Volume 3 Nomor 1 (2023) Pages 229 – 238
International Journal of Bunga Bangsa Cirebon (IJOBBA)
Email Journal : ijobba.bbc@gmail.com
Web Journal : http://journal.bungabangsacirebon.ac.id/index.php/ijobba

Comparison of Pneumonia Detection Using Convolutional Neural Network (CNN) Method on Balanced and Imbalanced Datasets

Nurhadiansyah[1], Ahmad Faqih[2]✉, Fidya Arie Pratama[3], Muhammad Afeefee Assalihee[4]

Affiliate
Institut Prima Bangsa [1], STMIK IKMI Cirebon[2], Universitas Bunga Bangsa Cirebon[3], Prince Songkla University Thailand[4]

Email : ahmadfaqih367@gmail.com[2]

Received: 2024-01-15; Accepted: 2024-02-28; Published: 2024-02-28

Abstract

Pneumonia is a serious respiratory disease that requires early detection to improve clinical outcomes and reduce mortality rates. Radiologists usually perform pneumonia detection through X-ray images, but this method is prone to human error and delays in diagnosis. Therefore, there is an urgent need to develop automated technology-based systems that can assist in the rapid and accurate diagnosis of pneumonia. This study aims to develop a Convolutional Neural Network (CNN) model to detect pneumonia in X-ray images. One of the main challenges in developing this model is the imbalance in the dataset, where the number of X-ray images labeled "pneumonia" may be significantly fewer compared to images labeled "normal." Dataset imbalance can cause the model to be biased toward the majority class and reduce detection accuracy. The CNN architecture was first tested on an imbalanced dataset. The test results of a 2-layer convolutional CNN achieved an accuracy of 93.10%, while a 3-layer convolutional CNN achieved an accuracy of 96.03%. Subsequently, the CNN architecture was tested on a balanced dataset using traditional augmentation methods. The test results of a 2-layer convolutional CNN achieved an accuracy of 93.84%, while a 3-layer convolutional CNN achieved an accuracy of 96.46%. It can be concluded that the CNN architecture using a balanced dataset has better accuracy compared to using an imbalanced dataset.

Keywords: *Deep Learning, traditional augmentation, Convolutional Neural Network (CNN)*

INTRODUCTION

Respiratory diseases, including pneumonia, represent a significant global health issue. Pneumonia, often caused by bacterial, viral, or fungal lung infections, can be life-threatening and requires immediate medical attention. Early detection is crucial to improve recovery chances and reduce negative impacts. Advances in image processing and deep learning technology in recent years have supported pneumonia detection. Recent studies in scientific journals reveal the potential of these methods. Research in the "Journal of Medical Imaging and Health Informatics" highlights the importance of image-processing technology in the diagnosis of pneumonia (Wang et al., 2020). This study demonstrates the challenges in disease diagnosis due to striking histopathological similarities between different but related conditions. There is an urgent need for computational methods that can help clinicians translate heterogeneous biomedical images into accurate and quantitative diagnoses. This need is particularly important in cases of small intestinal enteropathy, such as Environmental Enteropathy (EE) and Celiac Disease (CD). This research extends the initial analysis by developing an artificial intelligence (AI)--based image analysis platform that uses Convolutional Neural Networks (CNN) for these enteropathy cases. Data were obtained from three major studies at various locations. The image analysis platform for EE and CD was developed using a multi-zoom architecture CNN. Gradient-weighted Class Activation Mappings (Grad-CAMs) were used to visualize the decision-making process of the model in classifying diseases. A team of medical experts reviewed color-stained normalized biopsy images to reduce bias and Grad-CAMs to confirm structural continuity and biomedical relevance. The study involved 461 high-resolution biopsy images from 150 children with a median age of 37.5 months and nearly equal gender distribution (77 males, 51.3%). The ResNet50 and Shallow CNN models showed detection accuracy of 98% and 96%, respectively, which increased to 98.3% with the use of an ensemble. Grad-CAMs demonstrated the model's ability to learn various distinct microscopic morphological features for EE, CD, and controls.

Several risk factors that increase the risk of pneumonia infection include advanced age, smoking habits, environmental exposure, malnutrition, a history of previous pneumonia, chronic bronchitis, asthma, functional disorders, poor oral hygiene, the use of immunosuppressive therapy, oral steroids, and acid secretion inhibitors (Gabruseva et al., 2020). Rapid identification of suspected COVID-19 patients is crucial for controlling this disease. This study aims to develop a chest CT-based deep-learning algorithm for rapid triage in fever clinics. In an external validation set, using radiology reports as the reference standard, the AI triage achieved an area under the curve of 0,953 (95% CI 0,949–0,959), with a sensitivity of 0,923 (95% CI 0,914–0,932), specificity of 0,851 (95% CI 0,842–0,860), positive predictive value of 0,790 (95% CI 0,777–0,803), and negative predictive value of 0,948 (95% CI 0,941–0,954). This deep-learning algorithm has been developed and externally validated for the triage of COVID-19 patients in fever clinics. With high accuracy across various populations with diverse COVID-19 prevalence, integrating this system into standard clinical workflows can accelerate the identification of chest CT images indicative of COVID-19 (Akinyelu & Blignaut, 2022).

Findings from several research journals indicate that the diagnosis of pneumonia on X-ray images still heavily relies on the skills of doctors and radiologists, which can lead to human errors and delays in diagnosis. Deep learning technology, particularly the Convolutional Neural

Network (CNN) method, has great potential in pneumonia detection, although its use has not yet been fully optimized. Early detection of pneumonia is crucial as it significantly impacts patient recovery and treatment. Therefore, the development of an efficient and accurate automated system for detecting pneumonia is highly necessary.

This research aims to develop a Convolutional Neural Network (CNN) model to detect pneumonia in X-ray images by first balancing the dataset. Pneumonia is a serious respiratory disease that requires early detection to improve clinical outcomes and reduce mortality rates. Pneumonia detection through X-ray images is usually performed by radiologists, but this method is prone to human error and delays in diagnosis. One of the main challenges in developing this model is dataset imbalance, where the number of X-ray images labeled "pneumonia" may be significantly fewer compared to images labeled "normal." Dataset imbalance can cause the model to be biased towards the majority class and reduce detection accuracy.

This research employs traditional data augmentation to enhance the variation and quantity of data in the minority class. Traditional augmentation techniques include rotation, flipping, zooming, and translation of images. The CNN model will be designed with convolutional, pooling, and fully connected layers, then compiled with an appropriate loss function and effective optimizer. The model will be trained using training and validation data, and evaluated with test data to measure performance through metrics such as accuracy, sensitivity, specificity, and predictive value

This research is expected to produce a CNN model capable of detecting pneumonia in X-ray images with high accuracy and minimal bias, making it an effective diagnostic tool to support radiologists and expedite the pneumonia diagnosis process.

METHODOLOGY

This research initially adopts an experimental approach to establish the methodological foundation, which is then further developed through case studies to collect and analyze data. The data mining method used is Knowledge Discovery in Databases (KDD), a process that has become a standard in data analysis for identifying valuable patterns from complex datasets. KDD serves as the primary guide in this research, directing the process from data selection and preprocessing to model building and evaluation.

The implemented case study focuses on the problem of pneumonia detection using chest X-ray images. The initial step involves collecting a sufficient and representative dataset from diverse sources, ensuring a variety of pneumonia cases and normal conditions. The data is then prepared by performing normalization, noise removal, and resolution adjustment to ensure consistent quality before being fed into the analysis process.

The application of KDD involves the following steps: first, data selection and preprocessing to transform raw data into a format suitable for further analysis. Second, modeling using Convolutional Neural Network (CNN) techniques specifically designed for pneumonia detection from X-ray images. The CNN model is constructed with convolutional layers to extract important features from the images, followed by pooling layers to reduce data dimensions, and fully connected layers for final classification.

Finally, model evaluation is conducted using separate validation data to measure performance based on metrics such as accuracy, sensitivity, specificity, and positive and negative predictive values. The results of this research are expected to provide valuable guidance for the development of more efficient and accurate automated detection systems for pneumonia based on X-ray images..

RESULTS AND DISCUSSION

The dataset used for the experiments was sourced from www.kaggle.com and consists of 5216 X-ray images divided into two classes. The first class includes 1341 images showing normal lungs, while the second class consists of 3875 images showing lungs with pneumonia. Below are examples of images from both classes:

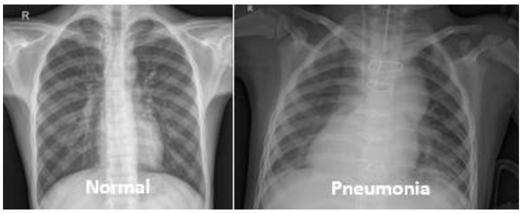

Image 1: X-ray of Normal Lungs and Pneumonia
source : https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia?resource=download

Dataset

The dataset used to train the CNN model was found to be imbalanced. Therefore, special treatment was required to balance the dataset. Data augmentation was performed using traditional methods. The dataset was expanded through traditional augmentation techniques, which include:

1. *Random rotation up to 10 degrees, both clockwise and counterclockwise.*
2. *Zooming within a range of 1 to 1.5.*
3. *Shifting height and width by (-10,10) pixels.*

Each image in the normal lung class was augmented three times. Initially, there were 1341 images in the normal lung class. After augmenting each image three times, the total increased to 4023 images. The final total for the normal lung class, after traditional augmentation, is 5364 images. Similarly, the pneumonia lung class had 497 images, which were also augmented three times, resulting in a total of 5364 images. Below are samples of the output from the traditional augmentation process.

CNN Architecture

In the initial phase of CNN model development in this research, both a 2-layer convolution and a 3-layer convolution model were proposed and compared in terms of accuracy.

The 2-layer convolution CNN architecture performs convolutional processes twice. The first convolutional process uses 32 filters with a 3x3 kernel. The second convolutional process uses 64 filters with a 3x3 kernel. Max-pooling is applied twice with a 2x2 matrix.

The 3-layer convolution CNN architecture performs convolutional processes three times. The first convolutional process uses 16 filters with a 3x3 kernel. The second convolutional process uses 32 filters with a 3x3 kernel. The third convolutional process uses 64 filters with a 3x3 kernel. Max-pooling is applied three times with a 2x2 matrix.

Subsequently, the architecture with the highest accuracy was selected for further development and is referred to as CNN 1.

CNN Architecture Test Results

The CNN architectures tested include a 2-layer convolution architecture and a 3-layer convolution architecture. These architectures were proposed by the author. They were tested on two datasets: imbalance and balance datasets

. • Imbalance Dataset

The first architecture tested was the 2-layer convolution architecture. This architecture was applied to the imbalance dataset and achieved an accuracy of 93.10%, a precision of 1, a recall of 1, and an F1-Score of 0.91.

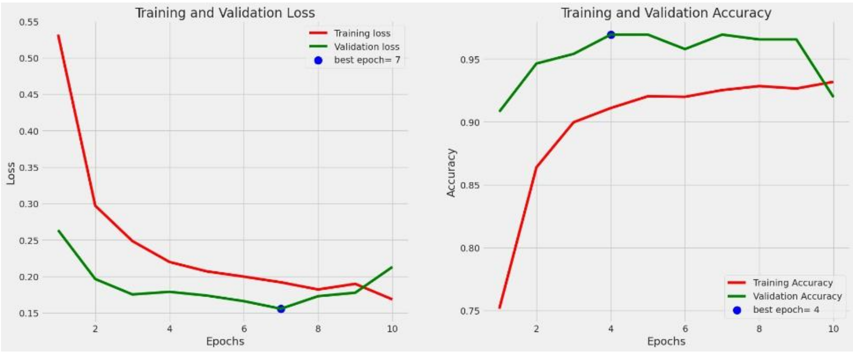

Image 2: Graph of the epoch and loss progression to achieve the best accuracy for the 2-layer convolution CNN architecture on the imbalance dataset

To ensure the accuracy of training and validation results, Figure 4.7 illustrates the epoch progression for the author's proposed CNN architecture. The training accuracy turns out to be higher than the validation accuracy, but the difference is insignificant, so the resulting model is still categorized as a fit model. The next architecture tested was the 3-layer convolution architecture. This architecture was applied to the imbalance dataset and achieved an accuracy of 96.03%, a precision of 0.99, a recall of 0.93, and an F1-Score of 0.93.

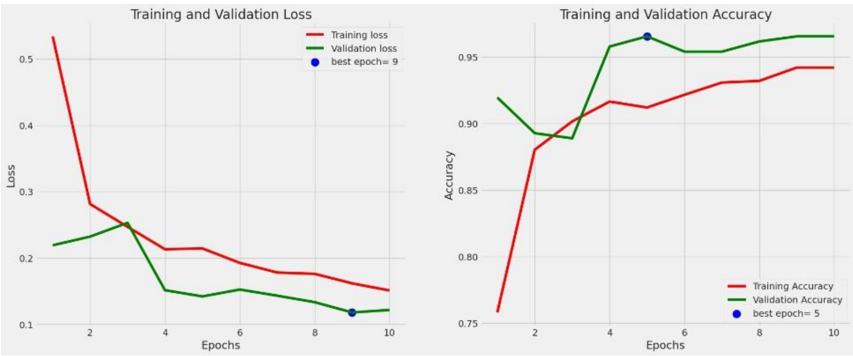

Image 3: Graph of the epoch and loss progression to achieve the best accuracy for the 3-layer

convolution CNN architecture on the imbalance dataset

To ensure the accuracy of training and validation results, Figure 4.9 illustrates the epoch progression for the author's proposed CNN architecture. The training accuracy is lower than the validation accuracy, indicating that the model obtained is an overfitting model.

• Balanced Dataset Using Traditional Augmentation

Next, these architectures were applied to the balanced dataset. The transformation from an imbalance to a balanced dataset was done using data augmentation. The data augmentation used is traditional augmentation. This dataset consists of a total of 10,768 images. There are 5,364 normal lung images and 5,364 pneumonia lung images. The 2-layer convolution

CNN architecture was applied to this balanced dataset. This architecture achieved an accuracy of 93.82%, a precision of 1.0, a recall of 1.0, and an F1-Score of 0.94.

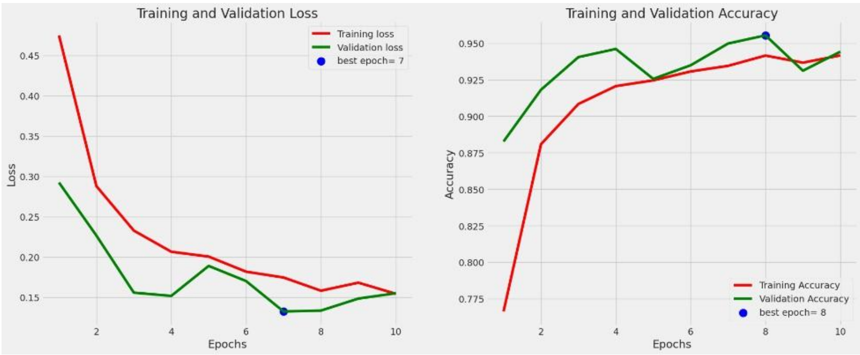

Image 4: Graph of the epoch and loss progression to achieve the best accuracy for the 2-layer convolution CNN architecture on the balanced dataset using traditional augmentation

The training accuracy turns out to be almost the same as the validation accuracy. This means that the model obtained is a fit model. The decrease in loss and the increase in accuracy occur gradually.

Next, the 3-layer convolution CNN architecture was applied to the balanced dataset. This architecture achieved an accuracy of 96.46%, a precision of 0.94, a recall of 0.93, and an F1-Score of 0.95.

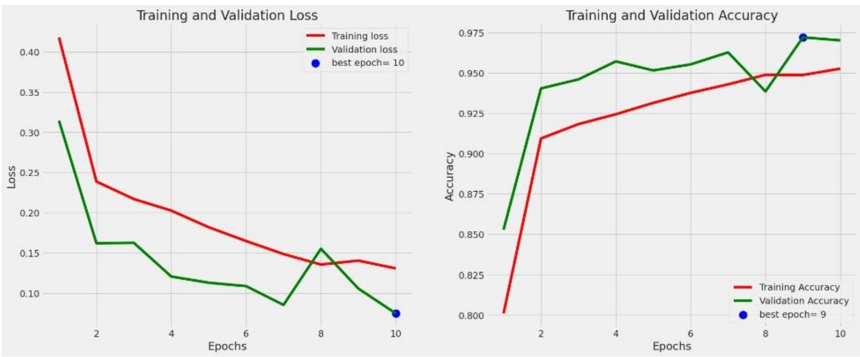

Image 5: Graph of the epoch and loss progression to achieve the best accuracy for the 3-layer convolution CNN architecture on the balanced dataset

The training accuracy is lower than the validation accuracy. However, the difference is still small, so the model obtained is categorized as a fit model. Accuracy improvement occurs gradually, with the best accuracy achieved at epoch 9.

The development of the CNN architecture began with testing the 2-layer convolution and 3-layer convolution configurations. After testing these architectures, their accuracy values were compared to determine the best architecture. These architectures were first tested on the imbalance dataset. The results showed that the 2-layer convolution CNN achieved an accuracy of 93.10%, while the 3-layer convolution CNN achieved an accuracy of 96.03%. Table 1 provides a comparison of the layer usage on the imbalance dataset.

Table 1: Results of testing accuracy for layer usage on the imbalance dataset

Layer	**Akurasi**
2	93,10%
3	96,03%

From Table 1, it is evident that the highest accuracy among the two architectures is achieved by the 3-layer convolution CNN. Based on these results, the chosen architecture is the 3-layer convolution CNN.

Next, these architectures were tested again on the balanced dataset. Initially, the architectures were tested on the balanced dataset using traditional augmentation methods. The results showed that the 2-layer convolution CNN achieved an accuracy of 93.84%, while the 3-layer convolution CNN achieved an accuracy of 96.46%. Table 2 shows the comparison of layer usage on the balanced dataset with traditional augmentation.

Table 2: Results of testing accuracy for layer usage on the balanced dataset with traditional augmentation

Layer	**Tradisional**
2	93.84%
3	96.46%

From Table 2, the comparison of accuracy for the balanced dataset using traditional augmentation at different layer configurations is shown. The 3-layer convolution CNN has higher accuracy compared to the 2-layer convolution CNN on the balanced dataset with traditional augmentation. Based on these results, it can also be concluded that the chosen architecture is the 3-layer convolution CNN

CONCLUSION

From the results of the proposed architecture testing, the 3-layer convolution CNN architecture has higher accuracy on both the imbalance and balanced datasets compared to the

2-layer convolution CNN. Additionally, the CNN architecture using the balanced dataset demonstrates better accuracy compared to using the imbalance data.

BIBLIOGRAPHY

Akinyelu, A. A., & Blignaut, P. (2022). COVID-19 diagnosis using deep learning neural networks applied to CT images. *Frontiers in Artificial Intelligence*, *5*. https://doi.org/10.3389/frai.2022.919672

Anshori, R. B., Tsp, H. F., & Siadari, T. S. (2022). Klasifikasi Citra Kanker Serviks Menggunakan Deep Residual Network. *E-Proceeding of Engineering*, *8*(6), 3163–3170.

Gabruseva, T., Poplavskiy, D., & Kalinin, A. (2020). Deep learning for automatic pneumonia detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, *2020-June*, 1436–1443. https://doi.org/10.1109/CVPRW50498.2020.00183

Gholamy, A., Kreinovich, V., & Kosheleva, O. (2018). Why 70/30 Or 80/20 Relation Between Training And Testing Sets : A Pedagogical Explanation. *Departmental Technical Reports (CS)*, *1209*, 1–6.

Gracia Moisés, A., Vitoria Pascual, I., Imas González, J. J., & Ruiz Zamarreño, C. (2023). Data Augmentation Techniques for Machine Learning Applied to Optical Spectroscopy Datasets in Agrifood Applications: A Comprehensive Review. *Sensors*, *23*(20), 8562. https://doi.org/10.3390/s23208562

Handayani, P., Susanto, T., Djamaris, A., & Novianti, M. D. (2022). Penerapan Data Mining untuk Menganalisis Faktor-Faktor yang Mempengaruhi Terhadap Kenaikan Harga Minyak Goreng. *Management, and Industry (JEMI)*, *05*(04), 251–268.

Hasanah, U., Dyah, Y., & Santik, P. (2021). Faktor Intrinsik dan Extrinsik yang Berhubungan dengan Kejadian Pneumonia di Wilayah Puskesmas Rembang. *Jurnal Kesehatan Masyarakat Indonesia*, *16*, 84. https://jurnal.unimus.ac.id/index.php/jkmi,

Husna, M., Pertiwi, F. D., & Nasution, A. S. (2022). FAKTOR-FAKTOR YANG BERHUBUNGAN DENGAN KEJADIAN PNEUMONIA PADA BALITA DI PUSKESMAS SEMPLAK KOTA BOGOR 2020. *Jurnal Media Kesehatan*, *5*(3), 127–133. https://doi.org/10.33088/jmk.v9i2.303

Motamed, S., Rogalla, P., & Khalvati, F. (2021). Data augmentation using Generative Adversarial Networks (GANs) for GAN-based detection of Pneumonia and COVID-19 in chest X-ray images. *Informatics in Medicine Unlocked*, *27*(October). https://doi.org/10.1016/j.imu.2021.100779

Mulyawan, H., Samsono, M. Z. H., & Setiawardhana. (2011). Identifikasi Dan Tracking Objek Berbasis Image Processing Secara Real Time. *EEPIS Repository*, 1–5. http://repo.pens.ac.id/1324/1/Paper_TA_MBAH.pdf

Natasya, F. A. (2022). Tatalaksana Pneumonia. *Jurnal Medika Hutama*, *03*(02), 2392–2399.

Nugroho, P. A., Fenriana, I., & Arijanto, R. (2020). Implementasi Deep Learning Menggunakan Convolutional Neural Network (CNN) Pada Ekspresi Manusia. *Algor*, *2*(1), 12–21.

Prasetio, R. T., & Ripandi, E. (2019). Optimasi Klasifikasi Jenis Hutan Menggunakan Deep Learning Berbasis Optimize Selection. *Jurnal Informatika*, *6*(1), 100–106. https://doi.org/10.31311/ji.v6i1.5176

Putra, F. A., Irawan, B., & Dinimaharawati, A. (2023). Pengenalan Ekspresi Wajah Menggunakan Arsitektur Xception Pada Algoritma Convolutional Neural Network (CNN). *E-Proceeding of Engineering*, *10*(1), 651–660.

Sandhopi, Lukman Zaman P.C.S.W, & Yosi Kristian. (2020). Identifikasi Motif Jepara pada Ukiran dengan Memanfaatkan Convolutional Neural Network. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, *9*(4), 403–413. https://doi.org/10.22146/jnteti.v9i4.541

Setiawan, W. (2022). *Deep Learning menggunakan Convolutional Neural Network*. Media Nusa Creative.

Sun, Y., Wong, A. K. C., & Kamel, M. S. (2009). Classification of imbalanced data: A review. *International Journal of Pattern Recognition and Artificial Intelligence*, *23*(4), 687–719. https://doi.org/10.1142/S0218001409007326

Suprapti, T., Dian Ade Kurnia, Doni Anggara, Rananda Deva Rian, & Aldi Setiawan. (2023). Implementasi Model Algoritma Generative Adversarial Network (Gan) Pada Sistem Presensi Berbasis Deteksi Wajah (SIDEWA). *Tematik*, *9*(2), 231–236. https://doi.org/10.38204/tematik.v9i2.1048

Wang, M., Xia, C., Huang, L., Xu, S., Qin, C., Liu, J., Cao, Y., Yu, P., Zhu, T., Zhu, H., Wu, C., Zhang, R., Chen, X., Wang, J., Du, G., Zhang, C., Wang, S., Chen, K., Liu, Z., … Wang, W. (2020). Deep learning-based triage and analysis of lesion burden for COVID-19: a retrospective study with external validation. *The Lancet Digital Health*, *2*(10), e506–e515. https://doi.org/10.1016/S2589-7500(20)30199-0

Yuhandri, Y., Ramadhanu, A., & Syahputra, H. (2022). Pengenalan Teknologi Pengolahan Citra Digital (Digital Image Processing) Untuk Santri Di Rahmatan Lil'Alamin International Islamic Boarding School. *Community Development Journal : Jurnal Pengabdian Masyarakat*, *3*(2), 1239–1244. https://doi.org/10.31004/cdj.v3i2.5868

Yuli Mardi. (2019). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*, *2*(2), 213–219.