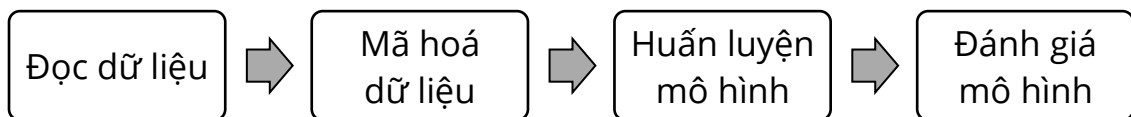


Bài thực hành 7. MÔ HÌNH PHÂN LỚP (Phần 3)

1. PHÂN LỚP ẢNH

Bài toán: Nhận diện bệnh nhiễm khuẩn phổi từ ảnh X-Quang

- Phát biểu bài toán:
 - Input: Một tấm ảnh X-quang.
 - Output: Nhãn dự đoán của tấm ảnh đó, gồm 1 trong 2 nhãn: NORMAL - bình thường và PNEUMONIA - bệnh.
- Bộ dữ liệu: **Chest X-ray dataset**.
 - Công bố khoa học: Kermay, Daniel S., et al. "**Identifying medical diagnoses and treatable diseases by image-based deep learning**." Cell 172.5 (2018): 1122-1131.
 - Link tải: <https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>.
- Các bước thực hiện tổng quát:



- Đọc dữ liệu:
 - Các file được tổ chức thành từng bức ảnh có định dạng là .jpeg.
 - Các bức ảnh thuộc nhãn NORMAL hoặc PNEUMONIA sẽ được gom chung thành 1 thư mục.
 - Code tham khảo đọc dữ liệu:

```
import glob
import numpy as np
import cv2
IMG_SIZE = 227
def load_dataset(path):
    X = np.array([])
    y = np.array([])
    classes = ['NORMAL', 'PNEUMONIA']
    for c in classes:
        files = glob.glob(path + c + "/*.jpeg")
        for f in files:
            print(f)
```

```
img = cv2.imread(f)
img = cv2.resize(img, (IMG_SIZE, IMG_SIZE))
if X.size == 0:
    X = np.array([img])
else:
    X = np.vstack([X,[img]])
y = np.append(y, c)
assert(X.size > 0), 'Cannot read file'
return (X,y)
```

– Để load dữ liệu từ tập train, ta sử dụng lệnh:

```
X_train, y_train = load_dataset("<đường dẫn tới dataset>")
```

– Thực hiện tương tự để đọc dữ liệu từ tập test.

2. BÀI TẬP

- **Bài tập 1.** Đọc dữ liệu và cho biết các thông tin sau:

a) Mục tiêu / tác vụ mà bộ dữ liệu hướng tới là gì?

b) Kích thước của dữ liệu là bao nhiêu trên tập train và test?

c) Phân bố nhãn theo từng tác vụ trên tập train và test. Vẽ biểu đồ barplot thể hiện phân bố nhãn trên từng tập dữ liệu.

d) Hiển thị 2 bức hình về ảnh X-Quang phổi lên màn hình (mỗi tập 1 ảnh).

- **Bài tập 2**

Xây dựng, huấn luyện và đánh giá mô hình phân lớp Logistic Regression theo hướng dẫn trong phần 1.

- **Bài tập 3**

a) Rút trích đặc trưng ảnh trong các tập dữ liệu bằng cách sử dụng các loại đặc trưng: HOG, CANNY và FAST.

b) Xây dựng các mô hình phân lớp SVM và huấn luyện các mô hình bằng các vector đặc trưng HOG, CANNY và FAST.

c) So sánh độ chính xác dự đoán của mô hình trong ba trường hợp:

➤ Sử dụng đặc trưng HOG.

➤ Sử dụng đặc trưng CANNY.

➤ Sử dụng đặc trưng FAST.

Lưu ý: Mỗi loại đặc trưng là một mô hình thử nghiệm.