

## ▼ Lab 7. Mô hình phân lớp (Phần 3)

```
1 from google.colab import drive
2 drive.mount('/content/drive')
```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force\_remount=True).

### ▼ 1. Đọc dữ liệu

```

1 import glob
2 import numpy as np
3 import cv2
4
5 IMG_SIZE = 227
6
7 def load_dataset(path):
8     X = np.array([])
9     y = np.array([])
10    classes = ['NORMAL', 'PNEUMONIA']
11    for c in classes:
12        files = glob.glob(path + c + '/*.jpeg')
13        for f in files:
14            print(f)
15            img = cv2.imread(f)
16            img = cv2.resize(img, (IMG_SIZE, IMG_SIZE))
17            if X.size == 0:
18                X = np.array([img])
19            else:
20                X = np.vstack([X, [img]])
21            y = np.append(y, c)
22
23    assert(X.size > 0), 'Cannot read file'
24    return (X,y)

```

```
1 X_train, y_train = load_dataset('/content/drive/MyDrive/Colab Notebooks/DS102 - Học máy thống kê/HK1 - 23-24/Lab 7. Mô hình phân lớp (
```

**Kết quả truyền trực tuyến bị cắt bớt đến 5000 dòng cuối.**

[illegible]



<matplotlib.image.AxesImage at 0x7be7a4feb2e0>



## ▼ 2. Tiền xử lý và mã hóa dữ liệu

```
1 from sklearn.preprocessing import LabelEncoder
2
3 le = LabelEncoder()
4 le.fit(y_train)
5
6 IMG_SIZE = 227
7
8 X1 = X_train.reshape(X_train.shape[0], IMG_SIZE*IMG_SIZE*3)
9 X3 = X_test.reshape(X_test.shape[0], IMG_SIZE*IMG_SIZE*3)
10
11 y1 = le.transform(y_train)
12 y3 = le.transform(y_test)
```

## ▼ 3. Huấn luyện mô hình

```
1 from sklearn.linear_model import LogisticRegression
2
3 lrc = LogisticRegression()
```

```
1 lrc.fit(X1, y1)
```

/usr/local/lib/python3.10/dist-packages/sklearn/linear\_model/\_logistic.py:458: ConvergenceWarning: lbfgs failed to converge (status=STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max\_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

[https://scikit-learn.org/stable/modules/linear\\_model.html#logistic-regression](https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression)

```
n_iter_i = _check_optimize_result(
```

```
    LogisticRegression
```

```
    LogisticRegression())
```

## ▼ 4. Đánh giá mô hình

```
1 y_pred_lrc = lrc.predict(X3)
```

```
1 from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
```

```
1 accuracy_score(y3, y_pred_lrc)*100
```

73.71794871794873

```
1 precision_score(y3, y_pred_lrc)*100
```

70.84870848708486

```
1 recall_score(y3, y_pred_lrc)*100
```

98.46153846153847

```
1 f1_score(y3, y_pred_lrc, average='macro')*100
```

65.25234964958983

## ▼ 5. Bài tập

### ▼ Bài tập 1

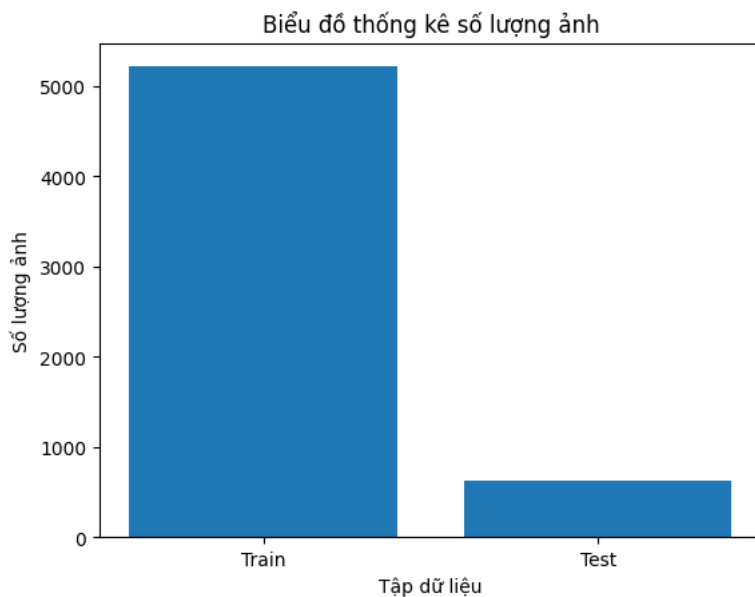
```
1 X_train[0].shape
```

(227, 227, 3)

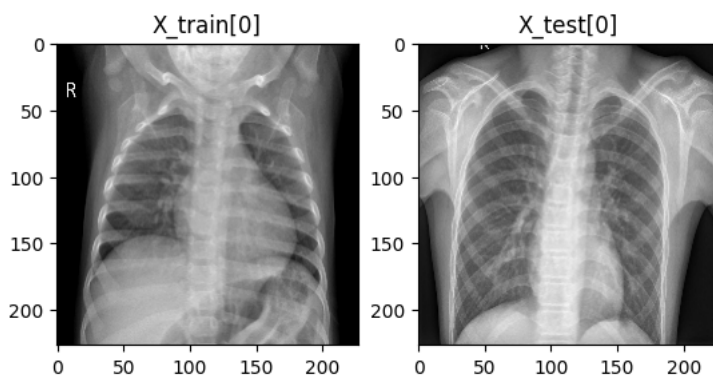
```
1 X_test[0].shape
```

(227, 227, 3)

```
1 import seaborn as sns
2 import matplotlib.pyplot as plt
3 plt.bar(['Train', 'Test'], [len(X_train), len(X_test)])
4 plt.xlabel('Tập dữ liệu')
5 plt.ylabel('Số lượng ảnh')
6 plt.title('Biểu đồ thống kê số lượng ảnh')
7 plt.show()
```



```
1 plt.subplot(1, 2, 1)
2 plt.imshow(X_train[0])
3 plt.title('X_train[0]')
4
5 plt.subplot(1, 2, 2)
6 plt.imshow(X_test[0])
7 plt.title('X_test[0]')
8
9 plt.show()
```



### ▼ Bài tập 3

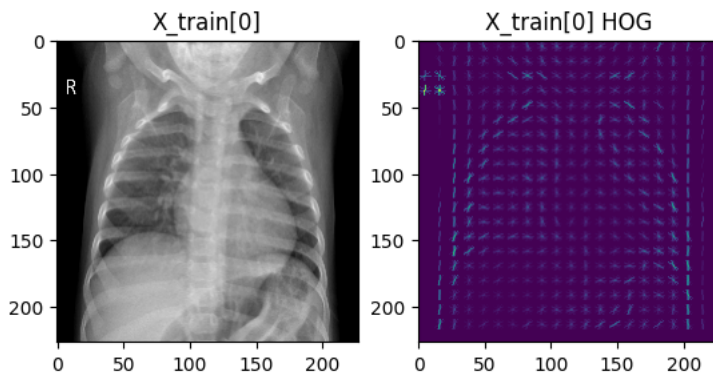
#### ▼ Đặc trưng HOG

```
1 from skimage.feature import hog
2
3 def feature_hog(img):
4     fd, hog_image = hog(img, orientations=9, pixels_per_cell=(11, 11),\
5                          cells_per_block=(2, 2), visualize=True, channel_axis=-1)
6
7     return hog_image
```

```
1 X1_hog = np.array([feature_hog(k) for k in X_train]).reshape(X_train.shape[0], IMG_SIZE
```

```
1 X3_hog = np.array([feature_hog(k) for k in X_test]).reshape(X_test.shape[0], IMG_SIZE*IMG_SIZE)
```

```
1 plt.subplot(1, 2, 1)
2 plt.imshow(X_train[0])
3 plt.title('X_train[0]')
4
5 plt.subplot(1, 2, 2)
6 plt.imshow(feature_hog(X_train[0]))
7 plt.title('X_train[0] HOG')
8
9 plt.show()
```



```
1 from sklearn import svm
2
3 svc_hog = svm.SVC(kernel='linear')
```

```
1 svc_hog.fit(X1_hog, y1)
```

```
▼ SVC
SVC(kernel='linear')
```

```
1 y_pred_svc_hog = svc_hog.predict(X3_hog)
```

```
1 accuracy_score(y3, y_pred_svc_hog)*100
```

74.51923076923077

```
1 precision_score(y3, y_pred_svc_hog)*100
```

71.11517367458866

```
1 recall_score(y3, y_pred_svc_hog)*100
```

99.74358974358975

```
1 f1_score(y3, y_pred_svc_hog, average='macro')*100
```

65.95277395532707

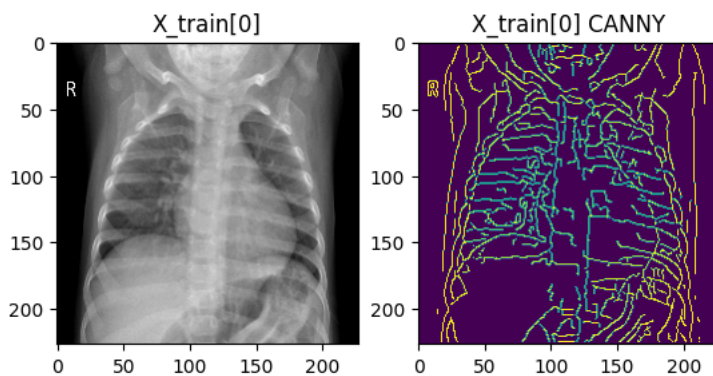
## ▼ Đặc trưng CANNY

```
1 from skimage.feature import canny
```

```
1 X1_canny = np.array([canny(k[:, :, 0]) for k in X_train]).reshape(X_train.shape[0], IMG_SIZE*IMG_SIZE)
```

```
1 X3_canny = np.array([canny(k[:, :, 0]) for k in X_test]).reshape(X_test.shape[0], IMG_SIZE*IMG_SIZE)
```

```
1 plt.subplot(1, 2, 1)
2 plt.imshow(X_train[0])
3 plt.title('X_train[0]')
4
5 plt.subplot(1, 2, 2)
6 plt.imshow(canny(X_train[0][:, :, 0]))
7 plt.title('X_train[0] CANNY')
8
9 plt.show()
```



```
1 from sklearn import svm
2
3 svc_canny = svm.SVC(kernel='linear')
```

```
1 svc_canny.fit(X1_canny, y1)
```

```
▼ SVC
SVC(kernel='linear')
```

```
1 y_pred_svc_canny = svc_canny.predict(X3_canny)
```

```
1 accuracy_score(y3, y_pred_svc_canny)*100
```

74.03846153846155

```
1 precision_score(y3, y_pred_svc_canny)*100
```

72.265625

```
1 recall_score(y3, y_pred_svc_canny)*100
```

94.87179487179486

```
1 f1_score(y3, y_pred_svc_canny, average='macro')*100
```

67.60955102982453

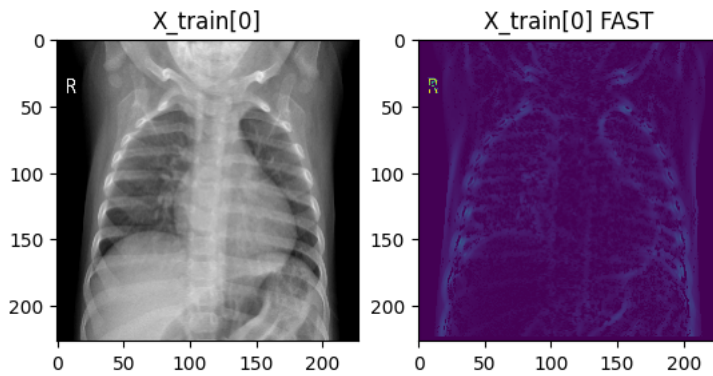
## ▼ Đặc trưng FAST

```
1 from skimage.feature import corner_fast
2
3 def feature_fast(img):
4     fast_image = corner_fast(img, n=8, threshold=0)
5
6     return fast_image
```

```
1 X1_fast = np.array([feature_fast(k[:, :, 0]) for k in X_train]).reshape(X_train.shape[0], IMG_SIZE*IMG_SIZE)
```

```
1 X3_fast = np.array([feature_fast(k[:, :, 0]) for k in X_test]).reshape(X_test.shape[0], IMG_SIZE*IMG_SIZE)
```

```
1 plt.subplot(1, 2, 1)
2 plt.imshow(X_train[0])
3 plt.title('X_train[0]')
4
5 plt.subplot(1, 2, 2)
6 plt.imshow(feature_fast(X_train[0][:, :, 0]))
7 plt.title('X_train[0] FAST')
8
9 plt.show()
```



```
1 from sklearn import svm
2
3 svc_fast = svm.SVC(kernel='linear')
```

```
1 svc_fast.fit(X1_fast, y1)
```

```
▼ SVC
SVC(kernel='linear')
```

```
1 y_pred_svc_fast = svc_fast.predict(X3_fast)
```

```
1 accuracy_score(y3, y_pred_svc_fast)*100
```

```
75.32051282051282
```

```
1 precision_score(y3, y_pred_svc_fast)*100
```

```
71.93308550185874
```

```
1 recall_score(y3, y_pred_svc_fast)*100
```

```
99.23076923076923
```

```
1 f1_score(y3, y_pred_svc_fast, average='macro')*100
```

```
67.64008620689656
```