

# DỰ ĐOÁN CHỮ CÁI KÝ HIỆU THỦ NGỮ BẰNG THUẬT TOÁN MÁY HỌC

Đồng Quốc Tuấn -

18520185 -

CS114.K21.KHTN



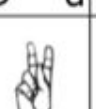
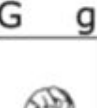
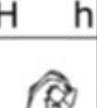
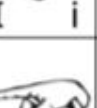
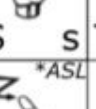


Link Github:

<https://github.com/furu-tuan-dong/CS114.K21.KHTN>



# I. Giới thiệu bài toán

- Bài toán dự đoán hình ảnh của bàn tay để đưa ra kết quả thuộc chữ cái nào trong bảng chữ cái thủ ngữ Việt Nam bằng thuật toán máy học.
- Dữ liệu dùng để dự đoán: hình ảnh của bàn tay.
- Bài toán phân loại 22 nhãn tương ứng với 22 ký hiệu trong bảng chữ cái thủ ngữ Việt Nam không bao gồm: “Đ”.

 A a	 B b	 C c	 D d	 Đ đ	 E e	 F f
 G g	 H h	 I i	 J j	 K k	 L l	 M m
 N n	 O o	 P p	 Q q	 R r	 S s	 T t
 U u	 V v	 W w	 X x	 Y y	 Z z	
 Horn ♂	 Breve ˘	 Circumflex ˆ	<b>Vietnamese Sign Language</b> Combining Diacritic Marks			
 Hook ˆ	 Acute ˙	 Grave ˘	 Tilde ˜	 Underdot ˙		

## II. Xây dựng ứng dụng máy học

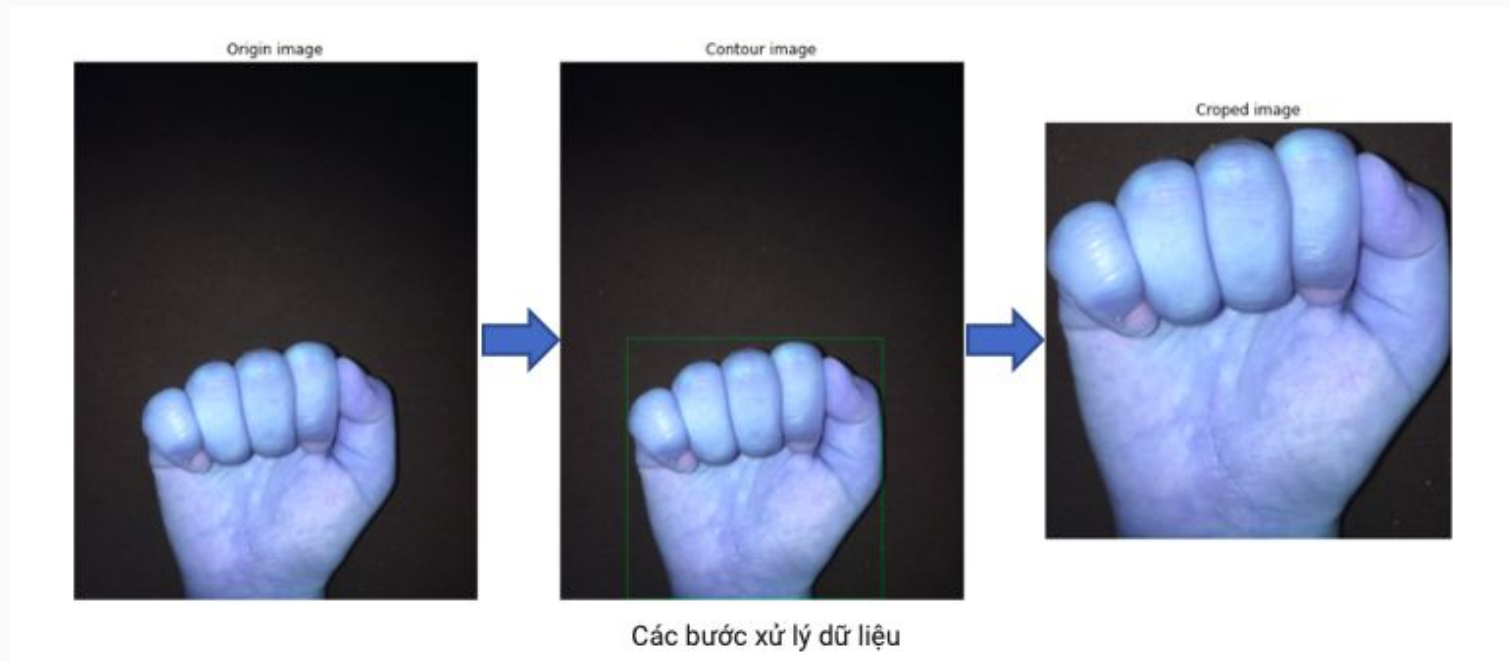
1. Mô tả dữ liệu
2. Tiền xử lý dữ liệu
3. Các thuật toán máy học
4. Training
5. Đánh giá
6. Tinh chỉnh (Tuning) các tham số
7. Dự đoán trên dữ liệu mới

# 1. Mô tả dữ liệu

- Ngôn ngữ kí hiệu tĩnh Việt Nam bao gồm 25 kí tự bao gồm 23 kí tự Việt Nam không bao gồm F, J, W, Z và 2 kí hiệu dấu 'râu' và 'mũ'.
- Bộ dữ liệu đồ án bao gồm 22 kí tự không bao gồm kí tự 'Đ'
- Mỗi kí tự bao gồm 50 hình ảnh đều được chụp bằng điện thoại, dưới ánh đèn Flash và nền đen.
- Kích thước bộ dữ liệu bao gồm  $22 * 50 = 1100$  hình ảnh

## 2. Tiền xử lý dữ liệu

Loại bỏ những thông tin không cần thiết như phần rìa đen của background



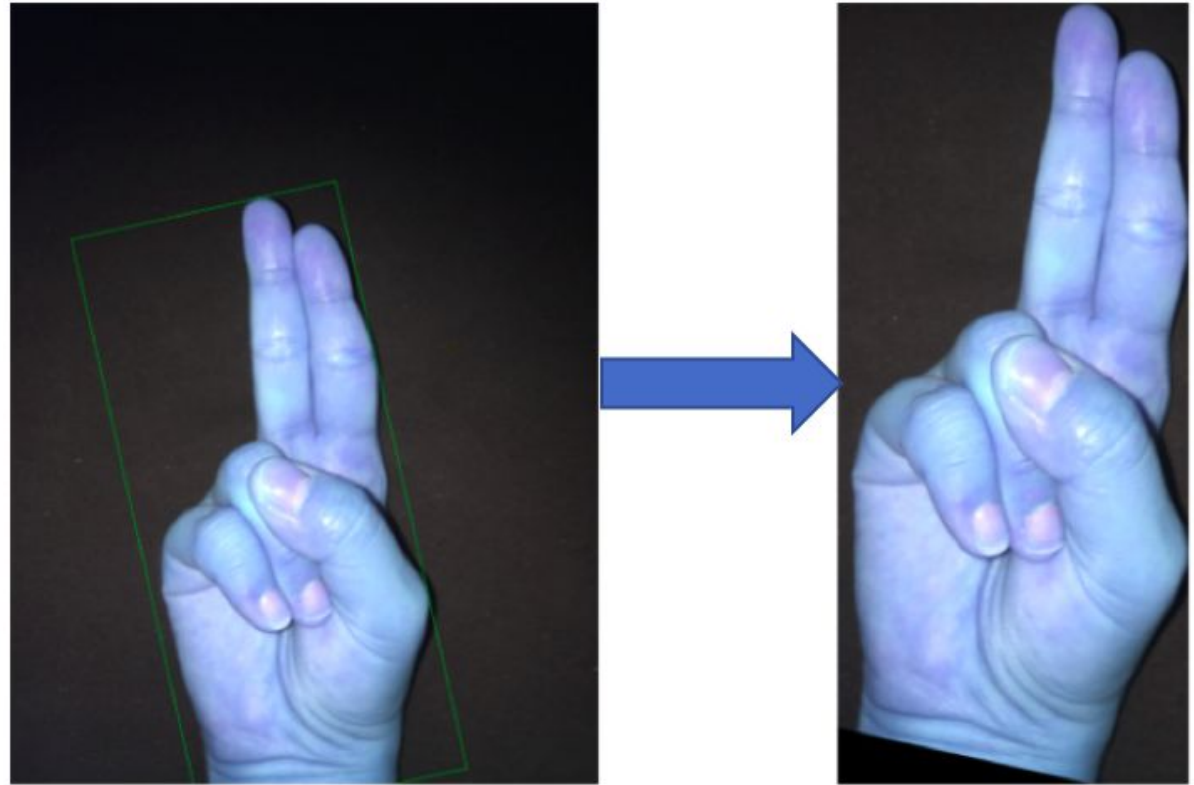
## 2. Tiền xử lý dữ liệu

Quá trình xử lý bao gồm 2 bước:

- Bước 1: Ảnh bằng cách đưa về ảnh gray -> lọc nhiễu bằng bộ lọc Gaussian -> áp dụng thresh hold để chuyển về ảnh binary (vì màu da và màu nền đối nhau). Tìm contour lớn nhất của ảnh binary và bounding box tương ứng (khung màu xanh của hình Countour image ở trên).
- Bước 2: Vì sẽ có một số bounding box có góc xoay, nên cần thực hiện một số phép biến đổi hình học để đưa về góc thẳng.

## 2. Tiền xử lý dữ liệu

Hình ảnh sau khi loại bỏ những thông tin không cần thiết được sử dụng để rút trích đặc trưng.



Áp dụng phép biến đổi hình học

## 2. Tiền xử lý dữ liệu

- Vì các hình sau khi xử lý có size khác nhau nên cần đưa về 90x90.
- Rút trích đặc trưng: Xem xét sự khác nhau giữa các ký hiệu chính là hình dáng và góc cạnh của ngón tay.
- Chuyển đổi hình ảnh thành vector (8100x1). Cũng chính là input của các model.





## 2. Tiền xử lý dữ liệu

- Phân chia dữ liệu: Dùng dữ liệu vừa được rút trích đặc trưng ở trên, để xây dựng model. Trong đó 80% dùng để train các thuật toán và 20% để thẩm định kết quả.
- Dữ liệu ở folder [test](#), bao gồm 5 ảnh cho từng 22 folder đại diện các ký tự, là bộ dữ liệu thô cần phải được tiền xử lý và rút trích đặc trưng, để kiểm tra hiệu quả của thuật toán tốt nhất chọn được.

### 3. Các thuật toán máy học

- K-Nearest-Neighbors (k-NN)
- Logistic Regression sử dụng kĩ thuật one-vs-rest (OvR)
- Softmax Regression

### 4. Training

Sử dụng các model mặc định của thư viện Scikit-learn

## 5. Đánh giá

- k-NN có kết quả thấp nhất với f1-score = 0.4
- Softmax Regression và OvR có kết quả gần giống nhau  
f1-score = 0.82
- Ký tự G có hiệu quả dự đoán thấp nhất ở cả Softmax Regression và OvR.

# 5. Đánh giá

Accuracy on training set: 100.0%				
Accuracy on test set: 40.45454545454545%				
	precision	recall	f1-score	support
A	0.08	0.75	0.14	12
B	1.00	0.78	0.88	9
C	1.00	0.07	0.12	15
D	0.58	0.78	0.67	9
E	0.71	0.56	0.63	9
G	1.00	0.12	0.22	8
H	1.00	0.08	0.14	13
I	1.00	0.33	0.50	9
K	1.00	0.10	0.18	10
L	0.00	0.00	0.00	9
M	0.42	0.73	0.53	11
N	0.78	0.70	0.74	10
O	1.00	0.17	0.29	12
P	1.00	0.07	0.12	15
Q	1.00	0.14	0.25	7
R	0.86	0.86	0.86	7
S	0.75	0.90	0.82	10
T	0.86	0.55	0.67	11
U	0.88	0.88	0.88	8
V	1.00	0.10	0.18	10
X	1.00	0.33	0.50	9
Y	1.00	0.43	0.60	7
accuracy			0.40	220
macro avg			0.81	220
weighted avg			0.81	220

k-NN

Accuracy on training set: 100.0%				
Accuracy on test set: 81.818181818183%				
	precision	recall	f1-score	support
A	0.57	0.67	0.62	12
B	0.82	1.00	0.90	9
C	1.00	0.87	0.93	15
D	0.90	1.00	0.95	9
E	1.00	0.89	0.94	9
G	1.00	0.25	0.40	8
H	0.64	0.54	0.58	13
I	0.71	0.56	0.63	9
K	1.00	0.90	0.95	10
L	1.00	1.00	1.00	9
M	0.53	0.82	0.64	11
N	0.67	1.00	0.80	10
O	0.92	1.00	0.96	12
P	0.85	0.73	0.79	15
Q	0.67	0.57	0.62	7
R	1.00	1.00	1.00	7
S	0.82	0.90	0.86	10
T	0.89	0.73	0.80	11
U	0.80	1.00	0.89	8
V	1.00	0.80	0.89	10
X	0.80	0.89	0.84	9
Y	1.00	1.00	1.00	7
accuracy			0.82	220
macro avg			0.84	220
weighted avg			0.84	220

one-vs-rest

Accuracy on training set: 100.0%				
Accuracy on test set: 82.272727272728%				
	precision	recall	f1-score	support
A	0.53	0.67	0.59	12
B	0.90	1.00	0.95	9
C	1.00	0.87	0.93	15
D	0.90	1.00	0.95	9
E	1.00	0.89	0.94	9
G	0.67	0.25	0.36	8
H	0.73	0.62	0.67	13
I	0.71	0.56	0.63	9
K	0.90	0.90	0.90	10
L	1.00	1.00	1.00	9
M	0.60	0.82	0.69	11
N	0.67	1.00	0.80	10
O	0.92	1.00	0.96	12
P	0.79	0.73	0.76	15
Q	0.80	0.57	0.67	7
R	0.88	1.00	0.93	7
S	0.90	0.90	0.90	10
T	0.90	0.82	0.86	11
U	0.80	1.00	0.89	8
V	1.00	0.70	0.82	10
X	0.80	0.89	0.84	9
Y	1.00	1.00	1.00	7
accuracy			0.82	220
macro avg			0.84	220
weighted avg			0.83	220

Softmax Regression

## 6. Tinh chỉnh (Tuning) các tham số

Sau khi tuning, Logistic Regression kỹ thuật one-vs-rest cho kết quả tốt nhất với f1-score = 0.83

Accuracy on training set: 100.0%				
Accuracy on test set: 82.72727272727273%				
	precision	recall	f1-score	support
A	0.53	0.67	0.59	12
B	0.90	1.00	0.95	9
C	1.00	0.87	0.93	15
D	0.90	1.00	0.95	9
E	1.00	0.89	0.94	9
G	0.67	0.25	0.36	8
H	0.73	0.62	0.67	13
I	0.71	0.56	0.63	9
K	1.00	0.90	0.95	10
L	1.00	1.00	1.00	9
M	0.60	0.82	0.69	11
N	0.67	1.00	0.80	10
O	0.92	1.00	0.96	12
P	0.79	0.73	0.76	15
Q	0.80	0.57	0.67	7
R	0.88	1.00	0.93	7
S	0.90	0.90	0.90	10
T	0.90	0.82	0.86	11
U	0.80	1.00	0.89	8
V	1.00	0.80	0.89	10
X	0.80	0.89	0.84	9
Y	1.00	1.00	1.00	7
accuracy			0.83	220
macro avg	0.84	0.83	0.82	220
weighted avg	0.84	0.83	0.82	220

## 7. Dự đoán trên dữ liệu mới

Sử dụng one-vs-rest với các tham số tốt nhất trên bộ dữ liệu test (khác nhau về cường độ sáng, góc chụp, ...), kết quả thu được: f1-score = 0.2

Accuracy: 20.9090909090907%				
	precision	recall	f1-score	support
A	0.00	0.00	0.00	5
B	0.00	0.00	0.00	5
C	0.33	0.60	0.43	5
D	0.22	0.40	0.29	5
E	0.20	0.40	0.27	5
G	0.00	0.00	0.00	5
H	1.00	0.20	0.33	5
I	0.25	0.40	0.31	5
K	0.00	0.00	0.00	5
L	1.00	0.20	0.33	5
M	0.07	0.20	0.11	5
N	0.20	0.40	0.27	5
O	0.00	0.00	0.00	5
P	0.00	0.00	0.00	5
Q	0.00	0.00	0.00	5
R	0.00	0.00	0.00	5
S	0.50	0.20	0.29	5
T	0.20	0.20	0.20	5
U	0.40	0.40	0.40	5
V	0.75	0.60	0.67	5
X	0.20	0.20	0.20	5
Y	0.50	0.20	0.29	5
accuracy			0.21	110
macro avg	0.26	0.21	0.20	110
weighted avg	0.26	0.21	0.20	110

# III. Kết luận

- Bài toán bị overfit có thể do bộ dữ liệu train chưa đa dạng về môi trường, model chưa hợp lý hoặc đặc trưng không chuẩn xác.
- Bài toán chưa xử lý được với ảnh có background không phải là màu tối. Chưa áp dụng vào được realtime.

**\*\* Hướng giải quyết \*\***

- Cần áp dụng các phương pháp loại bỏ background chỉ lấy mỗi bàn tay. Xem xét các đặc trưng quan trọng khác và model hợp lý cho đặc trưng đó.