Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 832−840
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7335

Comparative Analysis of CNN and SVM Algorithms for Pneumonia Classification from Chest X-Ray Images

Ar Ainun Bela, Kemas M Lhaksmana*

Faculty of Informatics, Telkom University, Bandung, Indonesia
Email: [1]arainunbl@students.telkomuniversity.ac.id, [2,*] kemasmuslim@telkomuniversity.ac.id
Correspondence Author Email: kemasmuslim@telkomuniversity.ac.id
Submitted: **12/05/2025**; Accepted: **30/06/2025**; Published**: 30/06/2025**

Abstract−Pneumonia significantly threatens human health, especially in children and the elderly. Diagnosing pneumonia using chest radiographs is time consuming and requires expert interpretation. This study proposes a comparative analysis of two algorithm models, namely Support Vector Machine (SVM) and Convolutional Neural Network (CNN), in CNN algorithm, it specifically uses DenseNet121 and InterceptionV3 architectures for the classification of chest X-ray images in pneumonia and normal categories. The methods used include data preprocessing with normalization and augmentation. The dataset is split into training and testing subsets, and implementation of SVM and CNN algorithms for classification. Kaggle provided the dataset for this study, comprising 5,863 chest X-ray images. Metrics such as accuracy, precision, recall, and F1-score calculated from the confusion matrix were used to evaluate the model's effectiveness. The test findings show that the DenseNet121 model has the best performance among the three models, with an accuracy, recall, and F1-score of 94%. The InceptionV3 model achieved 89% in accuracy, recall, and F1-score, which is higher than DenseNet121. Meanwhile, the SVM model showed the lowest performance with an accuracy of 81%, precision of 85%, recall of 81%, and F1-score of 79%. These outcomes signifies that Convolutional Neural Network (CNN) architectures, particularly DenseNet121, have superior capabilities in extracting complex features from chest X-ray images and show great potential to be applied in automatic and accurate pneumonia detection systems.

Keywords: Pneumonia; Image Classification; Support Vector Machine; Densenet121; InceptionV3; Chest X-Ray

1. INTRODUCTION

The rapid advancement of digital technology has significantly transformed the healthcare sector, particularly in medical imaging. Digitization of medical images facilitates more efficient diagnostic processes and supports the development of medical informatics systems [1]. Medical images have a crucial role in disease identification and classification. Chest X-rays are among the most widely used diagnostic tools for visualizing the condition of the lungs, heart, and respiratory tract. Doctors can detect gray-white areas in the lungs infected with viruses, bacteria, fungi, or other parasites to identify diseases such as pneumonia [2]. Pneumonia remains the top infectious killer of children globally. During 2019, pneumonia led to the deaths of 740,180 children under five years of age, making up 14% of all fatalities in that age group and 22% of all deaths among children aged one to five [3]. Pneumonia is a lung infection that causes inflammation in the alveoli, which become filled with pus and fluid, reducing the lungs' capacity to absorb oxygen. Clinically, it often presents with symptoms such as persistent cough, fever, chest pain, and difficulty breathing, especially in high-risk groups such as infants, elderly individuals, and immunocompromised patients [4]. Pneumonia classification can be approached using machine learning techniques, one of which is the CNN algorithm, a machine learning method that excels in automatic feature extraction from image data, making it an effective method in image recognition, especially in disease classification tasks such as pneumonia. The convolutional neural network (CNN) models learn spatial hierarchies of feature from the input images to extract complex visual patterns that are helpful in identifying pathological signs in the chest X-ray [5]. The machine learning method for pneumonia classification from chest X-ray images has been studied intensely. I Made Dendi May Sanjaya (2020) used CNN for pneumonia classification using lung chest X-ray images obtained from the Kaggle dataset with a total of 5,840 images. The CNN model has the RelU activation function and uses the Adam optimizer and was trained for 200 epochs. This model architecture consists of two convolution layers, two pooling layers, and two fully-connected layers. It achieved an average test accuracy of 89.58%. however, signs of overfitting were observed, likely caused by an imbalanced dataset and constrained model design, which may hinder its ability to generalize to new data and reduce its effectiveness in clinical applications[2]. The study by Eka Putra Agus Meidiawan and Muljono (2024) developed a combination of MobileNetV2 and SVM, trained on a SMOTE-Tomek-balanced CXR dataset. They tested three methods: Direct SVM (97% accuracy), Sobel edge detection + SVM (95%), and MobileNetV2 + SVM (98% accuracy), demonstrating the effectiveness of SVM when combined with CNN-based feature extraction. However, their study did not include comparisons with modern architectures such as DenseNet121 or InceptionV3 [6]. The research conducted by M.J. Sulistio, C. Lubis, and K.Kunci (2023) implemented Convolutional Neural Network (CNN) and MobileNet to detect pneumonia and COVID-19 using a smartphone-based application. Although this study is relevant in the context of mobile technology implementation for respiratory disease diagnosis, its focus was not on comparing CNN and SVM nor did it systematically evaluate the performance of modern Convolutional Neural Network (CNN) architectures on the same dataset [7].

The study by Salehi et al. (2023) emphasizes the importance of transfer learning in image processing for medical images such as DenseNet and Inception for complex classification tasks. However, they did not include a performance comparison with traditional algorithms like Support Vector Machine (SVM) under the same data

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 832−840
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7335

conditions [8]. Although CNN and SVM have been widely used individually, few studies have directly compared the performance of DenseNet121 and InceptionV3 representing modern CNN architectures with SVM in the task of pneumonia classification using the same chest X-ray dataset. In addition, the effectiveness of automatic feature extraction using CNN versus manual approaches such as HOG used in SVM has also rarely been explored quantitatively. The performance of CNN and SVM can be assessed using the confusion matrix through metrics such as accuracy, precision, recall, and F1-score. This classification process is intended to determine ehich algorithm performs better in assisting pneumonia diagnosis from medical images [9]. The evaluation is expected to offer valuable insights into the most effective approach for enchancing diagnostic accuracy and efficiency, ultimately halping to minimize the risk of misdiagnosis in pneumonia classification [10].

2. RESEARCH METHODOLOGY

2.1 Research Stages

To evaluate pneumonia classification using the SVM and CNN algorithms, several importance stages were carried out which included the data preparation process to model training. Figure 1 shows the workflow of the research conducted.

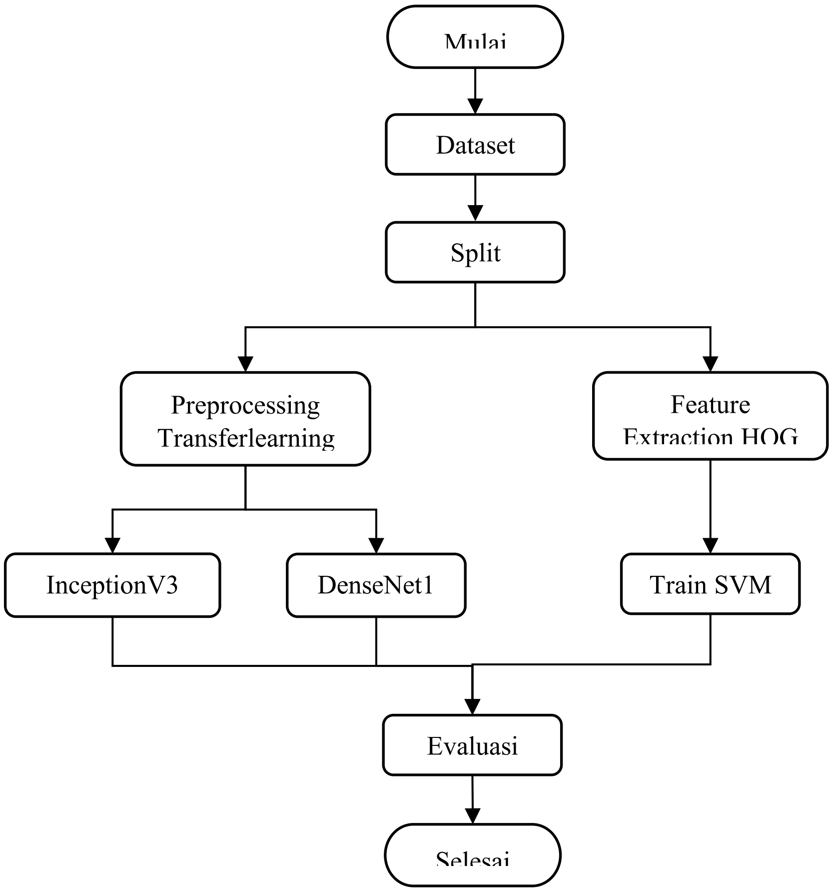


Figure 1. Research Stages in Pneumonia Classification Workflow.

2.1 Dataset

The dataset utilized in this study was sourced from Kaggle, compiled by Paul Mooney. the data is organized into three main folders, namely training, testing, and validation. The subfolders in the data represent each image category, such as pneumonia and normal. Derived from historical data of children aged between one and five years at Guangzhou Women and Children's Medical Center, 5,863 chest X-ray images in JPEG format with anteroposterior views were collected. Standard clinical procedures encompass image acquisition. Quality control measures are implemented for each radiograph before any analysis is carried out. During this phase, images that are rejected include those of low quality and unreadable images.

2.2 Feature Extraction

The method through which the feature extraction from SVM model is done is HOG which stands for Histogram of Oriented Gradients, widely known in image processing and computer vision, especially for object detection and

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 832−840
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7335

recognition. The method works based on the analysis of 'gradient orientations pattern' in an image to find patterns that can recognize visual forms and structures and to be robust even in the presence of variations in illumination and complex backgrounds [11].

2.3 Data Preprocessing

Data processing in this study involves a series of critical procedures, including normalization of the chest X-rays so that the dimensions are made uniform, data augmentation for increasing variation of data training, and the conversion of data to SVM and CNN-compatible formats. Moreover, The data is divided into the training and testing subsets to ensure that the process of model evaluation is conducted appropriately and accurately. These preprocessing operations are done to enhance the data quality and optimize the performance of the classification models.

2.4 Supprot Vector Machine Classification

The classification process in Support Vector Machines revolves around constructing a hyperplane that maximizes the distance between distinct data classes. A mathematical model guides the process which seeks an optimal margin between data points of different classes and the separating hyperplane [12].

$$f(x) = W \cdot X + b \tag{1}$$

$$(w \cdot x_i) + b > 1 \, for \, y_i = 1 \tag{2}$$

$$(w \cdot x_i) + b \leq -1 \, for \, y_i = -1 \tag{3}$$

The letter *w* within this framework stands for the normal vector which points at a right angle to the hyperplane. The letter *x* represents input vectors which include both data and features while letter *b* functions as the bias for establishing where the hyperplane exists inside the feature space [12].

$$Margin = \frac{1}{||w||} \tag{4}$$

SVM seeks to find the optimal hyperplane solution which creates the maximum gap between the two classes. The calculation of the margin involves computing the shortest distance from the hyperplane to the closest data points of each class through defined mathematical methods, where the margin is calculated as the inverse of the norm of the weight vector $\|w\|$, which means that the smaller the value of the greater the separation distance between the classes [13].

$$Min\, r(w) = \frac{1}{2}\, ||w||^2 \tag{5}$$

$$y_i(w \cdot x_i + b) \geq 1, \forall i \tag{6}$$

This concept can be formulated in the form of a quadratic programming problem, which is a mathematical approach involving a quadratic objective function with linear constraints. The purpose of this concepts is to find the minimum point of an equation subject to specific constraints, where the main goal is to minimize the objective function $\frac{1}{2}\, ||w||^2$ to obtain the maximum margin, provided that each data x_i is correctly classified so that it satisfies $y_i(w \cdot x_i + b) \geq 1, \forall i$. In SVM, the objective function used for this minimization process is expressed through a mathematical equation [14].

2.5 Transfer Learning Convolutional Neural Network

The deep learning program CNN has been purpose-built to analyze grid data which typically includes pictures. The technology stands as an improvement over the traditional Multilayer Perceptron model which processes two-dimensional datasets. Through their hierarchical layer structure CNNs handle both feature extraction and classification operations automatically [15]. A typical CNN model consists of three fundamental layer types which include the convolutional layer along with the pooling layer and fully connected layer [16].

a. Convolutional Layer
   The convolutional layer holds a central position within CNN architecture for handling input data through its convolution operation. Convolution refers to an arithmetic process that merges two mathematical functions to create a different function. The operation starts by using filters or kernels on input data which leads to the extraction of key details before these become feature maps [17].
b. Pooling Layer
   Pooling stands as a necessary operation in CNNs which people also call subsampling or downsampling because it reduces data complexity by decreasing its dimensions. This reduction process happens through filter application alongside a specified stride value for moving. Different types of pooling methods exist including max pooling that calculates average values within defined regions and average pooling which determines average values across varying input sizes [18].
c. Fully Connected Layer

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 832−840
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7335

A fully connected layer, often found in MLPs, works by connecting each neuron to those in the next layer. This helps transform the data into a format that supports linear classification. Neurons from the previous convolutional layer are converted into one-dimensional data, which are then passed into the fully connected layer. This process permanently removes spatial information from the data and is not reversible. The network employs fully connected layers solely at its terminal layer [7].

In this research, transfer learning methods will be applied. Transfer learning is a technique in which a model trained on a generic dataset is adapted to a specific dataset of interest. Typically, the pretrained models used in transfer learning have been trained on millions of images that are generic in nature and do not specifically correspond to the dataset of interest. The pretrained model is then customized to be applicable to the dataset that is the focus of the research or development. Such model are trained on millions of images covering thousands of object classes. Various filters (kernels) in the model will be active for different shapes, colors, and textures in the image. These filters can then be reused to learn features on new data. Once the features are learned, they can be connected to the hidden layer before the final classification layer to adapt the model to the new data [8].

a. The CNN model InceptionV3 demonstrates its ability to handle both image classification and image recognition functions. The Inception V3 structure includes different convolution operations that use various kernel sizes to extract features efficiently. The model employs regularization methods alongside reduction techniques for preventing overfitting [19].
b. The architecture of DenseNet121 includes dense connections between each layer in order to reuse features across the network [20].

2.6 Evaluation

The confusion matrix serves as a standard model evaluation tool for determining classification model accuracy. It organizes information into a table that shows how model predictions match actual results which helps users evaluate both accuracy and general model performance [21].

Table 1. Confusion Matrix

Class		Actual Value	
		Positive	Negative
Prediction Value	True	TP (True Positive)	FP (False Positive)
	False	FN (False Negative)	TN (True Negative)

The evaluation of a classification model's performance depends on the variables listed in table 1. The term TN represents the model's correct prediction of negative class instances which means the actual data belongs to the negative category and the model makes correct classifications. The model correctly identifies positive class data during True Positive (TP) cases. The model misclassifies negative class data as positive during False Positive (FP) cases. When the model fails to detect positive class data it incorrectly classifies such data as negative which is known as False Negative (FN) [22].

a. Accuracy is the ratio of correct predictions to the total predictions made by the model. The accuracy value is obtained by summing the correct predictions (TP + TN) and dividing by the total predictions (TP + TN + FP + FN) [23].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

b. Precision measures how accurate the model's positive predictions are. It is computed by dividing the count of accurately predicted positive cases (True Positives/TP) by the total number of positive predictions made by the model, which is TP + FP [24].

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

c. Recall indicates how effectively the model identifies all actual positive cases. This is calculated by dividing the count of TP by the sum of TP + FN[25].

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

d. In the world of machine learning and pattern recognition, the evaluation of performance of classification models usually relies on the commonly used F1-score. F1 scoring combines precision and recall in such a way that they are harmonically averaged into offering a singular score which provides a fairer assessment of the performance of the model using highly skewed datasets [26].

$$F1 - score = \frac{2 * Recall * Presisi}{Recall + Presisi} \tag{10}$$

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 832−840
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7335

3. RESULT AND DISCUSSION

This whole section here is about documenting the training and testing procedure of the models under consideration, namely, SVM and CNN architectures with special reference to InceptionV3 and DenseNet121. The training process is duly recorded with detailed tables and results of prediction made by each model. Also, evaluation metrics such as recall, precision, F1-score, and accuracy have been given to quantitatively assess the initial performance of the models before evaluation against unseen data.

The findings discussion part here delves into examining rather profoundly how well each of the models performed when it comes to uncovering patterns in medical imaging data of different types; specifically, in the differentiation between Normal and Pneumonia cases. Furthermore, their test performance is compared to be able to decide which model has the greatest level of precision, stability, and trustworthiness in the classification of the chest X-ray images.

3.1 Test Result

3.1.1 Support Vector Machine Models

In this section, the Support Vector Machine (SVM) algorithm is utilized to classify medical chest X-ray images into two categories: normal and pneumonia. The evaluation of the model's performance is conducted through two phases: validation and testing.

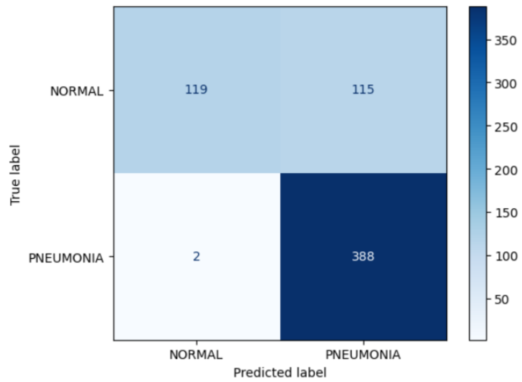
Validation Results:

	precision	recall	f1-score	support
NORMAL	1.00	0.88	0.93	8
PNEUMONIA	0.89	1.00	0.94	8
accuracy			0.94	16
macro avg	0.94	0.94	0.94	16
weighted avg	0.94	0.94	0.94	16

Test Results:

	precision	recall	f1-score	support
NORMAL	0.98	0.51	0.67	234
PNEUMONIA	0.77	0.99	0.87	390
accuracy			0.81	624
macro avg	0.88	0.75	0.77	624
weighted avg	0.85	0.81	0.79	624

(a)



(b)

Figure 2. (a) Classification SVM, (b) Confusion matriks

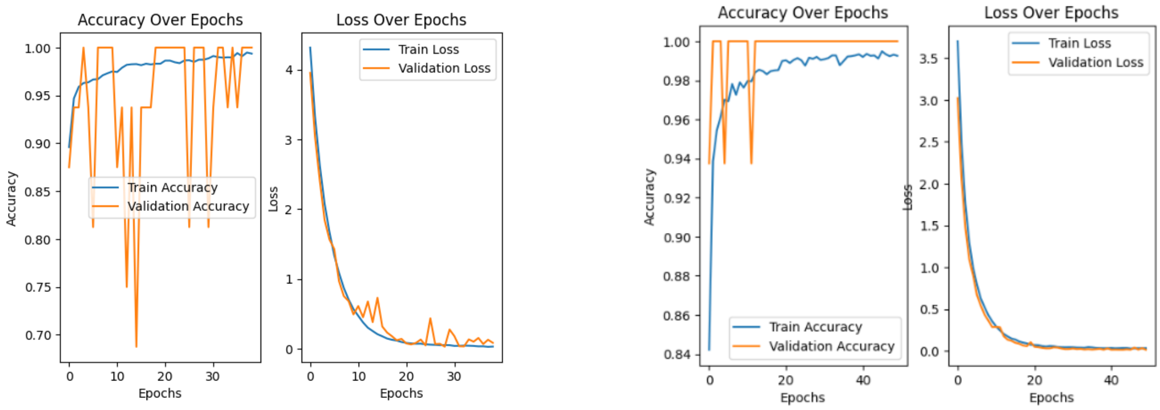
Based on Figure 2 (a), the evaluation result of the SVM model training indicates strong performance. Validation Result: the model achive an accuracy of 94%, indicating very good overall performance on the validation dataset. In terms of precision, the model reached 100% of the normal class, meaning all predicted normal cases were correct, and 89% for the pneumonia class, indicating that 89% of the predicted pneumonia cases were accurate. Regarding recall, the model successfully detected 88% of actual normal cases and achieved a perfect 100% recall for pneumonia, identifying all true pneumonia instances. The F1-score, which reflects the harmonic mean of precision and recall, was 93% for normal and 94% for pneumonia, showing a balanced and reliable classification performance across both classes. Test Result: Similarly, the model maintained an accuracy of 94% on the test data, again demonstrating excellent performance. The precision was 100% for normal, where every predicted normal case was correct, and 89% for pneumonia. The recall values mirrored those of the validation result, with 88% for normal and 100% for pneumonia. The F1-scores also remained consistent, at 93% for the normal class and 94% for the pneumonia class, indicating that the model generalized well and maintained its classification quality on unseen data.

Based on Figure 2 (b), the confusion matrix result from the SVM model indicated the following classification outcomes: the model correctly predicted 199 cases as normal, representing the True Negatives (TN). It misclassified 115 normal cases as pneumonia, which are counted as FP. Additionally, the model incorrectly predicted 2 pneumonia cases as normal, resulting in FN. Lastly, it accurately identified 388 pneumonia cases, which correspond to the True Positives (TP).

3.1.2 Convalutional Neural Network Models

This section presents the performance evaluation otcomes of two deep learning models, namely InceptionV3 and DenseNet121, in classifying chest X-ray images into two categories: normal and pneumonia. The evaluation focuses on three key aspects: accuracy and loss graphs during training, confusion matrix, and classification report for each model.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 832−840
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7335

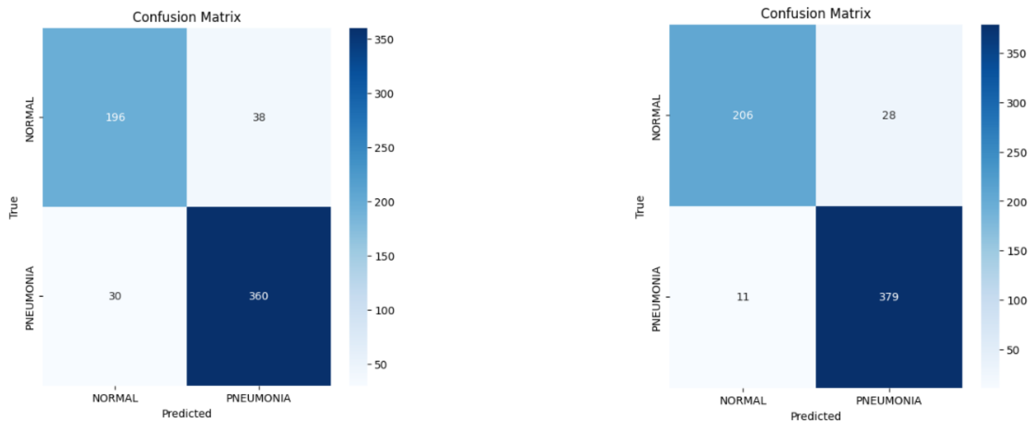


(a) (b)

Figure 3. Accuracy graph of (a) InceptionV3 and (b) DenseNet121

Based on Figure 3 (a), the training accuracy graph shows a steady increase from the beginning to the end of the training process. The validation accuracy, however, exhibits significant fluctuation throughout training, with several sharp drops and spikes. In the loss graph, the training loss decreases rapidly and consistently from start to finish. Similarly, the validation loss also decreases quickly at the beginning, mirroring the trend of the training loss. After several epochs, the validation loss begins to fluctuate, although not as drastically as the validation accuracy.

Based on Figure 3 (b), the training accuracy graph shows a steady increase from the beginning to the end of the training process. The validation accuracy begins to fluctuate after approximately 10 epochs, with notable variations, particularly between epoch 10 and 15. During this phase, there are several sharp drops and spikes in validation accuracy. In the loss graph, the training loss decreases rapidly and consistently from the start to the end of training. Similarly, the validation loss also decreases rapidly at the beginning, mirroring the trend of the training loss. After several epochs, the validation loss begins to fluctuate, although these fluctuations are not as pronounced as those seen in the validation accuracy.



(a) (b)

Figure 4. Confusion matriks of (a) InceptionV3 and (b) DenseNet121

Based on Figure 4 (a), the confusion matrix result from the InceptionV3 model show that model correctly predicted 196 cases as normal, which are categorized as True Negatives (TN). A total of 38 normal cases were misclassified as pneumonia, representing the False Positive (FP). Additionally, 30 pneumonia cases were incorrectly predicted as normal, classified as False Negatives (FN). Meanwhile, the model successfully identified 360 pneumonia cases correctly, which fall under the True Positives (TP).

Based on Figure 4 (b), the confusion matrix result from the DenseNet121 model indicate that the model correctly classified 196 cases as normal, which corresponds to the True Negatives (TN). A total of 38 normal cases were incorrectly predicted as pneumonia, classified as False Positives (FP). In addition, 30 pneumonia cases were misclassified as normal, which are considered False Negatives (FN). Meanwhile, the model accurately identified 360 pneumonia cases, falling under the True Positive (TP).

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 832−840
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7335

Classification Report:

	precision	recall	f1-score
NORMAL	0.87	0.84	0.85
PNEUMONIA	0.90	0.92	0.91
accuracy			0.89
macro avg	0.89	0.88	0.88
weighted avg	0.89	0.89	0.89

(a)

Classification Report:

	precision	recall	f1-score
NORMAL	0.95	0.88	0.91
PNEUMONIA	0.93	0.97	0.95
accuracy			0.94
macro avg	0.94	0.93	0.93
weighted avg	0.94	0.94	0.94

(b)

Figure 5. Classification report for (a) InceptionV3 (b) DenseNet121

Based on Figure 5 (a), the classification report from the InceptionV3 model demonstrates notable performance in distinguishing between normal and pneumonia classes. In terms of precision, the model achieved a score of 0.87 for the normal class, indicating that 87% of normal predictions were correct, and 0.90 for the pneumonia class, meaning that 90% of pneumonia predictions were accurate. Regarding recall, the model successfully identified 84% of actual normal cases and 92% of actual pneumonia cases. Defined as the harmonic mean, the F1-score balances precision and recall, was 0.85 for normal and 0.91 for pneumonia.

Based on Figure 5 (b), the DenseNet121 model demonstrates excellent classification performance based on the evaluation metrics. For precision, it achieved a score of 0.95 for the normal class, meaning 95% of all predicted normal cases were correct, and 0.93 for the pneumonia class, indicating that 93% of pneumonia predictions were accurate. Regarding recall, the model successfully identified 88% of actual normal cases and 97% of actual pneumonia cases. The F1-score, which reflects the harmonic mean of precision and recall, reached 0.91 for normal and 0.95 for pneumonia.

3.2 Discussion

The Support Vector Machine (SVM) model performed very well on the validation data with 94% accuracy. This shows that the model is able to learn patterns from the training data well. The model performance decreased on the test data, especially in recall, for the normal case. This suggests that the model may experience overfitting on the training data or lack of generalization on data not seen before. The SVM model performed very well in detecting pneumonia cases (TP = 388). This can be seen from the much larger number of TPs than FNs. Challenges in Detecting Normal Cases: The model had difficulty in detecting normal cases, with a high number of FPs (FP = 115). This indicates that the model tends to misclassify normal cases as pneumonia.

In the Inceptionv3 model, significant fluctuations are present in the validation accuracy graph indicating potential overfitting. Although the validation loss did not show severe overfitting. The InceptionV3 model showed good performance in differentiating normal from pneumonia cases. The model performed very well in detecting pneumonia cases (TP = 360). This can be seen from the much larger number of TPs than FNs. The model had difficulty in detecting normal cases, with a fairly high number of FPs (FP = 38). This indicates that the model tends to misclassify some normal cases as pneumonia. The InceptionV3 model showed good performance with 89% accuracy. High Performance in Detecting Pneumonia: The model had high precision and recall for pneumonia cases, indicating a good ability to detect pneumonia. The model had slightly lower precision and recall for normal cases compared to pneumonia, indicating room for improvement in detecting normal cases.

The DenseNet121 model has a graph that shows a steep drop during training and validation loss indicating that the model is learning well. The fluctuations in the validation loss, although not very large, indicate that there is noticeable variability in validation performance. The DenseNet121 model showed excellent performance in differentiating normal from pneumonia cases. A true positive count of 379 reflects the model's strong capability in detecting pneumonia cases. This can be seen from the much larger number of TPs than FNs. The model performed reliably in classifying normal cases (TN = 206), while keeping false positives to a minimum.

Table 2. Result of SVM and CNN Models (InceptionV3 and DenseNet121)

Models		Classification Result			
		Accuracy	Precision	Recall	F1-Score
SVM		0.81	0.85	0.81	0.79
CNN	InceptionV3	0.89	0.89	0.89	0.89
	DenseNet121	0.94	0.94	0.94	0.94

Table 2 shows a comparative performance of three pneumonia classification models based on chest X-ray images. The SVM model shows the lowest performance compared to other models. The test accuracy is 81%, with a precision of 85%, a recall of 81%, and an F1-score of 79%. This confirms the previous conclusion that SVM may experience overfitting or a lack of generalization on the test data. Although it has a low recall in detecting normal cases, SVM has a very high recall value in pneumonia detection. The InceptionV3 model shows a significant increase in performance compared to SVM. The test accuracy is 89%, with a precision of 89%, a recall of 89%, and an F1-score of 89%. This confirms the conclusion that CNN (including InceptionV3) is able to learn complex features from chest X-ray images better. InceptionV3 maintains relatively stable value across all evaluation metrics. In comparison to the other three models, DenseNet121 yielded the best results. Its test accuracy was 94%, with 94% precision, 94% recall, and 94% F1-score. This confirms the conclusion that DenseNet121 has great potential in classifying pneumonia from chest X-ray images. DenseNet121 reached the peak score compared to the other models, and has very stable values in each of its metrics.

4. CONCLUSION

Based on the results, it can be inferred that Convolutional Neural Networks (CNN) models, especially DenseNet121 and InceptionV3, perform far better than the Support Vector Machine (SVM) on pneumonia diagnosis of classified chest X-ray images. DenseNet121 yielded the best accuracy, precision, recall, and F1 score of 94%, with InceptionV3 following close at 89%, while SVM scored poorly on all parameters, at 81%. CNNs are advantageous for automatically extracting complex features from imbalanced datasets, though they work well with large datasets and require high computational capabilities. On the other hand, SVM is reasonably efficient in terms of resources and works effectively in high-dimensional spaces; however, it is incapable of learning complex visual features on its own without extensive feature engineering. The findings point toward the fact that CNN-based models seem more appropriate for automated diagnosis from medical images. Future work should explore the use of alternative pre-trained models such as ResNet or EfficientNet and conduct hyperparameter optimization to further enhance model performance and generalizability.

REFERENCES

[1] N. P. Ekananda and D. Riminarsih, "Identifikasi Penyakit Pneumonia Berdasarkan Citra Chest X-ray Menunggunakan Convolutional Neural Network" *Jurnal Ilmiah Informatika Komputer*, vol. 27, no. 1, pp. 79–94, 2022, doi: 10.35760/ik.2022.v27i1.6487.

[2] I. Md. D. Maysanjaya, "Klasifikasi Pneumonia pada Citra X-rays Paru-paru dengan Convolutional Neural Network," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 9, no. 2, pp. 190–195, May 2020, doi: 10.22146/jnteti.v9i2.66.

[3] World Health Organization, "Pneumonia in children." [Online]. Available : https://www.who.int/news-room/fact-sheets/detail/pneumonia

[4] M. M. Eid and Y. H. Elawady, "Efficient Pneumonia Detection for Chest Radiography Using ResNet-Based SVM," *European Journal of Electrical Engineering and Computer Science*, vol. 5, no. 1, pp. 1–8, Jan. 2021, doi: 10.24018/ejece.2021.5.1.268.

[5] B. Nugroho and E. Yulia, "Kinerja Metode CNN untuk Klasifikasi Pneumonia dengan Variasi Ukuran Citra Input," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 8, no. 3, pp. 533–538, 2021, doi: 10.25126/jtiik.202184515.

[6] E. Putra and A. Meindiawan, "Application of MobileNetV2 and SVM Combination for Enhanced Accuracy in Pneumonia Classification," *Journal of Applied Informatics and Computing (JAIC)*, Vol.8, No.2, pp. 332~340, December 2024, doi: 10.30871/jaic.v8i2.8426. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC

[7] M. J. Sulistio, C. Lubis, and K. Kunci, "Implementasi CNN dan MobileNet untuk Mendeteksi Penyakit Pneumonia dan COVID-19 dengan Menggunakan Aplikasi Smartphone," *Nusantara Journal of Multidisciplinary Science*, vol. 1, no. 4, 2023, [Online]. Available: https://jurnal.intekom.id/index.php/njms

[8] A. W. Salehi *et al.*, "A Study of CNN and Transfer Learning in Medical Imaging: Advantages, Challenges, Future Scope," *Sustainability*, vol. 15, no. 7, p. 5930, Mar. 2023, doi: 10.3390/su15075930.

[9] I. A. Firdaus, "Deteksi Infeksi Mycoplasma Pneumoniae Pneumonia Menggunakan Komparasi Algoritma Klasifikasi Machine Learning," *JOINTECS (Journal of Information Technology and Computer Science)*, vol. 7, no. 1, p. 35, Feb. 2022, doi: 10.31328/jointecs.v7i1.3242.

[10] I. Jawaz and Reni Rahmadewi, "Sistem Deteksi Pneumonia Paru-Paru dengan Pengolahan Citra Digital dan Machine Learning," *ELECTRON Jurnal Ilmiah Teknik Elektro*, vol. 5, no. 1, May 2024, doi: 10.33019/electron.v5i1.114.

[11] O. Ozaltin and O. Yeniay, "ECG Classification Performing Feature Extraction Automatically Using a Hybrid CNN-SVM Algorithm," in *2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, IEEE, Jun. 2021, pp. 1–5. doi: 10.1109/HORA52670.2021.9461295.

[12] O. J. Egwom, M. Hassan, J. J. Tanimu, M. Hamada, and O. M. Ogar, "An LDA–SVM Machine Learning Model for Breast Cancer Classification †," *BioMedInformatics*, vol. 2, no. 3, pp. 345–358, Sep. 2022, doi: 10.3390/biomedinformatics2030022.

[13] R. A. Wati, H. Irsyad, and M. E. Al Rivan, "Classification of Pneumonia Using Support Vector Machine," *Jurnal Algoritme*, vol. 1, no. 1, pp. 21–32, Oct. 2020, doi: 10.35957/algoritme.v1i1.429.

[14] N. Amaya-Tejera, M. Gamarra, J. I. Vélez, and E. Zurek, "A distance-based kernel for classification via Support Vector Machines," *Front Artif Intell*, vol. 7, 2024, doi: 10.3389/frai.2024.1287875.

[15] A. R. Pratama and A. F. Cobantoro, "Klasifikasi Citra Pneumonia Menggunakan Arsitektur Convolutional Neural Network (CNN)," *Networking Engineering Research Operation*, vol. 8, no. 2, pp. 133–144, Dec. 2023, doi: 10.21107/nero.v8i2.18992.

[16] I. Bakti and M. Firdaus, "Arsitektur CNN InceptionResNet-V2 Untuk Pengelompokan Pneumonia Chest X-Ray," *Jurnal Komputer dan Teknologi*, pp. 35–42, Jan. 2023, doi: 10.58290/jukomtek.v1i2.66.

[17] S. A. Widiarto, W. A. Saputra, and A. R. Dewi, "Klasifikasi Citra X-ray Toraks Dengan Menggunakan Contrast Limited Adaptive Histogram Equalization dan Convolutional Neural Network (Studi Kasus: Pneumonia)," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 6, no. 2, pp. 348–359, Nov. 2021, doi: 10.29100/jipi.v6i2.2102.

[18] I. Bakti, "Klasifikasi Pneumonia Chest X-Ray Dengan Arsitektur Inceptionresnet-V2," *Jurnal Teknologi Informasi (JUTECH)*, vol. 3, no. 2, pp. 107–121, Dec. 2022, doi: 10.32546/jutech.v3i2.2030.
[19] R. Kaur, R. Kumar, and M. Gupta, "Review on Transfer Learning for Convolutional Neural Network," in *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, IEEE, Dec. 2021, pp. 922–926. doi: 10.1109/ICAC3N53548.2021.9725474.
[20] M. Ahmed, N. Afreen, M. Ahmed, M. Sameer, and J. Ahamed, "An inception V3 approach for malware classification using machine learning and transfer learning," *International Journal of Intelligent Networks*, vol. 4, pp. 11–18, 2023, doi: 10.1016/j.ijin.2022.11.005.
[21] A. Fauziah, H. Hermanto, and M. A. Sukmarini, "Extreme Gradient Boosting pada Peramalan Pola Curah Hujan Bulanan Kabupaten Banyuwangi," *Jurnal Kridatama Sains dan Teknologi*, vol. 6, no. 02, pp. 430–440, Jul. 2024, doi: 10.53863/kst.v6i02.1154.
[22] W. Wijiyanto, A. I. Pradana, S. Sopingi, and V. Atina, "Teknik K-Fold Cross Validation untuk Mengevaluasi Kinerja Mahasiswa," *Jurnal Algoritma*, vol. 21, no. 1, May 2024, doi: 10.33364/algoritma/v.21-1.1618.
[23] M. I. Anugrah, J. Zeniarja, and D. S. Setiawan, "Peningkatan Performa Model Hard Voting Classifier dengan Teknik Oversampling ADASYN pada Penyakit Diabetes," *Edumatic: Jurnal Pendidikan Informatika*, vol. 8, no. 1, pp. 290–299, Jun. 2024, doi: 10.29408/edumatic.v8i1.25838.
[24] R. S. Y. P. Errissya Rasywir1, "Analisis dan Implementasi Diagnosis Penyakit Sawit dengan Metode Convolutional Neural Network (CNN) ," *Paradigma - Jurnal Komputer dan Informatika*, vol. 21, no. 2, Sep. 2020, doi:10.31294/p.v21i2.
[25] D. Marcella and S. Devella, "Klasifikasi Penyakit Mata Menggunakan Convolutional Neural Network Dengan Arsitektur VGG-19," *Jurnal Algoritme,* vol. 3, no. 1, pp. 60–70, 2022, doi:10.35957/algoritme.v3i1.3331.
[26] W. Ayadi, W. Elhamzi, I. Charfi, and M. Atri, "Deep CNN for Brain Tumor Classification," *Neural Process Lett*, vol. 53, no. 1, pp. 671–700, Feb. 2021, doi: 10.1007/s11063-020-10398-2.