

BIRKBECK, UNIVERSITY OF LONDON

Pneumonia Detection from Chest X-Ray Images

Author:
Baran Buluttekın

Supervisor:
Dr. George Magoulas



*A project report submitted in fulfillment of the requirements
for the degree of MSc Data Science*

in the

Department of Computer Science

June 17, 2020

Declaration

I have read and understood the sections of plagiarism in the College Policy on assessment offences and confirm that the work is my own, with the work of others clearly acknowledged. I give my permission to submit my report to the plagiarism testing database that the College is using and test it using plagiarism detection software, search engines or meta- searching software.

The report may be freely copied and distributed provided the source is explicitly acknowledged.

Abstract

New test text. New words

Contents

Declaration	i
Abstract	ii
1 Introduction	1
1.1 Aims and Objectives	1
1.2 CI/CD Pipeline	1
1.3 Engineering Related Challenges	2
1.4 Project layout and design	2
1.5 Reproducibility Guidance	2
2 Data	3
2.1 Data Augmentation	3
2.2 Limitations of the Dataset	3
References	4

1 | Introduction

Introduction related material. Explain overall aim.

1.1 Aims and Objectives

1.2 CI/CD Pipeline

Overall description of CI/CD and role in the project. CI pipeline setup with travis.

Continuous integration (CI) is a workflow strategy that helps ensure everyone's changes will integrate with the current version of the project. This lets you catch bugs, reduce merge conflicts, and increase overall confidence of your software is working. While the details may vary depending on the development environment, most CI systems feature the same basic tools and processes. In most scenarios, a team will practice CI in conjunction with automated testing using a dedicated server or CI service. Whenever a developer adds new work to a branch, the server will automatically build and test the code to determine whether it works and can be integrated with the code on the main development branch. The CI server will produce output containing the results of the build and an indication of whether or not the branch passes all the requirements for integration into the main development branch. By exposing build and test information for every commit on every branch, CI paves the way for what's known as continuous delivery, or CD, as well as a related process called continuous deployment. Difference between continuous delivery and continuous deployment is that CD is the practice of developing software in such a way that you could release it at any time. When coupled with CI, continuous delivery lets you develop features with modular code in more manageable increments. Continuous development is an extension of continuous delivery. It's a process that allows you to actually deploy newly developed features into production with confidence, and experience little, if any, downtime.

In more granular detail, this system works with central version control services and in this project central version control service used is Github. GitHub uses what are called *webhooks* to send messages to external systems about activity and events that occur in the project. For each event type, subscribers will receive message related to the event. Generally events refer to action involving the software such as new commit push or pull (merge) request and more. In this case whenever new commit is pushed to the any branch of the project. Message from Github will be

send out to third party system called *travis* ¹. Travis is a hosted CI service that allow build and test software hosted in version control services. Upon completing the build and test travis then passes the test results via Github API which will be visible to the developer.

1.3 Engineering Related Challenges

Role of hardware accelerators. Debugging related challenges of Neural Networks

1.4 Project layout and design

Overall project design such as folders and library interactions.

1.5 Reproducibility Guidance

Parts to consider before reproducing. Repo is not public Most runs carried out in colab notebooks should be positioned in same level as the src library running shell utility files requires kaggle api key

¹<https://travis-ci.org/>

2 | Data

Reminder of the dataset. [1]

2.1 Data Augmentation

Data augmentations applied and example results of the augmentation.

2.2 Limitations of the Dataset

Discussing different representation of the dataset will be covered in neural network section.

References

- [1] *Open Access Biomedical Image Search Engine*. <https://openi.nlm.nih.gov/>. Accessed: 2019-03-12.