Siddarth MV
2020A4PS1810G

Research Paper Formulation

Detection of Malignant Tumors in the Lung from Chest CT Scans

OVERVIEW

A model on detection of malignant lung tumors based on **A How-to-Model Guide for Neuroscience** by *Gunnar Blohm, Konrad P. Kording,* and *Paul R. Schrater.*

The 10 step model has been categorized into 4 parts

Framing the Question

Step 1 : Finding a Phenomenon and framing a question around it

- Phenomenon : Presence of Visually Identifiable (from CT Scans) microscopic tumors in lungs which may be malignant causing cancer to spread to other parts of the body
- What if the monotonous tasks can be automated with the help of Computer Vision and Machine Learning
- 3 Dimensional LDCT (Low Dose Computer Tomography) Scan Datasets of Human Torso can be taken as the input

Step 2 : Understanding the State of the Art

- Many different algorithms such as Convolutional Neural Networks (CNNs) , Vector Machines, Dynamic Bayesian Networks (DBNs), Machine Learning, Deep Learning and their variations and/or combinations have been used for achieve the goal

- An extensive knowledge of the various algorithms, their implementations and use cases will yield a better understanding on where to apply the algorithms to specific cases
- A strong foundation and set of skills in Programming and Data Science along with a good understanding of tumor cells is necessary to achieve the goal

Step 3 : Determining the basic ingredients

- The basic outlook would be to determine if a tumor is malignant or benign when a CT scan is fed into the model
- During the data loading process necessary transformations are performed so that it can be used in the model.
- The classification model can use algorithms such as CNN , Deep Learning etc.
- Metrics and critical analysis could be visually displayed at the end
- The dataset could also be labelled for certain conditions such as smoking status which could be later used in accuracy testing

Step 4 : Formulating Specific mathematically defined hypotheses

- A model could be created to use the fed CT scans and automatically detect malignant cancerous tumors which could use Deep Learning to train the model

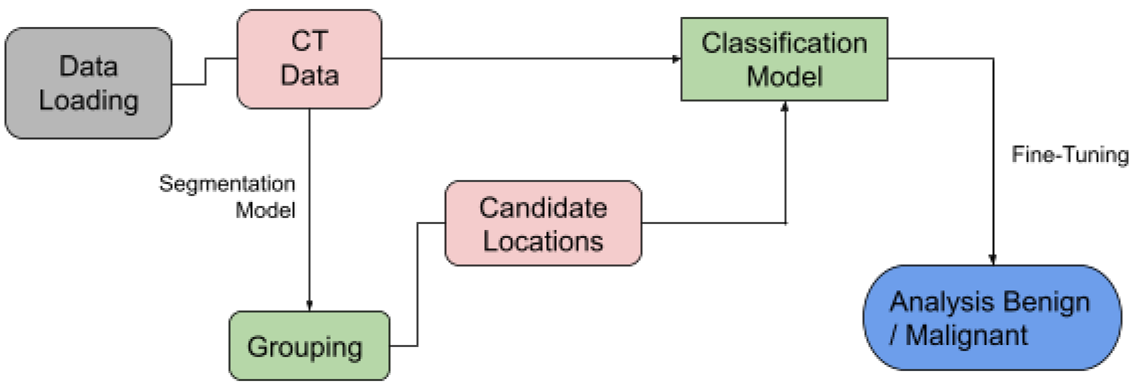
Implementing the Model

Step 5 : Selecting the toolkit

- There are many ways to implement the models and toolkits to use, programming languages such as C, C++, Python etc., could be used but are inefficient when compared to frameworks specifically crafted for Machine Learning such as Tensorflow, Torch, Scikit-Learn etc.

- An extensive practical knowledge and experience in programming and using these frameworks is necessary to implement the model in an efficient manner
- Torch framework along with python (Pytorch) has been used to implement this model since it is quite flexible and contains predefined functions such as transformations which need not be instantiated again , so code can be reused
- Torch framework use tensors which could yield faster results as tensors could work with GPUs which are optimized to process multiple computations simultaneously

Step 6 : Planning the Model



- The model would use datasets of CT scans. It would be transformed (like converting the raw data to tensors) in order to be fed into the model
- The complete transformed cata could be used or be further trimmed down to flag individual pixels to be fed to the classification algorithm
- The Classification algorithm uses Convolutional Neural Network (Deep Learning) , optimizers (SGD) and loss functions to train the model and hyperparameter tuning is done so as to get the optimal result and increase the accuracy
- Tensorboard and matplotlib libraries can be used to visualize the results and data

Step 7 & 8 : Implementation and completion of the model

The code and its implementation can be found at this link :

https://github.com/deep-learning-with-pytorch

Model Testing

Step 9 : Testing and Evaluating the Model

- The predictions from the classification model could be partitioned into 4 quadrants : True positives, False positives, True negatives and False Negatives which can provide the basis for our improved performance metrics
- The model accuracy could be further broken down into different categories upon related conditions such as smoking status to see how it performed on patients under certain conditions
- Different descriptive indices such as AUC, accuracy, sensitivity and specificity could be used to assess the performance of the model
- The model could be generalized so as to identify other diseases in and around the chest area such as COPD, Asthma, TB etc