

Ensemble-Based Modelling for Enhanced Detection of Pneumonia Disease

Mustafa Oguzhan Ozdemir

Department of Computer Engineering, Kastamonu University, Kastamonu, 37140, Turkey

E-mail: moguzhan578@gmail.com

ORCID iD: <https://orcid.org/0000-0003-0466-7278>

Kemal Akyol*

Department of Computer Engineering, Kastamonu University, Kastamonu, 37140, Turkey

E-mail: kakyol@kastamonu.edu.tr

ORCID iD: <https://orcid.org/0000-0002-2272-5243>

*Corresponding Author

Received: March 24, 2026; Revised: April 12, 2026; Accepted: June 8, 2026; Published: August 8, 2026

Abstract: Pneumonia is a lung condition that is rather prevalent and has the potential to be lethal. The early diagnosis of this disease is absolutely necessary to cut down on the number of fatalities. The aim of this research is to provide a decision-support tool that can help professionals in the field identify cases of pneumonia. The experimental studies used two publicly available datasets in the Kaggle repository. First, experiments were conducted with the pre-trained models. Then, hard and soft voting ensemble learning approaches were implemented using the five most successful deep learning models. According to the results, the soft voting approach outperformed others, with accuracies of 98.55% and 97.26% in two- and three-class datasets, respectively. With this result, a software included this approach has been developed to assist field experts in their decision-making.

Index Terms: Pneumonia, X-ray imaging, deep learning, ensemble learning, soft voting

1. Introduction

Pneumonia which results in lung ulcers [1] is an acute respiratory disease infection caused by viruses and bacteria [2,3]. This disease is one of the leading causes of death worldwide for children [1]. Lung infection causes damage to the pulmonary alveoli, which are tiny, balloon-shaped sacs found at the ends of the bronchioles [3]. Symptoms such as dry cough and chest burning are signs of a weakened immune system in smokers and stroke survivors [4]. Pneumonia's effect on humans can range from mildly uncomfortable to life-threatening. Children under 14, older people over 65, and people with lung problems or a weakened immune system are most at risk [5]. Radiologists manually performed visual inspections. The entire procedure is time-consuming and requires the presence of professional radiologists at all times [6]. Computer vision-based tools can provide decision support to field experts, enabling them to make the right decisions early diagnosis and intervention in health. An early diagnosis of pneumonia is crucial for ensuring effective treatment and improving survival chances. The automatic pneumonia diagnosis presents a considerable difficulty, as the symptoms of the condition are sometimes mild and arduous to detect from X-ray scans [7]. This study addresses this issue by presenting a voting approach that comprises several cutting-edge deep learning networks that detect pneumonia disease with high accuracy. In this context, this paper presents a comparison of the performances of hard voting and soft voting approaches. Furthermore, this study presents computer-aided software to assist field experts in making decisions. The identification of pneumonia at an early stage is absolutely necessary in order to stop the course of the disease. It also leads to a reduction in the number of patients and deaths. In light of this, the purpose of this work is to undertake extensive experiments that evaluate the effectiveness of pre-trained deep-learning models on X-ray pictures for the purpose of developing pneumonia detection methods. The main contribution of this study is to propose an ensemble-based modeling for enhanced detection of pneumonia disease with high performance and to develop a user-friendly, helpful application for field experts. In this context, the contributions of the paper are as follows:

- Ensemble learning-based modelling was introduced for detecting pneumonia disease.
- Comprehensive experiments confirm the proposed model's effectiveness.

- The proposed model employs a soft voting approach, which includes cutting-edge deep learning architectures, to achieve high accuracy.
- The proposed model offers high accuracies on the public two datasets and has potential as a valuable artificial intelligence tool in the medical field.

2. Related Works

There is a great deal of research that has been published on the subject of pneumonia detection. It has been observed that these studies are mainly based on two or three-class datasets. There is a clear indication that the COVID-19 class is also included in the research that was carried out using three-class datasets, despite the fact that the studies that were carried out with two-class datasets included picture samples that did not contain any pneumonia or illness results. Hu et al. proposed the DenseNet-121 network for pneumonia detection on a two-class dataset containing images with and without pneumonia. They discussed the importance of further research in the medical field to improve classification performance. The authors reported a 96.2% accuracy and 94.3% F1-score with the DenseNet-121 model [8]. Stephen et al. conducted experiments using pre-trained models to detect the presence or absence of pneumonia. The authors reported that data multiplexing operations such as rescaling, horizontal translation, width shift, height shift, and zooming positively affected the success of these models in improving classification performance [9]. Rahman et al. aimed to use deep learning methods to detect pneumonia disease using X-ray images in three two- and three-class datasets, including images of bacterial pneumonia, pneumonia patients, and patients with no signs of disease, in their study. The authors worked with AlexNet, ResNet-18, DenseNet-201, and SqueezeNet, pre-trained convolutional neural network models after the pre-processing datasets. The authors reported that the DenseNet-201 model provided the highest accuracy for each dataset in their experimental studies, and they achieved an average accuracy of 95% with this model [10]. Jain et al. employed convolutional neural networks on X-rays of patients with and without signs of pneumonia to detect pneumonia. The authors trained two models of their own design and commonly used convolutional neural network models, using different parameters and increasing the number of convolutional layers. The authors achieved an accuracy of 92.31% with the second model they developed [11]. Er employed several pre-trained models to extract feature maps from X-ray images and composed hybrid feature maps by combining the features and obtained 98.32% accuracy classification [12]. Ben Atitallah et al. applied the ensemble learning method, including VGG-16, Xception, Inception-V3, ResNet-50, and DenseNet-201 models, to a two-class dataset containing images of people with and without signs of pneumonia for pneumonia disease detection and achieved 97.3% accuracy [13]. In their study, Rai et al. applied a technique to validate the model performance based on the exclusion method on a dataset containing two classes. The authors obtained 87.00% accuracy with the ensemble learning method, including Random Forest, Extreme Gradient Boosting, Extra Decision Tree, and Gradient Boosting [14].

Since the COVID-19 outbreak that occurred in November 2019, it can be seen that many studies have been conducted to detect COVID-19 patients and pneumonia diseases with high success. Although the relevant studies focus on COVID-19 disease, the findings of pneumonia detection are also included in these studies. Some of these studies are listed below. Budak et al. analyzed several pretrained models on a three-class chest X-ray dataset. The authors reported that they achieved 98% accuracy with the DenseNet-121 and DenseNet-201 models [15]. Srivastava et al. combined the three-class 'Covid-19 radiography database' dataset, the two-class 'Sars-Cov-2 CT scan' dataset and the two-class 'Pneumonia chest X-ray' dataset. The authors achieved 98.18% sensitivity, 96.6% precision, 97.47% accuracy, and 95.36% F1 score results with the ensemble learning approach, including Inception-V4, ResNet-50, and EfficientNet-B0 deep learning models on the combined dataset [16]. Jabra et al. reported that 99.31% accuracy with ensemble learning on a three-class dataset including the several models [17]. Lanjewar et al. introduced an efficient forecasting framework based on Convolutional Neural Networks for COVID-19 and reported that the model they developed achieved 99.11% [18]. In their study, Kaur et al. applied image enhancement and data augmentation techniques to tomography images of individuals diagnosed with COVID-19, those diagnosed with pneumonia, and those with no disease findings in a three-class dataset. The authors also fine-tuned the parameters in their model and reported that the pre-processing they performed increased the accuracy of their original model from 94% to 98% [19]. Hashmi et al. used ResNet-18, DenseNet-121, Inception-V3, Xception, and MobileNet-V2 models on a two-class 'Chest X-Ray Images' dataset containing images of patients with pneumonia and images of patients without signs of the disease for pneumonia disease detection and reported that they achieved the best result with the DenseNet-121 model with 98% accuracy [20]. El Asnaoui employed binary and ternary model combinations with the ensemble learning method to the two-class 'Chest X-Ray Images' dataset, which includes images of patients with pneumonia and those with no signs of disease. The author achieved 95.17% accuracy with the combination of ResNet-50 and MobileNet-V2 [21]. Narin et al. merged three datasets containing images of COVID-19 patients, bacterial pneumonia patients, pneumonia patients, and patients with no evidence of disease. He analyzed the performance of the Inception-V3, ResNet-50, ResNet-101, ResNet-152, and ResNet-V2 models using 5-fold cross-validation on this newly created four-class dataset. The author reported that the ResNet-50 model was the most successful model with an accuracy of 99.70% [22]. Mujahid et al. used VGG-16 & original model, Inception-V3 & original model, and ResNet-50 & original model pairs for ensemble learning on a two-class dataset containing images of pneumonia patients and those without signs of disease for pneumonia detection and reported that Inception-V3 & original model pair gave the most

successful result with 99.29% accuracy under the exclusion validation technique [23]. Aldhahi and Sull reported that 99% value of average accuracy using 3-fold cross-validation on the three-class Sars-Cov-2 dataset using ensemble learning [24]. Zhang et al. proposed a new neural network model named CN2A-CapsNet on a three-class dataset. The authors achieved 96.71% accuracy with their model proposed using a 6-fold cross-validation method [25].

3. Material and Method

3.1. Datasets

Two datasets, which are publicly available in the Kaggle data repository, are used for hold-out experiments. The current datasets are referred to as Dataset-1 and Dataset-2 in the remainder of this study. Accordingly, 1172 images were allocated as test data, 3747 as training data, and 937 as validation data for hold-out experiments on Dataset-1. In addition, 1388 images were allocated as test data, 4440 as training data, and 1110 as validation data for Dataset-2. Table 1 summarizes the datasets.

Table 1. Statistical information of the datasets

	Class		Number of samples		
			Training	Validation	Test
Dataset-1	Pneumonia		2734	684	855
	No-finding		1013	253	317
		Total	3747	937	1172
Dataset-2	Pneumonia		1480	370	462
	No-finding		1480	370	463
	COVID-19		1480	370	463
		Total	4440	1110	1388

3.2. Deep learning

Convolutional Neural Networks (CNNs), which have hidden layers and neurons, have parameters that can be learned, such as weights, biases, etc. Computer vision is developing rapidly every day with deep learning. CNNs designed based on the visual systems of animals are widely used in computer vision studies. These networks incorporate the most widely used deep learning algorithms in computer vision and image processing works. The architecture of convolutional neural networks is designed and built to learn the distinctive features of images in a self-learning and adaptive manner.

3.3. Transfer learning

Transfer learning is reusing a previously trained model on a different problem dataset. In this approach, a model uses the knowledge gained from a previous task to improve its generalization to another task [26]. After the target task's output layer has been redesigned, the fully connected layers in the classification part are retrained [27].

3.4. Ensemble learning

Ensemble learning uses a strategy that involves multiple models to assign the final class label to an instance. The voting approach, an ensemble learning technique, uses the results of multiple base classifiers for the final class prediction. First, the base classifiers are trained separately on the training set. After that, the final prediction of the voting function is carried out by the process. When it comes to hard voting, the class label of an instance is established by using the majority voting method. On the other hand, when it comes to soft voting, the final class label is chosen by taking the maximum average prediction of the classifiers into consideration. In other words, the class having the highest probability of predicting the final class label is regarded as the final predicted class in soft voting. In contrast, in hard voting, the majority voting is considered [28].

3.5. Performance evaluation

The hold-out method divides the data set into training and test sets. The part allocated for training is used in the training phase of the designed model. The part allocated for the testing is used to validate whether the model performs well on data other than the training set. In the literature, the most common separation ratio when using the hold-out method is 80% of the data set for the training phase and the remaining 20% for the test phase [29]. In machine learning studies, precision, sensitivity, accuracy, and F1 score measurements are commonly used to evaluate the success of the proposed models. The confusion matrix parameters shown in Table 2 are used to calculate these measures [12].

Table 2. Confusion matrix structure for a two-class problem

		Predicts	
		Negative	Positive
Actuals	Negative	True Negative (TN)	False Positive (FP)
	Positive	False Negative (FN)	True Positive (TP)

Here, TP true positive, TN true negative, FN false negative and FP false positive. Accordingly, accuracy metric presents the ratio of the number of samples correctly classified by the classifier to the total number of samples, as shown in Eq. (1). Precision metric provides the percentage of samples that are correctly classified as positive to the total number of samples that are positively classified, as given in Eq. (2). Sensitivity metric represents the percentage of positive samples that are correctly identified out of the total number of positive samples, as shown in Eq. (3). Lastly, F1-score metric indicates the harmonic mean of the precision and sensitivity measures, as shown in Eq. (4).

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

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

$$Sensitivity = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - score = 2 \times \frac{Precision \times Sensitivity}{Precision + Sensitivity} \quad (4)$$

In multi-class problems, the interest class is considered positive, and the other classes are considered negative. Accordingly, class-wise measures are conducted using the metrics in Equations 1 to 4. Then the averages of the measurements obtained for each class are calculated using the metrics in Equations 5 to 8.

$$Average Accuracy(AA) = \frac{1}{N} \sum_{i=1}^N Accuracy_i \quad (5)$$

$$Average Precision(AP) = \frac{1}{N} \sum_{i=1}^N Precision_i \quad (6)$$

$$Average Sensitivity(AS) = \frac{1}{N} \sum_{i=1}^N Sensitivity_i \quad (7)$$

$$Average F1 - score(AF1) = \frac{1}{N} \sum_{i=1}^N F1 - score_i \quad (8)$$

where i and N indicate the related class and number of classes, respectively.

4. Experiments and results

4.1. Experimental environment

The experiments were performed on the Google Colaboratory environment with 12.7 GB RAM, 15 GB GPU RAM, 107.7 GB hard drive, and Python 3 Google Compute Engine (GPU) backend.

4.2. Proposed study

This study employs several deep learning networks in the experiments. Figures 1 and 2 show the graphical abstracts. Accordingly, Figure 1 shows the overall block diagram of the study, while Figure 2 shows a representative illustration of transfer learning approach. In the output layers of the pre-trained models, there are two neurons for Dataset-1 and three for Dataset-2. Considering the limitations of experimental studies in the Google Colaboratory environment, the epoch number was set to 20. Firstly, all images were resized according to the input size of each model's input layer. Using the

model checkpoint technique, the weights of the model for the training tour in which the best success was achieved on the validation set were retained, and the model's performance was measured on the test set after training. Hard and soft voting ensemble learning approaches were applied after determining the best five models among these models.

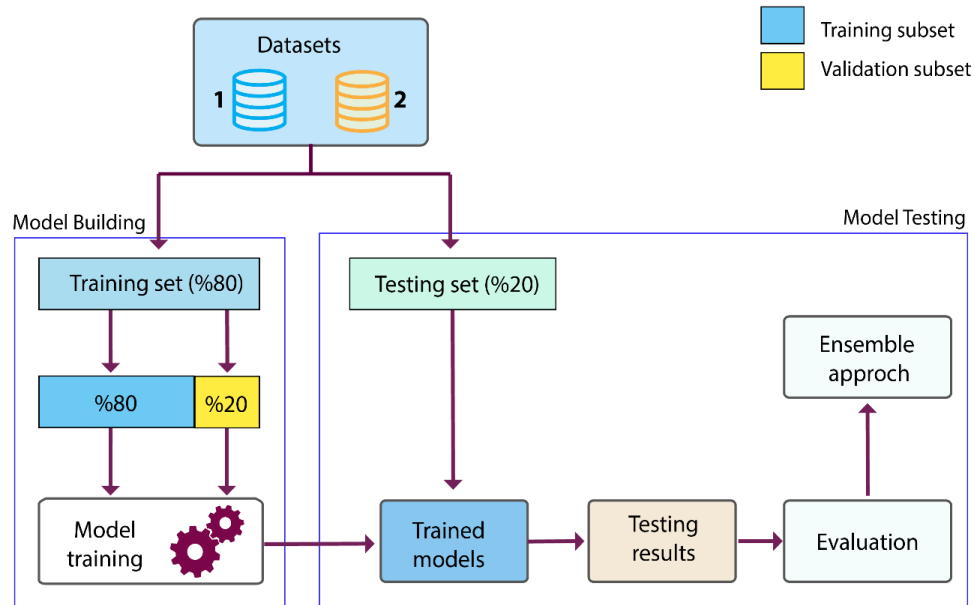


Fig. 1. General block diagram of the proposed study

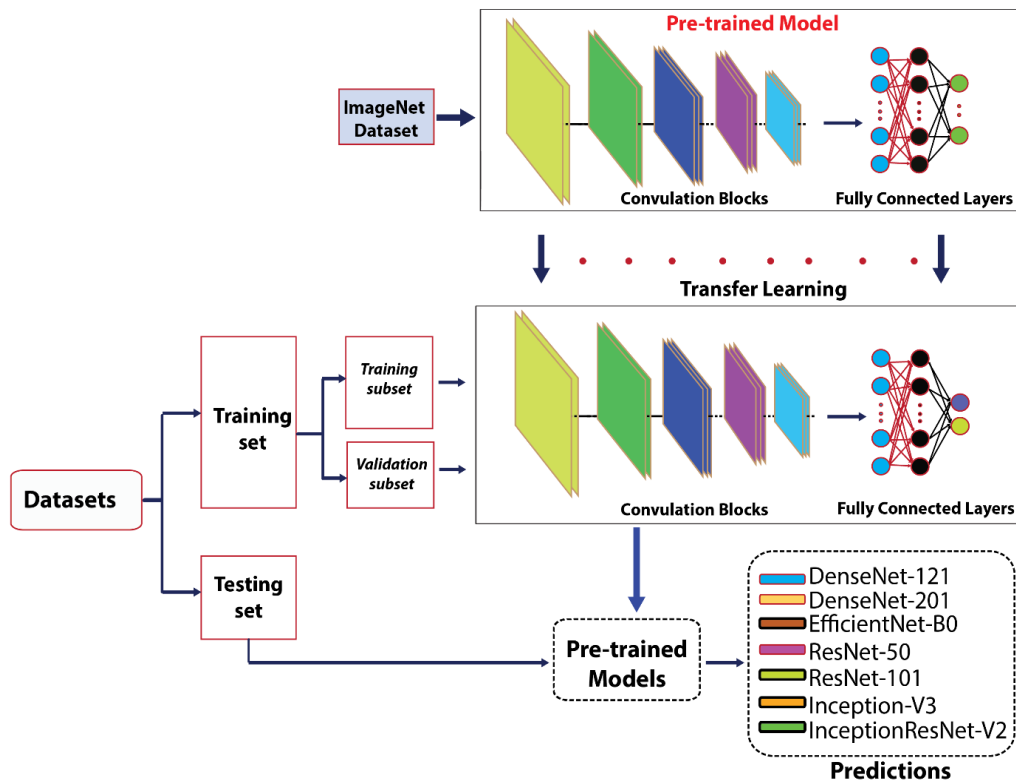


Fig. 2. Graphical abstract of transfer learning approach

4.3. Results

4.3.1. Dataset-1 results

Figures 3 and 4 show the accuracy plot of each pre-trained deep learning model on the training and validation set for Dataset-1.

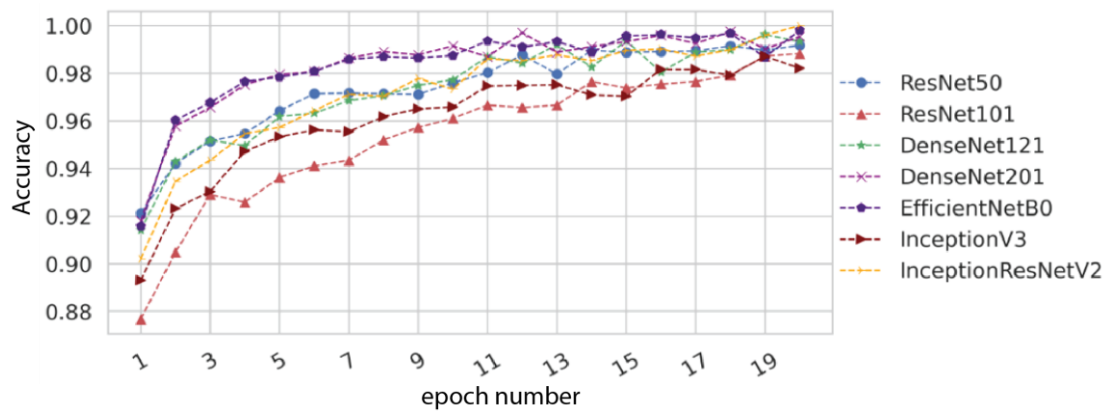


Fig. 3. Accuracy graphs of pre-trained models on the training dataset for Dataset-1

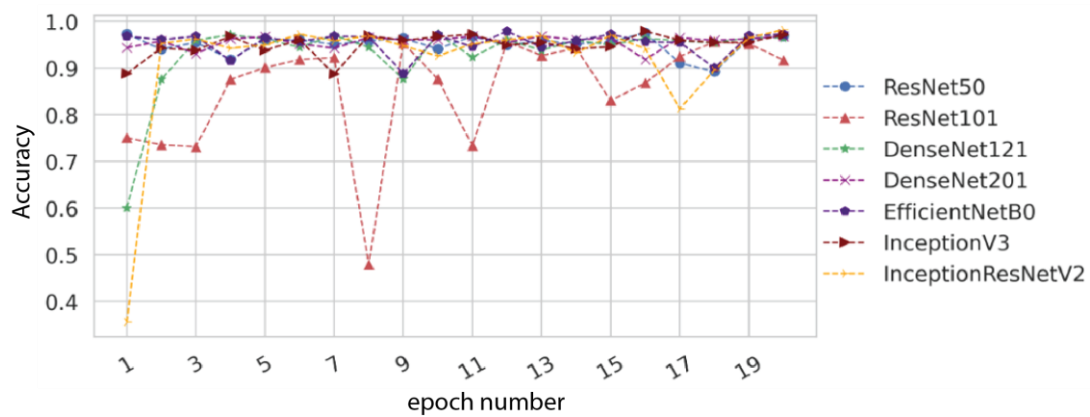


Fig. 4. Accuracy graphs of pre-trained models on the validation dataset for Dataset-1

After training, the performance of each model was calculated using the confusion matrix values of the models. Confusion matrices summarize the correct and incorrect predictions. The confusion matrices of the models for Dataset-1 are shown in Figure 5. Table 3 shows the accuracy, precision, sensitivity, and F1-score values of the models applied to Dataset-1. The accuracy values in bold in this table indicate the five most successful models. Accordingly, the top five most accurate models are ResNet-50, DenseNet-201, Inception-V3, DenseNet-121, and InceptionResNet-V2. These models gave accuracies of 97.95%, 97.87%, 97.7%, 97.7%, and 97.61%, respectively. For the Inception-V3 and DenseNet-121 models with the same accuracy values, the F1 score value was determined as a prioritization criterion.

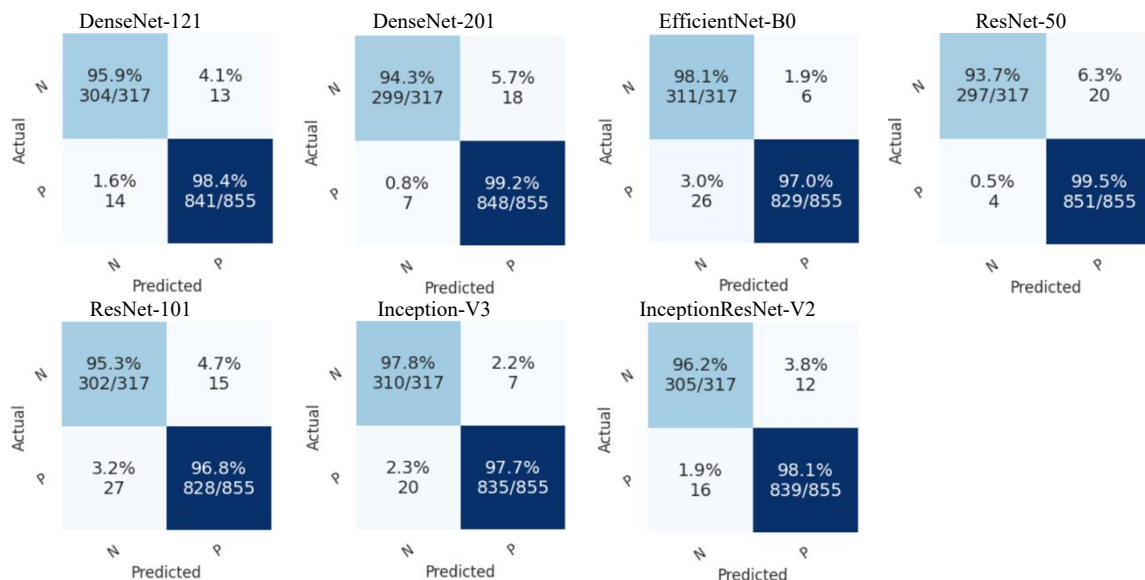


Fig. 5. Confusion matrices of the pre-trained models for Dataset-1 (N: Normal, P: Pneumonia)

Table 3. Experimental results on the Dataset-1

	Accuracy (%)	Precision (%)	Sensitivity (%)	F1-Score (%)
ResNet-50	97,95	97,7	99,53	98,61
ResNet-101	96,42	98,22	96,84	97,53
DenseNet-121	97,7	98,48	98,36	98,42
DenseNet-201	97,87	97,92	99,18	98,55
EfficientNet-B0	97,27	99,28	96,96	98,11
Inception-V3	97,7	99,17	97,66	98,41
InceptionResNet-V2	97,61	98,59	98,13	98,36

The confusion matrices shown in Figure 6 summarize the results obtained with the voting ensemble learning approaches, including the five models with the most successful results on Dataset-1.



Fig. 6. Confusion matrices of voting ensemble learning approaches for Dataset-1; a) hard voting, b) soft voting (N: Normal, P: Pneumonia).

The results obtained by both voting approaches are summarized in Table 4. According to this table, the soft voting approach provided 98.55% accuracy, 98.61% precision, 99.42% sensitivity, and 99.01% F1 score. The hard voting approach gave an accuracy value of 98.38%. Accordingly, the soft voting approach provided the highest accuracy on Dataset-1.

Table 4. Experimental results of voting approaches on the Dataset-1

	Accuracy (%)	Precision (%)	Sensitivity (%)	F1-Score (%)
Hard voting	98,38	98,38	99,42	98,90
Soft voting	98,55	98,61	99,42	99,01

4.3.2. Dataset-2 results

Figures 7 and 8 show the accuracy graph of each pre-trained deep learning model on the training and validation sets for Dataset-2.

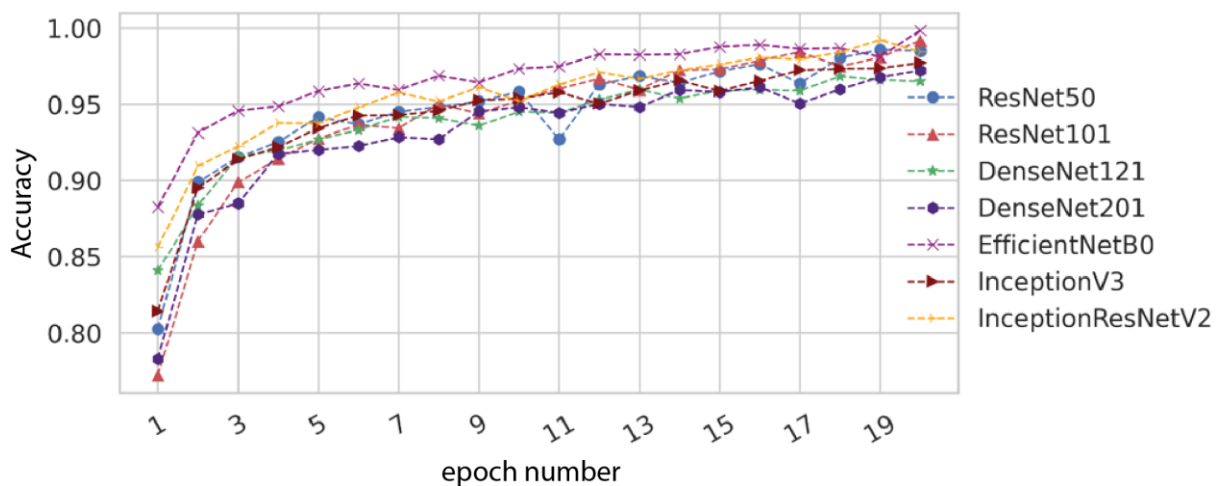


Fig. 7. Accuracy graphs of pre-trained models on the training dataset for Dataset-2

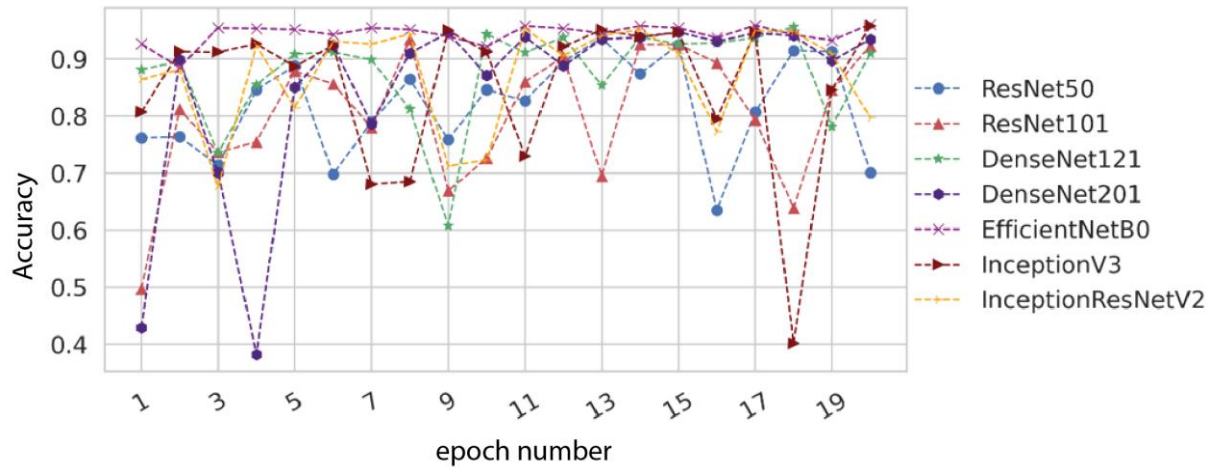


Fig. 8. Accuracy graphs of pre-trained models on the validation dataset for Dataset-2

After the training process, the performance of each model was calculated using the confusion matrix values of the models. The confusion matrices of the models for Dataset-2 are shown in Figure 9. Table 5 shows the models' accuracy, precision, sensitivity, and F1-score values to Dataset-2. The accuracy values in bold in this table indicate the five most successful models. The five most successful models were EfficientNet-B0 with 96.88% accuracy, InceptionResNet-V2 with 96.59% accuracy, Inception-V3 with 96.35% accuracy, DenseNet-121 with 96.11% accuracy and DenseNet-201 with 96.06% accuracy.

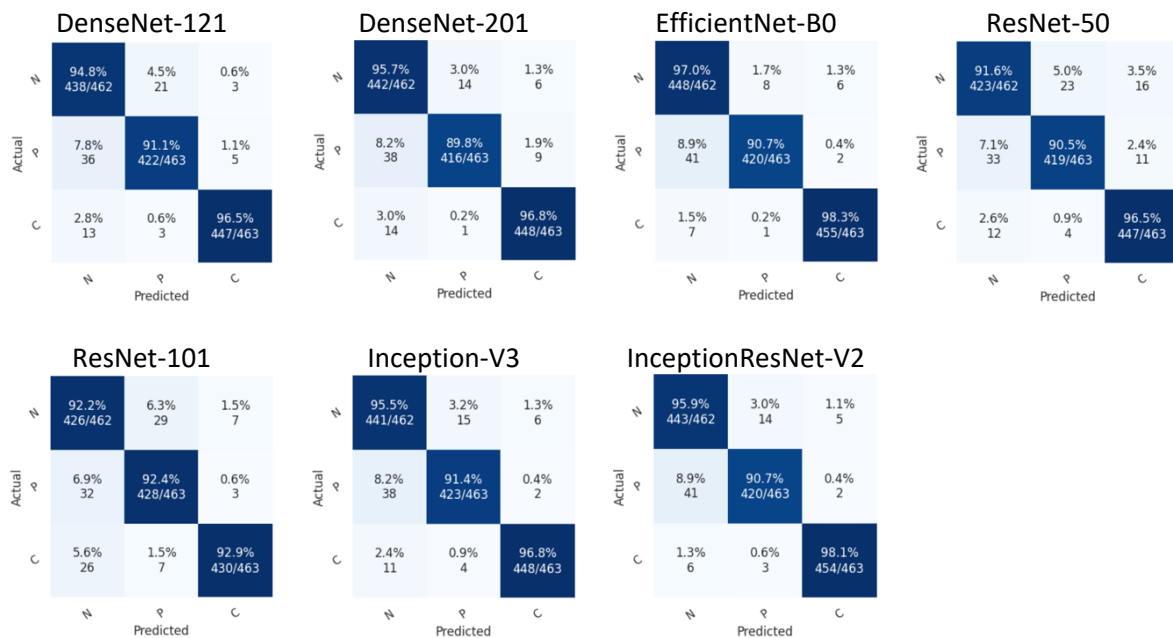


Fig. 9. Confusion matrices of pre-trained models for Dataset-2 (N: Normal, P: Pneumonia, C: Covid-19).

Table 5. Experimental results on the Dataset-2

	Accuracy (%)	Precision (%)	Sensitivity (%)	F1-Score (%)
ResNet-50	95,24	92,88	92,87	92,87
ResNet-101	95,00	92,66	92,51	92,58
DenseNet-121	96,11	94,27	94,16	94,21
DenseNet-201	96,06	94,25	94,09	94,17
EfficientNet-B0	96,88	95,5	95,32	95,41
Inception-V3	96,35	94,65	94,52	94,58
InceptionResNet-V2	96,59	95,00	94,89	94,94

The confusion matrices shown in Figure 10 summaries the results obtained with the voting ensemble learning approaches, including the five models with the most successful results on Dataset-2.

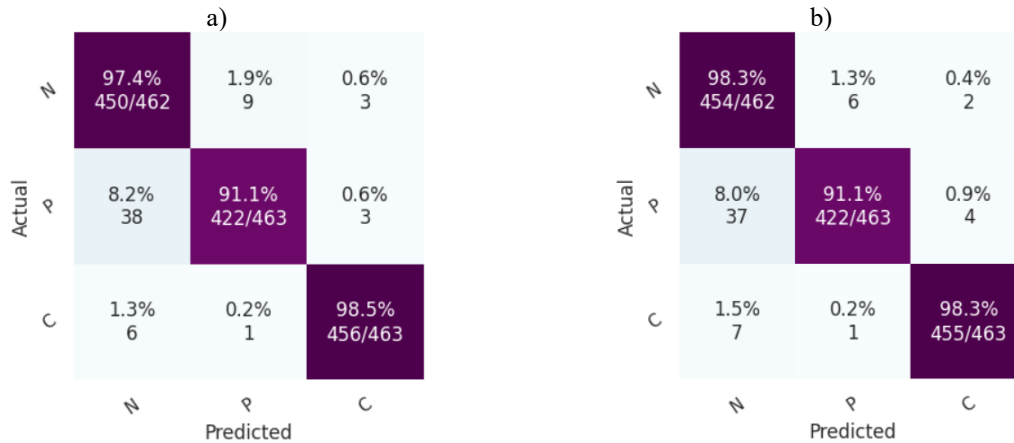


Fig. 10. Confusion matrices of voting approaches for Dataset-2; a) hard voting, b) soft voting (N: Normal, P: Pneumonia, C: Covid-19).

The results obtained with both voting learning approaches on Dataset-2 are summarized in Table 6. The soft voting approach gave the best results with 97.26% accuracy, 96.08% precision, 95.89% sensitivity, and 95.98% F1-score.

Table 6. Experimental results of voting approaches on the Dataset-2

	Accuracy (%)	Precision (%)	Sensitivity (%)	F1-Score (%)
Hard voting	97,12	95,83	95,68	95,75
Soft voting	97,26	96,08	95,89	95,98

Following experiments and findings, the decision support tool for field experts was developed using the soft voting approach that incorporates the most successful deep learning models. Figure 11 presents the user interface application. With this tool, the field expert uploads a patient's X-ray image to the system, and the system analyzes the X-ray image using our build soft voting-based models, for Dataset-1 (2-classes) and Dataset-2 (3-classes), respectively and the results are displayed on the screen and also saved to the system. Therefore, this system also allows the field expert to analyze patient information and medical history in detail. Model-1 and Model-2 in this user interface represent soft voting models built for 2-class and 3-class datasets, respectively.

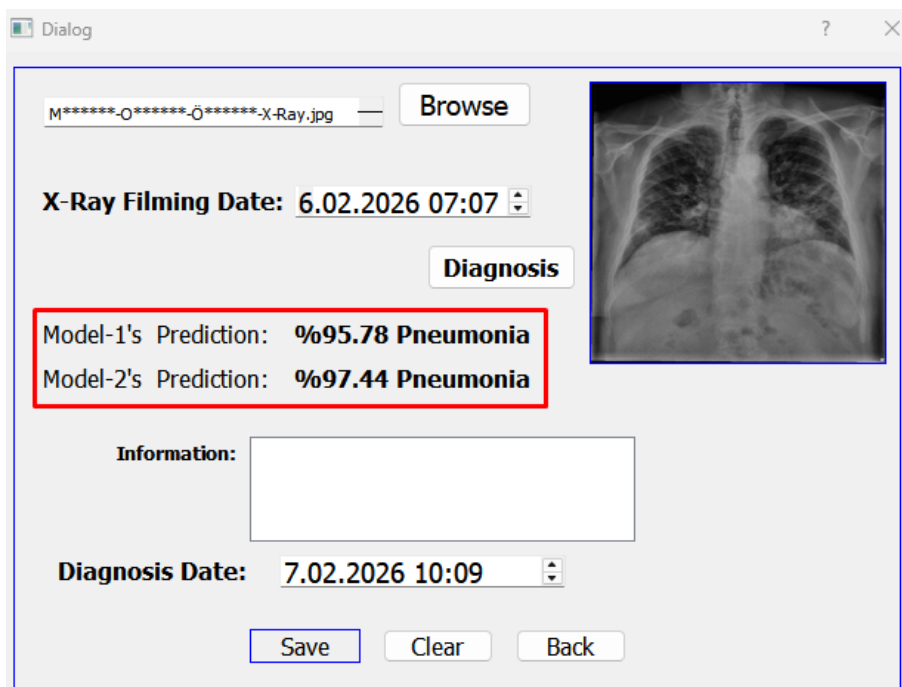


Fig. 11. User interface application for pneumonia detection

5. Discussion

Table 7 summarizes the experimental results of this paper and some works in literature. Accordingly, this table includes author names, dataset information, number of classes, number of instances, most successful model, and accuracy values. The bold fonts in this table indicate the results of this paper. For example, El Asnaoui, in his study on pneumonia disease detection, achieved a 95.17% success rate with hard voting in the ensemble learning approach he applied to the two-class dataset [21]. Mujahid et al. achieved 99.29% accuracy with Inception-V3 and the hard voting ensemble learning approach, including the original model they proposed and Inception-V3, on the two-class 'Pneumonia ChestX-ray Image' dataset, which includes images of pneumonia patients and those with no signs of disease for pneumonia disease detection [23]. In their study, Hashmi et al. achieved 98.00% accuracy with the DenseNet-121 model on the two-class 'Chest X-Ray Images' dataset, which includes images of patients with pneumonia and those without signs of disease for pneumonia disease detection [20]. Korkmaz achieved 97.70% accuracy with DenseNet-121 on the three-class dataset [30]. Narin combined three different datasets containing images of COVID-19 patients, patients with bacterial pneumonia, patients with pneumonia, and patients with no signs of the disease [22]. He achieved 99.70% accuracy with the ResNet-50 model using 5-fold cross-validation on this newly created four-class dataset. He achieved 99.70% accuracy with the ResNet-50 model using 5-fold cross-validation on this newly created four-class dataset [31]. Usman et al. demonstrated that their CNN model developed from scratch performed better than a pre-trained ResNet-50 model for pneumonia detection. Furthermore, the authors confirmed the importance of data augmentation and data size in their study [32]. Şimşek et al. applied resizing, cropping, bilinear interpolation, and histogram equalization preprocessing techniques to 1000 chest X-ray images obtained from various age groups in the clinics of Koç University Faculty of Medicine Hospital and performed pneumonia analysis with the CNN model they developed [33]. Shojib et al. applied Contrast-Limited Adaptive Histogram Equalization and Laplacian filtering preprocessing techniques for the detection of pneumonia and performed feature extraction with pre-trained deep learning models. The authors reported that they obtained the best results with the DenseNet201 feature extractor and the deep learning-based classifier they introduced [34]. Wu evaluated the performance of various convolutional neural networks for chest X-ray image classification and detection and reported that the lightweight ShuffleNetV2 model excelled in classification tasks, achieving high accuracy while minimizing computational load and memory consumption [35]. Nageye et al. investigated the performance of Simple CNN, DenseNet121, VGG16, and InceptionV3 models to distinguish between normal and pneumonia images. Furthermore, the authors employed data normalization and data augmentation techniques and reported that Simple CNN achieved the highest accuracy at 92% in their experiments [36]. In this study, the soft voting ensemble learning method achieved 98.55% accuracy for Dataset-1 and 97.26% accuracy for Dataset-2. As can be seen, the soft voting ensemble models presented in this study have provided acceptably high accuracies.

Table 7. The results of this paper and related works

Study	Dataset	Number of Classes	Number of samples	Data type	Model / Approach	Accuracy
Hashmi M.F. et al. [20]	"Chest X-ray Images"	2	5856	X-ray	DenseNet-121	98,00
El Asnaoui [21]	"Chest X-ray Images"	2	5856	X-ray	Ensemble learning	95,17
Mujahid M. et al. [23]	"Pneumonia Chest X-ray Image Dataset"	2	7750	X-ray	Ensemble learning	99,29
This study	Dataset-1	2	5856	X-ray	Soft voting	98,55
Korkmaz and Atila [30]	"COVID-19 Radiography Database"	3	3886	X-ray	DenseNet-121	97,70
Narin A. et al. [22]	"ChestX-ray" "Chest X-ray Images" "Chest X-ray Images (Pneumonia)"	4	7406	X-ray	ResNet-50	99,70
Usman et al. [32].	"RSNA pneumonia detection challenge dataset"	2	26,684	X-ray	CNN model	79
Şimşek et al. [33]	Koç University	2	1000	X-ray	CNN model	93,3
Shojib et al. [34]	"ChestX-ray"	2	5216	X-ray	DenseNet201 + Deep neural network classifier	99,68
Wu [35]	"COVID-19 Radiography Database"	4	21165	X-ray	Lightweight ShuffleNetV2	92,51
Nageye et al. [36]	"Chest X-Ray Images (Pneumonia)"	2	5863	X-ray	Simple CNN	92
This study	Dataset-2	3	6938	X-ray	Soft voting	97,26

6. Conclusion

This paper conducts deep learning experiments to assist field experts in detecting pneumonia, a disease that negatively affects human life. The results show that the soft voting, including the ResNet-50, DenseNet-201, Inception-V3, DenseNet-121, and Inception ResNet-V2, the best ones on the Dataset-1, achieved 98.55% accuracy. Also, the soft voting learning approach, which includes EfficientNet-B0, InceptionResNet-V2, Inception-V3, DenseNet-121, DenseNet-201, DenseNet-121, and DenseNet-201, the best ones on the Dataset-2, presented 97.26% accuracy. Compared to related state-of-the-art works, the classification accuracy of this paper is acceptably high. Also, this study introduced the decision support software for making the right decisions. Finally, the model proposed in this study can be incorporated into a decision support system designed to quickly analyze X-ray images with high accuracy, especially in hospitals with limited resources. Thus, the current work has the potential to be beneficial for radiologists and advantageous for usage in scanning conditions that are found in the real world. It could also be considered essential for future work in computer vision. On the other hand, it is also evaluated that this study has a dataset limitation. Validation experiments for models conducted in this paper were performed separately on two publicly available datasets. External validation works on different datasets are required for domain adaptation and, thus, model robustness. In the future, an investigation into more efficient ensemble procedures is going to be carried out with the intention of improving the diagnosis of pneumonia patients. In addition, one of the main efforts is to concentrate on different diseases for enhancing the robustness of the model that planned work.

All the Declarations and Statements

Author Contributions Statement

Mustafa Oguzhan Ozdemir – Data Curation and Software Implementation: Handled data acquisition, dataset preprocessing. Model Training, Validation, and Performance Evaluation: Led the model training process, validated results using standard metrics.

Kemal Akyol –Conceptualization, Methodology, and Supervision: Proposed research ideas, Constructed the overall framework.

All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding Declaration

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability Statement

The figures and tables supporting the results of this study are included in the article. Additionally, the datasets used for this paper are available at the following links:

- Dataset-1: <https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>
- Dataset-2: <https://www.kaggle.com/datasets/amanullahasraf/covid19-pneumonia-normal-chest-xray-pa-dataset>

Ethical Declarations

N/A.

Acknowledgements

We deeply appreciate the professional analysis provided by the editor and reviewers and their useful suggestions, which have helped to enhance the quality of the experiment and the validity of its findings.

Declaration of Generative AI in Scholarly Writing

Language polishing and grammar were done with the help of AI. There was no use of a generative AI to create content, generate ideas, analyse data, generate figures, or write scientifically.

Abbreviations

This manuscript uses the following abbreviations:

GPU – Google Compute Engine

TN – True Negative
 FP – False Positive
 TP – True Positive
 FN – False Negative

Appendix A/B/C..., with appendix title

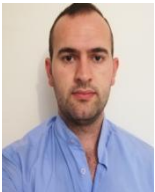
None.

References

- [1] N.M. Elshennawy, D.M. Ibrahim, Deep-Pneumonia Framework Using Deep Learning Models Based on Chest X-Ray Images, *Diagnostics* 2020, Vol. 10, Page 649 10 (2020) 649. <https://doi.org/10.3390/DIAGNOSTICS10090649>.
- [2] C. Ortiz-Toro, A. García-Pedrero, M. Lillo-Saavedra, C. Gonzalo-Martín, Automatic detection of pneumonia in chest X-ray images using textural features, *Comput. Biol. Med.* 145 (2022) 105466. <https://doi.org/10.1016/J.COMPBIOMED.2022.105466>.
- [3] M. Ali, M. Shahroz, U. Akram, M.F. Mushtaq, S.C. Altamiranda, S.A. Obregon, I.D.L.T. Diez, I. Ashraf, Pneumonia Detection Using Chest Radiographs with Novel EfficientNetV2L Model, *IEEE Access* 12 (2024) 34691–34707. <https://doi.org/10.1109/ACCESS.2024.3372588>.
- [4] J.F. Cordier, Organising pneumonia, *Thorax* 55 (2000) 318–328. <https://doi.org/10.1136/THORAX.55.4.318>.
- [5] O. Ruuskanen, E. Lahti, L.C. Jennings, D.R. Murdoch, Viral pneumonia, *Lancet* 377 (2011) 1264–1275. [https://doi.org/10.1016/S0140-6736\(10\)61459-6](https://doi.org/10.1016/S0140-6736(10)61459-6).
- [6] M. Grewal, M.M. Srivastava, P. Kumar, S. Varadarajan, RADnet: Radiologist level accuracy using deep learning for hemorrhage detection in CT scans, *Proceedings - International Symposium on Biomedical Imaging 2018-April* (2018) 281–284. <https://doi.org/10.1109/ISBI.2018.8363574>.
- [7] G.N. Meenakshi Sundara Desikar, R. Kadarkarai, K. Somasundaram, V. Kandasamy, P. Partheeban De, Optimized multi-dimensional attention spiking neural network for pneumonia detection in chest x-ray images, *Biomed. Signal Process. Control* 113 (2026) 108765. <https://doi.org/10.1016/J.BSPC.2025.108765>.
- [8] Z. Hu, Z. Yang, K.J. Lafata, F.F. Yin, C. Wang, A radiomics-boosted deep-learning model for COVID-19 and non-COVID-19 pneumonia classification using chest x-ray images, *Med. Phys.* 49 (2022) 3213–3222. <https://doi.org/10.1002/MP.15582>.
- [9] O. Stephen, M. Sain, U.J. Maduh, D.U. Jeong, An Efficient Deep Learning Approach to Pneumonia Classification in Healthcare, *J. Healthc. Eng.* 2019 (2019) 4180949. <https://doi.org/10.1155/2019/4180949>.
- [10] T. Rahman, M.E.H. Chowdhury, A. Khandakar, K.R. Islam, K.F. Islam, Z.B. Mahbub, M.A. Kadir, S. Kashem, Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection Using Chest X-ray, *Applied Sciences* 2020, Vol. 10, Page 3233 10 (2020) 3233. <https://doi.org/10.3390/APPI0093233>.
- [11] R. Jain, P. Nagrath, G. Kataria, V. Sirish Kaushik, D. Jude Hemanth, Pneumonia detection in chest X-ray images using convolutional neural networks and transfer learning, *Measurement* 165 (2020) 108046. <https://doi.org/10.1016/J.MEASUREMENT.2020.108046>.
- [12] M.B. Er, Classification of Pneumonia Using Pre-Trained Deep Networks With Chest X-Ray Images | GCRIS Database | Konya Technical University, *Konya Journal of Engineering Sciences* (2021). <https://gcris.ktun.edu.tr/handle/20.500.13091/2125> (accessed April 9, 2025).
- [13] S. Ben Atitallah, M. Driss, W. Boulila, A. Koubaa, H. Ben Ghézala, Fusion of convolutional neural networks based on Dempster–Shafer theory for automatic pneumonia detection from chest X-ray images, *Int. J. Imaging Syst. Technol.* 32 (2022) 658–672. <https://doi.org/10.1002/IMA.22653>.
- [14] N. Rai, N. Kaushik, D. Kumar, C. Raj, A. Ali, Mortality prediction of COVID-19 patients using soft voting classifier, *International Journal of Cognitive Computing in Engineering* 3 (2022) 172. <https://doi.org/10.1016/J.IJCCE.2022.09.001>.
- [15] C. Budak, V. Mençik, O. Varışlı, Online diagnosis of COVID-19 from chest radiography images by using deep learning algorithms, *Neural Comput. Appl.* 35 (2023) 20717–20734. <https://doi.org/10.1007/S00521-023-08867-5>.
- [16] N. Srivastava, G. Hinton, A. Krizhevsky, R. Salakhutdinov, Dropout: A Simple Way to Prevent Neural Networks from Overfitting, *Journal of Machine Learning Research* 15 (2014) 1929–1958.
- [17] M. Ben Jabra, A. Koubaa, B. Benjdira, A. Ammar, H. Hamam, COVID-19 Diagnosis in Chest X-rays Using Deep Learning and Majority Voting, *Applied Sciences* 2021, Vol. 11, Page 2884 11 (2021) 2884. <https://doi.org/10.3390/APPI1062884>.
- [18] M.G. Lanjewar, K.G. Panchbhair, P. Charanarur, Small size CNN-Based COVID-19 Disease Prediction System using CT scan images on PaaS cloud, *Multimed. Tools Appl.* 83 (2024) 60655–60687. <https://doi.org/10.1007/S11042-023-17884-4>.
- [19] B.P. Kaur, H. Singh, R. Hans, S.K. Sharma, C. Kaushal, M.M. Hassan, M.A. Shah, An augmentation aided concise CNN based architecture for COVID-19 diagnosis in real time, *Sci. Rep.* 14 (2024). <https://doi.org/10.1038/S41598-024-51317-Y>.
- [20] M.F. Hashmi, S. Katiyar, A.G. Keskar, N.D. Bokde, Z.W. Geem, Efficient Pneumonia Detection in Chest Xray Images Using Deep Transfer Learning, *Diagnostics (Basel)*. 10 (2020). <https://doi.org/10.3390/DIAGNOSTICS10060417>.
- [21] K. El Asnaoui, Design ensemble deep learning model for pneumonia disease classification, *Int. J. Multimed. Inf. Retr.* 10 (2021) 55–68. <https://doi.org/10.1007/S13735-021-00204-7/TABLES/6>.
- [22] A. Narin, C. Kaya, Z. Pamuk, Automatic Detection of Coronavirus Disease (COVID-19) Using X-ray Images and Deep Convolutional Neural Networks, *Pattern Analysis and Applications* 2021 24:3 24 (2020) 1207–1220. <https://doi.org/10.1007/s10044-021-00984-y>.
- [23] M. Mujahid, F. Rustam, R. Álvarez, J. Luis Vidal Mazón, I. de la T. Diez, I. Ashraf, Pneumonia Classification from X-ray Images with Inception-V3 and Convolutional Neural Network, *Diagnostics* 2022, Vol. 12, Page 1280 12 (2022) 1280. <https://doi.org/10.3390/DIAGNOSTICS12051280>.

- [24] W. Aldhahi, S. Sull, Uncertain-CAM: Uncertainty-Based Ensemble Machine Voting for Improved COVID-19 CXR Classification and Explainability, *Diagnostics (Basel)*. 13 (2023). <https://doi.org/10.3390/DIAGNOSTICS13030441>.
- [25] H. Zhang, Z. Lv, S. Liu, Z. Sang, Z. Zhang, Cn2a-capsnet: a capsule network and CNN-attention based method for COVID-19 chest X-ray image diagnosis, *Discover Applied Sciences* 6 (2024) 1–18. <https://doi.org/10.1007/S42452-024-05796-3/TABLES/7>.
- [26] L. Torrey, J. Shavlik, T. Walker, R. MacLin, Transfer Learning via Advice Taking, *Studies in Computational Intelligence* 262 (2010) 147–170. https://doi.org/10.1007/978-3-642-05177-7_7.
- [27] R. Ribani, M. Marengoni, A Survey of Transfer Learning for Convolutional Neural Networks, *Proceedings - 32nd Conference on Graphics, Patterns and Images Tutorials, SIBGRAPI-T 2019* (2019) 47–57. <https://doi.org/10.1109/SIBGRAPI-T.2019.00010>.
- [28] G. Dagnew, B.H. Shekar, Ensemble learning-based classification of microarray cancer data on tree-based features, *Cognitive Computation and Systems* 3 (2021) 48–60. <https://doi.org/10.1049/CCS2.12003>.
- [29] K. Pal, B. V. Patel, Data Classification with k-fold Cross Validation and Holdout Accuracy Estimation Methods with 5 Different Machine Learning Techniques, *Proceedings of the 4th International Conference on Computing Methodologies and Communication, ICCMC 2020* (2020) 83–87. <https://doi.org/10.1109/ICCMC48092.2020.ICCMC-00016>.
- [30] A. Korkmaz, U. Atila, *Artificial Intelligence Studies*, *Artificial Intelligence Studies* (2021) 1–13. <https://www.aistudies.org/index.php/ais> (accessed April 9, 2025).
- [31] K. He, R. Girshick, P. Dollar, Rethinking ImageNet Pre-training, *Proceedings of the IEEE International Conference on Computer Vision 2019-October* (2018) 4917–4926. <https://doi.org/10.1109/ICCV.2019.00502>.
- [32] Usman C, Rehman SU, Ali A, Khan AM, Ahmad B. Pneumonia Disease Detection Using Chest X-Rays and Machine Learning. *Algorithms* 2025, Vol 18, Page 82 2025;18:82. <https://doi.org/10.3390/A18020082>.
- [33] Şimşek Ç, Özkorukcuklu S, Işıldak B. Pneumonia detection in chest X-ray images using convolutional neural networks. *The European Research Journal* 2025;11:907–14. <https://doi.org/10.18621/EURJ.1641267>.
- [34] Shojib MSH, Mridha MMR, Nakib MFN, Islam S, Imtiaz S, Rahman MM, et al. Pneumonia Detection Using Deep Learning and Machine Learning Approaches from Chest X-Ray. 2025 IEEE 2nd International Conference on Computing, Applications and Systems, COMPAS 2025 2025. <https://doi.org/10.1109/COMPAS67506.2025.11381803>.
- [35] Wu C. Pneumonia Detection and Classification Using Lightweight ShuffleNetV2 Based on Transfer Learning, *Proceedings of the 2025 2nd International Conference on Generative Artificial Intelligence and Information Security, GAIIS 2025* 2025;1:621–5. <https://doi.org/10.1145/3728725.3728823>.
- [36] Nageye AY, Jimale AD, Abdullahi MO, Ahmed YA, Addow MA. Enhancing deep learning for pneumonia detection: developing web based solution for Dr. Sumait Hospital in Mogadishu Somalia. *Discover Applied Sciences* 2025 7:4 2025;7:309-. <https://doi.org/10.1007/S42452-025-06735-6>.

Authors' Profiles



Mustafa Oguzhan Ozturk received his B.Sc. from Computer Science Department of Karabuk University, Karabuk/Turkey in 2019. He received his M.Sc. degree from Natural and Applied Sciences, Kastamonu University, Kastamonu/Turkey in 2024.



Kemal Akyol received his B.Sc. from Computer Science Department of Gazi University, Ankara/Turkey in 2002. He received his M.Sc. degree from Natural and Applied Sciences, Karabuk University, Karabuk/Turkey and PhD degree from the same department. His research interests include machine learning, decision support systems and expert systems. He currently works at the Department of Computer Engineering, Kastamonu University.

How to cite this paper

Mustafa Oguzhan Ozdemir, Kemal Akyol, "Ensemble-Based Modelling for Enhanced Detection of Pneumonia Disease", *International Journal of Engineering and Manufacturing (IJEM)*, Vol.16, No.4, pp.58-70, 2026. DOI:10.5815/ijem.2026.04.04