International Journal of Computer Trends and Technology
ISSN: 2231 – 2803 / https://doi.org/10.14445/22312803/IJCTT-V68I10P104

Volume 68 Issue 10, 21-24, October 2020
© 2020 Seventh Sense Research Group®

Original Article

Lung Cancer Detection Using Convolutional Neural Network on Histopathological Images

Bijaya Kumar Hatuwal[1], Himal Chand Thapa[2]

[1]*Himalaya College of Engineering (of Tribhuvan University), Chyasal, Lalitpur, Nepal*
[2]*Lecturer, Himalaya College of Engineering (of Tribhuvan University), Chyasal, Lalitpur, Nepal*

Received Date: 15 September 2020
Revised Date: 17 October 2020
Accepted Date: 19 October 2020

Abstract *- Lung Cancer is one of the leading life taking cancer worldwide. Early detection and treatment are crucial for patient recovery. Medical professionals use histopathological images of biopsied tissue from potentially infected areas of the lungs for diagnosis. Most of the time, the diagnosis regarding the types of lung cancer are error-prone and time-consuming. Convolutional Neural networks can identify and classify lung cancer types with greater accuracy in a shorter period, which is crucial for determining patients' right treatment procedures and survival rates. Benign tissue, Adenocarcinoma, and squamous cell carcinoma are considered in this research work. The CNN model training and validation accuracy of 96.11 and 97.2 per cent are obtained.*

Keywords *- Convolutional Neural Network (CNN), Machine Learning, Lung Cancer, Histopathological Image*

I. INTRODUCTION

Lung cancer is prominent among both men and women, making up almost 25% of all cancer deaths [1]. The primary cause of death from lung cancer about 80% is smoking. Lung cancer in non-smokers can be caused by exposure to radon, second-hand smoke, air pollution, or other factors like workplace exposures to asbestos, diesel exhaust, or certain other chemicals lung cancers some people who do not smoke [2].

Various tests like imaging sets (x-ray, CT scan), Sputum cytology, and tissue sampling (biopsy) are carried out to look for cancerous cells and rule out other possible conditions. While performing the biopsy, evaluation of the microscopic histopathology slides by experienced pathologists is indispensable to establishing the diagnosis [3], [4], [5], and defines the types and subtypes of lung cancers [6]. For pathologists and other medical professionals, diagnosing lung cancer and its types is time-consuming. There is a significant change when the cancer types are misdiagnosed, which leads to incorrect treatment and may cost patients' lives.

Machine Learning (ML) is a subfield of Artificial Intelligence (AI) that allows machines to learn without explicit programming by exposing them to sets of data allowing them to learn a specific task through experience [7][8]. In previous research papers, most of the authors considered using x-rays, CT scans images with machine learning techniques such as Support Vector Machine (SVM), Random Forest (RF), Bayesian Networks (BN), and Convolutional Neural Network (CNN) for lung cancer detection and recognition purpose. Some papers also considered using histopathological images, but they distinguish between carcinomas and non- carcinomas images with lower accuracy. This research paper has considered using Convolutional Neural Network (CNN) architecture to classify benignly, Adenocarcinoma and squamous cell carcinomas. We have not found other papers using the CNN model to classify only the given three different histopathological images and the given model's accuracy.

In Section II, some previous related works are reviewed. The methodology and settings used are described briefly in Section III. Similarly, the research's obtained output is explained and shown with plots and tables in Section IV. The paper's conclusion is explained in Section V and cited sources mentioned in the References section.

II. LITERATURE REVIEW

The authors W. Ausawalaithong, A. Thirach, S. Marukatat, and T. Wilaiprasitporn [9] used deep learning with a transfer learning approach to predict lung cancer from the chest X-ray images obtained from different data sources. Image size of 224X224 with 121-layer Densely Connected Convolutional Network (DenseNet-121) and a single sigmoid node was applied in a fully connected layer. The proposed model achieved 74.43±6.01% mean accuracy, 74.96±9.85% mean specificity, and 74.68±15.33% mean sensitivity for different image source datasets.

T. Atsushi, T. Tetsuya, K. Yuka, and F. Hiroshi [10] applied Deep Convolutional Neural Network (DCNN) on cytological images to automate lung cancer type classification. They considered Small cell carcinoma, Squamous cell carcinoma, Adenocarcinoma images in their dataset. The DCNN architecture of 3 convolution and pooling layers and 2 fully connected layers with dropout 0f 0.5 were used. The model developed achieved an overall accuracy of 71.1%, which is quite low.

W. Rahane, H. Dalvi, Y., Magar, A. Kalane, and S. Jondhale [11] proposed using image processing and machine learning (Support Vector Machine) for lung cancer detection on computed tomography (CT) images. Image processing like grayscale conversion, noise reduction, and binarization was carried out. Features like area, perimeter, and eccentricity from the segmented image region of interest were fed to the support vector machine (SVM) model.

M. Šarić, M. Russo, M. Stella, and M. Sikora [12] proposed CNN architectures implementing VGG and ResNet for lung cancer detection using whole side histopathology images, and the output was compared using the receiver operating characteristic (ROC) plot. Patch level accuracy of 0.7541 and 0.7205 was obtained for VGG16 and ResNet50, respectively, which is quite low. The authors explained that the given models' low accuracy was due to large pattern diversity through different slides.

The authors S. Sasikala, M. Bharathi, B. R. Sowmiya [13] proposed using CNN on CT scan images to detect and classify lung cancer. They used MATLAB for their work and has two phases in training to extract valuable volumetric features from input data as the first phase and classification as the second phase. Their proposed system could classify the cancerous and non-cancerous cells with 96% accuracy.

SRS Chakravarthy, R. Harikumar [14], used Co-Occurrence Matrix (GLCM) and chaotic crow search algorithm (CCSA) for feature selection on computed tomography (CT) and applied probabilistic neural network (PNN) of the classification task. They found that the PNN model build on CCSA features performed better with 90% accuracy.

III. METHODOLOGY

Our proposed system followed data acquisitions, data formatting, model training, testing, and prediction, described in the below sections.

A. Data Acquisition

The histopathology images are obtained from LC25000 Lung and colon histopathological image dataset [15]. Three classes of benign tissue, Adenocarcinoma, and squamous carcinoma cells of lungs with 5000 histopathology images in each category, are considered for our work.

B. Data Formatting

The obtained dataset was RGB color histopathology images with .jpeg format. The images were resized to maintain a uniform aspect ratio of one with (180, 180) pixel size for the CNN operation. All the pixel values for the images were converted in range of (0, 1) to make convergence faster. We have implemented the image acquisition technique like horizontal and vertical flip and zooming to increase the image number and variation in the data pattern. The neural network tends to over-fit in case of a limited number of training data samples trained with a higher number of epochs [16]. Fig. 1(a) and Fig. 1(b) show Adenocarcinoma's histopathology image and its augmented images, respectively, with the horizontal and vertical flip and a zoom range of 0.2 applied.

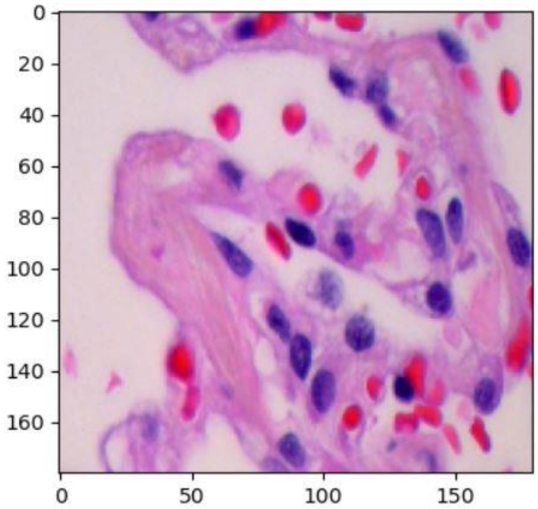

Fig. 1(a) Histopathology Image of Adenocarcinoma

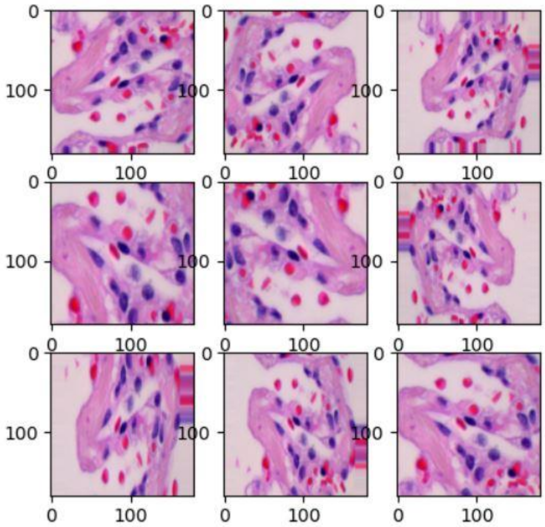

Fig. 1(b) Corresponding Augmented Histopathology Images of Adenocarcinoma

C. Model Training, Testing, and Prediction

A liner stack of layers was used to create the Convolutional Neural Network (CNNs or ConvNets) for the image classification and recognition. Training and testing images were passed through convolutional layers with kernel filters, max pooling, and fully connected layers. The softmax function was applied to classify the given object. The model was trained and tested using Google Colaboratory GPU named as a device: GPU: 0.

A neural network with three hidden layers, one input layer, and one fully connected layer was implemented for this task. Images are split in a ratio of 90:10 for training and validation purposes. Images of (180, 180) pixel size were passed to the input layer. Kernel matrix of (3, 3) with (ReLU(x) = max (0, x)) as an activation function was applied in each convolutional layer. Max pooling size of (2, 2) was implemented to reduce the computation parameters in the next convolution layer. A dropout value of 0.1 was applied to the model. A dense value of three with the sigmoid activation function was used to obtain the class probabilities for final output classes. An adaptive moment estimation (Adam) optimizer was used to calculate the learning rates for different parameters. Loss function calculates the discrepancy between the predicted output and the labeled output for the given input; categorical cross-entropy (CE) was used as a loss function for this task, which is calculated as:

$$CE = -\log\left(\frac{e^{S_p}}{\sum_j^C e^{S_j}}\right) \qquad (1)$$

C is the number of output class, S_p is the CNN score of the given positive class, and S_j is the score inferred by the net for each class C.

The performance of the developed CNN model was measured using the confusion matrix plot, and the metrics accuracy, precision, recall, and f1-score were also calculated as below:

$$Accuray = \frac{(TP + TN)}{(TP + FP + FN + TN)} \qquad (2)$$

$$Precision = \frac{TP}{(TP + FP)} \qquad (3)$$

$$Recall = \frac{TP}{(TP + FN)} \qquad (4)$$

$$F1 - Score = \frac{2 * (Recall * Precision)}{(Recall + Precision)} \qquad (5)$$

Where TP, FP, FN, and TN represents the output measures as true positive, false positive, false negative, and true negative values for the training and validation images of the models.

The trained model weights were saved into the hd5 file format and used to predict the future by loading the weights to the model architecture.

IV. RESULT AND DISCUSSION

The images were trained for 20 epochs with batch size 64 with 211 steps in each epoch. The model achieved a training accuracy of 96.11% and a validation accuracy of 97.20% in the final epoch. Below, Fig. 2(a) and Fig. 2(b) shows the plot of model accuracy vs. epoch and model loss vs. epoch for training and validation images.

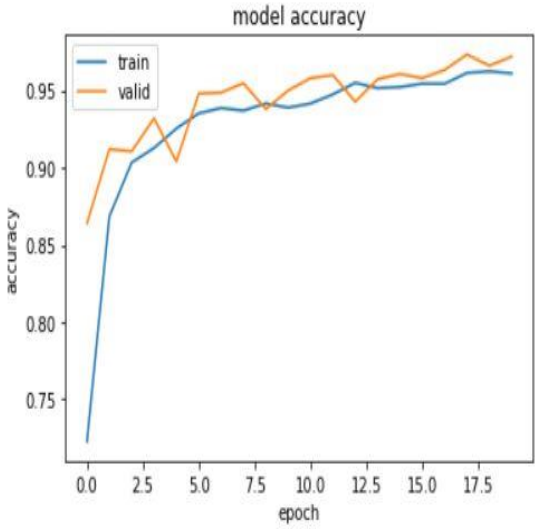

Fig. 2(a) Plot of Model Accuracy vs. Epoch for Training and Validation Images

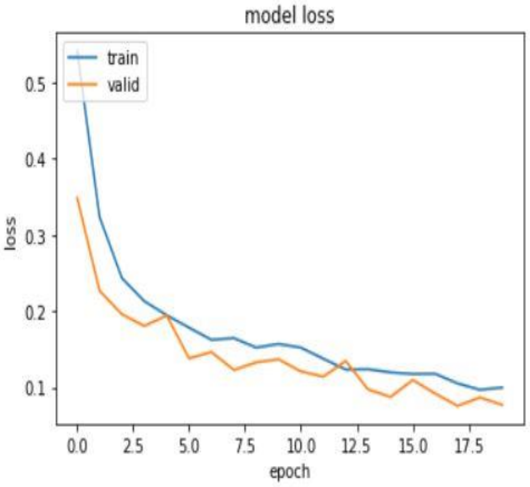

Fig. 2(b) Plot of Model Loss vs. Epoch for Training and Validation Images

Table I.
Precision, Recall, and F1-Score of Model for Different Categories

Category	**Precision**	**Recall**	**F1-score**
adenocarcinoma	0.95	0.97	0.96
benign tissue	1.00	1.00	1.00
Squamous Cell Carcinoma	0.97	0.95	0.96

The table shows the precision, recall, and f-score for the different histopathology image categories. The formula to calculate the given metrics is explained show in Section III-C.

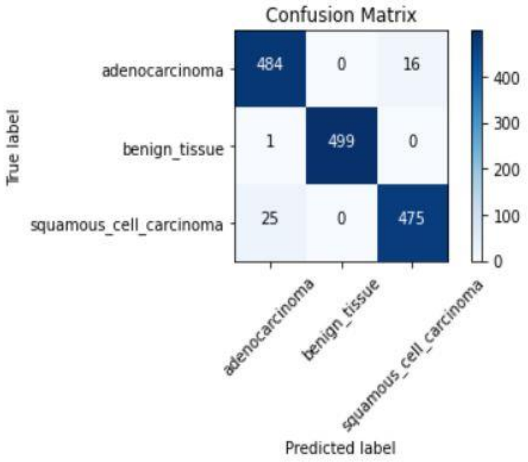

Fig. 3 Confusion Matrix of Different Image Categories for Validation Images

The confusion matrix shown in Fig. 3 depicts the true label vs. the predicted label of the images for the validation data in given labeled categories.

V. CONCLUSION

This research work presents lung cancer detection using histopathological images. A convolutional neural network (CNN) was implemented to classify an image of three different categories benign, Adenocarcinoma, and squamous cell carcinoma. The model was able to achieve 96.11% and 97.20% of training and validation accuracy. The precision, f1-socre, recall were calculated, and a confusion matrix plot was drawn to measure the model performance.

REFERENCES

[1] American Cancer Society, Lung Cancer Statistics. (2020) [Online]. Available: https://www.cancer.org/cancer/lung-cancer/about/key-statistics.html

[2] American Cancer Society, Lung Cancer Causes(2019). [Online]. Available: https://www.cancer.org/cancer/lung-cancer/causes-risks-prevention/what-causes.html

[3] G. A. Silvestri, et al. Noninvasive staging of non-small cell lung cancer: ACCP evidence-based clinical practice guidelines (2nd edition). Chest 132 (3)(2007) 178S-201S. doi:10.1378/chest.07-1360.

[4] W. D. Travis, et al. International association for the study of lung cancer/American thoracic society/European respiratory society international multidisciplinary classification of lung adenocarcinoma. Journal of thoracic oncology: official publication of the International Association for the Study of Lung Cancer 6 (2) (2011) 244-85.
doi:10.1097/JTO.0b013e318206a221

[5] L. G. Collins., C. Haines, R. Perkel & R. E. Enck. Lung cancer: diagnosis and management. American family physician 75 (1)(2007) 56-63.

[6] K. Yu, C. Zhang, G. Berry, et al. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. Nat Commun 7 (2016) 12474
doi: 10.1038/ncomms12474

[7] D. Bazazeh and R. Shubair, Comparative study of machine learning algorithms for breast cancer detection and diagnosis, 5th International Conference on Electronic Devices, Systems and Applications (ICEDSA), Ras Al Khaimah, (2016) 1-4
doi: 10.1109/ICEDSA.2016.7818560.

[8] E.D. Michie, D.J. Spiegelhalter, and C.C. Taylor, Machine Learning, Neural and Statistical Classification, Proceeding, (1994)

[9] W. Ausawalaithong, A. Thirach, S. Marukatat, and T. Wilaiprasitporn, Automatic Lung Cancer Prediction from Chest X-ray Images Using the Deep Learning Approach, 2018 11th Biomedical Engineering International Conference (BMEiCON), Chiang Mai, (2018) 1-5.
doi: 10.1109/BMEiCON.2018.8609997.

[10] T. Atsushi, T. Tetsuya, K. Yuka, and F. Hiroshi. Automated Classification of Lung Cancer Types from Cytological Images Using Deep Convolutional Neural Networks". BioMed Research International. (2017) 1-6. doi:10.1155/2017/4067832.

[11] W. Rahane, H. Dalvi, Y. Magar, A. Kalane and S. Jondhale, Lung Cancer Detection Using Image Processing and Machine Learning HealthCare, International Conference on Current Trends towards Converging Technologies (ICCTCT), Coimbatore, (2018)1-5.
doi: 10.1109/ICCTCT.2018.8551008.

[12] M. Šarić, M. Russo, M. Stella and M. Sikora, CNN-based Method for Lung Cancer Detection in Whole Slide Histopathology Images, 2019 4th International Conference on Smart and Sustainable Technologies (SpliTech), Split, Croatia, (2019)1-4.
doi: 10.23919/SpliTech.2019.8783041.

[13] S. Sasikala, M. Bharathi, B. R. Sowmiya. Lung Cancer Detection and Classification Using Deep CNN. (2019).

[14] SRS Chakravarthy and H. Rajaguru. Lung Cancer Detection using Probabilistic Neural Network with modified Crow-Search Algorithm. Asian Pacific Journal of Cancer Prevention, 20 (7) (2019) 2159-2166.
doi: 10.31557/APJCP.2019.20.7.2159.

[15] AA. Borkowski, MM. Bui, LB. Thomas, CP. Wilson, LA. DeLand, SM. Mastorides. Lung and Colon Cancer Histopathological Image Dataset. (LC25000). ArXiv: 1912.12142v1 [eess.IV], (2019)

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25 (2012).
doi: 10.1145/3065386.

[17] Narendra Mohan Tumor Detection From Brain MRI Using Modified Sea Lion Optimization Based Kernel Extreme Learning Algorithm International Journal of Engineering Trends and Technology 68(9)(2020)84-100.