

Lung Cancer Detection and Classification Using deep Learning

A. Reshma Bhavani¹, K. Chakravarthy², M. V. K. Sai Praveen³, Mr. T. Ravi Kumar⁴

1,2,3 Student, Dept of CSE, NS Raju Institute of Technology Visakhapatnam, Andhra Pradesh, India

4 Assistant Professor, Dept. of CSE, NS Raju Institute of Technology Visakhapatnam, Andhra Pradesh, India

Reshmaalluri666@gmail.com, Chakravarthykilaparthi2000@gmail.com, praveen1242000@gmail.com and ravitangudu.cse@gmail.com

ABSTRACT

In recent years, so many Computer Aided Diagnosis (CAD) systems are designed for diagnosis of several diseases. Lung cancer detection at an early stage has become very crucial and also very convenient with image processing and deep learning techniques. The lung patient Computer Tomography (CT) scan images used to detect and classify the lung nodules and to detect the malignancy level. The CT scan images are segmented using U-Net architecture. This paper proposes 3D multipath VGG-like network, which is evaluated on 3D cubes, extracted from Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI), Lung Nodule Analysis 2016 (LUNA16) and Kaggle Data Science Bowl 2017 datasets. Prediction from U-Net and 3D multipath VGG-like network are combined for final results. The lung nodules are classified and malignancy level is detected using this architecture with 95.60% of Accuracy and 0.387732 of log loss.

The main aim of this thesis is to detect lung cancer in a person. Nowadays lung cancer disease has become a main problem for every person and it needs to be detected in a very early stage otherwise it may lead to death of a person. The input for detecting the lung cancer is CT scan. The reports will take a photo copy as a training dataset and that will produce a valid dataset results of a person for lung cancer. The result summary will show value either Zero or One. we can know that when the value reaches to 1 the person is safe that the person does not have any lung cancer or any lung disease and when the result reaches to 0

the we can consider that the person has lung cancer and the person has to take the required treatment for the cancer for saving his life.

According to the survey of the World Health Organisation (WHO), Lung Cancer was the most leading cause of death in 2015 and it is on fifth rank in 2017. It is most common in smokers accounting 85% of cases among all. So many Computer Aided Diagnosis (CAD) Systems were developed in recent years. Detection of lung cancer in the early stage is necessary to prevent death and to increase survival rate. Lung nodules are the small masses of tissues which can be cancerous or noncancerous also called as malignant or benign. Benign tissues are commonly non-cancerous and does not have much growth where malignant tissues grow very fast and can effect to the other body parts and are dangerous to health.

Keywords: Lung Cancer, Computer Aided Diagnosis, Computer Tomography and World Health Organization.

INTRODUCTION

Cancer is a disease in which abnormal cells of the body divide very fast, and generate excessive tissue that forms a tumor. Cancer cells are capable of spreading to other parts of the body through the blood and lymph systems.

Cite this article as: A. Reshma Bhavani, K. Chakravarthy, M. V. K. Sai Praveen, Mr. T. Ravi Kumar, " Lung Cancer Detection and Classification Using deep Learning", International Journal & Magazine of Engineering, Technology, Management and Research (IJMETMR), ISSN 2348-4845, Volume 8 Issue 7, July 2021, Page 1-8.

When the uncontrolled cell growth occurs in one or both lungs, it is said to be Lung Cancer. Besides, developing into a healthy, normal lung tissue, these abnormal cells continue dividing and form lumps or masses of tissue called tumors. The main function of the lung which is to carry the bloodstream with oxygen to the entire body is disturbed by these tumors. There are many types of cancers.

Cancers that begin in the lungs are divided into two major types, non-small cell lung cancer and small cell lung cancer, depending on how the cells look under a microscope. Each type of lung cancer grows and spreads in different ways and is treated differently. This is usually believed to be a systemic disease at the time of diagnosis and thus surgery plays no part in the management of this disease. Limited disease: It is limited to one hemi thorax that can be included in a reasonable field of thoracic radiation therapy. Extensive disease: It is beyond one hemi thorax or that cannot be included in a reasonable field of thoracic radiation therapy. Non-small cell lung cancer (NSCLC) aggravates heavily later in its course than SCLC and consequently surgery is the best chance of cure. Patients those who are considered for surgical treatment must be carefully staged to determine tumor respectability. Positron Emission Tomography (PET) will help to assess modal involvement. But, only 15% of patients are appropriate for resection at diagnosis. The patient must also be carefully assessed pre-operatively to assure fitness for surgery.

Lung cancer is one of the most dangerous cancers in the world, with the smallest survival rate after diagnosis, with a gradual increase in the mortality rate every year. Figure 1 shows an example of cancerous lung image. The survival probability from lung cancer is indirectly proportional to its growth at its detection time. The chances of successful treatment are possible only if the disease is

detected at the earlier stage. An estimated result shows that 85% of lung cancer cases in males and 75% in females are caused by cigarette smoking (Singapore Cancer Society). The overall survival rate for all types of cancer is 63%. Though surgery, radiation therapy, and chemotherapy have been used in the treatment of lung cancer, the five-year survival rate for all stages combined is only 14%. This has not changed in the past three decades (American Cancer Society, 2005).



Figure 3.1 An example of cancerous lung image

Several thin-sectional CT images are produced in clinic for each patient and are estimated by a radiologist by looking at each image in the axial mode. Most of the images will be very tough to interpret and consume a lot of time that can cause high false-negative rates for detecting small lung nodules, and thus potentially miss a cancer. The fundamental idea of designing a CAD system is to make a machine algorithm that acts as a support to the radiologist and points out locations of doubtful objects, so that the overall sensitive rate is raised. CAD system must achieve following needs, improving the quality and accuracy of diagnosis, increasing success in therapy by early detection of cancer, avoiding unnecessary biopsies, Reducing radiologist interpretation time (Wiemker et al 2005). A CAD system for early detection of lung cancer by analyzing the images of CT is proposed in this chapter. Fuzzy Possibilistic C-Means (FPCM) is used for clustering the cancerous and non-cancerous nodules.

2.Requirements Specification

Use this Requirements Specification template to document the requirements for your product or service, including priority and approval. Tailor the specification to suit your project, organizing the applicable sections in a way that works best, and use the checklist to record the decisions about what is applicable and what isn't. The format of the requirements depends on what works best for your project. This document contains instructions and examples which are for the benefit of the person writing the document and should be removed before the document is finalized. To regenerate the TOC, select all (CTL-A) and press F9. Executive Summary

Project Overview

Describe this project or product and its intended audience, or provide a link or reference to the project charter.

Purpose and Scope of this Specification

Describe the purpose of this specification and its intended audience. Include a description of what is within the scope what is outside of the scope of these specifications. For example:

In scope

This document addresses requirements related to phase 2 of Project A: modification of Classification Processing to meet legislative mandate ABC. modification of Labor Relations Processing to meet legislative mandate ABC.

Out of Scope

The following items in phase 3 of Project A are out of scope: modification of Classification Processing to meet legislative mandate XYZ. modification of Labor Relations Processing to meet legislative mandate XYZ.

Product/Service Description

In this section, describe the general factors that affect the product and its requirements. This

section should contain background information, not state specific requirements (provide the reasons why certain specific requirements are later specified).

User Characteristics

Create general customer profiles for each type of user who will be using the product. Profiles should include: Student/faculty/staff/other experience technical expertise other general characteristics that may influence the product

Assumptions

List any assumptions that affect the requirements, for example, equipment availability, user expertise, etc. For example, a specific operating system is assumed to be available; if the operating system is not available, the Requirements Specification would then have to change accordingly.

Hardware Requirements

The project is intended to mine frequent itemsets in a faster way. Hence it is implemented as a desktop application. It requires a minimum of 4 GB RAM space with NTFS file system partition.

Software Requirements

Development Platform

The project is developed on Windows 7 platform. Windows 7 is one of the releases of Microsoft Windows, a series of operating systems produced by Microsoft for use on personal computers, including home and business desktops, laptops; netbooks, tablet PCs, and media center PCs.

System Configuration:-

- Processor - Pentium –IV
- RAM - 4 GB (min)
- Hard Disk - 20GB
- Key Board - Standard Windows Keyboard
- Mouse - Two or Three Button Mouse
- Monitor - SVGA

Software Requirements:

Operating system :Windows 10.

Coding Language :Python.

Python Packages

- ❖ NumPy
- ❖ Pandas
- ❖ Scikit-Learn
- ❖ Matplotlib
- ❖ Keras
- ❖ TensorFlow

Input Design

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- What data should be given as input?
- How the data should be arranged or coded?
- The dialog to guide the operating personnel in providing input.
- Methods for preparing input validations and steps to follow when error occur.

Objectives

1. Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the

management for getting correct information from the computerized system.

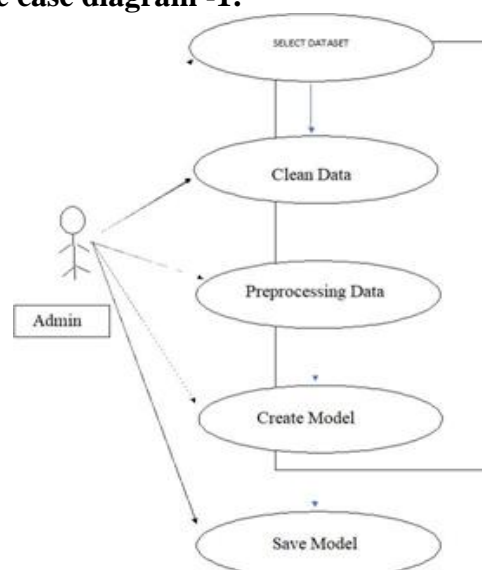
2. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.

3. When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow

Output Design

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output.

Use case diagram -1:



3. Sample Code

Algorithm:

Obtain the image of the target currency using one of the possible methods (e.g. : Camera, Scanner, etc)

Use Image Pre processing algorithms to change the nature of the image in order to extract required information.

Detect the boundaries and extract the ROI (Region of Interest) using cropping. Extract the desired features.

Compare the extracted feature values with ideal feature values that are calculated. Display the outputs.

```
Train.py import tflearn import h5py
import numpy as np
from cnn_model import CNNModel import
tensorflow.compat.v1 as tf
# Load HDF5 dataset
h5f = h5py.File('./data/train.h5', 'r')
X_train_images = h5f['X'] Y_train_labels =
h5f['Y'] h5f2 = h5py.File('./data/val.h5', 'r')
X_val_images = h5f2['X'] Y_val_labels =
h5f2['Y']
## Model definition convnet =
CNNModel() network =
convnet.define_network(X_train_images)
model = tflearn.DNN(network,
tensorboard_verbose=0,\
checkpoint_path='nodule3-classifier.tfl.ckpt')
model.fit(X_train_images, Y_train_labels,
n_epoch = 10, shuffle=True,\validation_set
= (X_val_images, Y_val_labels),
show_metric = True,\batch_size = 96,
snapshot_epoch = True, run_id = 'nodule3-
classifier')
model.save("nodule3-classifier.tfl")
print("Network trained and saved as nodule2-
classifier.tfl!") h5f.close()
h5f2.close()
Predict.py
import tflearn
```

```
from cnn_model import CNNModel import
tensorflow as tf
import pickle import pandas as pd import
numpy as np import h5py
from sklearn.metrics import roc_curve, auc,
confusion_matrix import itertools
import matplotlib.pyplot as plt hdf5_file =
'./data/test.h5'
def create_mosaic(image, nrows, ncols)M =
image.shape[1]
N = image.shape[2]
npad = ((0,0), (1,1), (1,1))
image = np.pad(image, pad_width = npad,
mode = 'constant',\constant_values = 0) M +=
2 N+=2
image = image.reshape(nrows, ncols, M, N)
image = np.transpose(image, (0,2,1,3)) image
= image.reshape(M*nrows, N*ncols) return
image
def format_image(image, num_images):
"""Formats images """
idxs = np.random.choice(image.shape[0],
num_images) M = image.shape[1] N =
image.shape[2]
imagex = np.squeeze(image[idxs, :, :, :]) print
(imagex.shape)
return imagex
def plot_confusion_matrix(cm, classes,
normalize=False, title='Confusion matrix',
cmap=plt.cm.Purples):
"""This function prints and plots the confusion
matrix. Normalization can be applied by
setting `normalize=True`."""
plt.imshow(cm, interpolation='nearest',
cmap=cmap) plt.title(title)
#plt.colorbar()
tick_marks = np.arange(len(classes))
plt.xticks(tick_marks, classes,
rotation=45)
plt.yticks(tick_marks, classes)
```

4. System Testing

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the

functionality of components, sub assemblies, assemblies and/or a finished product. It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

TYPES OF TESTS

Unit testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application. It is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

Integration testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfactory, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

Functional test

Functional tests provide systematic demonstrations that functions tested are

available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

Valid Input : identified classes of valid input must be accepted.

Invalid Input : identified classes of invalid input must be rejected.

Functions : identified functions must be exercised.

Output : identified classes of application outputs must be exercised.

Systems/Procedures: interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

System Test

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

White Box Testing

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.

Test strategy and approach

Field testing will be performed manually and functional tests will be written in detail.

Test objectives

- All field entries must work properly.
- Pages must be activated from the identified link.
- The entry screen, messages and responses must not be delayed.

Features to be tested

- Verify that the entries are of the correct format
- No duplicate entries should be allowed
- All links should take the user to the correct page.

Integration Testing

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.

The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level – interact without error.

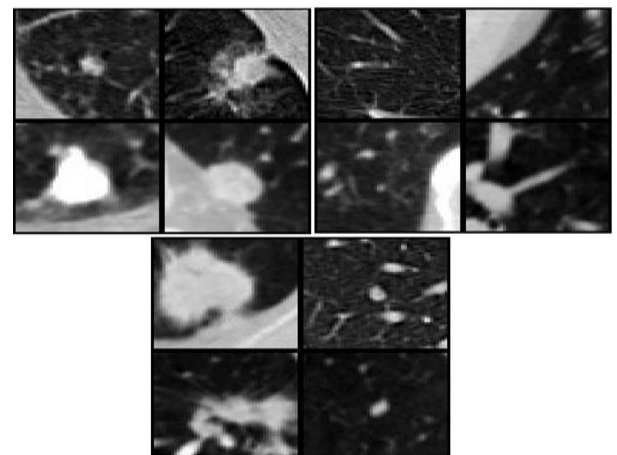
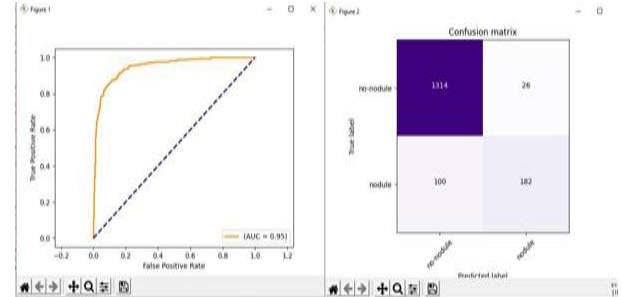
Test Results: All the test cases mentioned above passed successfully. No defects encountered.

Acceptance Testing

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

Test Result: All the test cases mentioned above passed successfully. No defects encountered.

5. Result and Discussion



6. Conclusion:

The convolution neural network method for detecting lung nodules in computer tomography was proposed. Lung nodule candidates were extracted using detection method. The candidates were ranked based on segmentation and AlexNet was used for FPs reduction. The proposed method shows promising performance in lung nodule detection. More than 80% and 95% of lung nodules that are in lung field in the TCIA database can be detected when the FPs per image were two and five, respectively.

References:

- [1] Li, Linlin Shen, and Suhuai Luo, "Solitary Feature-Based Lung Nodule Detection Approach for Chest X-Ray Radiographs," IEEE journal of biomedical and health informatics, vol. 22, no. 2, march 2018.
- [2] Hongyang Jiang, He Ma*, Wei Qian, Mengdi Gao and Yan Li An, Automatic

Detection System of Lung Nodule Based on Multi-Group Patch-Based Deep Learning Network, *IEEE Journal of Biomedical and Health Informatics*, vol.22, July 2018.

[3] Majd Zreik, Pim A. de Jong, Bob D. de Vos, Automatic Calcium Scoring in Low-Dose Chest CT Using Deep Neural Networks with Dilated Convolutions Nikolas Lessmann, Bram van Ginneken, Member, *IEEE transactions on medical imaging*, vol. 37, no. 2, February 2018.

[4] Prajwal Rao, Nishal Ancelette Pereira, and Raghuram Srinivasan Convolutional neural networks for lung cancer screening in computed tomography (CT) scans, vol 21. march 2017.

[5] X. Li, S. Luo, Q. Hu, J. Li, D. Wang, and F. Chiong, "Automatic lung field segmentation in X-ray radiographs using statistical shape and appearance models," *J. Med. Image. Health Informat.*, vol. 6, pp. 338–348, 2016

[6] Q. Dou, H. Chen, L. Yu, J. Qin, and P. A. Heng, "Multi- level contextual 3D CNNs for false positive reduction in pulmonary nodule detection," *IEEE Trans. Biomed. Eng.*, 2016.

[7] H. C. Shin et al., "Deep convolutional neural networks for computer aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.

[8] X. Li, S. Luo, Q. Hu, J. Li, and D. Wang, "Rib suppression in chest radiographs for lung nodule enhancement," in *Proc. 2015 IEEE Int. Conf. Inf. Autom.*, Lijing, China, 2015, pp. 50–55.

[9] E. J. van Beek, S. Mirsadraee, and J. T. Murchison, "Lung cancer screening: Computed tomography or chest radiographs?" *World J. Radiol.*, vol. 7, pp. 189–193, 2015.

[10] Z. Zhu, S. Jia, S. He, Y. Sun, Z. Ji, and L. Shen, "Three- dimensional Gabor feature extraction for hyperspectral imagery classification using a memetic framework," *Inf. Sci.*, vol. 298, pp. 274–287, 2015.

[11] H. Han, L. Li, F. Han, B. Song, W. Moore, and Z. Liang, "Fast and adaptive detection of pulmonary nodules in thoracic CT images using a hierarchical vector quantization scheme," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 2, pp. 648–659, 2015.

[12] J. Kuruvilla and K. Gunavathi, "Lung cancer classification using neural networks for CT images", *Computer Methods and Programs in Biomedicine*, vol. 113, no. 1, pp. 202-209, 2014.

[13] S. Chen and K. Suzuki, "Computerized detection of lung nodules by means of 'virtual dual-energy' radiography," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 2, pp. 369–378, Feb. 2013.

[14] M. Tan, R. Deklerck, B. Jansen, M. Bister, and J. Cornelis, "A novel computer-aided lung nodule detection system for CT images," *Med. Phys.*, vol. 38, no. 10, pp. 5630–5645, 2011.