JOURNAL OF
NATURAL SCIENCES AND MATHEMATICS RESEARCH

Available online at http://journal.walisongo.ac.id/index.php/jnsmr

Classification of Pneumonia in Thoracic X-Ray images based on texture characteristics using the MLP (Multi-Layer Perceptron) method

Latifatul Istianah[1], Heni Sumarti[1]
[1]Department of Physics, Faculty of Science and Technology, Universitas Islam Negeri Walisongo Semarang, Indonesia

Corresponding author: latifatulistianah@gmail.com
Received: 10 October 2020,
Revised: 24 October 2020,
Accepted: 23 December 2020.

Abstracts

One of the diseases that attack the lungs is pneumonia. This disease can attack someone with a weak immune system. Pneumonia is inflammation of the lungs that can be caused by pathogens, such as bacteria, viruses, and fungi. The purpose of this study was to classify fungal pneumonia, bacterial pneumonia, and lipoid pneumonia based on texture characteristics and the MLP method using machine learning WEKA. The method in this study has three stages including pre-processing, extraction of texture features consisting of Histogram and GLCM, and classification using the MLP (Multi Layer Perceptron) method. The results of the texture feature extraction showed that the three types of pneumonia were: lipoid pneumonia with brightness, sharp contrast & random distribution on correlation characteristics, bacterial pneumonia with high brightness, high contrast & random distribution on energy characteristics, and fungal pneumonia with brightness, sharp contrast, & random distribution of homogeneity attributes. The third similarity of pneumonia is in the gray level that collects each other in the middle. The method used in this study resulted in the same accuracy, sensitivity, and specificity, namely 100%. The image classification in this study shows the success of the texture features and the MLP method in classifying pneumonia images accurately so that they can be used as additional tools that make it easier for medical experts.

Keywords: pneumoniae, texture characteristics, MLP method, WEKA machine learning.

1. Introduction

The lungs are organs in human anatomy that have a function as a respiratory system, as a place for exchanging oxygen with carbon dioxide in the blood. Air pollution can cause lung problems. Inhaled polluted air contains many germs (Rahmadewi et al. 2013). Medical images have an important role in classifying a disease. X-rays (X-Rays) can be used to identify lung problems (Maysanjaya 2020).

One of the diseases that attack the lungs is pneumonia. This disease can attack a person with a weak immune system (Chouhan et al. 2020). Pneumonia causes inflammation of the lungs, especially the air sacs filled with fluid or pus (Sharma et al. 2020). Pneumonia is inflammation of the lungs that can be caused by pathogens, such as bacteria, viruses, and fungi (Gilani et al. 2012). Pneumonia is a lung disease in which there is an infection that occurs in the air sacs in the lungs caused by various microorganisms including fungal bacteria, mycobacteria and viral pneumonia classified as community-controlled pneumonia, hospital-acquired pneumonia (nosocomial), pneumonia in the host. immunocompromised, and aspiration pneumonia (FAP Scholastica, 2018).

Lipoid pneumonia is an unusual entity without special imaging such as CT X-ray (Roberts et al. 2020). A chest X-ray image can be used to see the condition of the patient's chest area. The results of the chest x-ray show the overall condition of the respiratory tract such as the chest, lungs, heart, and trachea. In image processing, examination using X-rays can describe the condition of the lungs affected by pneumonia (Wati, Irsyad, and Rivan 2020). A patient's chest X-ray is often used as an early indicator of pneumonia (Noor et al. 2010).

The classification method is the most commonly used technique in medical imaging (G. Litjens et al. 2017). Previous research conducted by Ramdhan (2016) discussed the classification of lung X-rays using histogram feature extraction and the Naive Bayes method to distinguish normal lungs, effusions, and lung cancer with an accuracy of 70%. Therefore, we used texture characteristics to differentiate fungal pneumonia, bacterial pneumonia, and lipoid pneumonia using the MLP (Multi Layer Perceptron) method. This research is expected to be a second opinion for medical experts in classifying cases of pneumonia using chest X-ray images.

2. Experiments Procedure

The images used in this study are 9 images, namely: 3 images of fungal pneumonia, 3 images of bacterial pneumonia, and 3 images of lipoid pneumonia. The pneumonia image database was taken from the ghitub site which was then processed by image processing. The method used is the texture feature method. Texture is an intrinsic feature of an image related to the level of roughness, granularity and regularity of the structural arrangement of pixels which is inherent in objects that are visible from the face and contains information about surface structure (Alfiani.dkk, 2011).

Pneumonia image processing aims to improve the quality contained in the image so that it can facilitate the processing of information (Ramdhan 2016). In this study there are three stages of pre-processing, namely cropping, texture feature extraction, and classification. At the cropping stage, unused images will be discarded so as to obtain the image information needed in the next process. This process was carried out in Matlab R2016a. Furthermore, the resizing stage was carried out which aims to equalize the images of fungal pneumonia, bacterial pneumonia, and lipoid pneumonia with a resizing result of 100 × 1100 pixels (Wati, Irsyad, and Rivan 2020). Image processing is done by converting an RGB image into a grayscale image (Ramdhan 2016). RBG image is a type of image that represents color in components R (Red), G (Green), and B (Blue) (Nugroho, 2015). While the grayscale image has a single value for each pixel, making it easier for further image analysis (Wahyuni 2015).

Image feature extraction is a step that extracts the characteristics of objects contained in the pneumonia x-ray image. In this research, extraction method based on Histogram and

GLCM (Gray-Level Co-occurrence Matrix). Histogram parameters consist of mean, standard deviation, variance, entropy, skewness, and kurtosis (Agussationo, Soesanti, and Najib 2018).

1) Standard deviation shows the root of a variation of image elements

$$\sigma = \sqrt{\sum_n (f_n - \mu)^2 p(f_n)} \quad (1)$$

2) Variant shows variations in image elements. f_n is the gray intensity and p(f_n) is the histogram value.

$$\sigma^2 = \sum_n (f_n - \mu)^2 p(f_n) \quad (2)$$

3) Entropy is a quantity that shows the complexity of the image.

$$H = -\sum_{i=0}^{L-1} p(i) log_2(p(i)) \quad (3)$$

4) Slope or Sweknes shows the level of relative slope/asymmetry of the histogram curve.

$$\alpha_3 = \frac{1}{\partial^3} \sum_n (f_n - \mu)^3 p(f_n) \quad (4)$$

5) Kurtosis is a value that shows the sharpness of the histogram curve.

$$\sigma_4 = \frac{1}{\sigma^4} \sum_n (f_n - \pi)^4 p(f_n) \quad (5)$$

Extraction GLCM (Gray-Level Co-occurrence Matrix) is a matrix that describes the frequency of occurrence of pairs of two pixels with a certain intensity in the distance d and orientation direction with a certain angle in the image. The GLCM feature extraction is carried out from an angle of 0^° and d = 1. The GLCM extraction method shows the co-occurrence of image data which is then used to determine the characteristics of an image that functions as a matrix. This extraction consists of energy, contrast, correlation, and homogeneity (Adi and Widodo 2016).

1) Energy is a measure of the homogeneity of the image.

$$Energy = \sum_i \sum_j p^2(i,j) \quad (6)$$

2) Contrast is the degree of gray between regions of the image.

$$Contras = \sum_i \sum_j (i-j)^2 p^2(i,j) \quad (7)$$

3) Correlation is a measure of the linear dependence of the gray degree of the image.

$$Correlastion = \frac{1}{\sigma_x \sigma_y} \sum_{i=1} \sum_{j=1} (i - \mu_x)(j - \mu_y) p(i,j) \quad (8)$$

4) Homogeneity is the uniformity of variations in the gray level of the image.

$$Homogenity = \sum_i \sum_j \frac{p(i,j)}{1+|i-j|} \quad (9)$$

Information Gain is information on the selection of feature attribute levels that are often used in image analysis. Selection of feature attributes is useful for obtaining attributes that affect the classification process. The selection process is carried out using Machine Learning WEKA (Ermawati 2020). WEKA, officially called the Waikato Environment for Knowledge Learning, is a computer program developed at the University of Waikato in New Zealand with the aim of identifying information from raw data collected from the agricultural domain. WEKA support may differ from standard data mining tasks such as data preprocessing, classification, clustering, regression, visualization and feature selection (Subbulakshmi, Deepa, and Malathi 2012)

The classification process uses the MLP method. The MLP (Multi Layer Perceptron) method is a method that uses one or more single screens (Rosita et al. n.d.). MLP is usually referred to as Artificial Neural Network (ANN) because MLP works as a human organ micro system. The advantage of MLP is that the algorithm applied is quite easy and has good accuracy (Ermawati 2020).

The confusion matrix analysis in WEKA consists of True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN) (Ermawati 2020). TP is positive data that is classified as positive in the system. FP is positive data that is classified as negative. TN is negative data classified as negative in the system and FN is negative data classified as positive in the system. Based on the results of the confusion matrix, the measurement index can be calculated using the following formula (Witten, et al., 2011; Khusna, 2016):

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \times 100\% \qquad (10)$$

$$Sensitifity = \frac{TP}{TP+FN} \times 100\% \qquad (11)$$

$$Specificity = \frac{TN}{TN+FP} \times 100\% \qquad (12)$$

3. Result and Discussion

X-rays are generally used to diagnose pneumonia early. Chest X-rays can be used to visualize and measure the structural and functional consequences of thoracic disease (Zhang et al. 2021). The results of pre-processing X-ray images of fungal pneumonia, bacterial pneumonia, and lipoid pneumonia are shown in Figure 1. The pre-processing that has been done has succeeded in removing the parts that are not needed in the process of increasing the image and uniformity of pixel size.

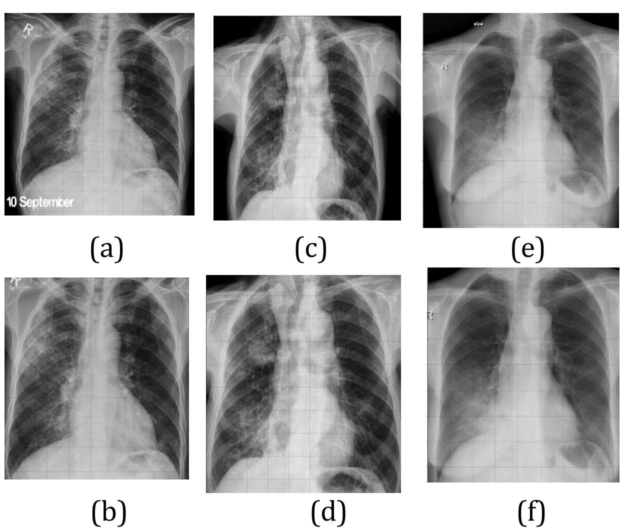

Figure 1. X-ray images of pneumonia in patients (a) Fungal pneumonia before pre-processing, (b) Fungal pneumonia after pre-processing, (c) Bacterial pneumonia before pre-processing, (d) Bacterial pneumonia after pre-processing, (e) Lipoid pneumonia before pre-processing, and (e) Lipoid pneumonia after pre-processing.

The histogram in Figure 2 shows the distribution of all gray levels unevenly and clustered in the middle. Of the three types of pneumonia that have been equalized, it is seen that all three have blurry contrasts. In addition, this first-order method changes the value of the gray level of certain pixels regardless of the location of the image (Rahmadewi et al. 2013).

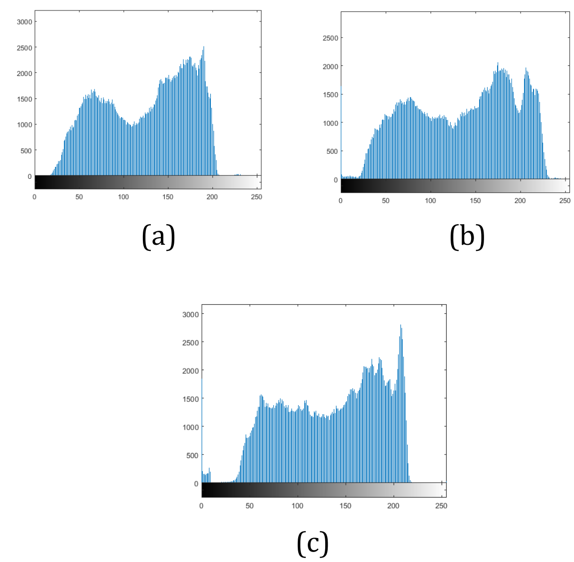

Figure 2. Histogram of X-ray images of pneumonia in (a) Fungal pneumonia, (b) Bacterial pneumonia, and (c) Lipoid pneumonia.

The next process is image selection and extraction. Image selection is done to determine the attributes that affect the classification. Based on the selection results, there are 10 characteristics. Texture feature values greater than zero are feature attributes that have a large effect. Table 1 shows the average gain ratio value, the average attribute value and the attribute average difference after extraction. The GLCM parameters consist of energy, contrast, correlation, and homogeneity, while the histogram parameters include the mean, standard deviation, variance, entropy, skewness, and kurtosis.

Table 1. Average Gain Ratio, attribute average, and attribute average difference

No	Feature Atributte	Gain Ratio	Average fungal pneumonia	Average bacteria pneumonia	Average lipoid pneumonia
1	Mean	0	124.8036± 3.53	124.8036± 18.6789	144.7722±14.1199
2	Contras	0	0.0939±0.0215	0.0827± 0.009	0.105±0.0103
3	Skewness	0	0.1106±0.2737	-0.2212±0.1564	-0.3594± 0.1034
4	Curtosis	0	2.052±0.2645	2.088±0.4857	2.052±0.1764
5	Standard Deviasi	0	44.8971±6.3494	46.3937±9.4058	51.6317±5.4987
6	Varian	0	8166303±11548896	1082919.5±371195.6	-479138.5±371195.6
7	Entropi	0	7.312± 0.1863	7.3522±0.274	7.4124±0.2255
8	Correlation	0.766	0.9847±0.0064	0.9896±0.0023	0.9896±0.0012
9	Energy	0.807	0.2458±0.0638	0.2978±0.0116	0.2884±0.0067
10	Homogenity	0.807	0.9633±0.0179	0.9801±0.0018	0.9749±0.0066

The results of testing 9 x-ray images of pneumonia in table 1 show the most influential feature attributes in the image processing classification process, namely correlation, energy, and homogeneity. Uniformity of correlation, uniformity of energy, and uniformity of homogeneity are used as classification references. Feature attribute ± standard deviation, where the standard deviation is the root of a variation of image elements (Agussationo, Soesanti, and Najib 2018).

In graph 3, there are 3 rank attribute traits that affect the classification process. Lines or linear tradelines on the chart cluster are used to see differences in the distribution of random texture feature data on fungal pneumonia, bacterial pneumonia, and lipoid pneumonia. The first attribute, namely correlation, shows that the contrast measure in patients with lipoid pneumonia is more contrasting than in patients with fungal pneumonia and bacterial pneumonia. The second attribute, namely energy, shows that the contrast measure in bacterial pneumonia patients is more contrasting than fungal pneumonia and lipoid pneumonia patients. And the third attribute, namely homogeneity, shows that the contrast size of fungal pneumonia patients is more

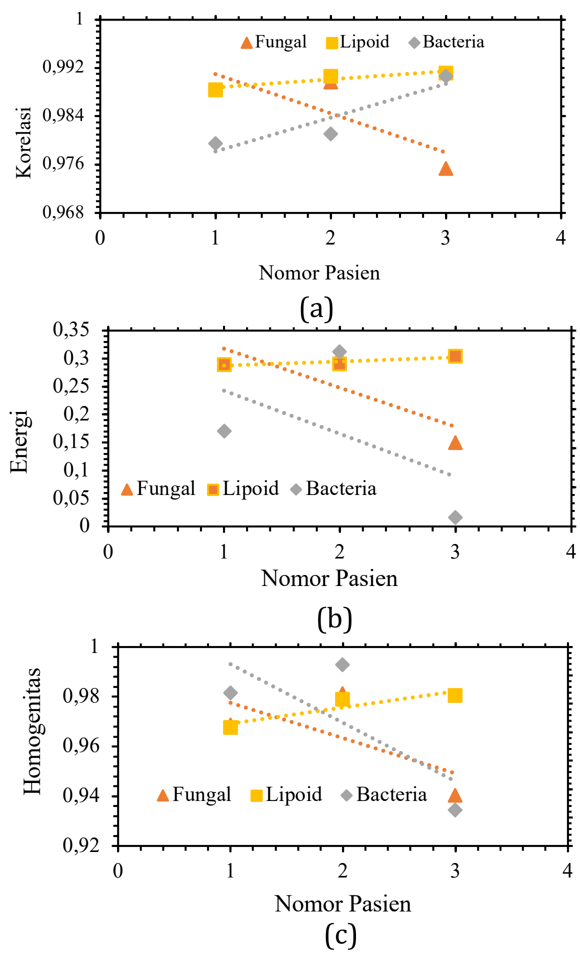

Figure 3. Cluster chart (a) Correlation, (b) Energy, and (c) Homogeneity

contrasting than bacterial pneumonia patients and lipoid pneumonia patients.

The next process is image selection and extraction. Image selection is done to determine the attributes that affect the classification. Based on the selection results, there are 10 characteristics. Texture feature values greater than zero are feature attributes that have a large effect. Table 1 shows the average gain ratio value, the average attribute value and the attribute average difference after extraction. The GLCM parameters consist of energy, contrast, correlation, and homogeneity, while the histogram parameters include the mean, standard deviation, variance, entropy, skewness, and kurtosis.

The results of the confused matrix are shown in table 2. The results of the Confused matrix are TP (True Positive) of 3, TN (True Negative) of 3, FP (False Negative) of 0, and FN (False Negative) of 0. TP is a pneumonia patient machine learning detected as pneumonia patients, TN was pneumonia negative patients detected as pneumonia negative patients, FP was pneumonia patients detected pneumonia negative patients, and FN was pneumonia negative patients but pneumonia patients were detected.

Table 2. Confused Matrix Results

TP	FP	TN	FN	Accu-racy	Sensi-tifity	Spesi-fisity
3	0	3	0	100%	100%	100%

Image classification using the MLP method provides an accuracy of 100%. When compared with previous research by Ramdhan (2016), the Naive Bayes method only produces an accuracy of 70%. The advantages of the MLP algorithm, which is easy to implement and good accuracy, can be used as a tool for medical experts in the classification process. The data presented in this study only amounted to 9 images. For further research, it will be improved again so that the grouping of images makes it easier to compare the results of image extraction with one another.

4. Conclusion

The classification of chest X-ray images in this study used 9 images including 3 fungal pneumonia images, 3 bacterial pneumonia images, and 3 lipoid pneumonia images based on texture characteristics and the MLP (Multi Layer Percetron) method. The MLP method produces 100% accuracy, sensitivity, and specificity. The results of the classification in this study indicate the success of texture features in classifying pneumonia images accurately so that they can become additional tools that make it easier for medical experts.

Acknowledgment

The authors thank the provider of the CXR pneumonia database, Github.

References

[1] Adi, Kusworo, and Edi Widodo. 2016. "Analisis Citra Ct Scan Kanker Paru Berdasarkan Ciri Tekstur Gray Level Co-Occurrence Matrix Dan Ciri Morfologi Menggunakan Jaringan Syaraf Tiruan Propagasi Balik." *Youngster Physics Journal* 5(4): 417–24.

[2] Alfiani,D.dkk. 2011. Klasifikasi Tekstur Parket Kayu dengan Menggunakan Metode Statistikal Grey Level Run Length Matrix. *Jurnal Teknologi Infonnasi Politeknik Telkom* 01 (01):09-15.

[3] Agussationo, Yudhi, Indah Soesanti, and Warsun Najib. 2018. "Klasifikasi Citra X-Ray Diagnosis Tuberkulosis Berbasis Fitur Statistis." *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)* 2(3): 736–45.

[4] Carreira, J. dan Sminchisesce, C.,2012, CPMC: Automatic Object Segmentation Using Constrained Parametric Min-Cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence.*

[5] Chouhan, Vikash et al. 2020. "A Novel Transfer Learning Based Approach for Pneumonia Detection in Chest X-Ray Images." *Applied Sciences (Switzerland)* 10(2).

[6] Ermawati, Eli. 2020. "Klasifikasi Nodul Payudara Berdasarkan Ciri Tekstur Pada Citra Ultrasonografi Menggunakan Scilab."

Universitas Islam Negeri Walisongo Semarang.

[7] F. A. P. Scholastica, Asuhan Keperawatan Pada Pasien Dengan Gangguan Sistem Pernapasan. Yogyakarta : Pustaka Baru Press, pp. 14–99, 2018.

[8] Gilani, Z.; Kwong, Y.D.; Levine, O.S.; Deloria-Knoll, M.; Scott, J.A.G.; O'Brien, K.L.; Feikin, D.R. A literature review and survey of childhood pneumonia etiology studies: 2000–2010. *Clin. Infect. Dis.* **2012**, *54* (Suppl. 2), S102–S108.

[9] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sanchez, ´ "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[10] Khusna, D.A. 2016. Klasifikasi Lesi Citra Ultrasonografi Payudara Berdasarkan Karakteristik Tepi. Tesis. Yogyakarta : Program Pascasarjana Universitas Gadjah Mada.

[11] Maysanjaya, I Md. Dendi. 2020. "Klasifikasi Pneumonia Pada Citra X-Rays Paru-Paru Dengan Convolutional Neural Network." *Jurnal Nasional Teknik Elektro dan Teknologi Informasi* 9(2): 190–95.

[12] Nugroho, A. 2015. *Klaaasifikasi Nodul Tiroid Berbasis Ciri Tekstur pada Citra Ultrasonografi.* Tesis. Yogyakarta : Program Pascasarjana Universitas Gadjah Mada Yogyakarta.

[13] Noor, Norliza Mohd, Omar Mohd Rijal, Ashari Yunus, and S. A.R. Abu-Bakar. 2010. "A Discrimination Method for the Detection of Pneumonia Using Chest Radiograph." *Computerized Medical Imaging and Graphics* 34(2): 160–66.

[14] Rahmadewi, Reni et al. 2013. "Metode Segmentasi Canny Pada Citra Rontgen Untuk Klasifikasi Penyakit Paru." *Jurnal Nasional Teknik Elektro* 1(2): 140–45. http://jti.respati.ac.id/index.php/jurnaljti/article/viewFile/1/1.

[15] Ramdhan, Alfiana. 2016. "Klasifikasi Citra Rontgen Paru-Paru Dengan Ekstraksi Fitur Histogram Dan Metode Naive Bayes Classifier, S1." 3(1): 1–41.

[16] Roberts, James, Yazeed Alwelaie, Gabriela Oprea-Ilies, and Eric Flenaugh. 2020. "Lipoid Pneumonia: A Rare Case." *Chest* 158(4): A2511. https://doi.org/10.1016/j.chest.2020.09.112.

[17] Rosita, Yesy Diah et al. "Penerapan Metode Multilayer Perceptron (Backpropagation) Dalam Penentuan Kesiapan Calon Transmigran."

[18] Sharma, Harsh, Jai Sethia Jain, Priti Bansal, and Sumit Gupta. 2020. "Feature Extraction and Classification of Chest X-Ray Images Using CNN to Detect Pneumonia." *Proceedings of the Confluence 2020 - 10th International Conference on Cloud Computing, Data Science and Engineering*: 227–31.

[19] Subbulakshmi, C. V., S. N. Deepa, and N. Malathi. 2012. "Comparative Analysis of XLMiner and WEKA for Pattern Classification." *Proceedings of 2012 IEEE International Conference on Advanced Communication Control and Computing Technologies, ICACCCT 2012* (978): 453–57.

[20] Wahyuni, Sri. 2015. "Penentuan Kondisi Tulang Femur Menggunakan Analisis Tekstur Pada Citra Digital." *Elkawnie* 1(2): 173–90.

[21] Wati, Risha Ambar, Hafiz Irsyad, and M Ezar Al Rivan. 2020. "Klasifikasi Pneumonia Menggunakan Metode Support Vector Machine." *Jurnal Algoritme* 1(1): 21–32.

[22] Witten, I.H., dkk. 2011. *Data Mining Practical Machine Learning Tools and Techniques*. Third Edition. USA: Morgan Kaufmann.

[23] Y. Dong, Y. Pan, J. Zhang, dan W. Xu, "Learning to Read Chest X-Ray Images from 16000+ Examples Using CNN," *2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, 2017, hal. 51-57.

[24] Zhang, Jianpeng et al. 2021. "Viral Pneumonia Screening on Chest X-Rays Using Confidence-Aware Anomaly Detection." *IEEE Transactions on Medical Imaging* 40(3): 879–90.