Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 7630 – 7641
Received 05 March 2021; Accepted 01 April 2021.

Design of Convolutional Neural Network for Lung Cancer Diagnosis

K. Karthikayani [1], T. Selvakumar [2], K. Ananthajothi [3].

[1]Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India. Email : karthikk3@srmist.edu.in.
[2]Department of Computer Science and Engineering, PSNA College of Engineering and Technolgy, Dindigul, India. Email : selvakumar@psnacet.edu.in
[3]Department of Computer Science and Engineering, Misrimal Navajee Munoth Jain Engineering College, Chennai, India. Email: kanandjothime@gmail.com

[*](Corresponding author): K. Ananthajothi

ABSTRACT

Computed Tomography(CT) lung screening is becoming more commonly used around the world. However, radiologists face a significant challenge in processing these videos. In a CT scan, there can be up to 600 slices. As a result, computer-aided-detection (CAD) systems are critical for a faster and more reliable data analysis. To minimize false positives, a system for CT lung screenings using convolutional neural networks (CNNs) is proposed in this study. When compared to conventional methods, the framework shows that even simple architectures are very effective at classifying 3D nodule data. We tested our model with various volume sizes and found that volume size has a significant impact on efficiency.

1. INTRODUCTION

Lung cancer, the most prevalent cancer in both men and women, is a significant cause of death around the world. According to some studies, the number of new cases of lung cancer in 2015 was projected to be around 221,200, accounting for approximately 13% of all cancer diagnoses. Lung cancer is responsible for about 27% of all cancer deaths. As a result, when lung nodules are discovered early on, they must be investigated and closely monitored. The 5-year survival rate of patients with lung cancer can be increased by around 50% if they are diagnosed early. Because of its ability to create three-dimensional (3D) images of the chest, computed tomography is the most powerful method of lung nodule detection. This results in better resolution of nodules and tumor

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 7630 – 7641
Received 05 March 2021; Accepted 01 April 2021.

pathology. The use of a CT image with computer processing to aid lung nodule diagnosis is common in clinics Tulasi Krishna et al.[2019].

A detection system and a diagnostic system are used in the computer-aided diagnosis (CAD) of lung cancer. The method classifies the candidate nodules from the previous stage as nodules or non-nodules (i.e., normal anatomic structures). The aim is to distinguish between benign and malignant nodules that have been discovered. Since the likelihood of malignancy is linked to the geometric scale, shape, and presence of pulmonary nodules, CAD may differentiate between benign and malignant pulmonary nodules using effective features like texture, shape, and growth rate. As a result, a CAD system's performance can be calculated in terms of diagnosis accuracy, speed, and automation level Ponnada et al.[2020].

One of the most difficult aspects of CT is that radiologists must interpret a large number of photographs. In a CT scan, there can be up to 600 slices. Radiologists face a significant challenge in analyzing such vast amounts of data. As a result, computer-assisted detection systems are critical for a quicker and more reliable data analysis. In recent years, neural networks, renamed deep learning, have surpassed conventional AI in every crucial task: speech recognition, image classification, and the generation of normal, readable sentences. Deep learning not only speeds up the vital mission, but it also increases the computer's precision and CT image detection and classification efficiency Shiri et al.[2018].

A nodule detection CAD system usually has two steps: nodule candidate detection and false positive reduction. The aim of the candidate detection phase is to generate nodule-like candidate points. This move necessitates a high level of vigilance, which results in a large number of false positives. A high number of false positives is undesirable since it raises the number of candidate nodules that radiologists must examine. As a result, the false positive reduction stage lowers the number of false positives while maintaining the same sensitivity. Detection scores have risen year after year as processing power has increased and new artificial intelligence (AI) algorithms have been developed. Machine learning techniques have improved detection accuracies

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 7630 – 7641
Received 05 March 2021; Accepted 01 April 2021.

significantly in recent years, but the datasets used to train the algorithms are still inadequate. These algorithms are expected to produce good results as CT image quality increases and more data is collected and labelled Zhang et al.[2019].

Convolutional neural networks (CNNs) have recently gained popularity in the machine learning field due to their superior results. CNNs are made up of neurons with weights and biases that can be learned. This algorithm is based on the structure of an artificial neural network (ANN), which is modelled after a biological neuron. Filters are trained by the machine itself, which gives CNNs an advantage over conventional neural networks. The CNN layer parameters are a series of learnable filters that enable the system to adjust to problems.These filters extract the spatial information in the input data using the convolution operation. Object identification, video analysis, speech recognition, natural language processing, and medical image analysis all benefit from CNNs Ahmed et al.[2020];Fadil et al.[2020].

The goal of this work is to use CNN architecture to create a CAD system with high sensitivity and low false positive rate. The accuracy of the classification is one of the most significant priorities in this goal. A low-accuracy quick algorithm is unsuitable for clinical use because losing true positives will be extremely dangerous to patients. Another aim is to see how the size of the input patch affects classification accuracy. Section II reviews the background, Section III describes the system model with results followed by conclusion in Section IV.

2. RELATED STUDY

Lung cancer is one of the leading causes of death in both men and women around the world, with an estimated five million deaths each year. In the diagnosis of lung disorders, a computed tomography (CT) scan may be very helpful. In many areas of medical imaging diagnosis, deep learning has proven to be a common and effective tool. Three forms of deep neural networks (e.g., CNN, DNN, and SAE) are used to model lung cancer calcification in this paper. With some modifications for benign and malignant lung nodules, these networks are added to the CT image classification mission. The LIDC-IDRI database was used to assess these networks.The CNN network had the best performance of the three networks, with an accuracy of 84.15 percent, sensitivity of 83.96 percent, and specificity of 84.32 percent Song et al.[2017]. Disparity measure is computed based on the frequency of dimensional match where the divergence is estimated from. Estimated according to the disease support measures using the data's value for disease support enhances the accuracy of the classification Ananthajothi et al[2019].

.

Lung cancer is one of the most dangerous cancers that results in a large number of deaths worldwide. The best way to increase a patient's chances of survival is to diagnose lung cancer early. A computed tomography scan is used to locate a tumour and determine the extent of cancer in the body. For computed tomography images of the lungs, the current study introduces a groundbreaking automated diagnosis classification process. The CT scan of lung images was analysed using Optimal Deep Neural Network and Linear Discriminant Analysis(LDA) in this paper.LDA is used to reduce the dimensionality of deep features derived from CT lung images in order to identify lung nodules as malignant or benign. The proposed classifier has a sensitivity of 96.2 percent, a specificity of 94.2 percent, and an accuracy of 94.56 percent when applied to CT images and then optimised using the Modified Gravitational Search Algorithm to classify lung cancer classification. The quantitative results show that the proposed classifier has a sensitivity of 96.2 percent, a specificity of 94.2 percent, and an accuracy of 94.56 percent Lakshmanaprabhu et al.[2018]. An overview of present and prospective AI applications in pathology image study is provided in this analysis, with a focus on lung cancer. The current opportunities and threats in lung cancer pathology image processing are addressed, as well as recent deep learning advances that could have an effect on digital pathology in lung cancer. Finally, the functional properties of deep learning algorithms in lung cancer prediction and diagnosis are summarized Wang et al.[2019].

The main goal of this project is to identify cancerous lung nodules in an input lung image and to distinguish lung cancer and its intensity. This study employs innovative Deep learning approaches to detect the position of cancerous lung nodules. The best

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 7630 – 7641
Received 05 March 2021; Accepted 01 April 2021.

feature extraction methods are used in this study, including Histogram of Directed Gradients, wavelet transform-based features, Local Binary Pattern, Scale Invariant Feature Transform, and Zernike Moment. The Fuzzy Particle Swarm Optimization (FPSO) algorithm is used to choose the best function after extracting shape, dimensional, gravimetric, and intensity features. Finally, Deep learning is used to classify these functions. The computational complexity of CNN is reduced thanks to a new FPSO CNN. Another dataset from Arthi Scan Clinic, which is a real-time data set, is subjected to additional valuation. According to the findings of the experiments, the novel FPSO CNN outperforms other strategies Asuntha et al.[2020]. The author presented a iterative influence measure based medical data classification algorithm which estimates the influence measure on multiple levels to identify the target class Ananthajothi et al[2019].

Lung cancer is one of the most deadly cancers in the world. The importance of early diagnosis and care in the rehabilitation of patients cannot be overstated. Histopathological photographs of biopsied tissue from potentially contaminated parts of the lungs are used by doctors to make diagnoses. The majority of the time, diagnosing the various forms of lung cancer is time-consuming and error-prone. Convolutional Neural Networks can more accurately distinguish and recognize lung cancer types in less time, which is critical for assessing a patient's treatment options and survival rate. This study takes into account benign tissue, glioblastoma, and colorectal cancer. The calibration and testing accuracy of the CNN model were 96.11 and 97.2 percent, respectively Hatuwal et al.[2020]. permutes the probability across each dimension according to different variables, the multi-attribute disease probability has been estimated for each disease class Ananthajothi et al[2019].

Deep learning methods are still evolving and improving their success on different tasks today. One of the most important factors is that as we use more technologies, we collect more data. This situation improves the algorithms' performance since deep learning algorithms are highly dependent on the volume of data. Another explanation is that, as CPU and GPU capacity increases, more data can be processed at once, resulting in faster performance than in the past. This allows for more trial and error with the

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 7630 – 7641
Received 05 March 2021; Accepted 01 April 2021.

algorithms, leading to a quicker convergence to the best model. In addition to the algorithms described above, multiple individuals and commercial company teams participated in the candidate detection challenge. Therefore, still there is a need for a better deep learning algorithm.

3. SYSTEM MODEL

The LUNA16 Challenge dataset is ideal for deep learning algorithms because it contains a large number of CT images. There are also named candidate points in the dataset, making it ideal for supervised learning algorithms. The size of the input tensor is one of the most critical parameters. Since these parameters must be consistent with each other, the size of the input affects all other parameters in the model. The size of the nodule varies from 3 mm to 34 mm in lung CT scans, making the input size decision much more difficult. There is no objective analysis in the literature that compares the impact of input tensor size on device efficiency. As a result, we wanted to observe and compare the impact of feedback on the system's output. Furthermore, we wanted to demonstrate whether or not different findings can be used in decision fusion to improve efficiency. To observe these results, we'll need to build a finely tuned CNN architecture that can handle all input sizes.

3.1 Dataset

The initial LUNA16 Challenge dataset provided by the organisers, as well as the participants' candidate detection algorithms, shaped the dataset used in this analysis. There are 554,670 candidates, but only 766 nodules out of 997 are included in this group. The algorithms failed to detect 20 nodules, so they are not included in the list. A.csv file is included with the CT dataset. Each candidate's position and corresponding class (nodule or non-nodule) are mentioned in this file. There are 557 true positives in the list since certain nodules were found several times in different locations. False positives were assigned to the remaining candidate points. In other words, just 0.1 percent of the candidates are true positives, resulting in a heavily biassed dataset (1:323).

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 7630 – 7641
Received 05 March 2021; Accepted 01 April 2021.

3.2 Preprocessing

The samples in the dataset were carefully analyzed before creating the network architecture. Since we decided to use CNN, sample parameters such as the distance between voxels were critical. The size of the neural network system's input must be the same for all samples during the training phase. Despite the fact that input measurements are measured in pixels, real-world dimensions are measured in millimeters. The architecture would not produce meaningful results if the same input dimensions correspond to different real-world dimensions because it was educated on different volumetric sizes. We studied the voxel spaces in all scans to deal with this case. While the horizontal axis in figure 1 represents the distance in millimeters, the vertical axis depicts the number of samples available for a given distance. This phase eliminated the inconsistencies between scans from different scanning machines. As a result of this move, the overall sizes of all scans differed. Since we concentrated on the nodules themselves in our network system, the overall scan size is unimportant in our algorithm.

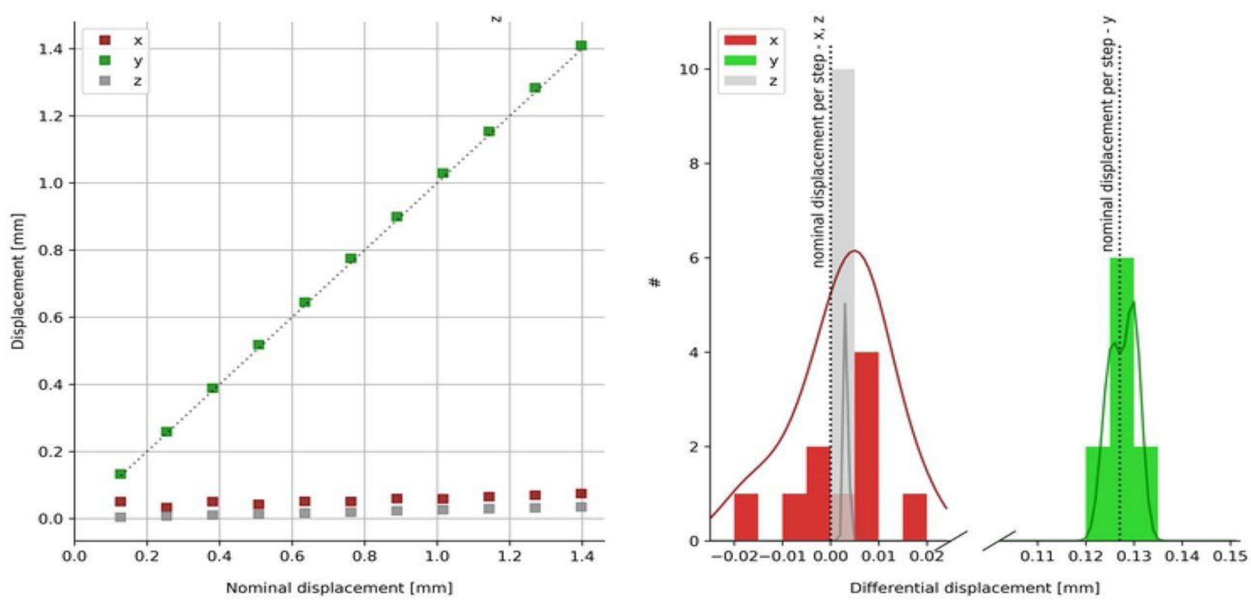

Figure 1. Distances between voxels in X, Y and Z axes.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 7630 – 7641
Received 05 March 2021; Accepted 01 April 2021.

The key idea is to use the same voxel scale for all scans so that the voxel sizes correspond to the same real-world dimensions. In a typical training scenario for this dataset, network parameters will learn the false positive structure since those cases would be experienced more often during the training phase. We used transfer learning and updated mini-batch techniques to avoid this issue.

3.3 Proposed CNN Architecture

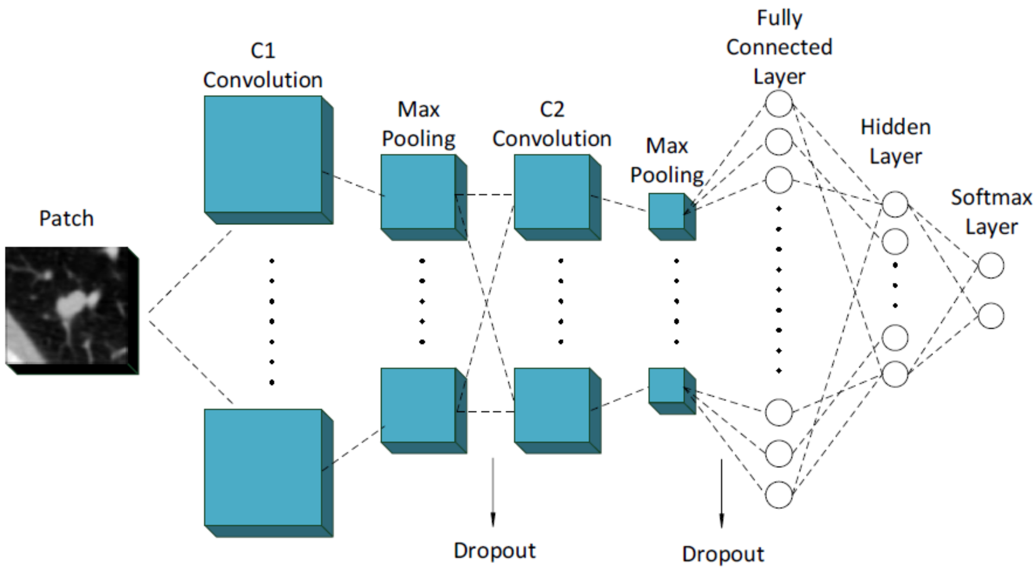

Figure 2. Proposed CNN Architecture

When tuning the CNN model, the number of convolution layers, accompanied by pooling layers and filter sizes, are the two most critical parameters. As a result, we ran multiple tests with various filter sizes and convolutional layer counts. The computational difficulties of CNN architectures were planned to be identical in these experiments. It was possible to build a sophisticated CNN architecture as shown in Figure 2, but the main goal was to fine-tune the model to distinguish input patch sizes. Furthermore, training a complex CNN model for this task will take far too long. Just one convolutional layer extracts insufficient features. Using more than three convolution layers, on the other hand, increases the difficulty and training time. As a result, we compared the models with two and three convolutional layers. In most instances, the max pooling layer produces better results. As a result, it is used as a phase after the convolutional layer. Finally,

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 7630 – 7641
Received 05 March 2021; Accepted 01 April 2021.

dropout is used to improve the model's generalization ability. The rest of the model is a traditional ANN. After the completely linked layer, there is only one secret layer. The number of nodes in the hidden layer was a critical step in this process.

3.4 Experimental results

The challenge organizers included a structure for evaluating model outcomes. The Python programming language was used to build the framework.

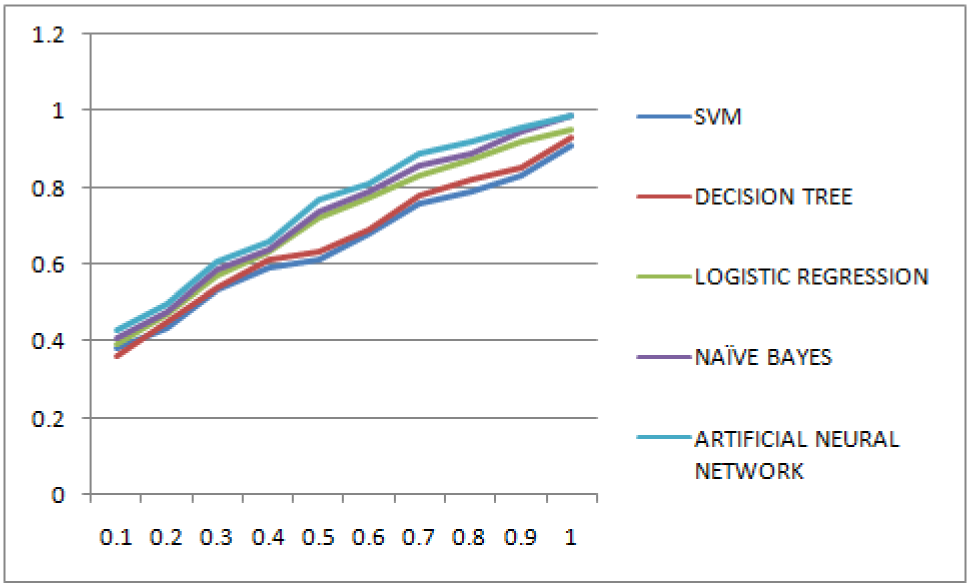

Figure 3. Sensitivities of the models at 10 false positive rates and scores

The system presented maximum, mean, and minimum false positive sensitivities per scan ranging from 0.1 to 1 as an output. Figure 3 depicts the scores for five different models. For similar CNN architectures, each patch produced different results. When SVM and ANN are compared, there is a 0.107 point difference in their ratings, which is a significant difference for models that are otherwise quite similar. This situation demonstrates the importance of qualified volume size in nodule detection problems. Even if a single model yields good performance, it must be compared to a similar model with different patch sizes. It can be shown that as the patch size grows, the model's output grows as well. Smaller patch sizes have a higher candidate detection probability for smaller nodules, despite their lower accuracy. Despite the fact that our primary goal was not to increase overall accuracy, the proposed model architecture is comparable to many

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 7630 – 7641
Received 05 March 2021; Accepted 01 April 2021.

other submissions. Overall efficiency can be enhanced by using broader and more diverse network architectures.

4.CONCLUSION

The aim of this study was to propose a false positive reduction scheme for candidates based on lung CT images, as well as to compare the effects of different patch sizes while training the algorithm to improve overall efficiency. Since interpreting these scans is a difficult task for radiologists, candidate identification and classification from lung CT images are critical. There are several slices to study for a single scan. Reviewing these slices takes time and is subject to oversight. As a result, in order to obtain a reliable result on these scans, there is a greater need than ever for a CAD device. To minimize false positives, a system for CT lung screenings using convolutional neural networks (CNNs) is proposed in this study. Another thing we learned from the experiments is that the input patch's receptive field is very significant in the training method. Small and large volumes concentrate on various characteristics during preparation, and ensembles of classifiers outperform individual classifiers.

REFERENCES

1. Ahmed, Tasnim & Parvin, Mst & Haque, Md & Uddin, Mohammad. (2020). Lung Cancer Detection Using CT Image Based on 3D Convolutional Neural Network. Journal of Computer and Communications. 08. 35-42. 10.4236/jcc.2020.83004.
2. Asuntha, A. & Srinivasan, Andy. (2020). Deep learning for lung Cancer detection and classification. Multimedia Tools and Applications. 79. 10.1007/s11042-019-08394-3.
3. Fadil, Hamdalla & Al-Huseiny, Muayed & Mohsen, Furat & Khalil, Enam & Hassan, Zainab. (2020). Diagnosis of Lung Cancer Based on CT Scans Using CNN. IOP Conference Series: Materials Science and Engineering. 928. 022035. 10.1088/1757-899X/928/2/022035.
4. Hatuwal, Bijaya & Thapa, Himal. (2020). Lung Cancer Detection Using Convolutional Neural Network on Histopathological Images. International Journal of

Computer Trends and Technology. 68. 21-24. 10.14445/22312803/IJCTT-V68I10P104.

5. Lakshmanaprabu, S.K. & Mohanty, Sachi & Shankar, Dr & N., Arunkumar & Ramirez-Gonzalez, Gustavo. (2018). Optimal deep learning model for classification of lung cancer on CT images. Future Generation Computer Systems. 92. 10.1016/j.future.2018.10.009.
6. Ponnada, Venkata Tulasiramu. (2020). Efficient CNN for Lung Cancer Detection. International Journal of Recent Technology and Engineering. 8. 3499-3505. 10.35940/ijrte.B2921.078219.
7. Ananthajothi.K & Subramaniam.M , 'CLDC: Efficient Classification of Medical Data Using Class Level Disease Convergence Divergence Measure', International Journal of Innovative Technology and Exploring Engineering (IJITEE), ISSN 2278-3075, Vol. 8, no. 10, pp. 2256-2262 (2019). https://doi:10.35940/ijitee.J1123.0881019
8. Song, QingZeng & Zhao, Lei & Luo, XingKe & Dou, XueChen. (2017). Using Deep Learning for Classification of Lung Nodules on Computed Tomography Images. Journal of Healthcare Engineering. 2017. 1-7. 10.1155/2017/8314740.
9. Shiri, Isaac & Geramifar, Parham & Bitarafan Rajabi, Ahmad & Mohammadzadeh, Ali & Pouraliakbar, Hamidreza. (2018). Deep convolutional neural network (D-CNN) for efficient and automatic lung cancer detection. Insights into Imaging. 9. 10.1007/s13244-018-0603-8/b-734.
10. Tulasi Krishna, Sajja & Devarapalli, Retz Mahima & Kalluri, Hemantha kumar. (2019). Lung Cancer Detection Based on CT Scan Images by Using Deep Transfer Learning. Traitement du Signal. 36. 339-344. 10.18280/ts.360406.
11. Ananthajothi.K & Subramaniam.M , 'Efficient Classification of Medical data and Disease Prediction using Multi Attribute Disease Probability Measure', Applied Mathematics & Information Sciences, ISSN 1935-0090, E.ISSN 2325-0399, Vol. 13, no. 5, pp. 783-789 (2019). https://doi:10.18576/amis/130511
12. Wang, Shidan & Yang, Donghan & Rong, Ruichen & Zhan, Xiaowei & Fujimoto, Junya & Liu, Hongyu & Minna, John & Wistuba, Ignacio & Xie, Yang & Xiao,

Guanghua. (2019). Artificial Intelligence in Lung Cancer Pathology Image Analysis. Cancers. 11. 1673. 10.3390/cancers11111673.

13. Zhang, Chao & Sun, Xing & Dang, Kang & Li, Ke & Guo, Xiao-wei & Chang, Jia & Yu, Zong-qiao & Huang, Fei-yue & Wu, Yun-sheng & Liang, Zhu & Zhang, Xue-gong & Gao, Xing-lin & Huang, Shao-hong & Qin, Jie & Feng, Wei-neng & Zhou, Tao & Zhang, Yan-bin & Fang, Wei-jun & Zhong, Wen-Zhao. (2019). Toward an Expert Level of Lung Cancer Detection and Classification Using a Deep Convolutional Neural Network. The Oncologist. 24. theoncologist.2018-0908. 10.1634/theoncologist.2018-0908.
14. Ananthajothi, K., Subramaniam, M. Multi level incremental influence measure based classification of medical data for improved classification. Cluster Comput 22, 15073–15080 (2019). https://doi.org/10.1007/s10586-018-2498-z