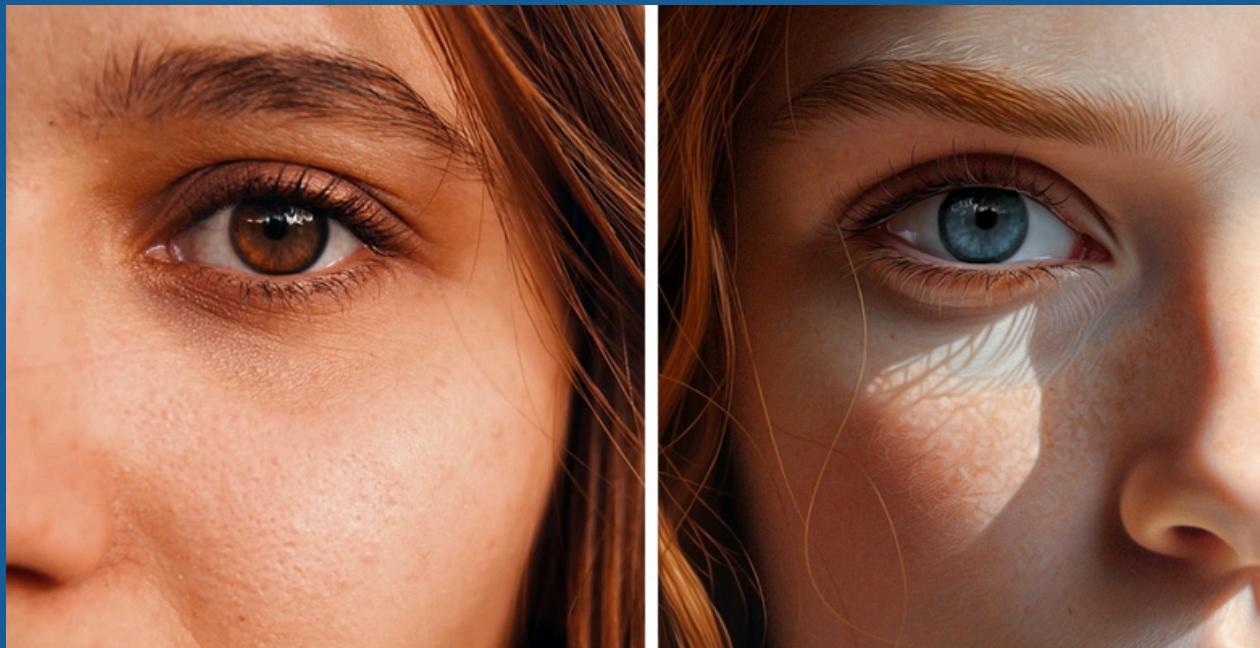


ART DETECTIVE

Học Máy - Semester 242

GIỚI THIỆU BÀI TOÁN

**Hình nào sau đây là ảnh thật?
Hình nào sau đây là ảnh sinh ra từ AI?**



A

REAL VS. A.I. PAIR 1

B



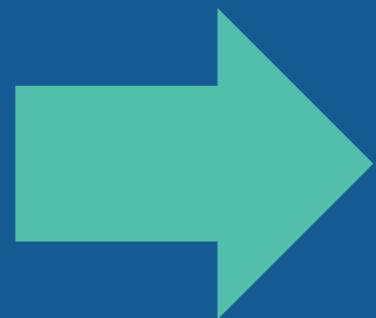
A

REAL VS. A.I. PAIR 2

B

Bài toán nghiên cứu

- Sự phát triển nhanh chóng của các mô hình sinh ảnh như GAN, Diffusion, v.v...
- Dẫn đến nhiều hệ lụy: giả mạo danh tính, lừa đảo, lan truyền thông tin sai lệch



Mục tiêu: Nghiên cứu và ứng dụng các thuật toán học máy để phân loại ảnh thật và ảnh do AI tạo ra

TIỀN XỬ LÝ DỮ LIỆU

Tập dữ liệu

- Tập dữ liệu được lấy trên Kaggle, được sử dụng trong cuộc thi Women in AI (WAI) 2025.
- Tập dữ liệu bao gồm 79.950 ảnh, được chia thành 2 tập huấn luyện (0.8) và tập kiểm tra (0.2)

The Kaggle logo is displayed in a large, blue, sans-serif font. The word "kaggle" is written in a lowercase, bold style, with a small trademark symbol (TM) at the top right of the letter "e".

Phân Tích Và Kiểm Tra Dữ Liệu

Toàn vẹn dữ liệu

- Kiểm tra giá trị Null trong dữ liệu
- Đổi chiều file CSV và ảnh có tồn tại đầy đủ không
- Kiểm tra xem số lượng ảnh thật và ảnh AI có cân bằng
- Phát hiện cặp ảnh bị thiếu hoặc dư

Đặc trưng hình ảnh

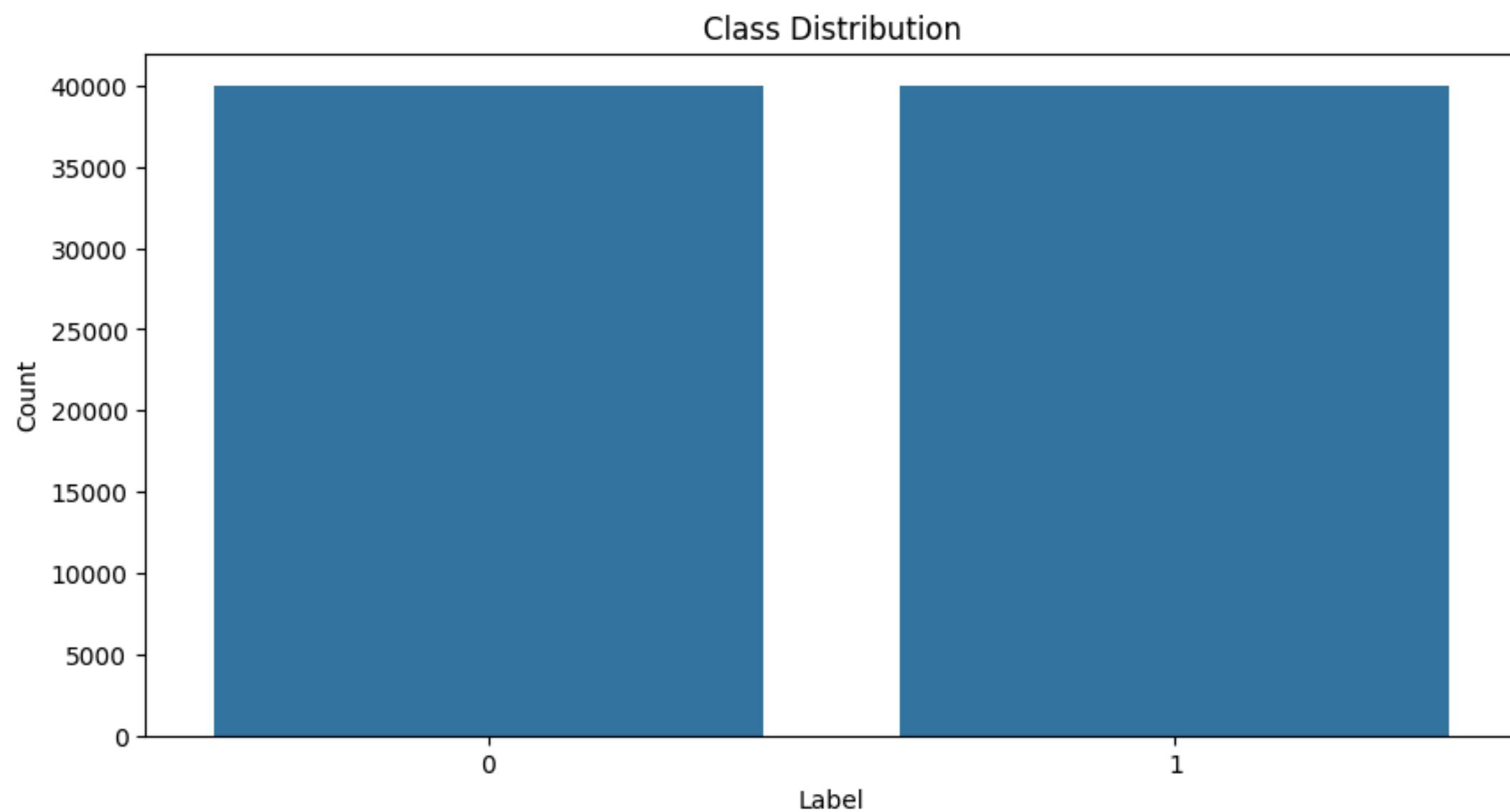
- Các ảnh có định dạng file giống nhau không?
- Kích thước ảnh gốc có đồng nhất hay không?
- Phân tích phân bố pixel giữa ảnh thật và ảnh AI để phát hiện khác biệt

```
[ ] print("Null counts: ")
print(train_df.isnull().sum())
```

→ Null counts:
id 0
label 0
dtype: int64

```
[ ] def check_file_existence(df):
missing = []
for fname in df['id']:
    if not os.path.isfile(os.path.join(DATA_DIR, fname)):
        missing.append(fname)
return missing
train_missing = check_file_existence(train_df)
print(f'Missing: {len(train_missing)}/{len(train_df)}')
```

→ Missing: 0/79950



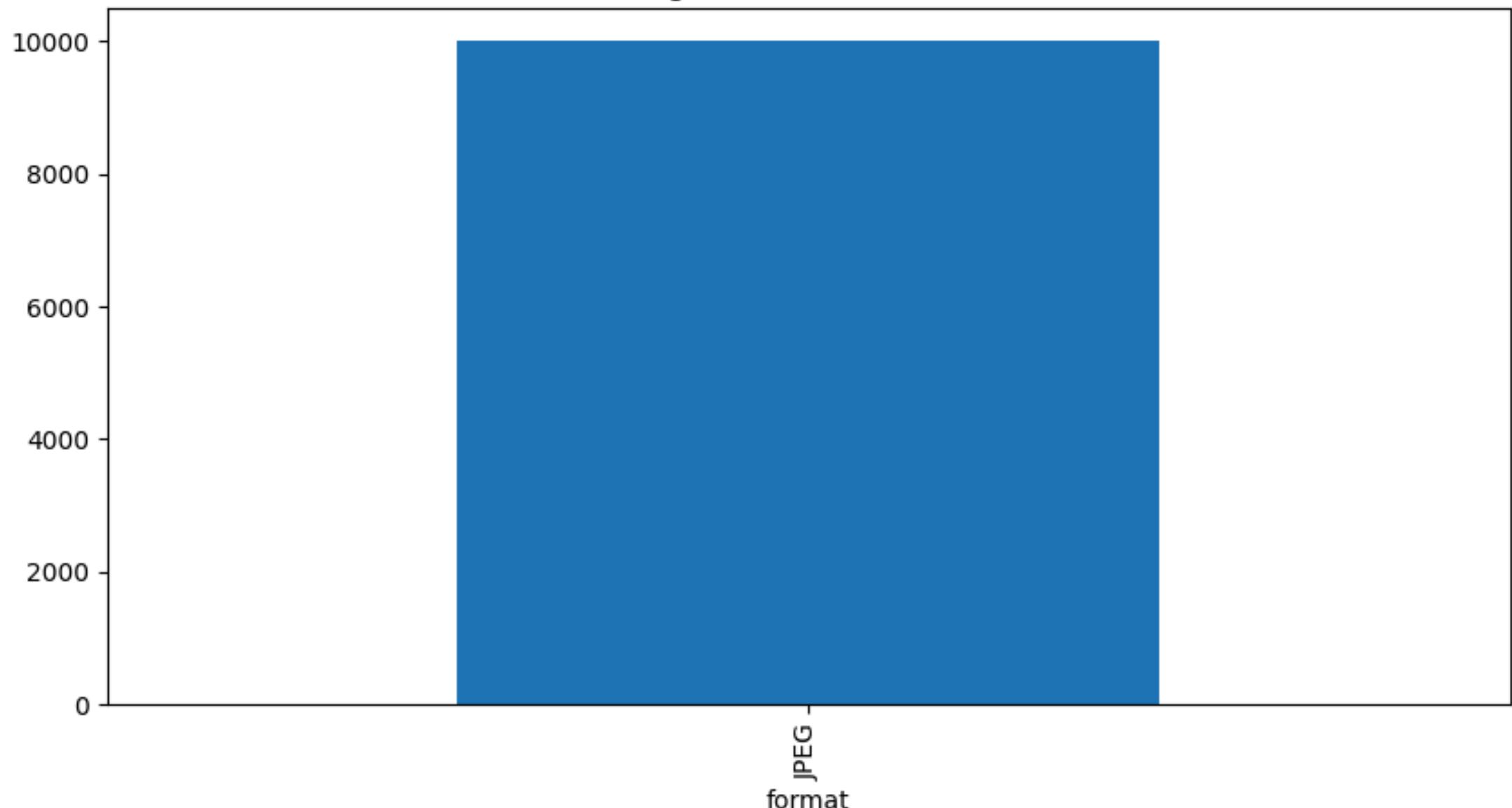
```
[ ] pair_violations = sum(train_df['label'].diff()[1::2] != -1) # think of  
print(f'Pairing violations: {pair_violations}/{len(train_df)//2}')
```

→ Pairing violations: 0/39975

```
[ ] # Pair completeness check  
pair_sizes = train_df.groupby("pair_id")["id"].count().value_counts()  
print("\nPair size distribution:")  
print(pair_sizes)
```

→
Pair size distribution:
id
2 39975
Name: count, dtype: int64

Image Format Distribution



AI-Generated Samples

Label 1
768x464



Label 1
768x416



Label 1
768x512



Label 1
768x512



Label 1
768x512



Label 1
768x512



Label 1
768x512



Label 1
768x512



Label 1
768x512



Label 1
768x512



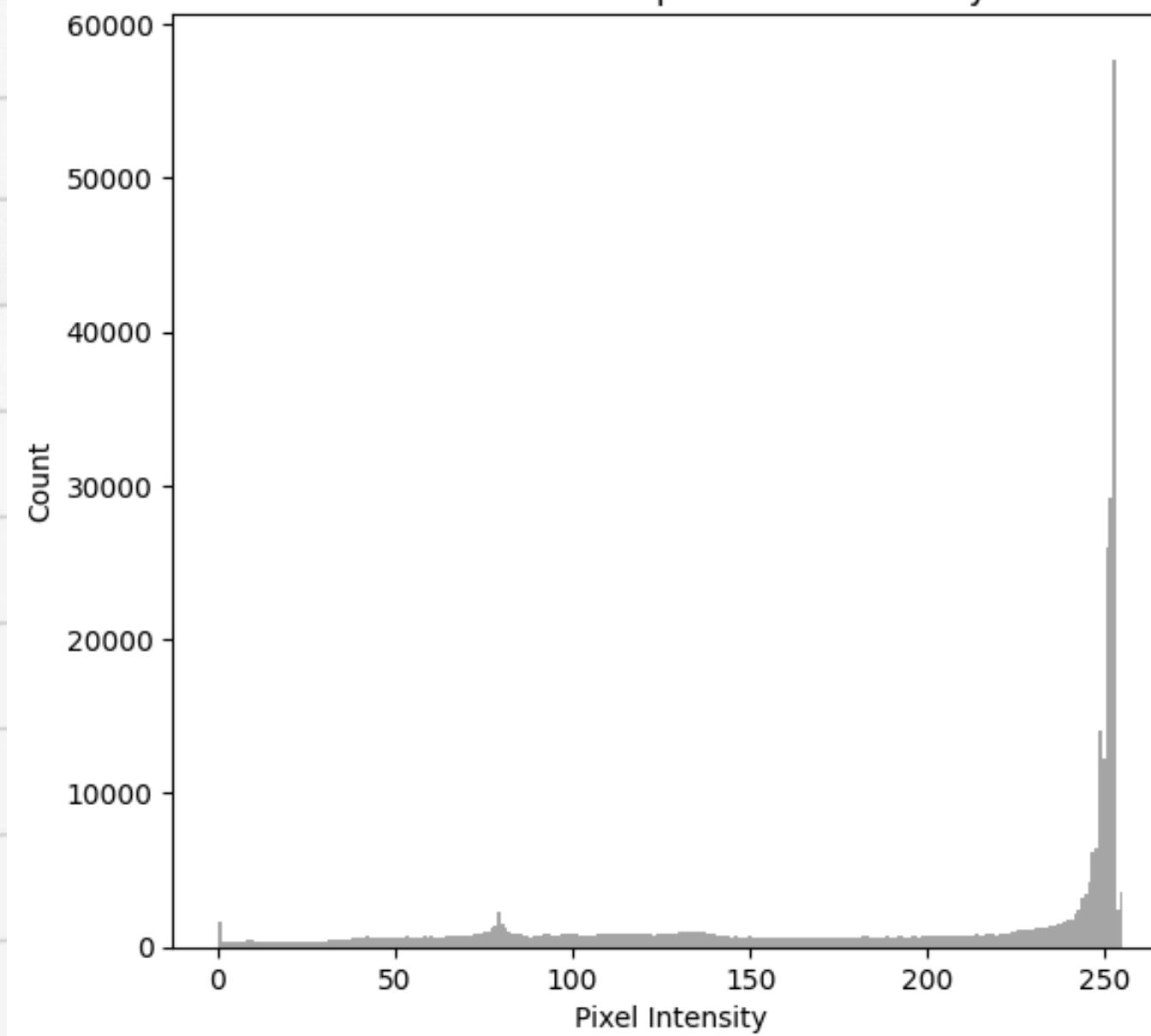
AI-Generated Samples



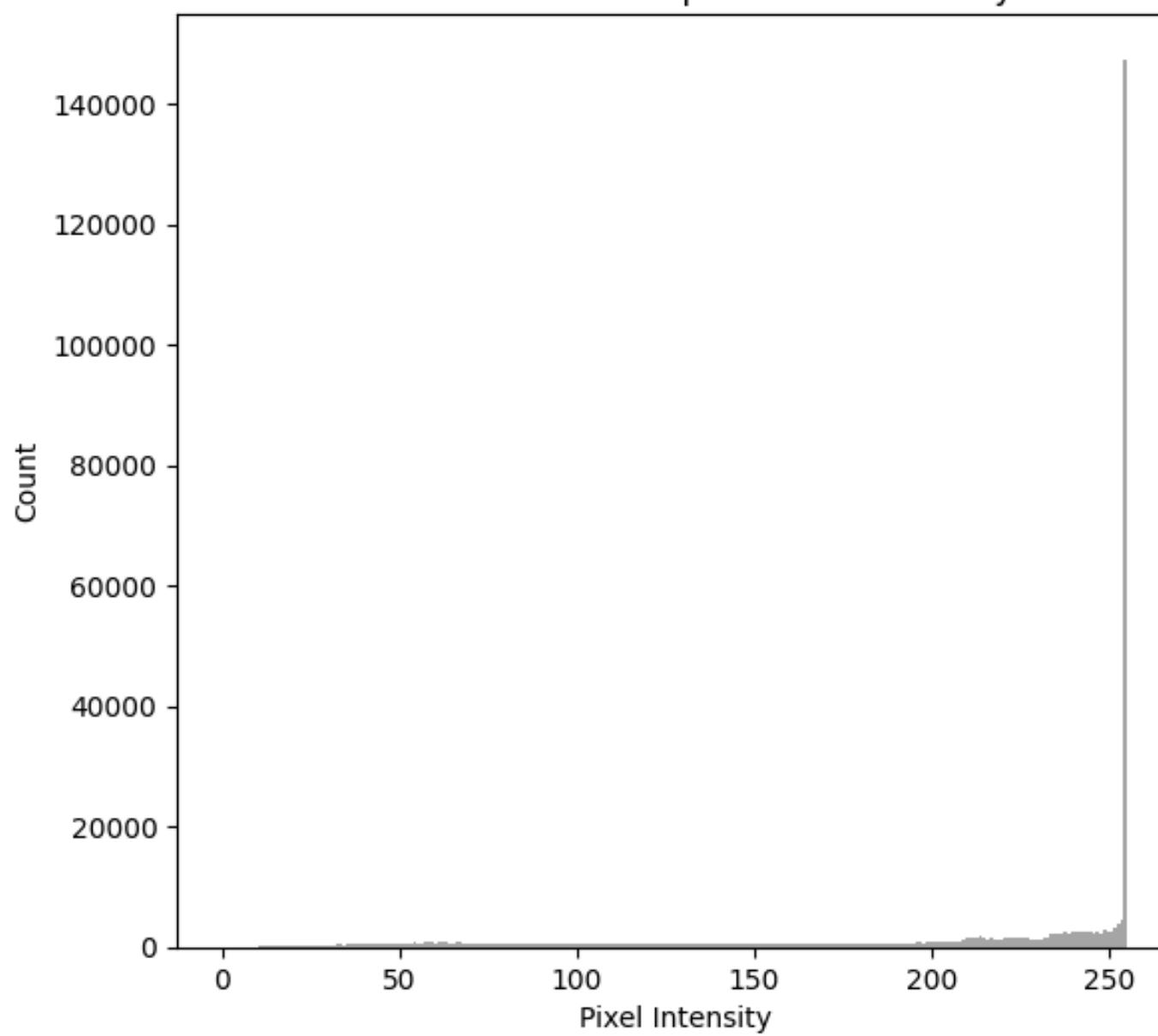
Human-Created Samples



AI-Generated Samples - Pixel Intensity



Human-Created Samples - Pixel Intensity



GIAI THUẬT HỌC MÁY

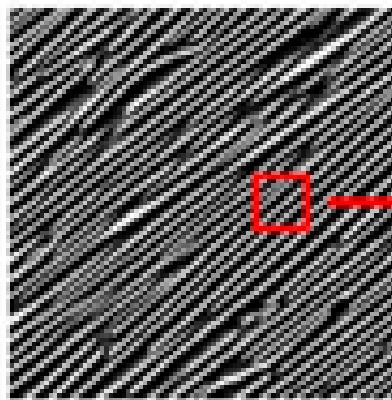
PHƯƠNG PHÁP



LOCAL BINARY PATTERNS

- Local Binary Patterns (LBP) là một kỹ thuật phân tích kết cấu hình ảnh.
- Output: Vector đặc trưng => dữ liệu đầu vào cho XGBoost

Original image



Original image
pixels

5	8	1
5	4	1
3	7	2

Binary Pattern
extraction

5	8	1
5	0	1
3	7	2

Binary Pattern

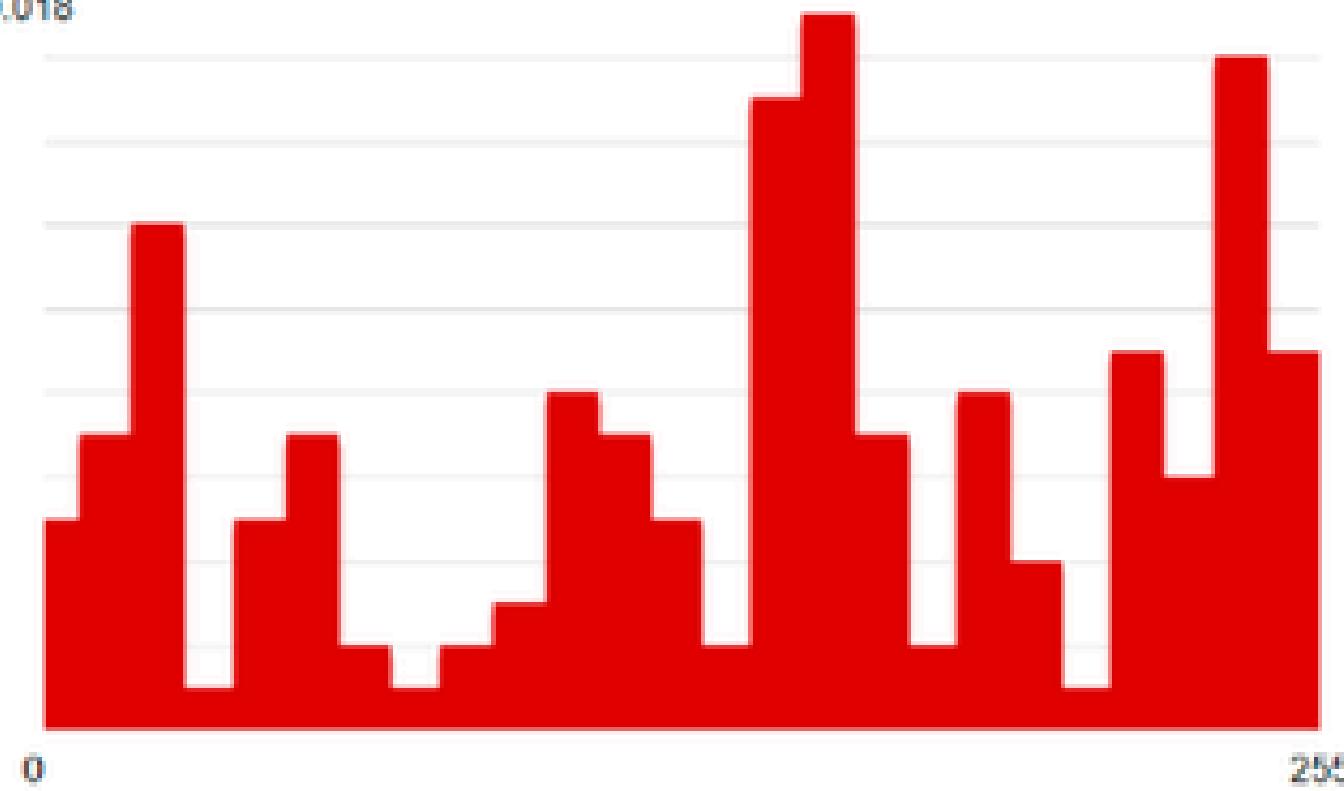
0	0	1
0	7	1
1	0	1

Decimal Value

0	0	0	1	0	1	1	1	0
2^4	2^3	2^2	2^1	2^0				

$16 + 4 + 2 + 1 = 23$

0.018



LBP Histogram = Feature Vector

Input Image

5	4	2	2	1
3	5	8	1	3
2	5	4	1	2
4	3	7	2	7
1	4	4	2	6

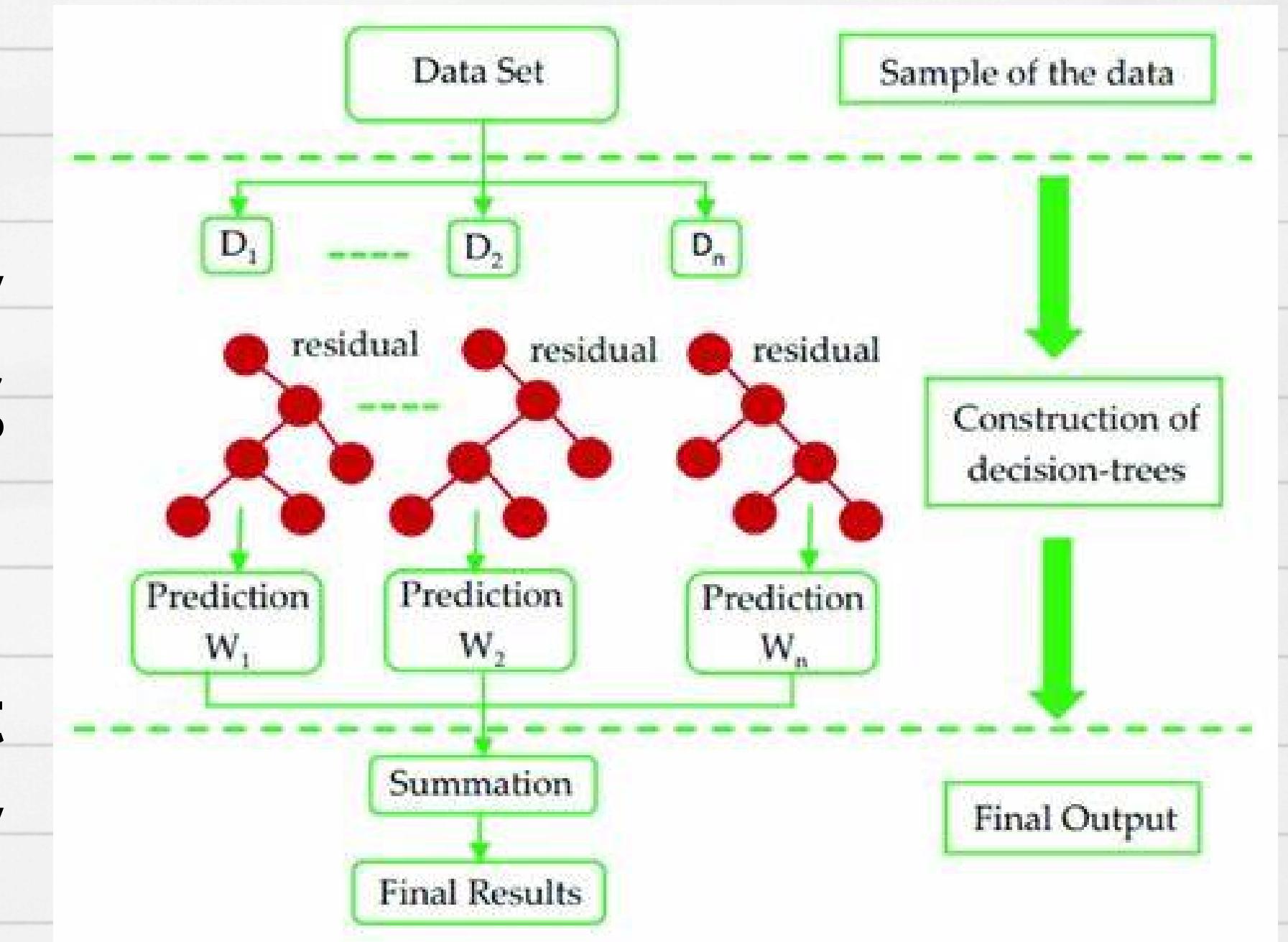
Output LBP Image

23

Storing of LBP Decimal Value in LBP Matrix

XGBOOST

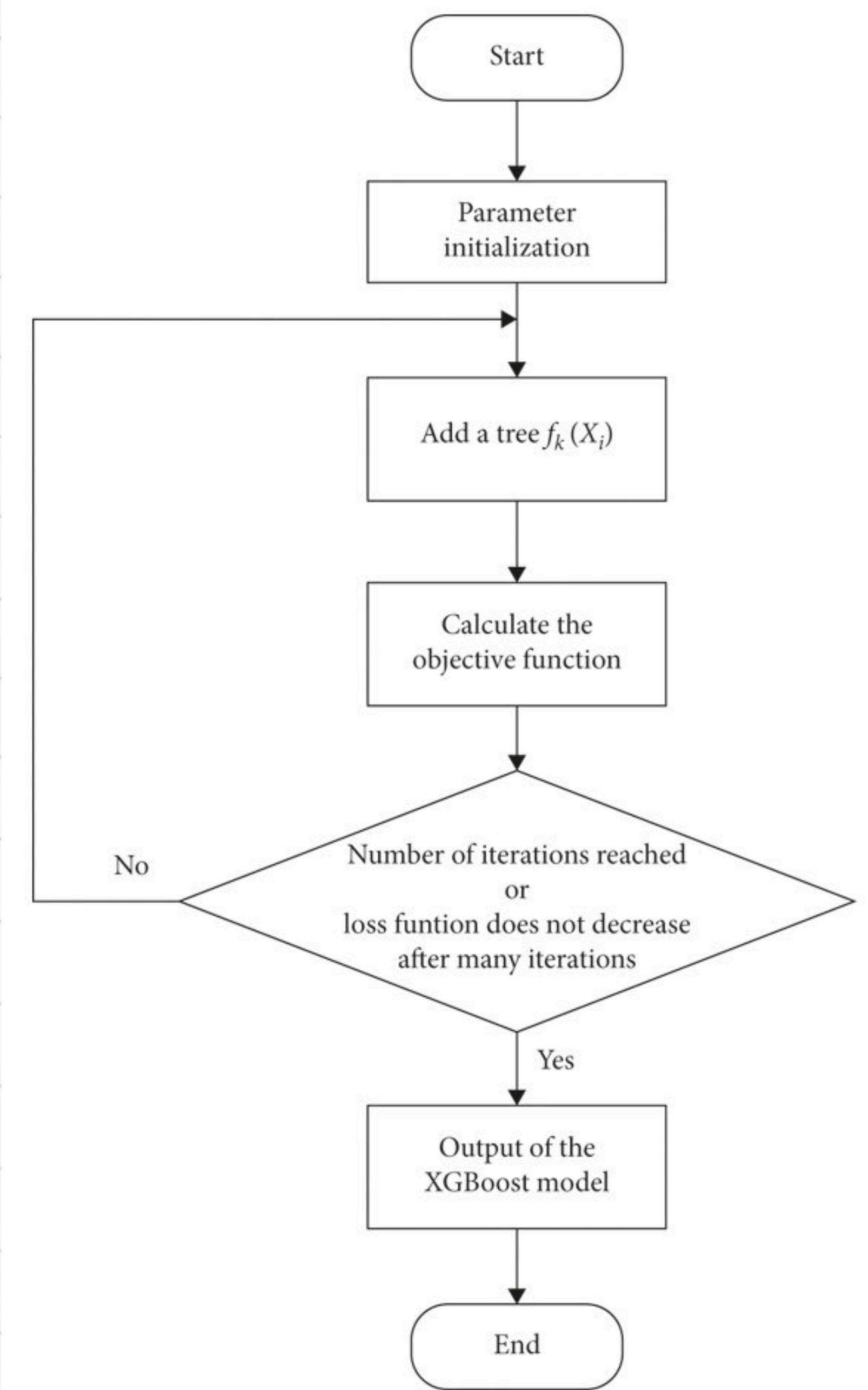
- XGBoost là một thuật toán học máy thuộc ensemble learning - phương pháp boosting.
- XGBoost kết hợp nhiều cây quyết định theo từng vòng lặp. Mỗi cây mới được huấn luyện nhằm sửa lỗi của cây trước đó, giúp mô hình học tốt hơn theo thời gian.



HÀM MỤC TIÊU TẠI ĐIỂM CỰC TIÊU

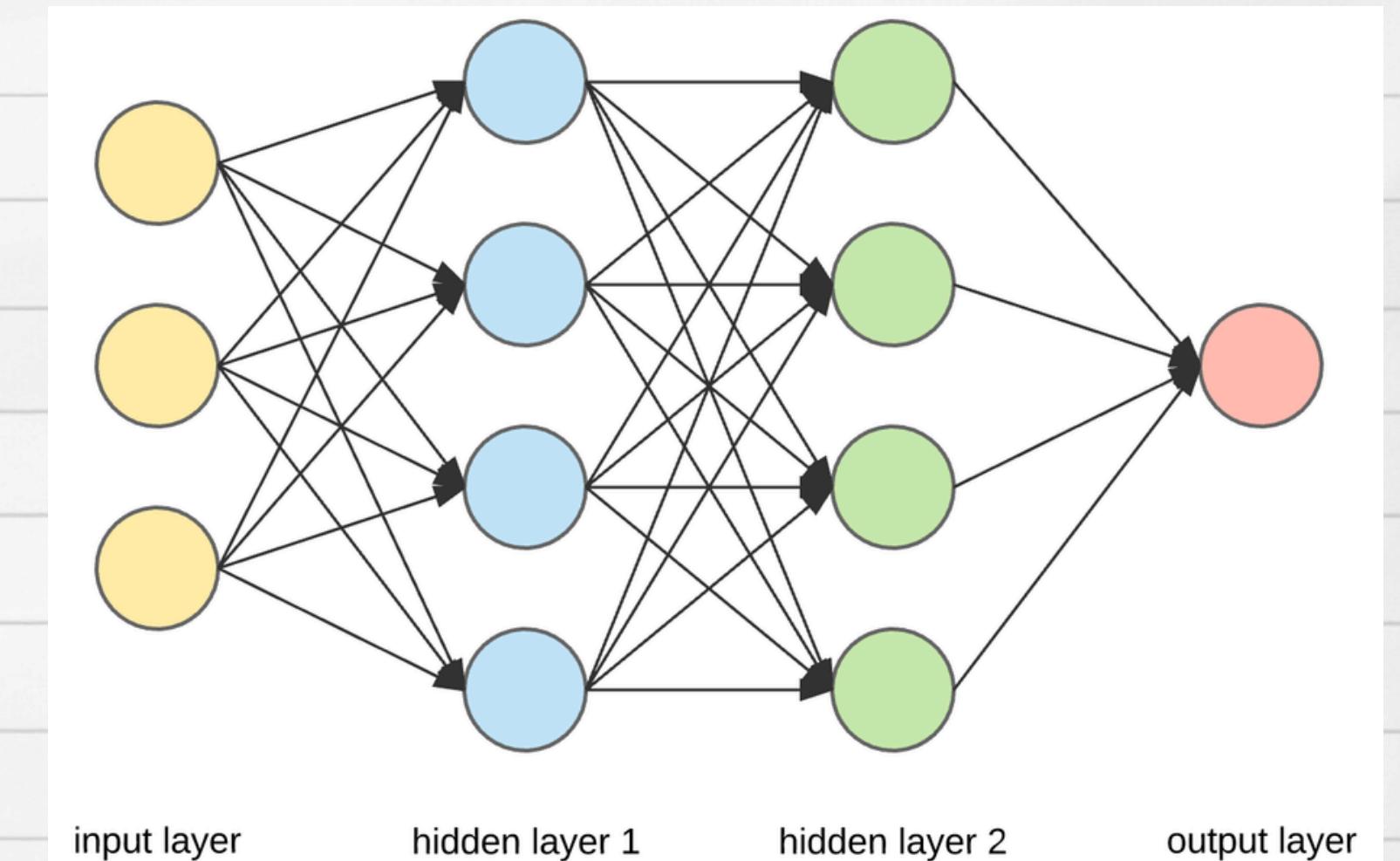
$$\mathcal{L}^{(t)} = -\frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T$$

- G_j là tổng gradient của các mẫu trong nút lá j
- H_j là tổng Hessian (độ cong) của các mẫu trong nút lá j
- λ là hệ số điều chỉnh (regularization) cho trọng số lá
- γ là hệ số phạt cho số lượng lá (control model complexity)
- T là tổng số lá (terminal nodes)

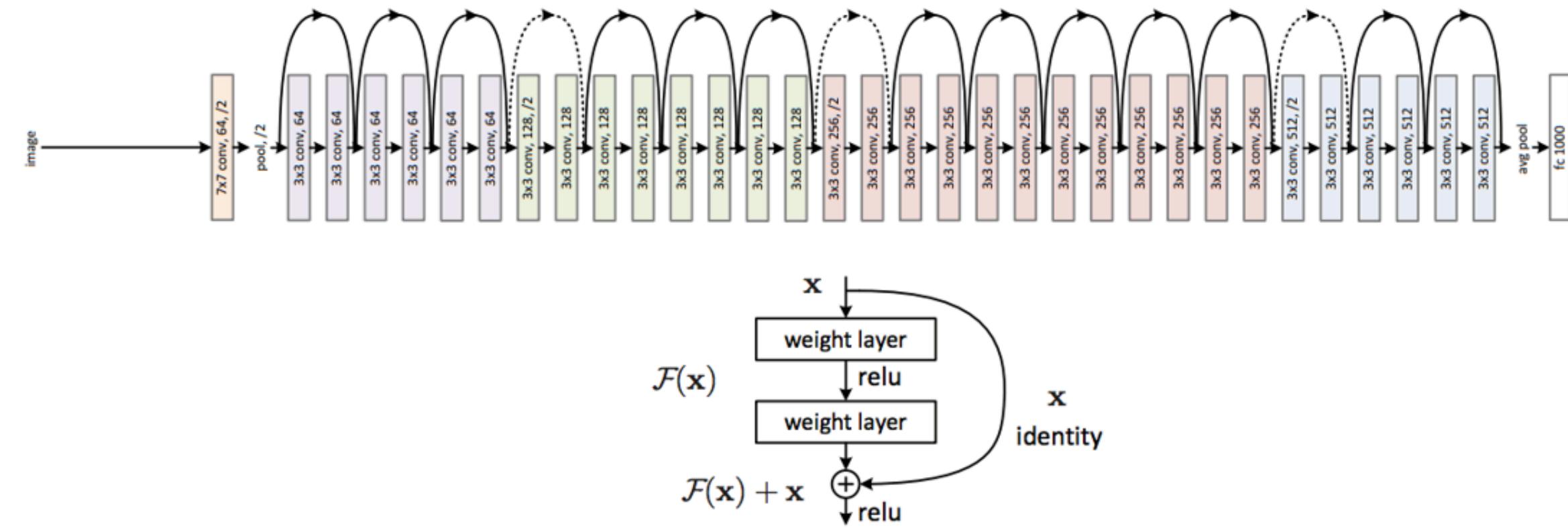


RESNET

- ResNet-50 là một mạng nơ-ron tích chập sâu sử dụng các kết nối residual để giải quyết vấn đề gradient biến mất



Residual Networks (ResNet50)



$$\text{Output} = \mathcal{F}(x) + x$$

- **x là đầu vào của lớp, $F(x)$ là phép toán được thực hiện bởi các lớp mạng**
- **Thay vì học trực tiếp $F(x)$, lớp residual sẽ học sự chênh lệch $F(x)-x$ và sau đó cộng lại với đầu vào ban đầu x**

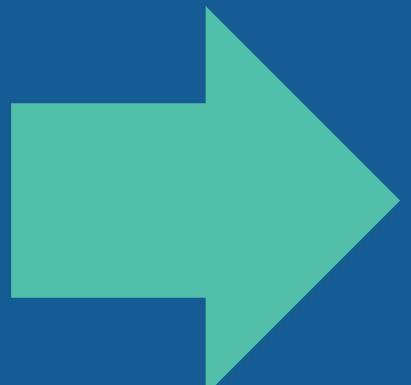
HUẤN LUYỆN MÔ HÌNH

Huân Luyện Mô Hình XGBoost

- Training Set = Dataset * 80% = 63960 tấm ảnh (64x64)
- Nhóm áp dụng một số phương pháp: GridSearchCV, 5-fold cross validation và hàm mất mát log

Các siêu tham số được tìm kiếm:

- n_estimators: [100, 300, 500]
- max_depth: [3, 6, 9]
- learning_rate: [0.01, 0.1, 0.2]
- gamma: [0, 0.1, 0.2]
- subsample: [0.7, 0.8, 1.0]



Bộ siêu tham số tốt nhất:

- n_estimators: 300
- max_depth: 6
- learning_rate: 0.1
- gamma: 0.2
- subsample: 1.0

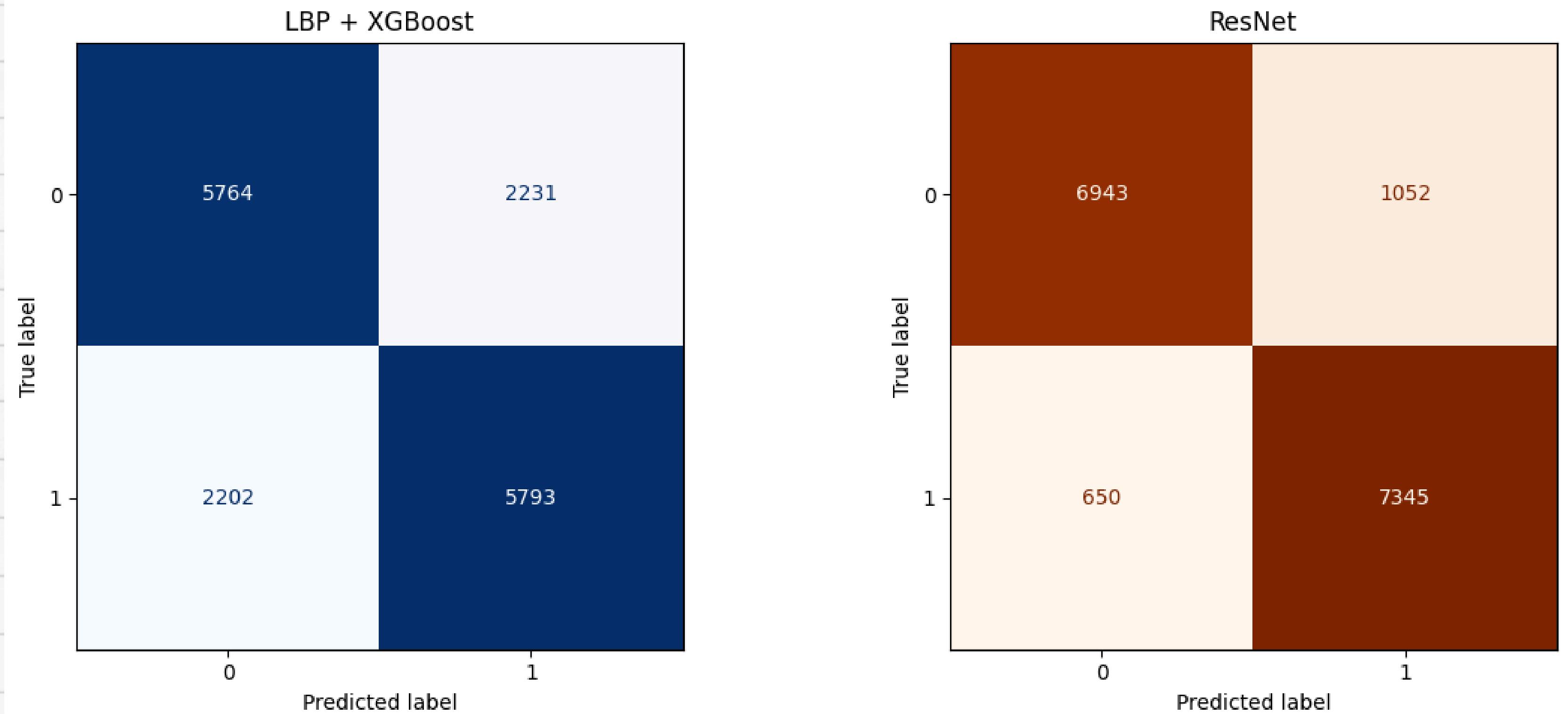
Huân Luyện Mô Hình RESNET

- Training Set = Dataset * 80% = 63960 tấm ảnh (64x64)
- Nhóm áp dụng một số phương pháp: AdamW, hàm mất mát Cross Entropy.
- Tham số huấn luyện: epoch=5, learning rate=0.001

```
Epoch 1/5, Loss: 0.4215, Accuracy: 81.48%, Time: 1310.60s
Epoch 2/5, Loss: 0.3666, Accuracy: 84.10%, Time: 728.54s
Epoch 3/5, Loss: 0.3438, Accuracy: 85.25%, Time: 728.74s
Epoch 4/5, Loss: 0.3279, Accuracy: 85.95%, Time: 727.33s
Epoch 5/5, Loss: 0.3154, Accuracy: 86.58%, Time: 728.66s
```

ĐÁNH GIÁ MÔ HÌNH

Confusion Matrix

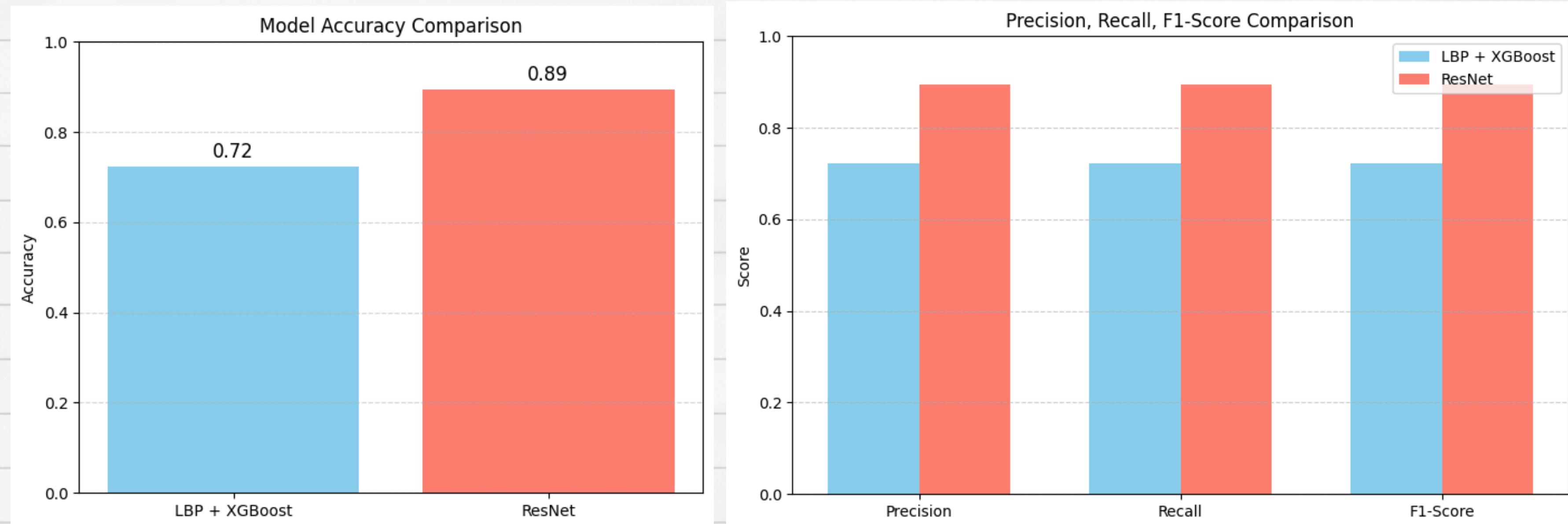


Đánh Giá Mô Hình

==== LBP + XGBOOST ===				
	precision	recall	f1-score	support
0	0.72	0.72	0.72	7995
1	0.72	0.72	0.72	7995
accuracy			0.72	15990
macro avg	0.72	0.72	0.72	15990
weighted avg	0.72	0.72	0.72	15990

==== RESNET ===				
	precision	recall	f1-score	support
0	0.91	0.87	0.89	7995
1	0.87	0.92	0.90	7995
accuracy			0.89	15990
macro avg	0.89	0.89	0.89	15990
weighted avg	0.89	0.89	0.89	15990

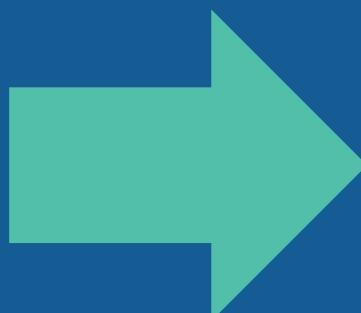
Biểu Diễn Bằng Đồ Thị



Kết Luận

- Nhóm tiến hành sử dụng 2 giải thuật học máy: ResNet và XGBoost.
- XGBoost có độ chính xác ở mức khá, trong khi đó ResNet có độ chính xác khá cao và hiệu quả hơn so với XGBoost.

Bài tập lớn giúp nhóm hiểu rõ hơn về các giải thuật học máy cũng như ứng dụng các giải thuật này vào các bài toán thực tế như Phân biệt ảnh thật và ảnh sinh ra từ AI



**THANK
YOU VERY
MUCH!**