Pneumonia Detection Using Convolutional Neural Network

UNIVERSITY OF HOUSTON

Chenxin Lyu Vincent Lin Xinyan Liu

Computer Science, University of Houston

Abstract

In modern medicine, a Chest X-ray image is the most important diagnosis for pneumonia. In another word, a pneumonia patient requires a chest X-ray to diagnose. In the Chest X-Ray - pneumonia detection project, we discussed the vectors that may help to detect pneumonia in X-ray images. We conduct a comprehensive analysis to list the differences between normal lung and infected (pneumonia) lung images. By incorporating with CNN image classification algorithm, we concluded the result that was able to prove the expectation. After analyzing and preprocessing all the chest X-ray images, we use the knowledge of transfer learning to improve diagnostic performance by feature fusion. CNN is an example of a deep neural network specialized in image analysis. Therefore, it is widely used in the field of computer vision. For instance, image classification, image clustering, object detection, and neural style transfer.

Background

Since 2020, the covid-19 has changed the way that we live and work, and it does damage some of us and the people around us. However, currently, chest X-ray is the best method for detecting pneumonia, and it plays an important role in clinical care and epidemiological studies. However, the detection of pneumonia by X-ray images is a challenging task that relies on the expertise of radiologists. However, with the rise of deep learning, convolutional neural networks can be accurately applied to image recognition and processing, providing new technical and methodological support for medical diagnosis. CT screening is helpful to detect and isolate the source of infection effectively, cut off the route of transmission, and detect and isolate asymptomatic infected persons in time, which has a positive effect on reducing severe patients and reducing the mortality rate. At the same time, when imaging features of pneumonia are detected in suspected patients, early isolation and treatment can accelerate the standard treatment of patients, which is beneficial to control the development of the epidemic. The chest image dataset was from Kaggle, it contains both test and train sections, and there are two types of images in each group which present normal lungs and lungs that are infected. Our goal is to perform the CNN method and use the model to predict pneumonia based on input chest X-ray images.

Methods

SVM:
Support vector machine, It is a supervised learning model and related learning algorithm for analyzing data in classification and regression analysis. Given a set of training instances, each marked as belonging to one or the other of the two classes, the SVM training algorithm creates a model that assigns new instances to one of the two classes, making it a non-probabilistic two Metalinear classifier.

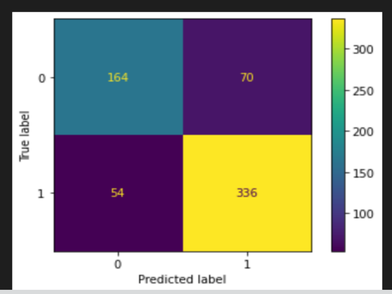
Logistic Regression:
It is a generalized linear regression analysis model, which is often used in data mining, automatic disease diagnosis, economic forecasting and other fields.

CNN:
Convolutional Neural Network (CNN) is a kind of deep neural network, specializing in image analysis, so it is widely used in computer vision applications, such as image classification and clustering, object detection, and neural style transformation.

Logistic Regression Classfication Report

	precision	recall	f1-score	support
0	0.87	0.87	0.87	1341
1	0.87	0.87	0.87	1342
accuracy			0.87	2683
macro avg	0.87	0.87	0.87	2683
weighted avg	0.87	0.87	0.87	2683

Logistic Regression Confusion Matrix



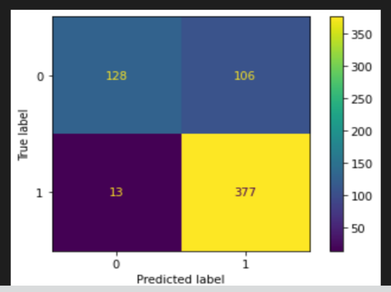

```
print(f'CNN with out image augmentation accuracy: {accuracy_score(test_y,y_pred)}')
```

```
CNN with out image augmentation accuracy: 0.8125
```

SVM Classfication Report

	precision	recall	f1-score	support
0	0.96	0.93	0.95	1341
1	0.93	0.96	0.95	1342
accuracy			0.95	2683
macro avg	0.95	0.95	0.95	2683
weighted avg	0.95	0.95	0.95	2683

SVM Confusion Matrix




```
print(f'CNN with image augmentation accuracy: {accuracy_score(test_y,y_pred)}')
```

```
CNN with image augmentation accuracy: 0.8670
```

Results

We used SVM and logistic models to predict whether the image is normal or pneumonia because the models are good at predicting classification outcomes. In SVM, we got 0.95 accuracy score. In logistics, we got 0.87 accuracy score. Then, we use CNN to do a deep learning model. We got 0.81 accuracy score in CNN and got 0.86 accuracy score with image augmentation.

Conclusion

The CNN model with image augmentation accuracy is better than without image augmentation. With SVM, we got 0.95 accuracy, which is high. Logistic regression also has about 0.87 accuracy. Compare the results of models, SVM is the most efficient method.

Acknowledgments

https://nnart.org/what-is-data-augmentation-in-a-cnn/
https://blog.csdn.net/weixin_40651515/article/details/105398458
https://hanifi.medium.com/sequential-api-vs-functional-api-model-in-keras-266823d7cd5e

Future Direction

We can use PCA and hyperparameters to improve our model accuracy score.

PCA: It is a statistical method. It is only necessary to measure the amount of information in terms of variance and is not affected by factors outside the dataset. The principal components are orthogonal, which can eliminate the mutual influence factors between the original data components.

Hyperparameter optimization: It is the topology and size of a neural network. It can reduce the value of the loss function, so that it maximizes the model performance.