

BÁO CÁO ĐỒ ÁN CUỐI KỲ

Lớp: CS114.(K21+K21.KHTN)

Môn: MÁY HỌC

GV: PGS.TS Lê Đình Duy - THS. Phạm Nguyễn Trường An
Trường ĐH Công Nghệ Thông Tin, ĐHQG-HCM

NHẬN DIỆN CÁC TÒA NHÀ TRONG UIT

TRẦN ĐÌNH KHANG - 18520072 - CS114.K21.KHTN

Link github:

<https://github.com/KhangTran2503/CS114.K21.KHTN>

Tóm tắt

- Ảnh của các thành viên của nhóm



Trần Đình Khang - 18520072

Tóm tắt

Bài toán: Nhận diện các tòa nhà trong UIT

- **Input:** Một bức ảnh các tòa nhà trong trường UIT
- **Output:** Đó là tòa nhà nào trong UIT

Kết quả đạt được:

- Đã xây dựng được nhiều model với các phương pháp lấy đặc trưng khác nhau
- Model tốt nhất đạt được kết quả hơn 50%

INPUT



OUTPUT



Tòa Nhà E

Mô tả bộ dữ liệu

- Dataset thu thập cùng với các bạn Trung, An, Toàn bằng cách đi chụp các tòa nhà trong trường.
- Bộ dữ liệu Train gồm 1096 ảnh file jpg được chụp bằng điện thoại Iphone 7. Các ảnh được chia vào 5 class là : B(Tòa B), C(Tòa C), D(Tòa D), E(Tòa E), CT(Căn Tin).

Trong đó:

- Class B(Tòa B): 277 ảnh
- Class C(Tòa C): 235 ảnh
- Class D(Tòa D): 157 ảnh
- Class E(Tòa E): 257 ảnh
- Class CT(Căn Tin): 170 ảnh

Mô tả bộ dữ liệu

- Bộ dữ liệu Test gồm 461 ảnh file jpg chia thành 5 class giống như tập train mỗi loại gồm:

- Class B(Tòa B): 101 ảnh
- Class C(Tòa C): 85 ảnh
- Class D(Tòa D): 95 ảnh
- Class E(Tòa E): 91 ảnh
- Class CT(Căn Tin): 89 ảnh

- Chia 2 Folder , một Folder lưu tập ảnh để Train và cái còn lại để Test.

Tiền xử lý dữ liệu & rút trích đặc trưng

- Tiền xử lý dữ liệu là resize các bức ảnh về cùng một kích thước 128x128, sau đó chuyển tất cả các ảnh thành màu grayscale.
- Lọc nhiễu ảnh, làm mịn ảnh sử dụng Gaussian Blur
- Cân bằng ánh sáng và độ tương phản (Histogram Equalization) vì có ảnh chụp ngoài trời dưới ánh mặt trời.
- Làm nét ảnh (Sharpening Images) để hiển thị rõ nét các góc, cạnh thuận tiện cho việc lấy đặc trưng ảnh

Tiền xử lý dữ liệu & rút trích đặc trưng

- Thử với các Feature sau:
 - + Xem các pixel sau khi Grayscale như là Feature.
 - + Sử dụng HOG để trích xuất.
 - + Sử dụng Local Binary Patterns (LBP).

Chọn mô hình thuật toán máy học

-Trong Classifications, có nhiều thuật toán phân loại khác nhau như SVM, Decision Tree, ...

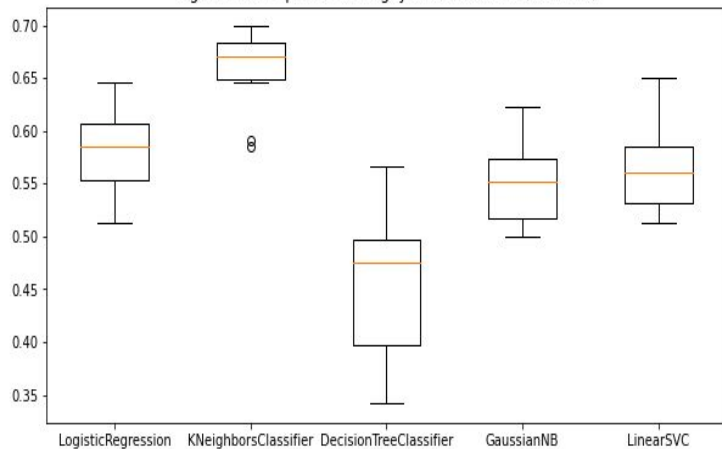
Trong đồ án, em chọn thuật toán:

- Linear SVM
- Logistic Regression
- Decision Tree
- Naive Bayes
- KNN

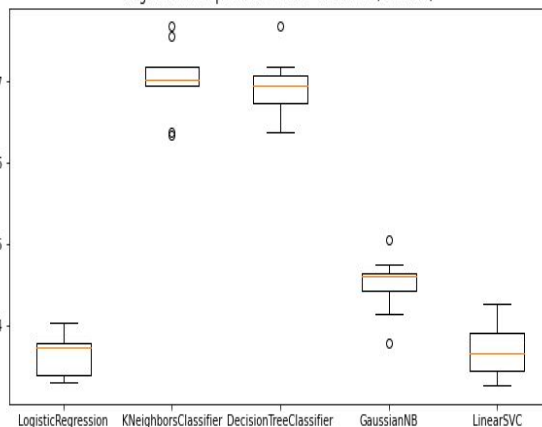
Chọn mô hình thuật toán máy học

1. Đối với mỗi cách rút trích, em thử với 5 model với StratifiedKFold để tìm model phù hợp nhất cho mỗi cách rút trích:

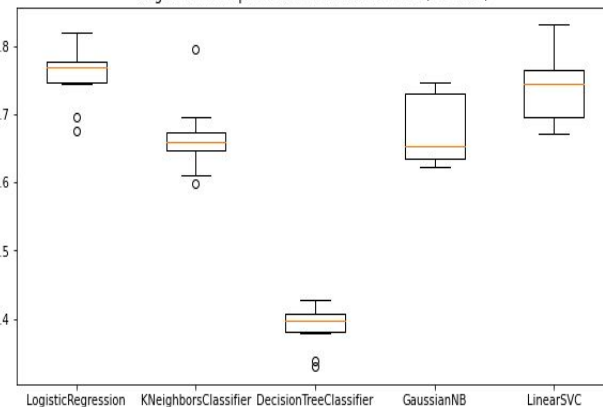
Algorithm Comparison with grayscale extraction(F1 score)



Algorithm Comparison with LBP extraction(F1 score)



Algorithm Comparison with HOG extraction (F1 score)



Rút trích: Grayscale ➡ KNN

Rút trích: LBP ➡ KNN

Rút trích: HOG ➡ Logistic Regression

Tuning mô hình thuật toán máy học

- Sử dụng RandomizedSearchCV cùng với StratifiedKFold(n_split=5) để tuning model.
- Đối với K-NN tinh chỉnh bộ tham số:

```
leaf_size = list(range(1,50))  
n_neighbors = list(range(1,30))  
p=[1,2]  
weights = ['uniform','distance']
```

- Đối với Logistic Regression tinh chỉnh bộ tham số:

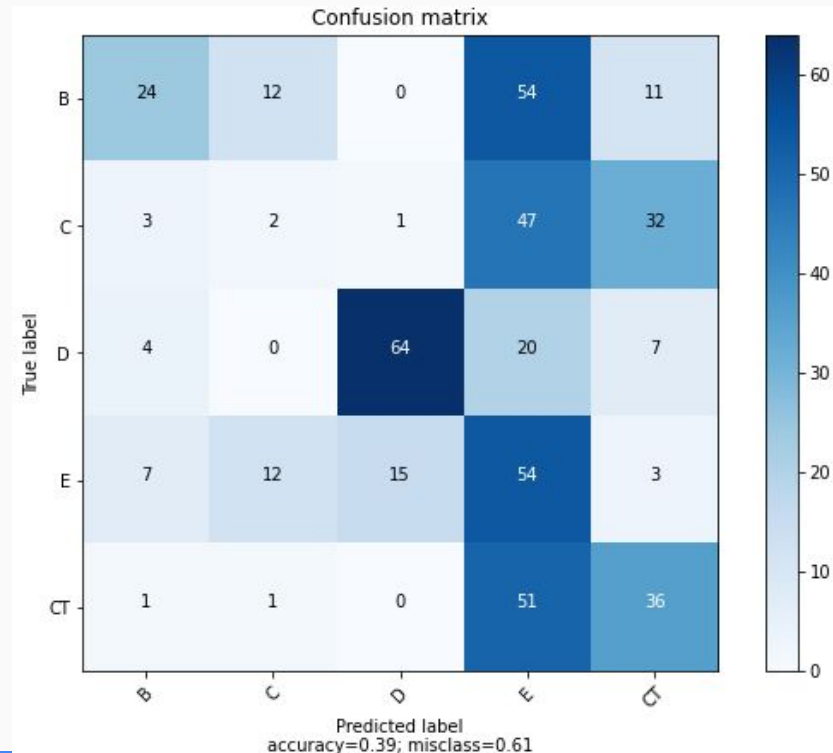
```
solver = ['newton-cg', 'lbfgs', 'liblinear']  
penalty = ['l1', 'l2', 'elasticnet']  
C = [10, 4.5, 3, 1.5, 1.0, 0.3, 0.1, 0.03, 0.01]
```

Sử dụng model cho Unseen data

1. Kết quả của phương pháp sử dụng ảnh grayscale chuyển sang vector đặc trưng.

	precision	recall	f1-score	support
B	0.62	0.24	0.34	101
C	0.07	0.02	0.04	85
D	0.80	0.67	0.73	95
E	0.24	0.59	0.34	91
CT	0.40	0.40	0.40	89
accuracy			0.39	461
macro avg	0.43	0.39	0.37	461
weighted avg	0.44	0.39	0.38	461

- Đối với Model(K-NN) thì kết quả dự đoán là 0.39, trong đó label E được dự đoán khá nhiều nhưng tỉ lệ dự đoán trúng label E khá thấp (0.26), nhầm lẫn tòa E và tòa B (54 bức ảnh).

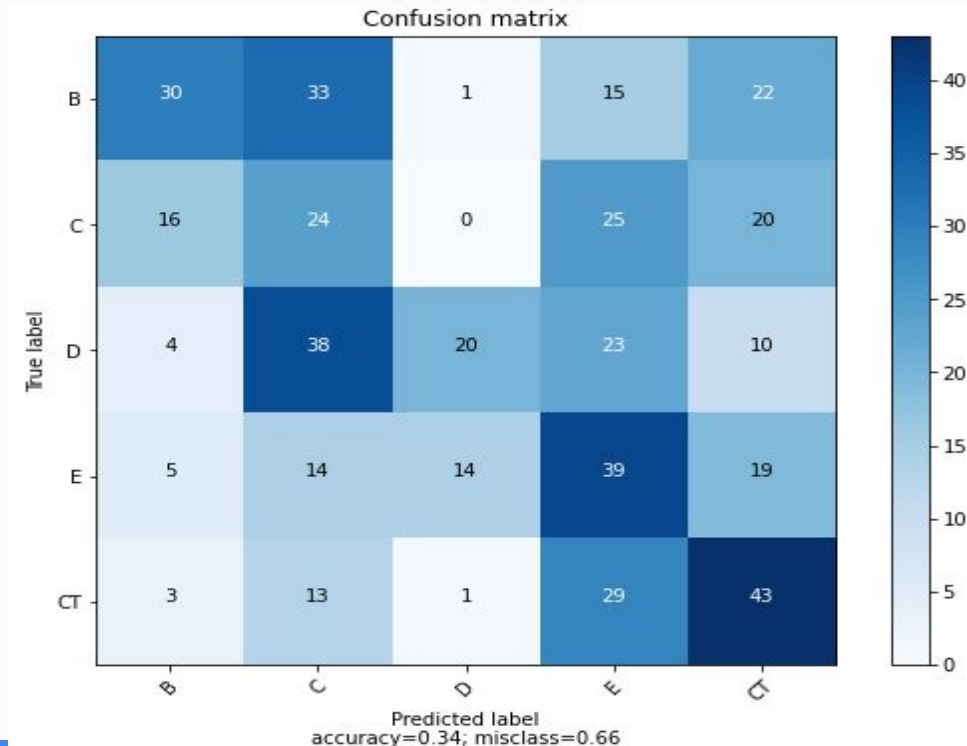


Sử dụng model cho Unseen data

2. Kết quả của phương pháp rút trích đặc bằng LBP

	precision	recall	f1-score	support
B	0.52	0.30	0.38	101
C	0.20	0.28	0.23	85
D	0.56	0.21	0.31	95
E	0.30	0.43	0.35	91
CT	0.38	0.48	0.42	89
accuracy			0.34	461
macro avg	0.39	0.34	0.34	461
weighted avg	0.40	0.34	0.34	461

- Đối với Model(K-NN) thì kết quả dự đoán là 0.34, model dự đoán nhiều vào tòa B, nhưng có 38 ảnh bị dự đoán C trong khi các bức ảnh này là tòa D.

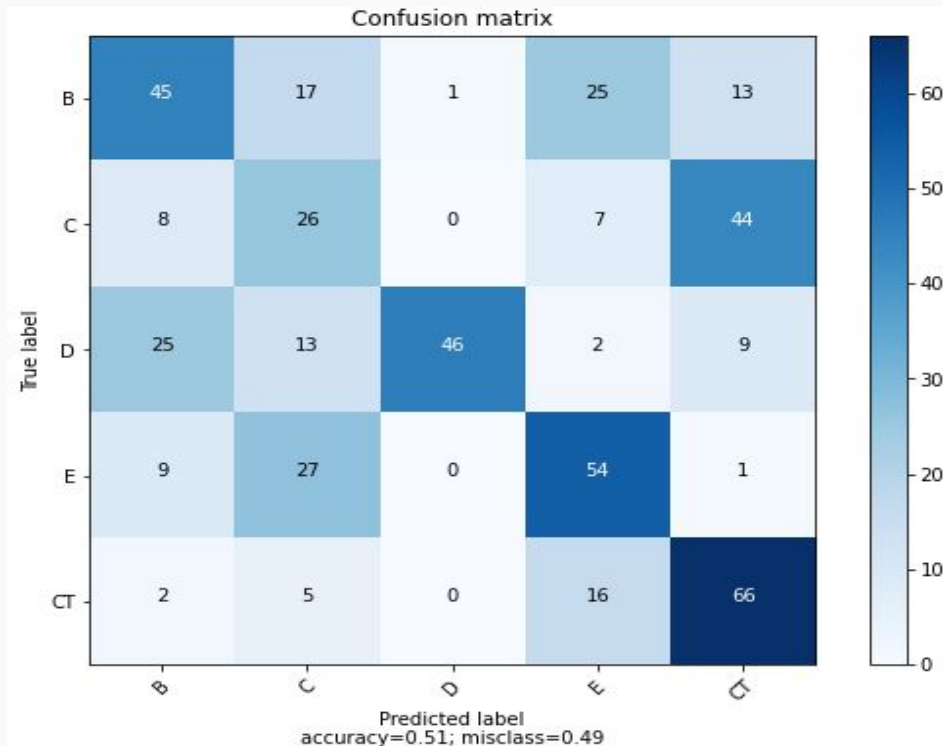


Sử dụng model cho Unseen data

3. Kết quả của phương pháp sử dụng ảnh HOG

	precision	recall	f1-score	support
B	0.51	0.45	0.47	101
C	0.30	0.31	0.30	85
D	0.98	0.48	0.65	95
E	0.52	0.59	0.55	91
CT	0.50	0.74	0.59	89
accuracy			0.51	461
macro avg	0.56	0.51	0.51	461
weighted avg	0.57	0.51	0.52	461

- Model logistic regression có acc (0.51) cao hơn các model khác nhưng vẫn còn thấp. Dự đoán sai nhiều (44) giữa tòa CT và tòa C.



Kết quả và đánh giá

- Các model sử dụng đã không có độ chính xác cao $< 52\%$. Nguyên nhân dẫn đến bởi các khó khăn sau:
 - Bộ dữ liệu còn quá ít (khoảng 1400 ảnh cho 5 class) chưa có được sự tổng quát cho mô hình. Có nhiều bức ảnh bị nhiễu bởi lá cây.
 - Phương pháp trích xuất đặc trưng còn hạn chế, chưa loại bỏ được những đối tượng không mong muốn trong hình. Chưa thử nghiệm với ảnh màu.
 - Hạn chế về kỹ năng thu thập dữ liệu. Các tòa nhà có kích thước lớn, không thể chụp toàn bộ một tòa nhà trong một khung hình nên phải chụp từng góc. Khi nhìn vào một vài góc thì sự có sự tương đồng giữa các tòa nhà nên đã gây nên sự nhầm lẫn cho mô hình dự đoán.
- **Hướng phát triển:**
 - Sẽ sử dụng CNN để rút trích đặc trưng, dùng các model deep learning.
 - Chụp thêm các ảnh sao cho ít bị nhiễu, lấy được toàn cảnh tòa nhà.