Jurnal Mantik, 7 (3) (2023), ISSN 2685-4236 (Online)

Published by:Institute of Computer Science (IOCS)

Jurnal Mantik

Journal homepage: www.iocscience.org/ejournal/index.php/mantik

Batch size and learning rate effect in covid-19 classification using CNN

Ni Komang Rai Mirayanti[1], Sariyasa[2], I Gede Aris Gunadi[3]

[1,3] S2 Ilmu Komputer, Pascasarjana, S2 Ilmu Komputer, Universitas Pendidikan Ganesha, Indonesia

[2] S2 Ilmu Komputer, Pascasarjana, S2 Ilmu Komputer, Universitas Pendidikan Ganesha, Indonesia

ARTICLE INFO

Article history:

Received Sep 18, 2023
Revised Sep 25, 2023
Accepted Oct 02, 2023

Keywords:

Batch Size;
CNN;
Covid-19;
K-fold validation;
Learning Rate.

ABSTRACT

This study evaluated the performance of a Convolutional Neural Network (CNN) model in classifying CT-Chest images of COVID-19 and non-COVID patients. The primary focus was to determine the influence of learning rate and batch size on the model's effectiveness. This research used 698 datas from Covid-19 and NonCovid-19 CT-Chest Image. Those dataset was obtained from medRxiv dan bioRxiv and has been approved by radiology expert in Tongji Hospital, China. In this research, COVID-16 dataset was classified by CNN with different batch sizes and learning rates for each iteration. Batch size used in this study were 1, 2, 4, 8, 16, 32, 64, and 128 with learning rates 0,00001; 0,0001; 0,001; 0,01; 0,1 and 1. This study found that this study showed that batch size and learning rate have a positive effect on CNN performance. The lower the learning rate, the lower the batch size will allow the network in the CNN model to perform better in classifying COVID-19 CT-Chest. In Addition, the best batch size for the classification is 64 with learning rate 0,01. These findings provided important insights into how parameters such as learning rate and batch size impacted the performance of the CNN model in classifying COVID-19 CT-Chest.

Corresponding Author:

Ni Komang Rai Mirayanti,
S2 Ilmu Komputer, Pascasarjana, S2 Ilmu Komputer,
Universitas Pendidikan Ganesha,
No. 11 Jalan. Udayana, Buleleng, Bali, 81116, Indonesia.
Email: kimangraimirayanti@gmail.com

1. INTRODUCTION

The advent of deep learning techniques, particularly Convolutional Neural Networks (CNN) has revolutionized various fields, including medical image analysis (Yamashita et al., 2018). The global health crisis caused by COVID-19 has especially benefited from these advancements. As of November 22, 2020, Indonesia has reported 493,000 COVID-19 cases, with 15,774 deaths, underscoring the need for effective diagnostic methods. In Indonesia, multiple COVID-19 tests are available, including Antigen Tests, Polymerase Chain Reaction (PCR), and Tes Cepat Molekuler (TCM) (Sugihantono et al., 2020). Besides these methods, CT-Chest imaging has proven valuable for diagnosing COVID-19 by identifying signs like ground-glass opacity patterns in the lungs (Susilo et al., 2020).

Ground-glass opacity is a radiological term that indicates an area in the lungs that is blurred or foggy so as to increase lung opacity (Infante et al., 2009). Previous research has highlighted CNN potential in medical image classification, including lung diseases and other respiratory conditions4. A literature review (Halmar, 2020) suggested that CT-Chest methods have higher sensitivity in diagnosing COVID-19 compared to PCR in early phases. A previous study employed CNN and achieved high classification accuracies, albeit with varying sensitivity and specificity rates (Swastika, 2020). Other works have also explored different CNN-based methodologies and achieved promising results (Hariyani et al., 2020). Despite the promise shown by CNN, there is an ongoing debate about optimizing model parameters for best performance, specifically concerning learning rate and batch size (Chen et al., 2023). There has been no previous research that specifically examines the impact of batch size and learning on CNN performance in classifying COVID-19 CT-Chest images. This study systematically investigating the impact of varying these hyperparameters on the performance of a CNN model explicitly built for classifying COVID-19 and non-COVID CT-Chest images, offering a more nuanced understanding of their impact on model performance, which is essential for fine-tuning CNN in the medical imaging domain. This study conducted by a series of experiments using different learning rates (0.00001, 0.0001, 0.001, 0.01, 0.1, 1) and batch sizes (1, 2, 4, 8, 16, 32, 64, 128) to ascertain the model's efficacy.

By doing so, this study seek to offer new insights into how varying these parameters influences CNN performance in COVID-19 CT-Chest imaging and to identify the best configurations leading to optimal performance. The Classification result will be analysed using some performance metrics. Previous studies often rely solely on accuracy, sensitivity and specificity as the primary performance metrics. However, in medical diagnostics, a more comprehensive evaluation is essential. Along with these metrics, this study employs a broader range of performance metrics, including precision, and F2 score. This expanded set of metrics provides better assessment of model performance regarding to its ability to correctly identify COVID-19 cases and avoid false positives. These metrics is calculated from values of the True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). This study gave particular emphasis to Sensitivity as a performance metric. Given that the dataset comprises CT-Chest images for COVID-19 diagnosis, high Sensitivity is crucial to ensure that as many actual cases of the disease are correctly identified as possible. Missing a positive COVID-19 case could have serious health implications, not only for the individual but also for public health due to the contagious nature of the virus. This study also incorporates k-fold cross-validation. This method divides the dataset into k subsets and conducts training and validation iteratively, ensuring that each subset serves as both training and validation data. This approach enhances the robustness and minimizing the risk of overfitting. By addressing these gaps in previous research and introducing novel elements, this study provides insights into the theoretical aspects of CNN model optimization and broader understanding of machine learning optimization strategies. This study offers valuable knowledge for future investigations in fine-tuning CNN and other deep learning models especially for COVID-19 CT-Chest images classification. In practical terms, this research holds substantial significance, particularly within the healthcare sector and the context of COVID-19 diagnosis. By optimizing batch size and learning rate in CNN, this study offer a valuable tool for healthcare professionals. The enhanced accuracy in COVID-19 diagnosis through CT-Chest images as demonstrated in this study, provide timely and reliable diagnoses. The emphasis on sensitivity as a performance metric and the addition of precision and F2-Score broaden the potential to identify more COVID-19 cases accurately, contributing significantly to public health efforts.

2. RESEARCH METHOD

The primary objective of this study was to evaluate the performance of a Convolutional Neural Network (CNN) model in classifying CT-Chest images of COVID-19 and non-COVID cases. Specifically, the study was designed to examine the effects of varying the learning rate and batch size on the model's performance metrics, such as accuracy, sensitivity, specificity, precision, and F2 score and to find the best learning rate and batch size values that leading to optimal performance.

This study was done repeatedly using different batch sizes and learning rates for each iteration. In addition, the experiment was carried out using k-fold validation with a value of k = 10. From each classification, TP (True Positive), TN (True Negative), FP (False Positive) and FN (False Negative) values were generated. These values then be used to calculate the values of accuracy, recall (sensitivity), specificity and precision. In general, the flow of this research can be seen from Figure 1,

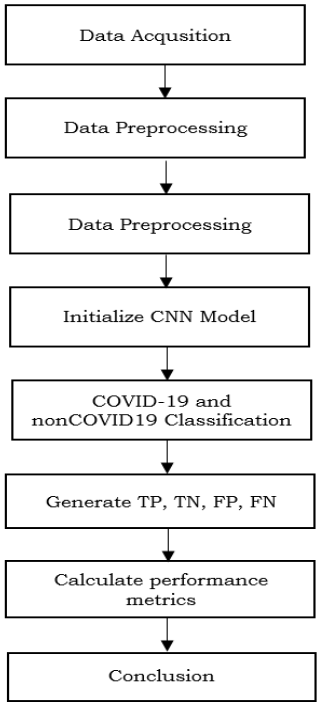

Figure 1. Flow of the research

The CNN classification model in this study used the DeepConvNet architecture and the Adam optimization algorithm as an optimizer with 10 epochs. The independent variables in this study were batch size and learning rate, while the dependent variables were accuracy, recall (sensitivity), specificity, precision and F2 score. These values will then be analyzed to find the best learning rate and batch size values that leading to optimal performance and to find the effects of varying the learning rate and batch size on the model's performance metrics. The following procedural steps were taken to achieve the research objectives:

2.1 Data Collection

The study leveraged a pre-existing dataset made up of CT-Chest images specifically labeled as COVID-19 and non-COVID. This dataset was rigorously curated from multiple reliable sources to include a diverse range of cases, enhancing the robustness of the study and was sourced from a reputable healthcare institution and consists of 746 labeled images, with each image having a dimension of 512x512 pixels. Rigorous quality control was performed on the dataset to ensure that the images met medical standards for clarity, contrast, and relevance. A total of 746 images were collected, split evenly between COVID-19 and non-COVID samples to ensure a balanced dataset and minimize any potential bias. Sample of the dataset used in this study can be seen in Figure 1.

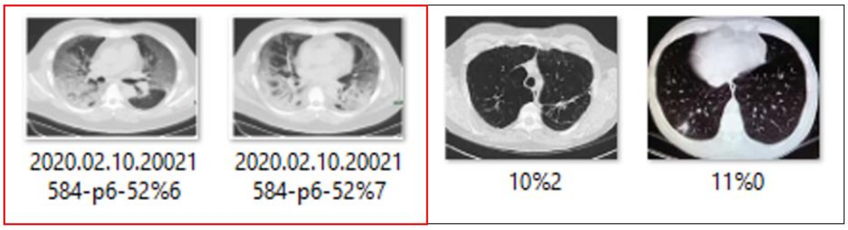

Figure 2. Sample of COVID-19 and nonCOVID-19 dataset

This dataset was obtained from COVID-19 related papers (Yang et al., 2020) and had been confirmed by senior radiologists at Tongji Hospital, China who carried out the diagnosis and treatment of COVID-19 patients during the outbreak. The CT-Chest of a Covid-19 patient will show something blurry called GGO (Ground Glass Opacity) covering the patient's lung area as seen in the red marked part in Figure 1, while the rest are the CT-chest of nonCOVID-19 patients.

2.2 Data Preprocessing

The first step in data preprocessing was to equalize the number of samples across labels to minimize data imbalance. Balancing data is crucial to avoid misconceptions when evaluating machine learning algorithm performance (Mooijman et al., 2023) which can lead to misleading performance metrics (Hensman & Masko, 2015). In this study, we aimed to overcome this issue by equalizing the number of samples across labels. Initially, the dataset contained 746 samples of data with COVID and non-COVID label. This dataset were randomly selected to match the number of COVID-labeled samples at 349. Hence, the total dataset size became 698 samples (2x349). Secondly, the images were resized to a uniform dimension of 128x128 pixels to make them compatible with the model and to speed up the computation (Goh et al., 2023), thereby maintaining consistency across the dataset (Parsania & Virparia, 2018). Thirdly, the pixel values were normalized to fall within the range of 0 and 1, rather than the typical 0-255, to facilitate faster and more stable training (Pei et al., 2023). These preprocessing steps were performed to maintain computational efficiency and improve the model's ability to learn the key features from the data (Jo, 2019). This normalization is crucial for speeding up the training process and avoiding potential numerical instability during backpropagation (Huang et al., 2023).

2.3 Model Architecture

The Convolutional Neural Network (CNN) model developed for this study featured a series of layers to facilitate image classification for COVID-19 and non-COVID cases using CT-Chest images and was carefully designed to maximize efficiency and effectiveness. It comprised three main convolutional layers, each equipped with a varying number of filters for intricate feature extraction. Specifically, the architecture comprised convolutional layers, pooling layers, activation layers, dropout layers, and fully connected layers (Hidaka & Kurita, 2017). The convolutional layers are essential for feature detection which is extract patterns from local regions of the input images, while pooling layers reduce dimensionality while retaining important features (Eom & Choi, 2021). ReLU (Rectified Linear Unit) was chosen as the activation function for the convolutional layers due to its proven efficacy in alleviating the vanishing gradient problem or non-linearity into the model (Alzubaidi et al., 2023)(Alzubaidi et al., 2023). Finally, the model was compiled using the Adam optimizer, a popular optimization algorithm widely adopted in deep learning applications for both computer vision and natural language processing. Moreover Adam optimizer are most commonly used to handle the medical images (Hassan et al., 2023). This optimizer combines the advantages of two other extensions of stochastic gradient descent: AdaGrad and RMSProp. Adam optimizer renowned for its efficient and adaptive learning rate management (Reyad et al., 2023).

2.4 Parameter Tuning

In this experiment, an extensive range of learning rates and batch sizes were investigated to scrutinize their potential impact on the Convolutional Neural Network (CNN) model's performance in classifying CT-Chest images into COVID-19 and non-COVID categories. The learning rates evaluated were 0.00001, 0.0001, 0.001, 0.01, 0.1, and 1, while the batch sizes varied from 1, 2, 4, 8, 16, 32, 64 to 128. These values were meticulously chosen based on a comprehensive review of existing literature as well as preliminary empirical tests (Sido & Konopik, 2019). The selection aimed to encapsulate a broad range of configurations, thereby providing a comprehensive analysis of how these hyperparameters could affect the model's training and subsequent performance.

The objective of such extensive hyperparameter tuning was not only to optimize the model's performance but also to gain insights into the trade-offs involved in selecting different learning rates and batch sizes. For instance, lower learning rates generally require more epochs to converge but can potentially lead to a more precise model, whereas higher learning rates may speed up the training process but risk overshooting the optimal solution (Jepkoech et al., 2021). Similarly, smaller batch sizes provide a more accurate estimation of the gradient but at the cost of increased computational time and resources (Bordelon et al., 2022), while larger batch sizes make the optimization process more computationally efficient but might lead to less accurate gradient estimates. Through this rigorous approach, the study aimed to identify which combinations of learning rates and batch sizes were most conducive to achieving a balance between computational efficiency and gradient quality, without compromising the ability of the CNN model to accurately classify the CT-Chest images.

2.5 Training and Validation

In this study, the dataset that has processed and consisted of 698 samples with 349 samples for each COVID and non-COVID label was split into training and validation sets using an 80:20 ratio. This proportion was applied to each label, meaning that 80% of the samples from each label were allocated to the training set, and the remaining 20% to the validation set. Specifically, the training set comprised 558 samples (279 data from each label), while the validation set included 140 samples (70 data from each label). This 80:20 split was chosen to maximize the training data while allowing for robust validation. It is noteworthy that the 80:20 ratio was chosen based on the premise that larger training datasets generally lead to better machine learning models (He et al., 2021). However, the proportion between training and validation sets may vary depending on the dataset and the specific task at hand.

To further ensure the model's robustness and generalizability, a K-Fold Cross-Validation technique was applied with k=10. In this approach, the training data was divided into 10 different subsets or 'folds.' The model underwent 10 rounds of training, each time excluding one of these subsets for validation while utilizing the remaining nine for training. This method assures that every data sample in the training set appeared at least once in a validation set (Cerqueira et al., 2020). The choice of k=10 was motivated by its common usage in machine learning for achieving a good balance between computational cost and model performance. Moreover, using 10 folds allowed for a comprehensive evaluation, as each fold had the opportunity to be a part of the validation set at least once. This exhaustive validation process served the study's primary objective well: investigating the effects of various batch sizes and learning rates on the CNN model's performance.

2.6 Performance Metrics

In alignment with the research's primary objective to investigate the influence of various batch sizes and learning rates on the CNN model, a meticulous evaluation was conducted using a comprehensive set of metrics including accuracy, sensitivity, specificity, precision, and F2 score. These metrics were chosen to provide a comprehensive understanding of the model's ability to identify both classes (COVID-19 and non-COVID) correctly. These metrics are important metrics in the medical domain (Althnian et al., 2021). They were calculated at each epoch during the validation phase to monitor the model's performance dynamically.

A. Accuracy : Measures the proportion of correct predictions in the total prediction set, providing a holistic view of the model's capability in classifying samples correctly.

$$Accuracy = \frac{(TP + TN)}{TP + FP + FN + TN} \quad (1)$$

B. Sensitivity (Recall): Particularly important for medical applications, sensitivity gauges the model's ability to correctly identify positive cases of COVID-19 among the dataset, thereby minimizing false negatives.

$$Recall = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Negative\ (FN)} \quad (2)$$

C. Specificity: Complements sensitivity by evaluating the model's performance in correctly identifying non-COVID cases, effectively minimizing false positives.

$$Specificity = \frac{True\ negative\ (TN)}{True\ negative\ (TN) + False\ positive\ (FP)} \quad (3)$$

D. Precision: Provides an understanding of how many of the predicted positive COVID-19 cases were actually positive, helping to mitigate the risk of false alarms.

$$Precision = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Positive\ (FP)} \quad (4)$$

E. F2 Score: A weighted harmonic mean of precision and recall, emphasizing the recall while still considering precision. It is particularly useful when the cost of false negatives is higher than that of false positives.

$$F_2 = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + 1/5\ (4\ x\ False\ Negative + False\ Positive)} \quad (5)$$

True Positives (TP) is the number of COVID-19 cases correctly identified by the model. True Negatives (TN) is the number of non-COVID cases correctly identified by the model. False Positives (FP) is the number of non-COVID cases incorrectly identified as COVID-19 by the model. False Negatives (FN) is the number of COVID-19 cases incorrectly identified as non-COVID by the model. These metrics were computed at each epoch during the validation phase, serving as an evaluative lens for the influence of different batch sizes and learning rates. The dynamic assessment of these metrics helped to identify the best hyperparameter configurations that led to optimal performance, thereby fulfilling the research's main aim. This evaluation process ensured that the model not only performed well on average but was also robust in classifying both COVID-19 and non-COVID cases correctly.

Explaining research chronological, including research design, research procedure (in the form of algorithms, Pseudocode or other), how to test and data acquisition (Cronje, 2020). The description of the course of research should be supported references, so the explanation can be accepted scientifically (Fryer & Dinsmore, 2020).

3. RESULTS AND DISCUSSIONS

This study used 698 datas of CT-Chest images labeled as COVID-19 and non-COVID-19 which was divided into training and validation subsets using an 80:20 ratio. CNN classification process began with data preprocessing to mitigate data imbalance for model training and optimizing model stability through pixel normalization. CNN architecture was thoughtfully designed to maximize efficiency in classifying CT-Chest images for COVID-19 diagnosis. It consisted of three primary convolutional layers, each equipped with a varying number of filters to extract intricate features from the images. Additionally, the architecture incorporated pooling layers to reduce dimensionality while retaining essential features. This study adopted the Rectified Linear Unit (ReLU) as the activation function for the convolutional layers which is applied to the output of each neuron in the network. It takes in the weighted sum of the input and produces an output that is then passed on to the next layer. This study employed the Adam optimizer which to calculate and update various network parameters to gradually approach and reach the optimal value. The model was trained over 10 epochs, optimizing two key hyperparameters learning rate and batch size. Several iterations were carried out with varying learning rates and batch sizes. This classification process used 10-fold validation to divides the dataset into 10 subsets and conducts training and validation iteratively, ensuring that each subset serves as both training and validation data at some point. Each iteration produced the value of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). These values then use to calculate performance metrics like Accuracy, Sensitivity, Specificity, Precision, and F2-Score based on the formula shown in Equation 1 to 5. The primary aims of this study are to assess how different learning rates and batch sizes affected the CNN's performance and to identify optimal batch size and learning rate that lead to the best performance of the CNN. This study demonstrated varying values of the performance metrics based on the learning rate and batch size combinations. The results shown in Figure 3 to 8. The metrics such as sensitivity, precision, specificity, accuracy, and F2-score shown in Figure 3 to 8 are derived from key diagnostic elements: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values which is calculated using formulas shown in Equation 1 to 5. These vital components are generated through systematic iterations of the CNN model's classification process, each employing various combinations of learning rates and batch sizes

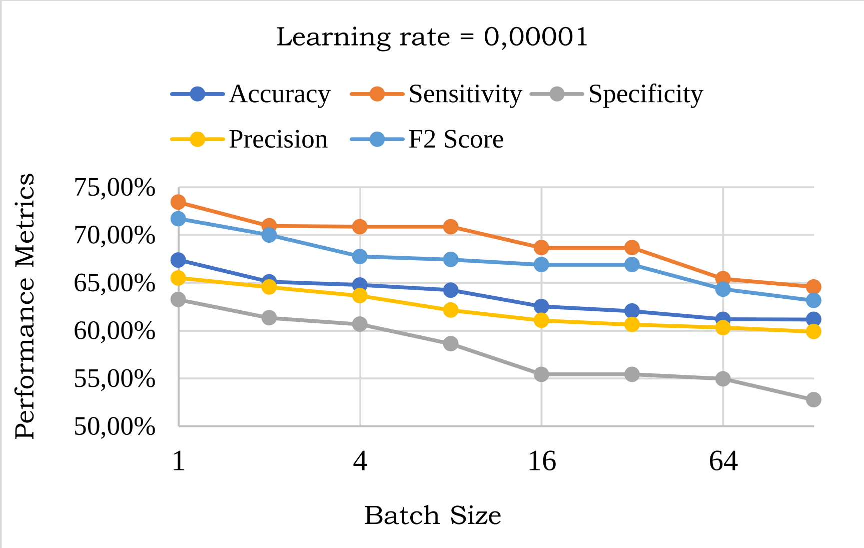

Figure 3. Performance metrics of learning rate 0,00001

Figure 3 shows that at a learning rate of 0.00001, the highest values for Accuracy, Sensitivity, Specificity, Precision, and F2-Score are achieved when using a batch size of 1. Specifically, the Accuracy reached 67.38%, Sensitivity was at 73.43%, Specificity scored 63.24%, Precision was at 65.51%, and the F2-Score reached 71.69%. Additionally, it was noted that the values for Accuracy, Sensitivity, Specificity, Precision, and F2-Score mostly decreased as the batch size increased.

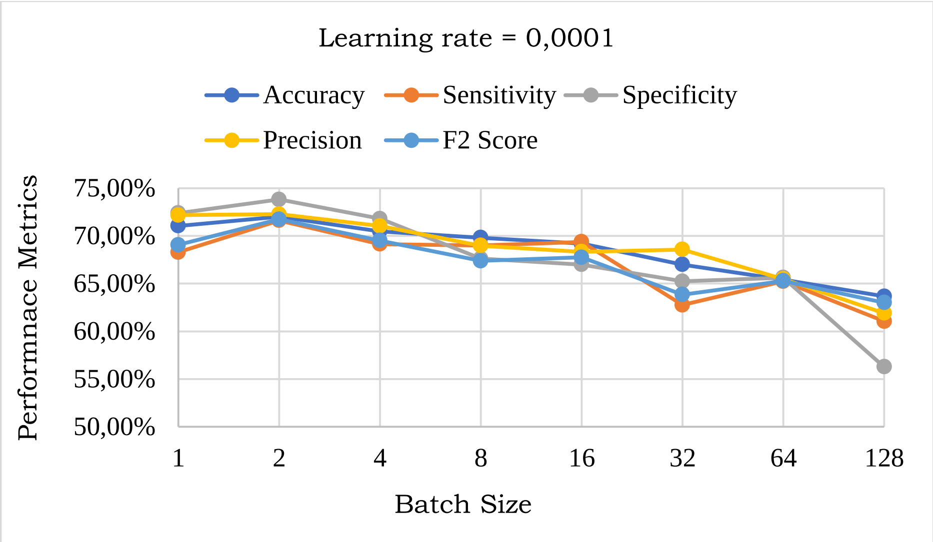

Figure 4. Performance metrics of learning rate 0,0001

Figure 4 shows that with a learning rate of 0.0001, the highest values for Accuracy, Sensitivity, Specificity, Precision, and F2-Score were achieved at a batch size of 1. Specifically, Accuracy reached 72.00%, Sensitivity was at 71.62%, Specificity stood at 73.81%, Precision was 72.28%, and F2-Score reached 71.73%. Furthermore, the values for Accuracy, Sensitivity, Specificity, Precision, and F2-Score generally decreased as the batch size increased.

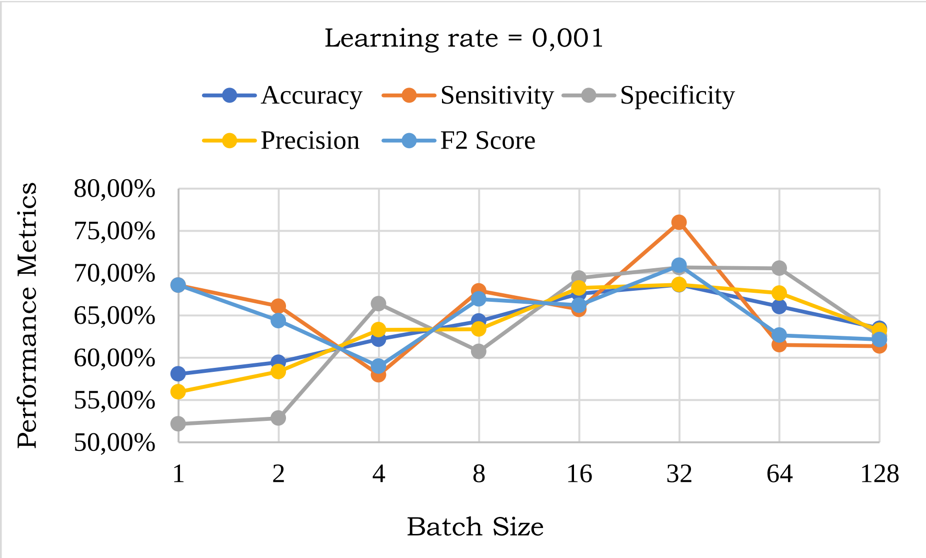

Figure 5. Performance metrics of learning rate 0,001

Figure 5 shows that with a learning rate of 0.001, the highest values for Accuracy, Sensitivity, Specificity, Precision, and F2-Score were achieved at a batch size of 32. Specifically, Accuracy reached 68.62%, Sensitivity was at 76.00%, Specificity stood at 70.67%, Precision was 68.64%, and F2-Score reached 70.92%.

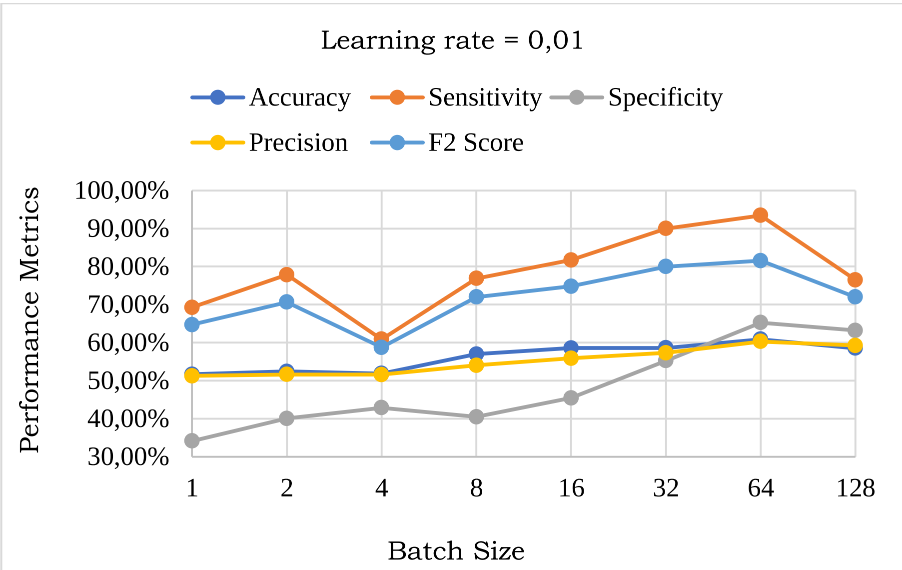

Figure 6. Performance metrics of learning rate 0,01

Figure 6 shows that the highest values for Accuracy, Sensitivity, Specificity, Precision, and F2-Score were attained with a learning rate of 0.01 at a batch size of 64. Specifically, the Accuracy was 60.86%, Sensitivity was 93.43%, Specificity was 45.24%, Precision was 58.27%, and the F2-Score was 81.53%.

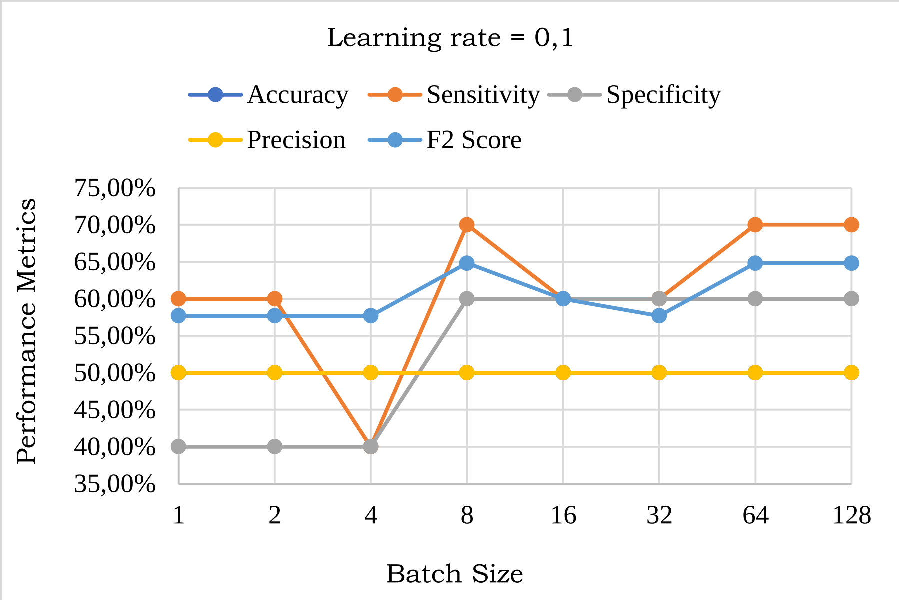

Figure 7. Performance metrics of learning rate 0,1

Figure 7 shows that the highest values for Accuracy, Sensitivity, Specificity, Precision, and F2-Score were achieved with a learning rate of 0.1 at a batch size of 64. Specifically, the Accuracy was 50.00%, Sensitivity was 70.00%, Specificity was 60.00%, Precision was 50.00%, and the F2-Score was 64.81%.

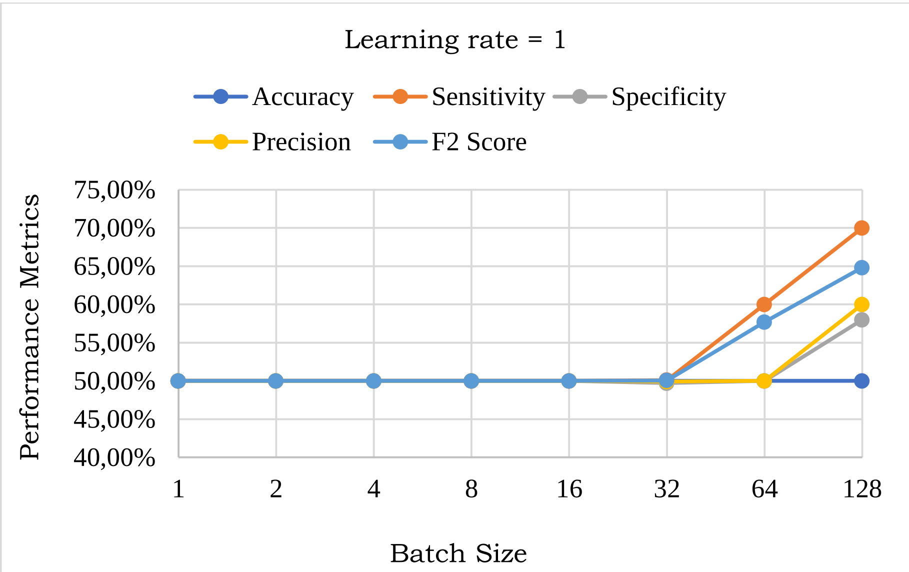

Figure 8. Performance metrics of learning rate 1

Figure 8 shows that the highest values for Accuracy, Sensitivity, Specificity, Precision, and F2-Score were achieved with a learning rate of 1 at a batch size of 128. Specifically, the Accuracy was 50.00%, Sensitivity was 70.00%, Specificity was 58.00%, Precision was 60.00%, and the F2-Score was 64.81%. Summary of the highest performance metrics results from each learning rate and batch size shown in Figure 3 to Figure 8 are presented in Table 1. These datas were generated from the classification result from each learning rate and batch size. This study used CNN to classify the COVID-19 CT-Chest images and used 10-fold validation to do the iteration for each learning rate and batch size. Each learning rate and batch size produced a lot of values of the TP, TN, FN, FP which then used to calculate the performance metrics. From all of those values, the highest result picked to be shown in Table 1.

Table 1. The highest performance metrics results from each learning rate and batch size

Learning Rate	Batch Size	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F2-Score (%)
0.00001	1	67.38	73.43	63.24	65.51	71.69
0.0001	1	72.00	71.62	73.81	72.28	71.73
0.001	32	68.62	76.00	70.67	68.64	70.92
0.01	64	60.86	93.43	45.24	58.27	81.53
0.1	64	50.00	70.00	60.00	50.00	64.81
1	128	50.00	70.00	58.00	60.00	64.81

Table 1 shows that lower batch sizes generally yielded highest results at lower learning rates, otherwise higher batch size yielded highest results at higher learning rate. The best performance yielded at learning rate 0.01 with batch size 64 with sensitivity 93.43%, accuracy 60.86% F2-Score 81.53%, specificity 45.24% and precision 58.27%. These values are considered as the best performance regarding to the aim of this study which is prioritize high sensitivity but still consider the values of accuracy, specificity, precision and F2-Score. These findings confirm that learning rate and batch size significantly influences the CNN model's performance. Therefore, selecting the appropriate learning rate and batch size is crucial, especially for cases that requiring high sensitivity such as COVID-19 diagnosis. Given that the dataset comprises CT-Chest images for COVID-19 diagnosis, high sensitivity is crucial to ensure that as many actual cases of the disease are correctly identified as possible. Missing a positive COVID-19 case could have serious health implications, not only for the individual but also for public health due to the contagious nature of the virus. This study highlights critical insights into optimizing CNN models for COVID-19 diagnosis which emphasize the significance of hyperparameter fine-tuning, particularly the learning rate and batch size. This aligns with growing research on hyperparameter optimization. This study incorporated *k*-fold cross-validation bolsters result robustness. Future work could explore alternative algorithms and data augmentation to enhance the CNN performance in COVID-19 classification and foster international collaboration to diversify datasets. Ultimately, this research advances CNNs for precise and reliable COVID-19 diagnosis during global health crises.

4. CONCLUSION

In conclusion, the study found that both the learning rate and batch size have a significant impact on the performance of a Convolutional Neural Network (CNN) model in classifying CT-Chest images into COVID-19 and non-COVID-19 CT Scan. Specifically, a lower learning rate was optimized with a smaller batch size, enabling more efficient classification by the CNN model. Conversely, a higher learning rate required a larger batch size for optimal performance. The most effective combination of learning rate and batch size was identified in the experiment, where a learning rate of 0.01 and a batch size of 64 yielded an Accuracy of 60.86%, Sensitivity of 93.43%, Specificity of 65.24%, Precision of 60.27%, and an F2-Score of 81.53%. These findings offer valuable insights for fine-tuning the parameters of CNN models in the classification of medical images, specifically for the diagnosis of COVID-19. The results suggest that smaller batch sizes are more effective at lower learning rates, while larger batch sizes perform better at higher learning rates. This research affirms the inherent capability of CNN in feature extraction, potentially simplifying the model architecture for similar future applications.

These findings provide valuable guidelines for optimizing the parameters of CNN models, particularly in the medical imaging domain where high Sensitivity is often a priority. Future research could focus on evaluating new machine learning algorithms for improved accuracy and exploring their real-time applicability in clinical settings. Data augmentation techniques could also be investigated to enhance model performance, especially when dealing with imbalanced or scarce data. The integration of multimodal medical data, like CT scans and X-rays, could offer a more robust and accurate system. Additionally, ethical considerations and data privacy need to be addressed, and collaborations with international institutions could be beneficial for diversifying datasets and improving model generalizability. While this study offers valuable insights into the impact of learning rate and batch size on CNN model performance in classifying CT-Chest images for COVID-19 diagnosis, it is essential to acknowledge its limitations. One notable limitation is the size of the dataset. Although this study meticulously curated a dataset of 698 COVID-19 CT-Chest images, a larger and more diverse dataset would further enhance the robustness of thesefindings. Future research could benefit from access to larger and diverse datasets encompassing a wider range of COVID-19 cases and varying demographics to enhance model robustness and real-world applicability. Future research could also be exploring novel machine learning and deep learning algorithms beyond CNN could lead to improved accuracy and efficiency in medical image classification. Comparative studies with emerging algorithms can shed light on their practicality. By addressing these limitations and pursuing these future research directions, this research can continue to advance the field of medical image analysis, contributing to more accurate and efficient diagnostic tools for COVID-19 and other medical conditions.

ACKNOWLEDGEMENTS

I would like to express my gratitude to God, my advisors, and my family for their unwavering support throughout this research. Special thanks to my partner Pio Pratama and everyone who contributed to this research.

REFERENCES

Althnian, A., AlSaeed, D., Al-Baity, H., Samha, A., Dris, A. Bin, Alzakari, N., Abou Elwafa, A., & Kurdi, H. (2021). Impact of dataset size on classification performance: An empirical evaluation in the medical domain. *Applied Sciences (Switzerland)*, *11*(2), 1–18. https://doi.org/10.3390/app11020796

Alzubaidi, L., Bai, J., Al-Sabaawi, A., Santamaría, J., Albahri, A. ., Al-dabbagh, B., Fadhel, M., Manoufali, M., Zhang, J., Al-Timemy, A., Duan, Y., Abdullah, A., Farhan, L., Lu, Y., Gupta, A., Albu, F., Abbosh, A., & Gu, Y. (2023). A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. *Journal of Big Data*, *10*. https://doi.org/10.1186/s40537-023-00727-2

Bordelon, B., Pehlevan, C., & Broeck, V. Den. (2022). *LEARNING CURVES FOR STOCHASTIC GRADIENT DESCENT ON STRUCTURED FEATURES*. 1–21.

Cerqueira, V., Torgo, L., & Mozetič, I. (2020). Evaluating time series forecasting models: an empirical study on performance estimation methods. *Machine Learning*, *109*(11), 1997–2028. https://doi.org/10.1007/s10994-020-05910-7

Chen, C. H., Lai, J. P., Chang, Y. M., Lai, C. J., & Pai, P. F. (2023). A Study of Optimization in Deep Neural Networks for Regression. *Electronics (Switzerland)*, *12*(14), 1–17. https://doi.org/10.3390/electronics12143071

Cronje, J. C. (2020). Designing Questions for Research Design and Design Research in e-Learning. *Electronic Journal of E-Learning*, *18*(1), pp13-24.

Eom, H., & Choi, H. (2021). Alpha-Integration Pooling for Convolutional Neural Networks. *Journal of KIISE*, *48*(7), 774–780. https://doi.org/10.5626/jok.2021.48.7.774

Fryer, L. K., & Dinsmore, D. L. (2020). The Promise and Pitfalls of Self-report: Development, research design and analysis issues, and multiple methods. *Frontline Learning Research*, *8*(3), 1–9.

Goh, E., Yepremyan, A., Wang, J., & Wilson, B. (2023). *MAESSTRO : Masked Autoencoders for Sea Surface Temperature Reconstruction under Occlusion. July*, 1–20.

Halmar, H. F. (2020). Pemeriksaan Diagnostik COVID-19. *Jurnal Keperawatan Muhammadiyah*, *5*(1), 222–230.

Hariyani, Y. S., HADIYOSO, S., & SIADARI, T. S. (2020). Deteksi Penyakit Covid-19 Berdasarkan Citra X-Ray Menggunakan Deep Residual Network. *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, *8*(2), 443. https://doi.org/10.26760/elkomika.v8i2.443

Hassan, E., Shams, M. Y., Hikal, N. A., & Elmougy, S. (2023). The effect of choosing optimizer algorithms to improve computer vision tasks: a comparative study. In *Multimedia Tools and Applications* (Vol. 82, Issue 11). Multimedia Tools and Applications. https://doi.org/10.1007/s11042-022-13820-0

He, X., Xue, F., Ren, X., & You, Y. (2021). *Large-Scale Deep Learning Optimizations: A Comprehensive Survey*. *1*(1). http://arxiv.org/abs/2111.00856

Hensman, P., & Masko, D. (2015). The Impact of Imbalanced Training Data for Convolutional Neural Networks. *PhD*. https://www.kth.se/social/files/588617ebf2765401cfcc478c/PHensmanDMasko_dkand15.pdf

Hidaka, A., & Kurita, T. (2017). Consecutive Dimensionality Reduction by Canonical Correlation Analysis for Visualization of Convolutional Neural Networks. *Proceedings of the ISCIE International Symposium on Stochastic Systems Theory and Its Applications*, *2017*(0), 160–167. https://doi.org/10.5687/sss.2017.160

Huang, L., Qin, J., Zhou, Y., Zhu, F., Liu, L., & Shao, L. (2023). Normalization Techniques in Training DNNs: Methodology, Analysis and Application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–20. https://doi.org/10.1109/TPAMI.2023.3250241

Infante, M., Lutman, R. F., Imparato, S., Di Rocco, M., Ceresoli, G. L., Torri, V., Morenghi, E., Minuti, F., Cavuto, S., Bottoni, E., Inzirillo, F., Cariboni, U., Errico, V., Incarbone, M. A., Ferraroli, G., Brambilla, G., Alloisio, M., & Ravasi, G. (2009). Differential diagnosis and management of focal ground-glass opacities. *European Respiratory Journal*, *33*(4), 821–827. https://doi.org/10.1183/09031936.00047908

Jepkoech, J., Mugo, D. M., Kenduiywo, B. K., & Too, E. C. (2021). The Effect of Adaptive Learning Rate on the Accuracy of Neural Networks. *International Journal of Advanced Computer Science and Applications*, *12*(8), 736–751. https://doi.org/10.14569/IJACSA.2021.0120885

Jo, J.-M. (2019). Effectiveness of Normalization Pre-Processing of Big Data to the Machine Learning Performance. *The Journal of the Korea Institute of Electronic Communication Sciences*, *14*, 547–552. https://doi.org/10.13067/JKIECS.2019.14.3.547

Mooijman, P., Catal, C., Tekinerdogan, B., Lommen, A., & Blokland, M. (2023). The effects of data balancing approaches: A case study. *Applied Soft Computing*, *132*, 109853. https://doi.org/10.1016/j.asoc.2022.109853

Parsania, P. S., & Virparia, P. V. (2018). Computational Time Complexity of Image Interpolation Algorithms. *International Journal of Computer Sciences and Engineering*, *6*(7), 491–496. https://doi.org/10.26438/ijcse/v6i7.491496

Pei, X., Zhao, Y. hong, Chen, L., Guo, Q., Duan, Z., Pan, Y., & Hou, H. (2023). Robustness of machine learning to color, size change, normalization, and image enhancement on micrograph datasets with large sample differences. *Materials and Design*, *232*, 112086. https://doi.org/10.1016/j.matdes.2023.112086

Reyad, M., Sarhan, A., & Arafa, M. (2023). A modified Adam algorithm for deep neural network optimization. *Neural Computing and Applications*, *35*, 1–18. https://doi.org/10.1007/s00521-023-08568-z

Sido, J., & Konopik, M. (2019). Deep Learning for Text Data on Mobile Devices. *International Conference on Applied Electronics*, *2019-Septe*, 1–4. https://doi.org/10.23919/AE.2019.8867025

Sugihantono, A., Burhan, E., & Samoedra, E. (2020). Pedoman Pencegahan dan Pengendalian CORONAVIRUS DISEASE (COVID-19) Revisi ke-5. *Jakarta: Kementerian Kesehatan Indonesia*, *5*(2), 1–214.

Susilo, A., Rumende, C. M., Pitoyo, C. W., Santoso, W. D., Yulianti, M., Herikurniawan, H., Sinto, R., Singh, G., Nainggolan, L., Nelwan, E. J., Chen, L. K., Widhani, A., Wijaya, E., Wicaksana, B., Maksum, M., Annisa, F., Jasirwan, C. O. M., & Yunihastuti, E. (2020). Coronavirus Disease 2019: Tinjauan Literatur Terkini. *Jurnal Penyakit Dalam Indonesia*, *7*(1), 45. https://doi.org/10.7454/jpdi.v7i1.415

Swastika, W. (2020). Studi Awal Deteksi COVID-19 Menggunakan Citra CT Berbasis Deep Learning. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *7*(3), 629. https://doi.org/10.25126/jtiik.2020733399

Yamashita, R., Nishio, M., Gian Do, R. K., & Togash, K. (2018). Convolutional neural networks: an overview and application in radiology. *Smart Innovation, Systems and Technologies*, *9*, 611–629. https://doi.org/10.1007/978-981-15-7078-0_3

Yang, X., He, X., Zhao, J., Zhang, Y., Zhang, S., & Xie, P. (2020). *COVID-CT-Dataset: A CT Scan Dataset about COVID-19*. 1–14. http://arxiv.org/abs/2003.13865