A Convolutional Neural Network Application for Predicting the Locating of Squamous Cell Carcinoma in the Lung

H.S. Nogay, and T.C. Akinci

Abstract—**Squamous cell carcinoma, one of the most common types of lung cancer types, usually occurs in the middle, right or left bronchi. Squamous cell carcinoma can be easily detected by imaging methods to determine the location within the lung. However, rarely the location of some tumor types cannot be determined. In this case, it may be delayed to obtain the results of the assay such as biopsy. This possible delay also means delayed diagnosis and delayed start of treatment. In order to solve this problem, it is possible to perform applications with machine learning methods. In this study, convolutional neural networks method was used to determine the location of cancerous tumor in squamous cell carcinoma of lung. With the designed convolutional neural network model, squamous cell carcinoma tumor location in lung cancer was estimated with an accuracy rate close to 100%.**

Index Terms—**Convolutional neuran networks, squamous cell carcinoma, classification, estimation, accuracy rate.**

I. INTRODUCTION

WHILE lung cancer is a rare disease at the beginning of the 20th century, the frequency has increased steadily in parallel with the increase in smoking habits and has become the most common type of cancer in the world [1]. According to the World Health Organization (WHO) report, lung cancer is the leading cause of death among males and the second type of cancer among females all over the world [2]. Squamous cell lung cancer, also called lung squamous cell carcinoma, accounts for about 30% of all lung cancers [3]. There are many treatment options for people diagnosed with squamous cell lung cancer, and doctors are constantly working to increase and improve these types of treatments. Squamous cell lung cancer is a type of non-small cell lung cancer. Squamous cell carcinoma starts as squamous cells that look like fish scales when viewed under a microscope. This cancer, which runs along the inside of the airways in the lungs, uses the most air channels to progress. Squamous cells in the tissue that make up the skin surface are also found on the surface of the body's organs and inside the digestive tract. Squamous cell carcinoma originating in the lungs is thought to be lung cancer.

H.S. NOGAY, works in University of Kayseri, Kayseri, Turkey, (e-mail: nogay@erciyes.edu.tr). ⓘD

T.C. AKINCI, is with Department of Electrical Engineering, Istanbul Technical University, Istanbul, Turkey, (e-mail: akincitc@itu.edu.tr). ⓘD

Squamous cell lung tumors usually begin in the middle of the lung or in one of the main airways (left or right bronchus) [3, 4]. The location of the tumor may affect the appearance of symptoms such as coughing, shortness of breath, chest pain and blood in the sputum. If the tumor is enlarged, a chest radiograph or computerized tomography scan may detect a space in the lung. This space is a gas or fluid-filled cavity in the tumor mass or nodule and is a classic sign of squamous cell lung cancer. Squamous cell lung cancer can spread to multiple sites such as the brain, spine, other bones, the adrenal glands, and the liver. Smoking affects this type of cancer more strongly than any other large-cell lung cancer. Other risk factors for squamous cell lung cancer include age, hereditary diseases, environmental pollution, mineral and metal dust, exposure to asbestos or radon. Positive results may be observed with appropriate treatment methods, although this type of cancer is sometimes hardly noticed by occult development [5-7]. Pathologically, lung cancer can be divided into 4 main groups; a) squamous cell carcinoma, b) large cell carcinoma, c) adenocarcinoma, d) small adenocarcinoma. Squamous cell carcinoma in 40% - 60% of males is more common in the middle and advanced age group. The main diagnostic methods of lung cancer are; a) CT (Computerized Tomography), b) Endoscopic Ultrasonography, c) Bronchoscopy, Thoracoscopy, Dermoscopy. d) Pulmonary Radiography e) Positron Emission Tomography (PET) f) Magnetic Resonance Imaging. The main treatment modalities are surgical treatment, radiotherapy, chemotherapy and cryotherapy. Squamous cell carcinoma shows different symptoms according to its tumor location. Identification of tumor location is of great importance in early diagnosis of the disease. The region of the lung where the tumor is located is an important piece of information to be used for the treatment process. The location of the tumor is immediately revealed by imaging techniques. In some cases, however, the tumor location may not be detected by imaging techniques. This leads to delays for biopsy and the later start of treatment [7-13].

Machine learning-based prediction methods have been used effectively in recent years for different purposes in cancer types [14-17]. In this study, convolutional neural network (CNN) model, which is the most popular method of machine learning approaches in recent years, is designed. Detection of tumor location in squamous cell carcinoma lung cancer is the main goal of the study. In the first section of the study, basic information on entry and lung cancer was included. In the second section, data set and materials used in the study were mentioned. In the third chapter, the design of the convolutional neural networks used in the study is briefly explained. In the fourth section, the estimation results obtained from the CNN

model are presented. In the fifth chapter, which is the last chapter, the results are evaluated.

II. MATERIALS

The data set used in the study was obtained from the website of The Cancer Imaging Archive (TCIA) [18,19]. The actual size of the image data in the data set is 512 x 512. All image data used in this study actually consist of 3-channel data. In order to train the CNN model in shorter time, all image data in the data set is reduced to 84x84 and single channel gray scale. According to the regions of the squamous cell carcinoma tumor dataset, a data set consisting of a total of 300 image data, 60 data for five regions, was used. Figure 1 shows a sample image at the actual size in the dataset.
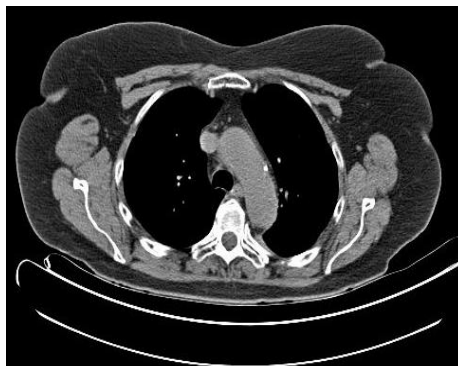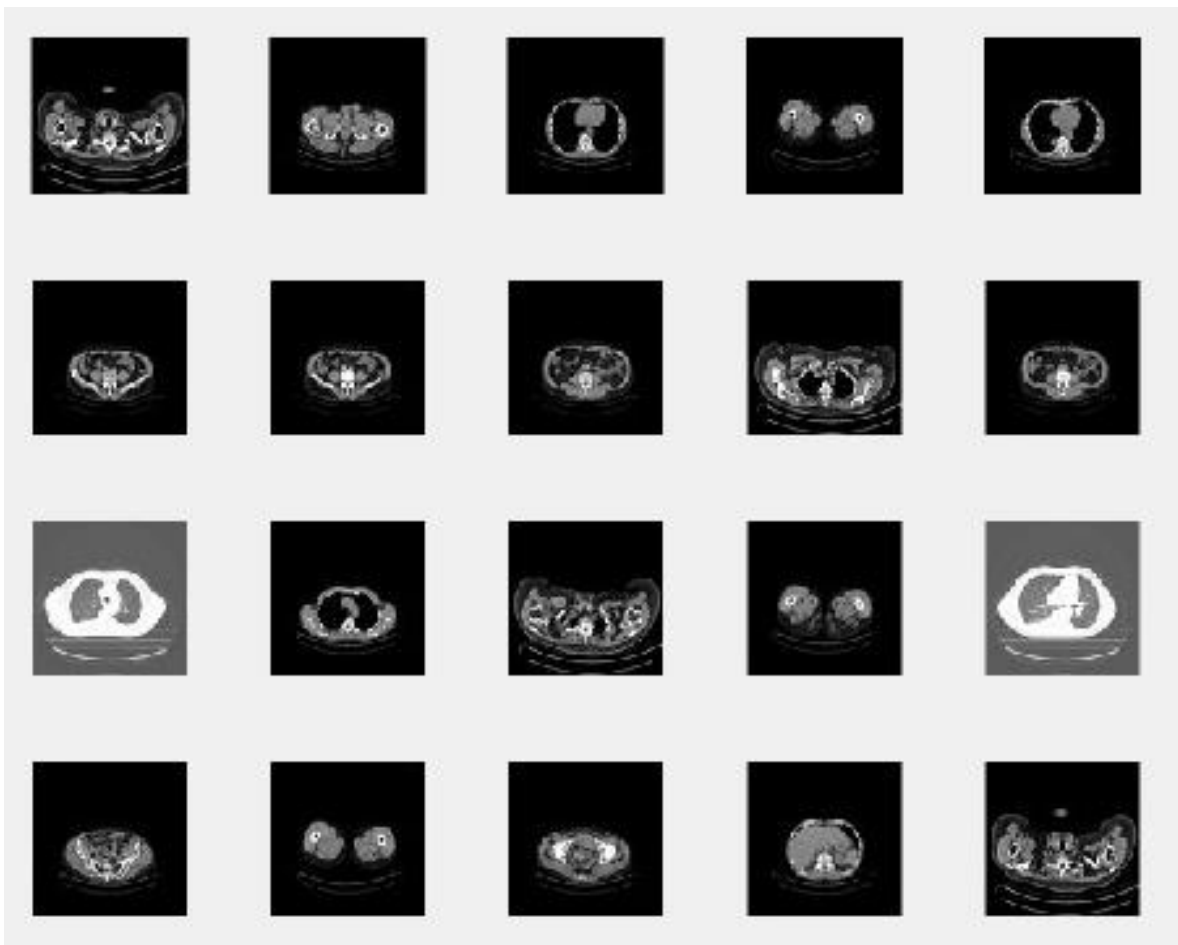

Fig. 1. Lung squamous cell carcinoma sample image

Fig. 2. Example of data set available for CNN

Figure 2 shows a sample data set consisting of 20 randomly selected images after reducing the image data to a usable size. The abbreviations and numbering used for the 5 regions of the lung where squamous cell carcinoma is present in the data set used in the study are shown in Table 1.

TABLE.1 SQUAMOUS CELL CARCINOMA LUNG LOCATION AND NUMBERING

Location	Abbreviation	Numbering
Left Lower Lobe	LLL	1
Left Upper Lobe	LUL	2
Right Upper Lobe	RUL	3
Right Lower Lobe	RLL	4
Right Middle Lobe	RML	5

III. METHODOLOGY

For the input layer of the CNN model, each image data is used in the form of a matrix form. If a randomly selected image data from the data set is redrawn by contouring the digitized matrix form, Figure 3 is obtained. In Figure 3, the data set for the input layer shows a contour of a randomly selected image data in the matrix form. The convolutional neural network designed and applied in the study is shown in figure 4. The CNN model shown in Figure 4 includes the process after the actual image data is reduced to 84x84 pixels and is mono-channelized.
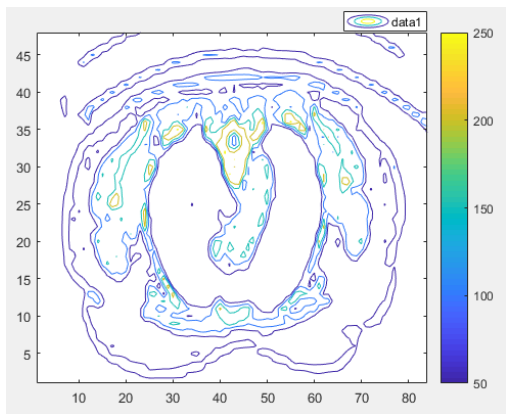

Fig. 3. Contour plot of a randomly selected sample image

The designed CNN model consists of the input layer, the convolution layer, the relu, the maxpooling layer, the fully connected layer and the softmax, respectively. As mentioned earlier, image data is used at the input layer. For the next layer, convoluted layer, 20 filters of 5x5 are used. After this filtration process, normalization with Relu transfer function is performed. In the Maxpooling layer, a 2x2 filter was used, shifted by 2 frames. The next processing layer is the 'fully connected layer' to which all neurons are connected. Classification is also performed with the last layer, the softmax layer.

In the study, 15 of 60 data used for each region were tested and the rest was used for training purposes. In this case a total of 75 data were used for the test and the remaining 225 data were used for the training of the model. The initial learning rate for the designed CNN model was accepted as 0.00001. 1000 is selected as the maximum epoch for training and testing in the study. The reason for choosing a very high Epoch number is to be able to achieve high accuracy.
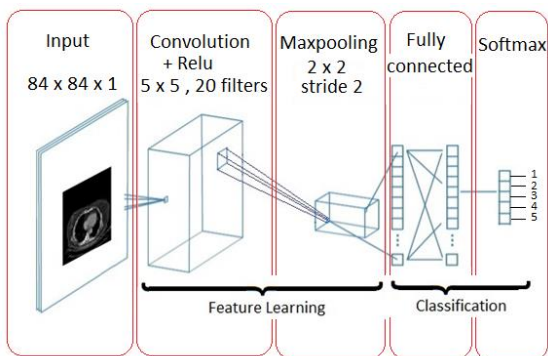

Fig. 4. Architecture of the designed CNN model

IV. RESULTS

With the training of the designed CNN model, it is estimated which region of the lung of the squamous cell carcinoma tumor is located. The accuracy of the model training and the results obtained for each 50 iterations are given in Table 2.

TABLE 2. THE EXTRACTED FEATURES

| Iteration	Epoch	Time elapsed (Second)	Mini-Batch Loss	Mini - Batch Accuracy (%)	Base Learning Rate
1	1	0.16	5.7198	18.67	1.00e-05
50	50	6.22	0.0501	99.9499	1.00e-05
100	100	12.15	0.0179	99.9821	1.00e-05
150	150	18.07	0.012	99.988	1.00e-05
200	200	23.63	0.0091	99.9909	1.00e-05
250	250	29.18	0.0073	99.9927	1.00e-05
300	300	34.72	0.0061	99.9939	1.00e-05
350	350	40.3	0.0052	99.9948	1.00e-05
400	400	45.93	0.0045	99.9955	1.00e-05
450	450	51.63	0.004	99.996	1.00e-05
500	500	57.26	0.0036	99.9964	1.00e-05
550	550	62.92	0.0033	99.9967	1.00e-05
600	600	68.63	0.003	99.997	1.00e-05
650	650	74.34	0.0027	99.9973	1.00e-05
700	700	79.97	0.0025	99.9975	1.00e-05
750	750	85.7	0.0024	99.9976	1.00e-05
800	800	91.42	0.0022	99.9978	1.00e-05
850	850	97.13	0.0021	99.9979	1.00e-05
900	900	102.87	0.0019	99.9981	1.00e-05
950	950	108.6	0.0018	99.9982	1.00e-05
1000	1000	114.45	0.0017	99.9983	1.00e-05
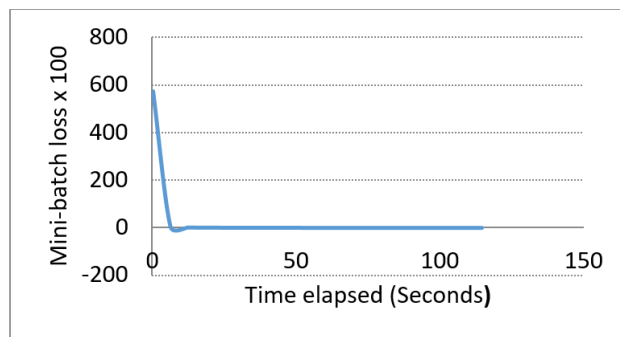
Fig. 5. The loss curve for the CNN model in the training process

At the end of each 50 epochs in the CNN model, the smallest loss ratio obtained can be viewed from the graph covering all epochs in figure 5, which are obtained according to the values given in table 2. If this graph is to be examined more clearly, the graph of Figure 6 is obtained by reducing the maximum value of the loss axis to 3. The change in losses is shown more clearly in this graph.
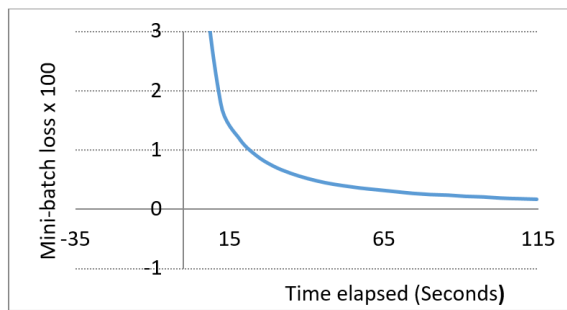
Fig. 6. Detailed loss curve

It is also possible to obtain an accuracy curve from the values in Table 2. Figure 7 shows the obtained accuracy curve. This accuracy curve is obtained by taking minimum values for every 50 epochs, as in the mini-batch loss curve. However, the total accuracy rate obtained is 99.11%.
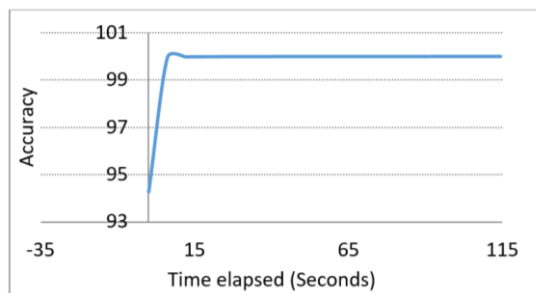

Fig. 7. Accuracy curve obtained from CNN model

V. CONCLUSION

In order to evaluate the results obtained from the CNN model designed in the study, it is seen that 1000 iterations in total are used in Table 2 when the training process is examined first. At the end of each 50 epoch, the error rate obtained by writing the lowest error rate shows an accuracy rate of 99.9983% after the last epoch. However, this accuracy is the lowest rate of the last 50 epochs. When all epochs are considered in this case, it is seen that the average accuracy rate is 99.11% of the true accuracy rate. This accuracy is an indication that the model is extremely successful. In the obtained error value, it is possible to say that the first epoch is a very high error of 5.71, but it is approaching zero by decreasing to the next epoch. Another important factor is the length of the education process. The training ended after 114.45 seconds at the end of a thousand epoch. This value may vary depending on the power of the processor, the size of the image, the number of images in the data set, and the number of convolutional and maxpooling layers used in the designed model. In fact, the number of filters used in the convolutional layer and the filter size affect both accuracy and training duration. As a result, with the designed CNN model, the estimation of which region of the lung of the squamous cell carcinoma tumor was performed in approximately 2 minutes, with an accuracy rate close to 100%.

REFERENCES

[1] Spiro S.G, Porter J.C. Lung cancer-Where are we today? Current advances in staging and nonsurgical treatment. American Journal of Respiratory and Critical Care Medicine, vol.166, no.9, pp.1166-1196, 2002.
[2] World Health Organisation, The World Health Report, 2004.
[3] Derman B.A., Mileham K.F., Bonomi P.D., Batus M., Fidler M.J. Treatment of advanced squamous cell carcinoma of the lung: A review, Transl Lung Cancer Res, vol.4, no.5, pp. 524-532, 2015.
[4] Wusheng Y., Ignacio I. Wistuba, Michael R. Emmert-Buck, Heidi S. Erickson, Squamous cell carcinoma – similarities and differences among anatomical sites, Am J Cancer Res, vol.1, no.3, pp.275-300, 2011.
[5] Wilkerson M.D., et al., Lung Squamous Cell Carcinoma mRNA Expression Subtypes Are Reproducible, Clinically Important, and Correspond to Normal Cell Types, Clinical Cancer Research. vol.16, no.19, 2010.
[6] Ateş İ., et al., Squamous Cell Cancer of The Lung with Synchronous Renal Cell Carcinoma, Turkish Thoracic Journal, vol.17, no.3, pp.125-127, 2016.
[7] Reck M. and Rabe K.F. Precision Diagnosis and Treatment for Advanced Non–Small-Cell Lung Cancer, The New England Journal of Medicine, vol.377, pp.849-861, 2017.
[8] Schild S.E., et al. Long-term results of a phase III trial comparing once-daily radiotherapy with twice-daily radiotherapy in limited-stage small-cell lung cancer, International Journal of Radiation Oncology Biology-Physics, vol.59, no.4, pp.943-951, 2004.
[9] Kulkarni A., Panditrao A. Classification of Lung Cancer Stages on CT Scan Images Using Image Processing, 2014 IEEE International Conference on Advanced Connnunication Control and Computing Teclmologies (lCACCCT).
[10] Sarker P., et al. Segmentation and Classification of Lung Tumor from 3D CT Image using K-means Clustering Algorithm, Proceedings of the 4th International Conference on Advances in Electrical Engineering (ICAEE) 8-30 September, Dhaka, Bangladesh, 2017.
[11] Usui S., et al., Differences in the prognostic implications of vascular invasion between lung adenocarcinoma and squamous cell carcinoma, An International Journal of for Lung Cancer and Other Thoracic Malignancies, vol.82, no.3, pp.407-412, 2013.
[12] Vakili M., Yousefghahari B., Sharbatdaran M. Squamous cell carcinoma of lung with unusual site of metastasis, Caspian Journal of Internal Medicine, vol.3, no.2, pp.440-442, 2012.
[13] Huang Z., Chen L.,Wang C. Classifying Lung Adenocarcinoma and Squamous Cell Carcinoma using RNA-Seq Data, Cancer Stud Mol Med Open Journal, vol.3, no.2, pp.27-31, 2017.
[14] Pearce C. Convolutional Neural Networks and the Analysis of Cancer Imagery, Stanford University, 2017.
[15] Fabio A. et al., Breast Cancer Histopathological Image Classification using Convolutional Neural Networks, Saint Etienne du Rouvray, France, 2017.
[16] Pratt H., Coenen F., Broadbent D.M., Harding S.P., Zheng Y. Convolutional Neural Networks for Diabetic Retinopathy, International Conference On Medical Imaging Understanding and Analysis, MIUA 2016, 6-8 July 2016, Loughborough, UK.
[17] Vallières, M. et al. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer, Scientific Report, vol.7, 2017.
[18] http://www.cancerimagingarchive.net/ , date of access: 10 Jan 2018.

BIOGRAPHY

H. Selçuk NOĞAY received B.S degrees in Electrical Education from Kocaeli University, and M.S and Ph.D degrees in Electrical Education from Marmara University respectively 2002, 2003 and 2008. His research interests include artificial neural networks, deep learning and biomedical engineering. He has been working as a Professor in Vocational School of Kayseri University in Kayseri, Turkey.

T. Cetin Akıncı received B.S degrees in Electrical Engineering. M.Sc. and Ph.D. degrees from Marmara University, Istanbul-Turkey. His research interests include artificial neural networks, deep learning, machine learning, image processing, signal processing and, biomedical engineering. He has been working as an Associate Professor in Electrical Engineering Department of Istanbul Technical University (ITU) in Istanbul, Turkey.