

Explainable optimization of deep learning model for COVID-19 detection using chest images



Slamet Riyadi ^{a,b,*}, Eka Nova Pramudya ^a, Cahya Damarjati ^{a,b}, Jose Manuel Molina Lopez ^b, Jesus Garcia Herrero ^b

^a Department of Information Technology, Faculty of Engineering, Universitas Muhammadiyah Yogyakarta, Indonesia

^b Department of Computer Science Department, Universidad Carlos III de Madrid, Spain

ARTICLE INFO

Keywords:
COVID-19
Deep learning
VGG
Optimizer
X-ray
Grad-CAM
Activation maps

ABSTRACT

The COVID-19 pandemic has now become endemic, yet it remains essential to identify and diagnose the virus. The X-ray images on chest evaluations have utilized numerous deep-learning techniques and optimization algorithms. The Visual Geometry Group (VGG) is a widely recognized deep-learning architecture for COVID-19 detection. This architecture incorporates various optimization algorithms, such as Adadelta, Adam, Adamax, Stochastic Gradient Descent (SGD), Adagrad, Adam, and Root Mean Square Propagation (RMSprop), to ascertain the most favorable learning parameters. Currently, researchers evaluate the performance of the deep learning and optimization algorithm to detect COVID-19 by computing the statistical parameters, i.e., the accuracy, precision, recall, specificity, and F-1 score. However, the way in which deep learning and optimization algorithms work to predict the outcome still needs a comprehensive explanation. This study aims to explain how deep learning and optimization predict COVID-19 or normal chest using visual and quantitative analysis of a Gradient Class Activation Mapping (Grad-CAM) and statistical evaluation. The methodology involved data collection of chest images from a public data set, preprocessing, training-testing, and result analysis on statistical parameters and Grad-CAM heatmap. The evaluation of statistical parameters yielded favorable outcomes from all the optimization techniques, with an average accuracy of up to 99 %. It was then followed by visually observing the Grad-CAM activation heatmap, which indicated crucial regions in the images that influenced the model's prediction outcomes. Grad-CAM allows the visualization and quantification of the activation map in each optimization. It reveals distinct activation maps for the lung area in both COVID-19 and normal pictures. The visual observation was also confirmed by quantitative analysis of Grad-CAM using root mean square error (RMSE) of correlation between each optimization algorithm. The researcher compares two heatmaps from different optimization methods by computing the RMSE of each heatmap pixel value. The RMSE confirmed that the heatmap supports the explanation of the optimization performance. In conclusion, this research explained how deep learning and optimization algorithms predict COVID-19 from the chest images.

1. Introduction

The COVID-19 outbreak has shifted from a pandemic to an endemic state; however, detection and diagnosis remain essential, especially with various new virus variants. Since the COVID-19 pandemic, many technologies, e.g., tele-consulting, video conference, virtual healthcare, and mobile-based self-care, have been used in society because the benefits align with its condition [1,2]. Regarding artificial intelligence technology, researchers have actively explore implementing deep learning methods to detect COVID-19 based on X-ray and CT-scan images. The

methods are also used to detect mortality risks of persons with COVID-19 based on experimental and medical parameters [3]. Deep learning is a widely used machine learning technique designed to handle vast amounts of data with high levels of complexity [4]. A Convolutional Neural Network (CNN) is a powerful deep learning technique used for image analysis tasks such as object recognition, object detection, and object classification [5]. The VGG network, a CNN architecture, has shown exceptional performance in image recognition, particularly in the detection of COVID-19 [6–11]. VGG provides cutting-edge models for object recognition that can accommodate up to 19 layers.

* Corresponding author. Department of Information Technology, Faculty of Engineering, Universitas Muhammadiyah Yogyakarta, Indonesia.
E-mail address: riyadi@umy.ac.id (S. Riyadi).

Furthermore, it surpasses the standard performance on other tasks and datasets not limited to ImageNet. It continues to be one of the most often used image recognition frameworks today. VGG consists of six models that vary in the number of layers they have. The VGG-16 and VGG-19 models are widely recognized as the most popular ones. Both models exhibit similarities, with the only distinction being the depths of their convolutional neural network layers, which are 16 and 19, respectively [12,13]. In this research, the researcher implement VGG-16 due to its simplicity compared to VGG-19.

Employing appropriate optimization strategies is essential for optimizing the training process of VGG-16. Optimization seeks to minimize or maximize the objective function, enhancing input parameters' performance in a statistical model or machine learning method [14]. Various optimization techniques have been extensively utilized, including Adadelta, Adam, Adamax, Stochastic Gradient Descent (SGD), Adagrad, Nadam, and RMSprop. Every optimization approach exhibits varying performance regarding convergence, training speed, and ability to avoid local minima [15]. Various optimization techniques would have varying effects on the ability to detect COVID-19, as measured by quantitative metrics such as accuracy, precision, recall, specificity, and F1 score [16–18]. The quantitative characteristics are anticipated to enable the assessment of the performance of each optimization and justify the selection of an optimization method.

Artificial intelligence techniques, such as deep learning and its optimization, are still perceived as enigmatic by certain individuals because of their intricate nature, which makes them difficult to comprehend intuitively. The deep learning model's prediction or judgment often lack a simple explanation due to their extensive parameters and capacity to uncover complicated pattern in data [19,20]. Researchers have attempted to introduce the concept of explainable artificial intelligence to elucidate the method's functioning. The Grad-CAM method can be utilized to ascertain the specific contribution of

individual pixels or picture components to the prediction outcome of a model. Therefore, Grad-CAM's visual representation of the model's attention toward different picture features or regions is valuable for analysis and interpretation [16,21]. Through the visual analysis of X-ray pictures using Grad-CAM, the researcher would like to elucidate the impact of optimization on image activation and identify critical regions in detecting COVID-19.

No prior research has explained the optimization of VGG-16 deep learning in the context of COVID-19 identification using X-ray pictures. Therefore, the objective of this study is to enhance the VGG-16 model and elucidate the techniques for optimizing deep learning in order to accurately detect COVID-19 using X-ray images. The combined investigation of optimization performance and visual interpretation using Grad-CAM [22,23] can comprehensively understand how optimization affects the model's ability to recognize COVID-19 infections. Visual observational analysis aids in understanding activation patterns and important locations, while quantitative analysis provides a measurable foundation for evaluating the statistical effectiveness of different optimization strategies. These findings enhance the comprehension of how various optimization strategies affect the model's capacity to identify and assess crucial attributes in X-ray pictures of diseased lungs caused by the COVID-19 virus.

2. Material and method

2.1. Experimental design

This empirical investigation was conducted using a series of five distinct phases, as illustrated in Fig. 1. The methodology comprised five stages: data collection, preprocessing, model training with seven optimization strategies, performance evaluation, and heatmap creation. The optimization performance was evaluated by comparing six statistical

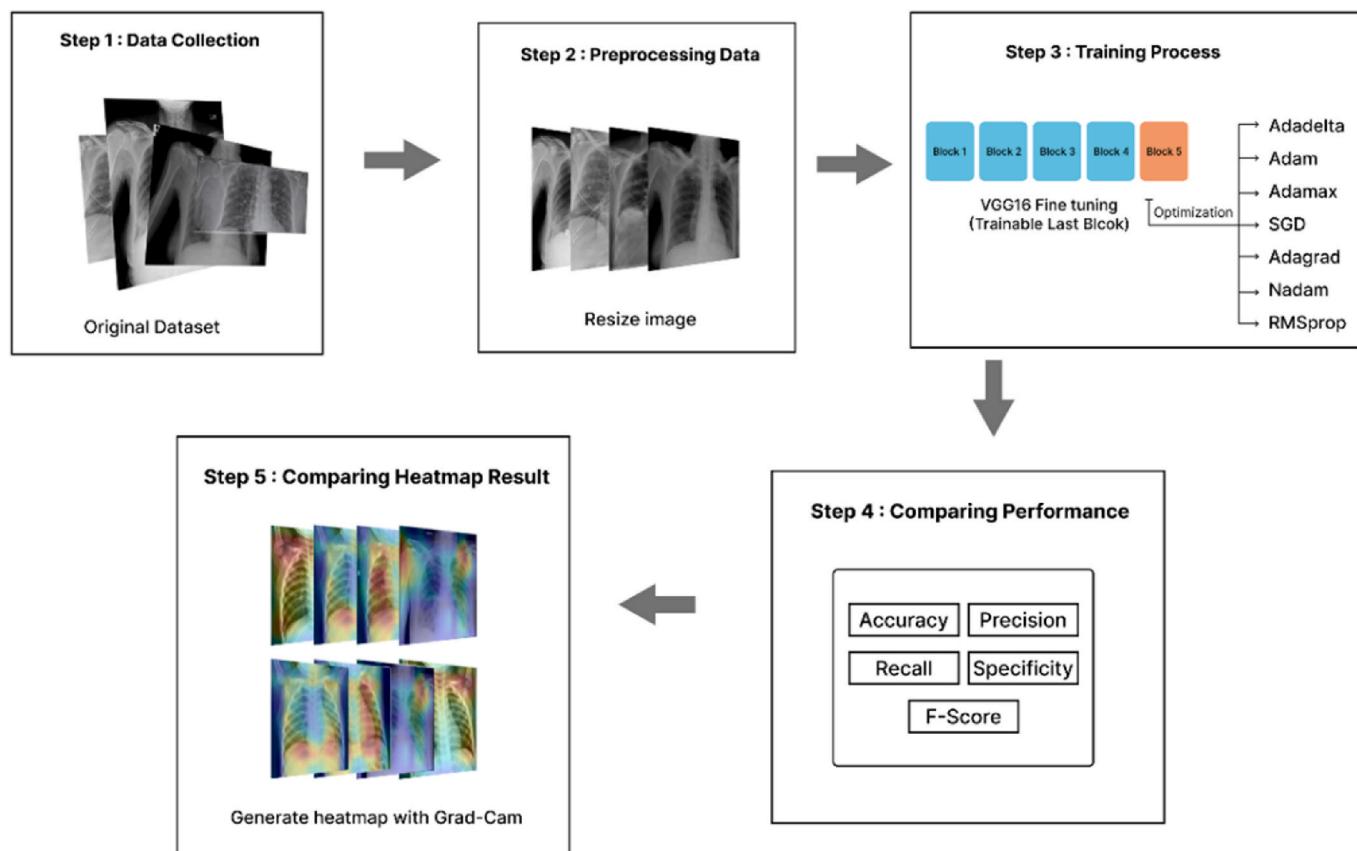


Fig. 1. The experimental design.

parameters, and the Grad-CAM algorithm was used to generate image heatmaps that visually depict the areas of focus for each optimization.

2.2. Dataset

The dataset used in this research consisted of 2159 X-ray lung scans, all obtained from Refs. [24,25]. The database was partitioned into two categories: normal and COVID-19. Fig. 2(a) and (b) display various X-rays depicting normal and COVID-19 conditions, respectively. Table 1 provides a breakdown of the image count for each class. The dataset was divided into training data (80 %) and testing data (20 %).

2.3. Image preprocessing

The VGG16 deep learning training utilized a standard image input size of 224 by 224 [26,27]. Choosing this size as the default was practical, aimed at optimizing computational efficiency, maintaining compatibility with pre-trained models, and capturing sufficient visual information for a classification task. The augmentation techniques were utilized to increase or decrease the size of photographs and make other modifications to them. The augmentations employed were tilt (shear), zoom (enlarge), and horizontal flip. Fig. 3 exhibits the enhancements.

2.4. Fine-tuning VGG-16

The researcher utilized the method of fine-tuning to enhance the efficacy of the VGG16 network in detecting COVID-19. VGG16 is a pre-trained convolutional neural network (CNN) model that has been trained on a large dataset known as ImageNet. The procedure of fine-tuning involved employing pre-trained weights from VGG16 and subsequently training the network on a specialized dataset specifically created for COVID-19 identification [28,29].

In VGG16 fine-tuning, the top-portion or fully connected layer section of the VGG16 model corresponded to the layers positioned at the end of the highest point of the VGG16 architecture, typically accompanied by the softmax layer for classification. During fine-tuning, it was common practice to exclude or remove the upper section of the VGG16 model from the retraining process, as depicted in Fig. 4. This was performed for two main reasons. Initially, these layers were predominantly focused on classifiers trained using the ImageNet dataset, which was not immediately applicable to the categorization of COVID-19. By excluding these layers, the model could concentrate on fine-tuning the lower layers to acquire more precise characteristics for COVID-19 identification [13].

Furthermore, by eliminating the upper section of the VGG16 model, the number of parameters that needed to be modified during fine-tuning

was reduced. The limited availability of the dataset for training proved advantageous. To minimize overfitting and get outstanding results, it is recommended to limit the number of adjustable parameters [30].

VGG16 consists of many blocks, each including a set of layers that include highly proficient low-level characteristics. While fine-tuning a pre-trained model, the objective is to modify the weights of a specific collection of layers to enable the model to adapt to a new dataset while retaining its capability to extract pre-existing low-level features. During the tuning process, one approach entailed immobilizing VGG16 blocks 1–4. By ceasing training in this specific region, the weights on these layers were conserved without undergoing further training. The low-level features of the initial blocks in VGG16, such as edge detection or simple lines, were generally broad and generic. These traits were not exclusive to a certain dataset and could be widely employed. Fig. 5 illustrates the technique of preserving well-trained low-level features by immobilizing blocks 1–4 of the VGG16 layer. This allows for fine-tuning subsequent blocks more pertinent to the new dataset [31,32]. This approach can impart knowledge on proficiency in a specific domain. The objective is to mitigate overfitting in the analysis of tasks with little data by reducing the disparity between the pre-existing model and the incoming task data [33].

2.5. Training

The creation and training of the VGG16 model were carried out using the Keras [34] and TensorFlow [35] libraries. Various parameters were fine-tuned during the construction of the VGG16 model to achieve optimal training outcomes. Initially, the model was trained using 50 epochs. An epoch is the total number of iterations throughout a model's training process. The weights and biases are updated in each iteration depending on the calculated results obtained from the training data.

Following that, the training process incorporated seven unique optimizations, namely Adadelta [36], Adam [37], Adamax [37], SGD [38,39], Adagrad [40], Nadam [41], and RMSprop [42]. Optimization algorithms used the gradients generated during training to modify the model's weights and biases. By employing distinct update policies and methods for each optimization, it is possible to get speedy convergence and accurate results through extensive experimentation with many possibilities.

The dataset in this experiment was divided into training and validation sets using an 80:20 ratio. 80 % of the data was used to train the model, while the remaining 20 % were used to test and evaluate the performance of the trained model. To reduce overfitting and increase model generalization to novel data, the dataset was partitioned in this manner.

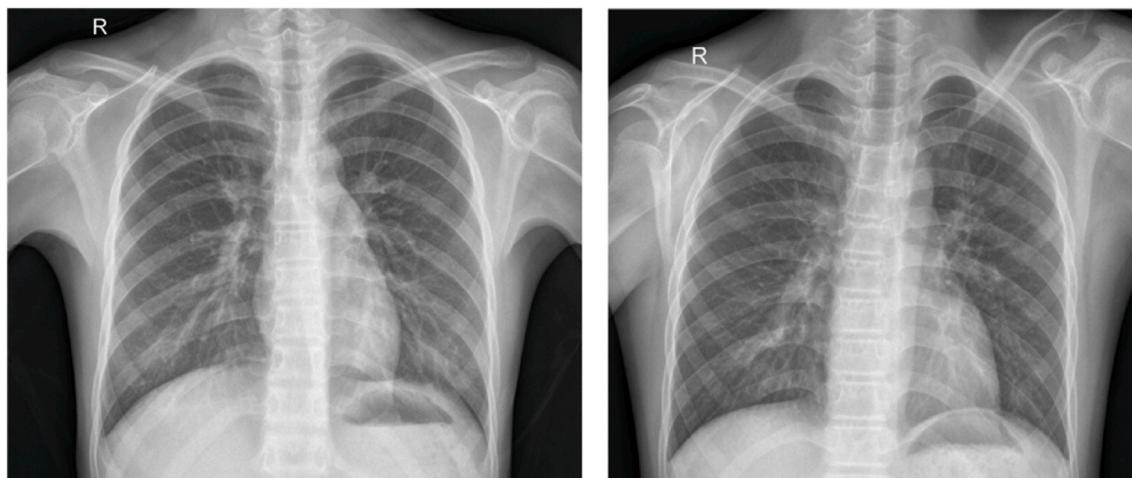


Fig. 2. (a) Sample of X-Ray for normal case.

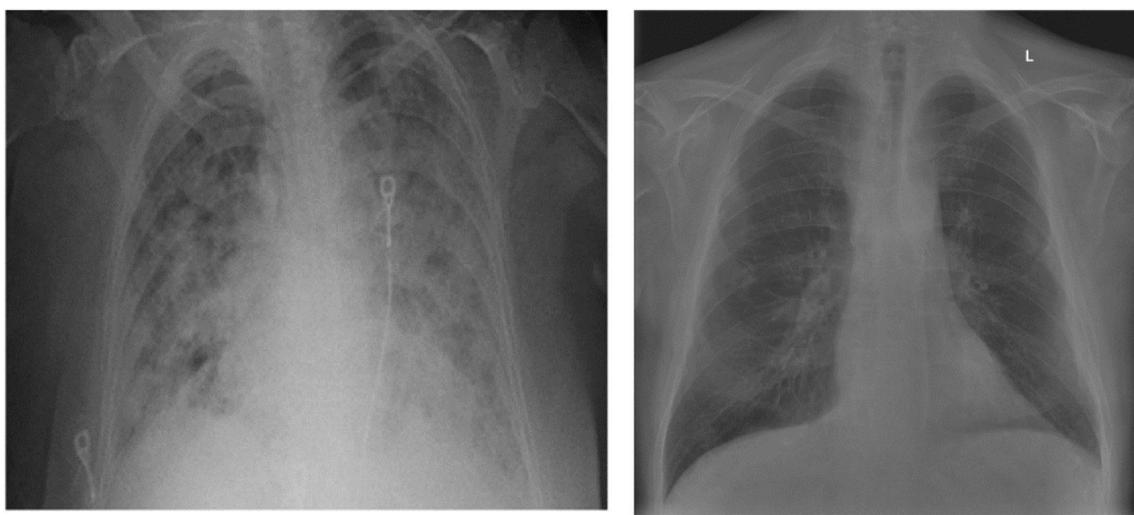


Fig. 2. (b) Sample of X-Ray for Covid-19 case.

Table 1
Detail of dataset split for training and testing.

Class	Training (80 %)	Testing (20 %)	Total
Normal	1266	317	1583
COVID-19	461	115	576
Total	1727	432	2159

2.6. Performance evaluation

The confusion matrix, presented in [Table 2](#), illustrates the parameters used to evaluate the training outcomes of each optimization. In this scenario, True Positive (TP) and False Positive (FP) represent the overall count of images for which the model correctly and incorrectly predicted. Similarly, the terms True Negative (TN) and False Negative (FN) refer to the count of images that the model accurately and inaccurately predicted, respectively.

The evaluation of model performance employed the confusion matrix to illustrate the model's ability to classify the correct class or label accurately. The performance matrix offered a comprehensive perspective on the model's accuracy in accurately predicting class or label assignments across many matrices. [Table 3](#) presents detailed information regarding the efficiency of the matrix.

2.7. Visual observation of Grad-CAM analysis

Researchers have devised a technique known as Gradient-weighted Class Activation Mapping (Grad-CAM) to ascertain the contribution of individual pixels or picture components to the prediction outcomes of a model. Grad-CAM facilitated the comprehension of the pivotal visual elements that influenced the model's decision-making process [16,21, 23]. This approach provided a visual representation that aided in understanding and analyzing the model's emphasis on characteristics or elements of the image. Grad-CAM employed gradient information to

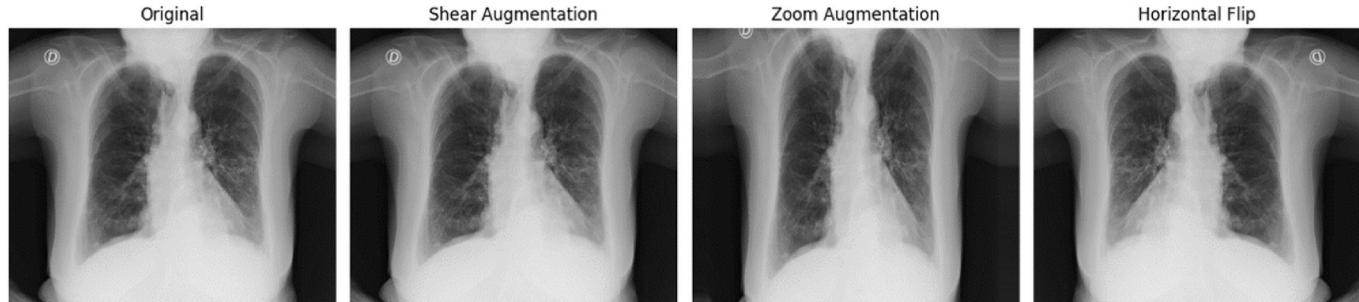


Fig. 3. Example of augmentation method.

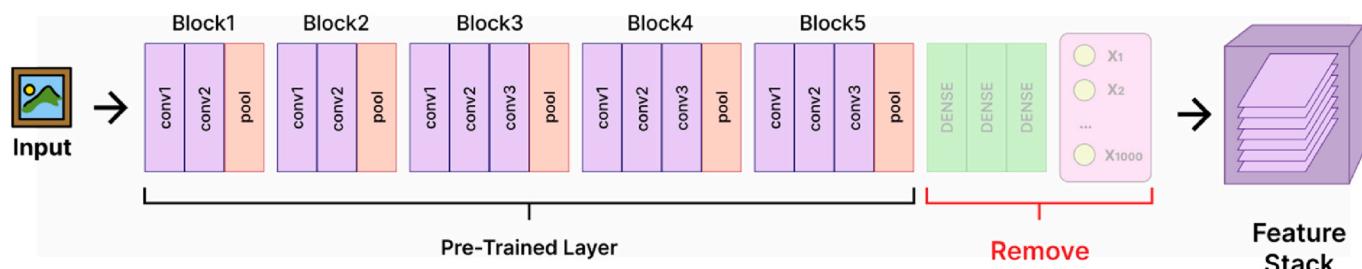


Fig. 4. Remove “top-portion” (default fully connected layer).

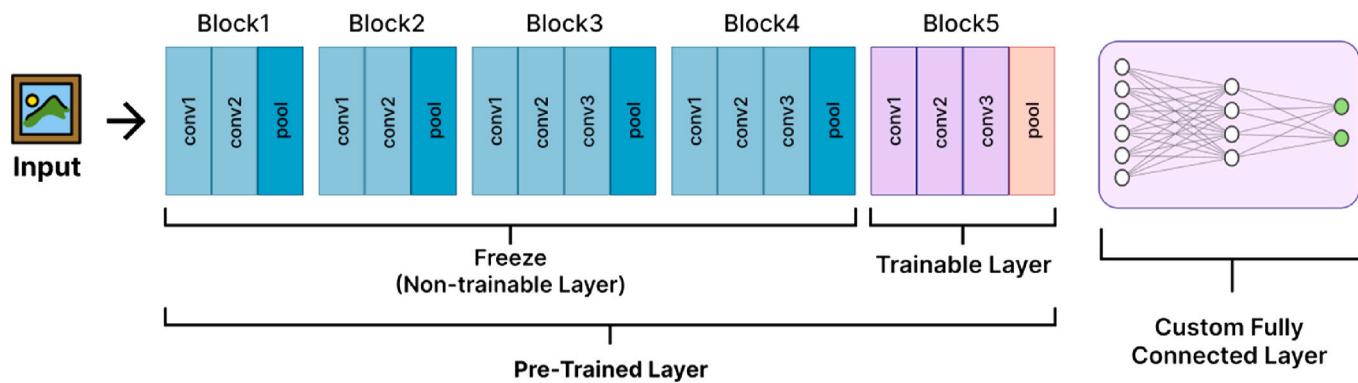


Fig. 5. Freezing block 1–4 and adding new custom fully connected layer.

Table 2
Confusion matrix of COVID-19 and normal classes.

	Actual COVID-19	Actual Normal
Predicted COVID-19	TP	FN
Predicted Normal	FP	TN

Table 3
Performance evaluation matrices.

Matrix	Formula
Accuracy	$\frac{TP + TN}{TP + FP + FN + TN}$
Precision	$\frac{TP}{TP + FP}$
Recall	$\frac{TP}{TP + FN}$
Specificity	$\frac{TN}{TN + FP}$
F-score	$\frac{2 * (Precision * Recall)}{(Precision * Recall)}$

generate a heatmap illustrating the areas where the VGG16 network's examination of lung pictures was advantageous. The network's COVID-19 detection assessments were visualized in a heatmap, illustrating the most significant areas of the images. Fig. 6 displays the overview of the application of Grad-CAM in deep-learning models.

Grad-CAM produced a map-like representation of the input images, highlighting important regions the VGG16 network used to achieve its

final classification determination [21]. This approach involved comparing the gradient of the output of the target class with the feature map produced by the convolution layers of the network. The input images were then visualized by displaying a heatmap highlighting the important areas. Grad-CAM can be utilized with VGG16 to provide insights into the network's image feature detectors and their role in categorization [21,37]. The performance aimed to elucidate the specific attributes of the COVID-19 images employed to generate the heatmap and demonstrate how a VGG16 deep learning model could discern these attributes.

2.8. Quantitative analysis of Grad-CAM

In addition to visual examination, the Root Mean Squared Error (RMSE) was employed for quantitative analysis of heatmaps generated by Grad-CAM. This statistic quantifies the disparity in pixel intensity between two heatmaps derived from distinct model analyses. Initially, employing the Grad-CAM technique, as previously elucidated, was employed to generate a heatmap of the designated layer. The presented heatmap illustrates the regions of the image that exert the greatest influence on the model's decision-making process in detecting COVID-19. The values in the heatmap were adjusted to a range of 0–1, guaranteeing uniformity in visual comparisons across various models.

The researcher evaluated the heatmap analysis findings from different models by comparing their RMSE values. The RMSE was computed by comparing the pixel intensity differences between the heatmaps of two models, Adadelta and Adagrad. The RMSE results offered insight into the level of visual similarity or dissimilarity between

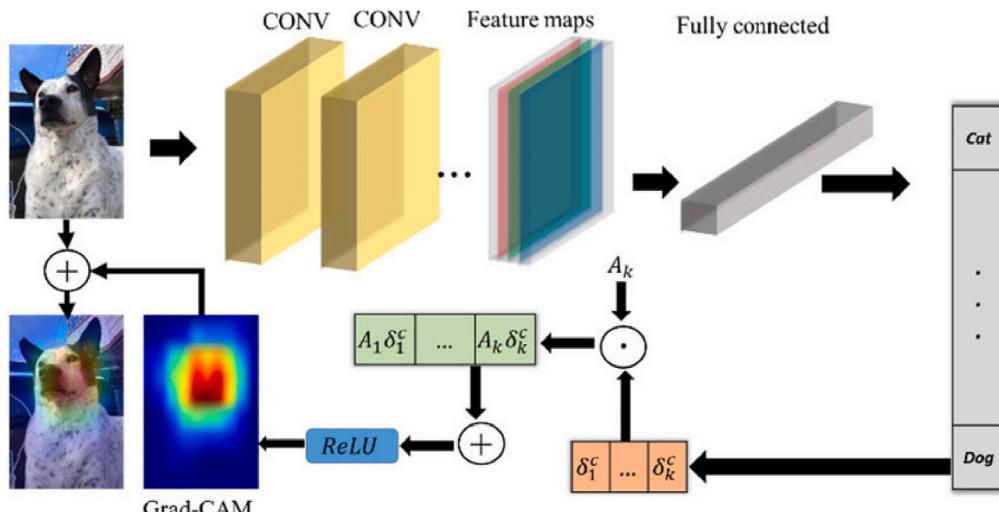


Fig. 6. Overview of the application of Grad-CAM in deep learning models [52].

the heatmaps of the models being examined. This procedure offers an initial assessment of the visual efficacy of each model in detecting COVID-19 in lung pictures.

This methodology conducted a comparative analysis of 7 distinct optimization strategies to provide a full visual comparison. Utilizing RMSE during the technique phase offered an unbiased criterion for choosing models with superior visual performance detecting COVID-19 in lung pictures. This information provided insights into the level of visual reliability of model analysis results in the context of COVID-19 detection.

Procedures for computing RMSE are as follows.

1. Selected the "block5_conv3" layer in the VGG16 model architecture as the main layer for heatmap analysis. This layer is a key component of a convolutional neural network and is crucial in detecting significant characteristics in lung pictures. This is backed by earlier identification, where this layer is the standard for identifying each optimization model.
2. The heatmap computation used the Grad-CAM technique to generate a heatmap based on the "block5_conv3" layer. The presented heatmap illustrated the regions of the image that exerted the most significant impact on the model's decision-making process in recognizing COVID-19.
3. Standardized heatmap values The values in the heatmap were normalized to fall within the range of 0–1. This facilitated the comparison of heatmaps generated by different models and guaranteed consistent visual intensity in the output.
4. Compared models by measuring the RMSE between the heatmaps generated by each pair of models, such as Adadelta and Adagrad. This procedure entailed computing the disparity in pixel intensity between two heatmaps and quantifying the extent of this disparity.
5. The RMSE values provided an interpretation of the similarity or visual differences between the heatmaps of the two models being evaluated. Smaller RMSE values suggested greater similarity in visual representations, while larger values indicated more pronounced contrasts.
6. The optimization analysis encompassed a total of 7 distinct optimization strategies in the comparison study. This procedure enabled the acquisition of outcomes about the visual efficacy of each model in detecting COVID-19 in lung pictures.

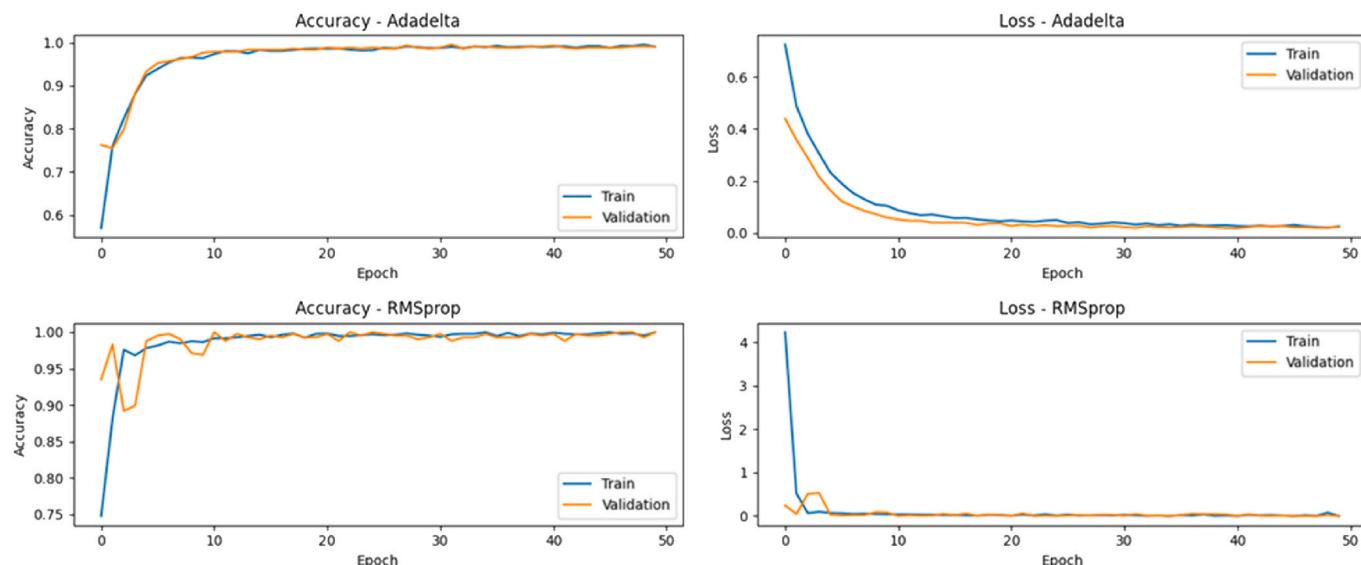


Fig. 7. Two optimizations with the lowest accuracy and the highest loss of the evaluation training.

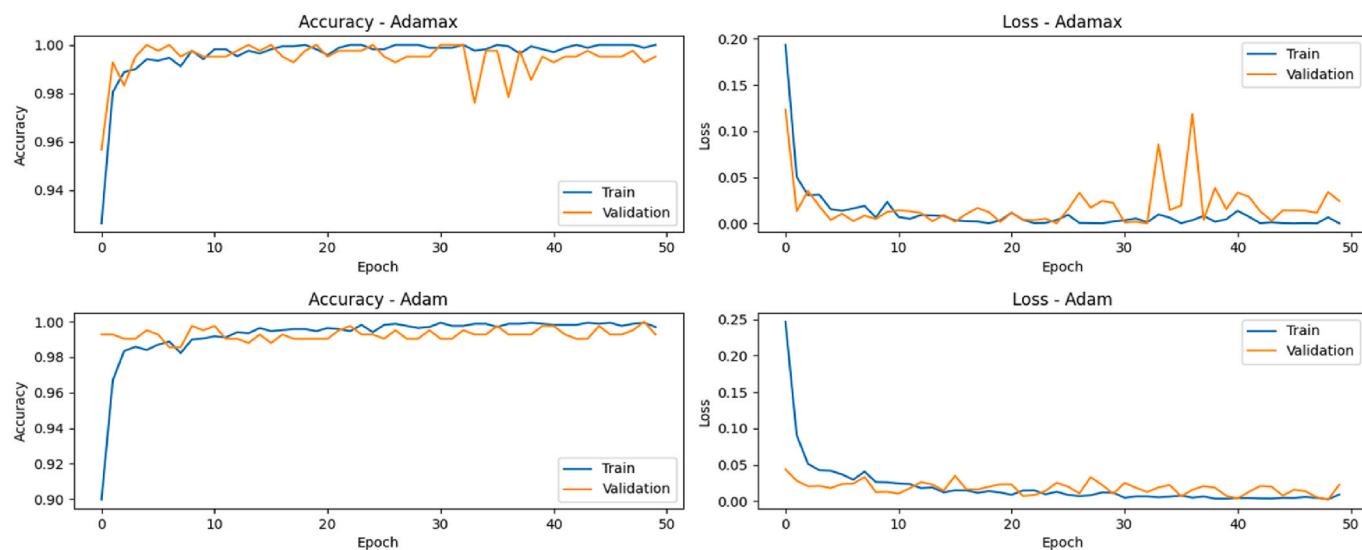


Fig. 8. Two optimizations with the highest accuracy and the lowest loss from the evaluation.

Table 4
Performance matrix evaluation for each optimization.

Optimization	Accuracy	Precision	Recall	Specificity	F1-Score
Adadelta	99,31	98,70	99,35	99,20	99,02
Adam	100,00	100,00	100,00	100,00	100,00
Adamax	99,54	100,00	98,72	99,02	99,36
SGD	99,31	100,00	98,10	98,55	99,06
Adagrad	99,77	99,35	100,00	99,85	99,68
Nadam	100,00	100,00	100,00	100,00	100,00
RMSprop	99,54	100,00	98,72	99,02	99,36

have exceptionally high specificity, ranging from 99.02 % to 100 %. The optimization method yields a notable F1 Score, ranging from 99.02 % to 100 %. The F1 Score is a metric that combines Precision and Recall, measuring the balance between the two. All optimization techniques achieve a favorable equilibrium between Precision and Recall for identifying COVID-19 infection.

The performance matrix indicates that employing the Adam and Nadam optimization methods yields a perfect score of 100 % for all assessment criteria (Accuracy, Precision, Recall, Specificity, and F1-Score) in both the COVID-19 and Normal classes. This demonstrates that both optimization approaches exhibit exceptional performance in accurately predicting both classes. Nevertheless, alternative optimization techniques such as Adadelta, Adamax, SGD, Adagrad, and RMSprop also exhibit exceptional performance, yielding evaluation scores that are nearly or exactly 100 %. This demonstrates the model's efficacy, optimized using different algorithms to accurately categorize cases as COVID-19 or normal within the provided dataset.

The observed similarities in the optimization outcomes of Adam and Nadam are not mere chance. This may be attributed to various factors, one of which is that Adam is a variant of the RMSprop approach that employs an adjustable learning rate. By utilizing this approach, Adam can address the challenge of determining learning rates manually and enhance the optimization of the learning process in deep neural networks [43]. The Nadam optimization method combines the Nesterov Accelerated Gradient (NAG) and Adam algorithms. The experimental findings indicate that Nadam exhibits superior convergence rates to Adam across various machine-learning tasks [37]. The Adam and Nadam optimization algorithm integrates momentum and adaptive techniques to optimize the learning process and address issues such as sluggish training [37]. The optimization of Adam and Nadam for training deep learning models with large batches demonstrates that Nadam outperforms Adam in terms of accuracy and efficiency [44].

The assessment results demonstrate that all optimization approaches employed in this study exhibited exceptional performance in identifying COVID-19 infection, exhibiting high accuracy, precision, recall, specificity, and F1 scores. Nevertheless, it is important to acknowledge that this research relies solely on the presented data and may not accurately represent performance in alternative datasets or varying circumstances.

3.2. Grad-CAM analysis on each VGG16 layer

Fig. 9 presents an overview of the VGG16 layers, serving as the basis for the Grad-CAM visualization study. The Grad-CAM study utilized a sequence of convolution layers followed by layer pooling to ascertain the relative significance of each layer type. Extracting features strongly depended on the convolution layer before the pooling step. The convolution layer utilized a convolution filter to detect more intricate features in the image, such as edges, corners, and textures. Before pooling, examining the Grad-CAM results on the convolution layer demonstrated how the layer focused on the essential features of the image [45–47].

The pooling layer decreases the spatial dimensions of the features generated by the preceding convolution layer. The pooling layer performs either maximum or average pooling on a specific feature region, reducing the feature's complexity and size. The Grad-CAM study demonstrates the impact of the pooling layer on the spatial representation of features discovered by preceding convolution layers [48–50]. By choosing the final convolution layer before pooling and the pooling layer, it is possible to observe the disparity in the impact of these two layers in emphasizing significant characteristics in the image. This aids in comprehending the function and impact of each layer type in extracting and manipulating features inside neural networks, as well as their contribution to developing the model's comprehension of a certain image.

Fig. 10 demonstrates the interpolation and placement of a Grad-CAM heatmap onto normal X-ray pictures. **Fig. 11** illustrates generating a Grad-CAM heatmap from COVID-19 X-ray images. It also demonstrates the visualization performed on many levels in the VGG16 model. This visualization was a sample outcome derived from Adam's optimization. Adam was chosen as the sample because the earlier quantitative analysis revealed that its optimization yielded flawless outcomes.

The installation of Grad-CAM on certain layers, namely 'block1_conv2', 'block1_pool', 'block2_conv2', 'block2_pool', 'block3_conv3', 'block3_pool', 'block4_conv3', 'block4_pool', 'block5_conv3', and 'block5_pool', unveiled the impact and disparities present in each layer.

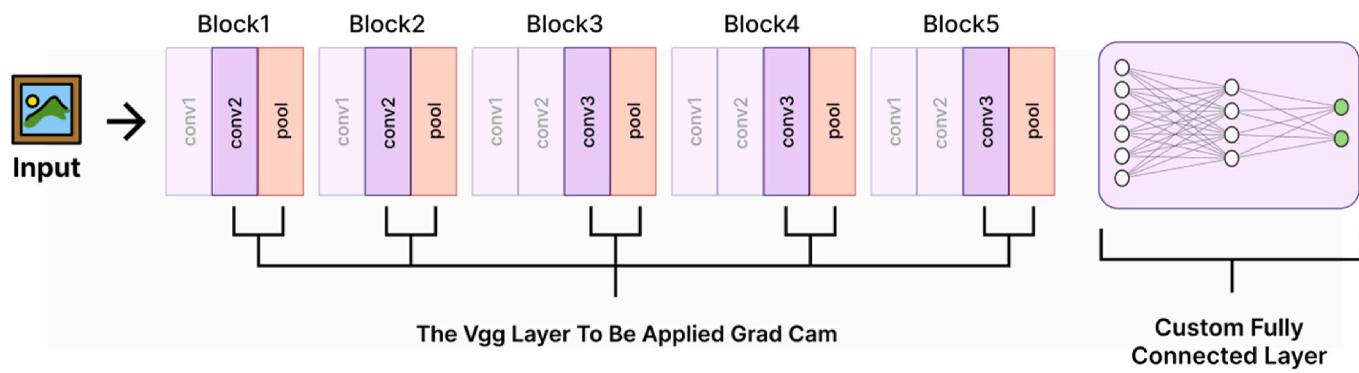


Fig. 9. VGG16 layers taken to apply Grad-CAM.

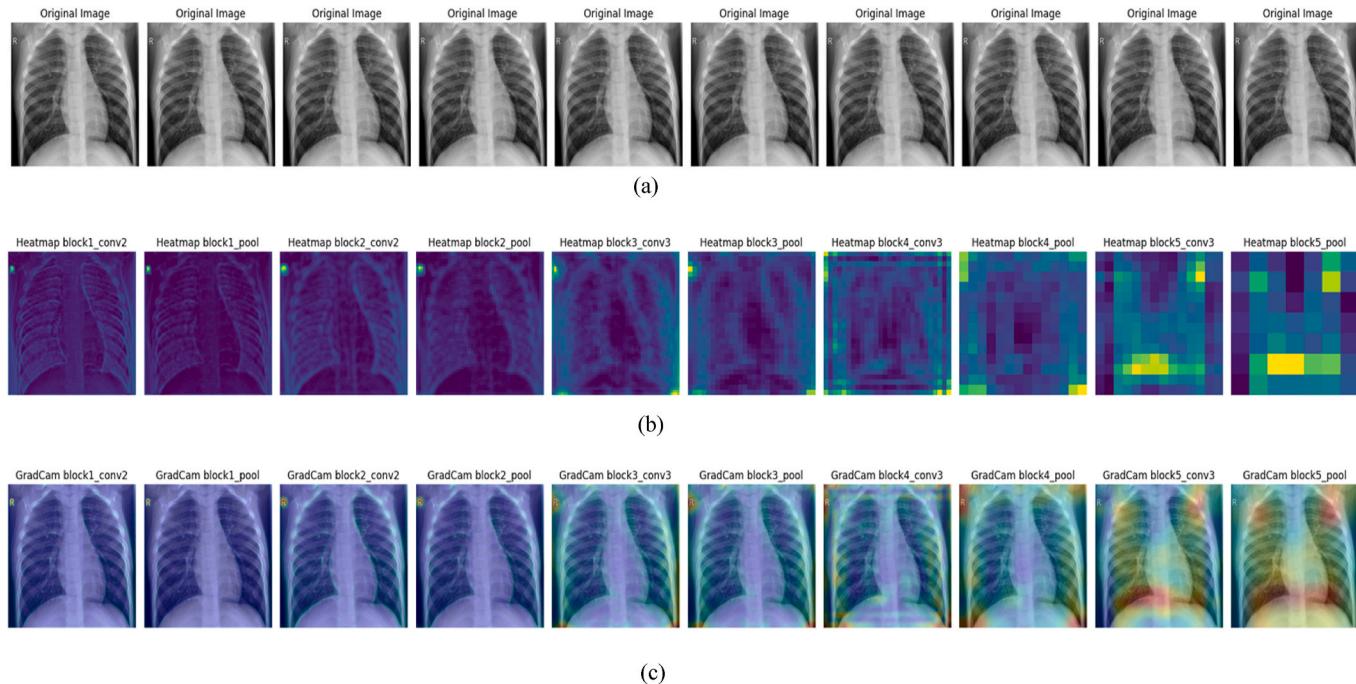


Fig. 10. (a) Sample NORMAL X-Ray images from the dataset, (b) Matrix heatmap of each layer in VGG16, (c) Grad-CAM visualization of each layer in VGG16.

Examining the 'block1_conv2' layer revealed the depiction of the initial feature representation in the photos. The initial layer of the CNN was responsible for detecting the fundamental characteristics of the images. It demonstrated how the optimization technique influenced the activation of these features throughout the early stages of image processing.

The pooling layers used in the model were 'block1_pool', 'block2_pool', 'block3_pool', 'block4_pool', and 'block5_pool'. These layers were utilized to decrease the spatial dimension of the feature representation. Utilizing Grad-CAM in these layers facilitated the comprehension of how optimization strategies impact diminished feature contributions via pooling procedures. The depiction of these layers illustrated the impact of the optimization strategy on the influence of characteristics in the pooling step.

The convolution layers 'block2_conv2', 'block3_conv3', 'block4_conv3', and 'block5_conv3' were positioned in the middle of the CNN network. These layers recorded the intricate and conceptual characteristics present in an image. By applying Grad-CAM to these layers, the researcher revealed the impact of the optimization strategy on the representation of higher-level features. The analysis provided valuable insights into the impact of the optimization strategy on the extraction of intricate aspects in the model's decision-making process.

3.3. Grad-CAM analysis layer Block5_conv3 on each optimization

Grad-CAM computed the gradient of the target class concerning the feature map of the final convolutional layer (block5_conv3) in VGG16. The "block5_conv3" layer was chosen for Grad-CAM because of its position as the final convolution layer before the last pooling layer in the model. The function of this layer typically involves capturing the most intricate and sophisticated characteristics present in an image.

Fig. 12 displays the uppermost row comprising the unaltered image, an X-ray of COVID-19, to be assessed. An activation map was constructed for each optimization in the middle row, showing the relative activation level of each pixel in the original COVID-19 image. At the bottom row, a stack of photos included a mixture of the original COVID-19 photographs and the activation map. The activation map was overlaid onto the original COVID-19 photos in this stack to assess its impact on the images.

To access the network's most sophisticated feature extraction level, the researcher uses the 'block5_conv3' optimization layer. It provided insight into how different adjustments to the model affect the depiction of important visual elements. The visual coverage of this layer was exceptional, offering a wealth of precise information regarding the properties of the image [17,32,51]. Grad-CAM seeks to establish a

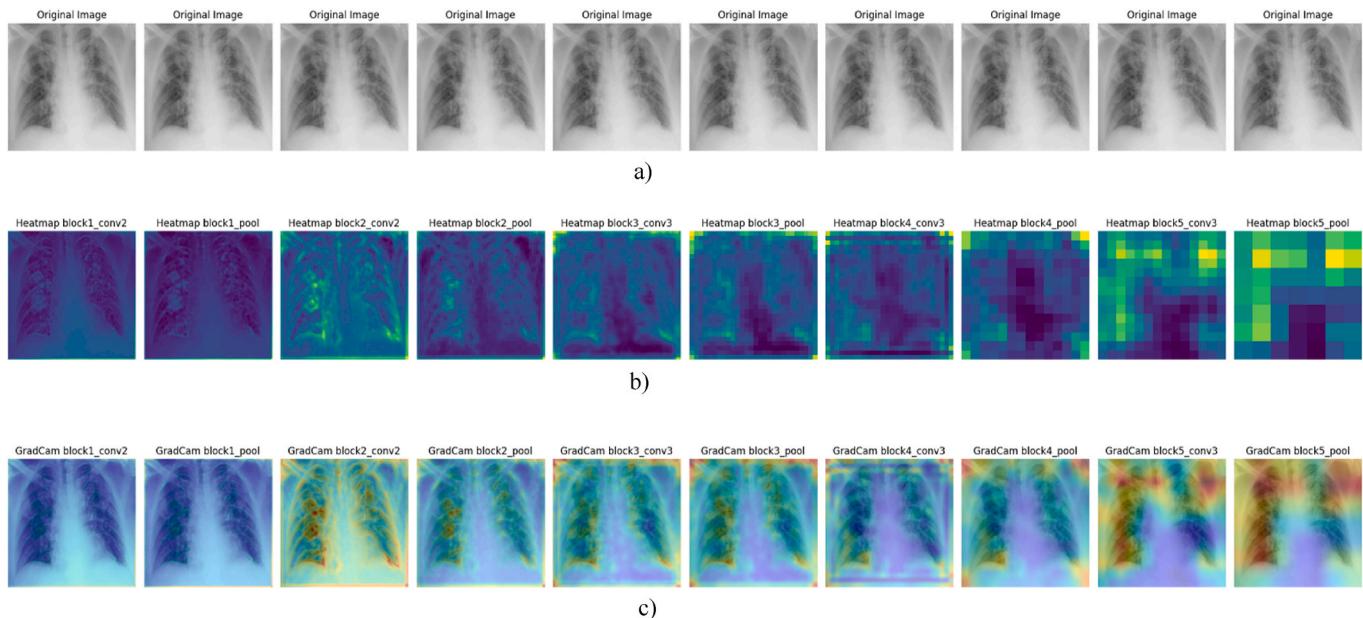


Fig. 11. (a) Example of X-ray Covid-19 images from the dataset, (b) Matrix heatmap of each layer in VGG16, (c) Grad-CAM visualization of each layer in VGG16.

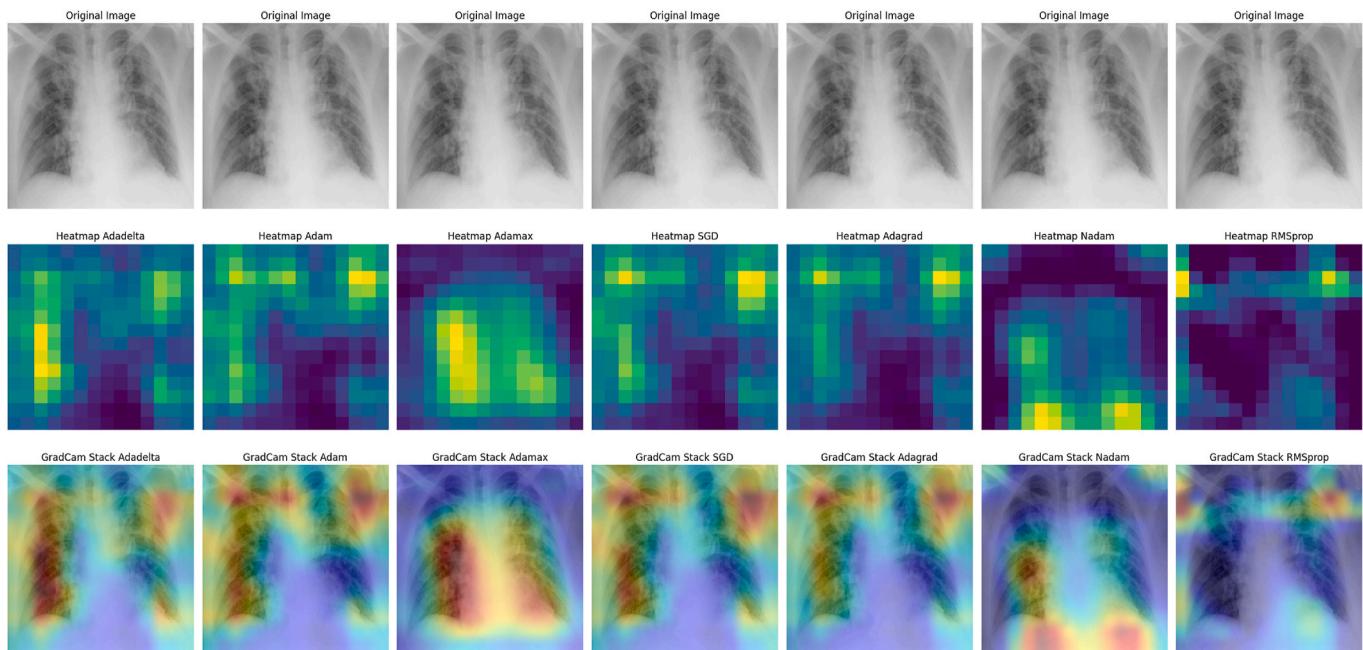


Fig. 12. Original image, heatmap matrix, and Grad-CAM activation of the Covid-19 X-ray images in each optimization.

connection between the convolution layer's feature map and the target class's output.

The layer "block5_conv3" can identify and display the specific regions of the images that substantially impact the final classification outcomes. The selection of this layer was based on its ability to produce a highly abstract visual representation that encompasses particular and crucial characteristics of the target class. As 'block5_conv3' was the final layer before the last pooling layer, the variations in visualization outcomes were more accurate and noticeable for each optimization. Therefore, it is straightforward to compare the impact of each optimization on feature extraction during the final step of image processing.

The Grad-CAM comparison sample exhibits conspicuous red or yellow regions, which signify pronounced or crucial activation areas. These regions typically emphasize the most significant or influential places

that have substantially impacted the model's predictions. The impact on the predicted outcomes becomes more significant as the hue becomes darker. Conversely, shades of blue or light green can be utilized with reduced intensity levels, particularly in areas surrounding the margins or less crucial sections of the image. This color represents a region with a lower degree of activation and does not substantially contribute to the prediction outcome.

Fig. 12 demonstrates that the optimization techniques (Adadelta, Adam, Adamax, SGD, and Adagrad) yielded concentrated identification outcomes in the dominant region of the left lung. The red intensity corresponds to the left lung. Both the Nadam and RMSprop optimization algorithms yielded different results. While Nadam focused on the lower region, the red region in RMSprop did not correspond to the lung area.

The top row of Fig. 13 shows the initial normal X-ray images that

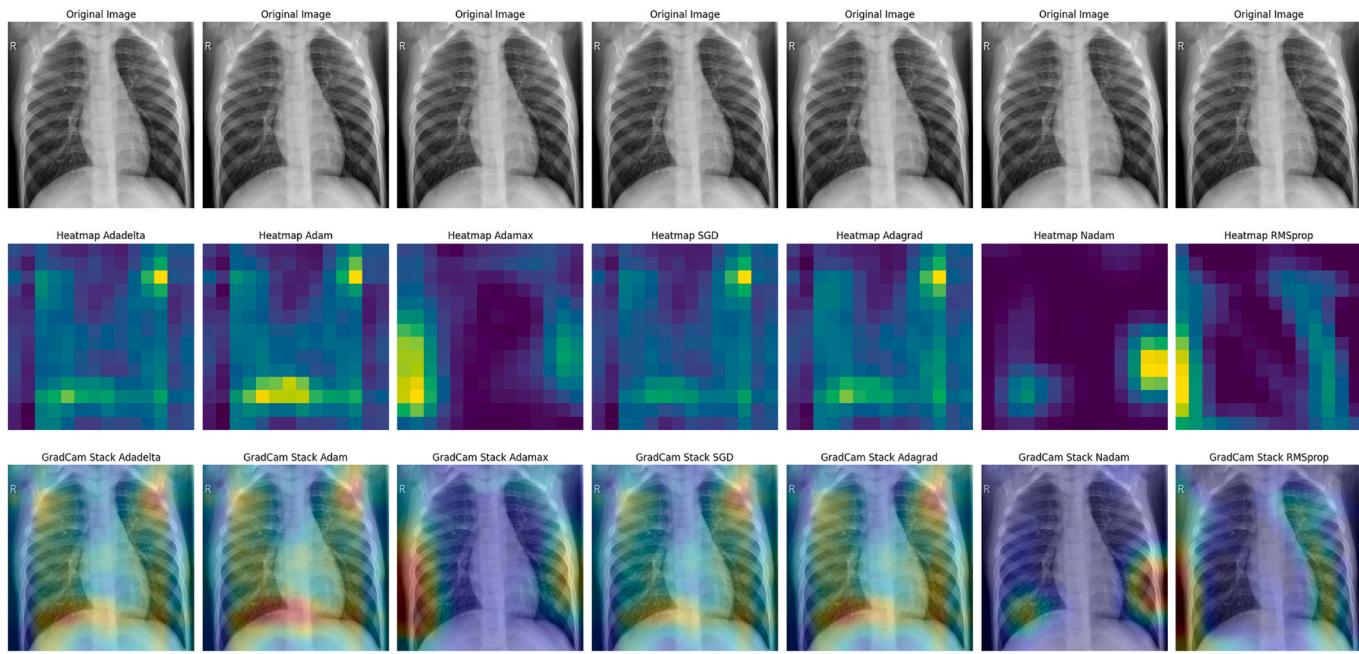


Fig. 13. Original image, heatmap matrix, and Grad-CAM activation of normal X-ray images in each optimization.

were examined. The activation map generated by these optimizations was displayed in the central column. The activation map illustrates the average activation of each pixel in the source images. The original photos and the activation map were merged into a unified image stack, displayed in the lower row. The activation map's colors were superimposed onto the original photos in this stack.

An activation heatmap aims to convey insights into the specific regions within an image that impact the model's prediction outcomes most. Here, the researcher are examining the activation heatmap of the COVID-19 X-ray image. This heatmap reveals the areas of the image that exhibit strong activation and significantly impact the model's prediction outcomes. By utilizing the activation heatmap and implementing these color guidelines, one can better understand the model's attention on particular regions within the COVID-19 image. This can aid in the identification and detection of COVID-19 infection. Fig. 13 displays a Grad-CAM comparison sample, where noticeable disparities between the image under normal settings and the prior condition may be observed. The colors exhibited in the activation region of regular images are comparatively less dim or vibrant than those depicted in Fig. 11 earlier. Nevertheless, the color employed continues to include a significant portion of the lung region, suggesting activation albeit with a lesser intensity than the preceding image.

The Adadelta, Adam, SGD, and Adagrad optimizations yield comparable outcomes when applied to regular pictures. The areas exhibiting higher color intensity are primarily localized in the dominant region of the lung, as seen in the lower section of the picture. The positioning of the red density signifies that this region holds utmost significance or substantially impacts the model's predictions regarding regular images.

The Adamax, Nadam, and RMSprop optimization algorithms yielded distinct outcomes when applied to regular pictures. The Adamax optimization method exhibited limited regions of strong activation, with a predominant blue tint indicating a lower activation level. Furthermore, Nadam and RMSprop optimization techniques exhibited reduced activation levels inside the predominant blue region. This indicated that the lung region did not exhibit substantial activation in typical photos or significantly impact the model's predictions throughout the optimization process.

3.4. Quantitative analysis of Grad-CAM

The RMSE was the assessment statistic typically employed in model comparison analysis in this research. It is a statistical metric used to quantify the magnitude of variations in pixel intensity between two given data sets or images. In utilizing neural network models to detect COVID-19 in lung pictures, the RMSE quantifies the visual disparities between heatmaps generated by various model analyses.

This visualization uses three matrices as a reference: a matrix containing minimum, average (mean), and maximum values. The selection of these three types of values was based on careful consideration to provide a comprehensive picture of the performance of various models in various aspects. The minimal value is selected to accurately represent a model's lowest and most optimal achievable value. A minimal value offers insights into the model's capacity to generate highly accurate predictions and is a benchmark for evaluating overall performance. The average number, or mean, is crucial since it represents the model's overall performance across different scenarios. Within this framework, the average value offers insight into each model's overall performance in examining COVID-19 identification in lung pictures. The maximum value measures the poorest performance or highest limit attainable by an optimization model. This maximum number indicates a hypothetical situation where a model may not perform at its best when making predictions.

The visualization is derived from a juxtaposition of two optimization methodologies, specifically optimization A and optimization B. Seven distinct optimization techniques were employed in this study, yielding 28 measurements in the form of RMSE. This comparative analysis allows for a more precise evaluation of the performance of each optimization method in generating predictions for detecting COVID-19 in lung pictures. Within this particular framework, each of the seven available models underwent seven separate comparisons. Initially, each model underwent self-comparison, yielding a value of 1 due to the same nature of the compared variables. Subsequently, each model underwent a comparison with six additional models, resulting in a value that accurately represented the comparison with the opposing model. The comparison procedure offered a profound understanding of the relative performance of each model in different scenarios and established a solid foundation for comprehending the merits and drawbacks of each

optimization in the analysis of COVID-19 detection in lung pictures.

Fig. 13 displays a matrix that presents the outcomes of model comparisons using pictures of COVID-19-infected lungs. This matrix presents a comprehensive summary of the Root Mean Square Error (RMSE) results from comparing various optimization methods. The matrix findings reveal distinct trends in the relative performance of several optimizations. The primary focus of the initial examination is the average (mean) root mean square error (RMSE) matrix, which is positioned in the center of **Fig. 14**. This number provides an overall measure of performance. It is important to observe that the diagonal elements in the matrix, which have a value of one, are omitted from the analysis as they arise from the same optimization comparison.

The mean matrix indicates that SGD optimization tends to provide lower RMSE values than other optimizations, particularly when contrasted with Adadelta and Adam. In this instance, the RMSE for SGD algorithm is approximately 0.07, but both the Adadelta and Adam algorithms exhibit RMSE values beyond 0.07. The Adagrad optimization method has strong performance, yielding an average the RMSE value of approximately 0.09. Adagrad's RMSE value consistently stays at a low level compared to Adadelta and Adam. In addition, it is observed that Nadam optimization often yields large RMSE values.

The mean matrix exhibits its highest achievement value when comparing the Nadam optimization technique with the Adadelta optimization technique, resulting in an RMSE value of 0.41. This value is the maximum among all comparisons in the average matrix. In addition, when comparing Nadam with Adam, SGD, and Adagrad, a significantly high RMSE value of approximately 0.40 is obtained. Despite the little discrepancy of approximately 0.01 from the current maximum value, it effectively indicates the inconsistency between the predictions generated by Nadam optimization and the original data.

By thoroughly examining the minimum and maximum matrices, the researcher may see intriguing patterns in the efficacy of the different optimizations. When examining the minimal matrix, it becomes evident that SGD and Adagrad optimization exhibit a propensity for generating the most minimal RMSE values. The RMSE value of 0.02 was the lowest in the comparison between SGD and Adadelta. These findings demonstrate that both SGD and Adagrad yield predictions that exhibit greater proximity to the original data in the context of this investigation.

The Adamax optimization algorithm yields varying outcomes when applied to the minimum matrix. Compared to other optimization methods, particularly SGD and Adagrad, Adamax optimization yields higher RMSE values, typically ranging from 0.20 to 0.23. These findings indicate that Adamax optimization may not be well-suited for assessing COVID-19 detection in lung pictures conducted in this work, as it generates predictions that deviate significantly from the original data. Upon

examining the maximum value matrix, the researcher can find that the SGD and Adagrad optimizations exhibit a similar pattern. Despite these two optimizations, the comparison yielded the lowest RMSE value, approximately 0.12. This demonstrates that SGD and Adagrad consistently generate predictions aligning with the data across different analysis scenarios.

The Nadam optimization exhibits varying trends in the matrix of maximum values. The RMSE value generated by the Nadam optimizer tends to be greater, and it even approaches the maximum value when compared to the Adagrad optimizer, with a value of 0.56. This implies that in certain instances, Nadam optimization could generate predictions that deviate from the actual data, as evidenced by elevated RMSE values.

By analyzing the outcomes of these comparisons, this research can offer a full overview of the performance of each optimization method in different analytical scenarios. SGD and Adagrad optimization methods tend to yield the lowest RMSE values in the minimum matrix. This indicates that these optimization techniques effectively generate forecasts that closely align with the original data. Nevertheless, Adamax optimization demonstrates elevated RMSE values, suggesting constraints in some analytical scenarios.

The findings of this research establish a robust basis for gaining a more profound comprehension of the merits and drawbacks of each optimization. This research offers valuable assistance for selecting the best suitable optimization method for more precise analysis by evaluating the relative performance of each optimization in different scenarios. This conclusion highlights the significance of selecting appropriate optimizations when applying COVID-19 detection analysis to lung images, aiming to enhance the overall accuracy and efficacy of the analysis.

Fig. 15 visually represents the correlation between the RMSE matrix and a normal lung picture. Similar to the previous heatmap representation, three correlation matrices encompass distinct values: the minimum, mean, and maximum. The average visualization matrix was selected as the primary reference for the initial study of the comparative findings of each optimization in the matrix.

The provided RMSE correlation matrix offers valuable insights for assessing the comparative effectiveness of several optimizations in the analysis of COVID-19 detection in lung pictures. Initially, upon examining the minimum value in the matrix, it is evident that SGD and Adadelta optimization exhibit superior performance, yielding an RMSE value of approximately 0.07. This suggests that both optimizations can generate predictions that approximate the true value in this particular analytical scenario. Conversely, Adamax optimization exhibits RMSE values that are generally higher (about 0.22) but still lower than Nadam

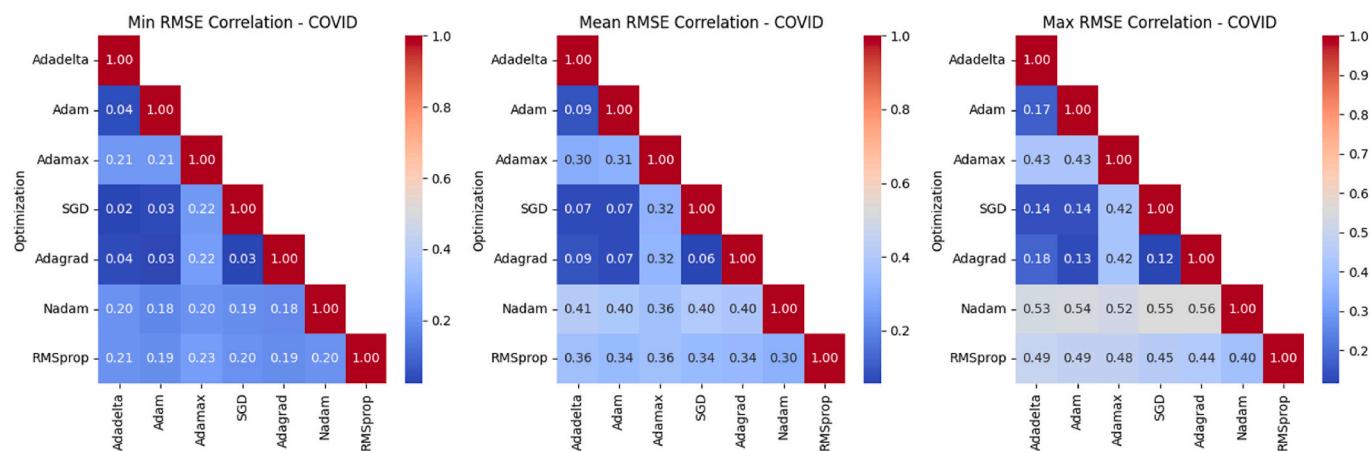


Fig. 14. The correlation of RMSE heatmap matrix of COVID-19 images of 7 optimization methods. The color scale ranges from blue (indicating low RMSE) to red (indicating high RMSE). Each number in the box represents the result of comparing the optimization model with. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

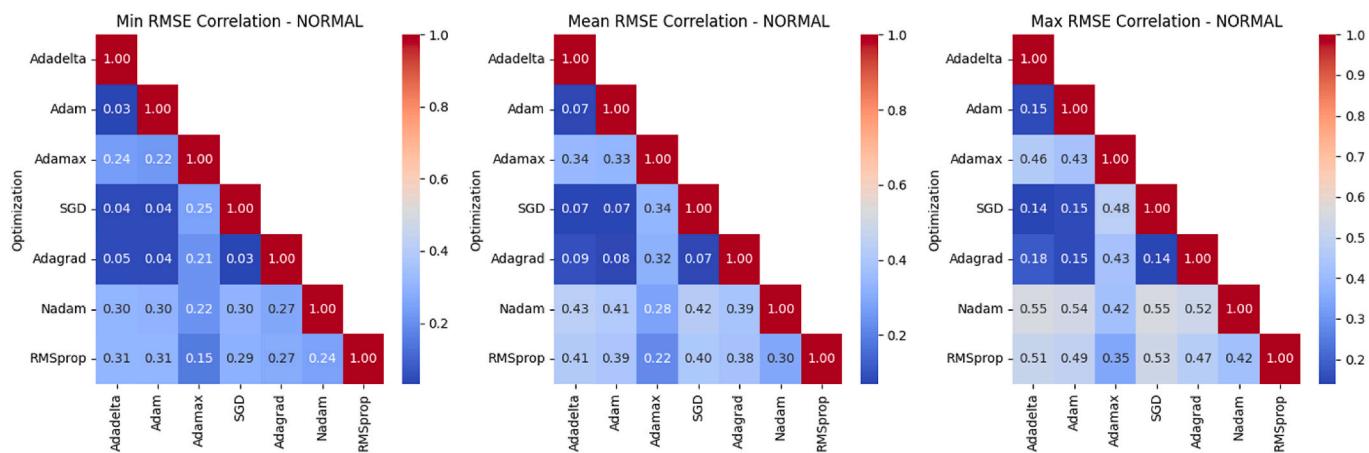


Fig. 15. The correlation of RMSE Heatmap Matrix of normal images of 7 optimization methods. The color scale is blue (low RMSE) to red (high RMSE), each value in the box is the result of comparing the optimization model with other optimizations according to row a. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

and RMSprop in numerous comparisons. This suggests difficulties in accurately forecasting COVID-19 detection data in lung pictures using Adamax optimization.

Moreover, the highest value in the matrix offers valuable information regarding the magnitude of performance variations that can arise from alternative improvements. Nadam and Adagrad's optimization provides the highest RMSE value, approximately 0.56, in the comparative results. This is a notable disparity in the forecasts generated by these two optimizations in the analysis of COVID-19 detection. Hence, to make accurate conclusions regarding the choice of optimizations, it is imperative to possess a profound comprehension of how these optimizations function in different scenarios and analytical contexts.

While these findings offer preliminary understanding, it is crucial to acknowledge that evaluating the effectiveness of an optimization only based on RMSE values is insufficient. When choosing an optimization method, it is important to consider additional criteria such as the rate at which it converges, its capacity to handle noisy data, and its stability. Hence, it is recommended that these findings be utilized as a preliminary basis for additional investigation and assessment to achieve the most effective outcomes in the analysis of COVID-19 detection in lung pictures.

4. Conclusion

His research study utilized optimization strategies to enhance the performance accuracy of deep learning models in detecting COVID-19 by analyzing X-ray images. The evaluation yielded highly favorable results by utilizing seven distinct optimization techniques, specifically Adadelta, Adam, Adamax, SGD, Adagrad, Nadam, and RMSprop. As the recorded average of 99 % indicates, the model's performance demonstrates the exceptional efficacy of all optimization methods. Using the Grad-CAM technique provided good information about significant areas in the model's decision-making, thereby helping scientists understand the work of the optimization methods. This process revealed distinct activation maps for the lung area in both COVID-19 and normal pictures.

Ethical statement

To the best of our knowledge, the research presented in the manuscript "Explainable Optimization of Deep Learning Model for Covid-19 Detection Using Chest Images" adheres to the ethical guidelines and standards for research conduct.

Compliance with Ethical Standards.

1. Conflict of Interest: The authors declare that there is no conflict of interest regarding the publication of this paper.
2. Authorship: All individuals listed as authors have contributed significantly to the conception, design, execution, or interpretation of the reported study. Furthermore, all authors have participated in the drafting or revising of the manuscript and have approved the final version submitted for publication.
3. Data Availability: The datasets used in this research were obtained from public datasets. All private information was not published in this manuscript.
4. Patient consent: This research does not conduct a direct treatment of patients.

CRediT authorship contribution statement

Slamet Riyadi: Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Eka Nova Pramudya:** Writing – original draft, Visualization, Software, Investigation, Data curation. **Cahya Damarjati:** Writing – original draft, Supervision, Methodology, Formal analysis. **Jose Manuel Molina Lopez:** Writing – review & editing, Validation, Methodology, Funding acquisition. **Jesus Garcia Herrero:** Writing – review & editing, Validation, Methodology.

Declaration of competing interest

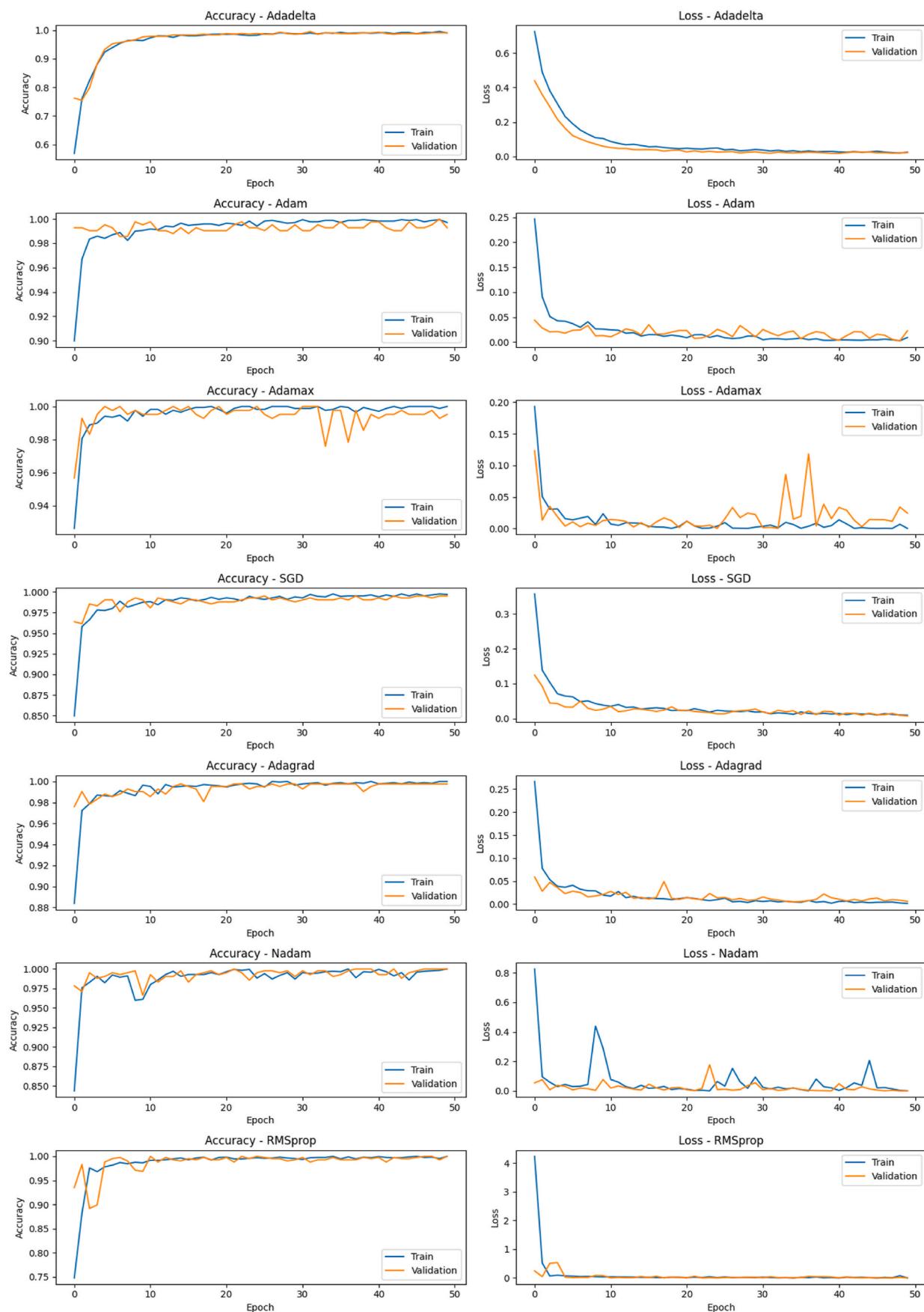
The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Slamet Riyadi reports financial support was provided by University of Muhammadiyah Yogyakarta. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors would like to acknowledge Universitas Muhammadiyah Yogyakarta for the International Research Collaboration Fund 2023 and Sabbatical Leave Program 2023; as well as Universidad Carlos III de Madrid (Applied of Artificial Intelligence Group, GIAA) for collaborating on these programs.

Appendices



References

- [1] Mehraeen E, et al. COVID-19 in pediatrics: a systematic review of current knowledge and practice. *Infect Disord: Drug Targets* Sep. 2021;22(5). <https://doi.org/10.2174/1871526521666210929121705>.
- [2] Mehraeen E, et al. The impact of COVID-19 pandemic on the levels of physical activity: a systematic review. *Infect Disord: Drug Targets* 2023;23(4). <https://doi.org/10.2174/1871526523666230120143118>.
- [3] Seyedalizaghi S, et al. Artificial intelligence in COVID-19 management: a systematic review. *J Comput Sci* Mar. 2023;19(5):554–68. <https://doi.org/10.3844/JCSP.2023.554.568>.
- [4] Hania AA. Mengenal artificial intelligence, machine learning, neural network, dan deep learning. *Jurnal Teknologi Indonesia*. 2017;(6):1–5. Available: <https://www.researchgate.net/publication/320395378>.
- [5] Yadav S, Jasminder, Sandhu K, Pathak Y, Jadhav S. Chest X-ray scanning based detection of COVID-19 using deepconvolutional neural network. Aug. 2020. <https://doi.org/10.21203/rs.3.rs-58833/V1>.
- [6] Binsawad M, Albahar M, Binsawad A. VGG-CovidNet: a Bi-branched dilated convolutional neural network for detection of COVID-19 cases from chest X-ray images. *Semantic Scholar* 2020. <https://doi.org/10.21203/rs.3.rs-133368/v1>.
- [7] Sitaula C, Hossain MB. Attention-based VGG-16 model for COVID-19 chest X-ray image classification. *Appl Intell May* 2021;51(5):2850–63. <https://doi.org/10.1007/S10489-020-02055-X/TABLES/11>.
- [8] Goyal L, Dhull A, Singh A, Kukreja S, Singh KK. VGG-COVIDNet: a Novel model for COVID detection from X-Ray and CT Scan images. *Procedia Comput Sci* Jan. 2023; 218:1926–35. <https://doi.org/10.1016/J.PROCS.2023.01.169>.
- [9] Kareem OS, Al-Sulaifanie AK. Wavelet and modified VGG-16 deep learning to identify COVID-19 based on X-ray images. 2022 international conference on data science and intelligent computing. ICDSC; 2022. p. 267–72. <https://doi.org/10.1109/ICDSIC56987.2022.10075730>.
- [10] Shibu George G, Raj Mishra P, Sinha P, Ranjan Prusty M. COVID-19 detection on chest X-ray images using Homomorphic Transformation and VGG inspired deep convolutional neural network. *Biocybern Biomed Eng* Jan. 2023;43(1):1–16. <https://doi.org/10.1016/J.BBEE.2022.11.003>.
- [11] Tang G, et al. A novel transfer learning network with adaptive input length selection and lightweight structure for bearing fault diagnosis. *Eng Appl Artif Intell* Aug. 2023;123:106395. <https://doi.org/10.1016/J.ENGAAPAI.2023.106395>.
- [12] Swapnarekha H, Behera HS, Nayak J, Naik B. Covid CT-net: a deep learning framework for COVID-19 prognosis using CT images. *J Interdiscipl Math* 2021;24 (2):327–52. <https://doi.org/10.1080/09720502.2020.1857905>.
- [13] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. 3rd international conference on learning representations, ICLR 2015 - conference track proceedings. Sep. 2014. <https://doi.org/10.48550/arXiv.1409.1556>.
- [14] Villamonte AF, Silva PJS, Ronquillo Di GD, Rosales MA, Bandala AA, Dadios EP. Python based defect classification of theobroma cacao bean using fine-tuned visual Geometry Group16. 2021 IEEE 13th international conference on humanoid, nanotechnology, information technology, communication and control, environment, and management, HNICEM 2021. 2021. <https://doi.org/10.1109/HNICEM54116.2021.9731887>.
- [15] Fatima N. Enhancing performance of a deep neural network: a comparative analysis of optimization algorithms. *ADCAIJ: Adv Distr Comput Artif Intell* J Jun. 2020;9(2):79–90. <https://doi.org/10.14201/ADCAIJ2020927990>.
- [16] Albahri S, Yar GNAH. Efficient grad-cam-based model for COVID-19 classification and detection. *Comput Syst Sci Eng* Aug. 2022;44(3):2743–57. <https://doi.org/10.32604/CSSSE.2023.024463>.
- [17] Lim Y, Takahashi H, Ramadhan AA, Baykara M. A novel approach to detect COVID-19: enhanced deep learning models with convolutional neural networks. *Appl Sci* Sep. 2022;12(18):9325. <https://doi.org/10.3390/APP12189325>. 2022, Vol. 12, Page 9325.
- [18] Subramanian M, Adhithiya GJ, Gowthamkrishnan S, Deepti R. Screening covid-19 infection from chest CT images using deep learning models based on transfer learning. *International Conference Computing Methodologies and Communication* 2022;1358–63. <https://doi.org/10.1109/ICCMC53470.2022.9754041>.
- [19] Papernot N, McDaniel P, Goodfellow I, Jha S, Celik ZB, Swami A. Practical black-box attacks against machine learning. *Asia CCS 2017 - proceedings of the 2017 ACM asia conference on computer and communications security*. Apr. 2017. p. 506–19. <https://doi.org/10.1145/3052973.3053009>.
- [20] AroraSanjeev. Opening the black box of deep learning: some lessons and take-aways. *Perform Eval Rev* May 2021;49(1):1. <https://doi.org/10.1145/3543516.3453910>. 1.
- [21] Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. *Int J Comput Vis* Feb. 2020;128(2):336–59. <https://doi.org/10.1007/S11263-019-01228-7/FIGURES/21>.
- [22] Ornek AH, Ceylan M. HayCAM: a novel visual explanation for deep convolutional neural networks. *Trai Du Signal* Oct. 2022;39(5):1711–9. <https://doi.org/10.18280/ts.390529>.
- [23] Cai X, Qiao Z, Gan Y, Wu J, Wu M, Dong H. Sound event detection on GradCAM method with complex pretrained model. *Proceedings - 2022 2nd international conference on computer graphics, image and virtualization. ICCGIV*; 2022. p. 94–7. <https://doi.org/10.1109/ICCGIV57403.2022.00024>.
- [24] Tartaglione E, Barbano CA, Berzovini C, Calandri M, Grangetto M. Unveiling COVID-19 from Chest X-ray with deep learning: a hurdles race with small data. *Int J Environ Res Publ Health* Apr. 2020;17(18):1–17. <https://doi.org/10.3390/ijerph17186933>.
- [25] Maguolo G, Nanni L. A critic evaluation of methods for COVID-19 automatic detection from X-ray images. *Inf Fusion* Apr. 2020;76:1–7. <https://doi.org/10.1016/j.inffus.2021.04.008>.
- [26] Bao Z, Wu Y, Pan G, Liu Y, Hu J, Liu Y. VGG16 fire image fast detection based on multi-layer features fusion. 2022 IEEE 5th international conference on electronics technology, ICET 2022. 2022. p. 1158–62. <https://doi.org/10.1109/ICET55676.2022.9824459>.
- [27] Azhar Y, Anugerah MR, Fahlopy MAR, Yusriansyah A. Image captioning using hybrid of VGG16 and bidirectional LSTM model. *Kinetik Nov*. 2022. <https://doi.org/10.22219/KINETIK.V7I4.1568>.
- [28] Lee KS, Kim JY, Jeon ET, Choi WS, Kim NH, Lee KY. Evaluation of scalability and degree of fine-tuning of deep convolutional neural networks for covid-19 screening on chest x-ray images using explainable deep-learning algorithm. *J Personalized Med* 2020;10(4). <https://doi.org/10.3390/jpm10040213>.
- [29] Sahinbas K, Catak FO. Transfer learning-based convolutional neural network for COVID-19 detection with X-ray images. In: *Data science for COVID-19 volume 1: computational perspectives*. Elsevier; 2021. p. 451–66. <https://doi.org/10.1016/B978-0-12-824536-1.00003-4>.
- [30] Mahmud T, Rahman MA, Fattah SA. CovXNet: a multi-dilation convolutional neural network for automatic COVID-19 and other pneumonia detection from chest X-ray images with transferable multi-receptive feature optimization. *Comput Biol Med* 2020;122. <https://doi.org/10.1016/j.combiom.2020.103869>.
- [31] Mahmud T, Rahman MA, Fattah SA. CovXNet: a multi-dilation convolutional neural network for automatic COVID-19 and other pneumonia detection from chest X-ray images with transferable multi-receptive feature optimization. *Comput Biol Med* Jul. 2020;122:103869. <https://doi.org/10.1016/J.COMBIOMED.2020.103869>.
- [32] Barakat B, Huang Q. Improving reliability of fine-tuning with block-wise optimisation. *Journal of LATEX CLASS FILES* Jan. 2023. <https://doi.org/10.48550/arXiv.2301.06133>.
- [33] Weiss K, Khoshgoftaar TM, Wang DD. A survey of transfer learning. *J Big Data* 2016;3(1). <https://doi.org/10.1186/s40537-016-0043-6>.
- [34] Chollet F, Keras: the Python deep learning library. *Astrophysics Source Code Library*; 2018 [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2018ascl.soft06022C/abstract>. [Accessed 13 June 2024].
- [35] Abadi M, et al. TensorFlow: large-scale machine learning on heterogeneous distributed systems. *Cluster Computing*; Mar. 2016. <https://doi.org/10.48550/arXiv.1603.04467>.
- [36] Zeiler MD. ADADELTA: an adaptive learning rate method. Dec. 2012. <https://doi.org/10.48550/arXiv.1212.5701>.
- [37] Kingma DP, Ba JL. Adam: a method for stochastic optimization. 3rd international conference on learning representations, ICLR 2015 - conference track proceedings. Dec. 2014. <https://doi.org/10.48550/arXiv.1412.6980>.
- [38] Bottou L, Curtis FE, Nocedal J. Optimization methods for large-scale machine learning. *SIAM Rev* 2018;60(2). <https://doi.org/10.1137/16M1080173>.
- [39] Reddi SJ, Kale S, Kumar S. On the convergence of Adam and beyond. 6th international conference on learning representations, ICLR 2018 - conference track proceedings. Apr. 2019. <https://doi.org/10.48550/arXiv.1904.09237>.
- [40] DuchiJohn HazanElad, SingerYoram. Adaptive subgradient methods for online learning and stochastic optimization. *J Mach Learn Res* Jul. 2011;12:2121–59. <https://doi.org/10.5555/1953048.2021068>.
- [41] Ruder S. An overview of gradient descent optimization algorithms. <https://doi.org/10.48550/arXiv.1609.04747>; Sep. 2016.
- [42] Reddi SJ, Kale S, Kumar S. On the convergence of Adam and beyond [Online]. Available: <http://arxiv.org/abs/1904.09237>; Apr. 2019.
- [43] Kingma DP, Ba JL. Adam: a method for stochastic optimization. 3rd international conference on learning representations, ICLR 2015 - conference track proceedings. Dec. 2014 [Online]. Available: <https://arxiv.org/abs/1412.6980v9>. [Accessed 29 December 2023].
- [44] Keskar NS, Nocedal J, Tang PTP, Mudigere D, Smelyanskiy M. On large-batch training for deep learning: generalization gap and sharp minima. 5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings Sep. 2016. <https://doi.org/10.48550/arXiv.1609.04836>.
- [45] Mahendran A, Vedaldi A. Visualizing deep convolutional neural networks using natural pre-images. *Int J Comput Vis* 2016;120(3). <https://doi.org/10.1007/s11263-016-0911-8>.
- [46] Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning deep features for discriminative localization. In: *Proceedings of the IEEE computer society conference on computer vision and pattern recognition*; 2016. <https://doi.org/10.1109/CVPR.2016.319>.

- [47] Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: visualising image classification models and saliency maps. 2nd international conference on learning representations, ICLR 2014 - workshop track proceedings. Dec. 2013. <https://doi.org/10.48550/arXiv.1312.6034>.
- [48] Wang J, Srikantha P. Stealthy black-box attacks on deep learning non-intrusive load monitoring models. IEEE Trans Smart Grid Jul. 2021;12(4):3479–92. <https://doi.org/10.1109/TSG.2021.3062722>.
- [49] de Sousa IP, Vellasco MMBR, da Silva EC. Explainable artificial intelligence for bias detection in COVID CT-scan classifiers. Italian National Conference on Sensors Aug. 2021;21(16). <https://doi.org/10.3390/S21165657>.
- [50] Junaidi A, Lasama J, Adhinata FD, Iskandar AR. Image classification for egg incubator using transfer learning of VGG16 and VGG19. 2021 IEEE international conference on communication, networks and satellite (COMNETSAT). Jul. 2021. p. 324–8. <https://doi.org/10.1109/COMNETSAT53002.2021.9530826>.
- [51] Li J, Liang Z, Xiao C. Transfer learning performance analysis for VGG16 in hurricane damage building classification. Proceedings - 2021 2nd international conference on big data and artificial intelligence and software engineering. ICBASE; 2021. p. 177–84. <https://doi.org/10.1109/ICBASE53849.2021.00041>.
- [52] Lin CJ, Jhang JY. Bearing fault diagnosis using a grad-CAM-based convolutional neuro-fuzzy network. Mathematics Jul. 2021;9(13). <https://doi.org/10.3390/MATH9131502>.