

Addressing Noise Challenges in CNN-based Pneumonia Detection: A Study using Primary Indonesian Thoracic Imagery

Wahyu Andi Saputra¹, Andi Prademon Yunus^{1*}

¹Telkom University, Banyumas, Indonesia

Email: andiwahyu@telkomuniversity.ac.id, andiaiy@telkomuniversity.ac.id

Abstract

Accurate pneumonia diagnosis is vital, especially in resource-limited areas like Indonesia. While CNNs show promise for automated detection using chest X-rays, real-world image quality affects their performance. This study addresses this challenge by using a primary dataset—images directly from Indonesian patients—to avoid the biases of pre-processed secondary data. This ensures our findings are relevant to the Indonesian context. We tested how different noise types (salt-and-pepper and Gaussian) impact the accuracy of several common CNN architectures. These noise types mimic common image imperfections. Our analysis reveals that noise degrades the CNN's ability for 3% to 5% performance. This highlights the need for better pre-processing methods and potentially specialized CNN designs to handle noisy images. Ultimately, our work improves our understanding of deploying CNNs for pneumonia diagnosis in real-world settings, leading to more reliable and helpful diagnostic tools. Using primary data from diverse populations is crucial for building trustworthy AI in healthcare.

Keywords: pneumonia classification, CNN, salt-and-pepper noise, gaussian noise

1. Introduction

Pneumonia remains a significant global health challenge, particularly in resource-limited areas such as Indonesia, where access to advanced diagnostic tools and medical expertise is often restricted. The World Health Organization (WHO) identifies pneumonia as one of the leading causes of mortality in children under five years old, underscoring the urgent need for efficient and accurate diagnostic methods [1]. In such contexts, the development and deployment of automated diagnostic systems, particularly those leveraging deep learning technologies like Convolutional Neural Networks (CNNs), offer a promising solution to enhance healthcare delivery and outcomes.

Convolutional Neural Networks have emerged as a powerful tool in the field of medical imaging, capable of learning complex patterns and features directly from raw data. Their application in diagnosing diseases from chest X-rays has been extensively researched, with numerous studies demonstrating their potential to achieve diagnostic accuracy comparable to that of human experts [2]. CNNs are particularly attractive due to their ability to process large datasets and their adaptability to various imaging modalities. However, the performance of CNNs is heavily contingent on the quality of the input data, a factor that poses significant challenges in real-world clinical settings.

In clinical environments, especially those with limited resources, the quality of medical images can be compromised by several factors, including equipment limitations, varying imaging protocols, and

environmental conditions [3][4]. These factors often introduce noise and artifacts into the images, which can adversely affect the performance of CNNs. Common types of noise encountered in medical imaging include salt-and-pepper noise, characterized by random occurrences of black and white pixels, and Gaussian noise, which manifests as variations in pixel intensity following a Gaussian distribution. Such noise can obscure critical features necessary for accurate diagnosis, leading to potential misclassification and reduced diagnostic reliability.

To ensure robust model performance, this study utilizes primary Indonesian patient image data, prioritizing authenticity over the potential biases of secondary datasets. However, the inherent limitations of primary data necessitate data augmentation. Augmentation artificially expands the training dataset, creating a diverse training environment for the CNN [5], [6]. This mitigates overfitting and improves the model's ability to generalize to unseen data, ultimately enhancing its reliability in real-world Indonesian clinical settings. The combined approach addresses both data quality and model generalizability.

The central focus of this study is to quantify the impact of different noise types on the accuracy of CNN architectures commonly used for pneumonia diagnosis. By simulating noise conditions that mimic real-world image imperfections, we aim to systematically evaluate how these disturbances affect CNN performance. The findings from our analysis highlight the critical need for improved pre-processing methods that can effectively mitigate the impact of noise on medical images. Techniques such as noise reduction algorithms, image

enhancement, and normalization can play a vital role in preserving the integrity of image features essential for accurate diagnosis. Additionally, there is a compelling case for the development of specialized CNN architectures that incorporate mechanisms for noise robustness, such as attention mechanisms or noise-resistant feature extraction layers.

By enhancing our understanding of how noise affects CNN performance, this study contributes to the broader goal of deploying reliable and effective AI tools in healthcare settings. The insights gained can inform the design of diagnostic systems that are not only accurate but also resilient to the variability and imperfections inherent in clinical imaging. This is particularly important in regions like Indonesia, where the successful integration of AI technologies into healthcare systems can significantly improve diagnostic capacity and patient outcomes.

The use of primary data from diverse populations is a crucial step toward building trustworthy AI systems in healthcare. It ensures that the models developed are representative and applicable to the populations they are intended to serve. This approach aligns with the broader objectives of ethical AI development, which prioritize transparency, fairness, and inclusivity. By focusing on primary datasets, this study addresses the critical issue of data representativeness, thereby enhancing the credibility and acceptance of AI-driven diagnostic tools in clinical practice.

This study provides valuable insights into the challenges and opportunities associated with using CNNs for pneumonia diagnosis in resource-limited settings. By systematically evaluating the impact of noise on CNN performance and emphasizing the importance of primary data, we lay the groundwork for future research and development efforts aimed at creating robust, reliable, and contextually relevant AI solutions for healthcare. The findings not only advance the field of medical imaging but also contribute to the broader discourse on the responsible and effective deployment of AI technologies in global health.

2. Noise Emergence in Primary Dataset Acquisition

The retrieval of primary datasets in medical imaging, particularly for tasks such as pneumonia classification, is inherently susceptible to the presence of noise. This noise can arise from a multitude of sources, each contributing to the degradation of image quality and the subsequent performance of Convolutional Neural Networks (CNNs). In the context of chest X-ray images, several key factors contribute to the emergence of noise.

One of the primary causes of noise in radiographic images is the inherent physical properties of the imaging process itself. For instance, in conventional X-ray techniques, secondary radiation and scattered radiation from the patient can introduce significant noise into the images. Additionally, digital radiography systems are prone to noise from various elements of the system, including the CCD camera, imaging screen, X-ray source,

and controller circuits. These sources can lead to the presence of Poisson noise, salt and pepper noise, and speckle noise, which manifest as random fluctuations in pixel intensity or as white and black pixels scattered across the image. Furthermore, electronic interferences in receiver circuits and radiofrequency emissions due to thermal effects can also contribute to noise in modalities like MRI. Fig. 1 illustrates the presence of noise based on different milliamperere-seconds (mAs) of the x-ray image asset [7].

The impacts of noise on the performance of CNNs in pneumonia classification are multifaceted and profound. Noise can obscure critical features necessary for accurate diagnosis, leading to misclassification and reduced diagnostic reliability. For example, salt and pepper noise can introduce random black-and-white pixels that may mimic or obscure pathological features, while Gaussian noise can blur the image, reducing the contrast and making it harder for the CNN to distinguish between different regions of the lung. This degradation in image quality can result in a significant drop in the accuracy of CNN models. Studies have shown that even moderate levels of noise can reduce the performance of CNNs by 2.5-5%, highlighting the need for robust pre-processing techniques and noise-robust learning methods to mitigate these effects.



Fig. 1 Noise presence on high milliamperere-second (left) and low milliamperere-second (right) [7]

Moreover, the presence of noise can exacerbate the issue of label noise, where incorrect annotations due to noise can further complicate the training process of CNNs. Label noise is a significant challenge in medical image analysis, as it can lead to overfitting to noisy labels, thereby reducing the generalizability and reliability of the models. The inter-observer variability among medical experts can also compound this issue, making it essential to develop methods that can estimate and mitigate noise rates in the dataset. By understanding the causes and impacts of noise, researchers can develop more effective strategies for noise reduction and robust learning, ultimately enhancing the performance and reliability of CNNs in real-world clinical settings.

In conclusion, the emergence of noise in primary dataset retrieval is a critical issue that must be addressed to ensure the accurate and reliable classification of pneumonia images using CNNs. By identifying the sources of noise and understanding its impacts on image quality and model performance, researchers can develop tailored approaches to mitigate these effects. This

includes the use of advanced noise reduction techniques, noise-robust learning methods, and sparse regularization to enhance the robustness of CNN models against various types of noise. Such efforts are crucial for the deployment of trustworthy AI-driven diagnostic tools in clinical environments.

3. Dataset's Noise Simulation

3.1. Dataset Collection

The primary dataset utilized in this study was meticulously curated from thoracic X-ray images obtained from three prominent medical institutions in Banyumas, Indonesia: RSUD Banyumas, RSUD Margono, and Klinik Utama Kesehatan Paru Kabupaten Banyumas. This dataset is composed of 4,156 images classified as NORMAL and 5,391 images classified as PNEUMONIA, forming a substantial and diverse collection that significantly enhances the robustness and generalizability of the study. The images were originally stored in both JPEG and DICOM formats, which are widely recognized standards in medical imaging due to their ability to preserve high-quality image data and associated metadata. To streamline the machine learning training process, all images were systematically converted into JPEG format (Fig. 2). This conversion was executed with meticulous attention to minimizing compression artifacts, ensuring that the clinical integrity and fine-grained details of the original images were retained. This step is of paramount importance, as excessive compression could obscure subtle yet clinically significant features that are critical for accurate pneumonia detection.

The pneumonia thoracic X-ray images in the dataset are stored in RGB format, with each file averaging approximately 500 KB in size. Despite being rendered in grayscale, these images exhibit a range of patterns indicative of pneumonia, such as consolidations, interstitial opacities, and other pathological findings. The majority of the images depict a standard thoracic view, focusing on the chest region of the patients. However, the dataset also includes a subset of images with unique characteristics that reflect the variability encountered in real-world clinical practice. These include the presence of a black border surrounding the thoracic area, variations in thoracic shape—particularly noticeable in pediatric cases where the lower neck and diaphragm are visible—and instances of skewing in thoracic images. These variations are not mere anomalies but rather represent the inherent diversity of clinical imaging, which poses significant challenges for automated diagnostic systems.



Fig. 2 Sample of Primary dataset acquired from medical facility in Indonesia

3.2. Data Cleaning and Preprocessing

Data cleaning and preprocessing are critical steps in preparing the dataset for analysis. To simulate noise, we artificially introduced known quantities of salt-and-pepper and Gaussian noise to clean images. This controlled simulation is essential for understanding how different types of noise affect the performance of CNNs.

For salt-and-pepper noise, we set the density to 5% and 15%, with densities above 15% considered to result in broken or damaged images. This range allows us to evaluate the impact of moderate to severe salt-and-pepper noise, which is commonly encountered in medical imaging due to factors such as sensor quality and transmission issues. The parameters for Gaussian noise were generated using the default settings from the Albumentation library, ensuring a realistic and consistent simulation of Gaussian noise.

This controlled simulation enables us to systematically evaluate the impact of noise on CNN performance, providing insights into how different noise types and intensities affect the accuracy and reliability of the models. By introducing noise in a controlled manner, we can isolate the effects of noise from other variables, thereby gaining a deeper understanding of the robustness of CNNs in real-world scenarios.

3.3. Noise Simulation Accomplishment

The implementation of noise simulation was carried out using the Albumentation library, a powerful tool for data augmentation and simulation. Here is a detailed overview of the steps involved:

Salt-and-Pepper Noise Simulation

Illustrated in Fig. 3, we used the SaltAndPepper transform from Albumentation to introduce salt-and-pepper noise at specified densities (5% and 15%). This process involved randomly selecting pixels and altering them to black or white, simulating the real-world noise conditions that can arise from various sources such as sensor malfunctions or data transmission errors. The random selection of pixels ensures that the noise is distributed uniformly across the image, mimicking the unpredictable nature of real-world noise.

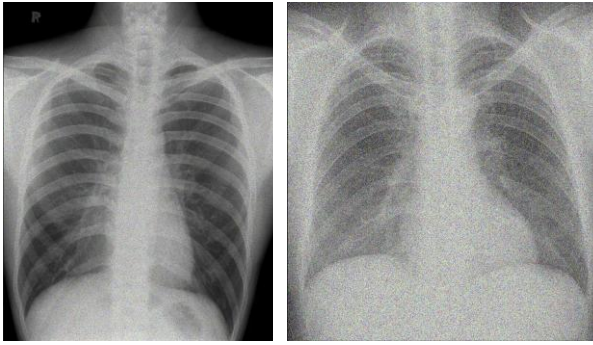


Fig. 3 Salt-and-Pepper simulation: 5% (left) and 15% (right)

GaussNoise Noise Simulation

The GaussNoise transform from Albumentation was employed to introduce Gaussian noise (Fig. 4). The default parameters of the GaussNoise transform were used to generate Gaussian noise, ensuring a realistic simulation of noise conditions that can occur due to factors like sensor imperfections or environmental interference. The use of default parameters helps in maintaining consistency and realism in the simulation, allowing for a more accurate assessment of how Gaussian noise affects CNN performance. By following these steps, we ensured that the noise simulation was both realistic and controlled, providing a robust framework for evaluating the impact of noise on CNN performance. This systematic approach to noise simulation is essential for developing CNN models that are resilient to the various types of noise encountered in real-world medical imaging.



Fig. 4 Gaussian Noise Simulation using kernel size: 0.5 (left) and 3 (right)

4. Experimental Evaluation dan Discussion

4.1. Experimental Evaluation

This experimental evaluation assesses the impact of various noise types on the performance of a Convolutional Neural Network (CNN) designed for pneumonia classification. The results organize the experimental data into five distinct noise categories: 'saltpepper05', 'gaussian3', 'gaussian05', 'non_augmented', and 'saltpepper15'. Each category represents a specific noise augmentation applied to the pneumonia image dataset.

Analysis of the training accuracy across epochs (Fig. 5) reveals varying trends depending on the noise type. For instance, the 'gaussian3' noise condition demonstrates a relatively consistent increase in training accuracy over the 25 epochs, suggesting the CNN adapts well to this type of noise. Conversely, the 'saltpepper05' and 'saltpepper15' conditions show more fluctuating accuracy, indicating a potentially more challenging adaptation process for the CNN. The 'non_augmented' condition also exhibits a steady increase in training accuracy, serving as a control for comparison. A detailed graphical representation of these trends would provide a clearer visualization of the learning curves for each noise type.

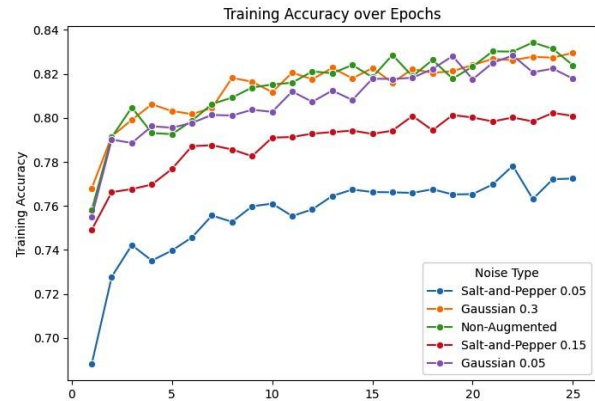


Fig. 5 Training accuracy over epochs

Similarly, the validation accuracy across epochs exhibits diverse patterns. The 'gaussian3' noise condition shows a steady increase in validation accuracy, indicating good generalization performance. The 'non_augmented' condition also displays a generally upward trend, although with some minor fluctuations. The 'saltpepper' conditions, however, show less consistent improvement in validation accuracy, suggesting potential overfitting to the training data in the presence of salt and pepper noise. Again, a visual representation of these validation curves would enhance the analysis (Fig. 6).

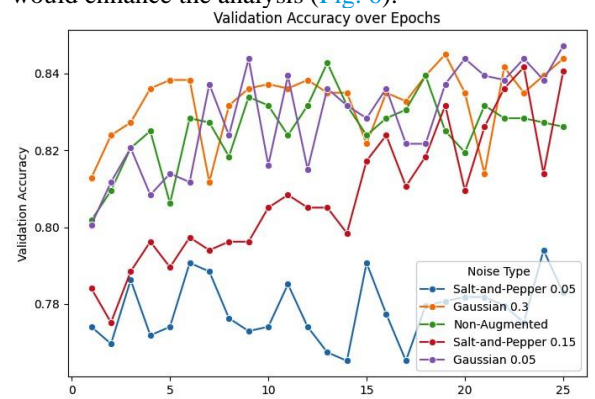


Fig. 6 Validation accuracy over epochs

Comparing the final training accuracies across the four noise conditions reveals that the 'gaussian3' noise condition achieves the highest accuracy, followed by 'non_augmented', 'saltpepper15', and 'saltpepper05'. This suggests that Gaussian noise, at the applied level, has a less detrimental effect on the CNN's learning process

compared to salt and pepper noise. The difference in performance between the two salt and pepper noise levels ('saltpepper05' and 'saltpepper15') also warrants further investigation (Fig. 7).

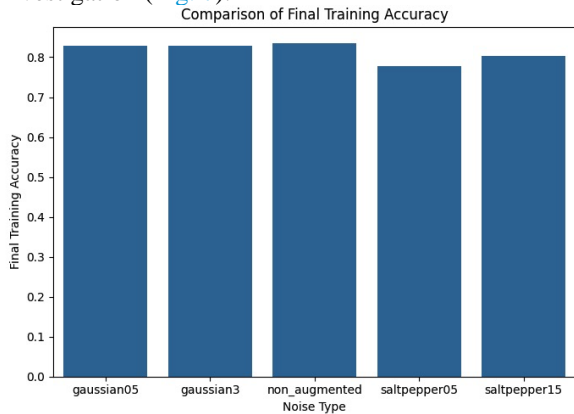


Fig. 7. Comparison of final training accuracy

A comparison of the final validation accuracies mirrors the trend observed in the training accuracies (Fig. 8). The 'gaussian3' condition achieves the highest validation accuracy, followed by 'non_augmented', 'saltpepper15', and 'saltpepper05'. This consistency between training and validation accuracies for the Gaussian noise condition reinforces the conclusion that the CNN generalizes well in Gaussian noise. However, the lower validation accuracies for the salt and pepper noise conditions highlight the challenges posed by this type of noise to the CNN's generalization ability. Further analysis, including statistical significance testing, would strengthen these observations.

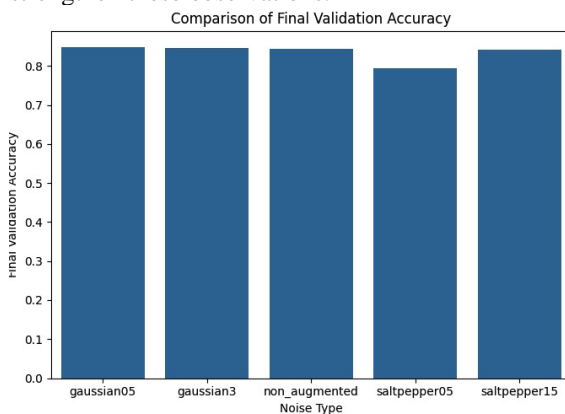


Fig. 8. Comparison of final validation accuracy

4.2. Discussion and Implications

The experimental results provide valuable insights into the effects of noise on CNN-based pneumonia classification. The most striking observation is the stark contrast between the impact of salt-and-pepper noise and Gaussian noise. Salt-and-pepper noise, with its abrupt and extreme pixel value changes, severely degrades the model's performance, leading to lower accuracy and validation accuracy. This type of noise disrupts the spatial structure of the images, making it difficult for CNN to

extract meaningful features. In contrast, Gaussian noise, which introduces smoother and more gradual variations, has a less detrimental effect and may even improve the model's robustness by simulating real-world variations in image quality. The lower training time observed in the Gaussian noise experiments further supports its suitability as a data augmentation technique, as it does not impose a significant computational overhead.

These findings have important implications for the design and implementation of CNN-based systems for medical image analysis, particularly in scenarios where image quality may be compromised by noise. While clean data remains the gold standard for achieving optimal performance, the use of Gaussian noise as a data augmentation strategy can help improve the model's robustness to real-world noise without significantly degrading accuracy. On the other hand, salt-and-pepper noise should be avoided or mitigated, as it has a pronounced negative impact on model performance. Future work could explore additional noise types and levels, as well as advanced noise reduction techniques, to further enhance the robustness of CNN-based pneumonia classification systems. Overall, this study underscores the importance of carefully considering the type and intensity of noise when designing and training deep learning models for medical image analysis.

Conclusion

This study quantitatively assessed the impact of salt-and-pepper and Gaussian noise on a CNN's pneumonia classification performance. Salt-and-pepper noise (0.05 and 0.15 intensity) significantly reduced accuracy (e.g., 0.778 training accuracy at 0.05 intensity vs. 0.828 for clean data), while training time remained consistent (250-260 seconds/epoch). Conversely, Gaussian noise (standard deviation 0.035) yielded comparable accuracy (0.829) to clean data, with a substantially reduced training time (90-100 seconds/epoch). These findings highlight the differential effects of noise types on CNN performance, suggesting Gaussian noise augmentation as a viable strategy to enhance robustness while emphasizing the need for salt-and-pepper noise mitigation in medical image analysis.

Acknowledgement

The successful completion of this research was contingent upon the collaborative efforts of Klinik Utama Kesehatan Paru Kabupaten Banyumas, RSUD Prof. Dr. Margono Soekarjo, and RSUD Banyumas, who provided crucial access to and support in the acquisition and analysis of the primary thoracic image dataset. Their medical and administrative staff's contributions were instrumental. Furthermore, the financial backing of the Direktorat Riset, Teknologi, dan Pengabdian kepada Masyarakat (DRTPM), Kemendikbudristek RI, enabled the execution of the research and the generation of the reported results.

References

1. “Effectiveness of lung ultrasonography for diagnosis of pneumonia in adults: a systematic review and meta-analysis - PMC.” Accessed: Feb. 04, 2025. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC5107554/>
2. S. Bashir, F. Firdous, S. Z. Rufai, and R. Bawa, “Chest disease detection from x-ray using machine learning: A review,” presented at the International Conference On Intelligent And Smart Computation (ICIASC-2023), Mohali, India, 2024, p. 020029. doi: 10.1
3. M. Bielecka, A. Bielecki, R. Obuchowicz, and A. Piórkowski, “Universal Measure for Medical Image Quality Evaluation Based on Gradient Approach,” in Computational Science – ICCS 2020, vol. 12140, V. V. Krzhizhanovskaya, G. Závodszy, M. H. Lees, J. J. Dong E. A. Krupinski, “Current perspectives in medical image perception,” *Atten. Percept. Psychophys.*, vol. 72, no. 5, pp. 1205–1217, Jul. 2010, doi: 10.3758/APP.72.5.1205.
4. K. Lakiotaki, N. Vorniotakis, M. Tsagris, G. Georgakopoulos, and I. Tsamardinos, “BioDataome: a collection of uniformly preprocessed and automatically annotated datasets for data-driven biology,” Database, vol. 2018, Jan. 2018, doi: 10.1093/database/bay01
6. M. Xu, S. Yoon, A. Fuentes, and D. S. Park, “A Comprehensive Survey of Image Augmentation Techniques for Deep Learning,” May 2022, [Online]. Available: <http://arxiv.org/abs/2205.01491>
7. W. Huda and R. B. Abrahams, “Radiographic Techniques, Contrast, and Noise in X-Ray Imaging,” *Am. J. Roentgenol.*, vol. 204, no. 2, pp. W126–W131, Feb. 2015, doi: 10.2214/AJR.14.13116.

Authors Introduction

Wahyu Andi Saputra



He obtained his Master degree from Universitas Gadjah Mada and is currently a Lecturer at the Department of Informatics from Telkom University, Indonesia. He is conducting research in the field of medical image processing since 2015.

Andi Prademon Yunus



He completed his Ph.D. degree at Mie University in 2023. Currently, he works at Telkom University, Indonesia as a Lecturer. His main research focus on the motion analysis and human behavioral computing. He excels on computer vision and human-computer interaction through his research experiments.