Available online at www.ijournalse.org

Emerging Science Journal

(ISSN: 2610-9182)

Vol. 7, Special Issue "COVID-19: Emerging Research", 2023

Analyzing the Effect of Basic Data Augmentation for COVID-19 Detection through a Fractional Factorial Experimental Design

Mateo Hidalgo Davila [1], Maria Baldeon-Calisto [1, 2*], Juan Jose Murillo [1], Bernardo Puente-Mejia [1], Danny Navarrete [1], Daniel Riofrío [2], Noel Peréz [2], Diego S. Benítez [2], Ricardo Flores Moyano [2]

[1] *Departamento de Ingeniería Industrial and Instituto de Innovación en Productividad y Logística CATENA-USFQ, Colegio de Ciencias e Ingenierías, Universidad San Francisco de Quito (USFQ), Diego de Robles s/n y Vía Interoceánica, Quito 170901, Ecuador.*

[2] *Applied Signal Processing and Machine Learning Research Group USFQ, Colegio de Ciencias e Ingenierías, Universidad San Francisco de Quito (USFQ), Quito 170901, Ecuador.*

Abstract

The COVID-19 pandemic has created a worldwide healthcare crisis. Convolutional Neural Networks (CNNs) have recently been used with encouraging results to help detect COVID-19 from chest X-ray images. However, to generalize well to unseen data, CNNs require large labeled datasets. Due to the lack of publicly available COVID-19 datasets, most CNNs apply various data augmentation techniques during training. However, there has not been a thorough statistical analysis of how data augmentation operations affect classification performance for COVID-19 detection. In this study, a fractional factorial experimental design is used to examine the impact of basic augmentation methods on COVID-19 detection. The latter enables identifying which particular data augmentation techniques and interactions have a statistically significant impact on the classification performance, whether positively or negatively. Using the CoroNet architecture and two publicly available COVID-19 datasets, the most common basic augmentation methods in the literature are evaluated. The results of the experiments demonstrate that the methods of zoom, range, and height shift positively impact the model's accuracy in dataset 1. The performance of dataset 2 is unaffected by any of the data augmentation operations. Additionally, a new state-of-the-art performance is achieved on both datasets by training CoroNet with the ideal data augmentation values found using the experimental design. Specifically, in dataset 1, 97% accuracy, 93% precision, and 97.7% recall were attained, while in dataset 2, 97% accuracy, 97% precision, and 97.6% recall were achieved. These results indicate that analyzing the effects of data augmentations on a particular task and dataset is essential for the best performance.

Keywords:

Medical Image Classification;
COVID-19 Detection;
Convolutional Neural Networks;
Image Data Augmentation;
Design of Experiments;
Fractional Factorial Design.

Article History:

Received:	28	May	2022
Revised:	20	August	2022
Accepted:	06	September	2022
Published:	24	September	2022

1- Introduction

In January 2020, Wuhan, China, reported the first case of the novel coronavirus disease (COVID-19) [1]. On March 12, 2020, the World Health Organization declared the SARS-CoV-2 virus a global pandemic due to its rapid spread and the thousands of deaths it caused [2]. The scientific community recognized the need for and urgency in developing new methods for detecting SARS-CoV-2. The initial methods appeared to successfully identify COVID-19 [3], ranging from nucleic acid amplification tests to antibody detection assays. However, these methods have limitations such as low detection sensitivity, long detection times, frequent false-negative nucleic acid results, and the need for professional technicians to perform them [3].

* **CONTACT**: mbaldeonc@usfq.edu.ec

DOI: http://dx.doi.org/10.28991/ESJ-2023-SPER-01

One of the primary symptoms of COVID-19 is viral pneumonia, which has led to the use of a different screening method that involves examining a potential patient's chest X-ray (CXR) [1]. However, due to the high inter-and intra-observer variability among practitioners, interpreting the image can be difficult and frequently inconsistent. By recognizing and classifying patterns in medical images, Deep Neural Networks, and in particular Convolutional Neural Networks (CNNs), have transformed automated disease detection [4-7]. As a result, several researchers have proposed CNNs designed specifically for COVID-19 chest X-ray classification. These CNNs analyze a chest X-ray image as input and determine whether or not the subject is SARS-CoV-2 virus-infected. The COVID-Net, a CNN architecture created by Wang et al. [8], introduced a lightweight projection-expansion-projection-extension design that enables an improved representation capacity. The model was tested on the open-access COVIDx dataset and achieved a 93% accuracy on the test dataset. Monshi et al. [9] developed CovidXrayNet, a CNN based on the EfficientNet-B0 with hyperparameters and a data augmentation strategy optimized for COVID-19 detection. With only 30 training epochs, CovidXrayNet obtained a 95.82% accuracy on the COVIDx dataset. Finding the best activation function and optimizer to create a model that can recognize COVID-19 from CXR and CT images was the focus of a study conducted by Algarni et al. [10]. The results of the experiments demonstrated that the best accuracy is obtained by combining the stochastic gradient descent algorithm with momentum and ReLU activation functions. Three different CNN architectures were tested by varying the hyperparameter values of the learning rate, batch size, and number of epochs according to a method proposed by Cohen et al. [11]. The results demonstrated that the Xception architecture [12] performed optimally. Khan et al. [13] proposed CoroNet, a deep CNN architecture based on the Xception architecture and pre-trained on the ImageNet dataset. CoroNet achieved an accuracy of 95% for a 3-class classification task.

CCNs typically have tens of millions of parameters and necessitate a large amount of data to avoid overfitting the training set. Overfitting occurs when a neural network models the training set perfectly but performs poorly on unseen data. Overfitting is a major issue in fields where large datasets are unavailable. This is the case with medical image analysis, where obtaining well-annotated data can be time-consuming, costly, and even impossible in some pathologies. Data augmentation is one of the most commonly used techniques to avoid overfitting. Data augmentation expands and diversifies the training set by altering the appearance of the original images or creating new ones.

Due to the scarcity of publicly available COVID-19 datasets, most CNNs developed for COVID-19 detection on chest X-rays employ diverse data augmentation techniques to improve classification accuracy. The most frequently reported operations were basic augmentation techniques. However, how the data augmentation operations affect model prediction is not discussed in detail. Although many studies have been conducted to investigate how data augmentation affects CNN performance on specific datasets for natural image classification [14-16] and medical image classification [17-19], very little research has been conducted to investigate its effect on COVID-19 classification. Furthermore, many authors regard data augmentation as necessary and use it without first analyzing its impact. Elgendi et al. [2] investigated the effectiveness of four geometric data augmentation strategies in COVID-19 detection and concluded that these strategies significantly reduce accuracy. However, the authors did not investigate how each data augmentation operation affected the model's detection accuracy. Given that data augmentation can be a powerful tool for improving a model's generalization and robustness, and that its effect varies across tasks and datasets, it is critical to understand the impact of each operation better when training a CNN for COVID-19 detection.

This study aims to determine the effect of basic augmentation techniques on detecting COVID-19 using CNNs on chest X-ray images. The analysis is carried out using a fractional factorial experimental design, which allows for the determination of which specific data augmentation techniques, and their interactions, have a statistically significant positive or negative effect on classification accuracy. The most commonly used basic augmentation techniques for COVID-19 detection found in the literature are tested using the CoroNet [13] architecture on publicly available COVID-19 datasets. The experiments show that the techniques of zoom, range, and height shift improve model accuracy in dataset 1. Meanwhile, none of the data augmentation techniques affected the performance of the CNN in dataset 2 [20, 21].

Furthermore, we achieve a new state-of-the-art performance on both datasets by training CoroNet with the optimal data augmentation operations and values obtained from the experimental design. This work makes the following contributions:

- Firstly, we statistically investigate how each data augmentation technique and their interactions affect the classification accuracy of a CNN on chest X-ray images for COVID-19 classification. To the best of our knowledge, this is the first study to examine data augmentation operations on the task of COVID-19 detection using a fractional factorial experimental design. Furthermore, the methodology presented here can be used to analyze data augmentation in other tasks and datasets.
- Secondly, we present the optimal data augmentation strategies for obtaining new state-of-the-art results on the two publicly available datasets tested.
- Thirdly, we demonstrate that implementing data augmentation during training does not always improve CNN performance and, in some cases, can even degrade it. Therefore, emphasizing the importance of always analyzing the impact of each augmentation operation on the problem at hand.

2- Literature Review

This Section reviews the data augmentation techniques applied for COVID-19 classification and how experimental design has been used to determine the most suitable machine learning hyperparameter values.

2-1- Data Augmentation Techniques for COVID-19 Classification

Data augmentation is a technique for artificially inflating the training set to avoid overfitting and improve performance. Due to the scarcity of data in medical image analysis, this technique is critical. According to Chalp et al. (2021) [22], data augmentation techniques are classified into three types: basic augmentation techniques, deformable augmentation techniques, and deep learning augmentation techniques. Basic augmentation modifies an image by applying a transformation that maps an image's points to distinct positions or by manipulating the pixel value intensities. Because it is generally quick and simple to implement, this technique is the most commonly used when training deep learning models. Deformable augmentation techniques, on the other hand, are used when basic augmentation does not provide enough variability. The user defines the deformation scale to ensure that the result is clinically plausible. Finally, in deep learning techniques, networks automatically learn image representations and generate new ones. Although data augmentation has been shown to improve test set accuracy in some tasks, other studies have found that using specific data augmentation techniques may negatively impact model performance [2, 23].

Several works have developed CNN architectures to automatically diagnose COVID-19 from chest X-rays. Table 1 compares highly cited studies on this subject, including the name of the neural network, data augmentation operations used, training set size, and accuracy achieved. Since this study aims to analyze how basic data augmentations affect the accuracy of a model, if mentioned in the manuscript, we include the ranges in which the augmentation operations are applied. Vertical flip, horizontal flip, rotation, zoom, width shift, height shift, and shear are the most frequently used data augmentation techniques, as demonstrated.

Table 1. Data Augmentation Techniques for COVID-19 detection in Chest X-ray images

Author	Neural Network Architecture	Data Augmentation Operations	Dataset size per class	Accuracy achieved
Abbas et al. [24]	AlexNet VGG19 ResNet GoogleNet SqueezeNet	Horizontal Flip Vertical Flip Width shift Rotation	COVID: 105 Sars: 11 Normal: 80 Total: 196	Multi-class classification: AlexNet: 89.10 VGG19: 93.10 ResNet: 93.10 GoogleNet: 89.65 SqueezeNet: 82.75
Baldeon et al. [25]	COVID-19 ResNet	Height Shift: -0.13 Width Shift: 0.23 Horizontal Flip Rotation: 187.5 Zoom: range 0.36	COVID: 140 Normal: 140 Pneumonia viral: 140 Total: 420	Multi-class classification: 94
Chowdhury et al. [26]	PDCOVIDNet	Height Shift: 0.15 Width Shift: 0.15 Shear: range 0.10 Rotation: 30 Zoom: 0.10	COVID: 175 Normal: 1072 Pneumonia viral: 1076 Total: 2323	Multi-class classification: 96.58
Goel et al. [27]	OptCoNet GWO-based CNN	Data augmentation operations not specified	COVID: 900 Normal: 900 Pneumonia: 900 Total: 2700	Multi-class classification: 97.78
Khan et al. [13]	CoroNet	Height shift Width shift Shear Re-scale Rotation Zoom	COVID: 284 Normal: 310 Pneumonia Bacteria: 330 Pneumonia Viral: 327 Total: 1251	Binary classification: 97.60
Kumar et al. [28]	SARS-Net	Elastic Deformation Intensity Shift Horizontal Flip Width shift Motion Blur Rotation Zoom	COVID: 258 Normal: 7966 Pneumonia: 5451 Total: 13675	Binary classification: 97.60

Marques et al. [29]	EfficientNetB	Rotation Flipping Noise Blur Shift Distortion	COVID: 404 Normal: 404 Pneumonia: 404 Total: 1212	Binary classification: 99.62 Multi-class classification: 96.70
Nishio et al. [30]	VGG16	Horizontal Flip Width Shift: 0.15 Height Shift: 0.15 Rotation: 15 Zoom: 0.15 Shear: 0.15	COVID: 215 Normal: 533 Pneumonia: 500 Total: 1248	Multi-class classification: 83.6
Rahimzadeh et al. [31]	Xception and ResNet50V2 concatenated	Horizontal Flip Vertical Flip Width Shift Height Shift Rotation Zoom	COVID: 149 Normal: 1634 Pneumonia: 2000 Total: 3783	Multi-class classification: Xception: 91.31 ResNet50V2: 89.79 Concatenated: 91.40
Yoo et al. [32]	Deep learning-based decision-tree classifier	Horizontal Flip Width Shift: 0.20 Height Shift: 0.20 Rotation: 15	COVID: 120 Non COVID: 120 Total: 240	Binary classification 95.0

2-2- Experimental Design

The experimental design, also known as the design of experiments (DOE), is a systematic method for studying multiple factors and a response variable. The objective is to determine if a set of factors and their interactions influence the response variable. DOE allows for the optimization of the response variable in addition to identifying the influencing factors [33]. To model the experimental design, the factors and their corresponding ranges, as well as the number of runs required to identify the relationship between the factors and the response variable, must be specified. The most basic type of experiment is the two-level factorial design, denoted as 2^k. The factors in this experiment have only two possible values: high level or low level. A 2^k design's effects model can be represented as follows:

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk} \quad (1)$$

where μ is the overall mean, τ_i is the effect of treatment i, β_j is the effect of block j, $(\tau\beta)_{ij}$ is the interaction effect of treatment i in block j, and ε_{ijk} is the random error term [33]. One drawback of the 2^k factorial design is that as the number of factors increases, the number of runs grows exponentially. Hence, the fractional factorial design is an alternative method that exploits the sparsity-of-effects principle and selects only a subset of the experiments from the full factorial design to run. Although conclusions about the most important factors can be obtained, some main effects and two-way interactions can be confounded [34]. Fractional factorial designs are a good choice when running an experiment can be costly or time-consuming. The notation used to identify a fractional design is 2^{k-p}, where k refers to the number of factors being investigated and p the number of generators.

Although previous research has not used experimental design to determine the best data augmentation strategies, it has been used to find the best hyperparameters in machine learning algorithms. Lujan-Moreno et al. [34] proposed using the design of experiments methodology to screen for the most significant hyperparameters, followed by a Response Surface Methodology to fine-tune their value. Staelin et al. [35] developed an algorithm inspired by the DOE methodology that iteratively refines the boundaries and resolution of a search grid. The algorithm is tested on the hyperparameter optimization of a least-squares SVM regression. F.Chou et al. [36] conducted research to determine the combination of hyperparameters that improves the performance of a CNN for image recognition. They define the Uniform Experimental Design (UED) concept as a space-filling design that can be used when the underlying model is unknown [37].

3- Research Methodology

This study investigates the effect of basic data augmentation techniques on the classification accuracy of a CNN for COVID-19 detection on chest X-ray images. The tests are performed on two publicly available COVID-19 datasets of varying sizes. Furthermore, the CoroNet CNN architecture [13] was chosen because it has good prediction performance, uses model parameters efficiently, and is open-source code. Finally, a $2^{(7-1)}$ factorial experimental design is used to understand the statistical impact of each data augmentation operation. This Section describes the COVID-19 datasets used, reviews the CoroNet architecture, explains the data augmentation factors examined, and presents the experimental design methodology.

3-1- Datasets Description

Two de-identified publicly available COVID-19 chest X-ray datasets are selected. The first dataset, denominated as dataset 1 throughout this work, is presented by Khan et al. [13] and is a recollection of images from the Github repository of Cohen et al. [11], images from the RSNA, Radiopedia, and Kaggle databases [21]. The images in the database are divided into four categories: COVID-19 positive, normal, bacterial pneumonia, and viral pneumonia. Following the approach of [13], the dataset is modified to have only three classes (COVID-19 positive, normal, and pneumonia) by combining bacterial and viral pneumonia observations into one class. The dataset contains 1678 images, of which 15% are used for testing, 17% for validation, and 68% for training. Zargari Khuzani et al. [20] made the second dataset available, referred to as dataset 2 throughout this paper. This dataset contains images from three classes: COVID-19 positive, normal, and pneumonia. The dataset is already balanced and has 381 images in total. 10% of the images are used for testing, 18% for validation, and 72% for training. For a fair comparison of methods, the partition used on both datasets is the same as that used by Khan et al. [13] and Zargari Khuzani et al. [20] in their respective works. Table 2 summarizes the number of images per dataset and class. In addition, randomly selected images from each dataset and class are shown in Figure 1. Both images are preprocessed by rescaling by 1/255, applying ZCA whitening, dividing pixel values by standard deviation, and setting the input mean to 0.

Table 2. Summary of classes and number of images per datasets

Data Repository	Images Classes	Train and Validation Size	Test Size
Khan et al. [13]	COVID-19	300	20
	Normal	368	77
	Pneumonia	768	145
Zargari Khuzani et al. [20]	COVID-19	126	14
	Normal	90	10
	Pneumonia	126	15

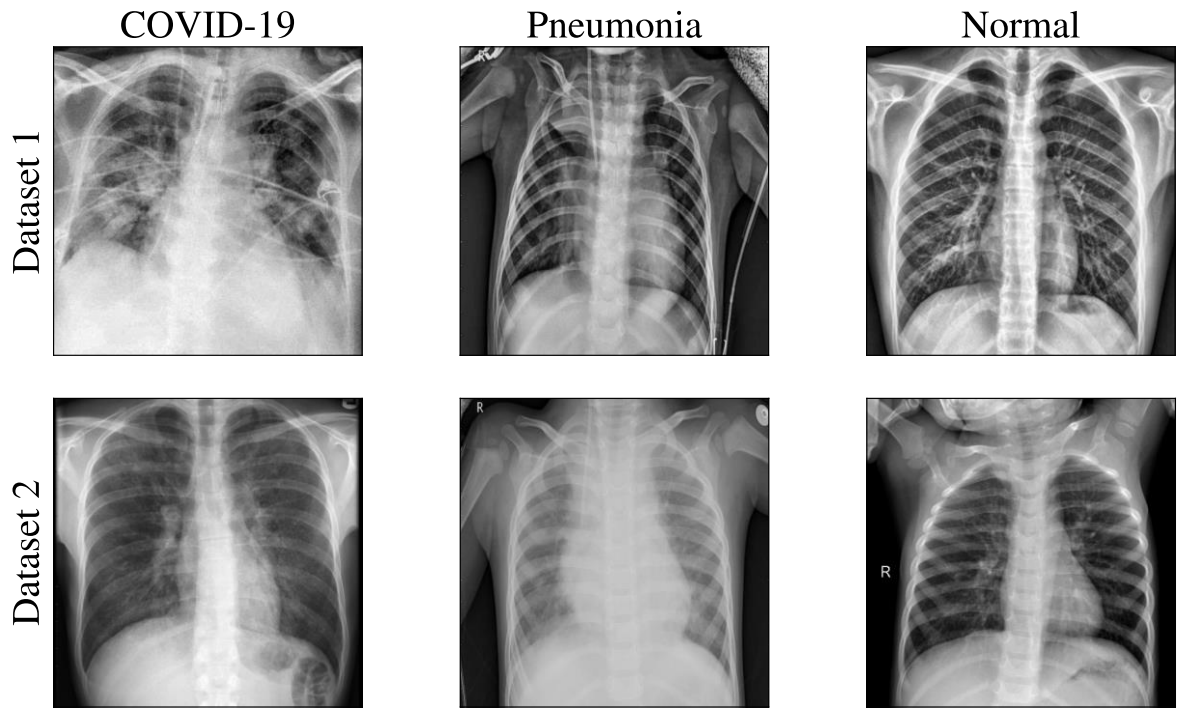

Figure 1. Randomly selected images from each dataset and class

3-2- Model Architecture

The CoroNet architecture [13] used in this study is a CNN designed to detect COVID-19 from chest X-ray images. CoroNet is based on the Xception architecture [12], an extreme version of the Inception model. The Xception model uses depthwise separable convolution layers with residual connections that replace the classic n×n×k convolutional operation with a 1×1×k pointwise convolution followed by an n×n channel-wise spatial convolution. The CoroNet architecture is based on the Xception architecture, with the addition of a dropout layer and two fully connected layers at the end. In addition, it includes a batch normalization layer after all convolutional and separable convolutional layers to reduce training time. Figure 2 depicts the architectural details of CoroNet.

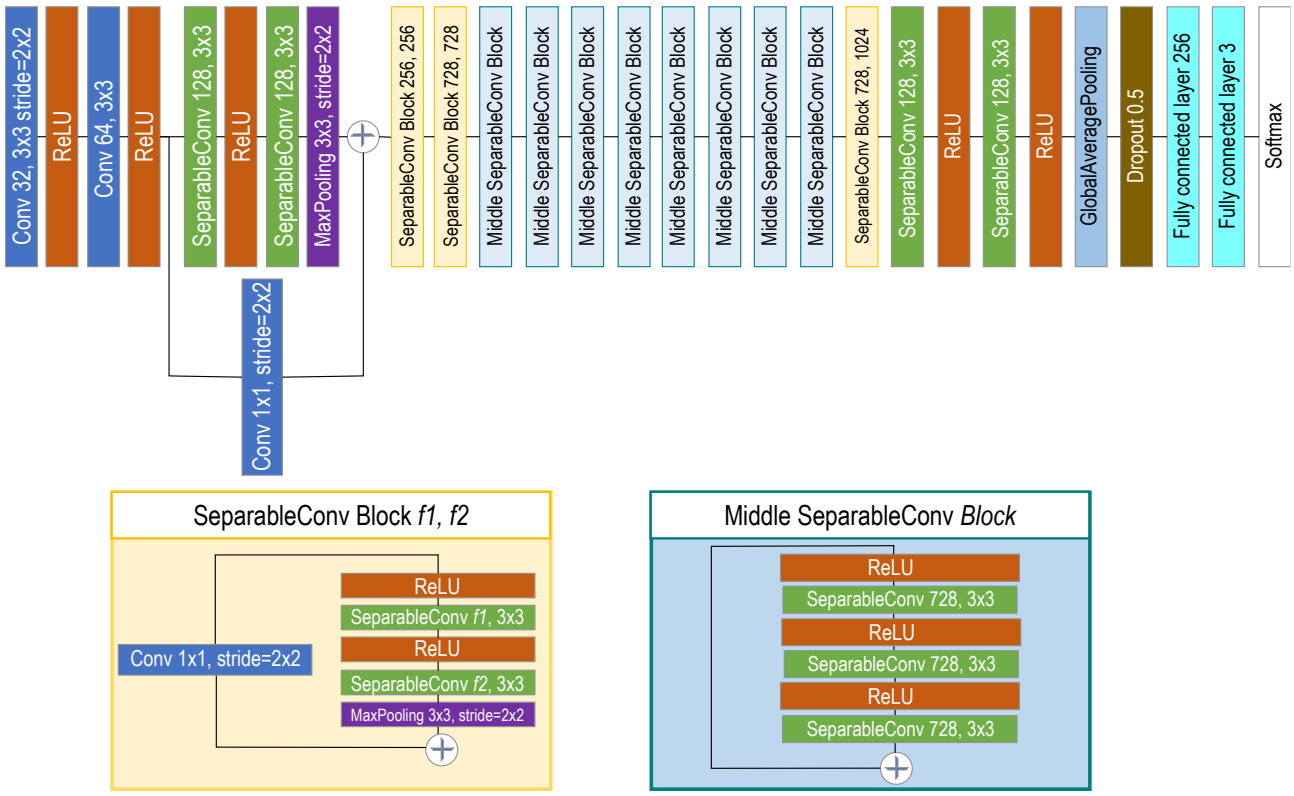

Figure 2. The CoroNet Architecture proposed by Khan et al. [13]. The numbers following the convolutional layers and separable convolutional layers refer to the number of filters and the size of the kernel window. Similarly, the number following the fully connected layers refers to the number of nodes in that layer. Note that all convolutional and separable convolutional layers are followed by a batch normalization layer not shown in the graph.

The architecture is relatively simple compared to competitor CNNs proposed for the same task, allowing for better evidence of the impact of different data augmentation operations on classification accuracy. Furthermore, CoroNet outperformed other studies in the literature in the task of COVID-19 multi-class classification. Finally, the CoroNet code is open source and is available on the author's Github repository [13].

3-3- Image Data Augmentation

The most commonly used basic data augmentation techniques for detecting COVID-19 found in the literature (refer to Table 1) were selected for testing in this work. Vertical and horizontal flip, rotation, zoom/scaling, width and height shift, and shear are the seven operations considered. Table 3 defines these operations, while Figure 3 depicts how each operation affects the appearance of an image. Table 4 also shows the high and low-level values tested for data augmentation operations. The values have also been determined based on a review of the literature. For each data augmentation operation, a high or low-level value is selected and used to train the CNN for a specific run. During training, the Data Image Generator tool from the TensorFlow library is used to augment the data on the fly.

Table 3. Definition of the Data Augmentation Operations

Data Augmentation Operation	Definition
Vertical flip	Performs a reflection of the original image along the vertical axis, swapping the upward and downward sections of the image. The function receives a Boolean as input data.
Horizontal flip	Performs a reflection of the original image along the horizontal axis, swapping the left and right sections of the image. The function receives a Boolean as input data.
Rotation range	The function receives an input angle in the form of an integer value and rotates the image by an angle randomly selected between zero and the input angle.
Zoom range	The function receives as input a float value and performs a random zoom between zero and the input value.
Width shift range	Shifts the image on the horizontal axis. The function receives as input a float value that represents a fraction of the total width the image can be shifted.
Height shift range	Shifts the image on the vertical axis. The function receives as input a float value that represents a fraction of the total height the image can be shifted.
Shear range	Performs a shear distortion in a counter-clockwise direction in degrees. The function receives an input angle in the form of a float value.

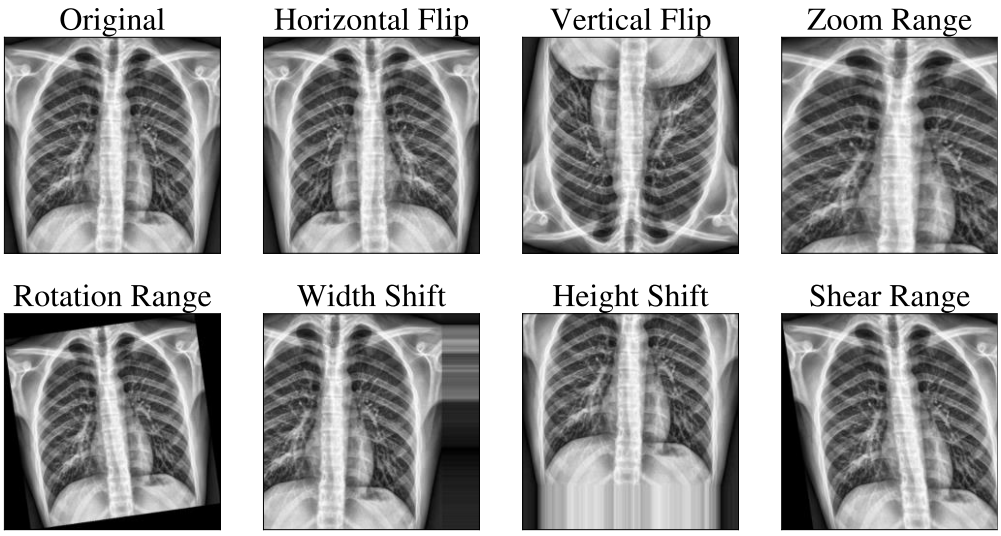

Figure 3. The Appearance of Images Altered by Data Augmentation

Table 4. Levels of the Data Augmentation Operations for the Experimental Factorial Design

Data Augmentation Operation	Low Level	High Level
A: Vertical flip	False	True
B: Horizontal flip	False	True
C: Rotation range	0	15
D: Zoom range	0	0.15
E: Width shift range	0	0.20
F: Height shift range	0	0.25
G: Shear range	0	0.20

3-4- Experimental Design Process

A factorial fractional experimental design is used, with the factors analyzed being the seven data augmentation operations listed in Table 3 and the response variable being the test set classification accuracy. The levels of the factors within the experimental model are shown in Table 4. As shown in Figure 4, the methodology used to implement the experimental design consists of four steps.

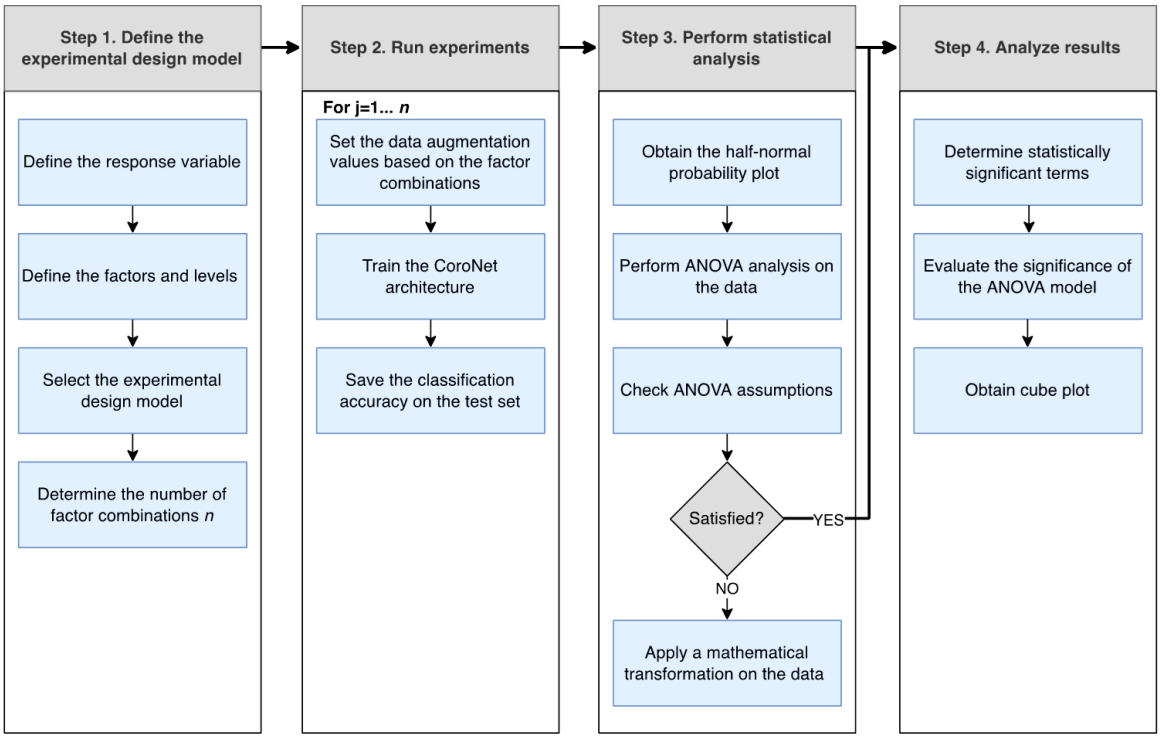

Figure 4. Flowchart of the research methodology

Step 1: The experimental design model is defined in this step by determining the response variable, factors, and levels. In addition, the experimental design model is selected. Implementing a fractional factorial $2^{(7-1)}$ model results in 64 distinct data augmentation parameter value combinations. The Design-Expert software V.13 is used to obtain the specific combinations to test. Given the computational resources and time required to train a CNN for COVID-19 classification, this fractional factorial design reduces the number of runs required. Furthermore, the model has a VII resolution, which means that the main effects, second and third-order interactions, will be aliased with high-order interactions, resulting in simple structures. As shown in Equation 2, the selected response variable is the classification accuracy of the test set.

$$Accuracy = \frac{Nº\ of\ images\ correctly\ classified}{Total\ Nº\ of\ images} \quad (2)$$

Step 2: The experiments are carried out in this step. CoroNet is trained with the specific data augmentation parameters and the response variable saved for each of the 64 data augmentation combinations obtained in step 1. The Adam optimizer is used to train the models, with a learning rate of 0.0001, a batch size of 25, and 300 training epochs. The models are implemented with Keras and Tensorflow 2.0 as the backend. Furthermore, the experiments are carried out on a Linux Ubuntu 18.04 platform equipped with an Nvidia Tesla V100 graphics card using an Nvidia Docker VM. To perform the statistical analysis, the classification accuracy of the test set is entered into the Design-Expert software.

Step 3: The results of the experiments are statistically analyzed, and significant factors and interactions are selected using a half-normal probability plot and the Analysis of Variance (ANOVA). Furthermore, the ANOVA assumptions are validated: a) the residuals are normally distributed, and b) the observations are chosen randomly and are independent. c) The variances are homogeneous. If the assumptions are not met, the data is mathematically transformed.

Step 4: The ANOVA table is examined to determine the statistically significant terms. In addition, the significance of the mathematical model developed to predict the response variable is evaluated. Subsequently, a cube plot is obtained to understand the relationship between the interactions and the response variable.

4- Results

The experimental design results for the two datasets are presented in this Section. Furthermore, the CoroNet architecture is compared to other proposals in the literature using the "optimal" data augmentation technique.

4-1- Classification Accuracy on Dataset 1

Figure 5 shows a half-normal probability plot with the experimental design results on dataset 1. The data augmentation operations of zoom (factor D), height shift (factor G), and the combination of the factors vertical flip, horizontal flip, and shear (third level interaction ABF) have effects that differ from 0 and will be statistically analyzed using the ANOVA. As this is an unreplicated design, the other terms will constitute the error portion of the experiment. Due to the hierarchy model building technique, non-significant terms such as shear (factor F) and horizontal flip and shear interaction will be added to the final model (second-level interaction BF).

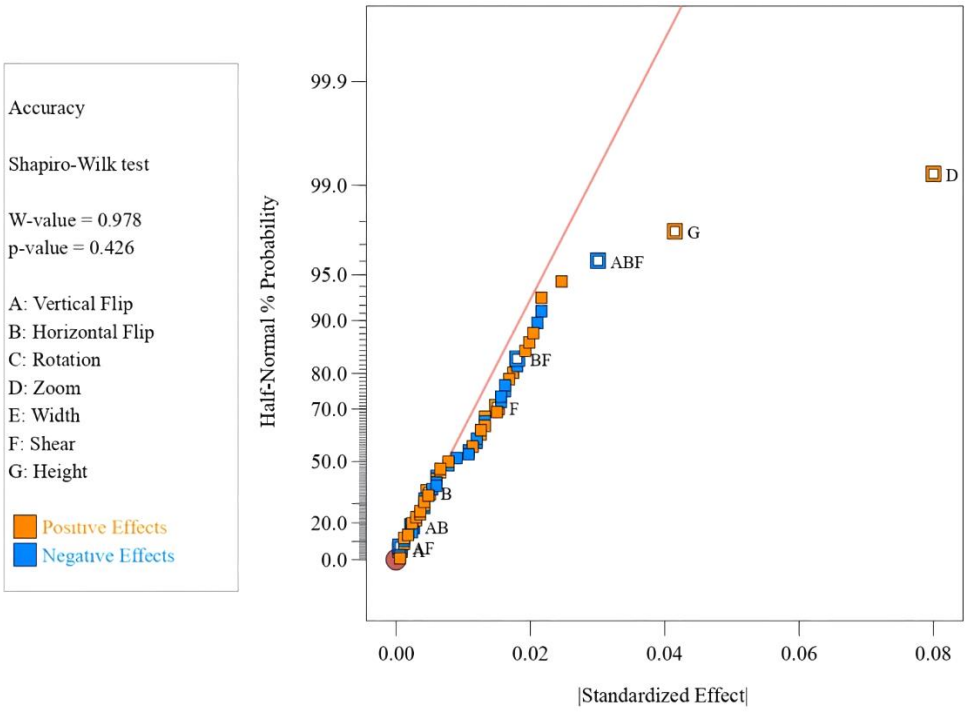

Figure 5. Half Normal Plot for Dataset 1

As shown in Table 5, the ANOVA is performed after selecting the significant factors and interactions with the suggested hierarchy terms. The results indicate that the data augmentation operations of zoom (factor D, p-value <0.0001), height-shift-range (factor G, p-value 0.0009), and the combination of the factors of vertical flip, horizontal flip, and shear (ABF interaction, p-value 0.0133) are not only significant but positively affect the accuracy of the model. In addition, a cube analysis is used to analyze the interaction, as shown in Figure 6. This plot shows the best combination of data augmentation operations. When vertical flip (factor A) is true, horizontal flip (factor B) is false, zoom (factor D) is high, shear (factor F) is high, and height shift (factor G) is high, the optimal data augmentation configuration is achieved. All other operations, such as rotation and width shift, are irrelevant and do not affect the model's accuracy. Table 6 summarizes the findings.

Table 5. ANOVA Results for Dataset 1

Source	Sum of Squares	df	Mean Square	F-value	p-value
Model	0.1660	9	0.0184	7.73	**< 0.0001**
A-Vertical Flip	6.250E-06	1	6.250E-06	0.0026	0.9594
B-Horizontal Flip	0.0004	1	0.0004	0.1677	0.6838
D-Zoom	0.1106	1	0.1106	46.35	**< 0.0001**
F-Shear	0.0039	1	0.0039	1.64	0.2061
G-Height	0.0298	1	0.0298	12.47	**0.0009**
AB	0.0001	1	0.0001	0.0419	0.8385
AF	6.250E-06	1	6.250E-06	0.0026	0.9594
BF	0.0056	1	0.0056	2.36	0.1305
ABF	0.0156	1	0.0156	6.55	0.0133
Residual	0.1288	54	0.0024		
Cor Total	0.2948	63			

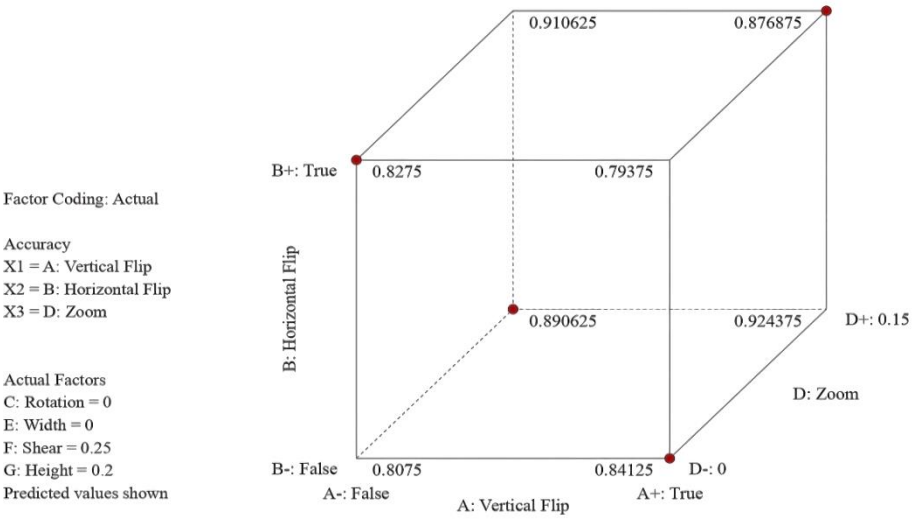

Figure 6. Cube analysis of the predicted classification accuracy on dataset 1. The optimal combination of data augmentation operations is when vertical flip (factor A) is true, horizontal flip (factor B) is false, zoom (factor D) is at its high level, shear (factor F) is at its high level, and height shift (factor G) at its high level.

Table 6. Data Augmentation Operations Effects on Datasets

Data Augmentation Operation	Dataset 1	Dataset 2
A: Vertical flip	Positively affect	Does not affect
B: Horizontal flip	Negatively affect	Does not affect
C: Rotation range	Does not affect	Does not affect
D: Zoom range	Positively affect	Does not affect
E: Width shift range	Does not affect	Does not affect
F: Height shift range	Positively affect	Does not affect
G: Shear range	Positively affect	Does not affect

4-2- Classification Accuracy on Data Set 2

As mentioned in step 3, the ANOVA assumptions must be met to validate the analysis. The assumptions were not met in this dataset. As a result, the response variable is transformed by applying an ArcSine to the square root of the response variable (Equation 3), which has been shown to provide a high percentage of non-normality, heterogeneity of variance, and nonadditivity correction [38].

$$Accuracy\ (Transformed) = ArcSine\left(\sqrt{Accuracy}\right) \quad (3)$$

Figure 7 shows the resulting half-normal probability plot for dataset 2. The graph shows that horizontal flip (factor B), the interaction of rotation range, zoom range, and sheer range (CDF interaction), and the interaction of vertical flip, width shift range, and height shift (AEG interaction) may be significant. Due to the hierarchy assumptions, more factors are included in the model, as in dataset 1, even though these terms are insignificant.

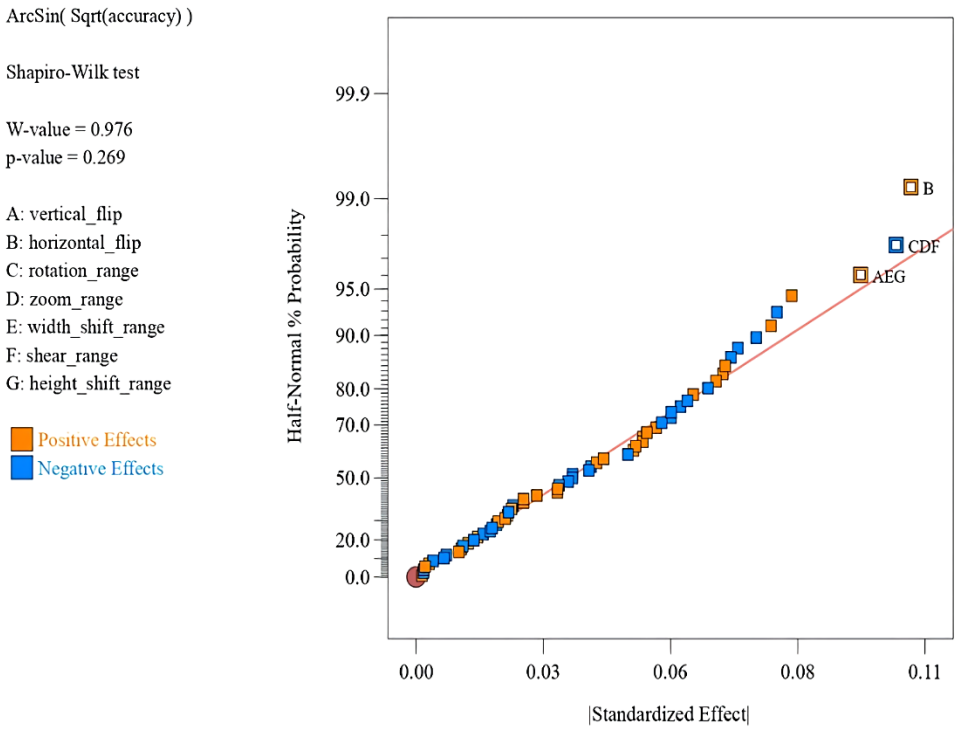

Figure 7. Half Normal Plot for Dataset 2

The ANOVA is performed with the significant factors and interactions. Table 7 shows that the model is not statistically significant, preventing us from determining the specific levels of each augmentation operation in dataset 2. Data augmentation should not be used on dataset 2 in this context. Table 6 summarizes these findings.

4-3- Data Augmentation Analysis

Table 8 summarizes the statistical significance of data augmentation operations and their effects on classification accuracy, whereas Table 8 depicts the best data augmentation strategy. Data augmentation has a significant effect on the classification accuracy in dataset 1. Vertical flip, zoom, height shift, and shear, in particular, improve the model's accuracy. Furthermore, zoom and height shift have the greatest effect, as evidenced by the small p-value in the ANOVA table. Rotation and width shifts are insignificant, meaning they do not affect the model's accuracy, either positively or negatively. Because its implementation raises training costs, it should not be included in the data augmentation strategy. Finally, horizontal flip harms the model's performance, so avoiding using it during training is essential.

However, the results in dataset 2 differ. None of the data augmentation operations have a statistically significant effect on the model's performance. This means that data augmentation during training does not affect the model's ability to generalize and should be avoided to eliminate unnecessary computations. The optimal data augmentation configurations discovered using the implemented methodology differ for the two datasets. When the data augmentation strategy for dataset 1 includes vertical flip, zoom with a range of 0.15, height shift with a range of 0.25, and shear with a range of 0.20, the best performance is obtained. It is best not to use data augmentation during training in dataset 2. These findings show that, even though we are analyzing the same classification task in similar datasets (same image modality and anatomical region portrayed), data augmentation strategies are not transferable between datasets and must

be performed independently for each case. In the case of medical images, data augmentation must produce images with a similar distribution to the original dataset. If this is not the case, the synthetic images can reduce the recognition capacity of a CNN by introducing noise. In the experiments presented, this is the case of the horizontal flip operation, which reduces test accuracy when applied to dataset 1. Another interesting finding is that data augmentation does not always benefit small datasets. The model trained in dataset 2, which contains 381 images, showed no improvement when any data augmentation operation was included. We believe that the dataset's limited information and diversity made it very difficult for the tested augmentation operations to produce plausible images that introduced "new" training information to the model. As a result, the model learned nothing beyond what the original dataset provided. It is recommended in this case to assess whether more advanced augmentation techniques can improve the model's performance.

Table 7. ANOVA Results for Dataset 2

Source	Sum of Squares	df	Mean Square	F-value	p-value
Model	0.6480	15	0.0432	1.36	**0.2062**
A-vertical_flip	0.0062	1	0.0062	0.1961	0.6599
B-horizontal_flip	0.1831	1	0.1831	5.76	0.0203
C-rotation_range	0.0028	1	0.0028	0.0892	0.7665
D-zoom_range	0.0109	1	0.0109	0.3430	0.5608
E-width_shift_range	0.0034	1	0.0034	0.1056	0.7466
F-shear_range	0.0384	1	0.0384	1.21	0.2769
G-height_shift_range	0.0015	1	0.0015	0.0478	0.8278
AE	0.0007	1	0.0007	0.0215	0.8841
AG	0.0020	1	0.0020	0.0639	0.8015
CD	0.0574	1	0.0574	1.81	0.1853
CF	0.0149	1	0.0149	0.4700	0.4963
DF	0.0050	1	0.0050	0.1588	0.6920
EG	0.0016	1	0.0016	0.0513	0.8218
AEG	0.1477	1	0.1477	4.65	0.0361
CDF	0.1722	1	0.1722	5.42	0.0242
Residual	1.53	48	0.0318		
Cor Total	2.17	63			

Table 8. Optimal Configuration of Data Augmentation Operations on Datasets

Data Augmentation Operation	Dataset 1	Dataset 2
A: Vertical flip	True	False
B: Horizontal flip	False	False
D: Zoom range	0.15	0
F: Height shift range	0.25	0
G: Shear range	0.20	0

4-4- Benchmarking

CoroNet is fully trained in the corresponding dataset using the optimal data augmentation strategies presented in Table 8, and its performance is evaluated on the test set. The metrics accuracy, precision, F1-score, and recall are used for evaluation. The metrics' formulation is presented in Equations 2 to 6. Meanwhile, the CoroNet results for each dataset and class are shown in Table 9.

$$Precision = \frac{True_{Positive}}{True_{Positive} + False_{Positive}} \quad (4)$$

$$Recall = \frac{True_{Positive}}{True_{Positive} + False_{Negative}} \quad (5)$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

Table 9. Evaluation metrics of the CoroNet with the optimized data augmentation strategy on dataset 1 and dataset 2

Class	Dataset 1 Precision	F1-Score	Recall	Dataset 2 Precision	F-Score	Recall
COVID-19	83%	91%	100%	100%	100%	100%
Normal	96%	96%	96%	91%	95%	100%
Pneumonia	100%	99%	97%	100%	97%	93%

The results from Table 9 demonstrate that the data augmentation strategies defined in this study positively affect classification accuracy. In dataset 1, the precision, F1- score, and recall values for COVID-19 class are 83%, 91%, 100% respectively. For dataset 2, the precision, F1-score, and recall values for the COVID-19 class are 100%, 100%, and 100%, respectively. Furthermore, Figure 8 shows the confusion matrix for each experiment, demonstrating that all of the mismatches for the COVID-19 class fall in the false-positive region, with no false negatives registered. The latter is an important discovery because, in this case, a false positive is a less costly error than diagnosing a person with COVID-19 as non-infected.

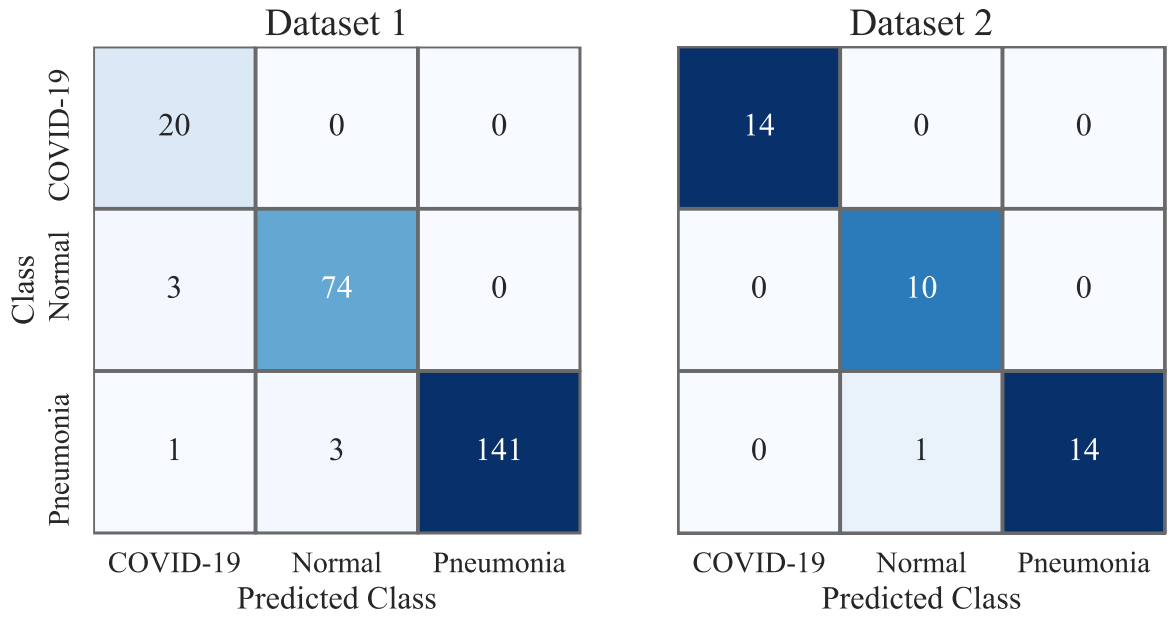

Figure 8. Confusion Matrix for Datasets 1, 2

Tables 10 and 11 show the average class-wise results for the proposed CoroNet and state-of-the-art models published using datasets 1 and 2, respectively. Furthermore, the results of CoroNet with the original data augmentation strategy implemented by Khan et al. [13] and tested on the corresponding datasets (denoted as Original DA in CoroNet) are shown in both Tables to provide a richer comparison. Finally, CoroNet without data augmentation, or CoroNet without DA, is evaluated for dataset 1.

Table 10. Evaluation metrics of the proposed CoroNet with the optimized data augmentation strategy (*) and competing state-of-the-art models on COVID-19 classification on dataset 1. The values presented are as reported in the published papers.

Model	Accuracy	Precision	F1-Score	Recall
Original DA in CoroNet [13]	95.0%	95.0%	95.6%	96.9%
CoroNet without DA	79.0%	79.0%	71.0%	78.3%
Optimized DA in CoroNet (*)	**97.0%**	**93.0%**	**95.3%**	**97.7%**

Table 11. Evaluation metrics of the proposed CoroNet with the optimized data augmentation strategy (*) and competing state-of-the-art models on COVID-19 classification on dataset 2. The values presented are as reported in the published papers; a dash means the specific metric has not been reported.

Model	Accuracy	Precision	F1-Score	Recall
COVID-Classifier [20]	94.0%	96.0%	94.3%	-
Baldeon et al. [25]	94.0%	93.6%	93.3%	93.6%
Original DA in CoroNet [13]	95.0%	96.0%	95.3%	95.6%
Optimized DA in CoroNet (*)	**97.0%**	**97.0%**	**97.3%**	**97.6%**

Although the purpose of this study is to investigate the impact of various data augmentation techniques on the classification accuracy of a CNN, with the optimal data augmentation strategy discovered, it was possible to improve the results of the base CoroNet architecture and achieve new state-of-the-art performance. On dataset 1, the optimized CoroNet reached a 97% accuracy, which increased by 2% the accuracy achieved by the original CoroNet [13]. This means that by improving the data augmentation strategy, 20000 more patients can be correctly diagnosed for every 1 million. CoroNet without DA reached a 79% accuracy. As a result, an adequate data augmentation strategy can correctly diagnose 180000 more patients for every million patients. These findings also show that, when used correctly, data augmentation can be an effective technique.

On database 2, the CoroNet architecture with the optimal configuration of data augmentation achieved a 97% accuracy. Compared with the COVID-Classifier architecture proposed by Zargari Khuzani et al. [20] and the COVID-19 ResNet proposed by Baldeon Calisto et al. [25], our network has a 3% increase in performance. With the improvement, 30000 more patients can be correctly diagnosed for every 1 million. The authors of Baldeon Calisto et al. [25] use a Bayesian hyperparameter optimization approach to determine the best data augmentation values. We improved that performance using a fractional factorial design, indicating that the proposed methodology can compete with well-established hyperparameter optimization approaches. Similarly, compared to CoroNet with the original data augmentation strategy, the proposed optimized CoroNet outperforms it in all evaluation metrics, demonstrating that data augmentation can reduce performance when used incorrectly.

5- Concluding Remarks

5-1- Discussion

Due to the rapid spread of the COVID-19 pandemic, a lack of nasopharyngeal testing materials and the limited availability of medical practitioners, accurate automated COVID-19 diagnosis methods are required. Chest images, such as X-rays, can be analyzed to provide a quick diagnosis. However, identifying pulmonary illnesses caused by a COVID-19 infection necessitates the use of skilled radiologists. Detecting COVID-19 automatically using CNN models is a promising direction. In this paper, we present a methodology for statistically identifying the effects of data augmentation operations on COVID-19 detection accuracy via chest X-ray analysis. The methodology is used in two publicly available COVID-19 datasets. Vertical flip, zoom, height shift, and shear positively affect accuracy in dataset 1. Horizontal flip has a negative effect, while rotation and width shift have no effect. These results indicate that the dataset contained substantial variation regarding image size and vertical placement. As a result, zooming in and out of the region of interest and vertically shifting it improved the model's generalization. None of the data augmentation operations in dataset 2 statistically affect the accuracy. This means that by introducing data augmentation, the CNN learns nothing new from the original dataset and thus should not be used during training.

In addition to the latter analysis, the implemented fractional factorial experimental design enables determining the optimal data augmentation strategy to maximize classification accuracy. On dataset 1, we outperformed Khan et al. [13] by 2% in average accuracy by training CoroNet with these optimal strategies. In dataset 2, we improved the average accuracy of the results obtained by Zargari Khuzani et al. [20] and Baldeon Calisto et al. [25] by 3%. We emphasize that the methodology and analysis presented here can be applied to other network structures or applications to help identify each augmentation operation's effect on the established evaluation metric and help define the optimal data augmentation policy to improve performance.

Although many studies have reported the success of data augmentation in classification or segmentation problems [22], little research has been conducted to statistically determine the positive or negative effects of data augmentation operations on COVID-19 detection performance. Elgendi et al. [2] investigated the impact of four data augmentation strategies for COVID-19 classification and proposed clinically based general guidelines for its use. Despite this, our findings are inconsistent with their conclusions. Vertical reflections and shearing, according to Elgendi et al. [2], should be avoided because they produce non-physiologic images in practice. Nevertheless, as demonstrated in Tables 10 and 9, our results indicate that vertical flip and shear have a statistically positive effect on the classification accuracy of dataset 1 and therefore produce better results for all evaluation metrics. According to Chlap et al. [22], many basic augmentation techniques do not aim to produce realistic images but encourage the model to learn more general features. The rotation and width shift operations do not appear to have a statistically significant influence on classification accuracy on either dataset and should not be included in the data augmentation strategy. This conclusion contradicts the guidelines of Elgendi et al. [2], which label rotation and translation as beneficial operations.

Several studies have shown that data augmentation is especially useful when the dataset is small [39, 40]. However, dataset 2, which only contained 381 images, did not benefit from data augmentation during training. As shown in Table 11, not using data augmentation improves the original CoroNet's performance (which implements a data augmentation strategy during training) and even outperforms [25], which uses a Bayesian hyperparameter optimization approach to set the data augmentation values. This finding was also found in skin lesion analysis [17], where data augmentation harmed the results for datasets with fewer than 500 images. As previously stated, this could be due to the original dataset's lack of diversity and information. Furthermore, it emphasizes the importance of considering the size of the dataset before implementing data augmentation.

Data augmentation strategies are frequently transferred between datasets without thoroughly evaluating their utility. For example, the data augmentation strategy developed by Krizhevsky et al. [41] in 2012 for the ImageNet dataset remains the standard for image classification [42]. Our findings, however, show that even when the same task (COVID-19 classification) and network structure are used, the optimal data augmentation policy can differ across datasets. Differences in image acquisition protocols, imaging equipment, and pixel value distribution can contribute to this behavior. As a result, developing methods that identify the optimal augmentation policy on a specific dataset and analyze the impact of each operation, as presented in our work, is critical. Moreover, due to the varying effects an augmentation operation may have on two distinct datasets, it is not prudent to recommend its use on all datasets.

Based on previous research, we chose a limited combination of basic data augmentation operations and levels, which is a limitation of the current investigation. Future research could investigate the effects of a broader range of augmentation operations (deformable and deep learning techniques) and levels on various applications and datasets.

5-2- Conclusion

A $2^{(7-1)}$ fractional factorial experimental design is used in this study to statistically examine the effect of basic data augmentation techniques on detecting COVID-19 on chest X-Ray images. This method also allows optimizing the optimal combination of data augmentation factors to maximize classification accuracy. The CoroNet architecture was selected for its performance and efficiency, and it was tested on two publicly available COVID-19 datasets. Vertical flip, zoom, shear, and height shift operations improve accuracy in dataset 1, while horizontal flip has the opposite effect. The dataset mentioned achieves a new state-of-the-art performance with the best combination of data augmentation parameters, achieving an accuracy of 97%. The experimental model in dataset 2 was not statistically significant. As a result, this dataset's best data augmentation parameter combination was to avoid using data augmentation operations. A 97% accuracy rate is achieved, outperforming all competing models by avoiding data augmentation during training. The experiments show that optimizing the values of the data augmentation operations for each dataset is essential. Finally, the findings presented show significant progress in incorrectly identifying COVID-19 from chest X-ray images, which will aid in a swift and accurate diagnosis.

6- Declarations

6-1- Author Contributions

Conceptualization, M.B.C.; methodology, M.H.D., M.B.C., J.J.M., and D.N.; software, M.H.D. and J.J.M.; validation, M.B.C., D.R., and D.N.; formal analysis, M.H.D. and J.J.M.; investigation, M.H.D., M.B.C., and J.J.M.; resources, D.S.B., D.R, N.P., and R.F.M.; data curation, M.H.D. and J.J.M.; writing—original draft preparation, M.H.D., M.B.C., J.J.M., and B.P.M.; writing—review and editing, M.B.C., D.N., D.S.B., D.R., N.P., and R.F.M.; visualization, M.H.D., J.J.M., and B.P.M.; supervision, M.B.C., D.N., and D.R.; project administration, B.P.M. All authors have read and agreed to the published version of the manuscript.

6-2- Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6-3- Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6-4- Acknowledgements

The authors thank the Applied Signal Processing and Machine Learning Research Group of USFQ for providing the computing infrastructure (NVidia DGX workstation) to implement and execute the developed source code. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

6-5- Institutional Review Board Statement

Not applicable.

6-6- Informed Consent Statement

Not applicable.

6-7- Conflicts of Interest

The authors declare that there is no conflict of interests regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

7- References

[1] Narin, A., Kaya, C., & Pamuk, Z. (2021). Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks. Pattern Analysis and Applications, 24(3), 1207–1220. doi:10.1007/s10044-021-00984-y.

[2] Elgendi, M., Nasir, M. U., Tang, Q., Smith, D., Grenier, J.-P., Batte, C., Spieler, B., Leslie, W. D., Menon, C., Fletcher, R. R., Howard, N., Ward, R., Parker, W., & Nicolaou, S. (2021). The Effectiveness of Image Augmentation in Deep Learning Networks for Detecting COVID-19: A Geometric Transformation Perspective. Frontiers in Medicine, 8. doi:10.3389/fmed.2021.629134.

[3] Ji, T., Liu, Z., Wang, G. Q., Guo, X., Akbar khan, S., Lai, C., Chen, H., Huang, S., Xia, S., Chen, B., Jia, H., Chen, Y., & Zhou, Q. (2020). Detection of COVID-19: A review of the current literature and future perspectives. Biosensors and Bioelectronics, 166, 112455. doi:10.1016/j.bios.2020.112455.

[4] Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a convolutional neural network. 2017 International Conference on Engineering and Technology (ICET). doi:10.1109/icengtechnol.2017.8308186.

[5] Baldeon-Calisto, M., & Lai-Yuen, S. K. (2020). AdaResU-Net: Multiobjective adaptive convolutional neural network for medical image segmentation. Neurocomputing, 392, 325–340. doi:10.1016/j.neucom.2019.01.110.

[6] Baldeon Calisto, M., & Lai-Yuen, S. K. (2020). AdaEn-Net: An ensemble of adaptive 2D–3D Fully Convolutional Networks for medical image segmentation. Neural Networks, 126, 76–94. doi:10.1016/j.neunet.2020.03.007.

[7] Baldeon Calisto, M., & Lai-Yuen, S. K. (2021). EMONAS-Net: Efficient multiobjective neural architecture search using surrogate-assisted evolutionary algorithm for 3D medical image segmentation. Artificial Intelligence in Medicine, 119, 102154. doi:10.1016/j.artmed.2021.102154.

[8] Wang, L., Lin, Z. Q., & Wong, A. (2020). COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. Scientific Reports, 10(1), 19549. doi:10.1038/s41598-020-76550-z.

[9] Monshi, M. M. A., Poon, J., Chung, V., & Monshi, F. M. (2021). CovidXrayNet: Optimizing data augmentation and CNN hyperparameters for improved COVID-19 detection from CXR. Computers in Biology and Medicine, 133, 104375. doi:10.1016/j.compbiomed.2021.104375.

[10] Algarni, A. D., El-Shafai, W., El Banby, G. M., Abd El-Samie, F. E., & Soliman, N. F. (2022). An efficient CNN-based hybrid classification and segmentation approach for COVID-19 detection. Computers, Materials and Continua, 70(3), 4393–4410. doi:10.32604/cmc.2022.020265.

[11] Cohen, J. P., Morrison, P., Dao, L., Roth, K., Duong, T. Q., & Ghassemi, M. (2020). Covid-19 image data collection: Prospective predictions are the future. arXiv preprint arXiv:2006.11988. doi:10.48550/arXiv.2006.11988.

[12] Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr.2017.195.

[13] Khan, A. I., Shah, J. L., & Bhat, M. M. (2020). CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. Computer Methods and Programs in Biomedicine, 196, 105581. doi:10.1016/j.cmpb.2020.105581.

[14] Taylor, L., & Nitschke, G. (2018). Improving Deep Learning with Generic Data Augmentation. 2018 IEEE Symposium Series on Computational Intelligence (SSCI). doi:10.1109/ssci.2018.8628742.

[15] Mikolajczyk, A., & Grochowski, M. (2018). Data augmentation for improving deep learning in image classification problem. 2018 International Interdisciplinary PhD Workshop (IIPhDW). doi:10.1109/iiphdw.2018.8388338.

[16] Shijie, J., Ping, W., Peiyi, J., & Siping, H. (2017). Research on data augmentation for image classification based on convolution neural networks. 2017 Chinese Automation Congress (CAC). doi:10.1109/cac.2017.8243510.

[17] Perez, F., Vasconcelos, C., Avila, S., Valle, E. (2018). Data Augmentation for Skin Lesion Analysis. In: , et al. OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis. CARE CLIP OR 2.0 ISIC 2018. Lecture Notes in Computer Science, 11041. Springer, Cham, Switzerland. doi:10.1007/978-3-030-01201-4_33.

[18] Safdar, M., Kobaisi, S., & Zahra, F. (2020). A Comparative Analysis of Data Augmentation Approaches for Magnetic Resonance Imaging (MRI) Scan Images of Brain Tumor. Acta Informatica Medica, 28(1), 29. doi:10.5455/aim.2020.28.29-36.

[19] Omigbodun, A. O., Noo, F., McNitt-Gray, M., Hsu, W., & Hsieh, S. S. (2019). The effects of physics-based data augmentation on the generalizability of deep neural networks: Demonstration on nodule false-positive reduction. Medical Physics, 46(10), 4563–4574. doi:10.1002/mp.13755.

[20] Zargari Khuzani, A., Heidari, M., & Shariati, S. A. (2021). COVID-Classifier: an automated machine learning model to assist in the diagnosis of COVID-19 infection in chest X-ray images. Scientific Reports, 11(1), 9887. doi:10.1038/s41598-021-88807-2.

[21] Kermany, D., Zhang, K., & Goldbaum, M. (2018). Labeled optical coherence tomography (oct) and chest x-ray images for classification. Mendeley data, 2(2). doi:10.17632/RSCBJBR9SJ.2.

[22] Chlap, P., Min, H., Vandenberg, N., Dowling, J., Holloway, L., & Haworth, A. (2021). A review of medical image data augmentation techniques for deep learning applications. Journal of Medical Imaging and Radiation Oncology, 65(5), 545–563. doi:10.1111/1754-9485.13261.

[23] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. Journal of Big Data, 6(1), 60. doi:10.1186/s40537-019-0197-0.

[24] Abbas, A., Abdelsamea, M. M., & Gaber, M. M. (2021). Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network. Applied Intelligence, 51(2), 854–864. doi:10.1007/s10489-020-01829-7.

[25] Baldeon calisto, M., Balseca Zurita, J. S., & Cruz Patiño, M. A. (2021). COVID-19 ResNet: Residual neural network for COVID-19 classification with bayesian data augmentation. ACI Avances En Ciencias e Ingenierías, 13(2), 19. doi:10.18272/aci.v13i2.2288.

[26] Chowdhury, N. K., Rahman, Md. M., & Kabir, M. A. (2020). PDCOVIDNet: a parallel-dilated convolutional neural network architecture for detecting COVID-19 from chest X-ray images. Health Information Science and Systems, 8(1). doi:10.1007/s13755-020-00119-3.

[27] Goel, T., Murugan, R., Mirjalili, S., & Chakrabartty, D. K. (2021). OptCoNet: an optimized convolutional neural network for an automatic diagnosis of COVID-19. Applied Intelligence, 51(3), 1351–1366. doi:10.1007/s10489-020-01904-z.

[28] Kumar, A., Tripathi, A. R., Satapathy, S. C., & Zhang, Y. D. (2022). SARS-Net: COVID-19 detection from chest x-rays by combining graph convolutional network and convolutional neural network. Pattern Recognition, 122, 108255. doi:10.1016/j.patcog.2021.108255.

[29] Marques, G., Agarwal, D., & de la Torre Díez, I. (2020). Automated medical diagnosis of COVID-19 through EfficientNet convolutional neural network. Applied Soft Computing Journal, 96, 106691. doi:10.1016/j.asoc.2020.106691.

[30] Nishio, M., Noguchi, S., Matsuo, H., & Murakami, T. (2020). Automatic classification between COVID-19 pneumonia, non-COVID-19 pneumonia, and the healthy on chest X-ray image: combination of data augmentation methods. Scientific Reports, 10(1). doi:10.1038/s41598-020-74539-2.

[31] Rahimzadeh, M., & Attar, A. (2020). A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2. Informatics in Medicine Unlocked, 19(100360). doi:10.1016/j.imu.2020.100360.

[32] Yoo, S. H., Geng, H., Chiu, T. L., Yu, S. K., Cho, D. C., Heo, J., Choi, M. S., Choi, I. H., Cung Van, C., Nhung, N. V., Min, B. J., & Lee, H. (2020). Deep Learning-Based Decision-Tree Classifier for COVID-19 Diagnosis From Chest X-ray Imaging. Frontiers in Medicine, 7. doi:10.3389/fmed.2020.00427.

[33] Montgomery, D. (2019). Design and Analysis of Experiments (10th Ed.). Wiley, Hoboken, United States.

[34] Lujan-Moreno, G. A., Howard, P. R., Rojas, O. G., & Montgomery, D. C. (2018). Design of experiments and response surface methodology to tune machine learning hyperparameters, with a random forest case-study. Expert Systems with Applications, 109, 195–205. doi:10.1016/j.eswa.2018.05.024.

[35] Staelin, C. (2003). Parameter selection for support vector machines. Hewlett-Packard Company, Tech. Rep. HPL-2002-354R1, 1. HP Laboratories, Haifa, Israel.

[36] Chou, F. I., Tsai, Y. K., Chen, Y. M., Tsai, J. T., & Kuo, C. C. (2019). Optimizing Parameters of Multi-Layer Convolutional Neural Network by Modeling and Optimization Method. IEEE Access, 7, 68316–68330. doi:10.1109/ACCESS.2019.2918563.

[37] Fang, K. T., & Lin, D. K. J. (2003). Ch. 4. Uniform experimental designs and their applications in industry. Handbook of Statistics, 22, 131–170, Elsevier, Amsterdam, Netherlands. doi:10.1016/S0169-7161(03)22006-X.

[38] Ahrens, W. H., Cox, D. J., & Budhwar, G. (1990). Use of the Arcsine and Square Root Transformations for Subjectively Determined Percentage Data. Weed Science, 38(4–5), 452–458. doi:10.1017/s0043174500056824.

[39] Wodzinski, M., Banzato, T., Atzori, M., Andrearczyk, V., Cid, Y. D., & Muller, H. (2020). Training Deep Neural Networks for Small and Highly Heterogeneous MRI Datasets for Cancer Grading. 2020 42nd Annual International Conference of the IEEE Engineering in Medicine; Biology Society (EMBC). doi:10.1109/embc44109.2020.9175634.

[40] Ogawa, R., Kido, T., & Mochizuki, T. (2019). Effect of augmented datasets on deep convolutional neural networks applied to chest radiographs. Clinical Radiology, 74(9), 697–701. doi:10.1016/j.crad.2019.04.025.

[41] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84–90. doi:10.1145/3065386.

[42] Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V., & Le, Q. V. (2019). AutoAugment: Learning Augmentation Strategies From Data. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). bdoi:10.1109/cvpr.2019.00020.