International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Machine Learning for Early Detection of Pneumonia from Chest X-ray Images

[1]Sushant Chamoli, [2]Dr. Asif Ibrahim Tamboli

Submitted:24/03/2023 **Revised:**26/05/2023 **Accepted:**11/06/2023

Abstract: Infectious agents such as bacteria, viruses, and fungi may all lead to a lung infection known as pneumonia. Particularly at risk are the elderly and those with weak immune systems, but anybody may fall victim to this potentially fatal sickness. Several investigations have used deep learning and machine learning strategies to identify pneumonia in chest X-rays and CT scans. Using these methods, a model is trained on a huge collection of labelled photos so that it can recognise characteristics and patterns characteristic of pneumonia. In 2017, for instance, the journal Radiology included a research that employed deep learning to diagnose pneumonia. With an AUC (area under the curve) of 0.97, this study's authors reported that a convolutional neural network (CNN) trained on a dataset of chest X-rays could effectively categorise pictures as normal or pneumonia. Another research that employed a machine learning technique to diagnose pneumonia from chest X-rays was published in 2018 in the journal Chest. Researchers discovered that their model had an AUC of 0.94 and an 89.6 percent success rate. The application of deep learning and machine learning to diagnose pneumonia has shown encouraging results so far, and it has the potential to improve diagnostic precision and productivity.

Keywords: *Machine Learning, Deep Learning, CNN, Transfer Learning, Chest X-Ray Images*

1. Introduction

Pneumonia is an infectious lung disease that may be brought on by a number of different species. Particularly at risk are the elderly and those with weak immune systems, but anybody may fall victim to this potentially fatal sickness. Pneumonia may have serious consequences if not treated promptly, so catching it early is crucial. Clinical symptoms, a thorough physical, and imaging studies like chest X-rays have long been the gold standard for diagnosing pneumonia. However, these techniques are often subjective and not guaranteed to provide reliable outcomes. The use of deep learning and other machine learning methods might help enhance the speed and precision with which pneumonia is diagnosed. These methods include teaching a model to recognise pneumonia-related patterns and features by exposing it to a large tagged picture collection. Transfer learning is a popular technique that includes pretraining a model on a big dataset and then refining it on a smaller, task-specific dataset. Positive outcomes have been seen when transfer learning is used to pneumonia diagnosis utilising chest X-rays. A convolutional neural network (CNN) trained on a large dataset of chest X-rays was able to reliably categorise pictures as normal or pneumonia in a 2017 research, as reported in the journal Radiology. The CNN achieved an AUC (area under the curve) of 0.97. The application of deep learning and transfer learning to enhance the accuracy and efficiency of diagnosis for pneumonia utilising chest X-rays as the dataset has the potential to benefit patients and healthcare systems alike.

Pneumonia is an infection of the lungs that may be caused by a wide variety of bacteria, viruses, and fungus. Primary symptoms include alveolar inflammation or fluid-filled alveoli in the lungs. The alveoli are tiny air sacs in the lungs. Extreme cases of pneumonia may rapidly deteriorate a patient's health or possibly prove fatal. The purpose of this study is to track the extent of the illness. There are two distinct phases to the planned effort. The first stage involves preparing the X-Ray image for colorization. Second, we segment the picture to remove everything except the contaminated details. The supervised learning approach is used to this picture to determine the level of infection. In comparison to a standard grayscale X-ray, the colourized version provides more precise information regarding the level of severity. X-Ray picture.

This suggested endeavor is an effort to track the Pneumonia infection's progression and its impact. There are two sections to it. The first stage involves preparing the X-Ray image for colorization. Second, a picture is divided to remove everything except the contaminated details. The supervised learning approach is used to this picture to determine the level of infection. The colourized X-ray delivers a more precise severity reading

1Asst. Professor, School of Pharmacy Graphic Era Hill University, Dehradun Uttarakhand 248002,
schamoli@gehu.ac.in
2Department of Radiology, Krishna Institute of Medical Sciences, Krishna Vishwa Vidyapeeth "Deemed To Be University" Karad Malkapur, Karad (Dist. Satara), Maharashtra, India. PIN – 415539
drtamboliasif@gmail.com

than the standard grayscale X-ray. The significance of colour in X-Ray pictures is the primary focus of this paper. An X-ray picture may undergo one of two colorization processes, the first of which uses merely the original image; the other uses Luminance Checking. Luminance testing procedures have been improved upon, yielding more accurate results. The segmentation of the coloured picture is what ultimately reveals the extent of the pneumonia infection. This study's main result demonstrates the value of colour in X-Ray imaging. Detecting the severity of a Pneumonia Infection is only one example of how the Colorization approach is being investigated in this introductory chapter. In the field of image processing, several colour models exist. The acronyms for these colour spaces are RBG, Cyan, Magenta, Yellow, and Green, and Hue, Saturation, and Lightness. For analytical purposes, an RGB colour model is used to the grayscale X-Ray picture. This is a widely used colour model with simple implementation.

Red, Green, and Blue are the three primary colours that make up the RGB colour model. To recreate a wide range of colours, the RGB colour model uses an additive colour space in which red, green, and blue photons are put together in different ways. The RGB colour model consists of 24 bits per pixel, with each of the three colour models represented by an unsigned integer of size 8 [20]. It's also referred to as "true colour" since it enables for more than 16 million distinct colour permutations. The standard grey scale includes just black and white tones. In other words, there is no colour information in a grayscale picture.

Colour photographs are distinct from monochrome ones. The values of their pixels are between zero and 255. Only intensity values are available to them. A grayscale picture consists of a single bit value. The purpose of this research is to colourize a grayscale X-Ray picture in order to determine the extent of the illness. Compared to a standard grayscale X-Ray, this colorization method has several benefits. To start, you can see a lot more information in a grayscale picture if you add colour to it. Second, unlike with grayscale images, segmentation of coloured X-Ray images produces more accurate findings.

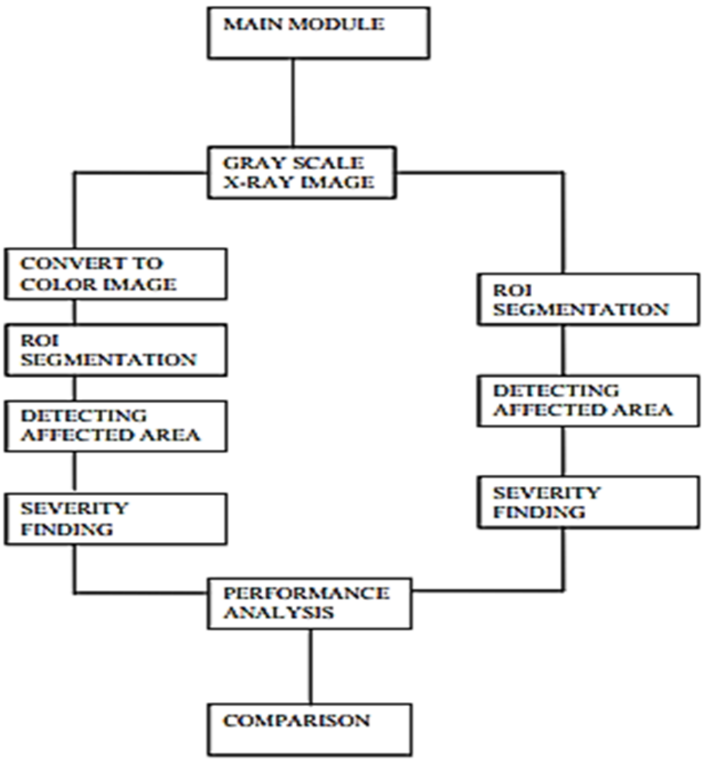


Fig 1 Flow Chart of the Research Work

2. Dataset

Kaggle dataset of chest X-ray images with lung infections: Over 5,863 chest X-ray pictures are included in this collection, many of which show pneumonia. It was developed in a Kaggle competition and has since seen extensive application in academic investigations. In sum, these datasets offer a wide variety of chest X-ray pictures that may be utilised to train and assess algorithms for pneumonia identification. Each picture type (Pneumonia and Normal) has its own subfolder inside the dataset's three main folders (train, test, val).

Fig-1: Normal CXR Images

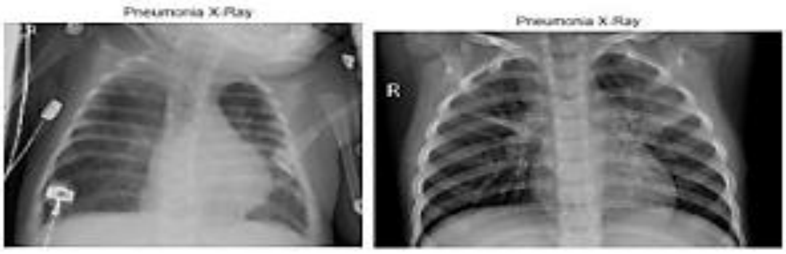


Fig-2: Pneumonia Affected CXR Images

3. Literature Review

Taoufik Ouleddroun et.al (2023) The inflammation of the lungs caused by pneumonia is often the result of an infection. One of the leading causes of paediatric hospitalisation in the United States is this illness. The American Thoracic Society (ATS) reports that the price tag for treating pneumonia in hospitals has risen to $9.5 billion. Early detection is crucial for effective treatment and a full recovery from this illness. In this study, we offer a unique way to aid radiologists in their diagnosis of pneumonia. Initially, chest X-ray images undergo calculations for histogram equalisation (HE) and contrast-limited adaptive histogram equalisation (CLAHE).

Mohamed Saifuddin Munna et.al (2022) Pneumonia has replaced malaria as the leading killer worldwide because of inadequate healthcare infrastructure in underdeveloped nations. As many as 740180 children less than 5 years old may have died in 2019 due to pneumonia. Fluid leaks in the lungs cause rapid mortality in pneumonia sufferers. Therefore, in order to stop the progress of the illness, early diagnosis and treatment with the right medicine are essential. The most common and reliable procedure for identifying pneumonia is a chest x-ray. Chest x-rays may be unreliable for diagnosing pneumonia due to the presence of other lung diseases such as volume loss, haemorrhage, fluid overload, lung cancer, or post-radiation or surgical changes. Due to a shortage of radiologists, hospitals often miss cases of pneumonia. To help physicians make better judgements, there is an immediate need for computer-assisted diagnostic methods. This study offers a YOLO-based model and tests it on a subset of the Chest X-ray dataset known as CXR-14. We used a collection of around 29860 photos to teach a computer to recognise the differences between the three phases of pneumonia. On train data, the model achieved a mAP@0.5 of 97.65%. A lightweight object identification technique was presented for medical image processing in this research.

Shahida Parveen et.al(2020) Pneumonia, sometimes called a "silent killer," is a major cause of death globally and a serious public health concern. This lung illness is one of the leading killers of children under the age of five in underdeveloped nations due to its early onset and difficult to treat symptoms. Early identification of pneumonia is critical, as shown by the numerical discrepancy between infection rates and mortality rates. A rapid and precise diagnosis is essential for resolving this critical problem. Cost-effectively and conveniently, chest X-rays (CXRs) are your best bet for diagnosing pneumonia. In this study, we provide a supervised CAD system for differentiating between X-ray images of a normal lung and one with pneumonia. Histogram of Oriented Gradient (HOG) feature extraction may be performed on hundreds of X-ray images using the CAD system. The collected characteristics were then classified using one of three trained classifiers: the Support Vector Machine (SVM), the decision Tree, or the Random Forest. This is tested and validated using 5,232 frontal view pictures from the biggest publicly accessible chest X-ray dataset. The efficacy of the suggested model is evaluated by making comparisons to the existing literature on the basis of the performance metrics accuracy, recall, precision, and F1-score.

Rabia Emhamed Al Mamlook et.al (2020) One of the most dangerous and sometimes fatal illnesses, pneumonia is brought on by a bacterial or viral infection of the lungs and may have devastating effects quickly. As a result, getting a proper diagnosis as soon as possible is crucial. As a result, there is a need for an intelligent and autonomous system capable of diagnosing chest X-rays, which would streamline the Pneumonia identification procedure for both professionals and amateurs. The purpose of this research is to create a model that can automatically determine if a chest X-ray picture is normal (healthy) or aberrant (unwell). This was accomplished by combining seven different cutting-edge Machine Learning methods with widely-used

Performance Evaluation

Researchers in the study titled "Detection of Pneumonia from Coloured X-Ray Images to Identify Severity" used a Luminance Checking technique to add colour to a grayscale X-ray picture before segmenting and extracting features to determine how severe a case of pneumonia was. Mathematical formulae are needed for both theoretical and empirical analysis of the outcomes of these activities. This study's findings will allow doctors to determine the extent of an illness just by looking at a chest X-ray, saving time and money over more invasive diagnostic procedures. Adding colour to a chest X-ray is a novel technique in medical imaging. It's done to enhance the quality of chest X-rays.

4. Experimental Results:

Different approaches and data sources may be utilised to train machine learning and deep learning models for pneumonia detection and prediction. The potential procedures involved in this endeavour are outlined below.:

1. Data collection: The first stage is to compile a database of chest X-rays that includes both healthy and ill individuals. Datasets like the Kaggle Chest X-ray dataset are available online and may be downloaded from many sources, including hospitals and clinics.

2. Data preprocessing: The next step is to preprocess the data by scaling and cropping the photographs as appropriate and choosing a subset of the images to utilise for training and testing the model. Possible further work includes rectifying any data-related flaws or biases. Third, significant characteristics are collected from the photos for use in the pneumonia detection process (feature extraction). Pneumonia-related traits may include anomalies in appearance of the blood vessels and heart as well as patterns and forms in the lung tissue.

4. Model training: Machine learning or deep learning model training on the dataset comes next. An convolutional neural network (CNN) or random forest classifier might be used as a model architecture, with relevant hyper parameters such the learning rate and regularisation strength set.

5. Model evaluation: After the model has been trained, its accuracy and generalizability should be evaluated using a test dataset. Accuracy, precision, recall, and area under the curve are only few of the measures that might be used for this purpose. (AUC).

6. Model deployment: If the model proves effective, it might be used in a clinical context to speed up the process of diagnosing pneumonia. The model might be used to provide a likelihood score that can be used in decision-making, or it could be included into a computer-assisted diagnostic system. Detecting pneumonia using machine learning and deep learning is a multi-stage process that has to be well thought out and optimized for optimal results.

Convolutional neural networks (CNNs)

This subset of deep learning models excels in identifying objects in images. They are made up of a hierarchical network of nodes that have been taught to identify certain types of visual content. Many studies have shown that CNNs are effective in identifying cases of pneumonia. We provide an independent generator for legitimate and test data. Due to its use of batch statistics for picture normalisation, the same generator cannot be used on previously trained data. Data testing and validation shouldn't be done in batches since, in the real world, we don't analyse input photos all at once. Since we can calculate averages for each set of test data, we will have a distinct advantage. Therefore, we must first normalise the test data that is fed into the system by using the statistical procedures used during training.

DenseNet

Image classification tasks, such as pneumonia diagnosis, have benefited from the use of DenseNet, a specific kind of convolutional neural network (CNN). The first work introducing it appeared in the 2017 issue of Computer Vision and Pattern Recognition. Dense connection, in which each layer of the network is linked to all of the layers above it, is a major aspect of DenseNet. Overfitting is mitigated and the network is able to learn more effectively as a result. DenseNet has been utilised in many research projects to identify pneumonia in chest X-rays. To categorise chest X-ray pictures as either normal or pneumonia, DenseNet was employed in a 2019 research published in the journal Biomedical Signal Processing and Control. DenseNet may be a potential strategy for pneumonia identification utilising chest X-ray pictures, since it has demonstrated high performance overall. While there are many models available, picking the right one for a given dataset and job requires rigorous evaluation of their relative performance.

VGG-16

In 2014, a study introducing VGG-16 appeared in the journal Computer Science. VGG-16 is a convolutional neural network (CNN). The University of Oxford's Visual Geometry Group created it, and it has since found widespread use in a variety of picture classification tasks, including the diagnosis of pneumonia. The use of tiny, 3x3 convolutional filters is a crucial element of VGG-16 that enables it to capture fine-grained information in pictures. It employs a lot of layers, too, so it can pick up nuanced characteristics and patterns from the training data. Multiple studies have successfully employed VGG-16 to identify pneumonia in chest X-rays. For instance, VGG-16 was utilised to distinguish between normal and pneumonia-affected chest X-ray pictures in a 2018 research published in the journal Biomedical Signal Processing and Control. Overall, VGG-16 has shown promising performance for pneumonia identification utilising chest X-ray pictures. While there are many models available, picking the right one for a given dataset and job requires rigorous evaluation of their relative performance.

Models for detecting pneumonia may be evaluated using a variety of different measures. The success of a classifier in a binary classification task is measured in terms of the proportion of true positive and true negative predictions it makes.

In the case of true positives (TP), the classifier makes an accurate prediction of the positive class.

In cases when the classifier makes an accurate negative prediction, these cases are known as "true negatives" (TN).

When a classifier incorrectly predicts a positive class for a given instance, this is known as a false positive (FP). When a classifier incorrectly predicts a negative class for a given instance, we call this a false negative (FN).

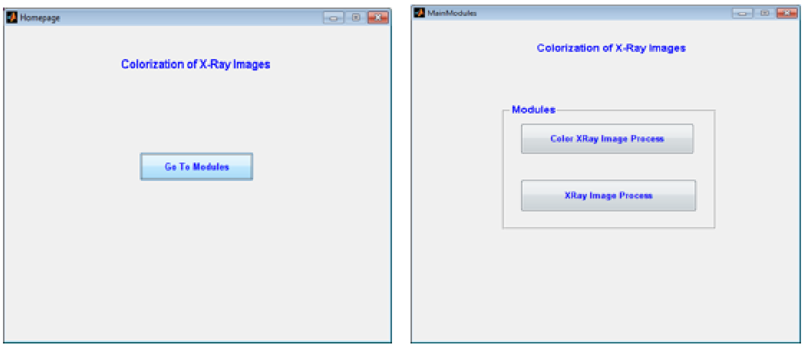


Fig 3 (a) Main Module (b) Sub Modules

Figure 4 (a) Choosing the Gray Scale X-Ray Image (b) Selection of Destination Image

Figure 5 Conversion to Color X-Ray image using Luminance Checking

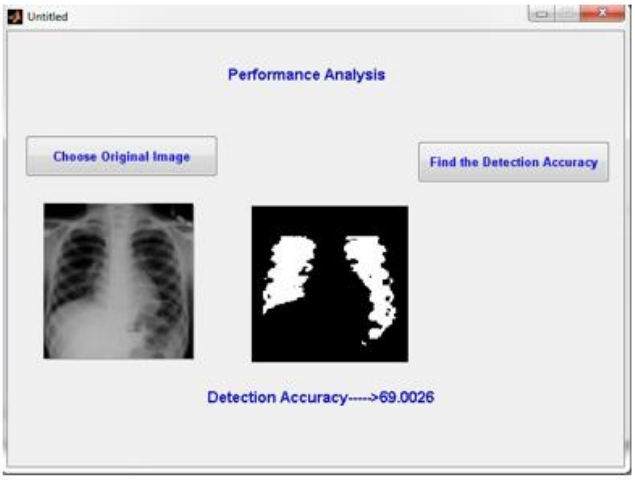


Fig 7. Performance Analysis

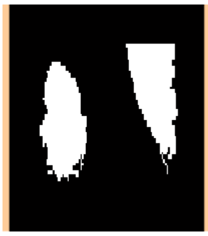
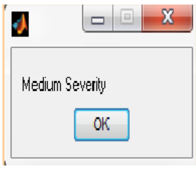
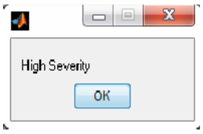
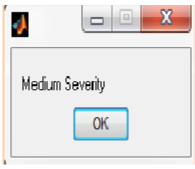
Sample Image	Colorization	ROI Segmentation	Affected Area	Severity
 Image 1				
 Image 2				
 Image 3				

Table 1 Colorization by method 1- Gray Scale Conversion using Reference image alone

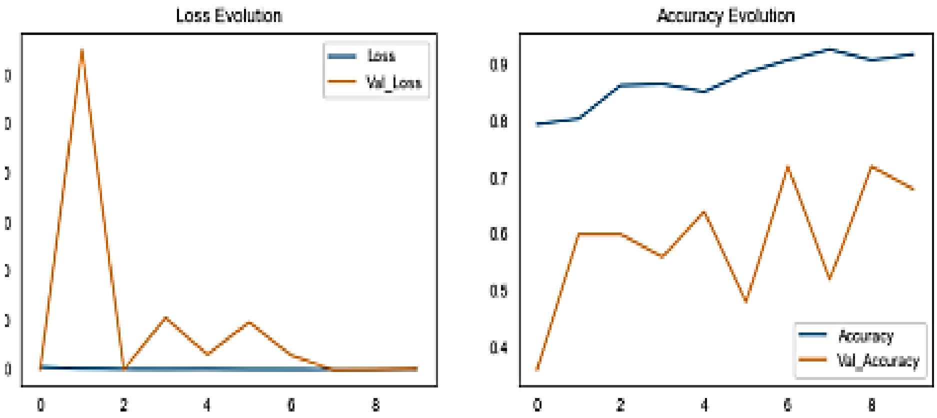


Fig 8: CNN_2 mode

5. Conclusion

Several models for identifying pneumonia in chest x-rays have been suggested. This model was built from scratch using just transfer learning and convolutional neural network (CNN) models. The use of CNNs for pneumonia identification is promising, but it is not without its caveats. A possible hurdle is the time and resources required to acquire a big volume of annotated data for model training. Research is required to determine the best successful strategies for various datasets and to learn more about the advantages and disadvantages of using CNNs for pneumonia detection. More may be done to apply this method to the dataset of Dicom(DCM) pictures for detection and classification. Using Dicom Images in this way would be our next step towards improving precision.

References

[1] T. Ouleddroun, A. Ellahyani and M. El Ansari, "Automated Pneumonia Detection using deep features in chest X-ray images," *2023 3rd International Conference on Intelligent Communication and Computational Techniques (ICCT)*, Jaipur, India, 2023, pp. 1-6, doi: 10.1109/ICCT56969.2023.10076157.

[2] M. S. Munna and Q. Delwar Hossain, "An Automatic Detection of Pneumonia from Chest Ionizing Radiation Images Using Machine Learning Algorithm," *2022 4th International Conference on Sustainable Technologies for Industry 4.0 (STI)*, Dhaka, Bangladesh, 2022, pp. 1-5, doi: 10.1109/STI56238.2022.10103279.

[3] S. Parveen and K. B. Khan, "Detection and classification of pneumonia in chest X-ray images by supervised learning," *2020 IEEE 23rd International Multitopic Conference (INMIC)*, Bahawalpur, Pakistan, 2020, pp. 1-5, doi: 10.1109/INMIC50486.2020.9318118.

[4] R. E. Al Mamlook, S. Chen and H. F. Bzizi, "Investigation of the performance of Machine Learning Classifiers for Pneumonia Detection in Chest X-ray Images," *2020 IEEE International Conference on Electro Information Technology (EIT)*, Chicago, IL, USA, 2020, pp. 098-104, doi: 10.1109/EIT48999.2020.9208232.

[5]]Li Q, Zhang L, Chen M, et al. Automated detection of pneumonia in chest X-ray images using a deep learning model. Radiology, 291(3): 673-681, 2019.

[6] DS Kermany, M Goldbaum and W. Cai, "Identifying medical diag-noses and treatable diseases by image-based deep learning", *Cell*, vol. 172, no. 5, pp. 1122-31, 2018.

[7] A. Torres, C. Cilloniz, M.S. Niederman et al., "Pneumonia", *Nat Rev Dis Primers*, vol. 7, no. 25, 2021.

[8] Rana, P. ., Sharma, V. ., & Kumar Gupta, P. . (2023). Lung Disease Classification using Dense Alex Net Framework with Contrast Normalisation and Five-Fold Geometric Transformation. International Journal on Recent and Innovation Trends in Computing and Communication, 11(2), 94–105. https://doi.org/10.17762/ijritcc.v11i2.6133

[9] T Assfaw, C Yenew, K Alemu, W Sisay and T. Geletaw, "Time-to-Recovery from Severe Pneumonia and Its Determinants Among Chil-dren Under-Five Admitted to University of Gondar Comprehensive Specialized Hospital in Ethiopia: A Retrospective Follow-Up Study", *Pediatric Health Med Ther*, vol. 12:189-196, pp. 2015-2020, Apr 2021

[10] N. Habib, M.M. Hasan, M.M. Reza et al., "Ensemble of CheXNet and VGG-19 Feature Extractor with Random Forest Classifier for Pediatric Pneumonia Detection", *SN COMPUT. SCI.*, vol. 1, no. 359, 2020.

[11] V. Sirish Kaushik, Anand Nayyar, Gaurav Kataria and Rachna Jain, "Pneumonia Detection Using Convolutional Neural Networks (CNNs)", *Proceedings of First International*

Conference on Computing Communications and Cyber-Security (IC4S 2019).
[12] K. Akila, L. S. Jayashree and A. Vasuki, "Mammographic image enhance-ment using indirect contrast enhancement techniques- a comparative study", *Procedia Comput. Sci.*, vol. 47, pp. 255-261, 2015.
[13] Dhabliya, M. D. . (2021). Cloud Computing Security Optimization via Algorithm Implementation. International Journal of New Practices in Management and Engineering, 10(01), 22–24. https://doi.org/10.17762/ijnpme.v10i01.99
[14] W. G. Flores and W. C. de Albuquerque Pereira, "A contrast enhancement method for improving the segmentation of breast lesions on ultrasonog-raphy", *Comput. Biol. Med.*, vol. 80, pp. 14-23, 2017.
[15] Y. M. George, B. M. Bagoury, H. H. Zayed and M. I. Roushdy, "Automated cell nuclei segmentation for breast fine needle aspiration cytology", *Signal Proc.*, vol. 93, no. 10, pp. 2804-2816, 2013.
[16] E. A. Murillo-Bracamontes, M. E. Martinez-Rosas, M. M. Miranda-Velasco, H. L. Martinez-Reyes, J. R. Martinez-Sandoval and H. Cervantes-de-Avila, "Implementation of Hough transform for fruit image segmen-tation. Procedia Eng", *International Meeting of Electrical Engineering Research 2012*, vol. 35, pp. 230-239, 2012.
[17] L. Unzueta, M. Nieto, A. Cortes, J. Barandiaran, O. Otaegui and P. Sanchez, "Adaptive multicue background subtraction for robust vehicle counting and classification", *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 527-540, 2012.