Deep Learning Applications in

Medical Image Analysis

**ABSTRACT:**

Brain tumor detection and classification is the most difficult and tedious task in the area of medicinal image preparing. MRI (Magnetic Resonance Imaging) is a medicinal procedure, generally adopted by the radiologist for representation of inner structure of the human body with no surgery. The tremendous success of machine learning algorithms at image recognition tasks in recent years intersects with a time of dramatically increased use of electronic medical records and diagnostic imaging. This review introduces the machine learning algorithms as applied to medical image analysis, focusing on convolutional neural networks, and emphasizing clinical aspects of the field. The advantage of machine learning in an era of medical big data is that significant hierarchal relationships within the data can be discovered algorithmically without laborious hand-crafting of features. We cover key research areas and applications of medical image classification, localization, detection, segmentation, and registration. We conclude by discussing research obstacles, emerging trends, and possible future directions.

**INDEX TERMS** Convolution neural networks, medical image analysis, machine learning, deep learning.

**EXISTING SYSTEM:**

There is a myriad of imaging modalities, and the frequency of their use is increasing. Smith-Bindman *et al.* looked at imaging use from 1996 to 2010 across six large integrated healthcare systems in the United States, involving 30.9 million imaging examinations. The authors found that over the study period, CT, MRI and PET usage increased respectively.

The symbolic AI paradigm of the 1970s led to the development of rule-based, expert systems. One early implementation in medicine was the MYCIN system by Short life , which suggested different regimes of antibiotic therapies for patients. Parallel to these developments, AI algorithms moved from heuristics-based techniques to manual, handcrafted feature extraction techniques. and then to supervised learning techniques. Unsupervised machine learning methods

are also being researched, but the majority of the algorithms from 2015-2017 in the published literature have employed supervised learning methods,

**DISADVANTAGES**

In image processing processing techniques used different types of filters and Fourier and discrete transform it increases the complexity the cost of those equipment also high. To know the result of the tumor concerned person has to be there. This diagnosis perform some particular equipment only

**PROPOSED SYSTEM:**

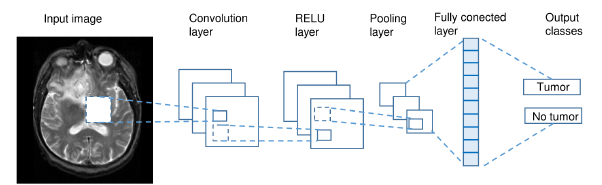
Detection, sometimes known as Computer-Aided Detection is a keen area of study as missing a lesion on a scan can have drastic consequences for both the patient and the clinician. The task for the Kaggle Data Science Bowl of 2017 involved the detection of cancerous lung nodules on CT lung scans. Approximately 2000 CT scans were released for the competition and the winner Fangzhou achieved a logarithmic loss score of 0.399. Their solution used a 3-D CNN inspired by U-Net architecture to isolate local patches first for nodule detection. Then this output was fed into a second stage consisting of 2 fully connected layers for classification of cancer probability. Shin *et al.*evaluated five well-known CNN architectures in detecting thoracoabdominal lymph nodes and Interstitial lung disease on CT scans. Detecting lymph nodes is important as they can be a marker of infection or cancer. They achieved a mediastinal lymph node detection AUC score of 0.95 with a sensitivity of 85% using GoogLeNet, which was state of the art. They also documented the benefits of transfer learning, and the use of deep learning architectures of up to 22 layers, as opposed to fewer layers which was the norm in medical image analysis. Overfeat was a CNN pre-trained on natural images that won the ILSVRC 2013 localization task . Ciompi applied Overfeat to 2-dimensional slices of CT lung scans oriented in the coronal, axial and sagittal planes, to predict

the presence of nodules within and around lung fissures. They combined this approach with simple SVM and RF binary classifiers, as well as a Bag of Frequencies, a novel 3-dimensional descriptor of their own invention.

**ADVANTAGES:**

It is evident that input data plays an important role in prediction along with machine learning techniques. As is seen in the dataset, provided, we have labels from 0 to 4 where the labels of 4 are hardly 13 and when we split the data into train and test, the number become very less which is nothing but noise and can be totally removed from the dataset by using filtering techniques and hence the linear model will be available to predict the outcome much better with absence of noise. It not only helps us in predicting the outcome but also gave us valuable insights about the nature of data, which can be used in future to train our classifiers in a much better way.

**SYSTEM ARCHITECTURE:**



**Modules:**

Importing Dataset:

Initial step in our project is to import image dataset which are having both tumor contained images, and non tumor x-ray images.

Splitting the data:

After importing the dataset need to split the data into two parts like training and testing datasets. Based on split size need to split the data

Building neural networks:

We need to feed the convolution neural networks with trained with different layers need to give,after that giving the testing the data for analyzing the result

Prediction:

Finally give on x-ray image to neural networks it will checks that weather that image having tumor or not and then gives the result

**Numpy:**

NumPy enriches the programming language Python with powerful data structures, implementing multi-dimensional arrays and matrices. These data structures guarantee efficient calculations with matrices and arrays. The implementation is even aiming at huge matrices and arrays, better know under the heading of "big data". Besides that the module supplies a large library of high-level mathematical functions to operate on these matrices and arrays.

**pandas**

Pandas is a high-level data manipulation tool developed by Wes McKinney. It is built on the Numpy package and its key data structure is called the DataFrame. DataFrames allow you to store and manipulate tabular data in rows of observations and columns of variables.

**Sklearn**

Scikit-learn provides a range of supervised and unsupervised learning algorithms via a consistent interface in Python.

It is licensed under a permissive simplified BSD license and is distributed under many Linux distributions, encouraging academic and commercial use

**matplotlib**

It is a plotting library used for 2D graphics in python programming language. It can be used in python scripts, shell, web application servers and other graphical user interface toolkits.

There are several toolkits which are available that extend python matplotlib functionality. Some of them are separate downloads, others can be shipped with the matplotlib source code but have external dependencies.

**Seaborn:**

Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.

**SYSTEM SPECIFICATIONS**

**SOFTWARE REQUIREMENTS:**

OS : Windows

Python IDE : python 2.7.x and above

Pycharm IDE,

Anaconda 3.5

Setup tools and pip to be installed for 3.6.x and above

**HARDWARE REQUIREMENTS:**

RAM : 4GB and Higher

Processor : Intel i3 and above

Hard Disk : 500GB: Minimum