

# FROM NOISY TO NOTABLE: RUMAH ADAT CLASSIFIER WITH SMART CLEANING

pHash de-dup & anti-leakage → EVA-02 fine-tuning with Focal Loss,  
Weighted Sampler & TTA

Tim wets belom dicoba

Evans Kizito

Franklin Daniel Situmorang

Bryant Farrel Titanius

# Pelestarian Rumah Adat Nusantara lewat Computer Vision



Identifikasi **gaya arsitektur** rumah adat **penting** untuk pelestarian budaya, edukasi, dan pengarsipan digital. **Tantangan utama:** kemiripan visual antarkelas (ornamen/atap), variasi sudut & pencahayaan, serta ukuran **dataset yang terbatas**.

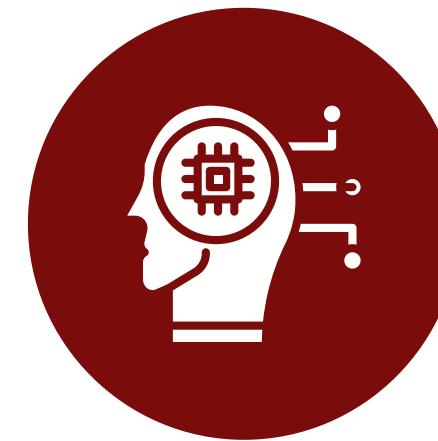
## Tujuan Penelitian



Mengembangkan sistem berbasis *Deep Learning* untuk mengidentifikasi 5 kategori rumah adat Nusantara dengan generalisasi tinggi.

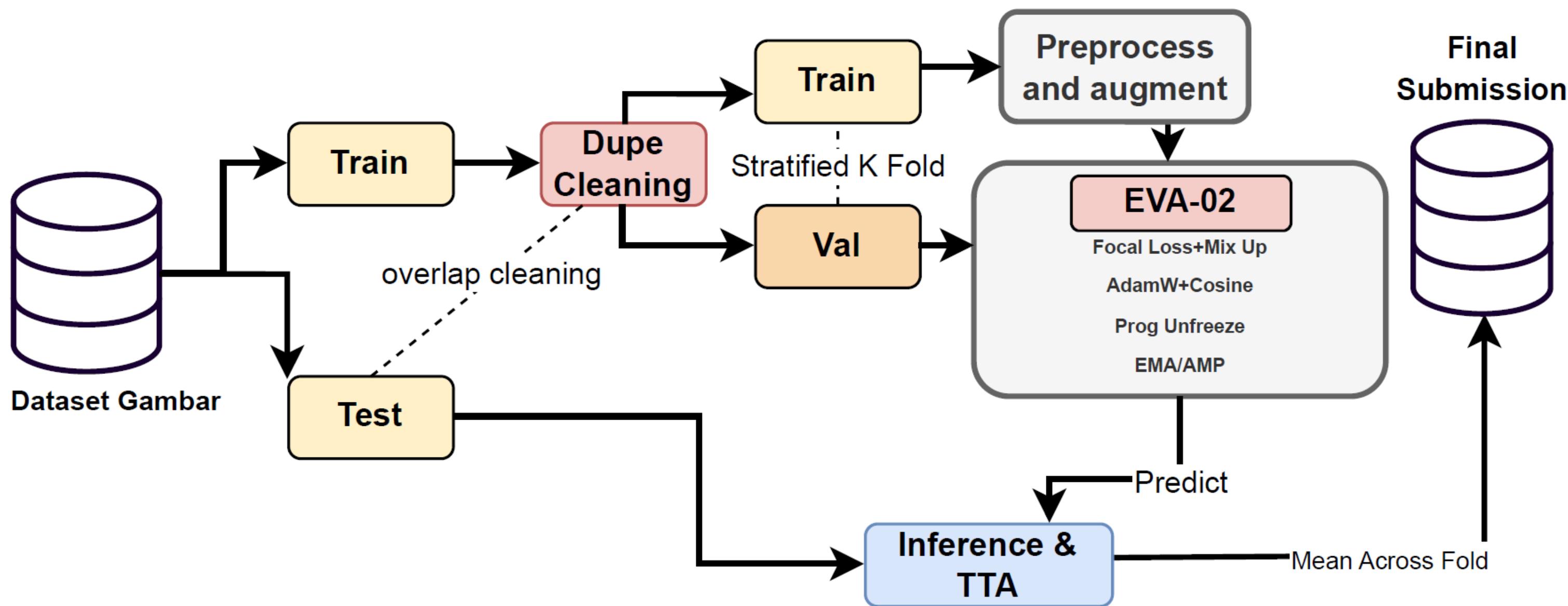


Mengevaluasi efektivitas model EVA-02 dalam menangani variabilitas dan kompleksitas visual arsitektur tradisional.



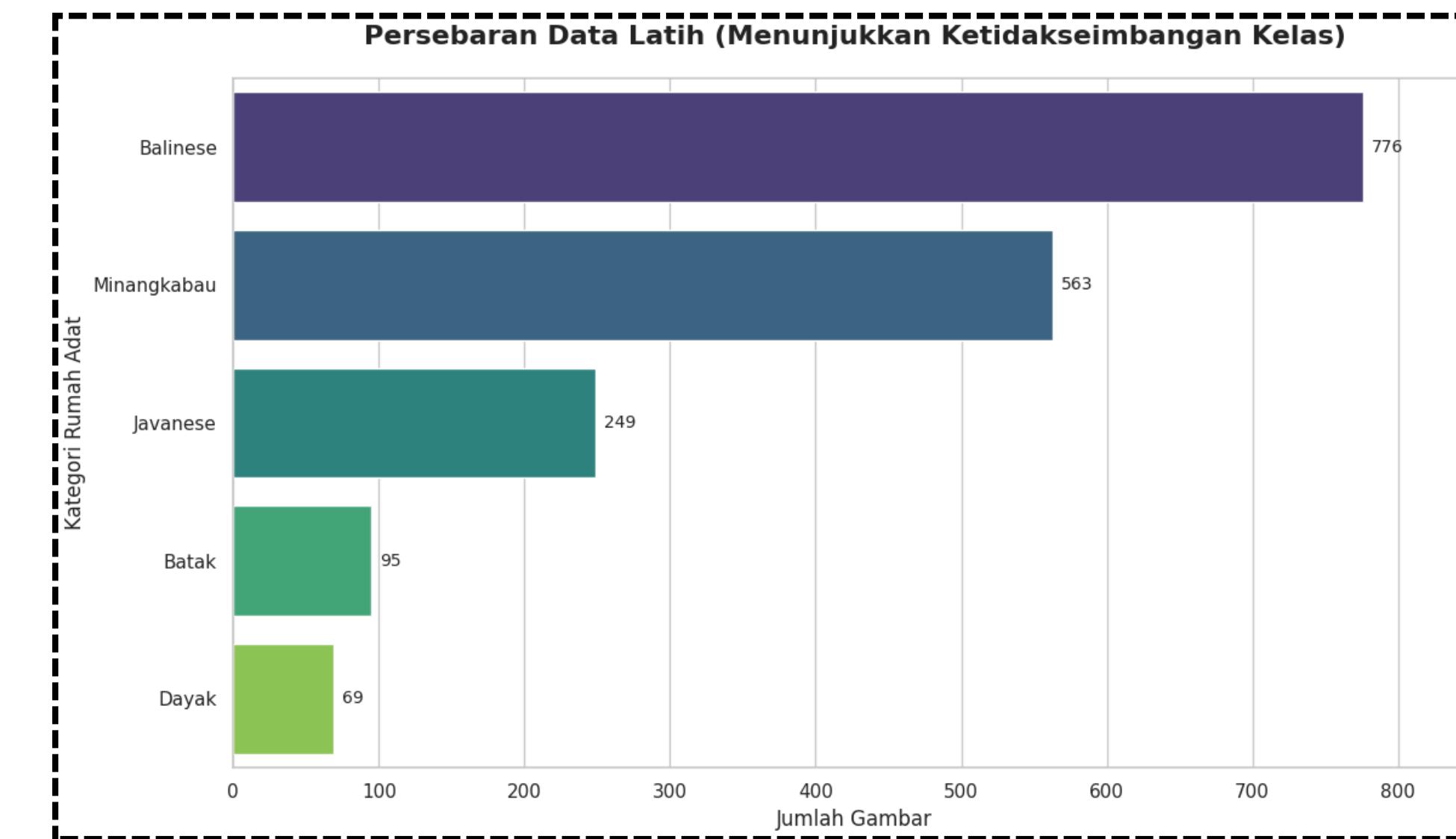
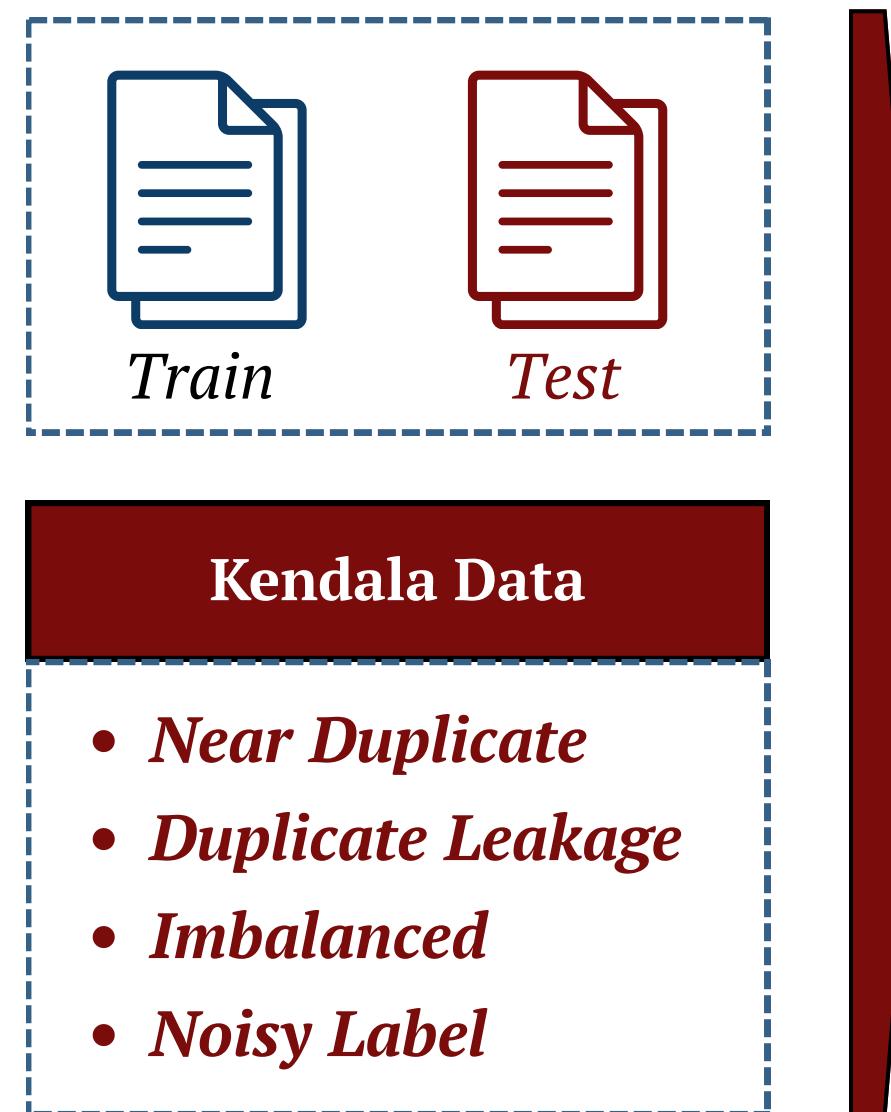
Menerapkan strategi optimasi untuk memaksimalkan *Macro F1-Score* pada dataset yang tidak seimbang.

## Alur Penelitian



## Deskripsi Dataset

Dataset terdiri dari **1752 gambar latih (*train*)** dan **444 gambar uji (*test*)** yang terbagi ke dalam lima kelas: **Balinese, Batak, Dayak, Javanese, dan Minangkabau.**



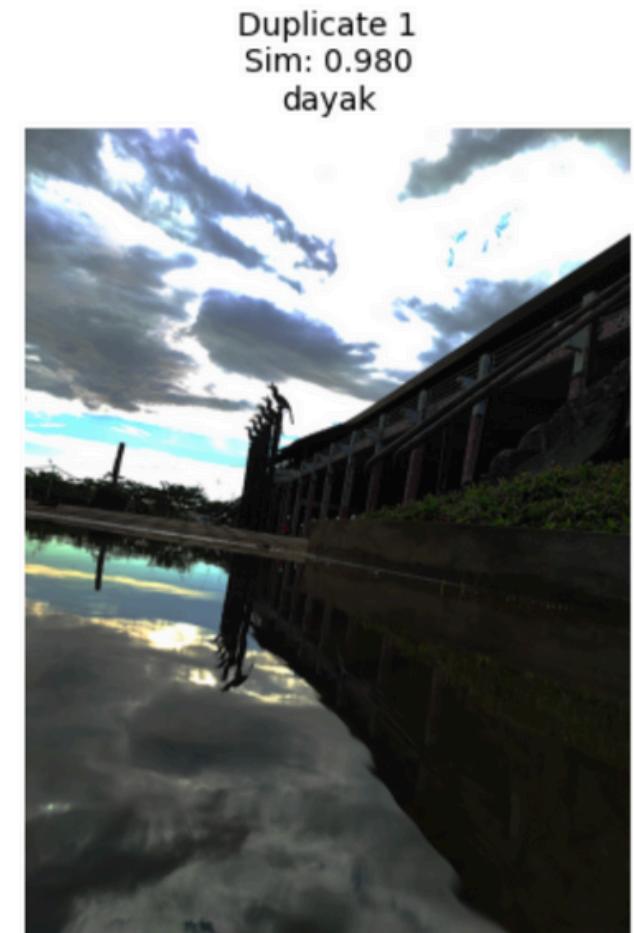
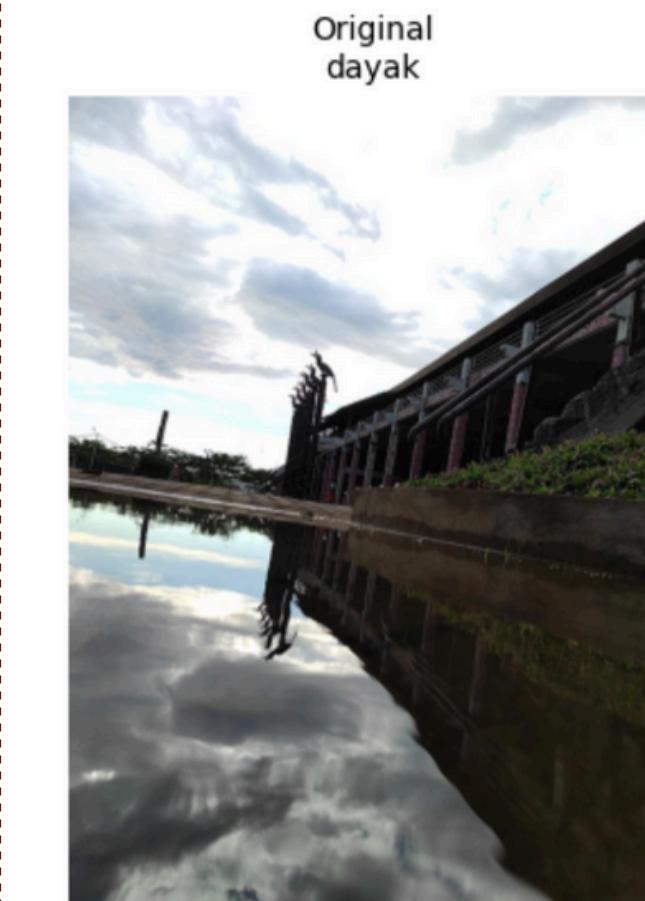
## Kendala Dataset

Gambar yang sama di **dua kelas berbeda** dapat menimbulkan **noise** pada Training model. Selain itu, adanya **gambar yang near duplicate** di dalam kelas maupun di **data test** juga dapat mempengaruhi **kualitas model**.

**Konflik antar kelas**



**Duplikat antar kelas**



## Karakteristik Dataset

Gambar yang sama di **dua kelas berbeda** dapat menimbulkan **noise** pada Training model. Selain itu, adanya **gambar yang near duplicate** di dalam kelas maupun di **data test** juga dapat mempengaruhi **kualitas model**.

Noisy Label

True: batak  
Pred: balinese



Duplikat antar train dan test

Train: batak\_train\_000094.jpg  
Original Class: batak

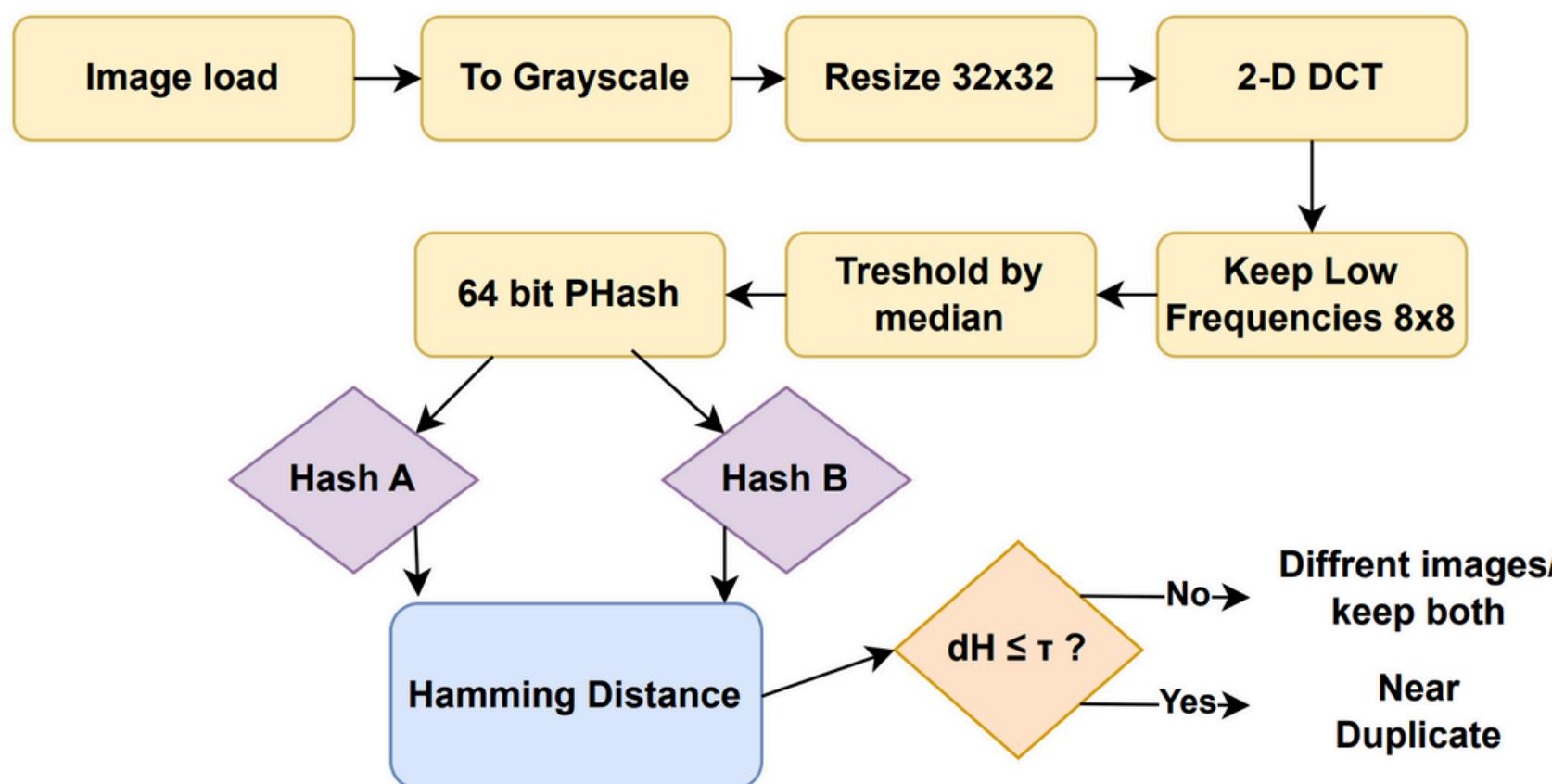


Test: Test\_296.jpg



## Pembersihan data duplikat

### Perceptual Hash (pHash)



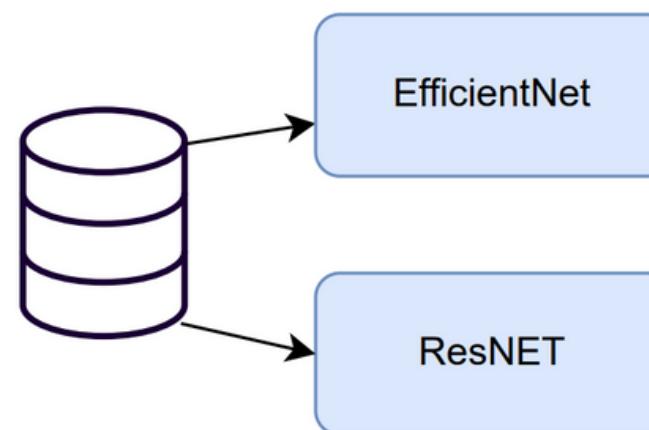
cross class conflict :68 → Keep True class : 12  
 near duplicate with test :31  
 single class duplicate :26

### Data Train Final

Before → After  
 1752 → 1639

## CNN Baseline (*EfficientNet-B0* dan *ResNet-50*)

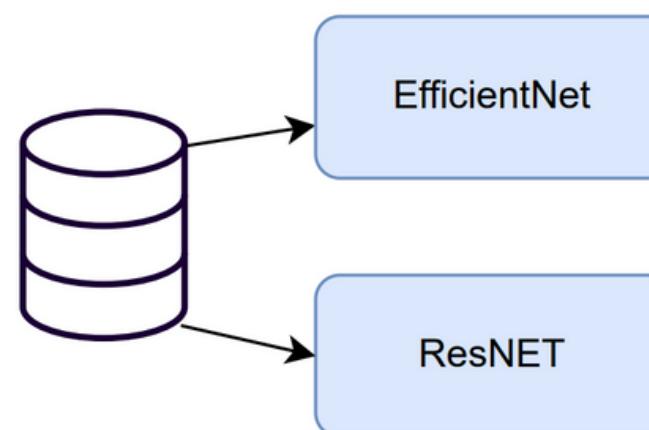
### Macro F1 (tanpa preprocessing)



0.52

0.54

### Macro F1 (dengan preprocessing)



0.65

0.66

### Problem Identification and Analysis

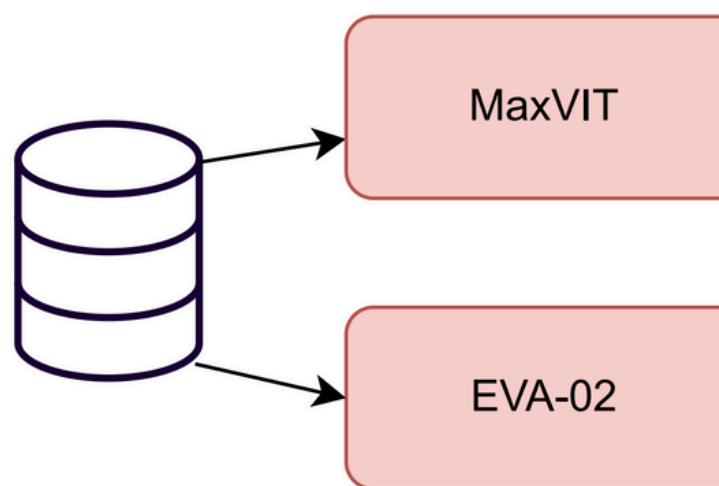
- 1 Arsitektur lokal → sulit menangkap relasi jarak jauh (proporsi bangunan, bentuk atap “gonjong” vs “joglo”).
- 2 Background shift tinggi (langit, pepohonan, wisatawan) → fitur tekstur mudah terdistraksi dan meski pakai WeightedRandomSampler/Focal, fitur lokal belum cukup memisahkan kelas mirip.

### Proposed Solution → Transformer Model. WHY?

- Self-attention = konteks global: setiap patch “melihat” patch lain → peka bentuk/siluet & tata massa.
- Multi-scale & hierarkis (MaxViT: local-window + grid/global; EVA-02: ViT modern + pretrain besar) → gabungkan detail lokal + struktur global.
- Pretraining skala besar → lebih robust terhadap variasi sudut, pencahayaan, background.
- Umumnya lebih merata per kelas → dorong kenaikan Macro-F1, bukan hanya akurasi.

## Transformer Model (*MaxVIT and EVA02*)

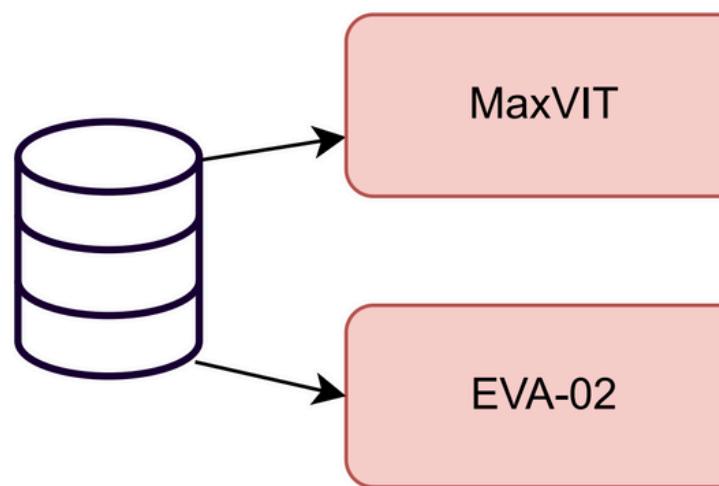
### Macro F1 (tanpa TTA,Focal Loss dll)



0.80

0.88

### Macro F1 (dengan TTA,Focal loss dll)

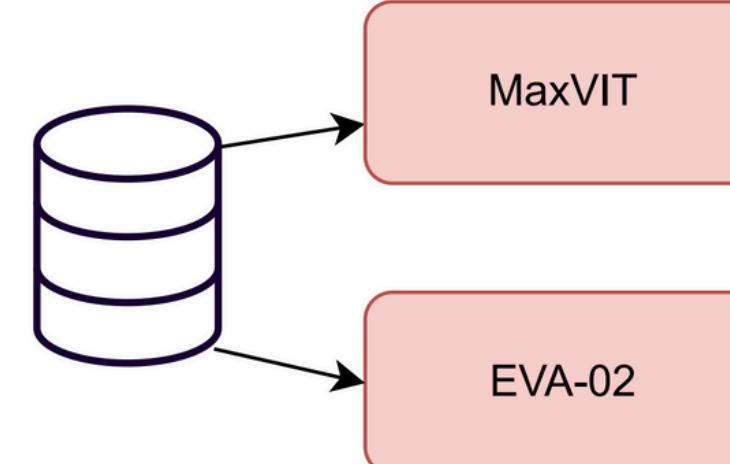


0.82

0.90



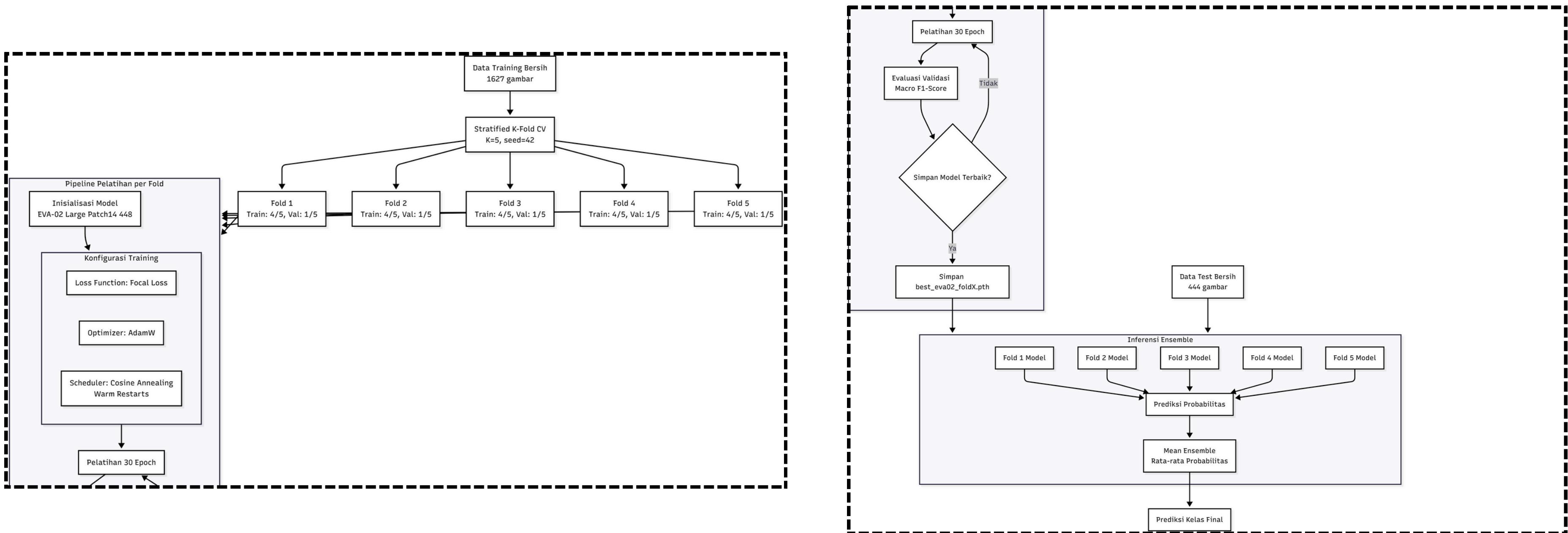
### Macro F1 (TTA,Focal loss+Preprocess)



0.87

0.98

## Pelatihan Model

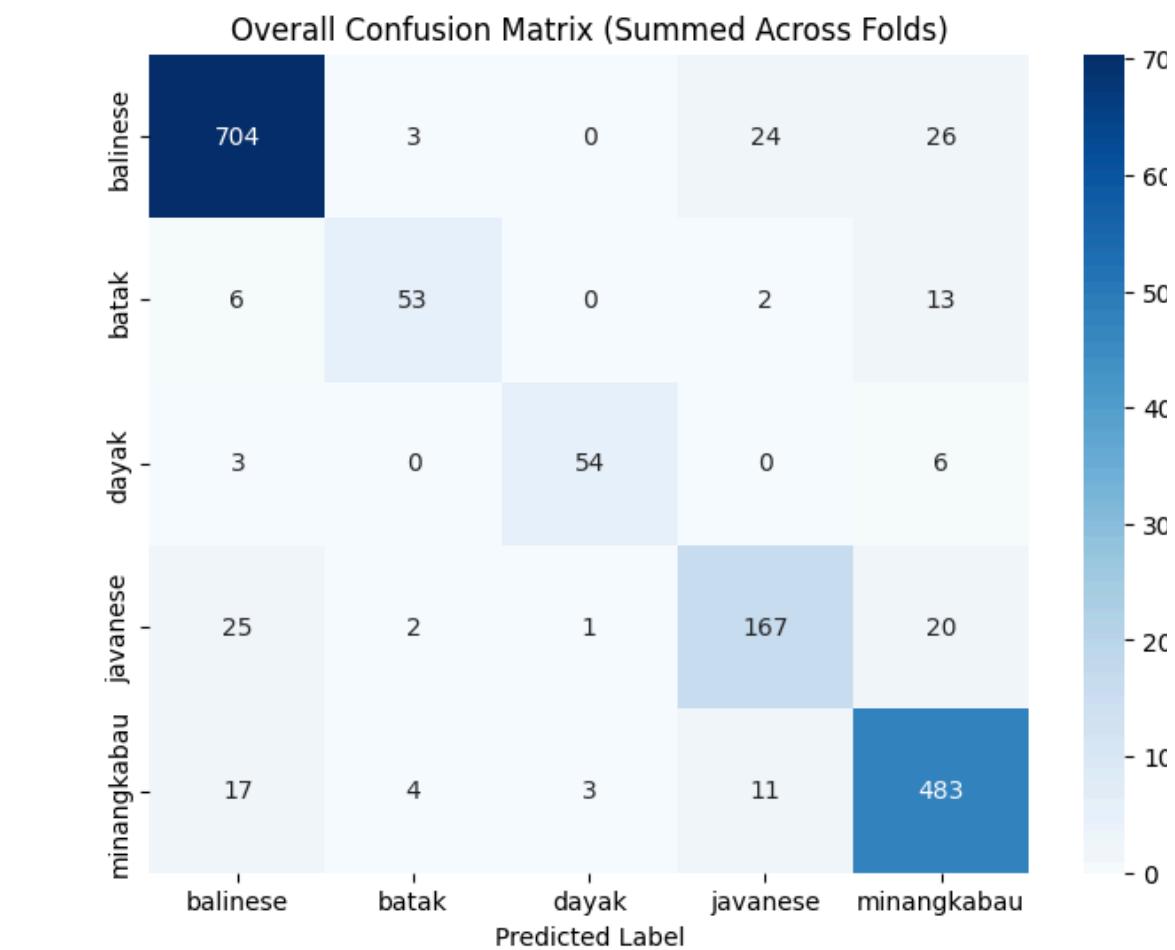


# Hasil akhir

Nama Model	Pre-Processing	Public Macro F1	Private Macro F1
EfficientNet Baseline	<b>False</b>	0.52312	0.48821
EfficientNet + Transform	<b>False</b>	0.74703	0.73107
ConvNeXt Baseline	<b>False</b>	0.77987	0.74005
ConvNeXt - Augmented + LDAM	<b>False</b>	0.82031	0.82586
Swin-Transformer Baseline	<b>False</b>	0.77482	0.69325
EVA-02 Baseline	<b>False</b>	0.86177	0.89046
EVA-02 + Pipeline	<b>False</b>	0.87174	0.90179
<b>EVA-02 + Pipeline + Pre-Processed</b>	<b>True</b>	<b>0.92937</b>	<b>0.98653</b>

## Evaluasi Metriks

class	precision	recall	f1_score
	0.932	0.932	0.932
balinese	0.932	0.932	0.932
batak	0.868	0.718	0.781
dayak	0.934	0.854	0.891
javanese	0.828	0.774	0.791
minangkabau	0.888	0.930	0.902



### Interpretasi:

- Balinese dan Minangkabau menunjukkan precision dan recall yang tinggi serta konsisten di semua fold.
- Batak dan Javanese mengalami fluktuasi dan penurunan F1-Score, mengindikasikan model lebih sering keliru.
- Kesalahan diduga kuat berasal dari kemiripan elemen arsitektural antar kelas dan variasi kualitas gambar yang masih menjadi tantangan.

- Balinese dan Minangkabau menunjukkan diagonal matriks yang kuat, mengonfirmasi performa klasifikasi yang andal.
- Kebingungan terbesar terjadi antara kelas Javanese dan Balinese, serta pada beberapa sampel Batak, yang mengindikasikan tumpang tindih fitur visual.

## Kesimpulan



Sistem berhasil dikembangkan dengan performa tinggi, membuktikan efektivitas EVA-02 dalam mengklasifikasikan rumah adat.



Keberhasilan didorong oleh strategi komprehensif mulai dari pembersihan data, penanganan ketidakseimbangan kelas, hingga validasi yang ketat.

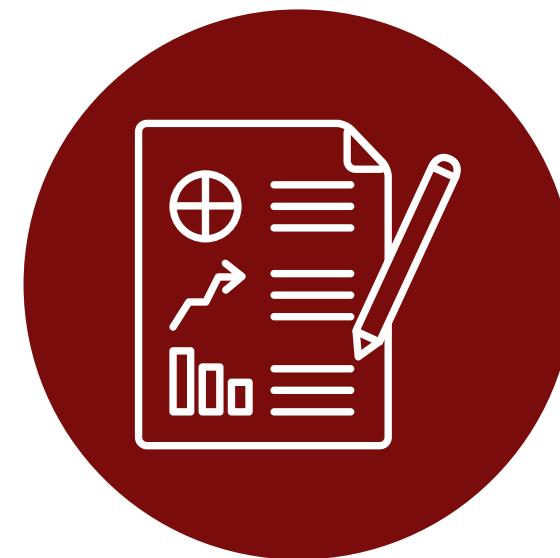


Analisis mengungkap tantangan utama pada kebingungan model antara kelas dengan kemiripan visual tinggi, seperti Jawa dan Bali.

## Rekomendasi



**Eksplorasi teknik Fine-Grained Visual Classification untuk meningkatkan kemampuan membedakan kelas yang mirip secara visual.**



**Melakukan verifikasi label manual untuk mengatasi noise dan ambiguitas dalam dataset.**



**Menerapkan Test-Time Augmentation yang lebih kompleks untuk meningkatkan stabilitas prediksi pada data baru.**

# Terima Kasih

© Tim wets belom dicoba