ELREI: Ensemble Learning of ResNet, EfficientNet, and Inception-v3 for Lung Disease Classification based on Chest X-Ray Image

by Indra kesuma

Submission date: 29-Jan-2024 09:36AM (UTC+0700)
Submission ID: 2280669333
File name: 2023103114_3.pdf (1.05M)
Word count: 7126
Character count: 38345

Received: May 19, 2023. Revised: June 26, 2023.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

ELREI: Ensemble Learning of ResNet, EfficientNet, and Inception-v3 for Lung Disease Classification based on Chest X-Ray Image

Lucky Indra Kesuma[1] Ermatita[2] Erwin[3]*

[1]*Doctoral Program of Engineering Faculty of Engineering Universitas Sriwijaya, Indralaya, Indonesia*
[2]*Information System Department, Computer Science Faculty, Universitas Sriwijaya, Indralaya, Indonesia*
[3]*Computer Engineering Department, Computer Science Faculty, Universitas Sriwijaya, Indralaya, Indonesia*
* Corresponding author's Email: erwin@unsri.ac.id

Abstract: A Chest X-ray (CXR) image can diagnose lung diseases. However, a diagnosis requires time and high accuracy, so an automatic system is needed. Convolutional neural network (CNN) is a reliable method for image classification and has many architectures. ResNet is a CNN architecture that can overcome gradient vanishing, but it has a deep network structure to detect errors. The EfficientNet CNN architecture can proportionally uniformize all depth, width, and resolution dimensions in each layer as needed, but it takes a long time in training. The Inception-v3 CNN architecture uses Inception blocks by reducing dimensions to small convolutions, but it has larger parameters than other architectures. ELREI is an acronym for ensemble learning of ResNet, EfficientNet, and Inception-v3 with weighted voting. ELREI combines the classification results on the ResNet, EfficientNet, and Inception-v3 architectures to overcome the limitations and combine the advantages of each architecture. ELREI works on the training stage at each epoch rather than the final results of each architecture. In addition to voting, ELREI uses a fully convolutional Network (FCN) at the final stage for the best weight determination and to prevent overfitting during training. The results of the accuracy, precision, recall, and F1-score of the ELREI method are excellent, above 98%. The training graph of the ELREI ensemble method proves that ELREI can overcome overfitting that occurs on the architectures. The results show the Ensemble ELREI method is excellent and robust for lung disease classification based on CXR images, which are carried out in 4 classes: normal, COVID-19, lung opacity, and pneumonia.

Keywords: Classification, EfficientNet, ELREI, Ensemble Learning, Inception-v3, ResNet.

1. Introduction

Coronavirus disease 2019 (COVID-19) is a respiratory infection caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus. Reverse transcription polymerase chain reaction (RT-PCR) tests can be used to diagnose COVID-19. However, RT-PCR has limitations, such as limited RT-PCR facilities and expensive test prices [1]. As a result, researchers are looking for newer and easier ways to diagnose COVID-19. One way is to use Chest X-ray (CXR) images. According to [2], CXR images are considered effective for diagnosing COVID-19 disease because they can show abnormalities in the patient's lungs. Unfortunately, CXR images usually have low image quality and noise, making it difficult for radiologists to detect COVID-19 disease. In this case, a fast and precise automatic diagnosis system is needed to classify CXR images.

The convolutional neural network (CNN) is now developing as a reliable method for image classification [3, 4]. Many architectures in CNN are often used for image classification, such as ResNet, DenseNet, EfficientNet, Inception-v3, and others. CNN can increase its performance by adding layers [5]. However, adding layers to CNN can result in increased layer complexity and gradient vanishing [6]. ResNet can be used to overcome this [6, 7]. ResNet uses skip connections in its layers to handle the gradient vanishing problem. A skip connection can accelerate the convergence of the network in ResNet. The advantage of ResNet is that performance

does not decrease even though the architecture is getting deeper. In addition, ResNet also has a lower network complexity compared to other CNN architectures [7]. The weakness of the ResNet architecture is that it is difficult to detect errors due to the deep layers [8]. Several studies have shown that using CXR images to classify COVID-19 is effective [9, 10]. Research by [9] used ResNet 4 classify COVID-19, which resulted in accuracy, recall, and F1-score values above 90%. Unfortunately, this research does not calculate the precision value. Research by [10] used ResNet-50 to classify COVID-19. The results showed an accuracy value above 90%. However, this research does not calculate recall, precision, and F1-score. 4

Another CNN architecture that can be used for image classification is EfficientNet. The EfficientNet architecture has the advantage of being able to scale dimensions such as depth, width, and resolution in layers uniformly. Scaling in EffcientNet is accomplished using a simple and effective coefficient combination [11]. Additionally, EffcientNet produces classification accuracy that is higher than that of other CNN architectures [12]. Research by [8] used EffcientNet-B4 to achieve accuracy and F1-score values above 86% for cancer image classification but did not measure recall and precision. Research by [13] applied an efficient net to Kalposcapi image classification, achieving an accuracy score of more than 90%, but did not measure other performance.

The Inception-v3 architecture is frequently used in image classification. Inception-v3 performs dimensionality reduction by using stacked 1×1 convolutions. Convolution will increase computational efficiency and make the network wider rather than deeper [7]. Several studies have successfully performed classification using Inception-v3 [21], [22]. Research by [14] successfully applied InceptionNet to the COVID-19 classification with accuracy above 90% but did not measure other performance. Research by [15] successfully applied InceptionNet to COVID-19 classification with accuracy, precision, and recall values that are still rather low.

In image classification, each CNN architecture must have advantages and disadvantages. The performance result of each single classification method can be maximized using the ensemble learning method. Ensemble learning is a method that can combine prediction results from various architectures or single classification methods into a new, more accurate model [16]. The advantage of ensemble learning is that it can take information from several classification methods and minimize the error rate of a single classification prediction. Several decision-making techniques in ensemble learning are appropriate and efficient, such as bagging, stacking, boosting, voting, and averaging [17]. Voting is a popular decision-making technique in Ensembles [17, 18]. One of the frequently used voting techniques is weighted voting. Weighted voting assigns different weights to classifiers based on specific criteria and chooses classifiers based on these weights [17]. Several studies have used ensemble learning with weighted voting techniques for COVID-19 classification. The study by [19] used a weighted voting technique to apply ensemble learning from DenseNet-201, ResNet-50, and Inception-v3, but the accuracy obtained was below 95% and did not measure precision performance. The study [20] uses a weighted voting technique to apply ensemble learning to five CNN architectures: Inception, ResNet-v2, Inception-v3, DenseNet121, and Xception. Unfortunately, the recall of the study was still below 90%. The study by [21] implemented an ensemble with a weighted majority vote of AlexNet, GoogleNet, and ResNet. The accuracy obtained from this study was excellent, namely 99%, but it did not measure the performance of recall. The study by [22] implemented an ensemble of GoogLeNet, ResNet-18, and DenseNet-121 with weighted voting, but the performance results for accuracy, recall, and F1-score were still below 90%. These studies only aggregated the final results on any single classifier for use in the test data. The ensemble methods from these studies did not work in the training process, so it cannot be known how the performance of the ensemble method was on the training data and validation data. The ensemble methods of these studies did not guarantee whether the weights obtained from the ensemble results were not overfitting.

This study proposes a new ensemble learning method that combines the results of ResNet, EfficientNet, and Inception-v3 called ELREI (Ensemble Learning of ResNet, EfficientNet, and Inception-v3) with weighted voting for lung disease classification based on CXR image. In contrast to the ensemble method in other studies. The ELREI method works on each epoch in the training stage, not on the final results of the training, so the weights of the ELREI results at each epoch on the training data and validation data can be checked on performance results to handle overfitting. The ELREI does not directly carry out weighted voting based on the results of ResNet, EfficientNet, and Inception-v3, but it is carried out using fully convolutional networks (FCN) to help the model learn the pattern of weights on each data and to avoid overfitting training for lung

Received: May 19, 2023. Revised: June 26, 2023.

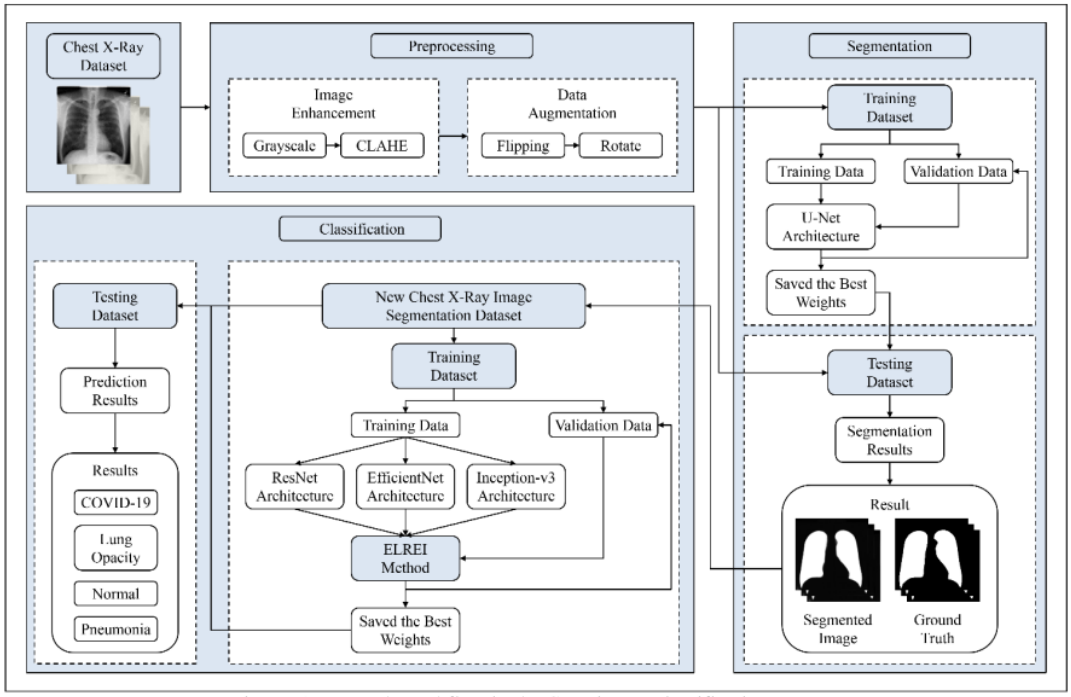


Figure. 1 Research workflow in the CXR image classification process

disease classification based on CXR images. Furthermore, testing the model obtained from the data training process is carried out on test data. The ELREI method with weighted voting is expected to perform COVID-19 classification using lung CXR images precisely and accurately compared to a single classification. The performance evaluation of the proposed method is calculated based on accuracy, recall, precision, and F1-score.

This paper is structured as follows: section 2 describes the proposed method, which consists of data description, image enhancement, image segmentation, image classification using ResNet, EfficientNet, and Inception-v3 architectures, ensemble learning using the ELREI method, and evaluation. Section 3 describes the research results and discussions. Section 4 concludes this paper.

2. Method

The workflow in this study is shown in Fig. 1.

2.1 Data description

The dataset used in this study is the COVID-19 radiography database, obtained from https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database. This dataset is a collection of CXR images from a research project at Qatar University and the University of Dhaka in collaboration with Pakistan and Malaysia. This dataset has four classes: 3616 COVID-19 images, 10192 normal images, 6012 lung opacity (non-COVID) images, and 1345 pneumonia images.

2.2 Image quality enhancement

The image quality enhancement stage removes unimportant parts or noise, increases dark contrast, and maintains all of the image's details. The first step in image quality enhancement is to convert CXR images into grayscale images. A grayscale image is used to simplify the image processing process. Next, increase the image contrast using CLAHE (Contrast limited adaptive histogram equalization). CLAHE is a low-contrast enhancement technique in images that can accentuate the details of an image effectively, especially in medical images [23].

2.3 Data augmentation

The data augmentation used is rotation and flipping. Rotation is done by rotating the image to the right or left according to the selected degree. Flipping is the process of enhancing an image by reversing it in a vertical or horizontal direction [24].

2.4 Image segmentation

Lung segmentation in CXR images can show lung abnormalities, facilitating the disease detection process. There are two processes at this stage: training data and data testing. The augmented CXR pictures are split into training and validation data throughout the training stage. Then training and validation are carried out using the U-Net architecture. The training process will yield the best weight used in the testing phase. Next, testing the model obtained from the data training process was carried out on the test data.

2.5 Image classification

Image classification consists of two processes, including the training and testing stages. The training stage is the process of building a classification model by training the model using predetermined training data. Meanwhile, the testing stage is the process of labeling image data using the classification model obtained from training. The training process uses the segmented image as input, which is divided into training data and testing data. The training data is further divided into training data and validation data. First, single classification methods, namely ResNet, EfficientNet, and Inception-v3, are trained. After that, classification was carried out using the ensemble learning method from the ResNet, EfficientNet, and Inception-v3 (ELREI). The ELREI method works on each epoch in the training stage rather than the final results of each architecture. This helps in handling overfitting by checking the performance results of the ELREI weights at each epoch on both the training and validation data.

2.6 ResNet

A ResNet uses a skip connection in its layer to solve gradient descent problems, which allows it to accelerate deep network convergence [25]. This architecture has lower computational complexity than other architectures, although it has greater depth [7]. Fig. 2 illustrates the ResNet architecture. Fig. 2 shows the CXR input image went through a 7×7 convolution process with stride 2. The convolution layer is the primary layer of CNN that performs convolution operations to study the representation of features from the input image. Stride represents the number of kernel shift steps when performing convolution operations. To calculate the result of the convolution layer, use Eq. (1) [26].

$$C = (A_p * K_q) + b_q \quad (1)$$

Where, $*$ is a convolution operation. To calculate the matrix entry from the $A_p * K_q$ result, use Eq. (2).

$$c_{i,j} = \sum_{u=0}^{n-1}\sum_{v=0}^{n-1}\left(a_{u+i,v+j} \times k_{u+1,v+1}\right) \quad (2)$$

So that,

$$c_{i,j} = \left(\sum_{u=0}^{n-1}\sum_{v=0}^{n-1}\left(a_{u+i,v+j} \times k_{u+1,v+1}\right)\right) + b_q \quad (3)$$

Where, i is the row, j is the column, n is the kernel height, C is the convolution result matrix (feature map) of the q-th kernel at the p-th input, A_p is the p-th input matrix, K_q is the q-th kernel matrix, b_q is the q-th kernel bias, $c_{i,j}$ is the entry matrix of the i-th row entry and j-th column in the matrix C, $a_{u+i,v+j}$ is the $u+i$-th row entry and $v+j$-th column in the A_p matrix, $k_{u+1,v+1}$ is the $u+1$-th row entry and $v+1$-th column in the K_q kernel matrix.

The results of the convolution process are then carried out through the batch normalization process. Batch normalization is an additional layer that helps to normalize the input value in the next layer, reduces the risk of overfitting, and speeds up the training process [27, 28]. Batch normalization calculations are defined mathematically by Eqs. (4), (5), and (6) [29].

$$\hat{c}_{i,j} = \frac{c_{i,j} - \mu_j}{\sqrt{\sigma_j^2 + \varepsilon}} \quad (4)$$

With,

$$\mu_j = \frac{1}{m}\sum_{i=1}^{m} c_{i,j} \quad (5)$$

$$\sigma_j^2 = \frac{1}{m}\sum_{i=1}^{m}\left(c_{i,j} - \mu_j\right)^2 \quad (6)$$

Where, μ_j is the average in each j-th column, m is the number of data in one mini-batch (rows), $c_{i,j}$ is the entry of the convolution result matrix C in the i-th row entry and j-th column, σ_j^2 is the variance in the j-th column hidden layer, $\hat{c}_{i,j}$ is the entry of the convolution result matrix C that has been normalized in the i-th row entry and j-th column, and ε is the smallest positive constant.

The ReLU activation function is then applied to overcome the complexity and speed of the training process. In ReLU, all negative pixel values in the image are 0, while for positive values they will be the same. ReLU calculations are defined mathematically by Eq (7) [29].

Received: May 19, 2023. Revised: June 26, 2023.

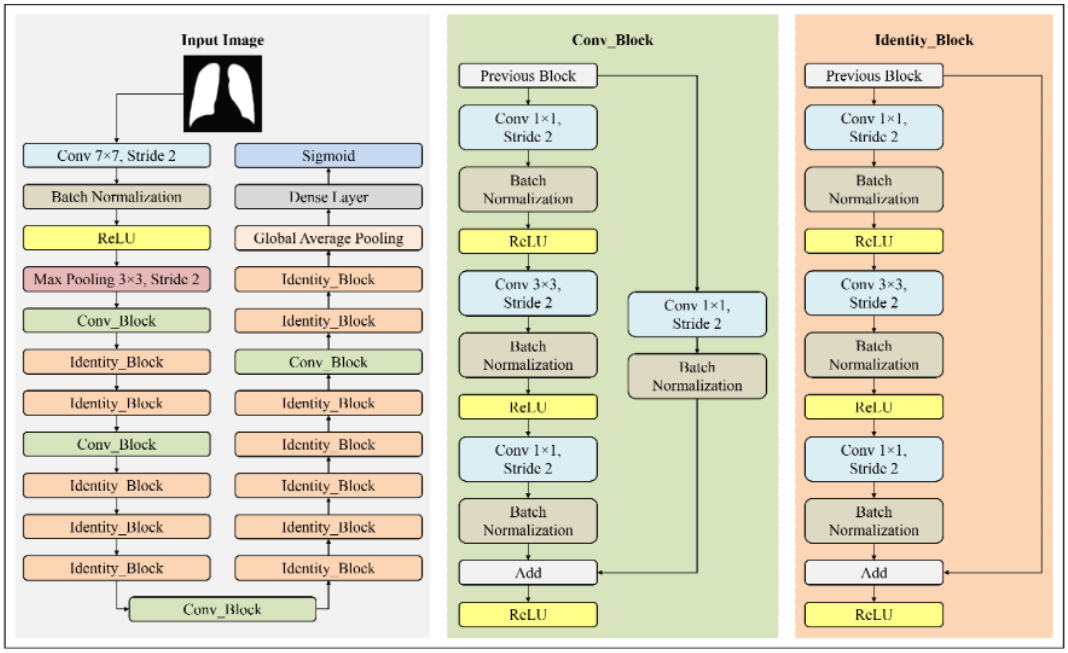


Figure. 2 Illustration of ResNet architecture in the CXR image classification process

$$h_{i,j} = r(\hat{c}_{i,j}) = \max(\hat{c}_{i,j}, 0) = \begin{cases} \hat{c}_{i,j} \; if \; \hat{c}_{i,j} \geq 0 \\ 0 \quad if \; \hat{c}_{i,j} < 0 \end{cases} \quad (7)$$

For, $\hat{c}_{i,j} \in (-\infty, +\infty)$. Where, $h_{i,j}$ is the result of the ReLU activation function and $\hat{c}_{i,j}$ is the entry of the convolution result matrix C that has been normalized in the i-th row entry and j-th column.

The next step in the dimension reduction process involved max polling 3×3 and stride 2. Max pooling selects the maximum value from each submatrix obtained through kernel pooling and stride shifting. The result of the max poll will then pass conv_block and identity_block. In the conv_block section, the input goes through a convolutional layer, followed by batch normalization. This process is repeated three times, with each iteration also applying the ReLU activation function. The input is then passed through the skip connection and processed into the convolution layer and batch normalization. Then, add the results of both processes using the add operation. In identity_block, input is processed into the convolution layer and batch normalization, followed by the ReLU activation function three times. Then, add the results of both processes using the add operation. The result of this process in identity_block is used in the multi-class classification stage. The multi-class classification stage consists of global average pooling, a dense layer, and a sigmoid activation function.

Global average pooling calculates the average value of the entire input image [30]. The dense layer is a layer that serves to classify images into predefined labels after passing through the feature extraction process [31]. The sigmoid activation function is an activation function with the input being a real number and the output being in $(0, 1)$ interval [25]. Sigmoid activation function calculations are defined mathematically by Eq (8) [25].

$$o_{i,j} = \sigma(h_{i,j}) = \frac{1}{1 + e^{-h_{i,j}}} \quad (8)$$

Where, $o_{i,j}$ is the result of the sigmoid activation function with $o \in (0,1)$ and $h_{i,j}$ is the entry of the ReLu activation function matrix in the i-th row entry and j-th column.

2.7 EfficientNet

EfficientNet is a CNN architecture with fewer parameters and improved classification accuracy, introduced by [11]. In their research, the EfficientNet architecture outperformed the number of parameters and accuracy of all previous architectures on the ImageNet dataset. The advantage of the EfficientNet architecture over other architectures is its ability to efficiently scale layer width, image resolution, layer depth, and other factors, resulting in reduced

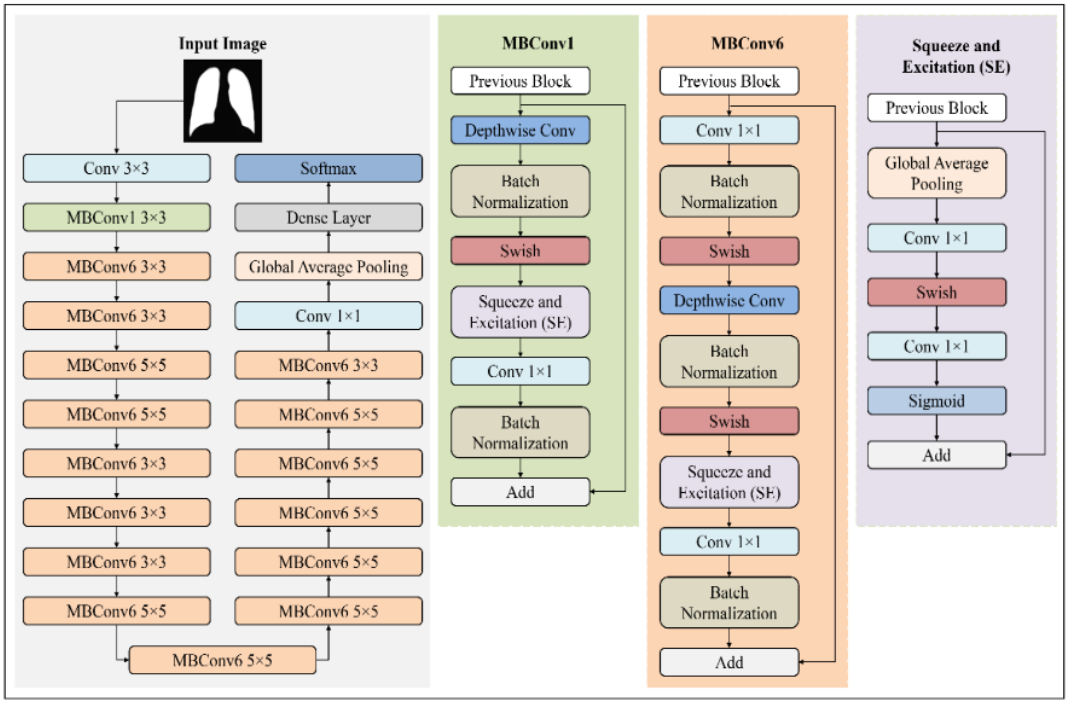


Figure. 3 Illustration of the EfficientNet architecture in the CXR image classification process

parameter sets and more efficient training [12]. Fig. 3 illustrates the EfficientNet architecture. Fig. 3 shows that the input CXR image is processed using a 3 × 3 convolution process. The primary building block of this architecture is the mobile inverted bottleneck convolution (MBConv), which includes two types: MBConv1 and MBConv6. MBConv1 is only used once with a 3×3 kernel, while MBConv6 is used later with different kernel sizes (3×3 and 5×5). MBConv consists of depthwise convolution, batch normalization, swish, squeeze and excitation (SE), convolution, batch normalization again, and add operation.

Depthwise convolution performs independent spatial convolutions on each input channel, allowing independent convolution on each channel [32, 33]. Depthwise convolution calculations are defined mathematically by Eq. (9).

$$D_l = \sum_{u=0}^{n-1} \sum_{v=0}^{n-1} \left(C_{l_{(u+i,v+j)}} \times K_{l_{(u+1,v+1)}} \right) \quad (9)$$

Where, D_l is the l-th depthwise convolution channel matrix, C_l is the l-th channel convolution result matrix, and K_l is the l-th channel kernel matrix.

Swish is a more efficient activation function for processing large data and deeper, complex networks. Swish activation function calculations are defined mathematically by Eq (10) [25, 29].

$$t_{i,j} = s\left(d_{l_{(i,j)}}\right) = d_{l_{(i,j)}} \left(\frac{1}{1 + e^{-\beta\left(d_{l_{(i,j)}}\right)}} \right) \quad (10)$$

Where, t is the output of the activation function of the swish with β being the parameter studied, and $d_{l_{(i,j)}}$ is the entry matrix D_l the i-th row entry and j-th column.

The Softmax is an activation function used in the output layer. Softmax has a function to produce an output in a probability distribution so that the output values are in the $(0, 1)$ interval and the total number is 1 [25]. Softmax is defined by Eq (11).

$$m_{i,j} = p\left(t_{i,j}\right) = \frac{e^{t_{i,j}}}{\sum_{k=1}^{n} e^{t_{i,j_k}}} \quad ; i = i, 2, \ldots, n \quad (11)$$

Where, $m_{i,j}$ is the output of the probability value of the softmax and $t_{i,j}$ is the entry of the Swish activation function matrix in the i-th row entry and j-th column.

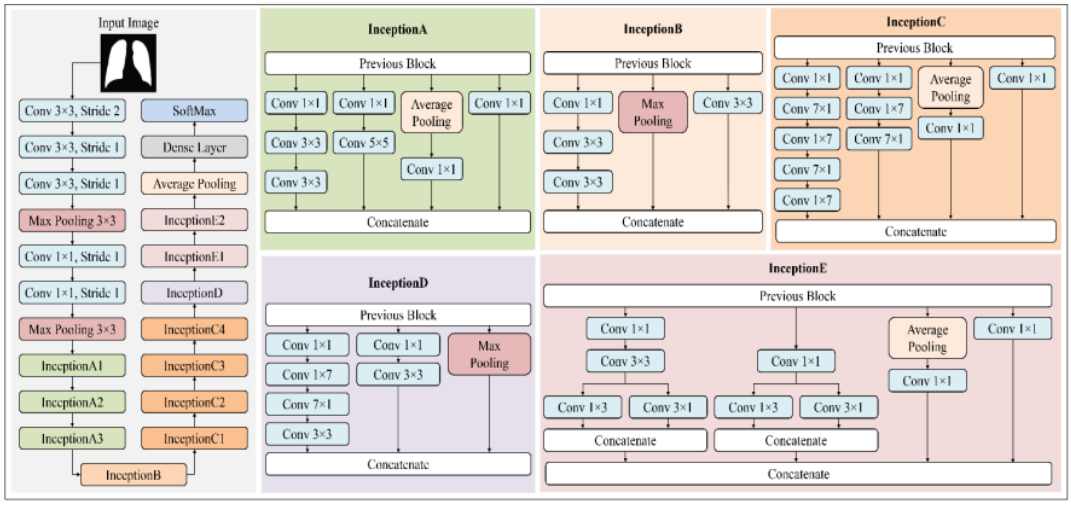


Figure. 4 Illustration of the Inception-v3 architecture in the CXR image classification process

2.8 Inception-v3

The Inception-v3 architecture was first introduced by [34], which uses a basic block, the inception block. This block reduces computational costs without affecting deeper networks or overfitting through stacked 1×1 convolutions. Fig. 4 illustrates the Inception-v3 architecture. Fig. 4 shows that the layers that make up the Inception-v3 architecture include the convolution layer, max pooling, inception block, global average pooling, concatenate, dense layer, and softmax. This architecture consists of 11 inception blocks (3 A blocks, 1 B block, 4 C blocks, 1 D block, and 2 E blocks), each with varying numbers of convolution layers and kernel sizes. The outputs from each path in the inception block are merged using concatenation.

2.9 Ensemble learning

The ensemble learning method used in this study is called ELREI (Ensemble Learning of ResNet, EfficientNet, and Inception-v3). The technique used to make decisions using the ELREI method is weighted voting. To determine the prediction results using the weighted voting technique, use Eq. (12).

$$f_i = \sum_{i=1}^{n} (w_i g_i) \quad (12)$$

Where, f_i is the result of label prediction for f_i, w_i is the weight of training results for the i-th architecture, and g_i is the result of each activation function on the i-th architecture.

The ELREI method works on each epoch in the training stage, not on the final result of each architecture. The weights of the ELREI results at each epoch on the training data and validation data can be checked on performance results to handle overfitting. To help the model learn the pattern of weights on each data and prevent overfitting training for lung disease classification based on CXR images, the ELREI uses Fully Convolutional Networks (FCN) rather than directly performing weighted voting based on the results of ResNet, EfficientNet, and Inception-v3. The results of FCN learning in the ELREI method produce the final weights used to classify lung diseases based on CXR images, consisting of 4 classes: normal, COVID-19, lung opacity, and pneumonia.

2.10 Evaluation

This stage will compare the performance results of a single classification (ResNet, EfficientNet, and Inception-v3) with the performance results of the ELREI method using model performance measures in the form of accuracy, sensitivity, specificity, and F1-score values. The performance results of the model are then compared with those of other studies.

3. Results and discussion

3.1 Image quality enhancement

Image quality enhancement is the initial stage carried out in this research. This stage aims to improve the image quality of the COVID-19 CXR.

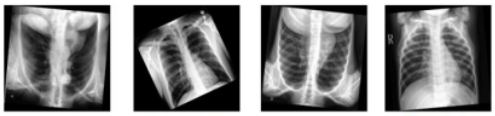


Figure. 5 Results of data augmentation using rotation transformation

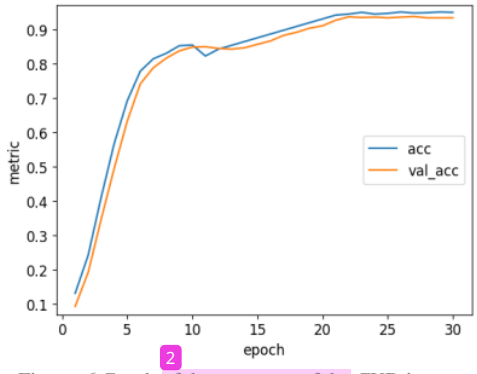


Figure. 6 Graph of the accuracy of the CXR image segmentation training process using the U-Net architecture

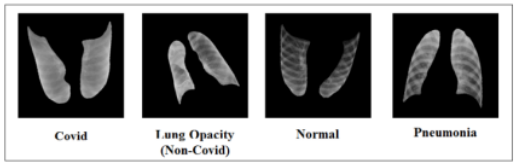


Figure. 7 An example of a lung segmentation image

CXR image is used as the initial input. After that, the image is converted to grayscale to facilitate image processing. Furthermore, the grayscale image is contrast-enhanced using the CLAHE method.

3.2 Data augmentation

In the CXR dataset used, the amount of data in each class is not balanced. Data augmentation is used to multiply data to overcome this problem. The oversampling technique is one of the techniques used to overcome data imbalance by increasing the amount of data in the minority class so that the amount of data is equal to that in the majority class. In this research, two transformations, rotation, and flipping, are applied in data augmentation. Fig. 5 illustrates some examples of CXR images obtained from data augmentation.

Fig. 5 shows a rotation transformation with an angle of 1°-359° applied to the image with the minority class. Starting with the pneumonia class, which originally contained 1345 images, the augmentation method replicated it to 10192 images. The total number of images generated from augmentation is 40768, with 10192 images in each class.

3.3 Image segmentation

Lung image segmentation can aid in the detection of abnormalities in the lungs. The CXR picture will be separated into two portions during lung image segmentation, with the lungs as the foreground and various regions other than the lungs as the background. The method used in this research is the U-Net architecture. The data generated from the segmentation process is separated into two types: training data and test data. The training data is separated again during the training process into training data and validation data. The segmentation training process is to build a model of the proposed architecture. The results of the training process are in the form of weights. Fig. 6 illustrates the segmentation stage accuracy graph during the training process. Fig. 6 shows that the accuracy graph for training data has increased with each epoch. For the first epoch, the accuracy value is 13%, then increases continuously to 95%. While the accuracy graph on validation data also increases with each epoch. The resulting graphs demonstrate that there is no overfitting in the training and validation processes, indicating that U-Net works very well in segmenting the lung CXR image.

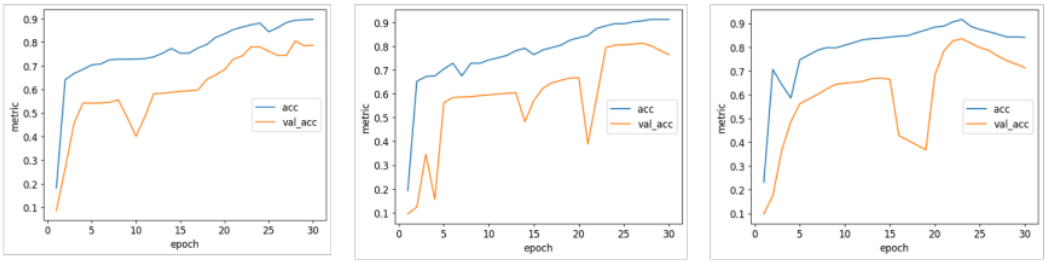
The testing stage is carried out to test the weights generated during training used in the test data. The best weights from the U-Net training process are utilized to test the segmentation of lung CXR pictures with additional data. The results of lung CXR segmentation are separated into two sections, with the lungs as the foreground and various regions other than the lungs as the background. Fig. 7 illustrates some examples of CXR images obtained from the segmentation process.

3.4 CXR image classification

In the image classification stage, the data used is the data from the CXR image segmentation. The data is divided into training data and test data. The training process for training data is divided again into training data and validation data. At this stage, 40768 training data is used, with 90% (36691 training data) and 10% (4077 validation data).

3.4.1. Single classification architecture

At this stage, image classification is performed separately using ResNet, EfficientNet, and Inception-



(a) (b) (c)

Figure. 8 Graph of the accuracy of the CXR image classification training process using the architectures: (a) ResNet, (b) EfficientNet, and (c) Inception-v3

v3 architectures. Fig. 8 illustrates the comparison of accuracy results obtained by each architecture on training data and validation data. The blue line in Fig. 8 indicates the accuracy value for the training data, while the orange line is for the validation data. In Fig. 8 (a), the accuracy graph for training data in the ResNet architecture increases with each epoch. For the first epoch, the accuracy value on the training data is 18%, then increases to 89%. In addition, the accuracy graph for validation data in ResNet architecture has increased and decreased. For the first epoch, it was 8% and increased to 55% in the eighth epoch. In the tenth epoch, the value decreased to 40% and then increased to 78% in the next epoch.

In Fig. 8 (b), the accuracy graph for training data in the EfficientNet architecture has increased at each epoch but has decreased in several epochs. At the first epoch, the accuracy value on the training data is 19%, which then increases to 91%. Meanwhile, the accuracy graph for validation data in the EfficientNet architecture has increased and decreased with each epoch. At the first epoch, it was at a value of 9%, then increased to 34%, and decreased again at the 4th, 14th, and 21st epochs. Subsequently, the accuracy increases and decreases again, reaching 76%.

Fig. 8 (c) illustrates the accuracy graph for training and validation data using the Inception-v3 architecture. Similar to the EfficientNet architecture, the accuracy values fluctuate with each epoch. At the first epoch, the accuracy value on the training data is 23% and continues to increase to 91% at the 23rd epoch. In addition, the accuracy value of validation data increased from 10% to 83%. However, for the next epoch, it decreased to 71%. The accuracy generated by the three architectures experienced overfitting. These architecture models can only detect previously trained data and can't learn patterns in untrained data. Consequently, the performance of the ResNet, EfficientNet, and Inception-v3 architecture models is not good enough to classify lung CXR images.

3.4.2. Ensemble learning uses the ELREI method

ELREI overcomes the limitations and capitalizes on the advantages of different CNN architectures (ResNet, EfficientNet, and Inception-v3) by combining their classification results. ELREI works during the training stage at every epoch. This allows for checking the performance results of the ELREI weights at each epoch on both the training and validation data to handle overfitting. Fig. 9 depicts an illustration of the ELREI method.

Fig. 9 shows that the ELREI method is an ensemble learning method that combines three CNN architectures: ResNet, EffcientNet, and Inception-v3. Each architecture generates a weight, namely weight 1, weight 2, and weight 3. The weights produced in each architecture are then calculated as the softmax value, and voting is continued using the weighted voting technique. After that, a fully connected layer is used at the next stage for the final weight determination and to handle overfitting during training.

Figs. 10 and 11 illustrate the accuracy and loss values generated during the training process using the ELREI method. Fig 10 shows that the accuracy graph for training data in the ELREI method has increased with each epoch. For the first epoch, the accuracy value is 15%, then increases continuously to 94%. While the accuracy graph for validation data in the ELREI method also increases at each epoch. For the first epoch, the accuracy value is 9%, then increases continuously to 92%. For the first epoch, the loss value is 56%, then decreases continuously until it reaches 1%. While the loss graph for validation data in the ELREI method also decreases at each epoch. For the first epoch, the loss value is 51%, then decreases continuously to 0%. The graph obtained shows that the ELREI method does not experience overfitting. ELREI can recognize and learn patterns in trained and untrained data. According to Figs 10 and 11, the model's performance in the ELREI

Received: May 19, 2023. Revised: June 26, 2023.

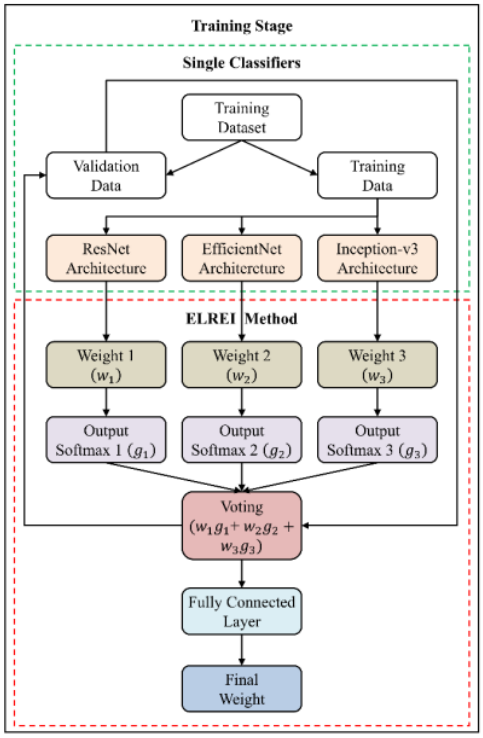


Figure. 9 Illustration of ELREI in the CXR image classification process

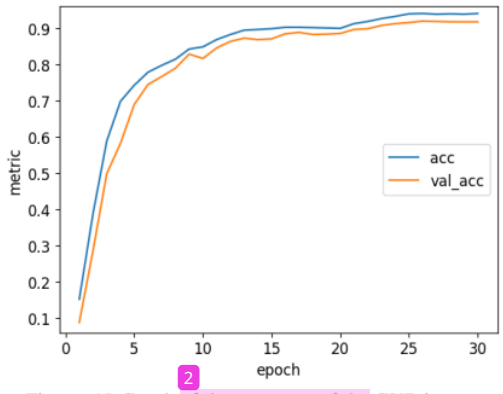


Figure. 10 Graph of the accuracy of the CXR image classification training process using the ELREI method

method is excellent in categorizing lung CXR pictures because the accuracy value is above 92% and the loss value is towards 0%. The testing process will employ the final weight determined during the ELREI method training process.

3.5 CXR image classification

In the testing stage, test data is used to see the

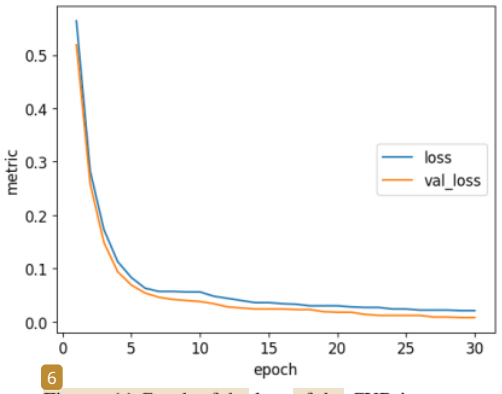


Figure. 11 Graph of the loss of the CXR image classification training process using the ELREI method

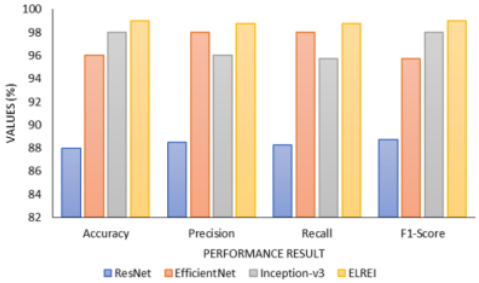


Figure. 12 Comparison of performance evaluation values in the CXR image classification process using the ResNet, EfficientNet, Inception-v3, and ELREI methods

extent of the success of the classification method in predicting lung CXR images. The accuracy, precision, recall, and F1-score values are employed to evaluate the classification method's performance. A comparison of performance evaluation values in a single classification method and the ELREI method is illustrated in Fig. 12. Fig. 12 shows that the performance results using the ELREI method on CXR image classification have improved performance compared to the ResNet, EfficientNet, and Inception-v3. The increase in accuracy value is 11% (ResNet), 3% (EfficientNet), and 1% (Inception-v3). The increase in precision value is 10.5% (ResNet), 1% (EfficientNet), and 3% (Inception-v3). Meanwhile, the increase in recall value is 10.75% (ResNet), 1% (EfficientNet), and 3.25% (Inception-v3). Additionally, the F1-score value increase is 10.25% (ResNet), 3.25% (EfficientNet), and 1% (Inception-v3).

3.6 Evaluation

In this research, the classification of lung CXR images using the ELREI method has been carried out.

Table. 1 Comparison of performance on CXR image classification of the lungs using ELREI with other research

Method	Year	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
EDL-COVID [35]	2021	97.8	97.83	97.77	97.77
DCNN-Based Ensemble [36]	2021	97.47	-	98.18	-
Ensemble on ResNet18 and DenseNet121 [37]	2022	94.1	94.5	94.1	94
CNN-Ensemble [38]	2022	98	98	98.25	97.75
Proposed Method (ELREI)	**2023**	**99**	**98.75**	**98.75**	**99**

The research evaluated the performance of this method and compared it with previous studies. A comparison of the classification results of the CXR image of the lungs in this study with other studies is shown in Table 1. Table 1 is a comparison of research results using the same dataset on CXR images for lung disease classification. The study by [35] and [36] have an accuracy and recall value equal to or more than 95% but did not measure the Precision and F1-score value. Research by [37] achieved the lowest accuracy, precision, recall, and F1-score with values below 95% from Table 1. The proposed method (ELREI) achieved the highest accuracy, precision, recall, and F1-score with values above 98%. The accuracy value of the proposed method demonstrates that it accurately classifies lung diseases in CXR images. The precision and recall values indicate that the model works well in identifying positive cases while minimizing false positives and false negatives. The F1-score value demonstrates a balanced performance between precision and recall. Based on the comparison results in Table 1, the ELREI method for lung CXR image classification has excellent performance. Judging from the performance value of the ELREI method, this method can classify CXR images of the lungs very well.

4. Conclusion

Based on the results and discussion previously presented, the proposed ELREI method is an ensemble learning approach that combines the classification results from ResNet, EfficientNet, and Inception-v3 to overcome these limitations and leverage the advantages of each architecture. ELREI employs a weighted voting technique to make decisions based on the learning outcomes of the three architectures. The method achieves excellent results in terms of accuracy, precision, recall, and F1-score, all above 98%. The training graph of ELREI demonstrates its ability to overcome overfitting. Overall, the ELREI method proves to be robust and excellent for classifying lung diseases based on CXR images, considering four classes: normal, COVID-19, lung opacity, and pneumonia.

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

We certify that all authors contributed to this study. Paper conceptualization and methodology, Lucky Indra Kesuma, Ermatita, and Erwin; software and validation; formal analysis and investigation; resources; data curation and writing-original draft preparation; writing-review editing and visualization; supervision; and project administration.

Acknowledgments

The authors thank the Doctoral Program in Engineering, Faculty of Engineering, Universitas Sriwijaya, for all the support for our research.

References

[1] H. Munusamy, J. M. Karthikeyan, G. Shriram, S. T. Revathi, and S. Aravindkumar, "FractalCovNet architecture for COVID-19 Chest X-ray image Classification and CT-scan image Segmentation", *Biocybern. Biomed. Eng.*, Vol. 41, No. 3, pp. 1025–1038, 2021.

[2] F. Song, N. Shi, F. Shan, Z. Zhang, J. Shen, H. Lu, Y. Jiang, Y. Shi, and S. P. Health, "Emerging Coronavirus 2019-nCoV Pneumonia", *Radiology*, 2019.

[3] I. R. I. Haque and J. Neubert, "Deep learning approaches to biomedical image segmentation", *Informatics Med. Unlocked*, p. 100297, 2020.

[4] A. Desiani, B. Suprihatin, S. Yahdin, A. I. Putri, and F. R. Husein, "Bi - path Architecture of CNN Segmentation and Classification Method for Cervical Cancer Disorders Based on Pap - smear Images", *IAENG Int. J. Comput. Sci.*, Vol. 48, No. 3, 2021.

[5] A. L. Aswathy, A. H. S., and V. C. Vinod, "COVID-19 diagnosis and severity detection from CT-images using transfer learning and back propagation neural network", *J. Infect. Public Health*, Vol. 14, No. 10, pp. 1435–1445, 2021.

[6] D. Sarwinda, R. H. Paradisa, A. Bustamam, and P. Anggia, "Deep Learning in Image Classification using Residual Network (ResNet) Variants for Detection of Colorectal Cancer", *Procedia Comput. Sci.*, Vol. 179, No. 2019, pp. 423–431, 2021.

[7] E. Paladini, E. Vantaggiato, F. Bougourzi, C. Distante, A. Hadid, and A. T. Ahmed, "Two Ensemble-CNN Approaches for Colorectal Cancer Tissue Type Classification", *J. Imaging*, Vol. 7, No. 3, p. 51, 2021.

[8] K. Ali, Z. A. Shaikh, A. A. Khan, and A. A. Laghari, "Multiclass skin cancer classification using EfficientNets – a first step towards preventing skin cancer", *Neurosci. Informatics*, Vol. 2, No. 4, p. 100034, 2022.

[9] A. Keles, M. B. Keles, and A. Keles, "COV19-CNNet and COV19-ResNet: Diagnostic Inference Engines for Early Detection of COVID-19", *Cognit. Comput.*, No. 0123456789, 2021.

[10] Z. Karhan and F. Akal, "Covid-19 Classification Using Deep Learning in Chest X-Ray Images", *TIPTEKNO 2020 - Tip Teknol. Kongresi - 2020 Med. Technol. Congr. TIPTEKNO 2020*, pp. 2021–2024, 2020.

[11] M. Tan and Q. V Le, "EfficientNet : Rethinking Model Scaling for Convolutional Neural Networks", In: *Proc. of the 36th International Conference on Machine Learning*, pp. 6105–6114, 2019.

[12] S. Chaudhary, V. Jakhetiya, B. N. Subudhi, U. Baid, and S. C. Guntuku, "Detecting Covid-19 and Community Acquired Pneumonia Using Chest CT Scan Images with Deep Learning", In: *Proc. of ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8583–8587, 2021.

[13] X. Chen, X. Pu, Z. Chen, L. Li, K. N. Zhao, H. Liu, and H. Zhu, "Application of EfficientNet-B0 and GRU-based deep learning on classifying the colposcopy diagnosis of precancerous cervical lesions", *Cancer Med.*, No. July 2022, pp. 1–10, 2023.

[14] S. Sharma, "COVID-19 prediction from chest X-ray images using deep convolutional neural network", *Artif. Intell. Mach. Learn. EDGE Comput.*, No. March 2020, pp. 315–324, 2022.

[15] M. Mujahid, F. Rustam, R. Álvarez, J. L. V. Mazón, I. D. L. T. Díez, and I. Ashraf, "Pneumonia Classification from X-ray Images with Inception-V3 and Convolutional Neural Network", *Diagnostics*, Vol. 12, No. 5, pp. 1–16, 2022.

[16] J. Gao, C. Leung, and C. Miao, "Diabetic Retinopathy Classification Using an Efficient Convolutional Neural Network", In: *Proc. of 2019 IEEE Int. Conf. Agents, ICA 2019*, pp. 80–85, 2019.

[17] V. C. Osamor and A. F. Okezie, "Enhancing the weighted voting ensemble algorithm for tuberculosis predictive diagnosis", *Sci. Rep.*, Vol. 11, No. 1, pp. 1–12, 2021.

[18] A. Desiani, E. S. Kresnawati, M. Arhami, Y. Resti, N. Eliyati, S. Yahdin, T. J. Charisa, and M. Nawawi, "Majority Voting as Ensemble Classifier for Cervical Cancer Classification", *Sci. Technol. Indones.*, Vol. 8, No. 1, pp. 84–92, 2023.

[19] A. Das, "Adaptive UNet-based Lung Segmentation and Ensemble Learning with CNN-based Deep Features for Automated COVID-19 Diagnosis", *Multimed. Tools Appl.*, Vol. 81, No. 4, pp. 5407–5441, 2022.

[20] S. A. Siddiqui, N. Fatima, and A. Ahmad, "Chest X-ray and CT Scan Classification using Ensemble Learning through Transfer Learning", *EAI Endorsed Trans. Scalable Inf. Syst.*, Vol. 22, No. 6, pp. 1–9, 2022.

[21] T. Zhou, H. Lu, Z. Yang, S. Qiu, B. Huo, and Y. Dong, "The Ensemble Deep Learning Model for Novel COVID-19 on CT Images", *Appl. Soft Comput.*, Vol. 98, p. 106885, 2021.

[22] R. Kundu, R. Das, Z. W. Geem, G. T. Han, and R. Sarkar, "Pneumonia Detection in Chest X-Ray Images using an Ensemble of Deep Learning Models", *PLoS One*, Vol. 16, No. 9 September, 2021.

[23] A. A. Siddiqi, G. B. Narejo, M. Tariq, and A. Hashmi, "Investigation of Histogram Equalization Filter for CT Scan Image Enhancement", *Biomed. Eng. Appl. Basis Commun.*, Vol. 31, No. 5, pp. 1–10, 2019.

[24] C. Shorten and T. M. Khoshgoftaar, "A Survey on Image Data Augmentation for Deep Learning", *J. Big Data*, Vol. 6, No. 60, 2019.

[25] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. A. Dujaili, Y. Duan, O. A. Shamma, J. Santamaría, M. A. Fadhel, M. A. Amidie, and L. Farhan, "Review of Deep Learning : Concepts, CNN Architectures, Challenges, Applications, Future Directions", *J. Big Data*, Vol. 8, No. 53, 2021.

[26] W. Chen, B. Yang, J. Li, and J. Wang, "An Approach to Detecting Diabetic Retinopathy based on Integrated Shallow Convolutional Neural Networks", *IEEE Access*, Vol. 8, pp. 178552–178562, 2020.

[27] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by

Reducing Internal Covariate Shift", In: *Proc. of the 32 nd International Conference on Machine Learning*, France: JMLR: W&CP, 2015.
[28] S. Bharati, P. Podder, M. R. H. Mondal, and V. B. S. Prasath, "CO-ResNet: Optimized ResNet model for COVID-19 diagnosis from X-ray images", *Int. J. Hybrid Intell. Syst.*, Vol. 17, Nos. 1–2, pp. 71–85, 2021.
[29] Y. Tian, "Artificial Intelligence Image Recognition Method Based on Convolutional Neural Network Algorithm", *IEEE Access*, Vol. 8, pp. 125731–125744, 2020.
[30] H. Gholamalinezhad and H. Khosravi, "Pooling Methods in Deep Neural Networks, a Review", *arXiv*, 2020. [Online]. Available: http://arxiv.org/abs/2009.07485
[31] D. R. Nayak, N. Padhy, P. K. Mallick, M. Zymbler, and S. Kumar, "Brain Tumor Classification Using Dense Efficient-Net", *Axioms*, Vol. 11, No. 1, 2022.
[32] Z. Younas and Z. Niu, "CNN with Depthwise Separable Convolutions and Combined Kernels for Rating Prediction", *Expert Syst. Appl.*, Vol. 170, p. 114528, 2021.
[33] R. Stolkin, "Dense Connection and Depthwise Separable Convolution based CNN for Polarimetric SAR Image Classification", *Knowledge-Based Syst.*, Vol. 194, 2020.
[34] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going Deeper with Convolutions", In: *Proc. of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, 2015.
[35] S. Tang, C. Wang, J. Nie, N. Kumar, Y. Zhang, Z. Xiong, and A. Barnawi, "EDL-COVID: Ensemble Deep Learning for COVID-19 Case Detection from Chest X-Ray Images", *IEEE Trans. Ind. Informatics*, Vol. 17, No. 9, pp. 6539–6549, 2021.
[36] P. Shrivastava, A. Singh, S. Agarwal, H. Tekchandani, and S. Verma, "Covid Detection in CT and X-Ray Images using Ensemble Learning", In: *Proc. of 5th International Conference on Computing Methodologies and Communication, ICCMC 2021*, pp. 1085–1090, 2021.
[37] T. K. K. Ho and J. Gwak, "Feature-Level Ensemble Approach for COVID-19 Detection using Chest X-ray Images", *PLoS One*, Vol. 17, No. 7 July, pp. 1–19, 2022.
[38] L. Visuna, D. Yang, J. G. Blas, and J. Carretero, "Computer-Aided Diagnostic for Classifying Chest X-ray Images using Deep Ensemble Learning", *BMC Med. Imaging*, Vol. 22, No. 1, pp. 1–16, 2022.

ELREI: Ensemble Learning of ResNet, EfficientNet, and Inception-v3 for Lung Disease Classification based on Chest X-Ray Image

ORIGINALITY REPORT

12%	9%	12%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to University of Al-Qadisiyah Student Paper	4%
2	Anita Desiani, Erwin, Bambang Suprihatin, Dwiza Riana, Muhammad Arhami, Indri Ramayanti, Yadi Utama. "Denoised Non-Local Means with BDDU-Net Architecture for Robust Retinal Blood Vessel Segmentation", International Journal of Pattern Recognition and Artificial Intelligence, 2024 Publication	2%
3	www.engineeringletters.com Internet Source	1%
4	ph.pollub.pl Internet Source	1%
5	www.researchgate.net Internet Source	1%
6	Anita Desiani, Erwin, Bambang Suprihatin, Filda Efriliyanti, Muhammad Arhami, Emma	1%

Setyaningsih. "VG-DropDNet A Robust Architecture for Blood Vessels Segmentation on Retinal Image", IEEE Access, 2022
Publication

---

7 "Proceedings of the 2nd International Conference on Electronics, Biomedical Engineering, and Health Informatics", Springer Science and Business Media LLC, 2022 — 1 %
Publication

---

8 trepo.tuni.fi — 1 %
Internet Source

---

9 Anita Desiani, Muhammad Adrezo, Nyayu Chika Marselina, Muhammad Arhami, Aulia Salsabila, Muhammad Gibran Al-Filambany. "A Combination of Image Enhancement and U-Net Architecture for Segmentation in Identifying Brain Tumors on CT-SCAN Images", 2022 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS), 2022 — 1 %
Publication

---

10 A F Agustin, H R Albarki, R S H Martin, A Jayanegara. "Identification of fish meal adulterated with rice bran by using an image analysis method", IOP Conference Series: Earth and Environmental Science, 2023 — 1 %
Publication

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		