Chest X-Ray Image Classification with Deep Learning

Flatiron School

Kamile Yagci 4/14/2022

Content

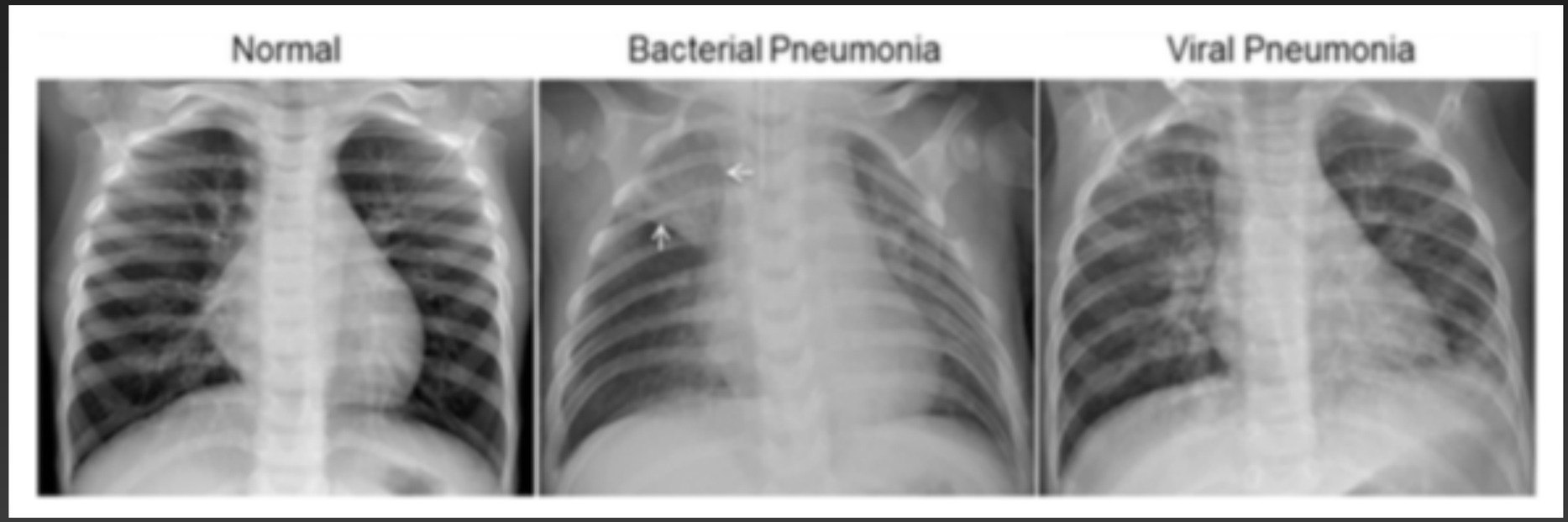
- Business Problem
- Data
- Method
- Analysis and Results
- Conclusion
- Next Steps
- Questions

Business Problem

- The Baylor Medical Center hired me to improve the accuracy in pneumonia diagnosis on pediatric patients.
- The study will use the chest X-ray Images of pediatric patients and do the image classification identifying whether the X-ray shows pneumonia or not.
- The outcome of this study will not only be used at Baylor Centers, but also in partner medical clinics in Africa, where the medical staff is limited.
- The automated identification system will provide early diagnosis of pediatric patients, so the treatment can start as soon as possible.
- Moreover it will decrease the human errors in pneumonia diagnosis.

Data (1)

- Dataset 'Chest X-Ray Images (Pneumonia)' is downloaded from Kaggle.
- Contains the chest x-ray images of pediatric patients from one to five years old, collected at Guangzhou Women and Children’s Medical Center.
- The diagnosis on chest x-ray images have three types: Normal, Bacterial pneumonia and Viral pneumonia.



Data (2)

- Dataset contains 5856 chest x-ray image files:
  - 1583 NORMAL images (no pneumonia)
  - 4273 PNEUMONIA images (bacterial or viral pneumonia)
- I distributed the data as:
  - ~70% training
  - ~15% validation
  - ~ 15% testing

	Normal	PNEUMONIA	ALL
Train	1107	2991	4098
Validation	238	641	879
Test	238	641	879
All	1583	4273	5856

Method

- In this study, I applied Image Classification with Deep Learning using the Convolutional Neural Networks (CNN) on chest x-ray images.
- Since the data has two classes (NORMAL and PNEUMONIA), this is a binary image classification.

Analysis and Result (1)

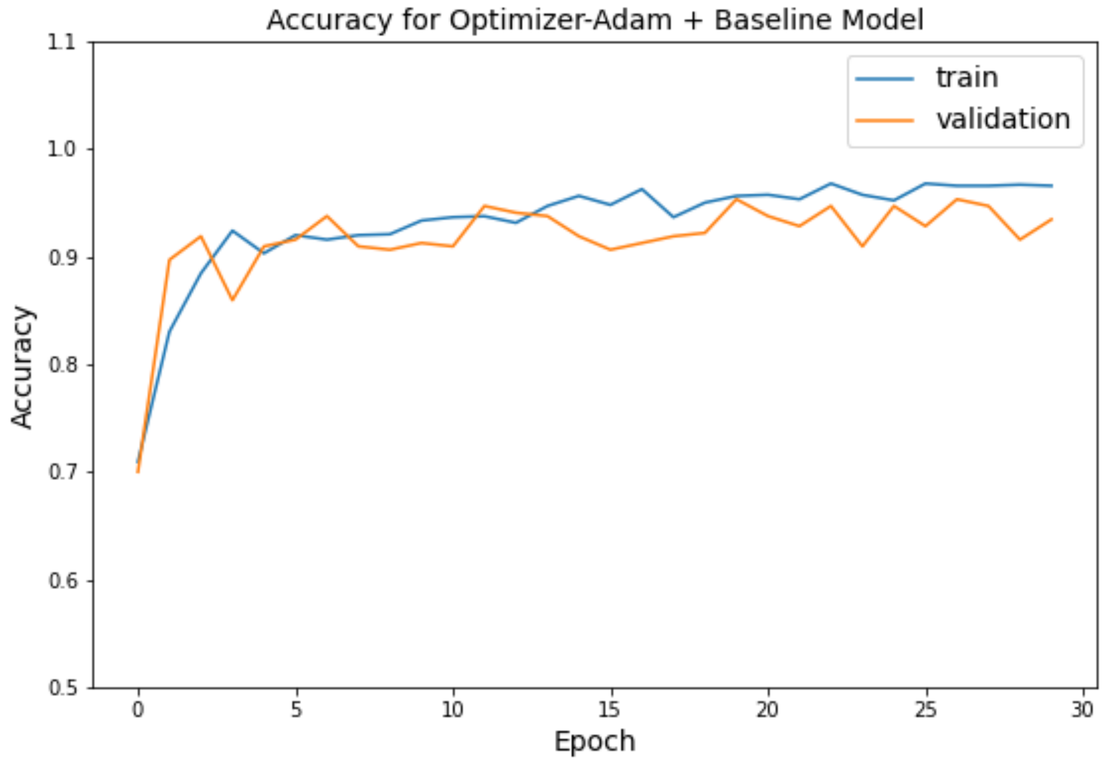
- In this study, limited computing power on my computer created a challenge.
- Image Classification Model built and trained considering this limitation:
  - Train model on subset data (~22% of whole train and validation data)
  - Use small image resolution
  - Limit the number of runs on training data (epoch)

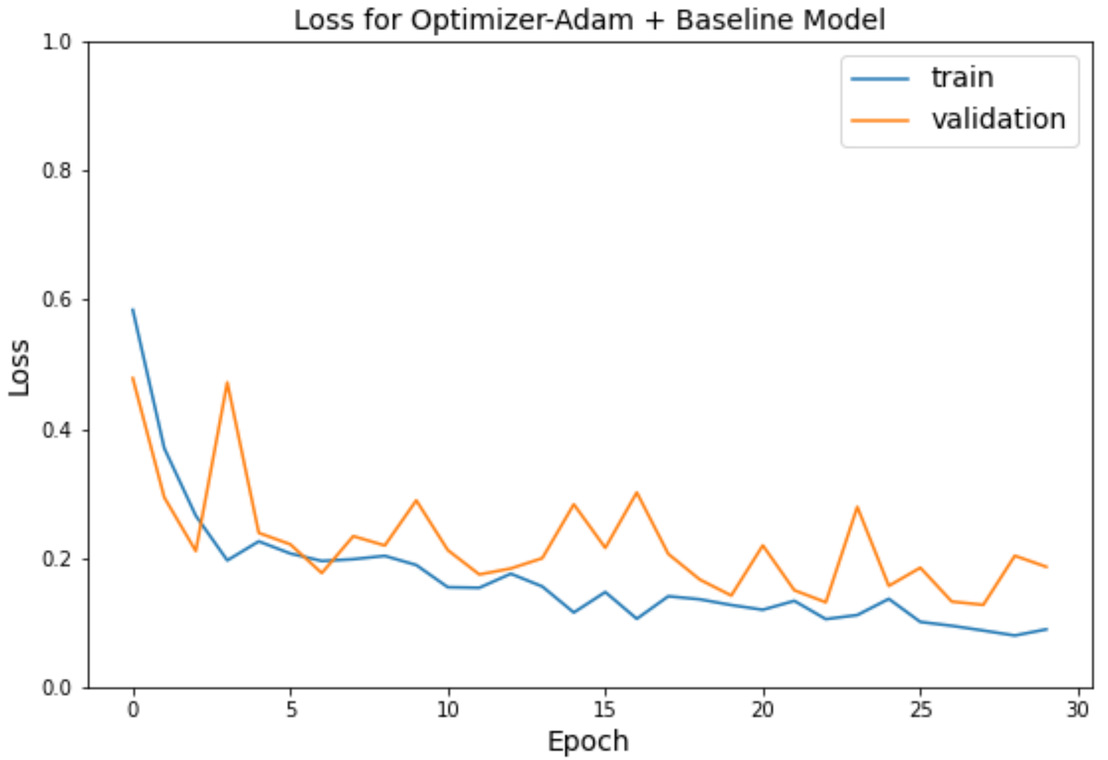
Analysis and Result (2)

- Model is evaluated according to:
  - Accuracy: Percentage of correct predictions
  - Loss: Measure of difference between predicted and true values
- Model is evaluated on whole training and testing data

Analysis and Result (3)

- Accuracy and Loss Graphs for Final Model
  - created in training process on subset data





Analysis and Result (4)

- Evaluation Results for Final Model:
  - Model prediction accuracy: 95.45%
  - evaluation calculated on whole train and test data
  - Trained on subset data

	Accuracy	Loss
Train	0.9736	0.0774
Test	0.9545	0.1214

- After final model is determined, I trained it on whole data. The prediction accuracy was similar.

Conclusion

- The final model has a very good performance.
- The model classifies NORMAL and PNEUMONIA chest-ray images with 95.45% accuracy.
  - 95.45% of the model predictions are excepted to be correct
- The final model performance on subset data and whole data is similar.

Next Steps

- Even though, the model performance is very good. There is always a room for improvement.
- I would like run all steps of this analysis on a powerful computer, or grid system.
  - use whole dataset for training
  - tune model on whole training data

Questions?