Model .....	47
4.4.1 Model CNN dengan Arsitektur MiniVGGNet.....	47
4.4.2 Model ViT dengan Konfigurasi ViT- <i>Tiny</i> .....	49

4.4.3 Model Hibrida CNN-ViT .....	51
4.5 <i>Training</i> dan <i>Validation</i> .....	52
4.6 <i>Testing</i> .....	54
4.7 Perbandingan Model .....	58
4.8 Interpretasi Hasil.....	62
4.8.1 Interpretasi Performa Model ViT .....	62
4.8.2 Interpretasi Performa Model CNN .....	63
4.8.3 Interpretasi Performa Model Hibrida CNN-ViT .....	65
BAB V PENUTUP.....	67
5.1 Kesimpulan .....	67
5.2 Saran.....	67
5.2.1 Saran untuk Peneliti Berikutnya.....	67
5.2.2 Saran untuk Industri Teknologi .....	68
5.2.3 Saran untuk Pemangku Kepentingan Kebijakan Pengelolaan Sampah.....	68
DAFTAR PUSTAKA .....	69
LAMPIRAN .....	76
RIWAYAT HIDUP.....	93

## DAFTAR GAMBAR

Gambar 1.1	Ilustrasi Perbandingan Cara Kerja CNN dan ViT.....	2
Gambar 2.1	Ilustrasi Proses Klasifikasi Citra .....	8
Gambar 2.2	Contoh Diagram Lingkaran.....	8
Gambar 2.3	Contoh Histogram.....	9
Gambar 2.4	Struktur Umum <i>Multilayer Perceptron</i> .....	13
Gambar 2.5	Arsitektur Umum <i>Convolutional Neural Network</i> .....	17
Gambar 2.6	Contoh Perhitungan pada Operasi Konvolusi .....	19
Gambar 2.7	Contoh Perhitungan pada Operasi <i>Max Pooling</i> .....	20
Gambar 2.8	Arsitektur Umum <i>Vision Transformer</i> .....	22
Gambar 2.9	Alur Pemrosesan pada <i>Transformer Encoder</i> .....	24
Gambar 2.10	Arsitektur Hibrida CNN-ViT .....	30
Gambar 2.11	Contoh Sampah Organik .....	33
Gambar 2.12	Contoh Sampah yang Dapat Didaur Ulang.....	33
Gambar 3.1	Alur Penelitian .....	38
Gambar 4.1	Perbandingan Proporsi Jumlah Citra Sampah Berdasarkan Kelas .....	39
Gambar 4.2	Distribusi (a) Tinggi Citra dan (b) Lebar Citra.....	39
Gambar 4.3	Perbandingan Citra Sebelum dan Sesudah <i>Resize</i> .....	41
Gambar 4.4	Perbandingan Citra <i>Training</i> Sebelum dan Sesudah Augmentasi.....	42
Gambar 4.5	Contoh Citra pada Kelompok (a) <i>Training</i> , (b) <i>Validation</i> , dan (c) <i>Testing</i> Sebelum Representasi Numerik.....	43
Gambar 4.6	Citra pada Kelompok <i>Training</i> Sebelum Augmentasi .....	45
Gambar 4.7	Perubahan (a) <i>Loss</i> dan (b) Akurasi Model Hibrida CNN-ViT	52
Gambar 4.8	Contoh Citra Data <i>Testing</i> Kelas <i>Organic</i> yang Diklasifikasi dengan Benar oleh Model Hibrida CNN-ViT pada <i>Epoch</i> Terbaik.....	54
Gambar 4.9	Contoh Citra Data <i>Testing</i> Kelas <i>Recyclable</i> yang Diklasifikasi dengan Benar oleh Model Hibrida CNN-ViT pada <i>Epoch</i> Terbaik.....	54
Gambar 4.10	Contoh Citra Data <i>Testing</i> Kelas <i>Organic</i> yang Diklasifikasi sebagai <i>Recyclable</i> oleh Model Hibrida CNN-ViT pada <i>Epoch</i> Terbaik.....	55
Gambar 4.11	Contoh Citra Data <i>Testing</i> Kelas <i>Recyclable</i> yang Diklasifikasi sebagai <i>Organic</i> oleh Model Hibrida CNN-ViT pada <i>Epoch</i> Terbaik.....	55
Gambar 4.12	<i>Confusion Matrix</i> Hasil Klasifikasi Model Hibrida CNN-ViT dengan <i>Epoch</i> Terbaik pada Data <i>Testing</i> .....	56
Gambar 4.13	Perbandingan Model Menurut Perubahan (a) <i>Training Loss</i> , (b) <i>Validation Loss</i> , (c) Akurasi <i>Training</i> , dan (d) Akurasi <i>Validation</i> .....	59

## DAFTAR TABEL

Tabel 2.1	Penelitian Terdahulu.....	5
Tabel 2.2	<i>Confusion Matrix</i> .....	32
Tabel 3.1	Variabel Data Pengamatan.....	34
Tabel 4.1	Distribusi Jumlah Citra Hasil <i>Data Splitting</i> .....	41
Tabel 4.2	Ringkasan Parameter Standardisasi per Kanal .....	45
Tabel 4.3	Arsitektur CNN MiniVGGNet yang Digunakan .....	49
Tabel 4.4	Konfigurasi Model ViT- <i>Tiny</i> yang Digunakan.....	50
Tabel 4.5	Arsitektur ViT dengan Konfigurasi ViT- <i>Tiny</i> yang Digunakan .	50
Tabel 4.6	Arsitektur Model Hibrida CNN-ViT yang Digunakan .....	51
Tabel 4.7	Performa <i>Training</i> dan <i>Validation</i> Model Hibrida CNN-ViT .....	53
Tabel 4.8	Performa <i>Testing</i> Model Hibrida CNN-ViT .....	57
Tabel 4.9	Perbandingan Model Menurut Performa <i>Training</i> dan <i>Validation</i> .....	60
Tabel 4.10	Perbandingan Model Menurut Performa <i>Testing</i> .....	61

## DAFTAR LAMPIRAN

Lampiran 1. Tampilan Citra Asli .....	76
Lampiran 2. Tampilan Citra Setelah <i>Resize</i> .....	77
Lampiran 3. <i>Syntax</i> Python untuk Seluruh Prosedur Analisis .....	78

PRODI STATISTIKA

## DAFTAR ISTILAH, LAMBANG, DAN SINGKATAN

$A$	<i>Array</i> berukuran $C \times H \times W$ hasil standardisasi dari <i>array</i> representasi citra yang telah di- <i>resize</i> dan diaugmentasi
$A^{(conv)}$	<i>Array feature map</i> hasil konvolusi pada CNN berukuran $N_c \times H_c \times W_c$
$A^{(pool)}$	<i>Array</i> hasil <i>max pooling</i> pada CNN berukuran $N_c \times H_p \times W_p$
<i>Array</i>	Struktur data terstruktur untuk menyimpan sekumpulan nilai bertipe sama dalam satu variabel, diakses melalui indeks, dan dapat berbentuk satu atau beberapa dimensi
<i>Batch</i>	Sekumpulan data yang diproses sekaligus dalam satu langkah pembaruan bobot saat proses <i>training</i> pada model
CNN	<i>Convolutional neural network</i>
$D_E$	Dimensi vektor <i>embedding</i> pada ViT
<i>Epoch</i>	Satu siklus penuh di mana seluruh data <i>training</i> diproses sekali oleh model
$L$	Banyaknya blok <i>encoder</i> pada ViT
<i>Learnable</i>	Parameter dalam <i>neural network</i> yang nilainya diperbarui selama proses <i>training</i> untuk meminimalkan kesalahan prediksi
MHA	<i>Multi-head attention</i>
MLP	<i>Multi layer perceptron</i>
$N_p$	Banyaknya <i>patch</i> yang dihasilkan pada ViT
<i>Padding</i>	Proses menambahkan bingkai nol di sekitar citra
RGB	<i>Red, green, and blue</i>
<i>Stride</i>	Jumlah langkah pergeseran <i>kernel</i> saat melakukan operasi konvolusi
<i>Token</i>	Unit terkecil dari data (teks atau citra) yang direpresentasikan sebagai vektor untuk diproses oleh model
<i>Token</i> [CLS]	<i>Classification token</i>
ViT	<i>Vision Transformer</i>
$X$	<i>Array</i> citra awal berukuran $C \times H \times W$ hasil <i>resize</i> dan augmentasi
$\hat{Y}_{CNN}$	Vektor probabilitas prediksi model CNN untuk kelas positif
$\hat{Y}_{ViT}$	Vektor probabilitas prediksi model ViT untuk kelas positif
$\hat{Y}_{HYB}$	Vektor probabilitas prediksi model hibrida CNN-ViT untuk kelas positif
$Z^{(l)}$	Matriks berukuran $(N_p + 1) \times D_E$ <i>output</i> blok <i>encoder</i> ke- $l$ pada ViT
$z_0$	Vektor <i>token</i> [CLS] berdimensi $D_E$ setelah melalui seluruh <i>encoder</i> ViT