Caption Generation from Medical Images

K.S. Seetharaman#1,P. Minu#2 M. Priyankaa#3, T. Mohana priya#4, F. Heena begaam#5
1Assistant Professor,Department of Computer Science and Engineering,Velammal College of Engineering and Technology,Tamil Nadu
2Final Year Student,Department of Computer Science and Engineering,Velammal College of Engineering and Technology,Tamil Nadu
3Final Year Student,Department of Computer Science and Engineering,Velammal College of Engineering and Technology, Tamil Nadu
4Final Year Student,Department of Computer Science and Engineering,Velammal College of Engineering and Technology, Tamil Nadu
5Final Year Student,Department of Computer Science and Engineering,Velammal College of Engineering and Technology, Tamil Nadu

Abstract

This paper proposes the neural network concept for caption generation from medical images by classifying the medical images. This project involves the process of detecting the occurrence of lung disease by producing the caption as normal or abnormal. GLCM (Gray level co-occurrence matrix) algorithm is used to extract the texture features of medical images. Artificial neural networks or connectionist systems are computing systems inspired by the biological neural networks that constitute animal brains. The neural network itself is not an algorithm, but rather a framework for many different machine learning algorithms to work together and process complex data inputs. The input to the caption generation model is a medical image(taking the organ as lungs).The goal of the task is to effectively identify the relevant medical concepts from medical images as a predictor of figure captions. The paper presents the procedure employed and provides an analysis of the obtained evaluation results.

Index Terms - *TB, X-ray images, chest radiographs, image processing, neural network, feature extraction, texture analysis.*

I. INTRODUCTION

Automatically describing the content of an image is a key challenge in artificial intelligence at the intersection of computer vision and natural language processing. This project produce the output as a caption about the occurrence of the TB in the input medical image. TB usually affects the posterior segment of the upper lobe. This could especially be beneficial to clinicians for useful insights and reduction of the significant burden on the overall workflow in patient care. Models are trained by using a large set of labeled data and neural network architectures that contain many layers. Various approaches are proposed in a number of papers dealing with the problems of automatic semantic tagging, and of automatic description generation of images. However, in medical imaging domain, this topic is yet to gain popularity. Obviously, medical image description poses its own set of problems.

II. LITERATURE SURVEY

Paper 1:- Texture analysis of TB X-ray images using image processing techniques.

The approach is to divide the separated lung fields in four different parts and analyze each part separately, with texture features extracted solely from these parts. In this way, the classifier should capture knowledge regarding the normal variation within that particular part. The avg. gray level, standard deviation (second moment), skew (third moment), uniformity, entropy etc., of each filtered image are computed as texture features.

As an initial step the images are obtained using image acquisition method and then the application of pre-processing algorithms including size normalization and filtering of the image is carried out. The features those are identified to be useful for diagnosis and analysis require separation of the lung fields from the background. Lung field masks are prepared to separate lung fields by estimating peripheral coordinates of lung fields manually, and using region based segmentation technique.

Paper 2:- Attention-based Medical Caption Generation with Image Modality Classification and Clinical Concept Mapping.

The recent advances in deep neural networks have been shown to work well for large scale image analysis tasks. . Hence, we use an encoder-decoder based deep neural network architecture, where the encoder uses a deep CNN to encode a raw medical image to a feature representation, which is in turn decoded using an attention-based RNN to generate the most relevant caption for the given image. We also utilize the same framework for clinical concept prediction to improve caption generation. Our experiments demonstrated that generating medical image captions by first predicting clinical concept IDs and then mapping them to all possible clinical terms in the ontology helps to improve the overall coverage of words in predicted captions.

Paper 3:- *Attentive linear transformation for image captioning.*

They propose a novel attention framework called attentive linear transformation (ALT) for automatic generation of image captions. Instead of learning the spatial or channel wise attention in existing models, ALT learns to attend to the high-dimensional transformation matrix from the image feature space to the context vector space. ALT attends to the high-dimensional transformation matrix from an image feature space to a context vector space. The advantage of ALT is that the weights in the linear transformation can capture information without a concrete form like spatial region or feature channel. Thus, ALT is able to attend to subtler and more abstract visual concepts than previous attention models. By using the proposed ALT, our caption model achieves state of art result on a few widely used benchmarks.

Paper 4:- *Generating captions for medical images with a deep learning multi-hypothesis approach: MedGIFT–UPB Participation in the ImageCLEF 2017 Caption Task*

The aim of the ImageCLEF concept detection task is to assign a set of medical concepts taken from a list of 20.463 possible pre-defined medical concepts to a biomedical image extracted from scholarly articles of the biomedical open access literature (PubMed Central). The challenge we addressed in our participation was the semantic gap between captions (text) and the image. Thus, the scientific challenge was to exploit these two types of information in order to automatically describe new medical images without any other resources than the training images provided. We investigate the network from a model-selection and optimization perspective. The task at hand is clearly very difficult without use of external resources.

Paper 5:- *Medical image captioning: learning to describe medical image findings using multi-task-loss CNN*

This paper presents a new multi-task-loss CNN based approach to joint automatic detection and semantic description of lesions in diagnostic images. The proposed approach outperforms the competing methods by up to 10%. . Automatic detection and classification of lesions in medical images remains one of the most important and challenging problems. The proposed CNN-based architecture is trained to generate and rank rectangular regions of interests (ROI's) surrounding suspicious areas. It has a clear advantage for supervised training on large datasets. The proposed approach generates standard radiological lexicon description which should help radiologists in understanding of the decision making process of CADx[Computer Aided Diagnosis] systems.

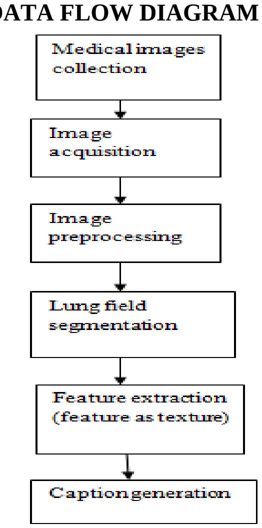
III. EXISTING SYSTEM

AdhiSusanto et al [5] have presented another paper which helps in lung tuberculosis detection. The data set used in this research work is thoracic X-ray images from Dr. Sarjito Hospital, Yogyakarta. Here, Image quality is enhanced using spatial filtering and histogram equalization techniques. Spatial filtering is done for noise reduction and histogram equalization is done for pixel intensity transformation. It also makes intensity level of the dataset uniform. Later, object isolation is done to get the ROI of the image. ROI is considered in different shapes. From these ROIs, features like mean, standard deviation, skewness, kurtosis and entropy are calculated. These five features are reduced to a single feature with the help of PCA.

IV. PROPOSED SYSTEM

In the proposed system, a caption is generated which indicates the occurrence of the lung disease such as normal or abnormal. This generation is done by performing some sequence of process. Initially image acquisition process is done which helps to het the input image. Then preprocessing is done to remove the noise or other irrelevant information from the input image. Followed by this, segmentation of lung field is done. This first change the input image into binary image and then the iamge is fragmented into four parts. Probabilistic neural network is applied on each fragmented part to extract the texture features. Based on the extracted features, a comparison is made with the some metrics. Finally the caption is generated which indicates the occurrence of the lung disease.

DATA FLOW DIAGRAM



V . MODULES

1. Image acquisition:

A particular image(medical image) is given as input to generate the caption. This method is used to obtain the input image.

```
[fn,pn]=uigetfile('*.jpg','Browse   an input Image');
I=imread([pn,fn]);
I=rgb2gray(I);
figure,imshow(I);
title('Input Image')
```

2. Image preprocessing:

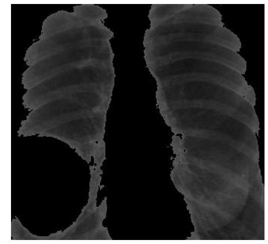
As an initial step the images are obtained using image acquisition method and then the application of pre-processing algorithms including size normalization and filtering of the image is carried out.

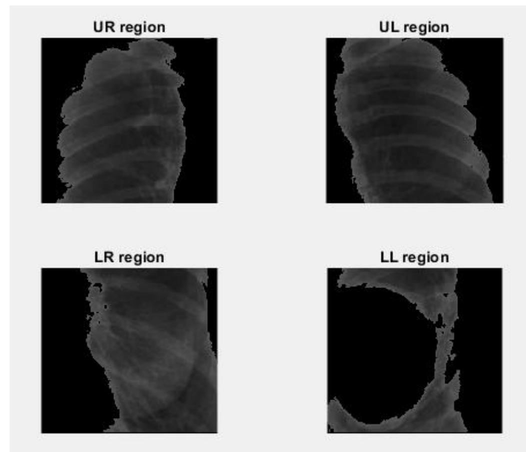
This process is also used to remove irrelevant information or noise using median filter.

```
L=graythresh(I);
I1=im2bw(I,0.35);
mask=imcomplement(I1);
figure,imshow(mask)
title('Binary Mask')
masked    =    bsxfun(@times,    I, cast(mask,class(I)));
figure,imshow(masked)
title('Seperated Lung Field')
```

3. Lung field segmentation:

The features those are identified to be useful for diagnosis and analysis require separation of the lung fields from the background. Lung field masks are prepared to separate lung fields by estimating peripheral coordinates of lung fields manually, and using region based segmentation technique.





4. Feature extraction:

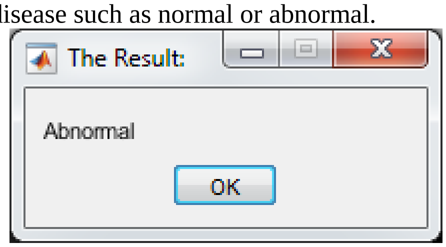
Every TB image (images are collected from public database) data is acquired with 256 gray levels (8 bits) and stored as JPEG (.jpg, .jpeg) data. Then the texture feature is extracted from the medical images which involves some concepts.

```
Features Extraction of UR region
Entropy         : 1.173097 1.211023 1.165618 1.231226
Avg. Gray level : 2276.000000 2695.000000 1983.000000 3157.000000
Std. Deviation  : 1.817447 1.784607 1.819395 1.764819
Smoothness      : 0.206652 0.244624 0.197511 0.270229
Third_moment    : 3.398398 3.400656 3.392368 3.400656
Uniformity      : 0.280327 0.340038 0.256451 0.389999

Features Extraction of UL region
Entropy         : 1.196801 1.293106 1.227146 1.272592
Avg. Gray level : 2378.000000 3476.000000 2218.000000 2892.000000
Std. Deviation  : 1.798387 1.716012 1.770430 1.737070
Smoothness      : 0.223011 0.316130 0.245578 0.290927
Third_moment    : 3.417141 3.417047 3.408786 3.417047
Uniformity      : 0.300673 0.454473 0.317692 0.399113
```

```
Features Extraction of LR region
Entropy         : 1.202904 1.283128 1.206375 1.248128
Avg. Gray level : 2169.000000 3152.000000 1877.000000 2
Std. Deviation  : 1.980538 1.907130 1.978908 1.941439
Smoothness      : 0.212126 0.292782 0.206388 0.251700
Third_moment    : 3.538203 3.543448 3.537884 3.543448
Uniformity      : 0.278750 0.411948 0.259356 0.339766

Features Extraction of LL region
Entropy         : 0.939681 0.995633 0.952051 1.000824
Avg. Gray level : 2376.000000 3273.000000 2253.000000 3
Std. Deviation  : 1.998002 1.971580 2.003808 1.969348
Smoothness      : 0.174793 0.233868 0.183729 0.242383
Third_moment    : 2.842560 2.843778 2.840051 2.843778
Uniformity      : 0.257803 0.361081 0.257888 0.372650
```

5. Caption generation:

Finally the caption is generated based on the comparison of the features which specify the occurrence of the disease such as normal or abnormal.



VI . CONCLUSION

In this paper, we represent the techniques and algorithms which is used to generate the caption for medical images. This greatly helps the patients to understand the state of their health. This automatic generation of caption saves the time for doctors to analyze the state of the particular organ. The neural network concept and GLCM algorithm used are helps to generate the caption in the manner which reduces the complexity and increases the computational speed.

REFERENCES

[1] http://www.who.int/trade/distance_learning/gpgh/gp gh3/en/index4.html

[2] https://www.nationaljewish.org/conditions/tuberculo sis-tb/types

[3] RatnasariNurRohmah, AdhiSusanto, Indah Soesanti, Computer Aided Diagnosis for Lung Tuberculosis Identification Based on Thoracic X-ray, IEEE 2016

[4] Poomimadevi.CS, Helen Sulochana.C, Automatic detection of pulmonary tuberculosis using image processing techniques, IEEE WiSPNET 2016

[5] RatnasariNurRohmah, AdhiSusanto, Indah Soesanti, Lung Tuberculosis Identification Based on Statistical Feature of Thoracic X-ray, IEEE 2015

[6] https://en.wikipedia.org/wiki/Median_filter

[7] R.C.Gonzalez, Digital Image Processing. Pearson, 2016, pp.689-795.

[8] Rohini Paul Joseph, M Manikandan, Brain tumor MRI image segmentation and detection in image processing, International Journal of Research in Engineering and Technology, volume 3, special issue 01, March 2015.

[9] https://en.wikipedia.org/wiki/Canny-edge-detector

[10] http://www.businessdictionary.com/definition/classical-linear-regression-CLR.html

[11] https://stats.stackexchange.com/questions/130293/sv m-and-smo-main-differences

[12] https://ceb.nlm.nih.gov/repositories/tuberculosischest-x-ray-image-data-sets/