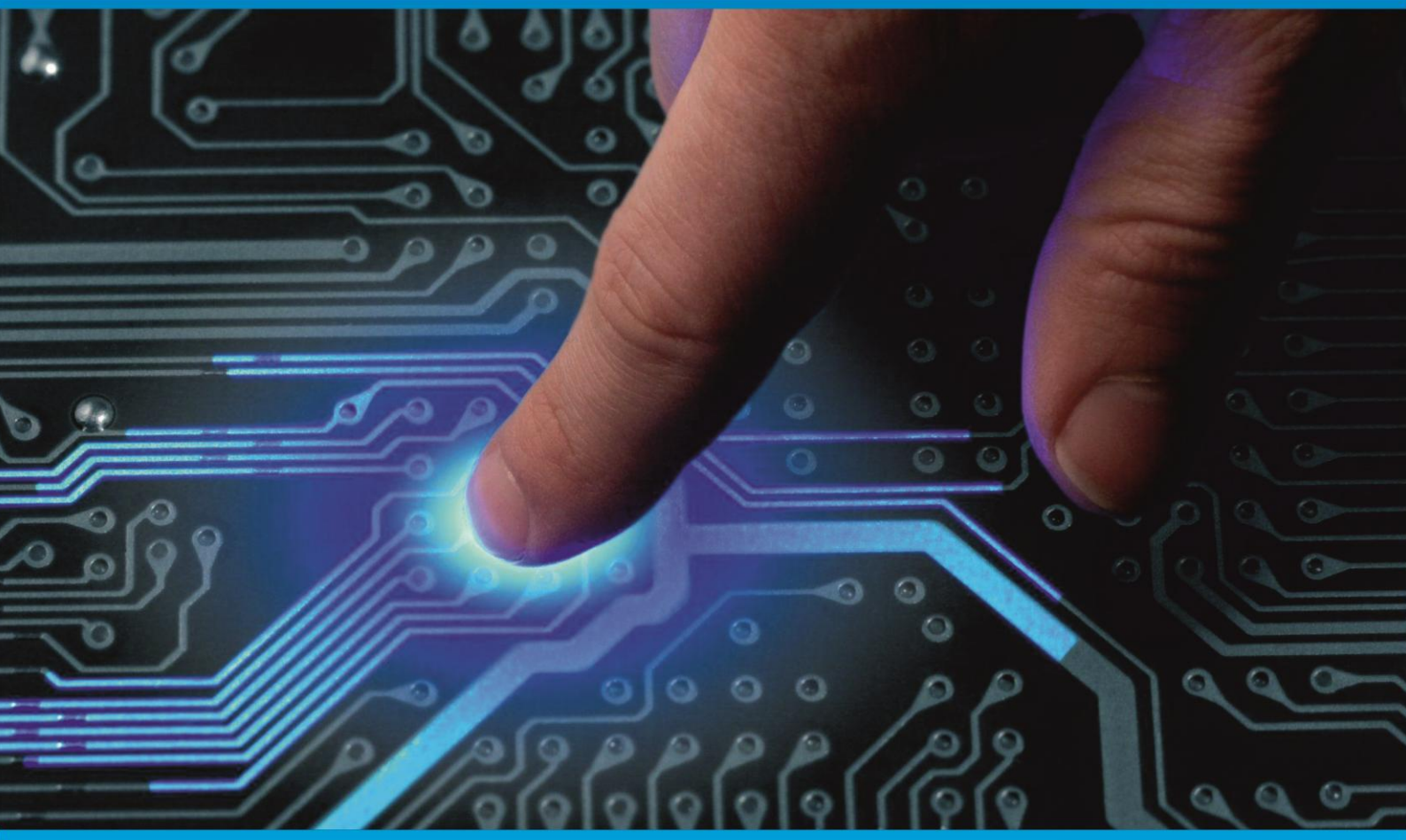
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 3, March 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903037 |

Lung Nodule Detection in X-Ray Image Using CNN

B.Sathiyapriya, K.Susmitha.,Msc.,M.Phil

Student, Dept. of Computer Science, Sakthi College of Arts and Science for Women, Oddanchatram, TamilNadu, India

Assistant Professor, Dept. of Computer Science, Sakthi College of Arts and Science for Women, Oddanchatram, TamilNadu, India

ABSTRACT: I *present* a novel procedure to apply deep learning techniques to medical image classification with increasing popularity of computed tomography (CT) lung screening, fully manual diagnosis of lung cancer puts a burden on the radiologists who need to spend hours reading through CT scanned images to identify Region of Interests (ROIs) to schedule follow-ups. Accurate computer-aided diagnosis of lung cancer can effectively reduce their workload and help training new radiologists. However, lung cancer detection is challenging because of the varying size, location, shape, and density of nodules. The current paper presents a methodology for exact diagnosing using statistical features from x-ray imaging. The potential future direction suggested could further improve the efficiency and increase the number of deep learning aided lung disease detection applications. In this project, I aim to experiment and implement various deep learning architectures. I would like to classify affected and non-affected region of interests with high sensitivity and low false positive rate using various types of convolution neural networks.

1. INTRODUCTION

Lung cancer is a leading cause of mortality than any other type cancer in the world. Cancer in itself has some scary implications fall victimized to it, and 1.6 million pass away as a result of it. In India, the number of new cases improved from approximate 65,000 in 2009 to 90,000 in 2013, registering 15-20% increase annually. It is troublesome in India where people hesitate to consult doctors at the earliest or do not have access to them. Medical professionals look time as one of the important factor to discover the cancer in the patient at the earlier stage; which is very important for successful treatment. Lungs are a unit set on the lateral sides of the body cavity and

separated from one another by the bodily cavity because the left respiratory organ is physically smaller than the correction respiratory organ this can be as a result of the correction and left lungs exhibit some obvious structural variations as the center come keen on the left aspect of the body part cavity, the accurate respiratory organ is partitioned as the superior, middle, and inferior lobes by 2 fissures. The left respiratory organ combined with a medial surface indentation, submitted to as the internal organ impression that's approached by the center. Lung illness refers to disorders that have an effect on the lungs.

The respiratory area caused by respiratory organ illness that could forestall the body from obtaining enough gas. Lung cancer may be an extremely aggressive and often fatal malignancy that originates within the epithelial tissue of the respiratory system. Smoking originates all respiratory organ cancers. Metastasis illustrate unfold of cancerous cells to alternative tissues, happens early within the course of the illness, creating a surgical cure unlikely for many patients. X-ray imaging is that the quickest, most typical, and least costly diagnostic. Production of digital X-rays from pictorial radiographs is turning into a typical follow to maximize info and cut back the amount of rejected X-rays. X-rays are a unit among the oldest sources of magnetic radiation used for imaging. The most effective use of X-rays in medical line is diagnosing the cancerous illness very beginning state.

II. LITERATURE SURVEY

LUNG NODULE DETECTION IN XRAY IMAGES USING CNN

We propose a new automated deep learning techniques to medical image classification was proposed. With increasing popularity of Chest X rays, fully manual diagnosis of lung nodules puts a burden on the radiologists who need to spend hours reading through Lund nodule images to identify Region of Interests (ROIs) to schedule follow-ups. Accurate computer-aided diagnosis of lung cancer can effectively reduce their workload and help training new radiologists. However, lung nodule detection is challenging because of the varying size, location, shape, and density of

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903037 |

nodules. Many studies have approached this problem using image-processing techniques with the intention of developing an optimal set of features. Convolutional neural network has demonstrated to learn discriminative visual features automatically and has beat many state-of-art algorithms in image-processing tasks, such as pattern recognition, object detection, segmentation, etc. In this report, we evaluate the feasibility of implementing deep learning algorithms for lung cancer diagnosis with the Lung Image Database Consortium (LIDC) database.. The performance of our best model is comparable with the state-of-art results in the lung nodule detection task. Furthermore, the proposed method can be used to nodule location detection. For nodule localization, the assessed Retina Net architecture achieved 43 true positives, 26 false-positives and 22 false-negatives. In comparison, performance of the two readers was 42 ± 2 true-positives, 28 ± 0 false-positives and 23 ± 2 false-negatives. Detailed results are shown in Table 1. If not otherwise stated, all results in this paper are given in the form mean ± standard deviation. Te nodule detection performance of Retina Net can be inspected visually. Lung segmental was used to exclude extra thoracic detections. For lung segmentation a Dice score of 0.67 was achieved.

III. SYSTEM ANALYSIS

3.1 EXISTING SYSTEM

With the advancement of Technology and Computer Aided Diagnosis (CAD), scientists have encouraged a lot of automated systems to address the issue of reducing false positive while estimating the presence of pulmonary nodules in the CT scans of the patients. Thus, today we have a surplus of data pertaining to the CT scan patients. From here, we have the opportunity to use current topics in image processing, data mining, and machine learning to identify hidden patterns in nodule size, location, structure, etc. and construct a model to increase the probability of malignant tumor detection. With the advent of the pattern recognition and machine learning, data scientists have proposed many approaches which were robust in finding the hidden patterns and reducing the false positives. Consequently, the deep learning has come into picture for the complex image classification in order to ensure that outliers and anomalies are properly handled in the model and thus, reducing the false positive rate for malignant pulmonary nodule detection.

3.2 PROPOSED SYSTEM

RELATED RESEARCH TO SOLVE THE PROBLEM:

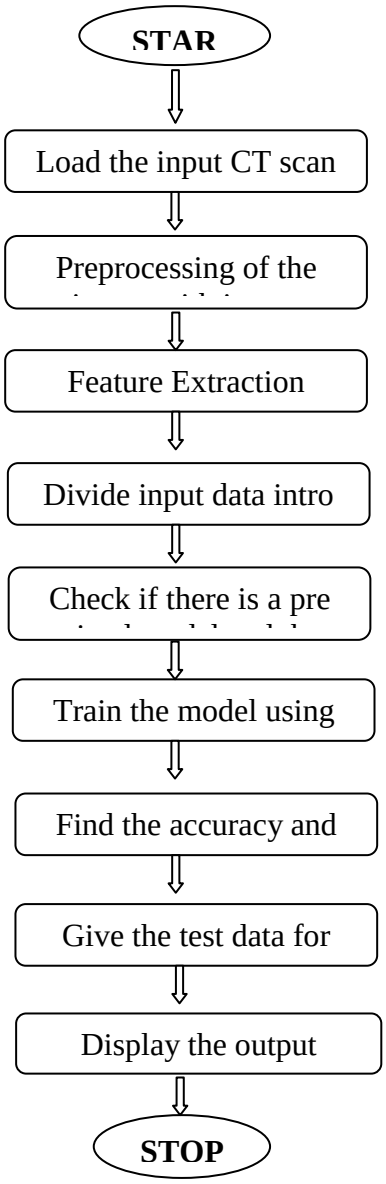
There has been a lot of research in recent times on the development of computer-aided diagnosis (CAD) systems for pulmonary nodule detection using CT imaging. Advancements in image processing field have increased the accuracy in the prediction of cancer from CT scans. There are plenty of research papers which discuss the various methods and outputs. A few examples include: ‘Recurrent Convolutional Networks for Pulmonary Nodule Detection in CT Imaging’, ‘Combining deep neural network and traditional image features to improve survival prediction accuracy for lung cancer patients from diagnostic CT’, ‘Computerized Detection of Lung Tumors in PET/CT Images’, and ‘Lung cancer classification uses neural networks for CT images’.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903037 |

3.3 ARCHITECTURE



IV. SYSTEM IMPLEMENTATION

4.1 PREPROCESSING

1**. Data Input**: Download data for around 1,500 patients from Kaggle – gives 2D slices of CT scan images for each patient, which are in the DICOM format. There are approximately 300 slices per patient.

2**. Image Processing**: Reconstruct the 3D structure of lungs, remove the noise from the image, and make the data set of all the patients' uniform and thus, eligible for feature extraction, which includes sorting, morphological, dilation, erosion, segmentation, masking etc.

3. **Classification**: Recurrent neural Network with MLP as its core is used for the classification. MLP, or multi-level perceptron, has a memory element, LSTM (long short term memory) associated with it. This further enhances the classification by emphasizing significant features while deemphasizing asspects which are unimportant.

4**. Validation**: 75% of training data is given to train the model and 25% is used for testing purposes.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903037 |

4.2FEATURE EXTRACTION&ENGINEERING

Using the 3D scan of the lungs allows us to retain as much of the information as possible. As explained, I then convert the CT scans (in Hounsfield Units) to pixel, allowing us to input a pixel array into our classifier as well as the features that I extract. Here, I will have utilized a two tiered approach. The first approach will focus on a simplistic method of feature extraction, then followed by our second approach which will place a heavy emphasis on deep learning. By doing so, I allow the simplistic model to set a baseline for accuracy, and we are then able to fully understand the impact that deep learning has on our precision. Analysis of tissue may be a method that is extremely self-made within the identification of medical pictures of abnormal respiratory organ, tissue and for this reason; the author describe concerning the x-rays and digital image process, which is that the means of recent identification defect as well as the illness.

The steps to discover the cancer are:-

(a) Cancelling the rips that seem white on the respiratory organs this could create mastics in detection the cancer by subtracting the background of the lung image from the cancer.

(b) Dividing the respiratory organ that carries the cancer into 2 elements (normal and abnormal). Examining these with one another by histogram and applied math notations.

(c) After specifying the partially hold the cancer in the image is processed by the subsequent steps:-

First, feature extraction is largely dependent on image identification. Thus any customized CNN model first have to fed with lot of training data unlike human beings for it to understand which category the image belongs to. In the process it does so by updating its weights i.e. applying appropriate filters, and thus emphasizing on the feature to be considered for recognition. This process is time consuming as with each input, it updates the weights and learns to identify the feature. The process outlined below was created by Visual Geometry Group (VGG) who have trained a CNN model with at least 22,000 categories to have the best possible weights for feature extraction. For our case, we can use the model directly for feature extraction without having to manually update the weights. With 16 layers, VGG16 is an extremely efficient image processing algorithm. Using a pre-trained model allows us to utilize the most up to date technologies in machine learning and data mining. By running this CNN on the data, we are able to auto detect features in each DICOM image. Thus, our output from the CNN is a data frame with approximately 40,000 processed features per patient.

4.3 CLASSIFICATION

Finally, I can input our data into a classification model. The data frame will have each individual patient as a row, and the different features as a column. The final column in the data frame will be a binary value: 1 if the patient has cancer, and 0 if the patient does not have cancer. For validation purposes, I will use 75% percent of our original data for training the classifier, and 25% for validation. From previous discussion, I mentioned, I are taking two different tracks in order to further analyze the accuracy differences between using a simple vs complex model. Thus, our classification steps for both approaches are as follows:

Approach 1: Once our images are processed and the pixel array is extracted from each CT scan from approach 1 mentioned above, I can feed the data frame into a clustering algorithm to group similar patients together. As such, using the four features previously generated, I utilized the K-Means clustering algorithm in order to partition the ~1,500 patients into k=2 clusters – positive or negative for cancer – where each patient will belong to the cluster with the nearest mean. By partitioning the data space as such, I are able to gain a preliminary understanding of the distinguishing characteristics between the two groups. The K-Means algorithm partitioned the data into the two groups with relatively high accuracy. I were then able to add the K-Means clustering classification as an input parameter to our next classification method: Random Forest. Since Random Forests are known to perform well on various tasks, including unscaled data and variable selection, I decided to employ this approach in order to understand the effectiveness of our feature extraction methodology. If our Random Forest returns a low accuracy, I can conclude that our feature extraction is not adequate, setting a basis for our deep learning approach.

Approach 2: The classification of the data set produced by our second approach using convolution neural networks is done by a multilevel perceptron (MLP) which follows the structure in the image below.

Using a multilevel perceptron has various advantages, including the capability to learn both linear and non-linear models, and most importantly, the capability to learn models in real-time (on-line learning) using partial fit. The MLP then follows the steps highlighted below:

- Checks if there a pre trained model to load it before training.
- Sequential model used for training the model.
- Model uses relu activation with various activation layers
- Dropout after every stage to prevent overfitting

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903037 |

• Model is compiled with 'binary_crossentropy' - loss which is the best for binary classification, optimizer – RMSPROP being used
• Checkpointer - used to save the best model based on accuracy metrics
• After training, evaluation is done to show the efficiency of the trained model.

V. CONCLUSION

In this project, I study the use of image processing, data mining, and machine learning techniques to predict lung cancer nodules in high risk patients. Based on the research and analysis conducted for this project using a publicly available data set of lung CT scans, I were able to develop a successful model for lung cancer nodule detection. By using a hybrid of approaches in image processing and classification, I were able to develop an end to end process that detects lung cancer nodules with high accuracy. Further, by placing a heavy emphasis on automation of image processing as well as a reduction of false positives, I were able to develop a full model that runs with 70% accuracy on test data. Given the difficult nature of the problem, I faced various challenges throughout the process. First, the segmentation of lungs is a very challenging problem due to inhomogeneity in the lung region, pulmonary structures of similar densities such as arteries, veins, bronchi, and bronchioles, and different scanners and scanning protocols and difference in quality of the CT scans, the images had to be made uniform before processing. The CT scans being in hundreds of images per patient had a memory constraint while processing and also was a time consuming process since data per patient was relatively high (around 300 images per patient with around 1,500 patients).

VI. FUTURE ENHANCEMENT

For feature extraction, we have use VGG16 the simplicity of the model makes it easy to implement. However, there are other pre-trained CNN models available too for feature extraction. For complicated process like image classification, many other deep learning technologies are proposed. At each stage, one could use a novel approach to retain the maximum features, which would overall lead to a better model. Other pre-trained models could be used like Resnet, Googlenet, etc. for feature extraction and other modules for deep learning could be combined with deferent loss function, layers and optimization techniques for better results. As previously mentioned, our classification model is potentially overfitting the data, and thus, there exists a clear opportunity to research different methodologies to combat this problem. Finally, the image classification being used is MLP classifier with sequential model, however, other regularization techniques and functions could be explored to update weights in order to increase the accuracy of the model.

REFERENCES

[1] J. Jiaying Shi, M Couprie: Lung Nodule Detection In Xray Images Using Cnn
[2] Petitjean C[1], Dacher Jn.: A Review Of Segmentation Methods In Short Axis Cardiac Mr Images," MIDAS.
[3] Yangming Ou, Jimit Doshi, Guray Erus: Multi-Atlas Segmentation Of The Prostate: A Zooming Process With Robust Registration And Atlas Selection,"
[4] M. Lynch, O. Ghita, and P. F. Whelan, "Segmentation Of The Left Ventricle In Cardiac Cine Mri Using A Shape-Constrained Snake Model,"
[5]C. Petitjean and J. N. Dacher, "Cardiac Lv And Rv Segmentation Using Mutual Context Information,"
[6] G. Ongun, U. Halici, K. Leblebicioglu, V. Atalay, M. Beksac, and S. Beksac, "Feature extraction and classification of blood cells for an automated differential blood count system," inProc. IJCNN, 2001, vol. 4, pp. 2461–2466.
[7] S. Mohapatra and D. Patra, "Automated leukemia detection using haus-dorff dimension in blood microscopic images," inProc. Int. Conf. Emerg. Trends Robot Commun. Technol., 2010, pp. 64–68.
[8] S. Mohapatra, S. Samanta, D. Patra, and S. Satpathi, "Fuzzy based blood image segmentation for automated leukemia detection," inProc. ICDeCom, 2011, pp. 1–5.
[9] S. Mohapatra, D. Patra, and S. Satpathi, "Image analysis of blood mi-croscopic images for acute leukemia detection," inProc. IECR, 2010, pp. 215–219.
[10] S. Mohapatra, D. Patra, and S. Satpathi, "Automated cell nucleus seg-mentation and acute leukemia detection in blood microscopic images," in Proc. ICSMB, 2010, pp. 49–54.
[11] MedlinePlus: Leukemia.National Institutes of Health. [Online]. Available: http://www.nlm.nih.gov/medlineplus/ency/article/001299.htm
[12] J. N. Jameson, L. K. Dennis, T. R. Harrison, E. Braunwald, A. S. Fauci, S. L. Hauser, and D. L. Longo, "Harrison's principles of internal medicine,"JAMA, vol. 308, no. 17, pp. 1813–1814, Nov. 2012.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 3, March 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0903037 |

[13] S. Serbouti, A. Duhamel, H. Harms, U. Gunzer, J. Mary, and R. Beuscart, "Image segmentation and classification methods to detect leukemias," in Proc. Int. Conf. IEEE Eng. Med. Biol. Soc., 1991, pp. 260–261.
[14] D. Foran, D. Comaniciu, P. Meer, and L. A. Goodell, "Computer-assisted discrimination among malignant lymphomas and leukemia using im-munophenotyping, intelligent image repositories, and telemicroscopy," IEEE Trans. Inf. Technol. Biomed., vol. 4, no. 4, pp. 265–273, Dec. 2000.
[15] K. S. Kim, P. K. Kim, J. J. Song, and Y. C. Park, "Analyzing blood cell image do distinguish its abnormalities," inProc. ACM Int. Conf. Multim., 2002, pp. 395–397.
[16] Q. Liao and Y. Deng, "An accurate segmentation method for white blood cell images," inProc. IEEE Int. Symp. Biomed. Imaging, Atlanta, GA, USA, 2002, pp. 245–248.
[17] S. Suri, S. Setarehdan, and S. Singh,Advanced Algorithmic Approaches to Medical Image Segmentation: State-of-the-Art Application in Cardiology, Neurology, Mammography and Pathology. Berlin, Germany: Springer-Verlag, 2001, pp. 541–558.
[18] N. Sinha and A. Ramakrishnan, "Automation of differential blood count,"in Proc. Conf. Convergent Technol. Asia-Pac. Region, 2003, vol. 2,pp. 547–551.
[19] W. Shitong, K. F. L. Chung, and F. Duan, "Applying the improved fuzzy cellular neural network IFCNN to white blood cell detection,"Neurocom-puting, vol. 70, no. 7–9, pp. 1348–1359, Mar. 2007.
[20] M. Oberholzer, M. Ostreicher, H. Christen, and M. Bruhlmann, "Methods in quantitative image analysis," Histochem. Cell Biol., vol. 105, no. 5, pp. 333–355, May 1996.

INNO SPACE

SJIF Scientific Journal Impact Factor

Impact Factor:
7.488

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details