Machine Learning Techniques for Lung Pneumonia Diagnosis
Aleksandar Skrbic RA153/2015
Soft Computing - E2 2018/2019

Problem Definition

A lung pneumonia classifier from X-ray scans is developed by applying different feature extraction methods and supervised machine learning algorithms to predicit whether lungs are healthy or have pneumonia.

Data Set

The dataset is organized into 3 folders (train, test, val) and contains subfolders for each image category (Pneumonia/Normal). There are 5,863 X-Ray images (JPEG) and 2 categories (Pneumonia/Normal).
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.
For the analysis of chest x-ray images, all chest radiographs were initially screened for quality control by removing all low quality or unreadable scans. The diagnoses for the images were then graded by two expert physicians before being cleared for training the AI system. In order to account for any grading errors, the evaluation set was also checked by a third expert.

Learning Methodology and Results

Learning Algorithms

- Support Vector Machine
- Gradient Boosting Classifier
- Multilayer Perceptron Neural Network
- Convolutional Neural Network

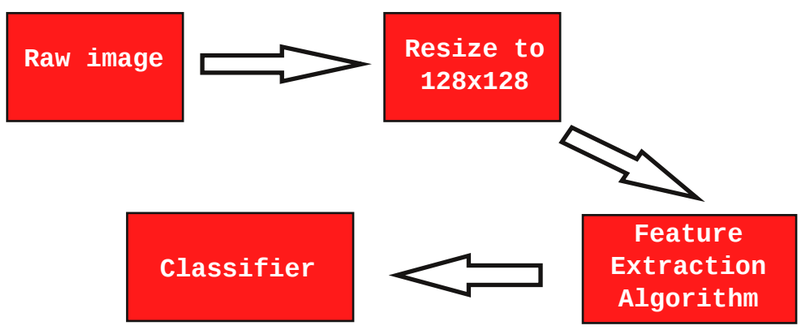
Results

Methodology	Test Accuracy
HOG + SVM	96.06%
LBP + SVM	72.09%
LBP + XGB	86.90%
LBP + MLP	86.90%
CNN	90.23%
AVG FILTER + SVM	93.49%
AVG FILTER + XGB	94.92%

Feature Extraction Algorithms

- Histogram of oriented gradients
- Local binary patterns
- Average filter
- Convolutional layers

Pipline



Confusion matrix for final model

Conclusion and Future Work

Final model get 96.06% accuracy and that is combination of HOG feature extractor and SVM classification algorithm. Results of classification are acceptable.

For future improvements, i'm planing to use autoencoder neural network for feature extraction and use output from autoencoder as input to SVM, XGB or MLP. Also there is a possibility that some more complex CNN architecture can give much better results that current architectures.