

Deep Learning and Probabilistic Neural Networks Based Detection and Classification of Lung Diseases for Pneumonia

P. V. Naga Lakshmi*¹, Dr. K. Vedavathi²

Submitted: 23/11/2023

Revised: 29/12/2023

Accepted: 08/01/2024

Abstract: Lung illness is a widespread problem in every region of the globe. Among these are asthma, pneumonia, chronic obstructive pulmonary disease (COPD), fibrosis, and tuberculosis. Detecting lung disease early on is crucial. Several models have been created that use a combination of machine learning and image processing to achieve this goal. Convolutional neural networks (CNN), vanilla neural networks (NN), visual geometry group-based neural networks (VGG), and capsule networks are only some of the well-known deep learning approaches used to predict lung illnesses. Since the publication of the novel Covid-19, numerous research projects focusing on the novel's ability to accurately foresee the future have been started all around the world. Since a number of individuals died from severe chest congestion, the early lung disease known as pneumonia is likely to have a tight connection to Covid-19 (pneumonic condition). The distinction between COVID-19 and other lung disorders, such as pneumonia, can be difficult for medical professionals to make. The most precise method of predicting lung disease is by X-ray imaging of the chest. Using patient chest X-ray images as the data source, we provide a novel framework in this study for the prediction of lung disorders including pneumonia and Covid-19. The system collects data, improves images, analyses regions of interest (ROIs) adaptively and precisely, extracts characteristics, and predicts diseases.

Keywords: Deep Learning, Hybrid Clustering, Classification, Probabilistic Neural Networks Lung Diseases, Pneumonia.

1. Introduction

Environmental changes, temperature changes, lifestyle changes, and other factors are contributing to a rapid increase in the adverse effects of disease on human health. This has resulted in an elevated risk to one's health. In 2016, the leading causes of COPD were responsible for the deaths of approximately 3.4 million people, smoking and pollution [1,2]. Asthma was responsible for the deaths of an additional 400,000 people worldwide. In developing countries and those with low to intermediate incomes, where a large number of individuals are battling against both destitution and air contamination, the gamble of lung sicknesses is particularly critical. According to the World Health Organization (WHO), indoor air pollution is the cause of nearly 4 million premature deaths annually. Asthma and pneumonia are two examples of such conditions. Therefore, it is essential to take measures to minimise environmental damage. Efficient diagnostic equipment is essential for helping in early diagnosis of lung issues. Since late December 2019, infected people have been experiencing

*1*Research Scholar, GITAM Institute of science, GITAM University, Vishakapatnam, Andhra Pradesh, India and Assistant Professor, Department of Computer Science Loyola Academy, Secunderabad, Telangana India.

*2*Professor, Department of Computer Science, GITAM Institute of Science, GITAM University, Vishakapatnam, Andhra Pradesh, India.

* Corresponding Author Email: nagmtechloyola@gmail.com

severe lung damage and breathing problems due to an emerging coronavirus known as COVID-19. The virus that causes COVID-19 and other viruses as well as bacterial infections can both cause pneumonia, a type of lung disease [3]. [Causes on pneumonia] As a result, diagnosing lung disorders at an earlier stage is more crucial than it has ever been. Machine learning and deep learning are two examples of learning algorithms that may make substantial contributions towards this goal. The importance of digital technology has grown over the past few years on a global scale. This paper's research, which makes use of the deep learning method, may help doctors and researchers figure out how to better diagnose lung illness. An enormous number of lung X-beams are utilized as a feature of the dataset. The provided approach may also aid in improving illness detection accuracy, which can save the lives of many susceptible individuals and reduce the total sickness rate. The expansion of the population is one of the reasons why the health plan has not yet been put into place [3,4]. Predicting diagnostic information from X-ray pictures using machine learning algorithms has been the subject of several studies. [5–7]. It is past time to find a solution to this conundrum given that computers are now in the hands of those who manage them and the vast majority of records are available to the general public. The development of computer science for use in medicine and health initiatives is one way that this solution can contribute to a decrease in the cost of receiving medical treatment. As a means of achieving Kaggle is used to acquire the NIH chest X-ray

image dataset for the implementation. [8,9], which a computing environment in which all of the source code is freely available to anyone. In order to diagnose lung illness, this work introduces a novel hybrid classification approach, which proved effective when applied to the dataset given above. The primary result of this research is a novel hybrid deep learning system well-suited for predicting lung disease from X-ray pictures. In November 2019, Covid-19 entered the world, posing a serious threat to human existence that quickly spread over the rest of the world. China was where this threat was first noticed. According to estimates, the sickness has now infected more than 63.2 million individuals worldwide, and 1.47 million people have passed away as a direct result of the infection. The information that nations need to protect themselves from Covid-19 is constantly being disseminated by the World Health Organization (WHO) (Fong et al., 2021). The US of America, India, Brazil, Russia, France, Italy, and China are only not many of the nations that have been seriously influenced by this danger. (Salepci et al. 2020) [10]. In spite of the fact that several nations were working on the production of the vaccine during the first six to eight months of the Covid-19, there was still no vaccination developed to protect against this threat. At long last, a couple of nations have made progress in fostering an immunization against the Right now in the testing or exploration transformative phases is the Coronavirus infection. Fever, cough, and shortness of breath are common symptoms of the COVID-19 virus in infected individuals. A severe pneumonic condition, on the other hand, affected people's lungs and ultimately caused their deaths. A significant decrease in oxygen level caused severe chest congestion (pneumonia) in the majority of Covid-19 victims, leading to a fatal heart attack. This was the primary cause of death for most fatalities (Chen et al., 2021). While pneumonia causes inflammation in the air sacs that are present in a person's lungs, it is a type of lung illness. This condition is known as a pneumonitis. Pneumonitis is the name given to this medical condition. It is possible that a significant amount of fluid will accumulate in your lungs, making it difficult for you to breathe as a result. If this occurs, you may experience respiratory distress. There are numerous disorders that can lead to pneumonia, including infections with viruses (like COVID-19 or the flu), infections with bacteria, and the common cold Diagnosing lung infections, such as viral, bacterial, or Covid-19 pneumonia, using chest X-ray pictures is now a very difficult assignment for medical practitioners. This is because the spread of the Covid-19 illness has made it more likely that lung infections will be misdiagnosed (Zhan et al. 2021); Bentivegna et al. (2020); Tabatabaei et al. (2020). Novel Coronavirus Infected Pneumonia (NCIP) is the name given to pneumonia brought on by a coronavirus that has just been discovered. The virus-caused lung disease is known by this term (Fong et al., 2021). The danger associated with lung cancer, on the other

hand, is unbearable (Sun et al. 2016; Zhou et al. 2002) [11]. The World Health Organisation (WHO) estimates that 8 million cases of lung cancer have been identified to date. However, this number is not higher than the number of people who have developed lung problems in the past 15 months as a result of Covid-19 and pneumonia. Several publications have been published on the subject of lung cancer, such as Using computer vision and soft computing, Kumar et al. (2015)[12], Li et al. (2018), and Makaju et al. (2018)[13] aim to approaches to predict the disease early. Medical imaging technologies such X-rays, MRIs, CT scans, isotope imaging, and a nuclear magnetic resonance imaging system variety of other imaging modalities, have been utilised quite frequently in the process of diagnosing lung disorders. When diagnosing patients with lung disease, radiologists and other medical professionals frequently use X-rays and CT scans, respectively, to look for indications of the disease. As a consequence of this, the chest X-ray is recommended by a significant number of medical professionals for the identification of pulmonary conditions, especially during the Covid-19 era. Using X-rays for imaging has been utilised by medical professionals for many decades to investigate and diagnose a wide variety of abnormalities that can occur organs of the human body (Khatri et al. 2020[14]). It's not a new practise by any means. Numerous studies have shown that X-rays can accurately and cheaply identify various disorders. In addition to its economical effectiveness and non-invasive characteristics, it also provides pathological modifications, which is a unique benefit of this diagnostic method [15] 2020 (Yasin et al.). Chest X-rays can reveal lung infections in a number of different ways. Consolidations, blunted costophrenic angles, widely dispersed nodules, cavitations, and infiltrates are a few of these manifestations (Angeline et al., 2020). This allows radiologists to utilise X-ray pictures to diagnose a wide range of conditions. According to Padma and Kumari, these ailments include pneumonia, pleurisy, nodule, effusion, infiltration, fractures, pneumothorax, and pericarditis. (2020; Rousan et al. 2020).[16]. Radiologists are able to do this because of the versatility of X-rays.

Radiologists have a difficult challenge when attempting to diagnose and categorise lung disorders from chest X-ray pictures. Due to this, researchers have focused heavily on creating tools for the automated identification of lung disorders (Avni et al. 2011, Jaeger et al. 2014, Patrapisetwong and Chiracharit 2016). Over the last decade, several CAD systems have been accessible with the goal of utilising X-ray pictures to identify lung disorders. Such systems, however, were unable to perform at the level necessary for the recognition and classification of lung illnesses. These responsibilities have become extremely challenging for CAD systems of this kind as a direct result of recent cases of COVID-19-assisted lung infections. It is absolutely necessary to recognize the onset of pneumonia in the lungs and to categorise the condition as either a bacterial

infection, an infection caused by the Covid-19 virus, or a viral infection [17].

As a result of this classification, pneumonic patients are in a position to receive the appropriate medical attention. At Covid-19, a number of studies that made use of CAD systems were presented, Using chest X-ray images, automated image processing and deep learning methods were created with the goal of identifying pneumonia disease. An innovation that fully automates deep learning automates the process of learning and extracting features. As a result, it takes more time to train on entire datasets and find anomalies when using deep learning. As a direct consequence of this, solutions of this kind are not reliable and do not hold up well when subjected to an increase in the total count of datasets. Convolutional neural network (CNN) and other deep learning techniques have received a lot of attention for their potential use in the identification of lung diseases (Asuntha and Srinivasan 2020)[18]. This is due to the fact that deep learning approaches possess superior accuracy and feature representation.

On the other hand, the computational complexity is not yet a topic of discussion and has not been addressed by any of the studies. Neither of these things has been done. The computational complexity of CNN is higher than that of other methods with a similar purpose because each X-ray image has a feature space that is high-dimensional.

II Pre-Processing

After the photos have been acquired, the image will be evaluated for a subsequent step of pre-processing, which will involve the consideration of noise and background information. Before continuing with the processing, it is necessary to apply techniques that eliminate noise in order to get rid of this extraneous information in the image. It is utilised to get rid of unnecessary information, such as noise, undesired backdrop component, depiction of the pectoral muscle on the skin, among other artefacts. Salt & Pepper, Gaussian, Speckle, and Poisson noise are the four types of noise that appeared in the photos of the skin [8].

The processed image is given a hybrid clustering strategy after modification so that the skin lesion region can be segmented Using K-Means and fuzzy C-Means clustering. This technique's name is segmentation. preliminary step in the K-means approach. The cost junction needs to be minimised as much as possible at the cluster centres, and this cost fluctuates at random based on the memberships of user inputs. The primary objective of segmentation is to divide a picture into distinct clusters depending on the area of interest, which in this case is the presence of skin cancer. According to the data shown in figure 9, once the specified edges have been found, uses the K-Means clustering algorithm. proposed system first, and fuzzy c means clustering is implemented afterwards.

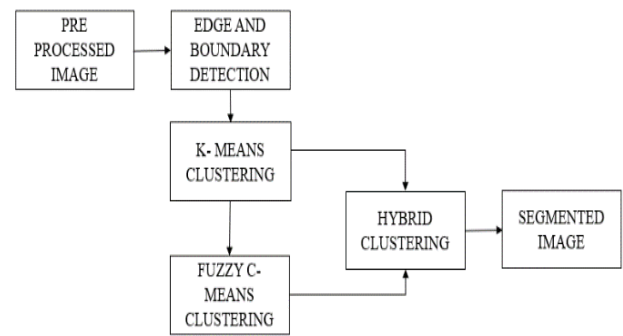


Fig. 1: Segmentation with Hybrid Clustering method

A. K-Means Clustering

In this section, we will go through the fundamental building blocks of the K-Means Clustering approach. Let $Z = z_j \mid j = 1, \dots, N$ be each of the data for Y , and let $Y = y_i \mid i = 1, \dots, P$ be the characteristics of P -dimensional vectors. K-means algorithm clusters where Z is $TP = T_j \mid j = 1, \dots, k$. Each section of the picture is given an average value based on its location in the centre of the cluster. Below, we'll take a closer look at each of the phases that make up K-Means clustering:

1. To begin, generate with centroids C , which will serve as the beginning points for the random distribution.
2. The distance between the pixel and the centre must be determined using the Euclidean method, which is written as "from X to C ."
3. Using the X_i distribution for the $i=1$ equation, $N(z_i, C)$ What is the value of the base?
4. Locating the new cluster centre is denoted by the symbol C_i , which is defined as follows:

$$C_i = \frac{1}{n_i} \sum_{j=1}^{n_i} (x_j) m_{ij}$$
5. Beginning at step 2, the process must be repeated in order to complete all of the clusters.

When the centroids settle into a particular cluster t at a threshold TH and stay there forever, we say that cluster t is stable, this is referred to as convergence. The positions must then be updated while taking the threshold value TH into account. This technique for K-Means clustering is depicted in detail in figure 1, which can be found here. Input photos that have been segmented using this approach are displayed in Figure 2.

$$\left| \frac{c^t - c^{t-1}}{c^t} \right| < TH \quad (1)$$

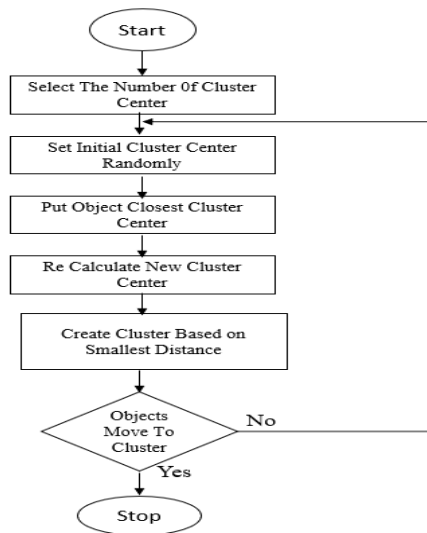


Fig. 2: Clustering with K-Means

B. Clustering using fuzzy C-Means

To separate the images into multiple cluster areas with parallel image pixel values, the fuzzy c-mean segmentation algorithm is typically utilised. This method will be applied on the image's space. This fuzzy version of the k-means segmentation method finds widespread use in the analysis of medical pictures. Fuzzy C-Means (FCM) is a clustering method that was initially presented in [12] and improved upon in [13]. This approach use iterative clustering to achieve a perfect C partition by minimising the weighted inner group sum of squared error. It finds widespread use in the pattern recognition and picture segmentation fields. The following steps make up the typical implementation of this method.

$$\forall x \left(\sum_{k=1}^{num.clusters} u_k(x) = 1 \right) \quad (2)$$

With a cluster of all the points, it is shown as algorithm centroids.

$$Center_k = \frac{\sum_x u_k(x)^m x}{\sum_x u_k(x)^m} \quad (3)$$

According to the aforementioned formulae, the cluster's centroid is located at a distance of:

$$u_k(x) = \frac{1}{d(center_k, x)} \quad (4)$$

Coefficient is then a valid parameter for a distribution that is greater than 1. The result is 1.

$$u_k(x) = \frac{1}{\sum_j \left(\frac{d(center_k, x)}{d(center_j, x)} \right)^{2/(m-1)}} \quad (5)$$

The amount equivalent to 2m for a linear normalisation of coefficients with a value of 1. The K-means approach [14] is similar to how this operates; when a cluster is 1 metre closer, it is significantly weighted more than other clusters. The phases that are described in the following paragraphs are part of the fuzzy C-means

clustering method, which is nearly identical to the K-means clustering methodology.

In terms of their number, the clusters must be selected at random. The process must be repeated exactly as it is shown in each step of equations 2,3,4,5. The previous equation (5) must be used to get the centre, and equations 3 and 5 must be used to calculate the coefficients using the center value from (5).

CLASSIFICATION

A wide variety of issue categories, including the successful use of neural networks has benefited a variety of fields, including economics, engineering, geology, physics, and biology. From a statistical perspective, neural networks are fascinating. due to the fact that they could potentially be used to solve problems involving prediction and categorization. Emulating the brain architecture present at birth is the basis for the PNN approach, which was developed. The connections between the neurons are made according to the established design so that the classification operation can be carried out successfully. The weights of the neurons are formed in a manner that is determined by the hybrid characteristics. After that, the characteristics of the hybrid that are characteristic of it are used to identify the weights' relationships with one another. The suggested network's layer count is based on the total number of weights used in the architecture. The representation of artificial neural network architecture here by the figure. In its most basic form, PNN may be broken down into two stages for categorization, namely training and testing. The layer-based architecture will be utilised throughout the training process that will be carried out. The input layer facilitates the mapping procedure on the input dataset, which is comprised of hybrid features ordered per weight distributions.

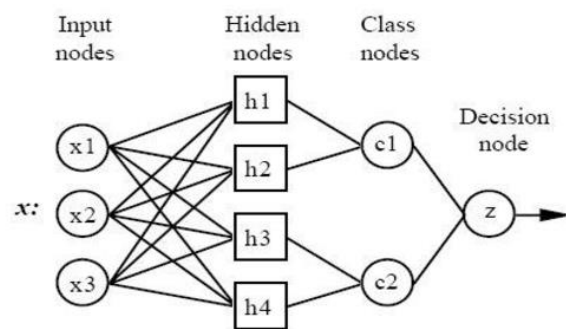


Fig 3. Layered architecture of PNN model.

The PNN configuration comprises of four discrete layers that are covered from view and each has a weight. After applying 96 11x11 filters at a stride of 4 pixels, the first convolutional2D hidden layer of the network reads in skin lesion images that are 224 by 224 by 3 pixels. The decision normalization layer and the class node activation layer come next. The two levels of concealed class nodes were subjected to the classification procedure. Information about the normalcy and anomalies of the characteristics of

skin cancer is stored separately by each of the two levels of the hidden layer. It is parted into a typical order stage and a distorted characterization stage in view of the division models. The result layer presently has marks for these two classes. The weights of both malignant and benign cancers are stored on the second level of the buried layer, while the weights of abnormal cancers are stored elsewhere. Both of these elements are again found in the hidden layer. These healthy and unhealthy weights are also labelled on the output layer. When the test image is employed, the classification stage of the testing procedure puts the test image's hybrid characteristics to the test. The way it will function is determined by the maximum feature matching criteria, which makes use of the Euclidean distance. The picture is deemed to be of normal skin if the labels hidden in layer 1 correspond to the characteristics. If the hidden layer C1 labels with the highest weight distribution match, it is determined to be a benign affected cancer image. If the buried layer C2 labels and the feature match occurred while the weight distribution was modest, then the image is categorised as having malignant impacted cancer.

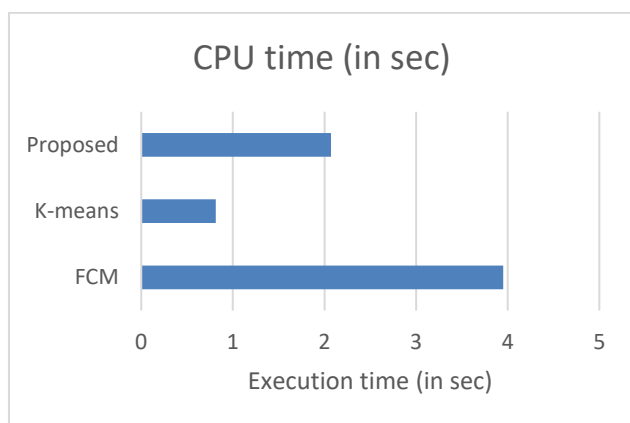


Fig. 4: CPU running time evaluation for multi-tissue detection performance

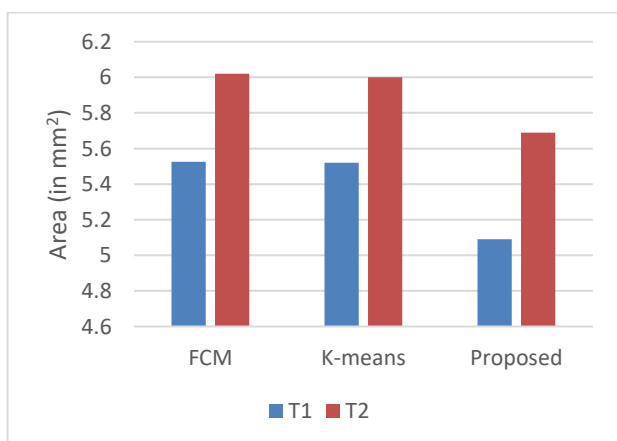
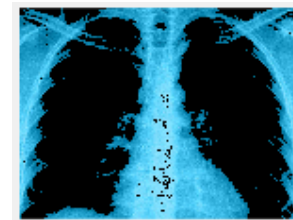
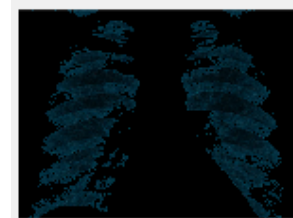


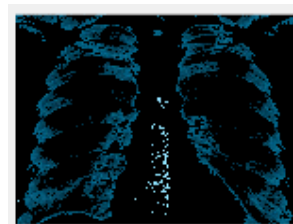
Fig. 5: Estimated area of segmented tissues



(a)



(b)



(c)

Fig 6: Input and Segmented Images

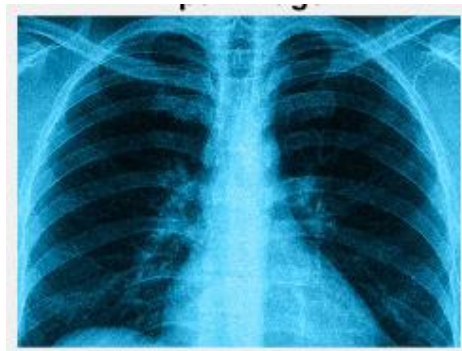


Fig 7: Input Image

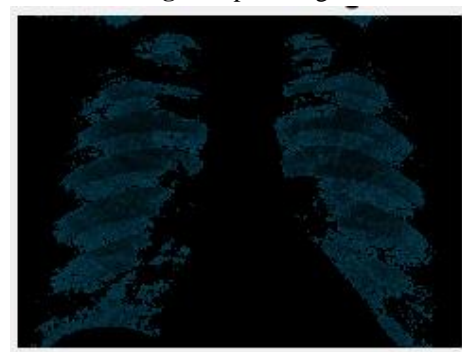


Fig 8: Segmented and Disease Effected region

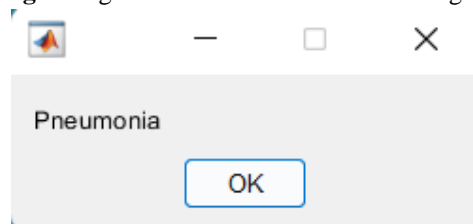


Fig 9: Disease Effect Identification

Table 1: Cpu Computation Time For Different Segmentation Techniques

S. No	Segmentation algorithm	Computation time (Seconds)			
		Sample Image 1	Sample Image 2	Sample Image 3	Sample Image 4
1	Fuzzy C Means	11.2633	3.4008	2.8348	2.9282
2	K-Means	0.4156	0.5248	0.4572	0.3532
3	Proposed	0.2596	0.2440	0.2812	0.2898

III. Conclusion

For the goal of categorising cases of pneumonia and Covid-19, we have presented a method in this body of work for identifying lung disease from chest X-ray pictures. The proposed framework incorporates methods from deep learning, machine learning and soft computing. While other approaches for identifying Covid-19 use image enhancement, ROI-based feature extraction, and normalisation, our model takes into consideration just the lung-specific characteristics. It is significantly distinct, in other words, in order to maximise computer efficiency and optimise classification performance.

References

[1] Bharati S, Podder P, Mondal R, Mahmood A, Raihan-Al-Masud M. Comparative performance analysis of different classification algorithm for the purpose of prediction of lung cancer. *Advances in intelligent systems and computing*, vol. 941. Springer; 2020. p. 447–57. https://doi.org/10.1007/978-3-030-16660-1_44.

[2] Coudray N, Ocampo PS, Sakellaropoulos T, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med* 2018;24:1559–67. <https://doi.org/10.1038/s41591-018-0177-5>.

[3] Mondal MRH, Bharati S, Podder P, Podder P. "Data analytics for novel coronavirus disease", *informatics in medicine unlocked*, 20. Elsevier; 2020. p. 100374. <https://doi.org/10.1016/j.imu.2020.100374>.

[4] Kuan K, Ravaut M, Manek G, Chen H, Lin J, Nazir B, Chen C, Howe TC, Zeng Z, Chandrasekhar V. Deep learning for lung cancer detection: tackling the Kaggle data science bowl 2017 challenge. <https://arxiv.org/abs/1705.09435>; 2017.

[5] Sun W, Zheng B, Qian W. Automatic feature learning using multichannel ROI based on deep structured algorithms for computerized lung cancer diagnosis. *Comput Biol Med* 2017;89:530–9.

[6] Song Q, Zhao L, Luo X, Dou X. Using deep learning for classification of lung nodules on computed tomography images. *Journal of healthcare engineering* 2017: 8314740. <https://doi.org/10.1155/2017/8314740>.

[7] Sun W, Zheng B, Qian W. Computer aided lung cancer diagnosis with deep learning algorithms. In: *Proc SPIE. Medical Imaging*, 2016, 9785. Computer-Aided Diagnosis; 2016. 97850Z. <https://doi.org/10.1117/12.2216307>.

[8] NIH sample Chest X-rays dataset. <https://www.kaggle.com/nih-chest-xrays/sample>. [Accessed 28 June 2020].

[9] Abbas A, Abdelsamea MM, Gaber MM (2020) Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. *Appl Intell*. <https://doi.org/10.1007/s10489-020-01829-7>

[10] Abiyev RH, Maaitah MKS (2018) Deep convolutional neural networks for chest diseases detection. *J Healthc Eng*. <https://doi.org/10.1155/2018/4168538>

[11] Angeline R, Mrithika M, Raman A, Warriar P (2020) Pneumonia detection and classification using chest X-ray images with convolutional neural network. In: Smys S, Iliyasa AM, Bestak R, Shi F (eds) *New trends in computational vision and bio-inspired computing*. ICCVIC. Springer, Cham

[12] Apostolopoulos ID, Mpesiana A (2020) Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. *Phys Eng Sci Med* 43:635–640. <https://doi.org/10.1007/s13246-020-00865-4>

[13] Asuntha A, Srinivasan A (2020) Deep learning for lung cancer detection and classification. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-019-08394-3>

[14] Avni U, Greenspan H, Konen E, Sharon M, Goldberger J (2011) X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words. *IEEE Trans Med Imaging* 30(3):733–46. <https://doi.org/10.1109/TMI.2010.2095026> (**Epub 2010 Nov 29. PMID: 2111876**)

[15] Bentivegna E, Luciani M, Spuntarelli V, Speranza ML, Guerritore L, Sentimentale A, Martelletti P (2020) Extremely severe case of COVID-19 pneumonia recovered despite bad prognostic indicators: a didactic report. *SN Compr Clin Med* 2:1204–1207. <https://doi.org/10.1007/s42399-020-00383-0>

[16] Butt C, Gill J, Chun D, Babu BA (2020) Deep learning system to screen coronavirus disease 2019 pneumonia. *Appl Intell*. <https://doi.org/10.1007/s10489-020-01714-3>.

[17] Chen X, Laurent S, Onur OA, Kleineberg NN, Fink GR, Schweitzer F, Warnke C (2021) A systematic review of neurological symptoms and complications of COVID-19. *J Neurol*. <https://doi.org/10.1007/s00415-020-10067-3>

[18] Dansana D, Kumar R, Bhattacharjee A, Hemanth DJ, Gupta D, Khanna A, Castillo A (2020) Early diagnosis of COVID-19-affected patients based on X-ray and computed tomography images using deep learning algorithm. *Soft Comput*. <https://doi.org/10.1007/s00500-020-05275-y>