UNIT 7: Final Capstone Proposal

September 2019

Overview

For my final capstone I will be using RSNA intracranial hemorrhage detection data set in order to identify acute intracranial hemorrhage and its subtypes from the CT study images of patients. The CT images can confirm bleeding and the evidence of trauma in head. Correct diagnosis of presence of hemorrhage and its type looking at the radiological report of the patient helps timely and effective care. The dataset consists of CT studies of patients and the probability of whether the type of hemorrhage exists or not. My objective is to build an algorithm that can detect hemorrhage and its type which could be a valuable information for the medical community to make data driven decisions.

Issues and Goals

1. What is the problem we are attempting to solve?

Intracranial hemorrhage, bleeding that occurs inside the cranium, is a serious health problem that requires rapid and often intensive medical treatment. Intracranial hemorrhages account for approximately 10% of strokes in the U.S., where stroke is the fifth-leading cause of death. Identifying the location and type of any hemorrhage present is a critical step in treating the patient.

Diagnosis requires an urgent procedure. When a patient shows acute neurological symptoms such as severe headache or loss of consciousness, highly trained specialists review medical images of the patient's cranium to look for the presence, location and type of hemorrhage. The process is complicated and often time consuming.

Through this capstone I aim to build an algorithm that can detect acute intracranial hemorrhage and its subtypes looking at the CT study images of the patients that can help the medical community identify the presence, location and type of hemorrhage in order to quickly and effectively treat affected patients.

2. How is your solution valuable?

The algorithm will detect the presence, location and type of hemorrhage based on computed tomography (CT) scan. On time diagnosis of presence, location, and type of hemorrhage informs the required urgent procedure for prompt and effective treatment of patients.

This piece of work can be an asset since the medical community sees the potential for AI to assist in detection and classification of hemorrhages, a condition that requires emergency care and treatment.

3. What is your data source and how will you access it?

The data source is https://www.kaggle.com/c/rsna-intracranial-hemorrhage-detection/overview/description where a collection of files summing up 156GB is available. The data consists of CT images as stage-1_train.zp and stage-1_test.zip and data for train and test. The dataset also consists of IDs of the patients as (stage_1_train.csv) (stage_1_sample_submission.csv) and multiple labels, one for each of five sub-types of hemorrhage, plus an additional label for any, which should always be true if any of the sub-type labels is true.

I will use a Kaggle kernel for this project and hence not necessary to download the data set to my local machine.

4. What techniques from the course do you anticipate using?

I will be preprocessing and cleaning the data set using pandas and balancing the data if required. Initially I will perform exploratory data analysis before modeling. Since my specialization was aimed to deep learning, I'll run Neural Network models such and Multi-Layer Perceptron(MLP) and Convolutional Neural Network (CNN) and evaluate the performance of the models according to the accuracy score. I will also use other traditional supervised learning algorithms to compare the scores. I will try using hyper parameter tuning techniques, such as GridSearchCV to optimize the model's performance.

5. What do you anticipate being the biggest challenge you will face?

The biggest challenge I will be facing is downloading this big data set and the preprocessing power of my personal computer to compute this amount of data and run the models accordingly. To overcome this challenge, I will be using Kaggle notebook which is a cloud computational environment that enables reproducible and collaborative analysis.

Conclusion

I believe that the objective of this project will help the medical community identify the presence, location and type of hemorrhage in order to quickly and effectively treat affected patients.

Potential next steps and ideas for productionization

We can create an app with more time/bandwidth for doctors/nurses to interactive with the model
We can make the model to self-learn/modify whenever there is new data.