Kurtosis and entropy for the automatic detection of pneumonia

Curtosis y entropía para la detección automática de neumonía

Uriel Calderon-Uribe[1], Salvador Calderon-Uribe[2]
[1]Technological Institute of the South of Guanajuato, Uriangato, Guanajuato, Mexico
[2]Mechatronics / Engineering Faculty, Autonomous University of Queretaro, San Juan del Rio, Queretaro, Mexico
urielcal92@gmail.com[1], scalderon2@alumnos.uaq.mx[2]

Abstract

Pneumonia is a lung infection caused by a virus or bacteria. It is characterized by symptoms such as high fever, severe pain in the chest, cough, and expectoration, among others. Today, the main way to detect pneumonia is through X-ray imaging, where an expert diagnoses a patient based on features that can be seen on the images. However, this process is susceptible to errors that can lead the expert to make a false diagnosis. Computer-assisted techniques have proven to be a great tool to support the diagnosis of pneumonia. In this work, a classifier (based on the logistic regression method) was developed that allows the detection of pneumonia based on signal processing features (Kurtosis and entropy). This system generates an accuracy comparable to that generated by more robust algorithms such as neural networks.

Keywords: Image processing; computer-aided diagnosis; signal processing features; machine learning; classification system; signal analysis.

Introduction

Pneumonia is an acute respiratory infection that affects the lungs, caused by either bacteria or viruses. It is considered one of the main causes of mortality in children around the world and about two billion people suffer from this disease each year (Elshennawy, N. M., & Ibrahim, D. M. (2020), Jaiswal, A. K., Tiwari, P., Kumar, S., Gupta, D., Khanna, A., & Rodrigues, J. J. (2019)).

Currently, the most reliable method for diagnosing pneumonia is a chest X-ray. However, the X-ray images are usually not clear and many times the expert can confuse pneumonia with some other benign abnormality in the chest (Aydogdu, M., Ozyilmaz, E., Aksoy, H., Gursel, G., & Ekim, N. (2010), Ayan, E., & Ünver, H. M. (2019, April)). Today, there is a shortage of radiological experts to help diagnose this disease (Saul, C. J., Urey, D. Y., & Taktakoglu, C. D. (2019), Islam, S. R., Maity, S. P., Ray, A. K., & Mandal, M. (2019, May)), which is why computational techniques have been used to assist the expert in making decisions. Currently, computer-assisted diagnosis based on intelligent algorithms has become a reliable solution for the detection of pneumonia (Hammoudi, K., Benhabiles, H., Melkemi, M., Dornaika, F., Arganda-Carreras, I., Collard, D., & Scherpereel, A. (2021), Yee, S. L. K., & Raymond, W. J. K. (2020, September), Singh, V., Poonia, R. C., Kumar, S., Dass, P., Agarwal, P., Bhatnagar, V., & Raja, L. (2020), Sun, X., Douiri, A., & Gulliford, M. (2022)). Among the most popular techniques, convolutional neural networks (CNNs) have shown great reliability for pneumonia detection and therefore have been widely adopted by the scientific community (Sharma, H., Jain, J. S., Bansal, P., & Gupta, S. (2020, January), Rahman, T., Chowdhury, M. E., Khandakar, A., Islam, K. R., Islam, K. F., Mahbub, Z. B., ... & Kashem, S. (2020)). However, although it is treatable in most cases, early detection plays a key role in the treatment of pneumonia. In this research, an innovative approach based on kurtosis and entropy is presented as an alternative to techniques that use convolutional neural networks. This technique combines the use of statistical features and the use of a classifier to differentiate between images with pneumonia and healthy images. The process consists of calculating the statistical properties and finally, using logistic regression, classify an image with pneumonia and a healthy image.

Kurtosis and entropy for pneumonia classification

The classification system developed in this study consists of three phases: input image, kurtosis and entropy measurement and image classification (Figure 1). In the input image phase, we receive an image without knowing to which class it belongs (health or pneumonia state). In the kurtosis and entropy measurement, the input image is taken, and the corresponding statistical features are extracted. In the classification phase, the extracted features are analyzed and evaluated with a knowledge base to make a decision, which is the label to be assigned. Finally, in the diagnosis phase, the result of our classifier is used to determine the person's diagnosis.

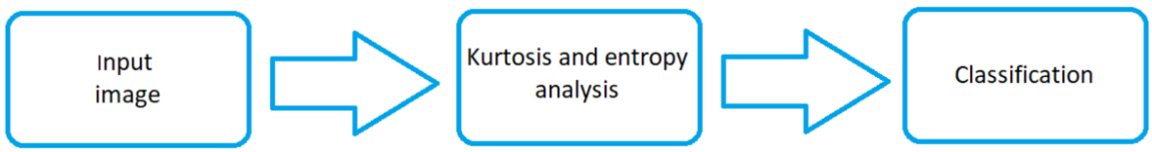

Figure 1: Overall process for pneumonia diagnosis.

Kurtosis and entropy measurement

For the statistical feature extraction, kurtosis and entropy are used. The kurtosis is a measurement used to calculate the shape of a probability distribution. In this study, kurtosis will serve as a threshold to separate people with pneumonia from healthy people. In image processing, the kurtosis is calculated according to the histogram. The histogram is the probability of observing an intensity value at a random location in an image. With this information, the kurtosis can be calculated as:

$$\gamma = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{x_i-\mu}{\sigma}\right]^4 - 3 \quad (1)$$

Where x_i is a sample of the histogram, μ is the mean of the pixels in the image, and σ es the standard deviation measurement in the test image. On the other hand, the entropy is calculated according to the texture features proposed by Unser (Unser, M. (1986)). In this approximation, the entropy is obtained from the sum and difference histograms measured from the test images. Once obtained the sum and difference histograms, the entropy is calculated according to the equation (2):

$$entropy \simeq -\sum_k P_s(k)\cdot\log\big(P_s(k)\big) - \sum_l P_d(l)\cdot\log\big(P_d(l)\big) \quad (2)$$

In this case, P_s and P_d are the sum and difference histograms calculated according to the algorithm proposed by Unser.

Logistic Regression for Pneumonia Classification

To define the logistic regression, we will denote as Y the binary response generated by the variables of interest $a_1, a_2, \ldots, a_p$ being in this case the kurtosis and the entropy features. The logistic regression model joins the conditional probability $P(Y = 1|a_1, a_2, \ldots, a_p)$ to $a_1, a_2, \ldots, a_p$ through:

$$\mathrm{P}(\mathrm{Y} = 1|\mathrm{a}_1, \mathrm{a}_2, \ldots, \mathrm{a}_\mathrm{p}) = \frac{\exp^{\beta_0+\beta_1 a_1+\cdots+\beta_p a_p}}{1+\exp^{\beta_0+\beta_1 a_1+\cdots+\beta_p a_p}} \quad (3)$$

Where $\beta_0, \beta_1, \ldots, \beta_p$ are denoted as the regression coefficients, which are estimated by maximizing the probability of the data set. Subsequently, the probability that $Y = 1$ is calculated for a new instance by replacing the β with their estimated values and the a with their realizations for the new instance. Finally, the following consideration is made:

$$\begin{aligned} &\boldsymbol{Y = 1\ if\ P(Y = 1) > c,} \\ &\boldsymbol{Y = 0\ otherwise} \end{aligned} \tag{4}$$

The most used threshold is $\boldsymbol{c = 0.5}$, which is also used in this study.

Results

Separability between classes

In this section, the experimentation and evaluation methods used in this methodology to test the classification process of the model are described. The set of X-ray images proposed by Kermany et. al. (Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C., Liang, H., Baxter, S. L., ... & Zhang, K. (2018)) was used to validate the system. Figure 2 shows an example of the images used in this approximation method. The images have a resolution of 1422x1152 pixels in grayscale and have 100 images for healthy state and 100 images for pneumonia.

Figure 2: X-ray images used in the classification system, a) Health, b) Pneumonia.

As a first approximation, the kurtosis and entropy features were calculated to observe the separability of the classes (health and pneumonia). Figure 3 shows the results obtained that generated the kurtosis and entropy features.

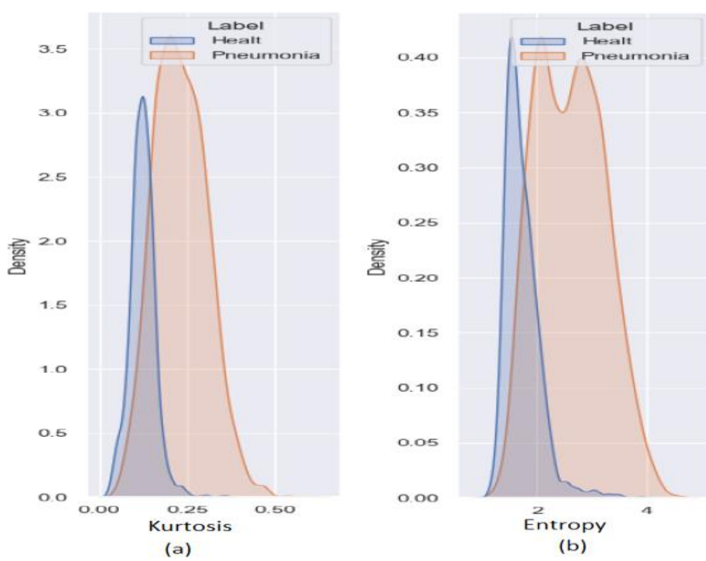

Figure 3: Kurtosis and entropy results.

Training system

To train and test the model, two image sets are created. The first image sets contain 160 elements of which 80 are healthy images, and the other one, contain 80 images with pneumonia. The remaining 60 images are used as a test, in where, 30 images are labeled as a healthy and the remaining as pneumonia.

Performance parameter

To measure the performance of the model, the accuracy metric has been used to describe the model. The accuracy as defined as the rate of predictions made correctly among the total number of samples obtained and it is defined according to the next equation (5):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{5}$$

Where TP and TN represents true positive (it was predicted health and the result was health) and true negative (it was predicted pneumonia and the result was pneumonia) respectively. FP and FN are false positive (it was predicted health and the result was pneumonia) and false negative (it was predicted pneumonia and the result was health). In Table 1 shows the result obtained using a confusion matrix. As it can be seen in this table, this model reached an 85% of accuracy.

Table 1. *Confusion matrix, pneumonia and health are the conditions of the patients.*

Condition	Health	Pneumonia
Health	35.25%	3.80%
Pneumonia	10.80 %	50.00%

To evaluate the performance of the system with other methodologies established in the state of the art, the system obtained was compared with other techniques related to the same task. Table 2 presents the method in the state of the art and the accuracy with which it can perform the classification of pneumonia (the accuracy results were obtained according to their research work).

***Table 2**. Comparison with different models of the state of the art.*

Author	Algorithm	Accuracy
Gang, P., Zhen et. al.	CNN	71%
Chowdary, G. J. (2021)	DCCN	76%
Demner-Fushman et. al.	Image segmentation	82%
Our proposal	Kurtosis and entropy analysis	85%

As can be seen in this table, this proposed method competes with the established methods in the state of the art, since high accuracy can be achieved by tuning few parameters.

Conclusions

A new method for the classification of pneumonia has been discussed. The system has multiple inputs and one output, where the inputs are based in kurtosis and entropy features, they are extracted from X-ray images and the output is a label prescription. A logistic regression algorithm has been constructed using a known set of X-ray images and has been evaluated on a small unseen set of X-ray images. The given image is useful to the class whose probability is highest.

A logistic regression classifier is an alternative to traditional systems because it is not a complex system and does not have many hyperparameters that must be tuned. One of the advantages of logistic regression is that only one hyperparameter needs to be tuned in order to perform its classification task. In this sense, we can see that this system can be an appropriate tool for the medical industry since it can serve as a tool that supports the decisions of a medical expert. Also, as it was shown, it presents a high precision that can compete with more complex systems.

References

Ayan, E., & Ünver, H. M. (2019, April). Diagnosis of pneumonia from chest X-ray images using deep learning. In *2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT)* (pp. 1-5). Ieee.

Aydogdu, M., Ozyilmaz, E., Aksoy, H., Gursel, G., & Ekim, N. (2010). Mortality prediction in community-acquired pneumonia - requiring mechanical ventilation; values of pneumonia and intensive care unit severity scores. *Tuberk Toraks*, *58*(1), 25-34

Chowdary, G. J. (2021). Impact of machine learning models in pneumonia diagnosis with features extracted from chest x-rays using vgg16. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(5), 1521-1530.

Demner-Fushman, D., Kohli, M. D., Rosenman, M. B., Shooshan, S. E., Rodriguez, L., Antani, S., ... & McDonald, C. J. (2016). Preparing a collection of radiology examinations for distribution and retrieval. *Journal of the American Medical Informatics Association*, *23*(2), 304-3.

Elshennawy, N. M., & Ibrahim, D. M. (2020). Deep-pneumonia framework using deep learning models based on chest x-ray images. *Diagnostics*, *10*(9), 649.

Gang, P., Zhen, W., Zeng, W., Gordienko, Y., Kochura, Y., Alienin, O., ... & Stirenko, S. (2018, March). Dimensionality reduction in deep learning for chest X-ray analysis of lung cancer. In *2018 tenth international conference on advanced computational intelligence (ICACI)* (pp. 878-883). IEEE.

Hammoudi, K., Benhabiles, H., Melkemi, M., Dornaika, F., Arganda-Carreras, I., Collard, D., & Scherpereel, A. (2021). Deep learning on chest X-ray images to detect and evaluate pneumonia cases at the era of COVID-19. *Journal of medical systems*, *45*(7), 1-10.

Islam, S. R., Maity, S. P., Ray, A. K., & Mandal, M. (2019, May). Automatic detection of pneumonia on compressed sensing images using deep learning. In *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)* (pp. 1-4). IEEE.

Jaiswal, A. K., Tiwari, P., Kumar, S., Gupta, D., Khanna, A., & Rodrigues, J. J. (2019). Identifying pneumonia in chest X-rays: a deep learning approach. *Measurement*, *145*, 511-518.

Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C., Liang, H., Baxter, S. L., ... & Zhang, K. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, *172*(5), 1122-1131.

Rahman, T., Chowdhury, M. E., Khandakar, A., Islam, K. R., Islam, K. F., Mahbub, Z. B., ... & Kashem, S. (2020). Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray. *Applied Sciences*, *10*(9), 3233.

Saul, C. J., Urey, D. Y., & Taktakoglu, C. D. (2019). Early diagnosis ofpneumonia with deep learning. *arXiv preprint arXiv:1904.00937*.

Sharma, H., Jain, J. S., Bansal, P., & Gupta, S. (2020, January). Feature extraction and classification of chest x-ray images using cnn to detect pneumonia. In *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 227-231). IEEE.

Singh, V., Poonia, R. C., Kumar, S., Dass, P., Agarwal, P., Bhatnagar, V., & Raja, L. (2020). Prediction of COVID-19 corona virus pandemic based on time series data using Support Vector Machine. *Journal of Discrete Mathematical Sciences and Cryptography*, *23*(8), 1583-1597.

Sun, X., Douiri, A., & Gulliford, M. (2022). Applying machine learning algorithms to electronic health records predicted pneumonia after respiratory tract infection. *Journal of clinical epidemiology*.

Unser, M. (1986). Sum and difference histograms for texture classification. *IEEE transactions on pattern analysis and machine intelligence*, (1), 118-125.

Yee, S. L. K., & Raymond, W. J. K. (2020, September). Pneumonia diagnosis using chest x-ray images and machine learning. In *Proceedings of the 2020 10th International Conference on Biomedical Engineering and Technology* (pp. 101-105).