X-ray's prediction

UNIVERSITY OF PIRAEUS – DEMOCRITUS INSTITUTE

ARTIFICIAL INTELLIGENCE, MASTER

MACHINE LEARNING, COURSE

THEODOROS ADAMANTIDIS, TN2001

ATHENS, FEBRUARY 2021

Content

Normal   Bacterial Pneumonia   Viral Pneumonia

Introduction

- Detect pneumonia on X-Ray images.

- 5,863 X-Ray images of two categories

- Chest X-ray images (anterior-posterior) were selected from retrospective cohorts of pediatric patients of one to five years old from Guangzhou Women and Children's Medical Center, Guangzhou. All chest X-ray imaging was performed as part of patients' routine clinical care.

Feature extraction (1/3)

- Corresponding notebook **ExtractFeatures.ipynb** in GitHub.

- Developed and support 3 different types of image extraction
  - Plain flatten image array
  - KAZE descriptors array
  - HOG descriptors array

- All images resized to 100x100 size.

- Data are stored after extraction into pickle files for fast accessing in the experiments.
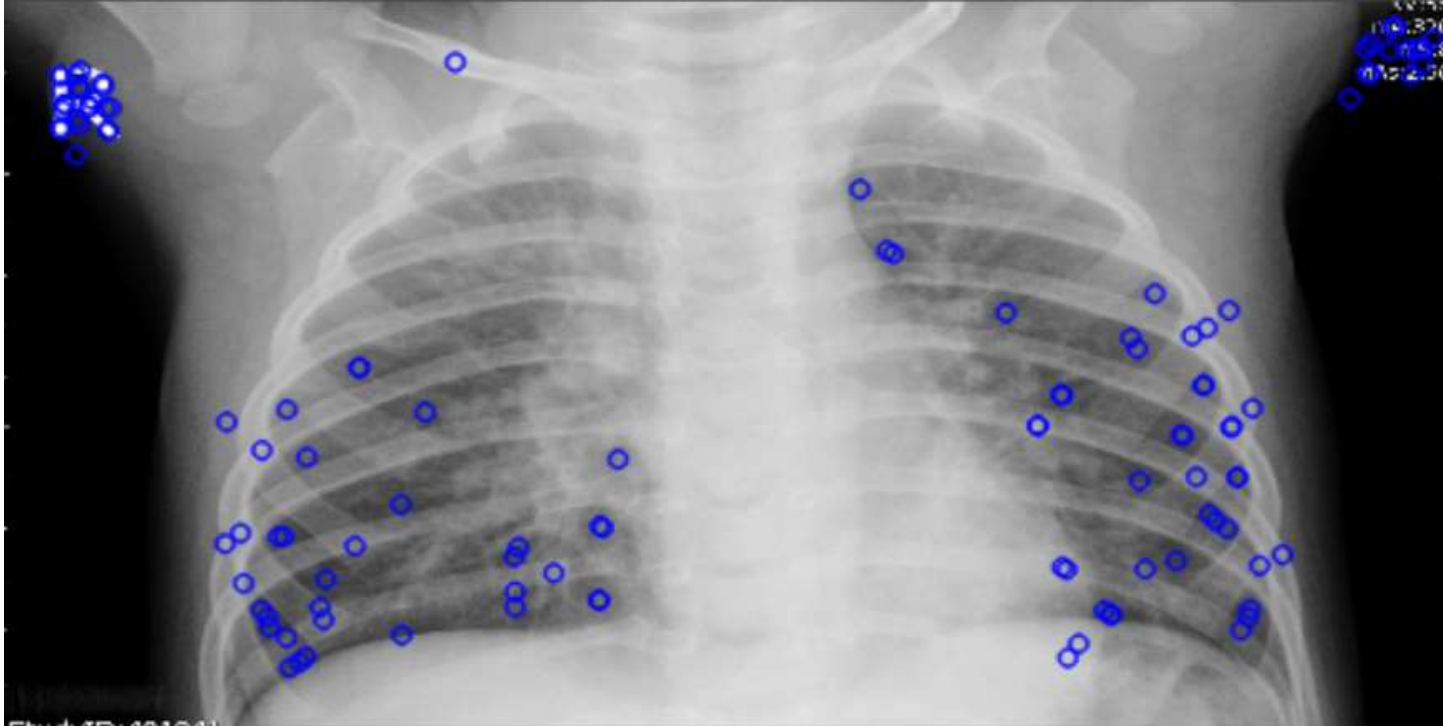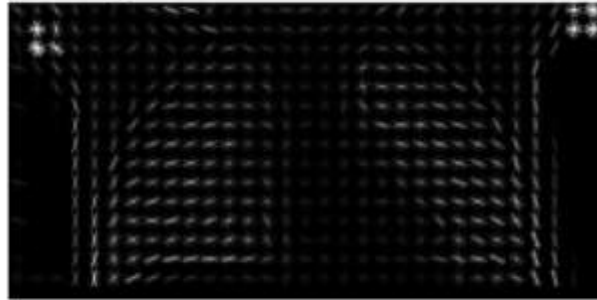
Feature extraction (2/3) - KAZE

- Detect and get the first 32 key points of the image ordered by their response value.

- Compute descriptor using these key points constructing a vector of 64 size.

- In case of the vector is less than 64, fill it with zeros.

Feature extraction (3/3) - HOG

Input image


Histogram of Oriented Gradients

- Cell size 10x10

- Block size 2x2

- Number of orientation bins 9

- Calculates descriptor and returns the array flatten.

Experiments (1/7)

- For the experiments used the pickle files extracted from the feature extraction step.

- Standardize features.

- Used to measure how "good" the model estimators are
  - Accuracy
  - Confusion matrix
  - Precision score

- Because of the quite difference in the number of each classes data, the accuracy by it self is not a good estimator.

- The models used are
  - Logistic regression
  - KNN
  - Discriminant analysis
  - Gaussian naive bayes
  - SVM

- Used GridSearchCV with RepeatedStratifiedKFold to tune models

Experiments (2/7)
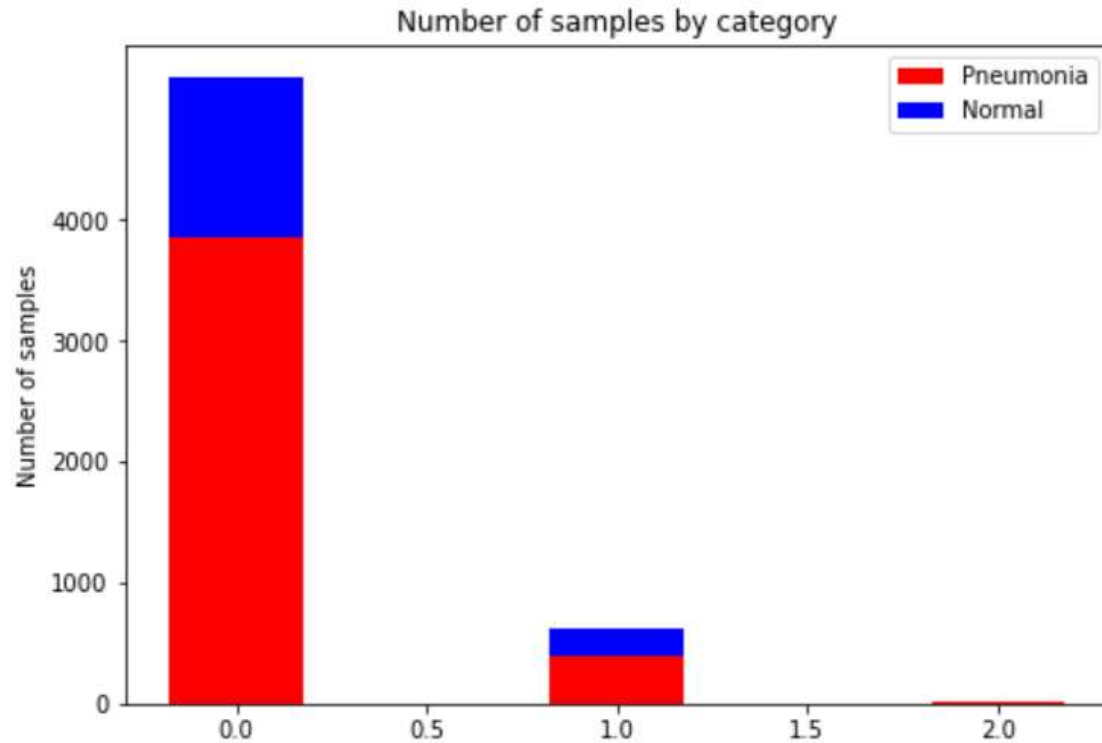
- Number of samples in training data: 5216
  - Number of samples in training data where category is Pneumonia : 3875
  - Number of samples in training data where category is Normal: 1341

- Number of samples in testing data: 624
  - number of samples in testing data where category is Pneumonia : 390
  - number of samples in testing data where category is Normal: 234

- Number of samples in validate data: 16
  - number of samples in validate data where category is Pneumonia : 8
  - number of samples in validate data where category is Normal: 8

Experiments (3/7) – Logistic regression

No tuned

- Accuracy LR classifier on training set: 1.00
- Accuracy of LR classifier on test set: 0.80
- Accuracy of LR classifier on validate set: 0.50
- Precision score 0.7661290322580645

Pred/Label	Normal	Pneumonia
Normal	118	116
Pneumonia	10	380

Tuned (C = 0.01)

- Accuracy LR classifier on training set: 1.00
- Accuracy of LR classifier on test set: 0.80
- Accuracy of LR classifier on validate set: 0.50
- Precision score 0.768762677484787

Pred/Label	Normal	Pneumonia
Normal	120	114
Pneumonia	11	379

Experiments (4/7) – KNN

No tuned

- Accuracy of K-NN classifier on training set: 0.96

- Accuracy of K-NN classifier on test set: 0.81

- Accuracy of K-NN classifier on validate set: 0.50

- Precision score 0.8096330275229358

Pred/Label	Normal	Pneumonia
Normal	151	83
Pneumonia	37	353

Tuned (neighbors = 11, weight = 'distance', alg = kd_tree)

- Accuracy of K-NN classifier on training set: 1.00

- Accuracy of K-NN classifier on test set: 0.82

- Accuracy of K-NN classifier on validate set: 0.50

- Precision score 0.768762677484787

Pred/Label	Normal	Pneumonia
Normal	158	76
Pneumonia	37	353

Experiments (5/7) – Discriminant analysis

No tuned

- Accuracy of LDA classifier on training set: 1.00
- Accuracy of LDA classifier on test set: 0.79
- Accuracy of LDA classifier on validate set: 0.50
- Precision score 0.766260162601626

Tuned (solver = 'lsqr')

- Accuracy of LDA classifier on training set: 1.00
- Accuracy of LDA classifier on test set: 0.79
- Accuracy of LDA classifier on validate set: 0.50
- Precision score 0.7676348547717843

Pred/Label	Normal	Pneumonia
Normal	119	115
Pneumonia	13	377

Pred/Label	Normal	Pneumonia
Normal	122	112
Pneumonia	20	370

Experiments (6/7) – Gaussian naive bayes

- Accuracy of GNB classifier on training set: 0.89

- Accuracy of GNB classifier on test set: 0.76

- Accuracy of GNB classifier on validate set: 0.50

- Precision score 0.780373831775701

Pred/Label	Normal	Pneumonia
Normal	140	94
Pneumonia	56	334

Experiments (7/7) – SVM

No tuned

- Accuracy of SVM classifier on training set: 1.00

- Accuracy of SVM classifier on test set: 0.79

- Accuracy of SVM classifier on validate set: 0.50

- Precision score 0.7589641434262948

Pred/Label	Normal	Pneumonia
Normal	113	121
Pneumonia	9	381

Tuned (C = 100, kernel = 'Linear')

- Accuracy of SVM classifier on training set: 1.00

- Accuracy of SVM classifier on test set: 0.80

- Accuracy of SVM classifier on validate set: 0.50

- Precision score 0.7703252032520326

Pred/Label	Normal	Pneumonia
Normal	121	113
Pneumonia	11	379

Results

- Model:: **Logistic regression** achieved:: 0.7661290322580645 precision score.

- Model:: **Logistic regression tuned** achieved:: 0.768762677484787 precision score.

- Model:: **KNN** achieved:: 0.8096330275229358 precision score.

- Model:: **KNN tuned** achieved:: 0.8228438228438228 precision score.

- Model:: **Discriminant analysis** achieved:: 0.7676348547717843 precision score.

- Model:: **Discriminant analysis tuned** achieved:: 0.766260162601626 precision score.

- Model:: **Gaussian naive bayes** achieved:: 0.780373831775701 precision score.

- Model:: **SVM** achieved:: 0.7589641434262948 precision score.

- Model:: **SVM tuned** achieved:: 0.7703252032520326 precision score.