Decision Science Letters 12 (2023) 199–210

Contents lists available at GrowingScience

Decision Science Letters

homepage: www.GrowingScience.com/dsl

A comparison between CNN and combined CNN-LSTM for chest X-ray based COVID-19 detection

Julio Fachrel[a], Anindya Apriliyanti Pravitasari[a*], Intan Nurma Yulita[b], Mulya Nurmansyah Ardhisasmita[c] and Fajar Indrayatna[a]

[a]*Department of Statistics, Faculty of Math and Natural Sciences, Universitas Padjadjaran, Indonesia*
[b]*Department of Informatics, Faculty of Math and Natural Sciences, Universitas Padjadjaran, Indonesia*
[c]*Department of Epidemiology, Faculty of Medicine, Universitas Padjadjaran, Indonesia*

C H R O N I C L E

Article history:
Received: November 28, 2022
Received in revised format: December 27, 2022
Accepted: February 8, 2023
Available online: February 8, 2023

Keywords:
COVID-19
X-ray
Deep Learning
Convolutional Neural Networks
Long Short-Term Memory

A B S T R A C T

COVID-19 detection through radiological examination is favoured since it is fast and produces more accurate results than the laboratory approach. However, when it has infected many people and put a strain on the healthcare system, the need for fast, automatic COVID-19 detection in patients has become critical. This study proposes to detect COVID-19 from chest X-ray (CXR) images with a machine learning approach. The main contributions of this paper are to compare two powerful deep learning models, i.e., convolutional neural networks (CNN) and the combination of CNN and Long Short-Term Memory (LSTM). In the combination model, CNN is recommended for feature extraction, and COVID-19 is classified using the features of LSTM. The dataset used in this study amounted to 4,095 CXR images, consisting of 1,400 images of normal conditions, 1,350 images of COVID-19, and 1,345 images of pneumonia. Both CNN and CNN-LSTM were executed in a similar experimental setup and evaluated using a confusion matrix. The experiment results provide evidence that the CNN-LTSM is better than the CNN deep learning model, with an overall accuracy of about 98.78%. Furthermore, it has a precision and recall of 99% and 98%, respectively. These findings will be valuable in the fast and accurate detection of COVID-19.

1. Introduction

SARS-CoV-2, also known as the coronavirus, is a virus that causes the disease COVID-19. It first emerged in late 2019 and has since spread globally, resulting in over 500 million confirmed cases and 6 million deaths until now (WHO Coronavirus (COVID-19) Dashboard, 2022). The majority of people who are infected with SARS-CoV-2 do not exhibit symptoms, which is known as being asymptomatic. However, some individuals may experience symptoms such as fever, coughing, and shortness of breath. These symptoms can range from mild to severe. (Wang et al., 2020). The vaccination is currently being widely administered in an effort to curb the spread of COVID-19 and has been shown to reduce severe symptoms and the risk of death. However, cases of COVID-19-related deaths are still common among adults over 40 years old, even after receiving two doses of the vaccine (Sheikh et al., 2020). Additionally, vaccines may cause allergies, making it impossible for some people to receive the vaccine (Schumaker et al., 2020). Therefore, preventive measures such as early detection of COVID-19 are still necessary.

Until now, no medicine has been discovered that can cure the COVID-19 infection. Therefore, testing and tracing are still the best solutions. The most commonly used method for diagnosing COVID-19 is Polymerase Chain Reaction (PCR) test. It is done by taking a sample of the patient's nasal swab, which is then analyzed and combined with a fluorescent dye to detect the presence of the virus (Wang et al., 2020). However, the PCR test has some weaknesses, like some cases that produce negative results even though the patient shows signs of being infected when looking at the results of lung scans (Ai et al., 2020).

* Corresponding author.
E-mail address: anindya.apriliyanti@unpad.ac.id (A. A. Pravitasari)

doi: 10.5267/dsl.2023.2.004

An alternative method for diagnosing COVID-19 infection is by using chest imaging. Narin et al. (2021) also found that images of the lower portion of the lungs had a higher accuracy in detecting COVID-19 than a nasal swab sample using the PCR method. X-ray and Computed Tomography (CT) are clinical diagnostic tools that can depict the condition of the lungs and identify COVID-19 in patients. Although CT has better detection sensitivity, X-ray is more commonly used in clinical practice due to its benefits and ease of use, such as its lower cost, minimal radiation dose, and wide availability in general hospitals (Narin et al., 2021).

The challenge of diagnosing COVID-19 through chest imaging is that it requires radiologists to interpret the images because they are not easily readable for non-experts. Additionally, the high number of patients compared to the number of radiologists can cause ineffective diagnosis as it requires more time and energy. To speed up the detection of COVID-19 through chest imaging, research aims to analyze chest x-ray images through the use of Artificial Intelligence (AI) features. The use of AI with a machine learning approach can help healthcare professionals predict COVID-19 through chest x-ray images with high accuracy and quickly. The detection of the coronavirus in X-ray images using an algorithm developed by Zein (2021) demonstrates that X-ray imagery can be used to detect COVID-19.

Some previous studies based on Deep Learning for the classification of COVID-19 thorax images were proposed by Reshi et al. (2021), Salman et al. (2020) and Gilanie et al. (2021), which discussed that a good CNN approach was used in classification. In other studies, Demir (2021) and Pustokhin et al. (2020) have tried to use Deep LSTM and Deep Bi-LSTM, which can produce accuracy comparable to CNN. Islma et al. (2020) and Dastider et al. (2021) proposed a combination of CNN-LSTM that can improve COVID-19 prediction accuracy on X-ray images. This study tried to propose an improved CNN, CNN-LSTM, and compared them to see the performance of both models in classifying 4,095 X-ray images into three categories: COVID-19, Pneumonia, and Normal.

The remainder of the paper is organized as follows. In Section 2, we review the scientific work related to this research. Section 3 explains the collection and split of the dataset as well as the proposed approach. Section 4 contains the results and a comparison of the experiments that were performed. Finally, in Section 5 and 6, we present the paper's discussion and conclusions as well as recommendations for further research.

2. Literature Review

Reshi et al. (2021) proposed a convolutional neural network (CNN) framework which used about 178 X-ray images as the dataset. The results showed that 42 of the 136 X-ray photos belonged to normal people or those with other illnesses, such as pneumonia, while the remaining 136 images belonged to confirmed COVID-19 patients. The testing results showed an overall accuracy of up to 99.5%, demonstrating the suggested CNN model's exceptional performance in the targeted application domain.

El Asnaoui and Chawki (2021) used deep learning models (VGG16, VGG19, DenseNet201, Inception ResNet V2, Inception V3, Resnet50, and MobileNet V2) to identify and categorize coronavirus pneumonia. The tests employed a chest X-ray and CT dataset of 6087 pictures (2780 images of bacterial pneumonia, 1493 of coronavirus, 231 of COVID-19, and 1583 normal) (2780 images of bacterial pneumonia, 1493 of coronavirus, 231 of COVID-19, and 1583 normal). The findings indicated that the employment of Inception Resnet V2 produced the greatest values of 92% accuracy, 92% sensitivity, 96 % specificity, 92% precision, and 92% F1 score. Even though Inception Resnet V2 provides a great result, it takes longer for the training and testing steps, which take 79,184.28 and 262 seconds, respectively. In addition, we find that Inception V3 is fast and offers powerful results (88% correctness and accuracy). As a result, the scientist can choose between computation time and accuracy when deciding which approach to use. However, since this study is in the medical field, the accuracy of the approaches remains a major selection consideration.

Wang, Lin, & Wong (2020) proposed COVID-Net, a deep convolutional neural network architecture for identifying COVID-19 instances from chest X-ray (CXR) pictures. The collection comprises 13,975 CXR pictures spanning 13,870 patient cases. The experimental results indicated that COVID-Net can achieve high accuracy (93.3%) and good sensitivity (91.0%) for identifying COVID-19 patients using chest X-rays. The COVID-Net network design may achieve a reasonable balance between computational efficiency and performance. The COVID-Net network exceeded the VGG-19 and ResNet-50 networks in accuracy and sensitivity. These findings highlight the advantages of selective long-range connections and the high architectural variety observed in COVID-Net, making it simpler to represent information and train on it.

Islam et al. (2020) created a suggested method to diagnose COVID-19 based on clinical photos, CT scans, and X-rays of the chest using a deep CNN-LSTM network. This architecture's structure was constructed by merging CNN and LSTM networks, with the CNN collecting sophisticated information from pictures and the LSTM functioning as a classifier. Combining retrieved attributes with an LSTM that separates COVID-19 instances from others increases the performance of the proposed system. The constructed system had 99.4% accuracy, 99.9% AUC, 99.2% specificity, 99.3% sensitivity, and a 98.9% F1-score. The recommended CNN-LSTM architecture outperforms the rival CNN network, based on the results of the testing.

Dastider et al. (2021) conducted research to estimate COVID-19 severity from lung ultrasonography. A deep CNN is described for frame-based classification of LUS pictures into four severity levels, followed by a long short-term memory

(LSTM) for handling the temporal features of the LUS films. The proposed model's performance is compared to three distinct baselines: the DenseNet-201 architecture, the new CNN architecture with the LSTM block, and without the LSTM block. The recommended hybrid network, which consists of CNN-LSTM blocks, is employed to produce the best results as it can identify slight alterations between the pictures and consistently anticipate the severity score. When compared to the CNN alone, the CNN-LSTM improves prediction accuracy by an average of 9 to 11%. The sensitivity, specificity, and F1 score all increase by 7 to 9%, 1%, and 6 to 8%, respectively, when compared to the CNN, which improves the sensitivity and specificity by 7 to 12% and 3 to 9%, respectively, over the DenseNet-201.

Table 1
Summary of previous research

Authors	Model	Dataset	Result
Reshi et al. (2021)	CNN	X-Ray Image (136 of COVID-19 and 42 of Normal or other diseases)	CNN is appropriate for identifying COVID-19. The experimental results have shown the overall accuracy is 99.5%,
El Asnaoui et al. (2021)	VGG16, VGG19, DenseNet201, Inception ResNet V2, Inception V3, Resnet50, and MobileNet V2	Chest X-ray & CT (2780 images of bacterial pneumonia, 1493 of coronavirus, 231 of Covid19, and 1583 normal	Inception Resnet V2 architecture is better than the other architectures cited in this work (with accuracy of 92.18%, sensitivity of 92.11%, specificity of 96.06%, and F1 score of 92.07%)
Wang et al. (2020)	COVID-Net	X-ray image (The dataset consists of 13,975 X-ray images across 13,870 patient cases)	COVID-Net achieves good accuracy with a test accuracy of 93.3%
Islam et al. (2020)	CNN-LSTM	X-ray image (1220 data for each class of COVID-19, pneumonia and normal)	The overall accuracy results for the CNN model are 99%, while the CNN-LSTM model obtains 99.4%.

The CNN model has the maximum accuracy, according to research by Reshi et al. (2021), as may be inferred from Table 1. But according to a study by Islam et al. (2020), the CNN-LSTM model is more accurate. This study seeks to evaluate the proposed CNN and CNN-LSTM models, choose the ideal number of deep learning layers, and apply the model to a larger X-ray image dataset, which is different from what was done by Islam et al. (2020).

3. Material and Methods

This section outlines the methodology we proposed as well as the resources we'll employ. The dataset that will be used in this investigation is first described. The proposed approach or architecture that will be used in this research will then be outlined, along with the metrics that will be assessed to determine how well our models perform. The experimental setup will next be described.

3.1. Dataset

The data used in this study is sourced from M.E.H. Chowdhury et al. (2020) and Rahman et al. (2020), which are easily accessible from Kaggle. The data is divided into three categories: normal X-ray image data, pneumonia, and COVID-19 cases. The normal X-ray image data used were 1,400 images, the COVID-19 case were 1,350 images, and the pneumonia X-ray image data were 1,345. The proportion of the training and testing data will be visualized in Table 2, while the sample of the dataset is shown in Fig. 1.

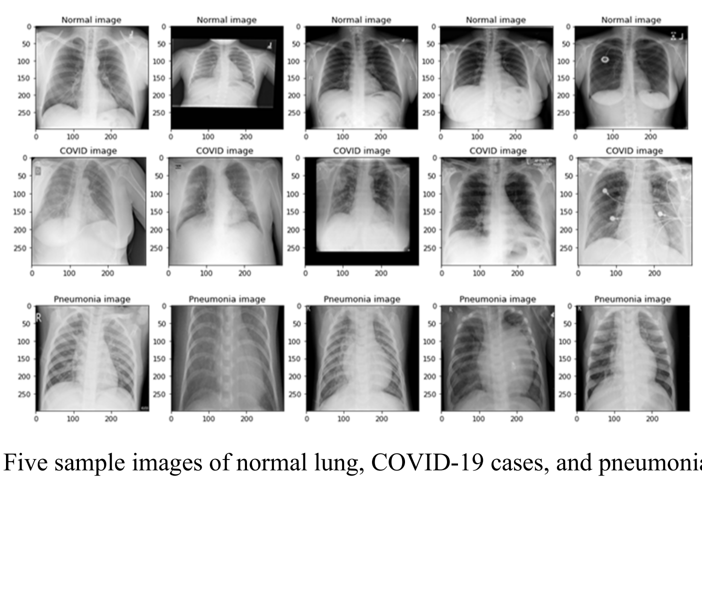


Fig. 1. Five sample images of normal lung, COVID-19 cases, and pneumonia cases.

Table 2
The partitioning description of used dataset

Data/Cases	Normal	COVID-19	Pneumonia	Overall
Training	1.120	1.080	1.076	3.276
Testing	280	270	269	819
Overall	1.400	1.350	1.345	4.045

3.2. Proposed Convolutional Neural Network (CNN)

This study used the Convolutional Neural Network (CNN) architecture to build the core of a hybrid CNN. CNN are inspired by the collection of animal visual cortex and usually consist of one or more convolutional layers, fully connected layers, ReLU layers, and pooling layers (Shen, Wu & Suk, 2017). CNN has neurons consisting of three dimensions: length, width, and height. Neurons in a given layer only connect to a small area of the previous layer. CNN can detect relevant image features, so the output from CNN is usually one or more probability labels or categories associated with the input image (Yamashita, 2018). Therefore, this study used CNN since it's the best model for image processing and computer vision.

A convolutional layer is a matrix-like filter that moves horizontally and then vertically for the following horizontal slide with a given step and kernel size until all pixels have been scanned. This filtered result will result in a new matrix called a "feature map" (Sun et al., 2020). A non-linear layer can be used after the convolution layer to adjust and constrain the output. The Rectified Linear Unit (ReLU) has been used more often because it has a constant gradient for the positive input, and when the gradient is zero, the result is a complete zero (Albawi, 2017).

The pooling layer used is a kind of non-linear down-sampling to decrease the size of the feature map and extract features. The pooling layer is an important factor of CNN in order to achieve an efficient training process and improve accuracy with minimal loss (Jie & Wanda, 2020). The two most commonly used pooling layers are max pooling and average pooling, where max pooling takes the maximum score of a matrix or image subregion and average pooling selects the average score of a matrix or image subregion.

Different from standard CNN architecture, in this study we proposed CNN that added batch normalization and a dropout scenario. Batch normalization is used to train deep neural networks faster and more consistently. Batch normalization can re-parameterize the underlying optimization issue to make it more stable and smooth, as well as give robustness to hyperparameter adjustments and minimize gradient bursts and losses (Santurkar et al., 2018).

The dropout layer is a regularization technique to reduce overfitting in neural networks that prevents complex co-adaptation of training data. In the dropout layer, the process of randomly eliminating connections on the neural network unit is carried out in the training process (Setiawan, 2021). Classification is carried out by the fully connected layer as an output layer based on the features obtained from the convolutional and pooling layers.

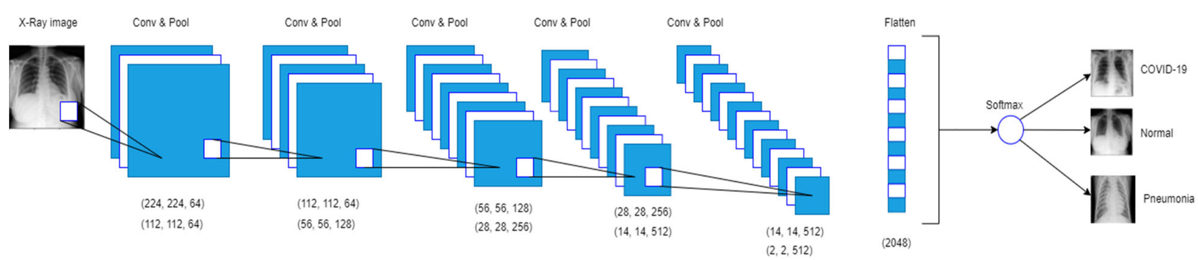


Fig. 2. An illustration of the CNN architecture

Fig. 2 illustrates the architecture of the CNN model used in this study. The best architectural design was chosen for this study as a result of the optimum blocks and layers based on maximum accuracy dan minimum loss. The chosen architecture consists of 12 convolutional layers, six pooling layers, six batch normalization layer, one dropout layer and one output layer. The architectural details are shown in the Table 3. Two or three 2D CNNs, a pooling layer, and a batch normalization layer are coupled with each convolution block. A convolution layer with a kernel size of 3×3 and activated by the ReLu function extracts essential features from images. There are two types of pooling layers used: the max-pooling layer with a kernel size of 2×2 and the average pooling layer with a kernel size of 4×4, which is used at the end of the convolution block. Before the output layer, there is a dropout layer with a 50% dropout rate. The output layer, which is activated by the softmax function, has three units that classify features into three classes (COVID-19, normal, and pneumonia), and there is a dropout layer before it with a 50% dropout rate.

Table 3
The summary of CNN model

Layer	Output Shape	Activation
2 × Convolution (3 × 3)	224 × 224 × 64	Relu
Batch Normalization	224 × 224 × 64	-
Max Pooling (2×2)	112 × 112 × 64	-
2 × Convolution (3 × 3)	112 × 112 × 128	Relu
Batch Normalization	112 × 112 × 128	-
Max Pooling (2x2)	56 × 56 × 128	-
2 × Convolution (3 × 3)	56 × 56 × 256	Relu
Batch Normalization	56 × 56 × 256	-
Max Pooling (2x2)	28 × 28 × 256	-
3 × Convolution (3 × 3)	28 × 28 × 512	Relu
Batch Normalization	28 × 28 × 512	-
Max Pooling (2×2)	14 × 14 × 512	-
3 × Convolution (3 × 3)	14 × 14 × 512	Relu
Batch Normalization	14 × 14 × 512	-
Max Pooling (2×2)	7 × 7 × 512	-
Average Pooling (4×4)	2 × 2 × 512	-
Flatten	2048	-
Batch Normalization	2048	-
Dropout	2048	-
Output	3	Softmax

3.3. Long Short-Term Memory (LSTM)

Long Short-Term Memory is a type of recurrent neural network (RNN) capable of storing data patterns and determining which data will be stored and discarded (Yan et al., 2021). The LSTM model consists of a memory cell with a gate structure that replaces the hidden layer neurons of the RNN (Aldi, Jondri, & Aditsania, 2018). In the structure of the LSTM model, there are input, forgotten, and output gates. The function of the gate is to deny or allow access to the LSTM memory. The input gate will block all small values (close to 0) from entering memory as shown in Fig. 3. Forgeting gate will remove all values from memory. Meanwhile, the output gate determines whether the values stored in the LSTM memory should be output. Each memory cell has three sigmoid layers and one tanh layer (Liu et al., 2021).

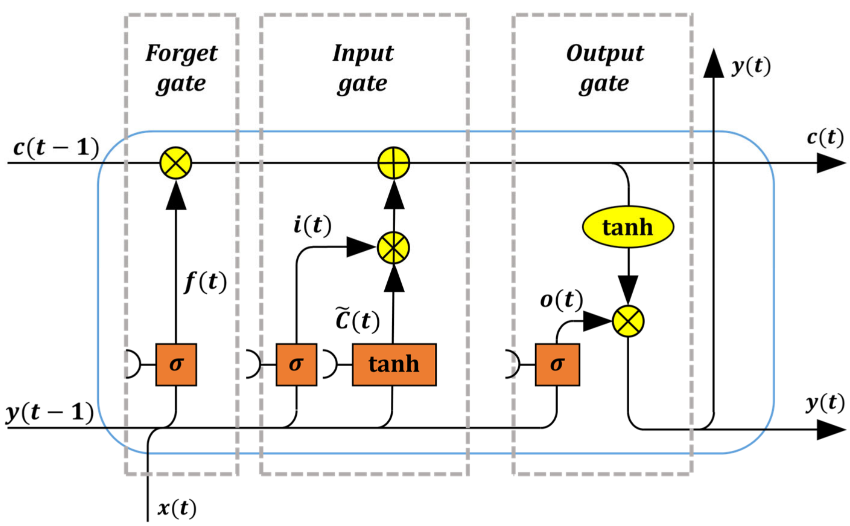


Fig. 3. Long short-term memory structure (modified from Yu et al., 2019)

The following are formula of LSTM calculation process (Qiu, Wang & Zhou, 2020).

1) The last moment output value y_{t-1} and the present input value x_t become the input of the forget gate, where W_f is the weight matrix, b_f is bias of the forgotten gate, and σ is the sigmoid function. The output value of forget gate f_t is obtained by using the formula (1).

$$f_t = \sigma(W_f[y_{t-1}, x_t] + b_f) \quad (1)$$

2) The last time output value and present input value are entered into the input gate. The formula (2) and (3) obtains output value i_t and candidate cell state $\tilde{C}_t$ of the input gate, where b_i and b_c are bias from input gate and cell state, and W_i and W_c are the weight input gate and cell state, respectively.

$$i_t = \sigma(W_i\,[y_{t-1}, x_t] + b_i) \tag{2}$$

$$\tilde{C}_t = tanh(W_c\,[y_{t-1}, x_t] + b_c) \tag{3}$$

3) Renew the current cell C_t state using the formula (4).

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \tag{4}$$

4) The output value of y_{t-1} and input value of x_t are accepted as input values from the output gate at time t. The results o_t from the output gate are obtained using the formula (5).

$$o_t = \sigma(W_o\,[y_{t-1}, x_t] + b_o) \tag{5}$$

5) LSTM results h_t are obtained using the formula (6).

$$y_t = o_t \times \tanh\,(C_t) \tag{6}$$

3.4. CNN-LSTM

In the CNN – LSTM structure, the CNN layer is used for feature extraction on input data which is then combined with LSTM to help predict the sequence. In general, Fig. 4 depicts the CNN-LSTM used in this study.

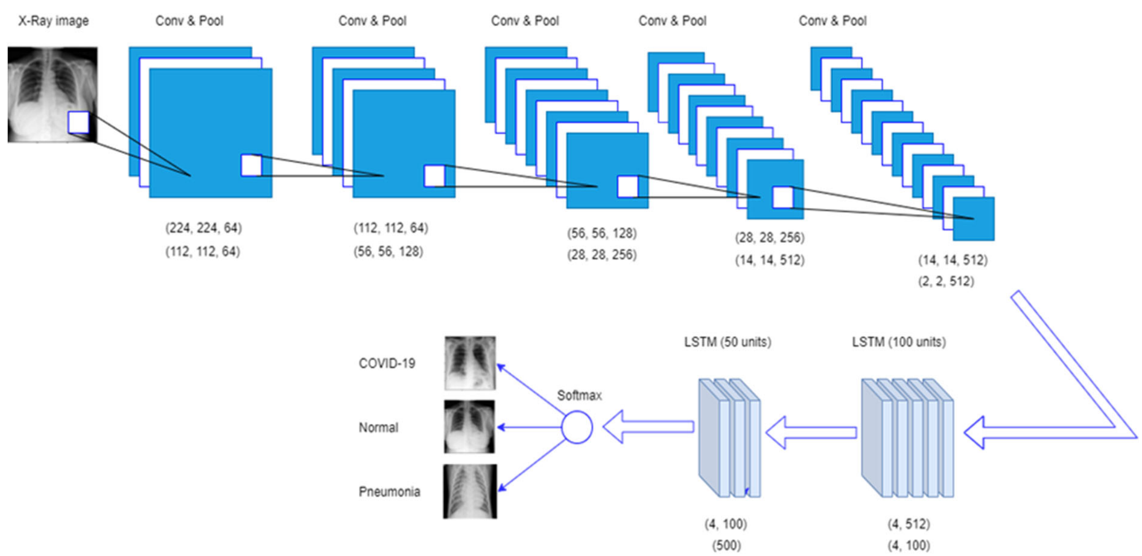


Fig. 4. An illustration of the CNN-LSTM architecture

CNN-LSTM consists of the Convolution, Pooling, LSTM, and Dense layers (Lu et al., 2020). The number of filters, kernel size, and the number of strides of each layer can be adjusted based on the data characteristics (Zhou et al., 2015). The effect of the adjustment to parameters is speed and performance.

The architecture of the CNN-LSTM used in this study is the hybrid of the proposed CNN with the LSTM model. This architecture connects the LSTM layer to the last part of the convolution block. There are two layers of LSTM used, the first LSTM layer has 100 output space dimension units, and the second layer has 50 units that only return the last hidden state output, which is an abstract representation of the input sequence. Each layer is followed by a dropout rate of 50%. The output layer has three units used to classify features into three classes (COVID-19, normal, and pneumonia) and activated by the softmax function. From Fig. 4, the layer on the CNN model is used to determine correlation and extraction of multidimensional data (Kim & Cho 2019).

The CNN-LSTM model summary is shown in Table 4. The last layer in the convolution block is the average pooling layer, and the output shape is found (None, 2, 2, 512), the reshape method is used, and the output shape becomes (None, 4, 512) as input to the LSTM layer. After the LSTM layer, there is a batch normalization layer and a dropout layer characterized by a 25% dropout rate before the output layer.

Table 4
The summary of CNN-LSTM model

Layer	Output Shape	Activation
2 × Convolution (3 × 3)	224 × 224 × 64	Relu
Batch Normalization	224 × 224 × 64	-
Max Pooling (2×2)	112 × 112 × 64	-
2 × Convolution (3 × 3)	112 × 112 × 128	Relu
Batch Normalization	112 × 112 × 128	-
Max Pooling (2×2)	56 × 56 × 128	-
2 × Convolution (3 × 3)	56 × 56 × 256	Relu
Batch Normalization	56 × 56 × 256	-
Max Pooling (2×2)	28 × 28 × 256	-
3 × Convolution (3 × 3)	28 × 28 × 512	Relu
Batch Normalization	28 × 28 × 512	-
Max Pooling (2×2)	14 × 14 × 512	-
3 × Convolution (3 × 3)	14 × 14 × 512	Relu
Batch Normalization	14 × 14 × 512	-
Max Pooling (2×2)	7 × 7 × 512	-
Average Pooling (4×4)	2 × 2 × 512	
LSTM (100)	4 × 100	Relu
LSTM (50)	50	Relu
Batch Normalization	50	-
Dropout	50	-
Output	3	Softmax

3.5. Experimental Setup

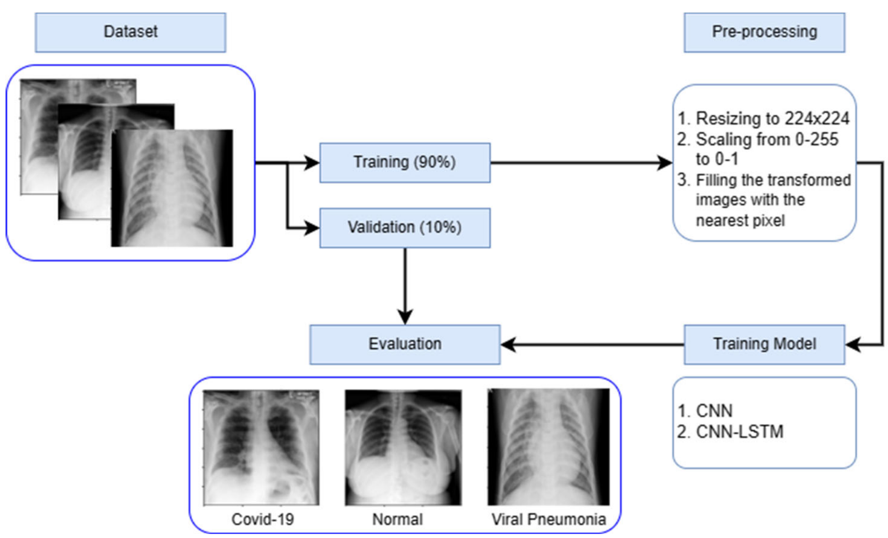


Fig. 5. Schema of model training

Fig. 5 visualizes the schema of model training. As seen in Fig. 5, we implemented the same experimental setup for those two deep learning models. In image preprocessing, the methods that will be used are changing the shape of the image's shape, rescaling, and image augmentation, which will later be implemented into the dataset. The preprocessing stage is carried out to ensure that the researcher can make better training decisions on the targeted features when entering data into the model. In the preprocessing stage, the image will be resized to 224×224 pixels. ImageDataGenerator TensorFlow is used to share and process the previous data in the preprocessing stage. In addition, image augmentation is done by rotating the image randomly by 5 degrees.

There are three categories of images: normal, COVID-19, and viral pneumonia. We have split images into 80% and 20% for training and testing, respectively. The training was run on a Kaggle notebook using a GPU accelerator. The model was compiled with the Adam optimizer using a learning rate 5×10^{-6} and a batch size of 63. The model's loss function in this study used categorical cross-entropy, while the model evaluation metric for training was defined as accuracy. The maximum number of epochs for training time was set to 85 epochs.

3.6. Evaluation

The confusion matrix can measure model performance or assess model feasibility. The confusion matrix produces accuracy scores, sensitivity, precision, recall, and F1-scores. These scores help evaluate the performance or feasibility of the model used. An $n \times n$ confussion matrix displays the predicted and actual classification, where n is the number of different classes (Pravitasari et al, 2020).

Table 5
Confusion matrix

Predicted values	Actual Values	
	Positive	Negative
Positive	True positive (TP)	False positive (FP)
Negative	False negative (FN)	True negative (TN)

Table 5 shows the structure of a 2×2 confussion matrix where accuracy, recall, specificity, and F1-scores are calculated by formula (7) to (10), respectively.

$$accuracy = \frac{TP+TN}{TN+FP+TP+FN} \tag{7}$$

$$recall = \frac{TP}{TP + FN} \tag{8}$$

$$specificity = \frac{TN}{TN + FP} \tag{9}$$

$$F1 - score = \frac{2 \times TP}{2 \times TP+FP+FN} \tag{10}$$

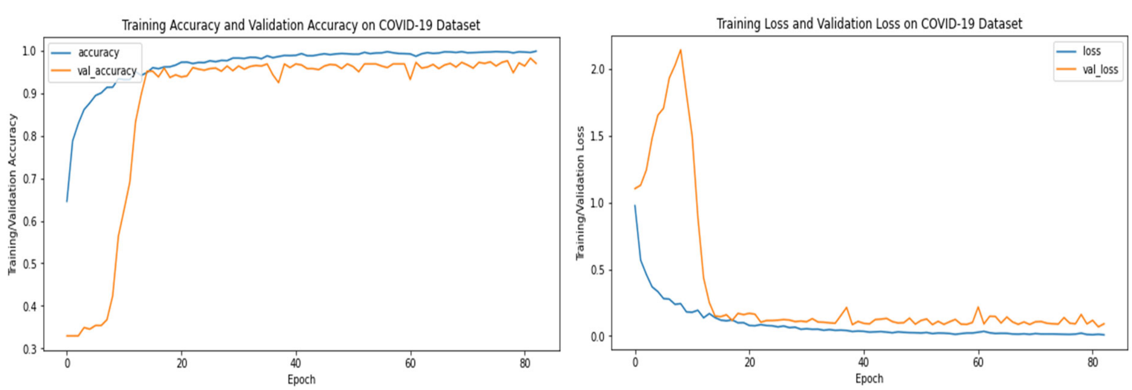
In this case, there are 3 classes. Then for one of the classes, COVID-19, TP represents a correctly classified COVID-19 case, and TN represents a non-COVID-19 case (normal or pneumonia) that is classified correctly. FP represents the incorrectly classified COVID-19 case as non-COVID-19, and FN represents the non-COVID-19 class that is incorrectly classified as COVID-19.

4. Result

In this section, we discuss the results obtained for classifying images into three classes: normal, COVID-19 and Pneumonia. We have split images into 80% for training and 20% for testing and compared two deep learning models, our proposed CNN and CNN-LSTM.

4.1. Proposed CNN

The performance of proposed CNN is shown in Fig. 6 with 85 epochs and the average time per epoch is 51 seconds.



(a) (b)

Fig. 6. The graph of (a) accuracy and (b) loss based on the proposed CNN model.

The accuracy graph shows that the accuracy of the training dataset increases rapidly until the 17th epoch and increases steadily until the 85th epoch, with an accuracy of 99%. The validation dataset starts to increase rapidly from the 8th epoch to the 17th epoch and runs steadily until the last epoch with an accuracy of 97%. The training and validation loss graphs show that the loss of the training data set decreases rapidly until the 17th epoch, while the validation data increases rapidly until the 8th epoch and then decreases until the 17th epoch.

Using the proposed CNN, as shown in Table 6, about 794 out of the 819 test cases were accurately detected and divided into three types. This model achieves accuracy and recall of 96.95% (794/819) and 98.89% (264/270), respectively. It appears that 5,32% of cases were incorrectly categorized as COVID-19 patients when they should have been classified as normal because there was the highest confusion in cases of normal (5,32%). Table 7 has a report on classification. The accuracy of the F1-score reaches 97%, while the COVID-19 case achieved 95% precision, 98% recall, and a 96% F1-score (all in rounded numbers).

Table 6
Confusion matrix of the CNN model

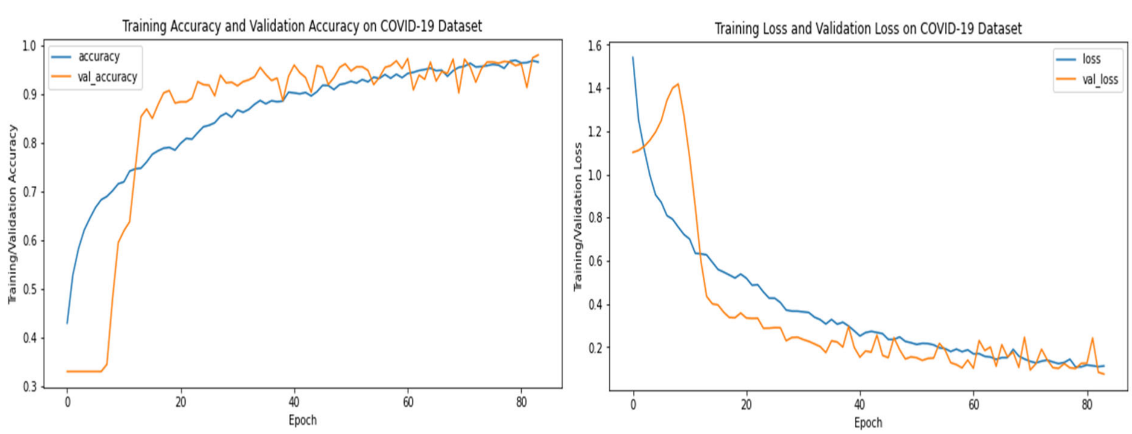
Actual	Predicted		
	COVID-19	Normal	Pneumonia
COVID-19	264	6	0
Normal	14	263	3
Pneumonia	1	1	267

Table 7
Classification report of the CNN model

	Precision	Recall	F1-Score	Support
COVID-19	0.95	0.98	0.96	270
Normal	0.97	0.94	0.96	280
Pneumonia	0.99	0.99	0.99	269
Accuracy			0.97	819
Macro average	0.97	0.97	0.97	819
Weighted average	0.97	0.97	0.97	819

4.2. CNN-LSTM

Fig. 7 depicts the CNN-LSTM model's performance over 85 epochs with the average time per epoch is 55 seconds. The accuracy graph indicates that the training dataset's accuracy steadily improves until the 85th epoch, when it reaches 98% accuracy. After the eighth epoch, the validation dataset's accuracy began to rise quickly, and by the 12th epoch, it had surpassed that of the training dataset, and it continued to run with 99% accuracy until the final epoch. The training and validation loss graphs demonstrate that the training dataset's loss continuously lowers until the 85th epoch, but the validation dataset's loss quickly grows until the 8th epoch, then rapidly drops until the 17th epoch, and then steadily decreases until the last epoch.



(a) (b)

Fig 7. The (a) training and (b) validation on the COVID-19 dataset based on the proposed CNN-LSTM model

Table 8 shows the confusion matrix as an evaluation of the CNN-LSTM network in classifying chest X-ray images. Among the 819 test dataset, 809 were detected correctly and classified into three classes. Overall accuracy was 98.78% (809/819), with 98.89% (264/270) recall. The biggest confusion was found in cases of COVID-19 (2.22%), meaning that 2.22 % occurred misclassification, which should belong to the COVID-19 case but was classified as normal case.

Table 8
Confusion matrix of the CNN-LSTM model

Actual	Predicted		
	COVID-19	Normal	Pneumonia
COVID-19	264	6	0
Normal	2	277	1
Pneumonia	1	0	268

The classification report can be seen in Table 9. The accuracy of the F1-score reaches 99%. While the COVID case achieved 99% precision, 98% recall, and 98% F1-Score (all in rounded numbers).

Table 9
Classification report of the CNN model

	Precision	Recall	F1-Score	Support
COVID-19	0.99	0.98	0.98	270
Normal	0.98	0.99	0.98	280
Pneumonia	1.00	1.00	1.00	269
Accuracy			0.99	819
Macro average	0.99	0.99	0.99	819
Weighted average	0.99	0.99	0.99	819

5. Discussion

In this study, we compared two powerful deep learning models, namely CNN and CNN-LSTM, to classify three classes of X-ray images (COVID-19, normal, and pneumonia). The results show that with 85 epochs, the combined CNN-LSTM model has a higher accuracy than the CNN model, where the accuracy reaches 98.78% and the F1-Score accuracy is 99%. CNN-LSTM also has a better performance for classifying COVID-19 cases, where precision, recall, and F1-score are 99%, 98%, and 98%, respectively. Otherwise, CNN is superior to CNN-LSTM based on the speed of training the model, where CNN takes 51 seconds/epoch and CNN-LSTM takes 55 seconds/epoch. This can be caused by the addition of 2 LSTM layers on the CNN-LSTM model so that the training time is slightly longer. However, the time difference is not too far.

The CNN-LSTM model can categorize X-ray images effectively. As seen in Table 10, CNN-LSTM can compete with existing models when compared to its accuracy. CNN-LSTM that we propose has higher accuracy than Inception Resnet V2 and COVID-Net models conducted by Reshi et al. (2021) and El Asnaoui et al. (2021). Successively, the accuracy they produced was 92.18% and 93.3%, while our model is able to obtain 98.78% accuracy.

Table 10
Comparison results of the proposed methods with the other deep learning methods on COVID-19 diagnosis

Approach	Data type	Cases number	method	Accuracy
Reshi et al. (2021)	X-Ray	178 (136 COVID-19 ,and 42 normal or people with other diseases)	CNN	99.5%,
El Asnaoui et al. (2021)	X-ray & CT	6087 (2780 of pneumonia, 1583 of normal, 1493 of coronavirus, and 231 of COVID-19)	Inception ResNet V2	92.18%
Wang et al. (2020)	X-Ray	13604 (8,066 of normal and 5,538 of non-COVID-19 or pneumonia.)	COVID-Net	93.30%
Islam et al. (2020)	X-Ray	1.525 for each case (COVID-19, normal, pneumonia)	CNN-LSTM	99.40%
Proposed CNN	X-Ray	4.095 (1400 of normal, 1350 of COVID-19, and 1345 of pneumonia)	CNN	96,95%
Proposed CNN-LSTM	X-Ray	4.095 (1400 of normal, 1350 of COVID-19, and 1345 of pneumonia)	CNN-LSTM	98.78%

Despite the positive findings, there are some limitations to this study. First, we only used 4,095 datasets divided into three categories. Second, this study only uses one type of image augmentation, rotation, whereas the next researcher can use a variety of methods. Due to time and equipment constraints, cross validation was not used in this study.

6. Conclusions

In this study, we tried to classify X-ray images, which are divided into three classes. We proposed CNN with 12 layers and added batch normalization and dropout scenarios. We compared the proposed CNN model with the CNN-LSTM to find the best performance. In general, these two models can find the features and characteristics of the lungs and perform the classification very well, even on a limited dataset. Based on accuracy, it found that the best model is a combination of CNN-LSTM with an accuracy of 98.78% and 99% for the F-1 score. Furthermore, it has a precision and recall of 99% and 98%, respectively. Based on the training time, the CNN model is faster than the CNN-LSTM model, where CNN takes 51 seconds per epoch while CNN-LSTM takes 55 seconds per epoch.

The performance of our proposed system does not use cross-validation. Therefore, for future studies, it would be better to use k-fold cross validation to compare models, and using more recent data would be beneficial in the context of COVID-19 as new variants continue to emerge.

Acknowledgment

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.

References

Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., ... & Xia, L. (2020). Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology.

Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017, August). Understanding of a convolutional neural network. In 2017 international conference on engineering and technology (ICET) (pp. 1-6). Ieee.

Aldi, M. W. P., Jondri, J., & Aditsania, A. (2018). Analisis dan Implementasi Long Short Term Memory Neural Network untuk Prediksi Harga Bitcoin. *eProceedings of Engineering*, *5*(2).

Chowdhury, M.E.H., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M.A., Mahbub, Z.B., Islam, K.R., Khan, M.S., Iqbal, A., Al-Emadi, N., Reaz, M.B.I., & Islam, M. T. (2020). Can AI help in screening Viral and COVID-19 pneumonia? *IEEE Access, 8*, 132665 - 132676.

Coronavirus disease - Answers. (n.d.). Retrieved from World Health Organization, https://www.who.int/ Retrieved 31 December, 2021

COVID-19 Radiography Database, https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database/version/4, Retrieved 14 June, 2022

Dastider, A. G., Sadik, F., & Fattah, S. A. (2021). An integrated autoencoder-based hybrid CNN-LSTM model for COVID-19 severity prediction from lung ultrasound. Computers in Biology and Medicine, 132, 104296.

Demir, F. (2021). DeepCoroNet: A deep LSTM approach for automated detection of COVID-19 cases from chest X-ray images. *Applied Soft Computing*, *103*, 107160.

El Asnaoui, K., & Chawki, Y. (2021). Using X-ray images and deep learning for automated detection of coronavirus disease. *Journal of Biomolecular Structure and Dynamics, 39*(10), 3615-3626.

Gilanie, G., Bajwa, U. I., Waraich, M. M., Asghar, M., Kousar, R., Kashif, A., ... & Rafique, H. (2021). Coronavirus (COVID-19) detection from chest radiology images using convolutional neural networks. *Biomedical Signal Processing and Control*, *66*, 102490.

Islam, M. Z., Islam, M. M., & Asraf, A. (2020). A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images. *Informatics in medicine unlocked, 20*, 100412.

Jie, H. J., & Wanda, P. (2020). RunPool: A Dynamic Pooling Layer for Convolution Neural Network. *Int. J. Comput. Intell. Syst.*, *13*(1), 66-76.

Kim, T. Y., & Cho, S. B. (2019). Predicting residential energy consumption using CNN-LSTM neural networks. *Energy, 182*, 72-81.

Liu, Y., Yu, X., Wu, Y., & Song, S. (2021). Forecasting Variation Trends of Stocks via Multiscale Feature Fusion and Long Short-Term Memory Learning. *Scientific Programming*, 2021.

Lu, W., Li, J., Li, Y., Sun, A., & Wang, J. (2020). A CNN-LSTM-based model to forecast stock prices. *Complexity*, *2020*, 1-10.

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.

Pravitasari, A.A., Iriawan, N., Almuhayar, M., Azmi, T., Irhamah, Fithriasari, K., Purnami, S. W., & Ferriastuti, W., (2020). UNet-VGG16 with transfer learning for MRI-based brain tumor segmentation. *TELKOMNIKA, 18*(3), 1310-1318.

Pustokhin, D. A., Pustokhina, I. V., Dinh, P. N., Phan, S. V., Nguyen, G. N., & Joshi, G. P. (2020). An effective deep residual network based class attention layer with bidirectional LSTM for diagnosis and classification of COVID-19. *Journal of Applied Statistics*, 1-18.

Qiu, J., Wang, B., & Zhou, C. (2020). Forecasting stock prices with long-short term memory neural network based on attention mechanism. *PloS one, 15*(1), e0227222.

Rahman, T., Khandakar, A., Qiblawey, Y., Tahir, A., Kiranyaz, S., Kashem, S. B. A., ... & Chowdhury, M. E. (2021). Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images. *Computers in biology and medicine*, *132*, 104319.

Reshi, A. A., Rustam, F., Mehmood, A., Alhossan, A., Alrabiah, Z., Ahmad, A., ... & Choi, G. S. (2021). An efficient CNN model for COVID-19 disease detection based on X-ray image classification. *Complexity*, *2021*, 1-12.

Salman, F. M., Abu-Naser, S. S., Alajrami, E., Abu-Nasser, B. S., & Alashqar, B. A. (2020). Covid-19 detection using artificial intelligence

Santurkar, S., Tsipras, D., Ilyas, A., & Madry, A. (2018). How does batch normalization help optimization?. *Advances in neural information processing systems, 31*.

Setiawan, W. (2021, May). Character Recognition using Adjustment Convolutional Network with Dropout Layer. In IOP Conference Series: Materials Science and Engineering (Vol. 1125, No. 1, p. 012049). IOP Publishing.

Sheikh, A., Robertson, C., & Taylor, B. (2021). BNT162b2 and ChAdOx1 nCoV-19 vaccine effectiveness against death from the delta variant. *New England Journal of Medicine, 385*(23), 2195-2197.

Shen, D., Wu, G., & Suk, H. I. (2017). Deep learning in medical image analysis. *Annual review of biomedical engineering, 19*, 221.

Sun, Y., Xue, B., Zhang, M., Yen, G. G., & Lv, J. (2020). Automatically designing CNN architectures using the genetic algorithm for image classification. *IEEE transactions on cybernetics, 50*(9), 3840-3854.

Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., & Peng, Z. (2020). Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. *Jama, 323*(11), 1061-1069.

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), 1-12.

WHO Coronavirus (COVID-19) Dashboard, https://covid19.who.int/ Retrieved 08 June, 2022.

Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights into imaging, 9*(4), 611-629.

Yan, Y., Yao, X. J., Wang, S. H., & Zhang, Y. D. (2021). A survey of computer-aided tumor diagnosis based on convolutional neural network. *Biology, 10*(11), 1084.

Yu, Y., Si, X., Hu, C., & Zhang, J. (2019). A review of recurrent neural networks: LSTM cells and network architectures. *Neural computation, 31*(7), 1235-1270.

Zein, A. (2021). Pendeteksian virus corona dalam gambar x-ray menggunakan algoritma artificial intelligence dengan deep learning python. *Jurnal ESIT (E-Bisnis, Sistem Informasi, Teknologi Informasi), 15*(1).

Zhou, C., Sun, C., Liu, Z., & Lau, F. (2015). A C-LSTM neural network for text classification. arXiv preprint arXiv:1511.08630.