JSCE (Journal of System and Computer Engineering)
Vol.6, No.1, January 2025, pp. 45-51
ISSN (online): 2723-1240
DOI : https://doi.org/10.61628/jsce.v6i1.1597

Automated Medical Image Processing for Lung Pneumonia Diagnosis Based on LS-SVM

Nursuci Putri Husain[1,a,]; Hamdan Arfandy[2,b]; Ryan Midzar Wiradinata Ramli [3,c*]
[1,2,3)] Program Studi Teknik Informatika, Fakultas Teknik, Universitas Islam Makassar
[a] nursuciputrihusain.dty@uim-makassar.ac.id; [b] hamdanarfandy@gmail.com; ryanmizarwiradinata@gmail.com
** Corresponding author*

Abstract

Pneumonia is an inflammation of the lungs that causes pain when breathing and limits oxygen intake. Pneumonia can be caused by bacteria, viruses, and fungi. Image processing, a branch of informatics or computer science, is a field highly related to the manipulation and analysis of digital images. This study aims to design a medical image processing system as an alternative to support the diagnosis of Pneumonia in the lungs using the Least Square Support Vector Machine (LS-SVM) method. LS-SVM is a simpler and modified model of the SVM method. Histogram of Gradient (HoG) is a commonly used feature extraction method in image processing and object detection. The objective of this study is to improve the quality of healthcare services and assist in faster and more accurate clinical decision-making. The results show that lung image analysis using the LS-SVM method has a good accuracy level in the image classification process, with 2000 training data inputs processed in the preprocessing stage, consisting of 1000 Pneumonia images and 1000 normal lung images, while the testing data used consisted of 500 images, with 250 Pneumonia images and 250 normal lung images. Based on the tested data, the system achieved an accuracy of 81% for 1300 tests, proving that the LS-SVM method is effective in image processing with satisfactory results.

Keywords— *Classification, Pneumonia, image processing, HoG, LS-SVM*

1. Introduction

The advancement of technology has driven the adoption of digital images as the primary format for recording and storing visual information. The production process of photos in digital image form not only requires an understanding of photography but also demands a specialized processing known as image processing (Jiang, 2020).

Pneumonia is an inflammation of the lungs that causes pain when breathing and limits oxygen intake (Cai et al., 2024). Pneumonia can be caused by bacteria, viruses, and fungi. The virus that causes pneumonia includes the Respiratory Syncytial Virus (RSV) (Lee et al., 2016). Early and accurate diagnosis is crucial to ensure timely treatment and improve patient outcomes.

Image processing, which is a branch of informatics or computer science, is a field closely related to the manipulation and analysis of digital images. In the modern technological era, the use of computers and specialized algorithms plays a central role in the development of image processing techniques. The main goals of image processing are to enhance image quality, optimize the information it contains, and even extract specific features that can be applied in various fields such as pattern recognition, object detection, and more.

By building a medical image processing system, the role of image processing extends beyond the artistic aspects of photography; it also serves as a critical component in the evolution of visual documentation technology across various fields of science and industry (Rayed et al., 2024). This research aims to design and implement a medical image processing system as an alternative tool to support the diagnosis of Pneumonia in the lungs using the LS-SVM (Least Squares Support Vector Machine) method. This system is envisioned to enhance diagnostic accuracy, reduce reliance on manual interpretations, and provide a faster, more reliable approach to identifying pneumonia cases.

2. Literature Review

The adoption of automated medical image processing systems has surged with advancements in artificial intelligence, especially for diagnosing lung diseases such as pneumonia (Ortiz-Toro et al., 2022). Pneumonia, a serious inflammation of the lungs caused by bacteria, viruses, or fungi, requires rapid and accurate diagnosis to improve patient outcomes. Traditional diagnosis methods, including manual X-ray interpretation by radiologists, can be time-consuming and prone to human error, which highlights the need for automated solutions. Image processing, a specialized branch of computer science, enhances diagnostic precision by manipulating digital images to optimize their quality and extract important features for diagnosis (Irwan Prasetya Gunawan et al., 2023).

Least Squares Support Vector Machine (LS-SVM) is a variant of Support Vector Machine (SVM) that simplifies the optimization process through a squared error minimization approach (Husain et al., 2017). LS-SVM has gained traction in medical imaging for its effectiveness in binary classification tasks, such as distinguishing between healthy and diseased lung tissues. Studies have demonstrated that LS-SVM is highly accurate in analyzing chest X-ray images, making it suitable for pneumonia detection. Additionally, LS-SVM’s lower computational requirements compared to traditional SVM make it a more practical choice for large-scale diagnostic applications.

Feature extraction is essential in medical image processing to capture distinctive attributes in an image that aid in classification (Husain, 2021). The Histogram of Gradients (HoG) technique, which captures edge orientation and texture, is widely used for extracting structural details in medical images. In lung pneumonia diagnosis, HoG enables the LS-SVM model to identify and classify irregularities indicative of infection, supporting a more precise diagnosis. Combined, these methods have shown promising results in providing accurate, reproducible, and efficient support for clinical decision-making, thus reducing diagnostic delays and improving healthcare service quality.

3. Method

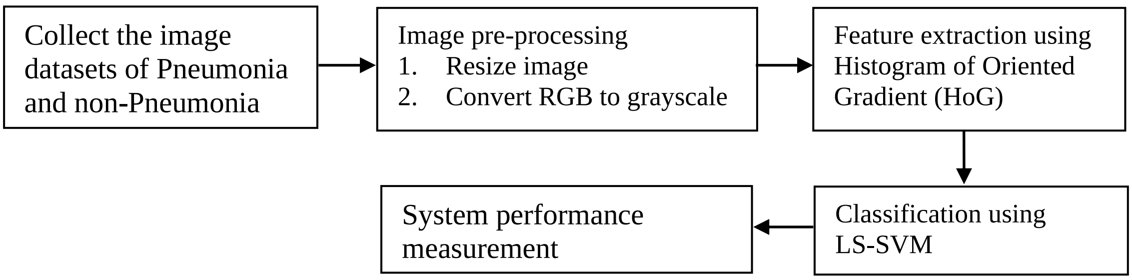

Figure 1. Design of automatic diagnosis of lung pneumonia

The proposed system architecture in this study is designed to support the automatic diagnosis of lung pneumonia through a series of medical image processing stages. The system consists of several main steps, including the collection of image datasets of Pneumonia and non-Pneumonia cases, feature extraction, classification using the LS-SVM method, and system performance measurement. The system's flow diagram is presented in Figure 1, which illustrates the systematic process from the initial stage to the evaluation of results.

3.1 Datasets of Pneumonia and non-Pneumonia

The image dataset of Pneumonia and non-Pneumonia cases used in this study is sourced from the Kaggle library. Published by Paul Mooney on March 25, 2018, the data was collected from pediatric patients aged 1-5 years at the Guangzhou Women and Children's Medical Center, Guangzhou.

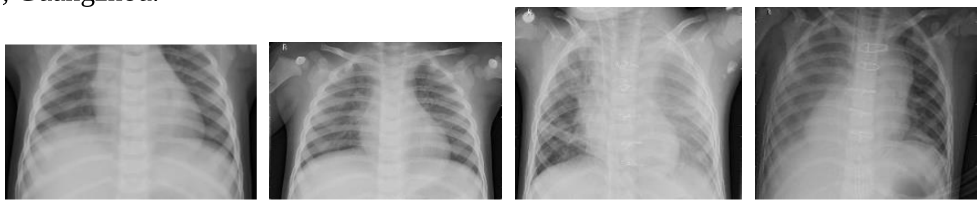

Figure 2. Example Dataset of Pneumonia and Non-Pneumonia Images

3.2 Preprocessing Step

The image pre-processing stage aims to improve image quality for optimal analysis. The first step is image resizing, which standardizes the dimensions of all images in the dataset to ensure consistency, such as resizing them to 224x224 pixels. This step helps reduce computational complexity and memory requirements while maintaining the integrity of image features, making the data suitable for processing by classification algorithms.

The second step is converting RGB to grayscale, which simplifies the images by removing color information and focusing solely on intensity levels. This reduces the data's dimensionality and highlights key patterns, such as lung structures and abnormalities, that are crucial for diagnosing Pneumonia. These pre-processing steps ensure the dataset is uniform and ready for further analysis and classification using LS-SVM.

3.3 Feature Extraction Step

Histogram of Oriented Gradients (HOG) is a feature extraction method used to represent images by capturing the distribution of gradient orientations, which helps in recognizing objects (Wang et al., 2018). In the context of classifying pneumonia and non-pneumonia images, HOG can effectively capture texture and structural patterns in lung images. The method works by dividing an image into small cells, calculating the gradient at each pixel, and then constructing histograms of gradient orientations for these cells. The result is a compact and robust representation of the image's local features, which is crucial for distinguishing between infected and healthy lung tissue in medical imaging tasks like pneumonia detection.

The process of extracting HOG features begins by computing the gradient magnitude and orientation for each pixel in the image. This step uses the following formulas for the gradient in the *x* and *y* directions:

$$G_x = I(x+1, y) - I(x-1, y) \quad (1)$$

$$G_y = I(x, y+1) - I(x, y-1) \quad (2)$$

Where *I(x,y)* is the intensity of the image at pixel (*x*,*y*). The magnitude and orientation of the gradient at each pixel are then calculated as:

$$Magnitude = \sqrt{G_x^2 + G_y^2} \quad (3)$$

$$Orientation = atan2(G_y, G_x) \quad (4)$$

These gradients are grouped into cells, and the histogram of gradient orientations is computed for each cell. The histogram typically uses nine bins, each corresponding to an orientation range. These cell histograms are then normalized over larger blocks to improve robustness against lighting variations and small geometric transformations. Once the HOG features are extracted, machine learning algorithms, such as Support Vector Machines (SVM), are used to classify the image as either pneumonia or non-pneumonia. The power of HOG in medical imaging lies in its ability to capture the important structural features of the lung tissue, which are critical for distinguishing between healthy and infected areas.

3.4 Classification Step

Once the features are extracted, LS-SVM is used to classify the pneumonia and non-pneumonia images. The LS-SVM aims to find a hyperplane (or decision boundary) that separates the two classes (pneumonia and non-pneumonia) in the feature space (Husain et al., 2017). Unlike traditional SVM, which minimizes the regularization term and hinge loss function, LS-SVM formulates the problem as a least squares optimization problem. This formulation simplifies the training process by using the following objective function.

$$\min_{w,b,\epsilon} \frac{1}{2} \parallel w \parallel^2 + \frac{1}{2}\lambda \sum_{i=1}^{n} \epsilon_i^2 \quad (5)$$

Where:

w is the weight vector (defining the decision boundary)
b is the bias term
ϵ_i are the slack variables
λ is a regularization parameter
n is the number of training samples

The LS-SVM finds a hyperplane that minimizes the sum of squared errors while maintaining a balance between the model's complexity and the margin of separation between the classes. This approach makes LS-SVM computationally more efficient than standard SVM in some cases, especially for large datasets.

3.5 System performance measurement

Accuracy is a fundamental metric for evaluating the performance of a classification system, including models designed to distinguish between pneumonia and non-pneumonia images (Husain & Aji, 2019). It measures the proportion of correctly classified samples (both pneumonia and non-pneumonia) out of the total samples evaluated. This metric is particularly useful in balanced datasets, where the number of samples in each class is approximately equal. High accuracy indicates that the model effectively captures patterns in the data and makes reliable predictions for both classes. The formula for accuracy is:

$$Accuracy = \frac{\text{True Positives (TP)} + \text{True Negatives (TN)}}{Total\ Samples\ (TP + TN + FP + FN)} \quad (6)$$

Where:

True Positives (TP): Pneumonia cases are correctly classified as pneumonia.
True Negatives (TN): Non-pneumonia cases are correctly classified as non-pneumonia.
False Positives (FP): Non-pneumonia cases are misclassified as pneumonia.
False Negatives (FN): Pneumonia cases are misclassified as non-pneumonia.

4. Results And Discussion

In this study, the dataset was processed to prepare a total of 2,000 training images, consisting of 1,000 pneumonia cases and 1,000 normal lung images. This balanced dataset ensures that the model has an equal representation of both classes, reducing the risk of bias during training. For evaluation, a testing dataset of 500 images was used, comprising 250 pneumonia cases and 250 normal lung images. Preprocessing steps included resizing images for uniformity, grayscale conversion to reduce complexity, and normalization to enhance the contrast of important features.

The visualization in Figure 3 illustrates the extracted Histogram of Oriented Gradient (HoG) features from a sample image in the dataset. The HoG method captures critical texture and structural patterns by analyzing the local gradients and their orientations. The grid-like appearance with directional lines reflects the gradient magnitudes and orientations, which are essential in identifying features such as lung consolidations in pneumonia cases or normal texture patterns in healthy lungs. These HoG features form the basis for classification, providing the LS-SVM model with robust and discriminative information for accurately distinguishing between pneumonia and non-pneumonia images. This visualization validates that the feature extraction step successfully captures relevant information needed for effective classification.

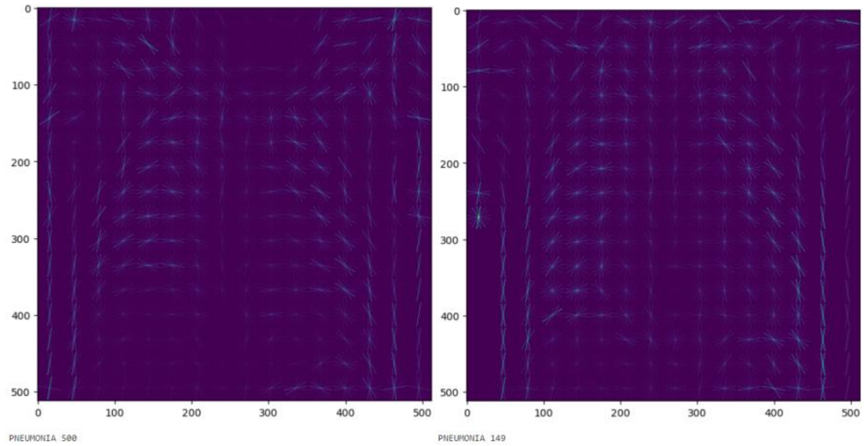

Figure 3. Histogram of Oriented Gradients (HoG) feature extraction visualizations

The table shows examples of classification results from the proposed pneumonia detection system. It includes three cases: one correctly classified normal image, one correctly classified pneumonia image, and one misclassified pneumonia image. The first case, labeled as "Normal 053," was accurately predicted as normal, demonstrating the system's ability to identify healthy lung patterns. Similarly, the second case, "Pneumonia 077," was correctly classified as pneumonia, reflecting the system's effectiveness in detecting abnormal lung conditions. However, in the third case, "Pneumonia 078," the system misclassified the pneumonia image as normal, highlighting a limitation in distinguishing certain pneumonia patterns.

Table 1. Example of Classification Results Using the Proposed Method

Image	Label	Prediction
	Normal 053	Normal

	Pneumonia 077	Pneumonia
	Pneumonia 078	Normal

In this study, two tests were conducted. In the first scheme, the system utilized 1,000 training data to train the model and 300 test data to evaluate its performance. Based on the testing results in Table 1, the system achieved a prediction accuracy rate of 75.0%, with a breakdown of 225 correct predictions and 75 incorrect predictions. Although an accuracy rate of 75% demonstrates a reasonably good performance, it still leaves room for further improvement. The potential for enhancement can be pursued by increasing the amount of training data.

Table 2. Confusion Matrix for LS-SVM Classification Testing Using 300 Test Data

TP	FP	FN	TN	Accuracy (%)
88	62	13	137	75%

In the results of the second scenario testing, the researchers expanded the dataset by adding 2,000 images for training and 500 images for testing. From this experiment, the accuracy rate increased to 81.0%. Out of the 500 test data, 405 were correctly classified, while 95 were misclassified. This improvement in accuracy demonstrates that with more data, the model can learn better and produce more accurate predictions.

Table 3. Confusion Matrix for LS-SVM Classification Testing Using 500 Test Data

TP	FP	FN	TN	Accuracy (%)
138	92	3	267	81%

5. Conclusions

This study developed an automated medical image processing system for the diagnosis of Pneumonia in the lungs using the LS-SVM (Least Squares Support Vector Machine) method. The system aims to enhance accuracy and speed in clinical decision-making related to pneumonia diagnosis. Through image processing using the HoG (Histogram of Oriented Gradients) feature extraction, the system processed 2,000 training data and 500 test data, consisting of pneumonia and normal lung images. The test results showed an accuracy of 81%, demonstrating the effectiveness of the LS-SVM method in medical image classification with satisfactory results. This system can serve as an alternative to support pneumonia diagnosis, reduce human error, and speed up the diagnostic process, although there is still room for improvement, particularly in increasing the training data.

References

Cai, K., Wang, Z., Yang, X., Fu, W., & Zhao, X. (2024). Predicting Clinical Outcomes in COVID-19 and Pneumonia Patients: A Machine Learning Approach. *Viruses*, *16*(10), 1624. https://doi.org/10.3390/v16101624

Husain, N. P. (2021). Best First Feature Selection Dan Radial Basis Function Untuk Klasifikasi Penyakit Diabetes. *Iltek : Jurnal Teknologi*, *16*(01), 29–33. https://doi.org/10.47398/iltek.v16i01.41

Husain, N. P., & Aji, N. B. (2019). Klasifikasi Sinyal EEG Dengan Power Spectra Density Berbasis Metode Welch Dan MLP Backpropagation. *Jurnal ELTIKOM*, *3*(1), 17–25. https://doi.org/10.31961/eltikom.v3i1.99

Husain, N. P., Arisa, N. N., Rahayu, P. N., Arifin, A. Z., & Herumurti, D. (2017). Least Squares Support Vector Machines Parameter Optimization Based On Improved Ant Colony Algorithm For Hepatitis Diagnosis. *Jurnal Ilmu Komputer Dan Informasi*, *10*(1), 43. https://doi.org/10.21609/jiki.v10i1.428

Irwan Prasetya Gunawan, Ratnadewi, Safrizal, Nursuci Putri Husain, & Guson Prasamuarso Kuntarto. (2023). *Pemrosesan Citra* (P. 143). Pt Mafy Media Literasi Indonesia. https://books.google.co.id/books?hl=id&lr=&id=_f_uEAAAQBAJ&oi=fnd&pg=PA1&ots=sDEslbXYNo&sig=dIkcpAXmRjKYbhSEqVUXOMCrD8s&redir_esc=y#v=onepage&q&f=false

Jiang, N. (2020). Role of Digital Image Processing in Image Art under the Background of Big Data. *IOP Conference Series: Materials Science and Engineering*, *750*(1), 012095. https://doi.org/10.1088/1757-899X/750/1/012095

Lee, K. H., Gordon, A., & Foxman, B. (2016). The role of respiratory viruses in the etiology of bacterial pneumonia: An ecological perspective. *Evolution, Medicine, and Public Health*, *2016*(1), 95–109. https://doi.org/10.1093/emph/eow007

Ortiz-Toro, C., García-Pedrero, A., Lillo-Saavedra, M., & Gonzalo-Martín, C. (2022). Automatic detection of pneumonia in chest X-ray images using textural features. *Computers in Biology and Medicine*, *145*, 105466. https://doi.org/10.1016/j.compbiomed.2022.105466

Rayed, Md. E., Islam, S. M. S., Niha, S. I., Jim, J. R., Kabir, M. M., & Mridha, M. F. (2024). Deep learning for medical image segmentation: State-of-the-art advancements and challenges. *Informatics in Medicine Unlocked*, *47*, 101504. https://doi.org/10.1016/j.imu.2024.101504

Wang, C., Li, Z., Dey, N., Li, Z., Ashour, A. S., Fong, S. J., Sherratt, R. S., Wu, L., & Shi, F. (2018). Histogram of Oriented Gradient Based Plantar Pressure Image Feature Extraction and Classification Employing Fuzzy Support Vector Machine. *Journal of Medical Imaging and Health Informatics*, *8*(4), 842–854. https://doi.org/10.1166/jmihi.2018.2310