IEC
Institute of Electronics and Computer

Journal of Artificial Intelligence and Systems, 2022, 4, 107-126
https://iecscience.org/journals/AIS
ISSN Online: 2642-2859

Classification Performance of Deep Transfer Learning Methods for Pneumonia Detection from Chest X-Ray Images

Osman Güler[1], Kemal Polat[2, *]

[1]Tusaş Şehit Hakan Gülşen Vocational and Technical High School, Ankara 06890, Turkey
[2]Department of Electrical and Electronics Engineering, Bolu Abant Izzet Baysal University, Bolu 14280, Turkey
* Corresponding Author: Kemal Polat, Email: kpolat@ibu.edu.tr

How to cite this paper: Osman Güler, Kemal Polat (2022). Neural Network Learning of Context-Dependent Affordances. Journal of Artificial Intelligence and Systems, 4, 107–126.
https://doi.org/10.33969/AIS.2022040107.

Received: September 7, 2022
Accepted: November 18, 2022
Published: December 12, 2022

Abstract

Pneumonia is one of the leading diseases of child mortality in the world. The fastest imaging method for detecting pneumonia in chest X-rays. Examining X-ray images is carried out by expert radiologists. It is important to develop computer-aided diagnosis systems due to the difficulty of the images examined. In this study, DenseNet121, DenseNet169, ResNet50, ResNet101, MobileNetV2, VGG16, Xception and InceptionV3 deep learning models were used to classify chest X-ray images. Experiments were done on the chest X-ray dataset of 5856 labeled images with the proposed models, and the results were compared. Transfer learning models have generally achieved high success rates in the problem of detecting pneumonia from chest X-ray images. The Xception model performed best with 96.16% validation and 95.73% test accuracy. It has been seen that transfer learning models are successful in classification problems.

Keywords

Transfer Learning, Machine Learning, Deep Learning, Chest X-Ray, classification.

1. Introduction

Pneumonia is a respiratory infection that affects the lungs, allowing people to get oxygen from the outside to live a healthy life. Pneumonia causes the lungs to fill with pus, making breathing difficult and reducing oxygen uptake. Pneumonia, a contagious disease, can result in death if not treated in time. According to the World Health Organization (WHO) data, the rate of pneumonia in child deaths was 14% in 2019 and 22% in child deaths aged 1-5 years [1]. Therefore, early diagnosis and

diagnosis of pneumonia and appropriate treatment are very important, especially in order to reduce child mortality and save more children's lives. Lung infections can be detected with computerized medical imaging techniques. The most commonly used medical imaging techniques are chest X-ray, computed tomography (CT), and magnetic resonance imaging (MRI) [2, 3]. Chest X-ray is the most used imaging method due to cost and rapid imaging. Manual examination of the obtained medical images by a radiologist and correct diagnosis is important, but the disadvantage is time cost [4].

They are developing computer-based diagnostic systems and decision support systems that assist expert radiologists in playing an essential role in early diagnosis and treatment. Especially in recent years, intensive studies have been carried out on computerized diagnostic systems supported by artificial intelligence. By using machine learning and deep learning methods, artificial intelligence-supported medical diagnosis and classification studies are carried out in cancer [5], Covid19 [6], diabetes [7], Alzheimer's [8] and all other disease types in health fields.

In recent years, many different studies have been carried out on the classification of chest X-ray images with deep learning methods. Especially in the last few years, it has been focused on the images of the lungs damaged by the covid19 disease. In [3], Covid19, Pneumonia, and Healthy conditions were classified with VGG19 architecture on MongoDB dataset and 97.11% accuracy was obtained. In [9], a 22-layer convolutional neural network (CNN) as feature extractor and Support Vector Machine (SVM), Random Forest (RF) and KNearest Neighbor (KNN) as classifier were used on Mendeley Data v2 dataset. As a result of the study, an accuracy rate of 99.52% with CNN+RF, 96.55% with CNN+SVM and 97.32% with CNN+KNN was obtained. In [10], an accuracy rate of 0.9578 was obtained by using the combination of InceptionResnetV2 and Xception for detection of Covid19. In [11], two ensemble learning backbones, namely DenseNet201, VGG16 and GoogleNet models and A, namely DenseNet201, InceptionResNetV2 and Xception models and B, were combined with the Transformer Encoder mechanism. As a result, group A achieved 97.22% accuracy and group B 96.44% accuracy. In [12], using 7 CNN models, a voting algorithm was developed for Covid19 disease that classifies chest X-ray images as binary, and as a result, it was stated that a diagnostic accuracy rate close to 100% was obtained. In [13], VGG16 and MobileNetV2 models were combined with the fusion method without classifier layers for the classification of Covid-19, Pneumonia and healthy images, resulting in an accuracy of 96.48%. In [14], a multi-scale attention network mechanism was used to classify Covid19 and pneumonia variants, resulting in an accuracy of 97.46%. In [15], feature channel attention block Squeeze and Excitation structure was added to DenseNet model and 92.8% accuracy rate was obtained as a result of the experiment. In [16] machine learning and KNN, SVM and RF classifiers were used and various image processing methods were studied. As a result, accuracy rates between 95% and 99% were obtained. In [17], the optimal hyperparameter set of the ResNet50 and VGG16 architectures was automatically defined with the Genetic Fine Tuning method, and an accuracy rate of

97% was obtained as a result of the experiments. In [18], the number of images was increased with cGAN, and deep transfer learning models such as ResNet50, Xception and DenseNet169 were used. As a result, the proposed model achieved a test accuracy of 93.67%. In [19], data balancing was performed using the smote algorithm, and an accuracy rate of 97.4% for tuberculosis and pneumonia, and 88% for bacterial, viral and Covid19 classifications was obtained with multi-level classification to classify tuberculosis, pneumonia and Covid19. In [20], a structure consisting of a combination of transfer learning approach and capsule network is proposed. Finally, an accuracy of 94.84% was achieved with the InceptionV3 model, which added capsule layers. In [21], a convolutional neural network approach including residual network evaluations and extended convolution is proposed. With the CNN based pre-trained models VGG19, ResNet50 and inceptionv3 architectures, 95.61%, 96.15% and 95.16% accuracy rates were obtained, respectively. In [22], a combination of convolutional neural networks and VGG19 model was proposed and 99.10% accuracy rate was achieved in classifying different chest diseases. In [23], Chest X-ray8 dataset was preprocessed, data augmentation techniques were applied, VGG19 network and Deep Convolutional Generative Adversarial Network (DCGAN) were used for classification. As a result, an accuracy rate of 99.34% was obtained. In [24], long short-term memory (LSTM) deep learning model and SVM, KNN, ensemble classifier and deep learning classifiers were used. In [25], DenseNet121 and ResNet50 models are combined with a layer created with CNN blocks. It was stated that the proposed model was successful in classification.

In this study, transfer learning methods were used to classify chest X-ray images in two classes as healthy and pneumonia. Various image preprocessing techniques were applied on chest X-ray images and classification was made with DenseNet121, DenseNet169, ResNet50, ResNet101, MobileNetV2, VGG16, Xception and InceptionV3 transfer learning models in the keras library. In this study, the performance and accuracy rate of transfer learning models on chest X-ray images were compared.

In the continuation of the study, the methodology of the research, the transfer learning methods, and the data set are explained in the second part. In the third part, the experimental results were given. In Chapter 4, the result of the study is explained.

2. Research Methodology

The first study, which is accepted as the basis of deep learning, is the model called LeNet, which is a gradient-based convolutional neural network used by Lecun et al. in 1998 to classify handwritten numbers from 0 to 9 [26]. Later on, Deep learning started to become popular with the Large Scale Visual Recognition Challenge (ILSVRC) competition held by ImageNet in 2012 and has reached wide audiences with the same name. Deep learning models have become a widely used machine learning method for solving many problems due to their success in the face of artificial intelligence problems. In the widely used deep learning studies, we can reuse

the same model in solving different problems. This method is called transfer learning. Transfer learning is a machine learning method that uses the knowledge and weights of the model that has successfully solved a previous problem and then applies it to the solution of a different problem. The advantages of transfer learning can be listed as it costs less time to learn new tasks, uses less information from experts, and evaluates more situations effectively [27].

In this study, pneumonia was detected from chest X-ray images with transfer learning. In this section, transfer learning architectures, dataset and image preprocessing steps are explained. The study methodology is as shown in Figure 1.

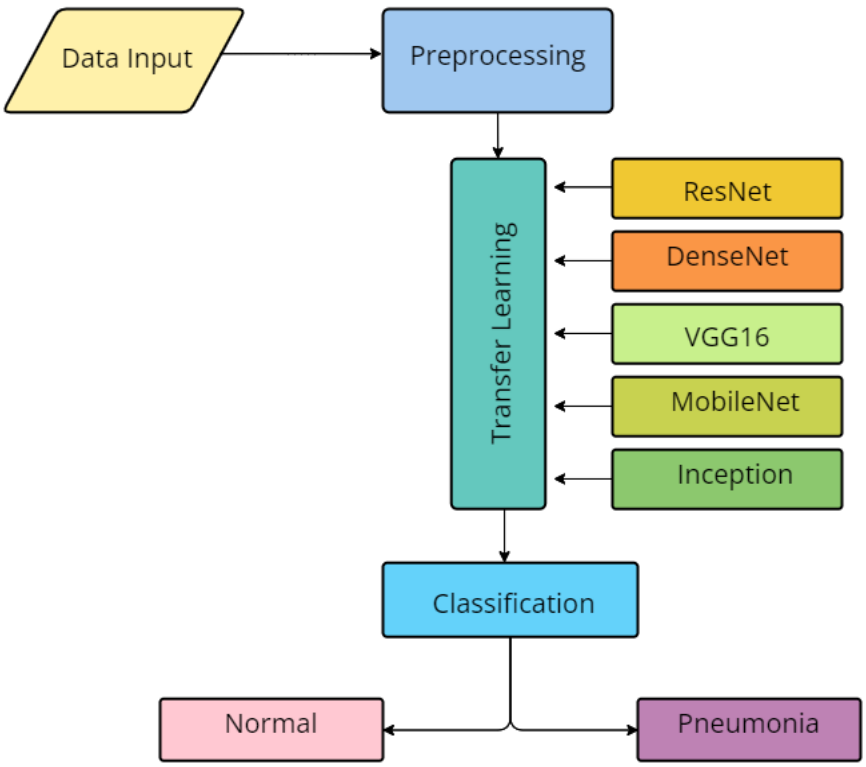

Figure 1. Research Methodology

This section explains ResNet, DenseNet, VGG16, MobileNet and Inception structures used as transfer learning models for pneumonia detection in chest X-ray images.

2.1. ResNet

The ResNet model was developed by the Microsoft research team and won the ILSVRC competition held in 2015 with a success rate of 95.51% [28]. In the ResNet model, the weight values of the previous layer are directly transferred to the next layer. The most important feature of this architecture is that it consists of residual blocks. In the residual blocks, an F(x) result is calculated after applying convolution, ReLU activation function and reconvolution to input x. The first input value of the network, x, is added to the result of F(x) and the output value is found as H(x) = F(x) + x.

2.2. DenseNet

In DenseNet architecture, each layer uses the properties of all previous layers as input values and sends the properties of its own layer as input values to the next layers [29]. L-layer traditional convolutional neural networks have L connections while DenseNet network has L(L+1)/2 direct connections. For each layer, the feature maps of all previous layers are used as input and the feature maps of all subsequent layers are used as inputs.

2.3. VGG16

The VGGNet model is the deep learning model proposed by the Visual Geometry Group (VGG) at Oxford University, which came second in the ILSVRC competition held in 2014 [30]. It has two widely used models, VGG16 and VGG19, one with 16 layers and one with 19 layers. The VGG16 model has 13 convolution layers and 3 fully connected layers, while the VGG19 model has 16 convolution layers and 3 fully connected layers. In the VGGNet model, the maximum pooling method in the pooling layer, ReLU as the activation function, and the softmax function in the classification layer are used.

2.4. MobileNet

MobileNet is modern neural network architecture developed for use in image classification applications in mobile devices and embedded systems, using deeply separable convolutions to form light deep neural networks [31]. The MobileNet structure consists of the first layer, which is a full convolution, and the deeply separable convolution layers.

2.5. Inception

It is the convolutional neural network model that won the ILSVRC competition held in 2014 with a success rate of 93.33% [32]. The convolutional neural network model, which has an initial module called the inception module in its structure, has a depth of 22 layers and consists of 144 layers. With the Inception module, convolution operation is performed with filter sizes of 1×1, 3×3, 5×5 and maximum pooling is performed with the size of 3×3.

InceptionV2 and InceptionV3 models include optimization and more effective use methods to increase efficiency on the Inception model [33]. For more effective use of dimensional reduction and parallel structures of Inception modules, the convolution layers are divided into smaller factors, and the asymmetric convolution layers are factored, thus reducing the number of parameters without reducing their efficiency.

2.6. Dataset

In this study, chest X-ray dataset prepared by Kermany, Zhang, and Goldbaum [34] was used to detect pneumonia. The dataset consists of 5856 labeled images, of which 1583 are normal and 4273 pneumonia. In Figure 2, sample images of healthy and pneumonia patients in the Chest X-Ray data set are given.

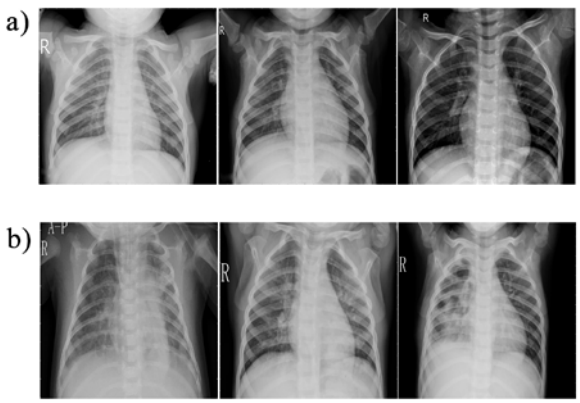

Figure 2. a) normal labeled samples, b) pneumonia labeled samples

2.7. Preprocessing

The images in the chest X-ray data set that we will work on within the scope of the study are of different sizes. Images have been resized to 140x140 pixels in order to make the training faster. Then, 5x5 kernel Gaussian filter was applied to remove the noise in the images.

3. Experimental Results

In this study, transfer learning-based classification was applied using DenseNet121, DenseNet169, ResNet50, ResNet101, MobileNetV2, VGG16, Xception and InceptionV3 models in the keras library to detect pneumonia from Chest X-Ray images. To evaluate the training results with the models, 20% of the dataset was reserved for testing, and the remaining images were divided into 80% training and 20% validation to be used in training. As a result of the trainings, the metric values of measurement accuracy, precision, recall, f1 score, Matthews Correlation Coefficient (MCC) and Balanced Accuracy Score (BAS) were calculated. Equations

for accuracy, precision, recall, f1 score, MCC and BAS metrics are given below. Here; TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives.

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{TP+TN}}{\boldsymbol{TP+TN+FP+FN}} \quad (1)$$

$$\boldsymbol{Precision} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FP}} \quad (2)$$

$$\boldsymbol{Recall\ (Sensivity)} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FN}} \quad (3)$$

$$\boldsymbol{F1\ score} = \boldsymbol{2\ x}\ \frac{\boldsymbol{(PrecisionxRecall)}}{\boldsymbol{(Precision+Recall)}} \quad (4)$$

$$\boldsymbol{Specificity} = \frac{\boldsymbol{TN}}{\boldsymbol{TN+FP}} \quad (5)$$

$$\boldsymbol{BAS} = \frac{\boldsymbol{(Sensitivity + Specificity)}}{\boldsymbol{2}} \quad (6)$$

$$\boldsymbol{MCC} = \frac{\boldsymbol{(TPxTN - FPxFN)}}{\sqrt{\boldsymbol{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}} \quad (7)$$

Table 1 shows accuracy rates for train, validation and test dataset for DenseNet121 model. The model achieved an accuracy of 96.85%, 94.34%, and 94.03% for training, validation, and test data, respectively. Figure 3 shows the accuracy and loss graphics for DenseNet121 model.

Table 1. DenseNet121 accuracy rates

Data	**Accuracy**
Training	96.85%
Validation	94.34%
Test	94.03%

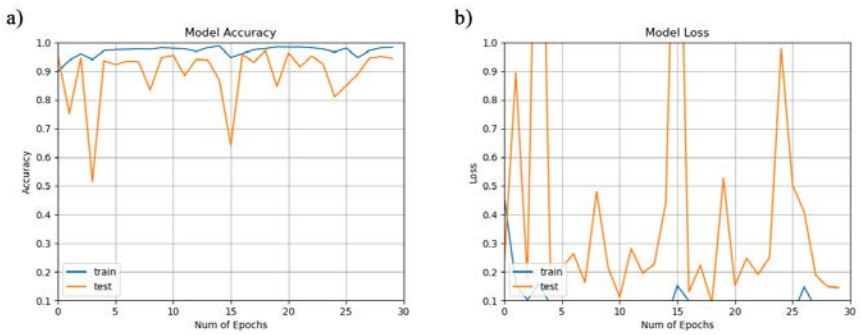

Figure 3. DenseNet121 a) accuracy graphic, b) loss graphic

Table 2 represents the metric for performance measuring. DenseNet121 model achieved precision of 98% and 93% for Normal and Pneumonia classes respectively.

Table 2. DenseNet121 performance metrics

class	precision	recall	f1-score	support
Normal	0.98	0.82	0.89	341
Pneumonia	0.93	0.99	0.96	831

Table 3 shows accuracy rates for train, validation and test dataset for DenseNet169 model. The model achieved an accuracy of 94.77%, 92.74%, and 92.41% for training, validation, and test data, respectively. Figure 4 shows the accuracy and loss graphics for DenseNet169 model.

Table 3. DenseNet169 accuracy rates

Data	Accuracy
Training	94.77%
Validation	92.74%
Test	92.41%

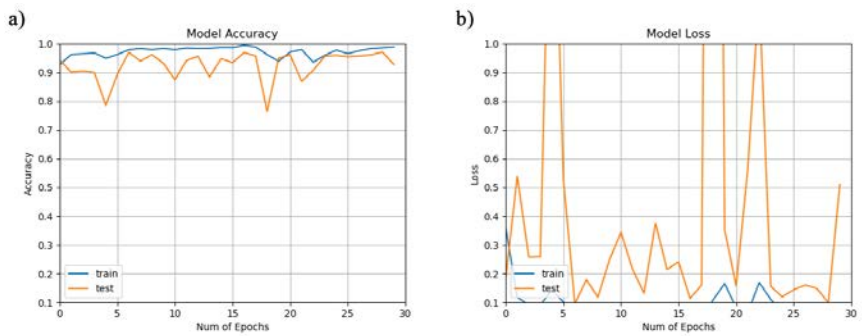

Figure 4. DenseNet169 a) accuracy graphic, b) loss graphic

Table 4 represents the metric for performance measuring. DenseNet169 model achieved precision of 99% and 91% for Normal and Pneumonia classes respectively.

Table 4. DenseNet169 performance metrics

class	precision	recall	f1-score	support
Normal	0.99	0.74	0.85	341
Pneumonia	0.91	1.00	0.95	831

Table 5 shows accuracy rates for train, validation and test dataset for ResNet50 model. The model achieved an accuracy of 88.90%, 86.98%, and 87.37% for training, validation, and test data, respectively. Figure 5 shows the accuracy and loss graphics for ResNet50 model.

Table 5. ResNet50 accuracy rates

Data	Accuracy
Training	88.90%
Validation	86.98%
Test	87.37%

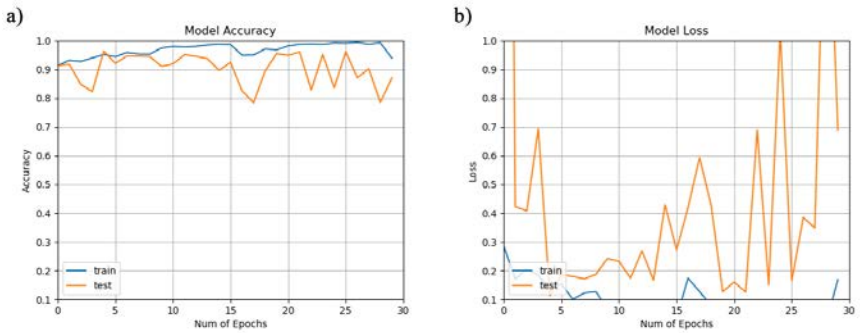

Figure 5. ResNet50 a) accuracy graphic, b) loss graphic

Table 6 represents the metric for performance measuring. ResNet50 model achieved precision of 70% and 99% for Normal and Pneumonia classes respectively.

Table 6. ResNet50 performance metrics

class	precision	recall	f1-score	support
Normal	0.70	0.97	0.82	341
Pneumonia	0.99	0.83	0.90	831

Table 7 shows accuracy rates for train, validation and test dataset for ResNet101 model. The model achieved an accuracy of 95.81%, 93.28%, and 92.75% for training, validation, and test data, respectively. Figure 6 shows the accuracy and loss graphics for ResNet101 model.

Table 7. ResNet101 accuracy rates

Data	Accuracy
Training	95.81%
Validation	93.28%
Test	92.75%

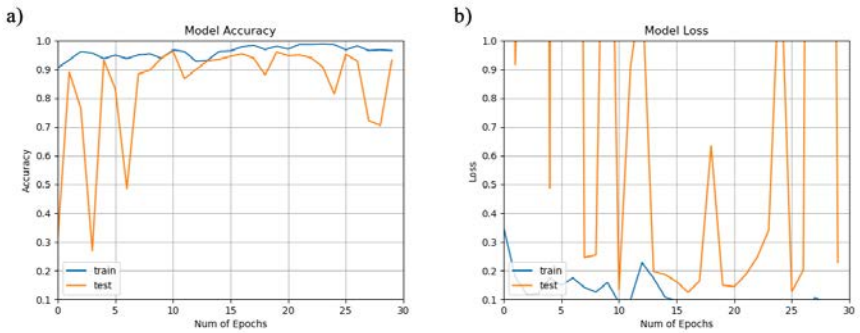

Figure 6. ResNet101 a) accuracy graphic, b) loss graphic

Table 8 represents the metric for performance measuring. ResNet101 model achieved precision of 98% and 91% for Normal and Pneumonia classes respectively.

Table 8. ResNet101performance metrics

class	precision	recall	f1-score	support
Normal	0.98	0.77	0.86	341
Pneumonia	0.91	0.99	0.95	831

Table 9 shows accuracy rates for train, validation and test dataset for MobileNetV2 model. The model achieved an accuracy of 90.61%, 89.33%, and 88.99% for training, validation, and test data, respectively. Figure 7 shows the accuracy and loss graphics for MobileNetV2 model.

Table 9. MobileNetV2 accuracy rates

Data	Accuracy
Training	90.61%
Validation	89.33%
Test	88.99%

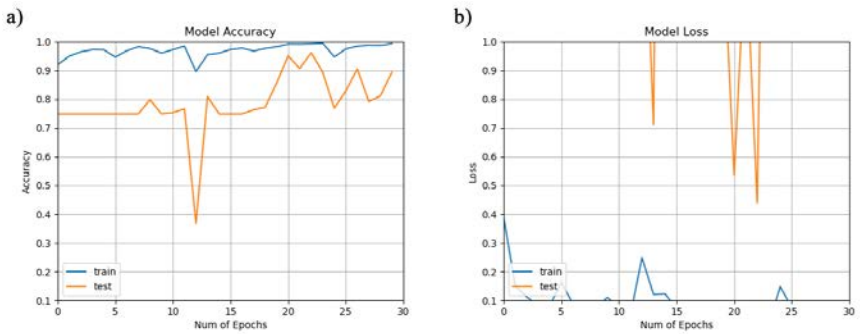

Figure 7. MobileNetV2 a) accuracy graphic, b) loss graphic

Table 10 represents the metric for performance measuring. MobileNetV2 model achieved precision of 73% and 99% for Normal and Pneumonia classes respectively.

Table 10. MobileNetV2 performance metrics

class	precision	recall	f1-score	support
Normal	0.73	0.98	0.84	341
Pneumonia	0.99	0.85	0.92	831

Table 11 shows accuracy rates for train, validation and test dataset for VGG16 model. The model achieved an accuracy of 98.05%, 93.49%, and 94.45% for training, validation, and test data, respectively. Figure 8 shows the accuracy and loss graphics for VGG16 model.

Table 11. VGG16 accuracy rates

Data	Accuracy
Training	98.05%
Validation	93.49%
Test	94.45%

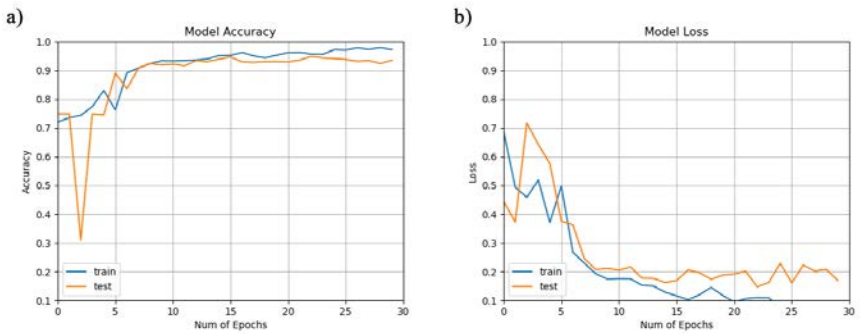

Figure 8. VGG16 a) accuracy graphic, b) loss graphic

Table 12 represents the metric for performance measuring. VGG16 model achieved precision of 90% and 96% for Normal and Pneumonia classes respectively.

Table 12. VGG16 performance metrics

class	precision	recall	f1-score	support
Normal	0.90	0.91	0.91	341
Pneumonia	0.96	0.96	0.96	831

Table 13 shows accuracy rates for train, validation and test dataset for Xception model. The model achieved an accuracy of 99.73%, 96.16%, and 95.73% for training, validation, and test data, respectively. Figure 8 shows the accuracy and loss graphics for Xception model.

Table 13. Xception accuracy rates

Data	Accuracy
Training	99.73%
Validation	96.16%
Test	95.73%

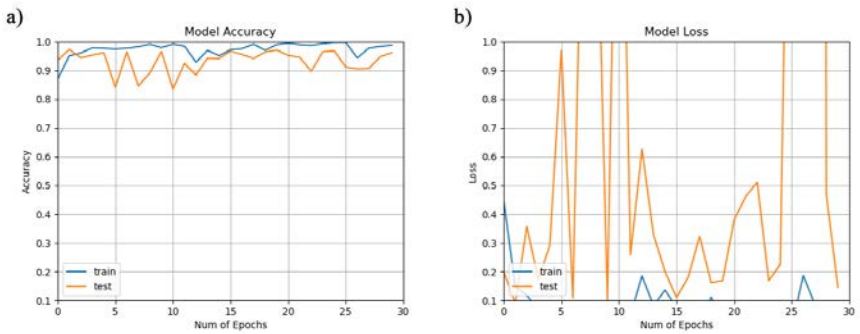

Figure 9. Xception a) accuracy graphic, b) loss graphic

Table 14 represents the metric for performance measuring. Xception model achieved precision of 90% and 96% for Normal and Pneumonia classes respectively.

Table 14. Xception performance metrics

class	precision	recall	f1-score	support
Normal	0.94	0.91	0.93	341
Pneumonia	0.97	0.97	0.97	831

Table 15 shows accuracy rates for train, validation and test dataset for InceptionV3 model. The model achieved an accuracy of 98.08%, 94.13%, and 95.48% for training, validation, and test data, respectively. Figure 10 shows the accuracy and loss graphics for InceptionV3 model.

Table 15. InceptionV3 accuracy rates

Data	Accuracy
Training	98.08%
Validation	94.13%
Test	95.48%

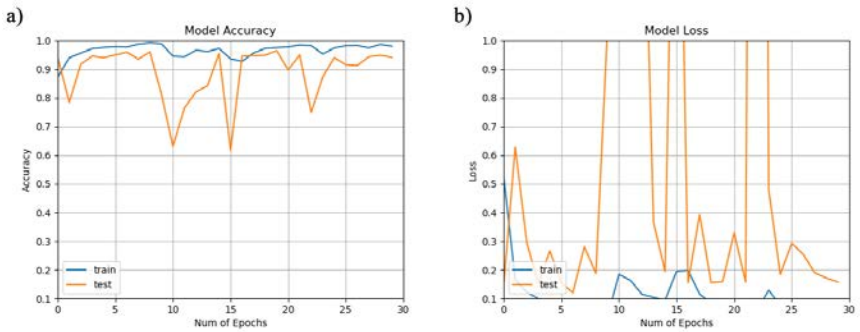

Figure 10. InceptionV3 a) accuracy graphic, b) loss graphic

Table 16 represents the metric for performance measuring. InceptionV3 model achieved precision of 90% and 96% for Normal and Pneumonia classes respectively.

Table 16. InceptionV3performance metrics

class	precision	recall	f1-score	support
Normal	0.94	0.90	0.92	341
Pneumonia	0.96	0.98	0.97	831

Table 17 represents the BAS and MCC metrics for performance measuring. Xception model achieved BAS of 94.48% and MCC of 89.6% respectively.

Table 17. Bas And Mcc Results

Model	BAS	MCC
DenseNet121	90.34 %	85.42 %
DenseNet169	87.12 %	81.62 %
ResNet50	90.32 %	74.7 %
ResNet101	88.06 %	82.31 %
MobileNetV2	91.72 %	77.67 %
VGG16	93.41 %	86.59 %
Xception	94.48 %	89.6 %
InceptionV3	93.87 %	88.94 %

Figure 11 shows the confusion matrices for transfer learning models. For testing purposes, 341 Normal images, 831 Pneumonia images were used.

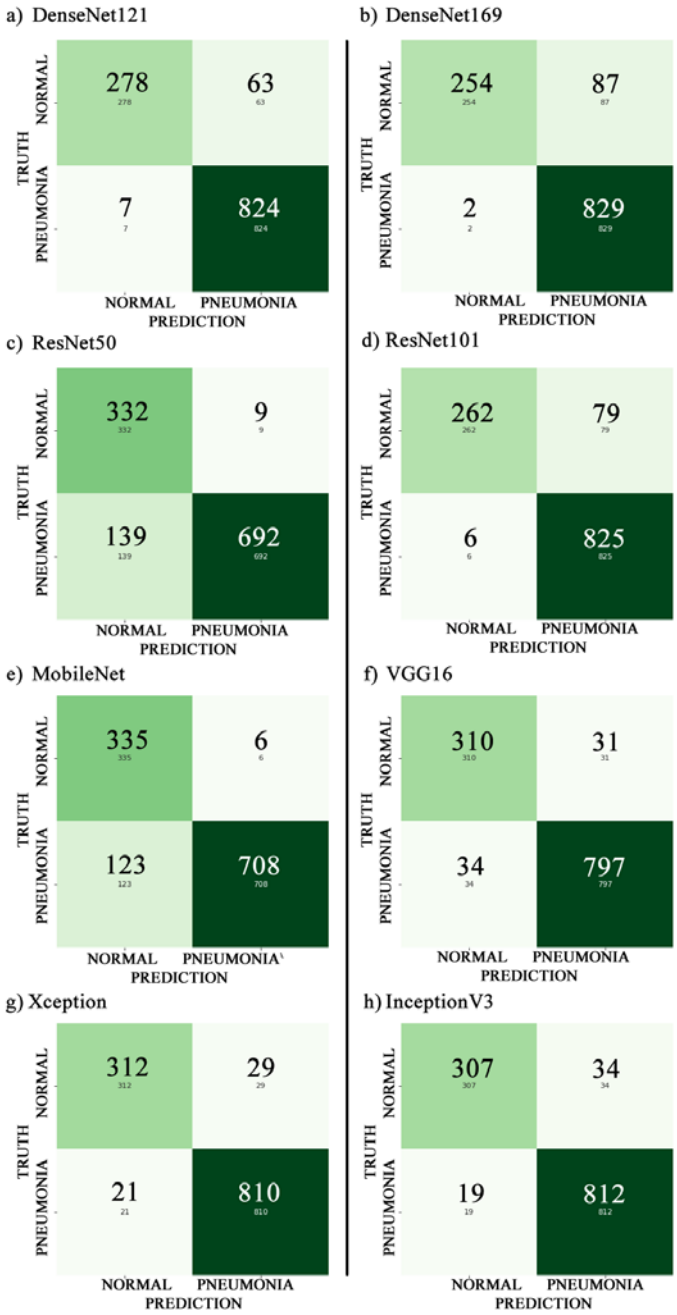

Figure 11. Confusion matrices a) DenseNet121, b) DenseNet169, c) ResNet50, d) ResNet101, e) MobileNetV2, f) VGG16, g) Xception, h) nceptionV3

Table 18 represents a brief summary of test results for transfer learning models for classification of pneumonia detection from Chest X-Ray images.

Table 18. Summary of Test Results

Model	Training Accuracy	Validation Accuracy	Test Accuracy	BAS	MCC
DenseNet121	96.85%	94.34%	94.03%	90.34%	85.42%
DenseNet169	94.77%	92.74%	92.41%	87.12%	81.62%
ResNet50	88.90%	86.98%	87.37%	90.32%	74.70%
ResNet101	95.81%	93.28%	92.75%	88.06%	82.31%
MobileNetV2	90.61%	89.33%	88.99%	91.72%	77.67%
VGG16	98.05%	93.49%	94.45%	93.41%	86.59%
Xception	**99.73%**	**96.16%**	**95.73%**	**94.48%**	**89.60%**
InceptionV3	98.08%	94.13%	95.48%	93.87%	88.94%

4. Discussion and Conclusion

This research aims to detect pneumonia from chest X-ray images with deep transfer learning and to see the classification success of transfer learning models. In this study, DenseNet121, DenseNet169, ResNet50, ResNet101, MobileNetV2, VGG16, Xception and InceptionV3 models in the keras library were used as transfer learning models. As a result of the training, training success, validation and test accuracies, precision, recall and f1-score values for classification, bass and mcc calculations were made to compare the models.

As a result of the training, the DenseNet121 model achieved training, validation, and test accuracy rates of 96.85%, 94.34%, and 94.03%, respectively. The DenseNet169 model achieved training, validation and testing accuracy rates of 94.77%, 92.74%, and 92.41%, respectively. The ResNet50 model achieved training, validation and testing accuracy rates of 88.90%, 86.98%, and 87.37%, respectively. The ResNet101 model achieved training, validation and testing accuracy rates of 95.81%, 93.28%, and 92.75%, respectively. The MobileNetV2 model achieved training, validation and testing accuracy rates of 90.61%, 89.33%, and 88.99%, respectively. The VGG16 model achieved training, validation and testing accuracy rates of 98.05%, 93.49%, and 94.45%, respectively. The Xception model achieved training, validation and testing accuracy rates of 99.73%, 96.16%, and 95.73%, respectively. The InceptionV3 model achieved training, validation and testing accuracy rates of 98.08%, 94.13%, and 95.48%, respectively. When the accuracy values are compared, the first three most successful transfer learning models are Xception, InceptionV3 and VGG16, respectively. In this study, ResNet50 and MobileNetV2 models showed the lowest performance.

To measure the success of classification models, it is not enough to look at the accuracy rates. Therefore, BAS and MCC metrics were also calculated to evaluate the performance of each model. The BAS and MCC values calculated for the

DenseNet121 model are 90.34% and 85.42%, respectively. The BAS and MCC values calculated for the DenseNet169 model are 87.12% and 81.62%, respectively. The BAS and MCC values calculated for the ResNet50 model are 90.32%, and 74.70%, respectively. The calculated BAS and MCC values for the ResNet101 model are 88.06%, and 82.31%, respectively. The BAS and MCC values calculated for the MobileNetV2 model are 91.72%, and 77.67%, respectively. The calculated BAS and MCC values for the VGG16 model are 93.41%, and 86.59%, respectively. The BAS and MCC values calculated for the Xception model are 94.48%, and 89.60%, respectively. The BAS and MCC values calculated for the InceptionV3 model are 93.87% and 88.94%, respectively. According to BAS and MCC calculations, it is seen that the three most successful transfer learning models are Xception, InceptionV3 and VGG16, respectively.

Because our dataset is not evenly distributed, precision, recall and f1-score metrics are also calculated to compare the classification success of the transfer learning models. In DenseNet121 model, precision, recall and f1-score metrics calculated for Normal class are 0.98, 0.82 and 0.89, respectively, precision, recall and f1-score metrics calculated for Pneumoni class are 0.93, 0.99 and 0.96, respectively. Precision, recall and f1-score metrics calculated for Normal class in DenseNet169 model are 0.99, 0.74 and 0.85, respectively, and precision, recall and f1-score metrics calculated for Pneumoni class are 0.91, 1.00 and 0.95, respectively. Precision, recall and f1-score metrics calculated for Normal class in ResNet50 model are 0.70, 0.97 and 0.82, respectively, and precision, recall and f1-score metrics calculated for Pneumoni class are 0.99, 0.83 and 0.90, respectively. Precision, recall and f1-score metrics calculated for Normal class in ResNet101 model are 0.98, 0.77 and 0.86, respectively, and precision, recall and f1-score metrics calculated for Pneumoni class are 0.91, 0.99 and 0.95, respectively. In the MobileNetV2 model, the precision, recall and f1-score metrics calculated for the Normal class are 0.73, 0.98 and 0.84, respectively, and the precision, recall and f1-score metrics calculated for the Pneumoni class are 0.99, 0.85 and 0.92, respectively. Precision, recall and f1-score metrics calculated for Normal class in VGG16 model are 0.90, 0.91 and 0.91, respectively, precision, recall and f1-score metrics calculated for Pneumoni class are 0.96, 0.96 and 0.96, respectively. Precision, recall and f1-score metrics calculated for Normal class in Xception model are 0.94, 0.91 and 0.93, respectively, and precision, recall and f1-score metrics calculated for Pneumoni class are 0.97, 0.97 and 0.97, respectively. Precision, recall and f1-score metrics calculated for Normal class in InceptionV3 model are 0.94, .090 and 0.92, respectively, while precision, recall and f1-score metrics calculated for Pneumoni class are 0.96, .098 and 0.97, respectively. According to the precision, recall and f1-score metrics obtained on the classes, it can be said that the Xception, InceptionV3, and VGG16 models have a stable classification over 0.90 in both classes.

In our experiments to compare the performance of transfer learning models in chest X-ray dataset classification, it was seen that the Xception model was the

most successful model. The InceptionV3 model follows it in second place. Experimental results have shown us that models with inception structure perform well in training, classification, and recall on an unbalanced dataset. Although inception architecture achieves the highest performance, it is seen that other transfer learning models also achieve successful results. This shows us that the success of transfer learning in classification problems is undeniable. Therefore, We can say that his proven models can achieve high performance in different problems.

With transfer learning, we can achieve successful results on different data sets. Deep transfer learning models with different structures can be used alone or in combination with various methods to solve various problems. Future studies can be conducted on the performance of a deep learning architecture consisting of a combination of different transfer learning models in classification problems, especially in medical imaging.

References

[1] World Health Organization, Pneumonia [Fact Sheet], 2021. https://www.who.int/news-room/fact-sheets/detail/pneumonia

[2] Liang, G., & Zheng, L. (2020). A transfer learning method with deep residual network for pediatric pneumonia diagnosis. *Computer methods and programs in biomedicine*, 187, 104964.

[3] Chakraborty, S., Paul, S., & Hasan, K. M. (2022). A transfer learning-based approach with deep cnn for covid-19-and pneumonia-affected chest x-ray image classification. *SN Computer Science*, 3(1), 1-10.

[4] Moujahid, H., Cherradi, B., El Gannour, O., Bahatti, L., Terrada, O., & Hamida, S. (2020). Convolutional neural network based classification of patients with pneumonia using X-ray lung images. *Transfer*, 2(99.41), 16.

[5] Dabass, M., Vashisth, S., & Vig, R. A convolution neural network with multi-level convolutional and attention learning for classification of cancer grades and tissue structures in colon histopathological images. *Computers in biology and medicine*, 147, 105680.

[6] Alafif, T., Tehame, A. M., Bajaba, S., Barnawi, A., & Zia, S. (2021). Machine and deep learning towards COVID-19 diagnosis and treatment: survey, challenges, and future directions. *International journal of environmental research and public health*, 18(3), 1117.

[7] Li, T., Gao, Y., Wang, K., Guo, S., Liu, H., & Kang, H. (2019). Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening. *Information Sciences*, 501, 511-522.

[8] Savaş, S. (2022). Detecting the stages of Alzheimer's disease with pre-trained deep learning architectures. *Arabian Journal for Science and Engineering*, 47(2), 2201-2218.

[9] Sourab, S. Y., & Kabir, M. A. (2022). A comparison of hybrid deep learning models for pneumonia diagnosis from chest radiograms. *Sensors International*, 3, 100167.

[10] Gayathri, J. L., Abraham, B., Sujarani, M. S., & Nair, M. S. (2022). A computer-aided diagnosis system for the classification of COVID-19 and non-COVID-19 pneumonia on chest X-ray images by integrating CNN with sparse autoencoder and feed forward neural network. *Computers in Biology and Medicine*, 141, 105134.

[11] Ukwuoma, C. C., Qin, Z., Heyat, M. B. B., Akhtar, F., Bamisile, O., Muaad, A. Y., Addo, D., & Al-Antari, M. A. (2022). A hybrid explainable ensemble transformer encoder for pneumonia identification from chest X-ray images. *Journal of Advanced Research*.

[12] Li, D., & Li, S. (2022). An artificial intelligence deep learning platform achieves high diagnostic accuracy for Covid-19 pneumonia by reading chest X-ray images. *Iscience*, 25(4), 104031.

[13] Sharma, A., Singh, K., & Koundal, D. (2022). A novel fusion based convolutional neural network approach for classification of COVID-19 from chest X-ray images. *Biomedical Signal Processing and Control*, 77, 103778.

[14] Wong, P. K., Yan, T., Wang, H., Chan, I. N., Wang, J., Li, Y., Ren, H., & Wong, C. H. (2022). Automatic detection of multiple types of pneumonia: Open dataset and a multi-scale attention network. *Biomedical Signal Processing and Control*, 73, 103415.

[15] Wang, K., Jiang, P., Meng, J., & Jiang, X. (2022). Attention-Based DenseNet for Pneumonia Classification. *IRBM*, 43(5), 479-485.

[16] Ortiz-Toro, C., García-Pedrero, A., Lillo-Saavedra, M., & Gonzalo-Martín, C. (2022). Automatic detection of pneumonia in chest X-ray images using textural features. *Computers in Biology and Medicine*, 145, 105466.

[17] Vieira, P. A., Magalhães, D. M., Carvalho-Filho, A. O., Veras, R. M., Rabêlo, R. A., & Silva, R. R. (2021). Classification of COVID-19 in X-ray images with Genetic Fine-tuning. *Computers & Electrical Engineering*, 96, 107467.

[18] Mehta, T., & Mehendale, N. (2021). Classification of X-ray images into COVID-19, pneumonia, and TB using cGAN and fine-tuned deep transfer learning models. *Research on Biomedical Engineering*, 37(4), 803-813.

[19] Venkataramana, L., Prasad, D., Saraswathi, S., Mithumary, C. M., Karthikeyan, R., & Monika, N. (2022). Classification of COVID-19 from tuberculosis and pneumonia using deep learning techniques. Medical & *Biological Engineering & Computing*, 60(9), 2681-2691.

[20] Bodapati, J. D., & Rohith, V. N. (2022). ChxCapsNet: Deep capsule network with transfer learning for evaluating pneumonia in paediatric chest radiographs. *Measurement*, 188, 110491.

[21] Malik, H., Anees, T., Din, M., & Naeem, A. (2022). CDC_Net: multi-classification convolutional neural network model for detection of COVID-19, pneumothorax, pneumonia, lung Cancer, and tuberculosis using chest X-rays. *Multimedia Tools and Applications*, 1-26.

[22] Malik, H., & Anees, T. (2022). BDCNet: multi-classification convolutional neural network model for classification of COVID-19, pneumonia, and lung cancer from chest radiographs. *Multimedia Systems*, 28(3), 815-829.

[23] Rajasenbagam, T., Jeyanthi, S., & Pandian, J. A. (2021). Detection of pneumonia infection in lungs from chest X-ray images using deep convolutional neural network and content-based image retrieval techniques. *Journal of Ambient Intelligence and Humanized Computing*, 1-8.

[24] Goyal, S., & Singh, R. (2021). Detection and classification of lung diseases for pneumonia and Covid-19 using machine and deep learning techniques. *Journal of Ambient Intelligence and Humanized Computing*, 1-21.

[25] Mamalakis, M., Swift, A. J., Vorselaars, B., Ray, S., Weeks, S., Ding, W., Clayton, R.H., Mackenzie,L.S., & Banerjee, A. (2021). DenResCov-19: A deep transfer learning network for robust automatic classification of COVID-19, pneumonia, and tuberculosis from X-rays. *Computerized Medical Imaging and Graphics*, 94, 102008.

[26] Y. LeCun, L. Bottou, Y. Bengio ve P. Haffner, "Gradient-based learning applied to document recognition", *Proceedings of the IEEE*, c. 86, sayı 11, ss. 2278-2324, 1998.

[27] Cook, D., Feuz, K. D., & Krishnan, N. C. (2013). Transfer learning for activity recognition: A survey. *Knowledge and information systems*, 36(3), 537-556.

[28] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[29] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).

[30] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv*:1409.1556.

[31] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv*:1704.04861

[32] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vonhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).

[33] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016).

Rethinking the inception architecture for computer vision. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818-2826).

[34] Kermany, D., Zhang, K., Goldbaum, M. (2018), "Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification", *Mendeley Data*, V2, doi: 10.17632/rscbjbr9sj.2