

# Lung CT Image Classification using Deep Neural Networks for Lung Cancer Detection

Prathyusha Chalasani, S. Rajesh



**Abstract:** In Recent years, image processing strategies are broadly utilized in a few restorative territories for image improvement in prior division and treatment stages, where the time factor is imperative to find the variation from the norm issues in target pictures, particularly in different malignant growth tumors, for example, lung disease. Lung cancer is the most important disease cause high mortality rate. And computer-aided diagnosis can be useful for physicians to accurately identify the cancer cells. Many computer-aided methods have been studied and applied using image processing and machine learning. But, they are not acceptable for a health-based classification models to have high false positive and true negative rates as it they can devastate lives through false diagnosis. To reduce the effect of them in classification, to perform experiments JSRT data set is considered as it is the most widely used benchmark data set. The proper segmentation of lung tumor from X-ray, CT-scan or MRI these are the stepping stone towards automated diagnosis system for lung cancer detection. Our detection is to train this neural network using volumes with tumor size and position. In recent techniques like machine learning can predict lung cancer but this technique is not suitable for predicting segmentation of images in that particular area.

**Keywords:** Classification, Data Augmentation, Deep learning, Lung Cancer, Prediction.

## I. INTRODUCTION

Lung cancer is the leading cause of death from cancer, and the world's second most common cancer in both men and women [1]. Because lung cancer is most frequently diagnosed at an advanced stage, the overall survival for 5 years is poor. Early recognition is vital to early intercession prompting improved endurance. Contrasted and radiographs, low-portion CT can give progressively definite data and has been accounted for to prompt a 22% lower death rate [2]. Studies of the screening of lung cancer based on human expert analysis have identified false positive rates as high as 54 % [3]. This not just builds the expense for additional tests and surgeries, yet in addition purposes pointless nervousness for patients and their families. The advancement of incredible PC supported methodologies for lung cancer early screening is basic to improve the present 3 clinical act of CT imaging evaluation. Computer-assisted methods are expected to produce automated solutions for

early screening of lung cancer and a decreased false positive diagnosis rate. A number of studies on image processing methods are available in the literature to detect early stage cancer detection. But the early-stage cancer detection impact ratio isn't greatly improved. With the advent of machine learning methods, several researchers are seeking to detect cancer early. Neural network plays a key role in the identification of cancer cells in normal tissues, which in turn provides an effective tool for assisting in the detection of deep learning cancer.

## II. RELATED WORK

Brahim et.al., [1] proposed a process known as Lung CT Image Segmentation using Methodology of Deep Neural Networks. A segmentation of a lung CT image using U-net architecture in deep learning for segmentation of images. The dataset used in this experiment is the Lung Image Database Registry image database, which consists of computed tomography (CT) scans with marked-up annotated lesions for diagnostic and lung cancer screening.

A system called Multi-Stage Lung Cancer Detection and Prediction Using Multi-class SVM Classifier was developed by Alam J et.al., [2]. Using multi-class SVM classifier, they provide an effective the identification and prediction of lung cancer algorithms. Image enhancement and segmentation were performed separately at each classification process. Deep learning With Lung Segmentation and Bone Shadow Exclusion Techniques for Lung Cancer's Chest X-Ray Analysis. proposed by Gordienko Y et.al., [3]. In this, they used a lung segmentation process and bone shadow exclusion techniques to examine photos using a deep learning approach. And, pre-processed bone-free datasets are much more reliable and result in loss relative to other pre-processed datasets.

Hesamian MH et.al., [4] introduced a method titled Deep Learning Techniques for Segmentation of Medical Image. A very popular methods used in this are CNN, FCN, U-Net that used deep learning techniques for segmentation of medical images. Nevertheless, the approach proposed helps to more efficiently compare the nodule containing areas to which hard conclusions derived from statistical information on the size and shape of the nodule have been further developed. Lobo P et.al., [5] proposed a method known as classification and segmentation procedures from CT Images for Lung Cancer Detection. In this, they used a method which is called Support Vector Machine (SVM) to identify the image and Fuzzy c means (FCM) to segment image.

Revised Manuscript Received on February 05, 2020.

\* Correspondence Author

Prathyusha Chalasani, CSE, V R Siddhartha Engineering College, Vijayawada, India. Email: prathyusha020@gmail.com

S. Rajesh, CSE, V R Siddhartha Engineering College, Vijayawada, India. Email: [rssingam99@gmail.com](mailto:rssingam99@gmail.com)

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Vector support software was used to classify the image as normal or as cancer.

Yanfeng Lil et.al., [6] build a 3D Thoracic MR Images method named Deep Learning Lung Nodule Detection. In this methodology, a lung nodule detection method for thoracic MR images is proposed based on profound learning. The Faster R-CNN is designed with optimized parameters for the identification of lung nodules, spatial three-channel feedback and transfer learning. techniques for the diagnosis and segmentation of lung cancer from CT images developed by Uday Kamal et.al., [7]. In this the methodology consists of the stage of preprocessing, the stage of segmentation, the stage of extraction of features and the stage of classification. The images are classified in classification according to the features which are given to them. We try to present the picture as something real for easier analysis.

Senthil Kumar K et.al., [8] suggested a Lung cancer detection using the different evolutionary algorithms to segment images. They also suggested an algorithm for detecting and predicting lung cancer and providing extremely encouraging results compared to other techniques used. The SVM classifies a collection of textural features that are derived from the isolated ROIs.

Lakshmana Prabu S.K et.al., [9] developed a method known as the Optimal Deep Learning Model for Lung Cancer Classification on CT Images. In this, the proposed ODN with feature reduction showed improved classification compared with other classification strategies in the case of lung CT Images. An automatic approach to classifying Lung cancer reduces hand marking time and avoids human error. The researchers planned to achieve greater accuracy and accuracy in recognizing a normal and abnormal image of the lung through machine learning.

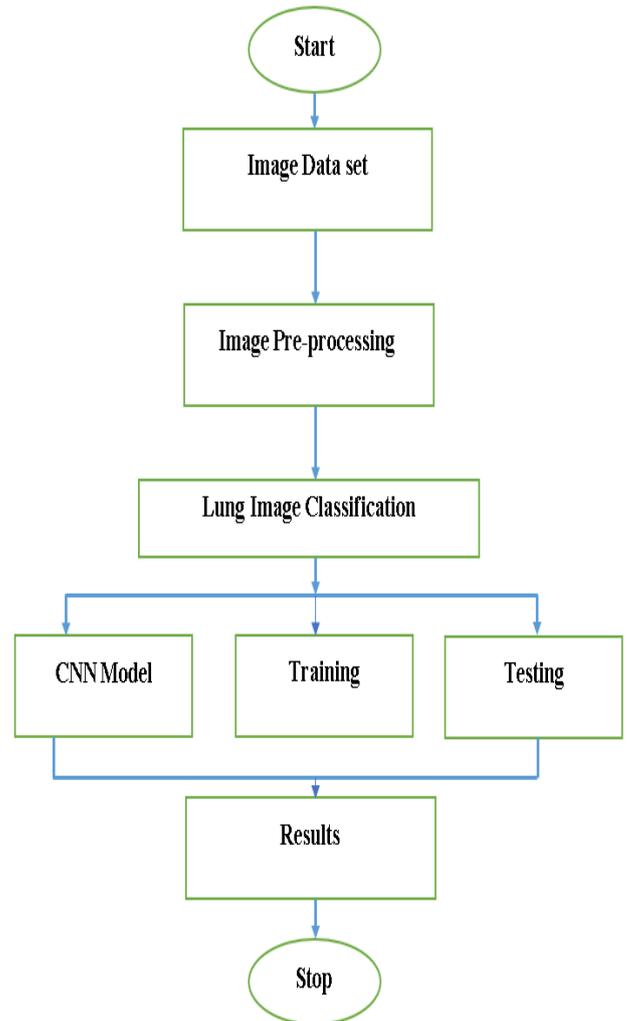
Sneha Potghan et.al., [10] Classification of Lung tumors based on multi-layer perceptron. In this, they proposed a segmentation of the volume of the lung is carried out using the clustering algorithm k-means. The five statistical characteristics of the volume of the lungs are extracted to describe the CT images of the non-tumor as well as tumorous lungs. The characteristics were extracted using an artificial neural network algorithm called MLP that is capable of effectively distinguishing non-tumor and tumor characteristics.

Moffy Vas et.al., [11] suggested a method called a lung cancer detection device using the analysis of lung CT scans. The approach adopted in this paper is aimed at creating an automatic lung cancer detection system. Median Filter application to remove impulse noise in the images has proven a success. The morphological operations also helped to achieve satisfactory results in the segmentation process.

Emine CENG et.al., [12] presents Lung Cancer Detection A Deep Learning Based Approach. In this, they roped a draws attention to early lung cancer diagnosis. The classification of lung nodules is both benign and malignant. Deep learning architecture CNN is particularly famous for its performance in the classification of images. It also obtains successful results for biomedical image classification process.

## III. METHODOLOGY IMPLEMENTATION

The method of classifying lung CT images using deep neural networks for detecting lung cancer as shown in below fig 1.



**Fig 1. Overview of Proposed Methodology**

The suggested technique consists of various modules used by deep neural networks to identify the lung CT images. They are,

- Data set Collection
- Image Pre-processing
- Classification
- Training
- Testing

### A. Data set Collection

In this methodology, there are 247 images of 154 cases with lung nodules and 93 cases without lung nodules in our data set called as JSRT image data set. BSE-JSRT data set contains 247 images of the JSRT data set, but the special algorithms remove no clavicle and rib shadows. For this data set, we applied classification techniques, respectively. The database contains the following information about each patient in a text file: nodule size (mm), age, gender, final

diagnosis, anatomical nodule position, x-and y-coordinates of the 20 nodule centers in a digital image. And, the sample pictures of the given data set as shown in below fig 2.

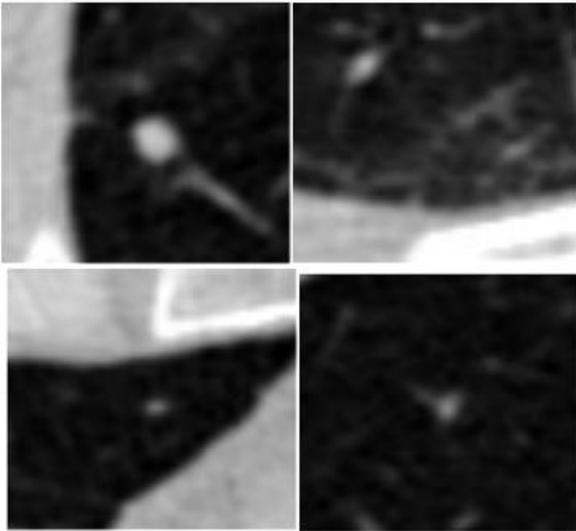


Fig 2. Sample input image data set

We also use data from the Lung Image Database Consortium and the Institute for Research on Infectious Diseases [ (LIDC / IDRI) database]. This dataset was composed of 888 CT scans with annotations defining positions and labels of ground truth. First phase was the development of an image training database.

**B. Image Pre-processing**

This process is aimed at improving image data, suppressing undesirable distortions or enhancing some important features to further process. Pre-processing stage of the image ends with smoothing of the image.

**a. Image Enhancement**

Image enhancement is the method of modifying digital images to making the effects more suitable for displaying or further analyzing images. You can remove noise, sharpen or brighten an image for example, making it easier to identify key features. It suppresses noise or other minor variations in the image; similar to suppressions of high frequency in the frequency domain. Median filtering is used to eliminate the noise from the images. It reduces blurring of the edges. The benefits of median filtration are that it is not affected by individual noise bursts, decreases impulsive noise quite well and does not blur the edges much and can be applied iteratively Image enhancement is intended to improve the capacity for human viewers to interpret or perceive information in images or to provide better feedback for other automated image processing techniques.

**C. Classification**

Machine learning algorithms such as vector supporting machines are often used in this step to identify and recognize tumors. But when we define features, they're often limited by the assumptions we make. It results in decreased sensitivity. Deep learning, however, could be an ideal solution, as these algorithms can learn features from the raw image data. One difficulty in applying such algorithms is the lack of medical image data which is labelled. While this is a drawback for all deep learning applications, it is more so for medical image

data due to confidentiality issues with patients. In this, a neural convolution network will be developed, equipped, and has lung nodules detected. We use the data from the Lung Image Database Consortium and Infectious Disease Research Initial move was to construct a training image database.

**a. Creating an image Database**

In this, they formatted the images as data of. mhd and. raw. The header data is in the .mhd files and the image data is stored in the. Raw files. We used the SimpleITK library for file-reading. Mhd. Mhd. Every CT scan has dimensions of 512x 512 x n, where n is the number of the axial scans. That CT scan has about 200 images. We could train the CNN on all the pixels potentially, but that would increase the time and the computational cost. So, we decided instead to crop the images around the coordinates given in the annotations.

**b. Building a CNN Model**

We are ready to build a CNN in this. Since dabbling a little with TensorFlow, we settled for something incredibly simple that was way too much work. We opted to use tflearn. Tflearn is a wrapper of high-level APIs around TensorFlow. It has made coding much more palatable. And, in this architecture we used 3 convolutionary layers as shown in fig. 3 below.

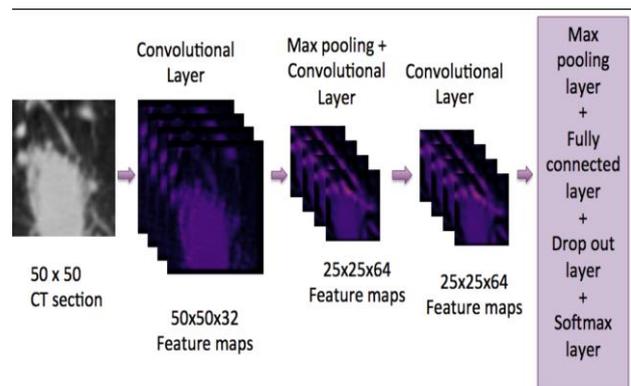


Fig 3. CNN Architecture

**c. Training the Model**

In this, the data required to train a CNN is very high. Training the model for batches is often beneficial. It is not always possible to load all training data into memory, because you need enough memory to manage it and the apps as well. We worked out of a MacBook Pro from 2012.

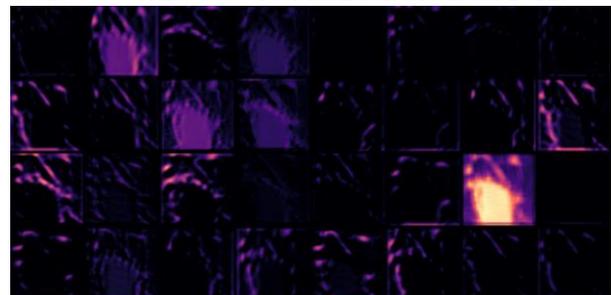
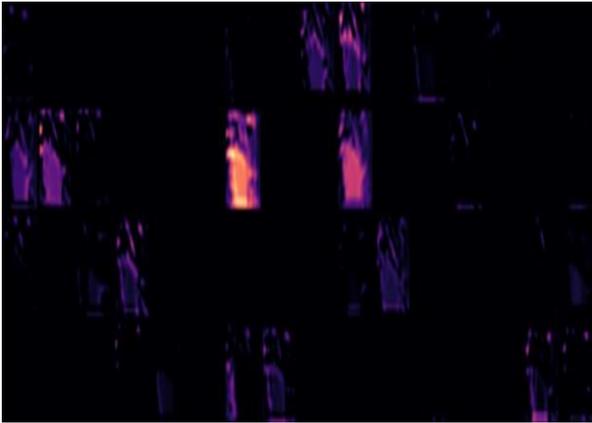


Fig 4. First convolution layer

So, we chose to use h5py library to load all the images into a hdfs dataset.

As the filters are of low resolution (5x5), visualizing the created features maps would be more useful. So, if we go through this image through the first convolutionary layer (50 x 50 x 32), a feature map is created that looks like this as below shows fig 4.

The max pooling layer is after the first layer down sampled the feature map by 2. So, the resulting feature map is: when the down sampled feature map is passed into the second convolutionary layer of 64 5x5 filters:



**Fig 5. Max Pooling layer**

The feature map generated by the third convolutional layer containing 64 3x3 filters is:



**Fig 6. Feature Map generation**

#### d. Testing the data

In this, we tested our CNN model on 1623 images. Then, we get a validation accuracy of 86 %. Our model has a precision of 82.3 % and recall of 71.2 %. The model has a specificity of 88.2 %.

#### e. Experimental Setup

In this implementation, we use different tools and libraries respectively. They are,

- Google Colab -

It is a free cloud service, and now supports free GPU to improve the coding skills of your Python language programming. Google Colab is an online virtual cloud machine, and we can use popular libraries such as Keras, TensorFlow to develop deep learning applications

- Keras -

Keras is a high-level, Python-written neural network API that can run on top of TensorFlow.

- TensorFlow

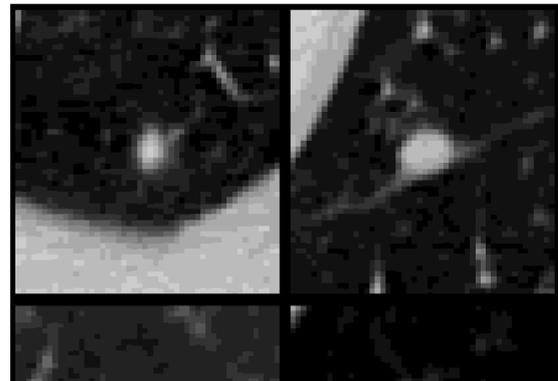
TensorFlow is a Python library, developed and published by Google for fast numerical computing. It is a foundation library that can be used either directly to create Deep Learning models or by using wrapper libraries that simplify the process built on top of TensorFlow.

- PyTorch

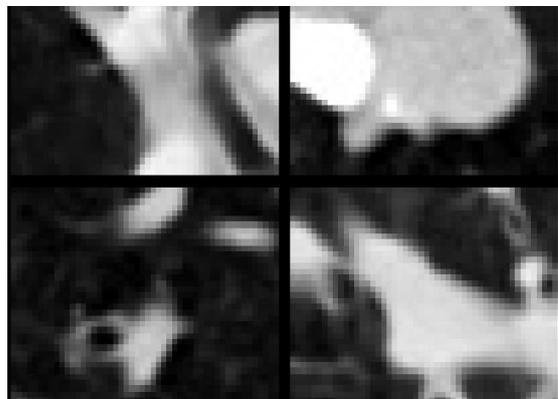
PyTorch is a Torch software-based, open source machine learning platform, used for applications such as computer vision and natural language processing.

## IV. RESULTS

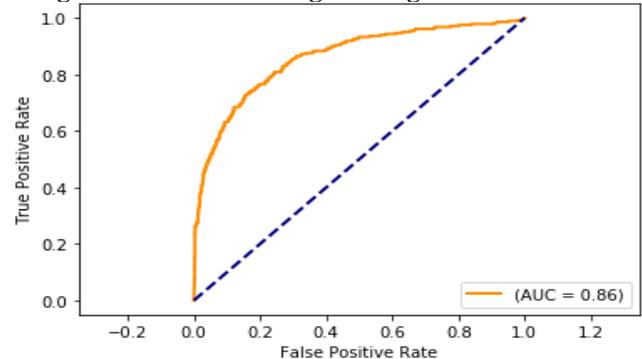
In this, when we do all those experiments using CNN. We get CNN, that was their finest algorithm or design. We present all of the algorithms with the results and performance analysis needed for accuracy. We employed three levels of convolution in our architecture.



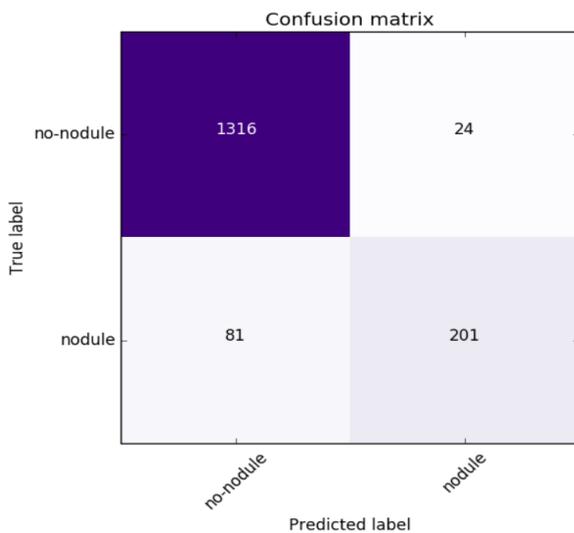
**Fig 7. Classification image of lung cancer detection 1**



**Fig 8. Classification image of lung cancer detection 2**



**Fig 8. Analysis of True and Positive Predictions**



**Fig 9. Confusion Matrix**

## V. CONCLUSION

The experimental results indicate that in many pictures, the techniques used in this method were sufficiently effective to segment the tumor. They consider the intensity variance of the image separately for each image. It was found that, by achieving high accuracy, the developed system was able to segment tumors of different tumor and non-tumor images from the used database. we get a validation accuracy of 86 %. Our model has a precision of 82.3 % and recall of 71.2 %. The model has a specificity of 88.2 %. The framework proposed is checked on JSRT data set and the findings have been found to be correct. In the future, a greater number of images will be tested on our proposed system and accuracy will be measured. They should equate our work with existing works.

## REFERENCES

1. Brahim AIT SKOURT, Abdelhamid EL HASSANI, Aicha MAJDA "Lung CT Image Segmentation Using Deep Neural Networks "Volume 127, 2018, Pages 109-113.
2. Alam J, Alam S, Hossan A. Multi-stage lung cancer detection and prediction using multi-class svm classifie. In2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2) 2018 Feb 8 (pp. 1-4). IEEE.
3. Gordienko Y, Gang P, Hui J, Zeng W, Kochura Y, Alienin O, Rokovyi O, Stirenko S. Deep learning with lung segmentation and bone shadow exclusion techniques for chest X-ray analysis of lung cancer. InInternational Conference on Computer Science, Engineering and Education Applications 2018 Jan 18 (pp. 638-647). Springer, Cham.
4. Hesamian MH, Jia W, He X, Kennedy P. Deep learning techniques for medical image segmentation: Achievements and challenges. Journal of digital imaging. 2019 Aug 15;32(4):582-96.
5. Lobo P, Guruprasad S. Classification and segmentation techniques for detection of lung cancer from ct images. In2018 International Conference on Inventive Research in Computing Applications (ICIRCA) 2018 Jul 11 (pp. 1014-1019). IEEE.
6. Li Y, Zhang L, Chen H, Yang N. Lung nodule detection with deep learning in 3D thoracic MR images. IEEE Access. 2019 Mar 18; 7:37822-32.
7. Kamal U, Rafi AM, Hoque R, Hasan M. Lung cancer tumor region segmentation using recurrent 3D-DenseUNet. arXiv preprint arXiv:1812.01951. 2018 Dec 5.
8. Senthil Kumar K, Venkata Lakshmi K, Karthikeyan K. Lung cancer detection using image segmentation by means of various evolutionary algorithms. Computational and mathematical methods in medicine. 2019;2019.

9. Lakshmana Prabu SK, Mohanty SN, Shankar K, Arunkumar N, Ramirez G. Optimal deep learning model for classification of lung cancer on CT images. Future Generation Computer Systems. 2019 Mar 1; 92:374-82.
10. Potghan S, Rajamenakshi R, Bhise A. Multi-Layer Perceptron Based Lung Tumor Classification. In2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA) 2018 Mar 29 (pp. 499-502). IEEE.
11. Vas M, Dessai A. Lung cancer detection system using lung CT image processing. In2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA) 2017 Aug 17 (pp. 1-5). IEEE. .
12. CENGIL E, ÇINAR A. A Deep Learning Based Approach to Lung Cancer Identification. In2018 International Conference on Artificial Intelligence and Data Processing (IDAP) 2018 Sep 28 (pp. 1-5). IEEE.

## AUTHORS PROFILE



**Ms. Prathyusha Chalasani**, Pursuing M.Tech, Department of Computer Science and Engineering, V R Siddhartha Engineering College, Vijayawada.



**Mr. S. Rajesh** B.Tech, M.Tech. is currently working as an Assistant Professor in the Department of Computer Science and Engineering, Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada. He is pursuing Ph.D, and he has twelve years of teaching experience, and he published 10 international journal and conference papers.