



# Article Enhancing an Imbalanced Lung Disease X-ray Image Classification with the CNN-LSTM Model

Julio Fachrel <sup>1</sup><sup>(b)</sup>, Anindya Apriliyanti Pravitasari <sup>1,\*</sup><sup>(b)</sup>, Intan Nurma Yulita <sup>2</sup><sup>(b)</sup>, Mulya Nurmansyah Ardhisasmita <sup>3</sup><sup>(b)</sup> and Fajar Indrayatna <sup>1</sup><sup>(b)</sup>

- <sup>1</sup> Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Bandung 45363, Indonesia; julio19001@mail.unpad.ac.id (J.F.); fajar.indrayatna@unpad.ac.id (FI.)
- <sup>2</sup> Department of Computer Science, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Bandung 45363, Indonesia; intan.nurma@unpad.ac.id
- <sup>3</sup> Department of Public Health, Faculty of Medicine, Universitas Padjadjaran, Bandung 45363, Indonesia; mulya@unpad.ac.id
- \* Correspondence: anindya.apriliyanti@unpad.ac.id

Abstract: Lung diseases have a significant impact on respiratory health, causing various symptoms and posing challenges in diagnosis and treatment. This research presents a methodology for classifying lung diseases using chest X-ray images, specifically focusing on COVID-19, pneumonia, and normal cases. The study introduces an optimal architecture for convolutional neural network (CNN) and long short-term memory (LSTM) models, considering evaluation metrics and training efficiency. Furthermore, the issue of imbalanced datasets is addressed through the application of some image augmentation techniques to enhance model performance. The most effective model comprises five convolutional blocks, two LSTM layers, and no augmentation, achieving an impressive F1 score of 0.9887 with a training duration of 91 s per epoch. Misclassifications primarily occurred in normal cases, accounting for only 3.05% of COVID-19 data. The pneumonia class demonstrated excellent precision, while the normal class exhibited high recall and an F1 score. Comparatively, the CNN-LSTM model outperformed the CNN model in accurately classifying chest X-ray images and identifying infected lungs. This research provides valuable insights for improving lung disease diagnosis, enabling timely and accurate identification of lung diseases, and ultimately enhancing patients' outcomes.

**Keywords:** lung diseases; COVID-19; pneumonia; X-ray; convolutional neural networks; long short-term memory

## 1. Introduction

Lung diseases encompass a wide range of disorders that are prevalent and linked to significant morbidity and mortality [1]. These disorders can significantly impact the respiratory system, including the lungs, airways, and pulmonary blood vessels. As critical organs responsible for the respiration process, the lungs play an essential role in providing oxygen to the body and eliminating carbon dioxide. However, lung diseases can interrupt this vital process, causing uncomfortable symptoms, such as fatigue, shortness of breath, wheezing, coughing, and chest pain. Diagnosing and treating lung diseases can be challenging due to the wide range of possible causes and symptoms [2]. Nevertheless, it is critical to identify and treat them early to enhance patient outcomes and lessen the burden of these conditions on society. With ongoing research and advances in technology, there is optimism for improving our understanding and management of various lung diseases.

One of the most widely discussed topics in the field of lung diseases today is COVID-19. The global impact of COVID-19 on lung health has been significant, resulting in a large number of hospitalizations and fatalities worldwide. COVID-19 has served as a painful reminder of the importance of respiratory health and the need for continued study,



Citation: Fachrel, J.; Pravitasari, A.A.; Yulita, I.N.; Ardhisasmita, M.N.; Indrayatna, F. Enhancing an Imbalanced Lung Disease X-ray Image Classification with the CNN-LSTM Model. *Appl. Sci.* 2023, 13, 8227. https://doi.org/10.3390/ app13148227

Academic Editor: Jan Egger

Received: 2 June 2023 Revised: 9 July 2023 Accepted: 13 July 2023 Published: 15 July 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). preparation, and cooperation to successfully handle lung diseases. The global response to the COVID-19 pandemic has stressed the interconnectedness between lung health and public health, pressing on the importance of ongoing study and joint efforts in the field of lung diseases. It is worth mentioning that many people who died from COVID-19 had severe chest congestion and a consequent large decrease in oxygen levels, which raised the risk of massive heart attacks [3]. On another note, pneumonia is also a type of lung disease characterized by inflammation in the small air sacs within the lungs. It can be caused by different pathogens, including bacteria, viruses, or fungi [4]. Interestingly, the signs and symptoms of pneumonia have similarities with COVID-19 [5]. Given this similarity and the fact that different diseases require different treatments [6], it becomes crucial to accurately identify specific diseases. This ensures that appropriate and distinct treatment approaches can be employed based on the specific lung disease. Because of that, this research aims to identify the classes as COVID-19, pneumonia, and normal.

Radiological images of the lungs provide an alternative approach to diagnosing lung infections. Clinical diagnostic tools, such as X-rays and computed tomography (CT), can effectively assess and describe the condition of the lungs. Although CT scans offer better detection sensitivity, X-ray radiography is more commonly utilized in clinical settings due to its advantages and conveniences, including lower cost and widespread availability in general hospitals [7]. Thus, X-ray radiography is preferred in many cases, serving as a practical and efficient method for diagnosing lung infections.

Detecting and classifying lung diseases using chest X-ray images is a challenging and complex task. To aid radiologists and accelerate the identification process, researchers have developed deep-learning models [8–10]. These models utilize advanced machine learning algorithms and neural networks to analyze and categorize X-ray images based on distinct patterns and features associated with various lung diseases. Furthermore, the emergence of COVID-19 has significantly increased the patient load, placing additional demands on radiologists to accurately identify and diagnose cases. This surge in cases has necessitated more time and energy from radiologists, highlighting the need for efficient and effective diagnostic tools. Deep learning is hoped to be a valuable tool in this context, enabling radiologists to handle the growing workload.

The goal of this study is to create an image classification model that uses deep-learning techniques to speed up the identification of lung diseases, thereby reducing the effort and time involved in the diagnostic process. This model will aid in effectively identifying patients exposed to COVID-19 and those with pneumonia. The major objective is to reduce mistakes and misdiagnoses in lung disease treatment, thereby improving patient care and outcomes. By enhancing the accuracy and efficiency of disease detection, this study endeavors to contribute to the overall optimization of lung disease management and reduce the potential for mishandling of such conditions.

The main contributions of this research are two-fold:

- 1. Implementation and selection of the optimal architecture of convolutional neural network (CNN) and long short-term memory (LSTM) deep-learning models for the classification of lung diseases using chest X-ray images. The selection process is based on evaluation metrics and training time, ensuring that the models are efficient and effective in accurately identifying different lung diseases;
- 2. Addressing the challenge of an imbalanced dataset by applying various image augmentation techniques. Imbalanced datasets, where certain classes have significantly fewer samples than others, can pose challenges in model training. By employing appropriate image augmentation methods, this research aims to improve the performance of the deep-learning models by artificially expanding the dataset and creating a more balanced representation of different lung diseases.

The remainder of this paper is structured as follows. Section 2 provides a comprehensive review of the existing scientific literature related to this research. Section 3 outlines the methodology employed in this research, including the collection and splitting of the dataset. Section 4 presents the results obtained from the experiments conducted in this study. Section 5 offers a comprehensive discussion of the findings, conclusions, and recommendations for future research.

#### 2. The Literature Review

Tekerek and Al-Rawe [11] introduced a classification method based on deep learning to identify lung diseases from chest X-ray images, with a specific emphasis on detecting COVID-19. This approach aims to categorize chest X-ray images into three groups: normal; COVID-19; and viral pneumonia. It employs an eight-layer convolutional neural network that combines MobileNet [12] and DenseNet [13] models. The research findings indicate a precision value of 1.00 for COVID-19 and normal cases while achieving a precision of 0.79 for viral pneumonia. The recall values are 1.00 for normal and viral pneumonia and 0.69 for COVID-19. The F1 score is found to be 1.00 for normal, 0.79 for COVID-19, and 0.85 for viral pneumonia. The proposed method achieves an impressive accuracy of 96% and a ROC AUC score of 0.94. These outcomes showcase the remarkable accuracy of the proposed approach in diagnosing and classifying chest X-ray images, surpassing the performance of traditional CNN and MobileNet methods. The method's high precision and F1 score are particularly important for minimizing false negatives, thereby aiding in the prevention of disease transmission.

Gupta et al. [14] presented a method that uses deep-learning models, pre-processing techniques, and lung segmentation to improve the precision of COVID-19 detection in chest X-ray images. The study uses InceptionV3 [15] and U-Net [16], which are deep-learning models, to process and identify chest X-ray images as either COVID-19-negative or positive. By adding lung segmentation during pre-processing, this study aims to remove irrelevant surrounding information that could introduce bias and create inaccurate results. The results of this study show an amazing accuracy rate of approximately 99% for the most effective models in spotting COVID-19. However, this study also shows the effect of visual noise on model bias and underscores the value of lung segmentation in reducing bias and ensuring more consistent results. This study admits that the current models strongly rely on visible abnormalities in the lungs as signs of COVID-19, and further improvements are necessary to address this weakness.

Badrahadipura et al. [17] conducted a study utilizing the Inception ResNet-v2 [18] architecture and transfer learning to classify chest X-ray images into three categories: normal; viral pneumonia; and COVID-19. The dataset consisted of 3616 COVID-19 cases, 10,192 of normal cases, and 1345 cases of viral pneumonia. The model underwent two rounds of training. Initially, the Inception ResNet-v2 layers were frozen, preserving the weights and biases learned from the ImageNet dataset. Only the additional layers were added after the Inception ResNet-v2 was trained. In the second training phase, all layers were unfrozen, allowing for further fine-tuning of the entire model. This research highlighted that the model performed better in classifying images belonging to viral pneumonia and normal classes compared to the COVID-19 class, as indicated by higher precision, recall, and F1 scores. The overall accuracy of the model was reported to be 0.966, with an F1 score of 0.97. These findings demonstrate the potential of using the Inception ResNet-v2 architecture and transfer learning for accurate classification of chest X-ray images, particularly in distinguishing between viral pneumonia, normal, and COVID-19 cases, contributing to advancements in medical imaging and healthcare applications.

Abbas et al. [19] conducted a study to investigate the application of transfer learning using the DeTraC (Decompose, Transfer, and Compose) deep CNN architecture for COVID-19 chest X-ray classification. DeTraC incorporates a class decomposition mechanism to address irregularities presented in the image dataset. This study demonstrates the effectiveness of DeTraC in accurately classifying COVID-19 cases while also showcasing its robustness in handling data irregularities and the limited availability of training images. Through validation experiments with various pre-trained CNN models, VGG19 [20] emerged as the most successful model within the DeTraC framework. The experimental results highlight the impressive performance of DeTraC in detecting COVID-19 cases, achieving an accuracy of 93.1% with a sensitivity of 100% in accurately distinguishing COVID-19 X-ray images from both normal and severe acute respiratory syndrome cases.

Goyal and Singh [5] proposed a framework for detecting COVID-19 and pneumonia in chest X-ray images. The framework is divided into multiple steps, which include dataset gathering, picture quality improvement, ROI estimation, feature extraction, and illness classification. Two publicly accessible chest X-ray datasets are used, and picture quality is improved by utilizing such techniques as median filtering and histogram equalization. Various characteristics, such as visual, shape, texture, and intensity, are retrieved and normalized from each ROI picture. Soft computing approaches, such as ANN [21], SVM [22], KNN [23], ensemble classifiers [24], and a deep-learning classifier dubbed F-RNN-LSTM, are used for classification. The F-RNN-LSTM deep-learning architecture combines RNN and LSTM for enhanced disease categorization. Experiment findings show that the suggested framework is successful. When compared to previous approaches, the F-RNN-LSTM model achieves an accuracy of roughly 95% while requiring less computing effort.

Demir [25] introduced an innovative method for detecting COVID-19 from X-ray images by utilizing a deep LSTM model. The model is developed from scratch, offering a unique architecture specifically designed for this purpose. To enhance the model's performance, the study incorporates such pre-processing techniques as the Sobel gradient and marker-controlled watershed segmentation. This research conducts experiments on a combined public dataset consisting of 361 COVID-19, 500 pneumonia, and 200 normal chest X-ray images. The dataset is divided randomly into training and testing sets, with different ratios tested. The most favorable results are obtained when using an 80% training and 20% testing split. Impressively, the proposed model achieves a perfect 100% success rate across all performance metrics, including accuracy, sensitivity, specificity, and F-score. These findings are particularly remarkable considering the small size of the dataset used in the study.

Pustokhin et al. [26] introduced the RCAL-BiLSTM model, which combines ResNet [27], a class attention layer (CAL) [28], and a Bi-LSTM. The model comprises several stages, including preprocessing, using bilateral filtering [29], feature extraction, using RCAL-BiLSTM, and classification employing SoftMax. Feature extraction involves ResNet for extracting features, CAL for capturing discriminative class-based features, and Bi-LSTM for modeling class dependencies in both directions. The SoftMax layer is then used to classify the feature vectors into their respective feature maps. Experimental validation is performed on a dataset of chest X-ray images, and the results illustrate the superior performance of the RCAL-BiLSTM model. It achieves high sensitivity (93.28%), specificity (94.61%), precision (94.90%), accuracy (94.88%), F-score (93.10%), and kappa value (91.40%), highlighting the effectiveness of the proposed model for COVID-19 diagnosis.

Hamza et al. [30] proposed a CNN-LSTM architecture combined with an improved optimization algorithm to address the challenges of multisource fusion and redundant features. The dataset consisted of four classes: COVID-19; normal, viral pneumonia, and lung opacity. The framework includes contrast enhancement and data augmentation to improve the quality and quantity of training samples. Deep transfer learning is utilized in training a CNN-LSTM model and fine-tuning an EfficientNet [31] model for feature extraction. The overall accuracy achieved was 98.5%.

Fachrel et al. [32] compared two deep-learning models, namely, convolutional neural networks (CNN) and a combination of CNN and long short-term memory (LSTM). The dataset consists of 4095 CXR images (1400 of normal conditions, 1350 of COVID-19, and 1345 of pneumonia). Both CNN and CNN-LSTM models are evaluated using a confusion matrix and compared in terms of performance. The experimental results demonstrate that the CNN-LSTM model outperforms the CNN model, achieving an overall accuracy of approximately 98.78%. It also exhibits high precision and recall, reaching 99% and 98%, respectively. These findings suggest that the proposed CNN-LSTM model can contribute to fast and accurate COVID-19 detection.

The previous studies primarily focused on the development of deep-learning algorithms and certain preprocessing methods to classify lung diseases. Furthermore, the utilization of LSTM networks has been recognized as an effective approach to achieving higher performance scores [5,25,26,30,32]. To the best of our knowledge, the problem of imbalanced datasets has been given limited consideration in these studies. Therefore, our research aims to address this gap by focusing on selecting the optimal architecture for the CNN-LSTM model and tackling the challenges associated with imbalanced datasets. We plan to employ various image augmentation techniques to improve the model's performance and enhance its ability to handle imbalanced data.

## 3. Materials and Methods

In this section, we will provide a description of the dataset utilized, the models employed, the experimental setup, and the evaluation metrics used to assess the performance of the models.

## 3.1. Dataset

This study used a dataset from Kaggle [33,34] that contained three classes (COVID-19, normal, and pneumonia) shown in Figure 1. A total of 15,153 images (10,912 normal; 3616 COVID-19; and 1345 pneumonia) were used in this study, with 90% used for training and 10% used for validation. The proportion of splitting data can be seen in Table 1.



Figure 1. Images of normal, COVID-19, and pneumonia.

	Normal	COVID-19	Pneumonia	Total
Training	9173	3255	1211	13,639
Validation	1019	361	134	1514
Total	10,912	3616	1345	15,153

Table 1. Separating training and test data for each class.

#### 3.2. Convolutional and Recurrent Networks: CNN and LSTM

The convolutional neural network (CNN) is a widely used algorithm for processing image data. It employs mathematical operations to learn from a network called a "convolution" [35]. The convolution layer, which is a crucial component of CNN, plays a vital role in feature extraction by utilizing local connections and shared characteristic weights [36]. It consists of linear and non-linear operations that work together to extract meaningful patterns from the input data [37]. The convolution layer's ability to extract important features is a key reason why CNN is well-suited for image data. Multiple filters within the convolution layer learn distinct features through a variety of weights [38]. These filters can be represented mathematically as  $n \times n$  matrices, where each element corresponds to a weight. The filtering process is illustrated in Figure 2, where each pair of elements in the filter is multiplied and added to generate a single output value. This process is repeated for each filter, resulting in a "feature map" that highlights specific features in the input data. By employing multiple filters with different weights, the convolution layer effectively learns a diverse set of features, enabling it to identify and differentiate various objects or patterns in the input data.



Figure 2. Illustration of the filtering process on the convolution layer (modified from [39]).

LSTM, a type of recurrent neural network (RNN), can effectively retain information over long periods and learn from inputs that are widely separated in time [40]. Unlike traditional RNNs, LSTM overcomes the problem of vanishing gradients and captures longterm dependencies by employing memory cells with specialized gating mechanisms [41]. Figure 3 illustrates the architecture of LSTM, where each memory cell consists of three types of gates (forget, input, and output) that control the information flow. The forget gate filters out small values, the input gate determines new information to be added to the memory, and the output gate determines whether to output the stored value. Additionally, each memory cell incorporates three sigmoid activation functions and one tanh activation function [42]. The tanh function maintains values within the range from -1 to 1, while the C(t-1) C(t) tanh forget gate: input gate: h(t) X output  $f_t$ i, ĉ. gate: Ot tanh Ø a 0 h(t-1) X(t)

sigmoid function transforms the range from -1 to 1 to 0 to 1, preventing zero values from entering the memory cells. For a more detailed explanation of LSTM, refer to [43].

Figure 3. The architecture of the long short-term memory (LSTM) (modified from [44]).

## 3.3. CNN-LSTM Architecture

In the CNN-LSTM architecture, CNN is used to extract important features from an image, while LSTM replaces the fully connected role, which is to classify based on the features that have been extracted by CNN. By combining CNN and LSTM networks in this way, the CNN-LSTM architecture can handle both spatial and temporal information in an image, making it useful for classification. Figure 4 shows the architecture of the CNN-LSTM, where the LSTM layer is placed after the convolution layer and receives the output value from the last convolution layer for further classification.



Figure 4. The architecture of CNN-LSTM.

In a fully connected layer, the connections between nodes in different layers are not specific to any particular sequence of inputs, and each node can only process one input at a time. On the other hand, the nodes in an LSTM layer are connected along the sequence of inputs, allowing the network to capture dependencies between elements in the sequence [45]. Additionally, LSTM employs a gating mechanism that enables it to selectively recall or forget information from previous time steps, allowing it to simulate long-term dependencies in the input sequence effectively. Furthermore, because of its ability to manage the problem of vanishing gradients (gradients that become very small during backpropagation), LSTM is more easily optimized than typical fully-connected networks [46].

## 3.4. Evaluation

## 3.4.1. Confusion Matrix

Confusion matrix is a method for evaluating the performance of a classification model in making predictions [47]. The confusion matrix generates scores for accuracy, precision, recall, and F1 score. This score is used to assess the model's performance. Predicted and actual classifications are shown in a  $n \times n$  confusion matrix, where n is the number of classes [48]. There were three classes employed in this study: normal, COVID-19, and pneumonia. The form of the confusion matrix for the three classes can be seen in Table 2.

Table 2. Confusion matrix for three classes.

Actual		Predict	
Actual	Α	В	С
А	T <sub>AA</sub>	F <sub>BA</sub>	F <sub>CA</sub>
В	F <sub>AB</sub>	T <sub>BB</sub>	F <sub>CB</sub>
С	F <sub>AC</sub>	F <sub>BC</sub>	T <sub>CC</sub>

In the three-class confusion matrix, there is a slight difference in calculating true positive, true negative, false positive, and false negative. In the context of class A in Table 2, the calculations for those values are as follows:

$$True \ Positive = T_{AA} \tag{1}$$

$$True \ Negative = T_{BB} + T_{CC} + F_{BC} + F_{CB}$$
(2)

$$False \ Positive = F_{AB} + F_{AC} \tag{3}$$

$$False \ Negative = F_{BA} + F_{CA} \tag{4}$$

From the confusion matrix, we can calculate the evaluation metric as follows:

$$precision = \frac{TP}{TP + FP}$$
(5)

$$\operatorname{recall} = \frac{TP}{TP + FN} \tag{6}$$

F1 score = 
$$2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$
 (7)

In this case, we did not use accuracy because, when dealing with imbalanced data, where one class has significantly fewer instances than the others, accuracy can be misleading. In such cases, the F1 score is a more suitable evaluation metric as it takes into account both precision and recall. The F1 score is particularly useful in cases of unbalanced data. This is because the F1 score considers both false positives and false negatives, which are crucial in such cases to identify the model's performance in correctly predicting the minority class [49].

#### 3.4.2. K-Fold Cross-Validation

K-fold cross-validation means performing data splits repeatedly k times. In this method, the data is first partitioned into k segments. Then k training iterations were carried out, and validation was carried out in such a way that in each iteration, there were k different test data conditions [50]. An illustration of the k-fold cross-validation process can be seen in Figure 5.



Figure 5. Illustration of K-fold cross-validation.

#### 3.5. Experimental Setup

As shown in Figure 6, this study consists of several steps, including pre-processing, dataset splitting, selecting an appropriate architecture, and evaluating the best architecture. To ensure accurate and reliable predictions from a machine learning model, it is crucial to perform data pre-processing before training the model. This process involves transforming the data into a format suitable for the machine learning model. In this specific research, the images are resized to  $224 \times 224$  pixels, and the pixel values are normalized by dividing them by 255. This transformation converts the range from 0–255 to 0–1, preparing the data for the model.

This research aims to explain the process of training a model to achieve accurate classification of X-ray images. The training will be conducted in two phases:

- 1. Determining the optimal number of deep-learning layers;
- 2. Applying image augmentation techniques.

To select the optimal method for achieving high classification accuracy, both stages will be evaluated using the F1-score metric. The experiment will maintain the same setup as presented in Table 3, with 65 epochs and a batch size of 32, ensuring consistency. Additionally, the Adam optimizer will be used with an initial learning rate of  $5 \times 10^{-6}$ . These hyperparameters were chosen based on previous experimentation and have been found to be effective for the task at hand [32]. By maintaining consistent hyperparameters across experiments, the obtained results can be easily compared and evaluated, leading to more reliable and accurate conclusions.



Figure 6. Research scheme.

Table 3. Hyperparameters Setup.

Hyperparameters			
Batch Size	32		
Optimizer	Adam		
Learning rate	$5  imes 10^{-6}$		
Epochs	65		

In the first stage, experiments with various layer configurations will be conducted to establish the ideal number of deep-learning layers. This will involve building models with different numbers of layers. To increase the diversity of the training data, image augmentation techniques will be applied in the second step to the X-ray images. This will include transformations, such as rotation, flipping, and cropping, to generate new versions of the images for the model to learn from. The best augmentation techniques will be selected for use in the final model. Our aim is to develop a highly accurate model for X-ray image classification that can support medical diagnosis by incorporating the outcomes of both training phases.

#### 4. Results and Discussion

#### 4.1. Determining the Optimal Number of Deep-Learning Layers

In this section, there are two processes, determining the number of convolutional blocks and the number of LSTM layers. Two training scenarios were employed to determine the optimal number of convolutional blocks: one with four convolutional blocks and another with five convolutional blocks. Each convolutional block in this study consisted of two convolutional layers with a kernel size of  $3 \times 3$ . Following the convolution-al layers, there was one layer of batch normalization and, subsequently, a layer of max pooling with a pool size of  $2 \times 2$ . Additionally, a dropout layer with a parameter of 0.5 was added to the output layer in both scenarios.

Figure 7 illustrates the architecture consisting of four convolutional blocks, with each block employing a distinct number of convolutional filters. Specifically, the convolutional filters used were 64, 128, 256, and 512 for the corresponding blocks. Figure 8 illustrates the architecture with five convolutional blocks. This architecture is similar to the one with four convolutional blocks, with the addition of an extra block that has a convolutional filter size of 512.



Figure 7. Convolutional neural network with 4 convolution blocks.



Figure 8. Convolutional neural network with 5 convolution blocks.

From Table 4, it can be concluded that the second scenario, which utilizes a convolutional neural network with five convolutional blocks, achieves better performance than scenario 1, with an F1 score of 0.97. This indicates that increasing the number of convolutional blocks in the model allows for deeper and more comprehensive learning of the input data, resulting in improved performance. As the best scenario involves using five convolutional blocks, for the next step, which is determining the number of LSTM layers, we will proceed with the five convolutional layers in the model.

Table 4. Results for the Number of Convolution Blocks.

Scenario	F1 Score
4 convolutional blocks	0.96
5 convolutional blocks	0.97

The next step involves determining the optimal number of LSTM layers. Two scenarios are conducted, one with one LSTM layer and another with two LSTM layers. Since the LSTM layer accepts two-dimensional input while the output of the convolutional layer is three-dimensional, we need to reshape the convolutional output value to match the input shape required by the LSTM layer. The output value of the initial convolution, which is (2, 2, 512), is reshaped (4, 512) to serve as the input for the LSTM layer.

Figure 9 presents the architecture of the CNN-LSTM model with one layer of LSTM. This architecture utilizes five convolutional blocks and incorporates an additional layer of LSTM with a unit size of 100. Consequently, the output shape of this model will be (100). Figure 10 presents the architecture with two layers of LSTM. This architecture is similar to the one with one layer of LSTM but includes an additional LSTM layer with a unit size of 50. As a result, the output shape will be (50).



#### Figure 9. CNN-LSTM with 1 LSTM Layer.



Figure 10. CNN-LSTM with 2 LSTM Layer.

Based on the findings presented in Table 5, it can be concluded that using two layers of LSTM leads to improved performance compared to using only one layer. The model with two LSTM layers achieves an F1 score of 0.99, indicating its effectiveness. This suggests that the deeper architecture with two layers of LSTM enables the model to capture more complex patterns and dependencies in the data, resulting in higher accuracy in classifying lung diseases. The additional layer allows for a more comprehensive analysis of sequential information and enhances the model's ability to make accurate predictions. Therefore, at this stage, it can be inferred that the optimal model utilizes five convolutional blocks and two LSTM layers.

Table 5. Results for the Number of LSTM layers.

Scenario	F1 Score
1 layer of LSTM	0.96
2 layers of LSTM	0.99

## 4.2. Applying Image Augmentation Techniques

This stage will involve model training by applying image augmentation techniques. Image augmentation is a method used to perform oversampling, which helps address imbalanced datasets by resampling the imbalanced classes [51]. In this study, the following three types of augmentations will be applied: rotation; shifting; and zooming. The test results were obtained by applying image augmentation techniques, as shown in Table 6. Three augmentations yielded the best results. Firstly, a rotation of five degrees achieved an F1 score of 0.98. Secondly, a height shift, which involves vertically moving the image with a range of 0.1 (10% of the size of the image height), resulted in an F1 score of 0.98. Thirdly, a width shift, which involves horizontally shifting the image with a range of 0.1 (10% of the size of the image width), yielded an F1 score value of 0.98.

Augmentation	F1 Score
Without augmentation	0.99
Rotate (5)	0.98
Rotate (10)	0.97
Rotate (15)	0.97
Zoom (0.1)	0.95
Zoom (0.2)	0.91
Heigh Shift (0.1)	0.98
Width Shift (0.1)	0.98

 Table 6. Image augmentation test results.

After obtaining the best augmentation results, we attempted to combine the three augmentations and recorded the results in Table 7. Surprisingly, the model without augmentation yielded better results than the augmented model, with an F1 score of 0.99. This could be attributed to the fact that the augmented model requires more training to achieve high accuracy. By examining the training process depicted in Figures 11 and 12, it is evident that the model without augmentation achieved higher accuracy and lower loss within the same number of epochs. On the other hand, the augmented model shows potential for improved results if trained for a greater number of epochs.

Table 7. Result of combined augmentation.

Augmentation	F1 Score
Without augmentation	0.99
Rotate (5) Heigh Shift (0.1) Width Shift (0.1)	0.92



Figure 11. Loss and accuracy model without augmentation.



Figure 12. Loss and accuracy model with augmentation.

To further investigate this, an experiment was conducted by training the augmented model for 85 epochs. The findings, illustrated in Figure 13, indicate that increasing the number of epochs only benefits the training data, as the loss consistently decreases while accuracy steadily increases. However, there is no significant difference observed in the validation data when compared to the 65-epoch model. The graph of the validation data exhibits fluctuations and even shows signs of overfitting around the 80th epoch. This observation may be attributed to the lack of variation in the dataset and the standardized format of many X-ray datasets, rendering augmentation unnecessary.





Models with augmentation also require a longer training time. Table 8 presents a comparison of the training process time using the same machine on Kaggle and the GPU P100 accelerator. It turns out that the model without augmentation has the fastest training process, with a duration of 91 s per epoch.

The most optimal model, based on the findings of stages 1 and 2, consists of five convolutional blocks and two layers of LSTM. Interestingly, this model achieves outstanding accuracy and also boasts the fastest training process.

250

100

200

Augmentation	F1 Score	Training Time (s/Epoch)
Without augmentation	0.99	91
Rotate (5)	0.98	199
Heigh Shift (0.1)	0.98	199
Width Shift (0.1)	0.98	199
Rotate (5) Heigh Shift (0.1) Width Shift (0.1)	0.92	202

Table 8. Comparison of model training time.

## 4.3. Evaluation

An evaluation was conducted to assess the performance of the optimal model. The model was validated using a dataset of 1514 observations, including 361 for the COVID-19 class, 1019 for the normal class, and 134 for the pneumonia class. The classification results are presented in Table 9, demonstrating that the model performs well in accurately classifying X-ray images, with a significant number of correct classifications. Out of the total 1514 images, 1497 were classified correctly, resulting in an overall accuracy of 98.88% (1497/1514). However, some errors were observed in the COVID-19 class, with 11 images misclassified as normal. Figure 14 displays the misclassified images for the COVID-19 class, highlighting that some of these images do not meet the required standards, such as being too small, leading to their misclassification.

Table 9. Confusion matrix.

Actual		Predict	
	COVID-19	Normal	Pneumonia
COVID-19	350	11	0
Normal	2	1017	0
Pneumonia	0	4	130



Figure 14. Image of COVID-19 misclassified as normal.

The classification report is presented in Table 10, revealing that the CNN-LSTM model effectively classifies X-ray images of the lungs with an overall F1 score of 0.99. The precision score for the pneumonia class was found to be the highest at 1.00, indicating a low probability of misclassifying COVID-19 or normal images as pneumonia. On the other hand, the recall score for the normal class was the highest at 1.00, indicating a low error rate in misclassifying normal images into other classes. The F1 score for the normal class was also the highest at 0.99, demonstrating that the CNN-LSTM model performs exceptionally well in predicting the normal class compared to the other two classes. These findings suggest that the CNN-LSTM model exhibits a high level of accuracy in classifying X-ray images of the lungs, which could potentially assist medical professionals in diagnosing respiratory diseases.

Table 10. Classification report.

	Precision	Recall	F1 Score
COVID-19	0.99	0.97	0.98
Normal	0.99	1.00	0.99
Pneumonia	1.00	0.97	0.98
Accuracy			0.99

To validate the performance of the constructed model architecture, the k-fold crossvalidation method will be employed. This method aims to assess the model's performance when tested with different datasets. In this study, the test was conducted 10 times using 10 different validation data sets. The results, as presented in Table 11, indicate an average precision value of 0.97, a recall value of 0.96, and an F1 score of 0.9776.

Fold	Precision	Recall	F1 Score
1	0.99	0.97	0.9887
2	0.96	0.95	0.9788
3	0.95	0.99	0.9808
4	0.99	0.96	0.9854
5	0.98	0.94	0.9739
6	0.94	0.98	0.9764
7	0.99	0.96	0.9874
8	0.96	0.98	0.9835
9	0.97	0.96	0.9614
10	0.99	0.92	0.9592
Average	0.97	0.96	0.9776

Table 11. Results using k-fold cross-validation.

To assess the performance of our proposed model, we conducted a thorough comparison with other existing models using the same dataset. Table 12 presents a comparison of different architectures for the classification of lung diseases using chest X-ray images. Our proposed architecture (CNN-LSTM) outperforms the other architectures with an impressive F1 score of 0.9887. Despite having a similar training time of 91 s per epoch, the CNN-LSTM model demonstrates superior classification performance in accurately identifying lung diseases. These results highlight the potential of deep-learning models, particularly CNN-LSTM, in improving the accuracy of lung disease classification. The findings of this research emphasize the importance of selecting appropriate architectures for specific tasks and underscore the benefits of utilizing CNN-LSTM models in the field of medical image analysis.

Architectures	F1 Score	Training Time (s/Epoch)
Inception-ResNet-v2 [18]	0.8687	147
ResNet-50 [27]	0.9485	82
VGG16 [20]	0.9728	93
VGG19 [20]	0.9679	109
CNN (5 convolutional blocks)	0.9721	92
Proposed architectures (CNN-LSTM)	0.9887	91

Table 12. Comparison of the proposed architectures with existing architectures.

## 5. Conclusions

The analysis of chest X-ray images for COVID-19, pneumonia, and normal cases was conducted using a combination of convolutional neural network (CNN) and long short-term memory (LSTM) models. These findings highlight the significant potential of deep-learning models, particularly the CNN-LSTM architecture, in greatly enhancing the accuracy of lung disease classification. This research emphasizes the critical importance of selecting appropriate architectures tailored to specific tasks and underscores the numerous advantages of employing CNN-LSTM models in medical image analysis.

Among the evaluated models, the one with five convolutional blocks, two LSTM layers, and no augmentation emerged as the most effective, achieving an impressive F1 score of 0.9887 with a training duration of 91 s per epoch. It is worth noting that the main source of misclassifications was observed in 11 COVID-19 datasets mistakenly labeled as normal, accounting for 3.05% of the COVID-19 data. The pneumonia class demonstrated the highest precision (1.00), while the normal class exhibited the highest recall (1.00) and F1 score (0.99). Through k-fold cross-validation with 10 folds, the average precision, recall, and F1 score were calculated to be 0.97, 0.96, and 0.9776, respectively. Overall, incorporating an LSTM layer into the CNN model can improve the classification of chest X-ray images and effectively identify infected lungs. The addition of the LSTM layer allows the model to capture temporal dependencies and effectively model sequential information present in the images. By considering both spatial and temporal information, the CNN-LSTM architecture enhances the accuracy and robustness of the classification task. This combination of CNN and LSTM networks proves to be valuable in medical image analysis, particularly for the detection and diagnosis of lung diseases.

The results of this research provide valuable insights for future studies. To improve the accuracy and reliability of the model, it is recommended that the next study focus on cleaning and organizing the data prior to modeling. This step can help reduce misclassification and enhance the model's performance. Additionally, for further development and to broaden the scope of the study, it is suggested to include more types of lung data, such as lung cancer and tuberculosis. This would enable the model to identify a wider range of lung diseases and provide more comprehensive and accurate results.

**Author Contributions:** J.F., A.A.P., I.N.Y., M.N.A. and F.I., understood and designed this research; J.F. analyzed the data and drafted this paper. All authors critically read and revised the draft and approved the final paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** The authors are grateful to the Directorate for Research and Community Service (DRPM) Universitas Padjadjaran, which supports this research under RPLK Grant no. 2203/UN6.3.1/PT.00/2022.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

**Data Availability Statement:** Data is available at Kaggle [33,34]. The source code of this study is available at https://www.kaggle.com/code/jfachrel/covid19-detection-from-chest-x-ray-images-cnn-lstm, accessed on 5 July 2022.

Acknowledgments: The authors would like to thank the Artificial Intelligence and Big Data Research Center of Universitas Padjadjaran for the facilities provided to carry out this study. We would also like to acknowledge Mohammad Hamid Asnawi for his valuable contribution in reviewing and editing the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Antoniou, K.M.; Margaritopoulos, G.A.; Tomassetti, S.; Bonella, F.; Costabel, U.; Poletti, V. Interstitial Lung Disease. *Eur. Respir. Rev.* 2014, 23, 40–54. [CrossRef] [PubMed]
- Postow, L.; Noel, P.; Lin, S.; Zhou, G.; Fessel, J.; Kiley, J.P. Diagnosing and Treating Lung Disease at the Cellular Level. Am. J. Physiol.-Lung Cell. Mol. Physiol. 2020, 319, L541–L544. [CrossRef] [PubMed]
- Chen, X.; Laurent, S.; Onur, O.A.; Kleineberg, N.N.; Fink, G.R.; Schweitzer, F.; Warnke, C. A Systematic Review of Neurological Symptoms and Complications of COVID-19. *J. Neurol.* 2021, 268, 392–402. [CrossRef] [PubMed]
- 4. McIntosh, K. Community-Acquired Pneumonia in Children. N. Engl. J. Med. 2002, 346, 429–437. [CrossRef] [PubMed]
- Goyal, S.; Singh, R. Detection and Classification of Lung Diseases for Pneumonia and COVID-19 Using Machine and Deep Learning Techniques. J. Ambient. Intell. Humaniz. Comput. 2021, 14, 3239–3259. [CrossRef]
- Gattinoni, L.; Chiumello, D.; Caironi, P.; Busana, M.; Romitti, F.; Brazzi, L.; Camporota, L. COVID-19 Pneumonia: Different Respiratory Treatments for Different Phenotypes? *Intensive Care Med.* 2020, 46, 1099–1102. [CrossRef] [PubMed]
- Narin, A.; Kaya, C.; Pamuk, Z. Automatic Detection of Coronavirus Disease (COVID-19) Using X-ray Images and Deep Convolutional Neural Networks. *Pattern Anal. Appl.* 2021, 24, 1207–1220. [CrossRef]
- Reshi, A.A.; Rustam, F.; Mehmood, A.; Alhossan, A.; Alrabiah, Z.; Ahmad, A.; Alsuwailem, H.; Choi, G.S. An Efficient CNN Model for COVID-19 Disease Detection Based on X-ray Image Classification. *Complexity* 2021, 2021, 6621607. [CrossRef]
- Salman, F.M.; Abu-Naser, S.S.; Alajrami, E.; Abu-Nasser, B.S.; Ashqar, B.A.M. COVID-19 Detection Using Artificial Intelligence. Int. J. Acad. Eng. Res. 2020, 4, 18–25.
- Gilanie, G.; Bajwa, U.I.; Waraich, M.M.; Asghar, M.; Kousar, R.; Kashif, A.; Aslam, R.S.; Qasim, M.M.; Rafique, H. Coronavirus (COVID-19) Detection from Chest Radiology Images Using Convolutional Neural Networks. *Biomed. Signal Process. Control* 2021, 66, 102490. [CrossRef]
- 11. Tekerek, A.; Al-Rawe, I.A.M. A Novel Approach for Prediction of Lung Disease Using Chest X-ray Images Based on DenseNet and MobileNet. *Wirel. Pers. Commun.* 2023, 12, 1–15. [CrossRef] [PubMed]
- 12. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861.
- 13. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. *arXiv* 2016, arXiv:1608.06993.
- 14. Gupta, A.; Mishra, S.; Sahu, S.C.; Srinivasarao, U.; Naik, K.J. Application of Convolutional Neural Networks for COVID-19 Detection in X-ray Images Using InceptionV3 and U-Net. *New Gener. Comput.* **2023**, *41*, 475–502. [CrossRef]
- 15. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. *arXiv* 2015, arXiv:1512.00567.
- 16. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv* 2015, arXiv:1505.04597.
- Badrahadipura, R.; Nur Septi, S.Q.; Fachrel, J.; Yulita, I.N.; Pravitasari, A.A.; Agustian, D. COVID-19 Detection in Chest X-rays Using Inception Resnet-V2. In Proceedings of the 2021 International Conference on Artificial Intelligence and Big Data Analytics, ICAIBDA 2021, Bandung, Indonesia, 27–29 October 2021. [CrossRef]
- Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. In Proceedings of the 31st AAAI Conference on Artificial Intelligence, AAAI 2017, San Francisco, CA, USA, 4–9 February 2017. [CrossRef]
- 19. Abbas, A.; Abdelsamea, M.M.; Gaber, M.M. Classification of COVID-19 in Chest X-ray Images Using DeTraC Deep Convolutional Neural Network. *Appl. Intell.* **2021**, *51*, 854–864. [CrossRef]
- Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015-Conference Track Proceedings, San Diego, CA, USA, 7–9 May 2015.
- 21. Priddy, K.L.; Paul, E. Keller. Artificial Neural Networks: An Introduction; SPIE Press: Bellingham, WC, USA, 2005; Volume 68.
- 22. Boser, B.E.; Guyon, I.M.; Vapnik, V.N. Training Algorithm for Optimal Margin Classifiers. In Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory, Pittsburgh, PA, USA, 27–29 July 1992. [CrossRef]
- 23. Cover, T.; Hart, P. Nearest Neighbor Pattern Classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27. [CrossRef]
- 24. Rahman, A.; Tasnim, S. Ensemble Classifiers and Their Applications: A Review. arXiv Prepr. 2014, arXiv:1404.4088. [CrossRef]
- Demir, F. DeepCoroNet: A Deep LSTM Approach for Automated Detection of COVID-19 Cases from Chest X-ray Images. *Appl. Soft. Comput.* 2021, 103, 107160. [CrossRef]

- Pustokhin, D.A.; Pustokhina, I.V.; Dinh, P.N.; Phan, S.V.; Nguyen, G.N.; Joshi, G.P.; Shankar, K. An Effective Deep Residual Network Based Class Attention Layer with Bidirectional LSTM for Diagnosis and Classification of COVID-19. *J. Appl. Stat.* 2023, 50, 477–494. [CrossRef]
- 27. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
- 28. Touvron, H.; Cord, M.; Sablayrolles, A.; Synnaeve, G.; Jégou, H. Going Deeper with Image Transformers. In Proceedings of the IEEE International Conference on Computer Vision, Montreal, QC, Canada, 11–17 October 2021. [CrossRef]
- 29. Paris, S.; Kornprobst, P.; Tumblin, J.; Durand, F. Bilateral Filtering: Theory and Applications. *Found. Trends Comput. Graph. Vis.* **2009**, *4*, 1–73. [CrossRef]
- Hamza, A.; Attique Khan, M.; Wang, S.-H.; Alqahtani, A.; Alsubai, S.; Binbusayyis, A.; Hussein, H.S.; Martinetz, T.M.; Alshazly, H. COVID-19 Classification Using Chest X-ray Images: A Framework of CNN-LSTM and Improved Max Value Moth Flame Optimization. *Front. Public Health* 2022, 10, 948205. [CrossRef] [PubMed]
- 31. Tan, M.; Le, Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the 36th International Conference on Machine Learning, ICML 2019, Long Beach, CA, USA, 9–15 June 2019; Volume 2019.
- 32. Fachrel, J.; Pravitasari, A.A.; Yulita, I.N.; Ardhisasmita, M.N.; Indrayatna, F. A Comparison between CNN and Combined CNN-LSTM for Chest X-ray Based COVID-19 Detection. *Decis. Sci. Lett.* **2023**, *12*, 199–210. [CrossRef]
- 33. Chowdhury, M.E.H.; Rahman, T.; Khandakar, A.; Mazhar, R.; Kadir, M.A.; Bin Mahbub, Z.; Islam, K.R.; Khan, M.S.; Iqbal, A.; Al Emadi, N.; et al. Can AI Help in Screening Viral and COVID-19 Pneumonia? *IEEE Access* 2020, *8*, 132665–132676. [CrossRef]
- Rahman, T.; Khandakar, A.; Qiblawey, Y.; Tahir, A.; Kiranyaz, S.; Kashem, S.B.A.; Islam, M.T.; Al Maadeed, S.; Zughaier, S.M.; Khan, M.S.; et al. Exploring the Effect of Image Enhancement Techniques on COVID-19 Detection Using Chest X-ray Images. *Comput. Biol. Med.* 2021, 132, 104319. [CrossRef]
- 35. Goodfellow, I.; Bengio, Y.; Courville, A. Deep Learning; MIT Press: Cambridge, MA, USA, 2016.
- 36. Guo, T.; Dong, J.; Li, H.; Gao, Y. Simple Convolutional Neural Network on Image Classification. In Proceedings of the 2017 IEEE 2nd International Conference on Big Data Analysis, ICBDA 2017, Beijing, China, 10–12 March 2017. [CrossRef]
- Yamashita, R.; Nishio, M.; Do, R.K.G.; Togashi, K. Convolutional Neural Networks: An Overview and Application in Radiology. Insights Imaging 2018, 9, 611–629. [CrossRef]
- Lee, H.; Song, J. Introduction to Convolutional Neural Network Using Keras; An Understanding from a Statistician. *Commun. Stat. Appl. Methods* 2019, 26, 591–610. [CrossRef]
- 39. Aparna, S.; Ekambaram Naidu, M. Applying FIR and IIR Digital Filters over Video Image Processing. *Int. J. Appl. Eng. Res.* 2016, 11, 7624–7632.
- Yadav, A.; Jha, C.K.; Sharan, A. Optimizing LSTM for Time Series Prediction in Indian Stock Market. *Procedia Comput. Sci.* 2020, 167, 2091–2100. [CrossRef]
- Liu, Y.; Yu, X.; Wu, Y.; Song, S. Forecasting Variation Trends of Stocks via Multiscale Feature Fusion and Long Short-Term Memory Learning. Sci. Program 2021, 2021, 5113151. [CrossRef]
- 42. Qiu, J.; Wang, B.; Zhou, C. Forecasting Stock Prices with Long-Short Term Memory Neural Network Based on Attention Mechanism. *PLoS ONE* 2020, *15*, e0227222. [CrossRef]
- 43. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. Neural Comput. 1997, 9, 1735–1780. [CrossRef] [PubMed]
- Taroon, G.; Tomar, A.; Manjunath, C.; Balamurugan, M.; Ghosh, B.; Krishna, A.V.N. Employing Deep Learning in Intraday Stock Trading. In Proceedings of the 2020 5th International Conference on Research in Computational Intelligence and Communication Networks, ICRCICN 2020, Bangalore, India, 26–27 November 2020. [CrossRef]
- 45. Wang, Y.; Wu, Q.; Dey, N.; Fong, S.; Ashour, A.S. Deep Back Propagation–Long Short-Term Memory Network Based Upper-Limb SEMG Signal Classification for Automated Rehabilitation. *Biocybern. Biomed. Eng.* **2020**, *40*, 987–1001. [CrossRef]
- 46. Sagheer, A.; Kotb, M. Unsupervised Pre-Training of a Deep LSTM-Based Stacked Autoencoder for Multivariate Time Series Forecasting Problems. *Sci. Rep.* **2019**, *9*, 19038. [CrossRef] [PubMed]
- Ting, K.M. Confusion Matrix. In Encyclopedia of Machine Learning and Data Mining; Springer: Boston, MA, USA, 2017; p. 260. [CrossRef]
- 48. Visa, S.; Ramsay, B.; Ralescu, A.; Van Der Knaap, E. Confusion Matrix-Based Feature Selection. Maics 2011, 710, 120–127.
- Ibrahim, M.; Torki, M.; El-Makky, N. Imbalanced Toxic Comments Classification Using Data Augmentation and Deep Learning. In Proceedings of the 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018, Orlando, FL, USA, 17–20 December 2018. [CrossRef]
- Refaeilzadeh, P.; Tang, L.; Liu, H. Cross-Validation. In *Encyclopedia of Database Systems*; Springer: Boston, MA, USA, 2009; pp. 532–538. [CrossRef]
- 51. Shorten, C.; Khoshgoftaar, T.M. A Survey on Image Data Augmentation for Deep Learning. J. Big Data 2019, 6, 1–48. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.