MATHEMATICAL MODELING AND COMPUTING, Vol. 12, No. 1, pp. 22–30 (2025)

Comparison of some CNN architectures for detecting cardiomegaly from chest X-ray images

Regragui H.[1], Ziani M.[1], Moussaid N.[2]

[1] *LMSA Laboratory, Department of Mathematics, Faculty of Sciences, Mohammed V University in Rabat, Morocco, hind-regragui@um5r.ac.ma, m.ziani@um5r.ac.ma*

[2] *LMCSA, Hassan II University of Casablanca, P.O. Box 146, Mohammedia, Morocco, noureddine.moussaid@fstm.ac.ma*

(Received 22 December 2023; Revised 22 July 2024; Accepted 18 November 2024)

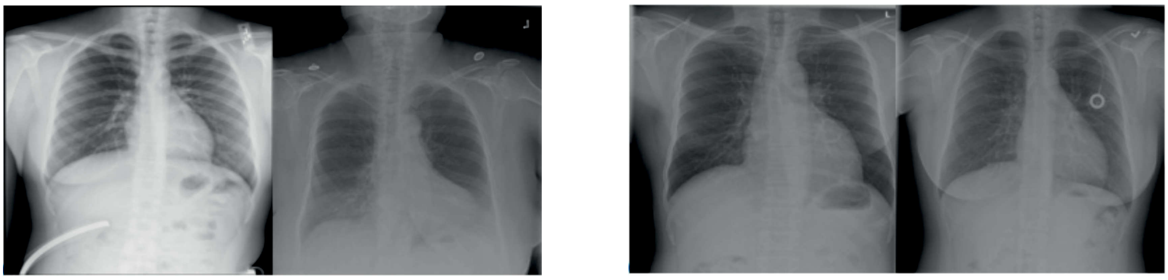
In medical image analysis, deep learning and convolutional neural networks (CNN) are widely employed, particularly in tasks such as classification and segmentation. This study specifically addresses their application in healthcare for detecting cardiomegaly, a condition characterized by an enlarged heart, often related to factors such as hypertension or coronary artery diseases. The primary objective is to develop an algorithm to identify cardiomegaly in chest X-ray images, constituting a binary classification problem (whether the image exhibits cardiomegaly or not). Using the CXR8 dataset from the National Institute of Health Clinical Center, comprising 2 776 cardiomegaly images and 60 361 no finding images, the inputs are labeled images, and the outputs are the corresponding labels (Cardiomegaly or No Finding). Employing Keras and TensorFlow Python libraries, we aim to construct a CNN model that excels in binary classification, distinguishing between cardiomegaly and no finding in chest X-ray images.

Keywords: *deep learning; convolutional neural networks; cardiomegaly; binary classification.*

2010 MSC: 92B20, 68T05, 92C42, 92C55, 68U10 **DOI:** 10.23939/mmc2025.01.022

1. Introduction

Cardiomegaly refers to a medical situation where the heart becomes enlarged. Often termed as an "enlarged heart," its origins can differ. Frequently, this condition emerges due to hypertension or coronary artery diseases. When the heart enlarges, it may struggle to pump blood efficiently, leading to the onset of congestive heart failure. The X-ray imagery provides insights into the state of the lungs and heart [1], see Figure 1.



a ***b***

Fig. 1. (***a***) X-ray images for patients without cardiomegaly, (***b***) X-ray images for patients with cardiomegaly.

X-ray tests are the initial procedure for identifying cardiomegaly in a patient [2]. After obtaining the X-ray, a radiologist examines it to determine the presence of the condition [3]. If an X-ray reveals an enlarged heart, additional tests are typically conducted to determine the cause. The cardio-thoracic ratio, measured from X-rays, compares the heart transverse diameter to that of the thoracic cage. This measurement use the broadest part of the chest, extending to the lung pleura, excluding the outer skin edges. Pathology is considered when the cardiothoracic ratio exceeds 50%, provided the X-ray is taken accurately. A modern technique involves assessing heart health by comparing the heart area to the chest area, termed the two-dimensional cardiothoracic ratio, see Figure 2. The cardiothoracic is defined as

$$\text{Cardiothoracic ratio} = \frac{MRD + MLD}{ID},$$

where MRD is greatest perpendicular diameter from midline to right heart border, MLD is greatest perpendicular diameter from midline to left heart border, ID is internal diameter of chest at level of right hemidiaphragm.

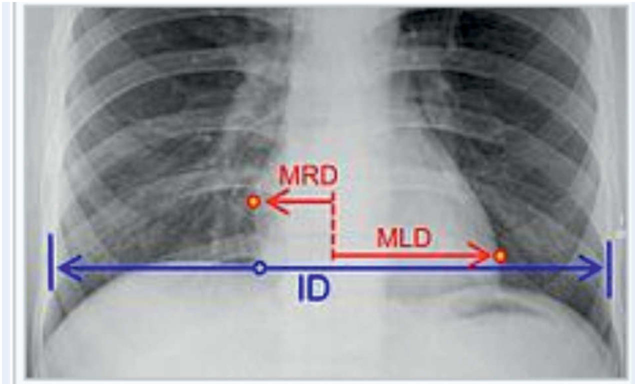


Fig. 2. Cardiomegaly.

Deep neural networks, particularly Convolutional Neural Networks (CNN), have emerged as the dominant machine learning architecture in recent years [4]. Their widespread adoption can be attributed to their unique ability to hierarchically extract and process information from images [5]. This inherent capability allows CNN to identify patterns and features at various levels of complexity [6]. As a result, they have been instrumental in the medical field, where accurate image classification is paramount [7]. From detecting anomalies in X-rays to differentiating between healthy and malignant cells in histopathological slides, CNN have revolutionized the way medical images are analyzed, aiding in more accurate and timely diagnoses [8].

The objective of our work is to conduct a comprehensive comparison between three prominent architectures, namely CNN, ResNet-50, and MobileNet. By contrasting their performances, we aim to determine which model delivers the highest accuracy in our specific application.

The paper is organized as follows. The next section offers a concise review of recent advancements in applying deep learning to medical image processing. In the third section, we delve into the materials and methods, starting with a detailed presentation of the "Chest X-ray" dataset. The fourth section showcases the results obtained from deploying this technique to detect Cardiomegaly disease from CXR images.

2. Literature survey

Prior studies have demonstrated significant success rates in exposure. For instance, [9], the authors employed a CNN based on ImageNet to diagnose various diseases from CXR images, achieving an accuracy of 89%. Another study, [10], introduced DualNet, an innovative architecture that processes both the back and front of CXR images [11]. Despite using a large dataset consisting of thousands of simulated images, they achieved a 91% accuracy rate. In [12], the team attained a 92% accuracy rate using established models such as ResNet-101.

A group from Taishan Medical University carried out research, see [13], where they developed an automated method to determine a patient position or body part from digital radiography images using a CNN algorithm. While they relied solely on frequency curve classification and gray matching to achieve their results, they needed over 7 000 images to attain a prediction accuracy of 90%.

3. Material and methods

3.1. Chest X-Ray 8 dataset

The dataset "cardiomegaly-disease-prediction-using-cnn" is processed and taken form the original dataset (CXR8, National Institutes of Health – Clinical Center) of 5552 frontal CXR pictures from distinct people, see Figure 3. We have 4438 training images belonging to 2 classes (True and False Class, Size 128×128 and shape (128,128,1)) and 1114 testing images belonging to 2 classes (True and

False Class, Size 128×128 and shape (128,128,1)). This dataset is publicly available for research and for evaluating various computer-aided detection methods. We use the framework of CNN, ResNet-50, and MobileNet for model construction and evaluation, with the aim of discerning which one yields superior accuracy.

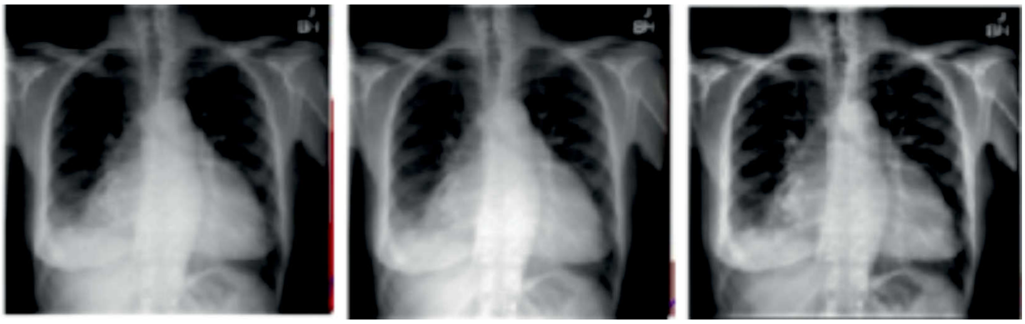


Fig. 3. Sample images of the Dataset X-Ray 8.

3.2. Convolutional neural network (CNN)

As presented in Figure 4, the CNN architecure is initialized using a sequential model.

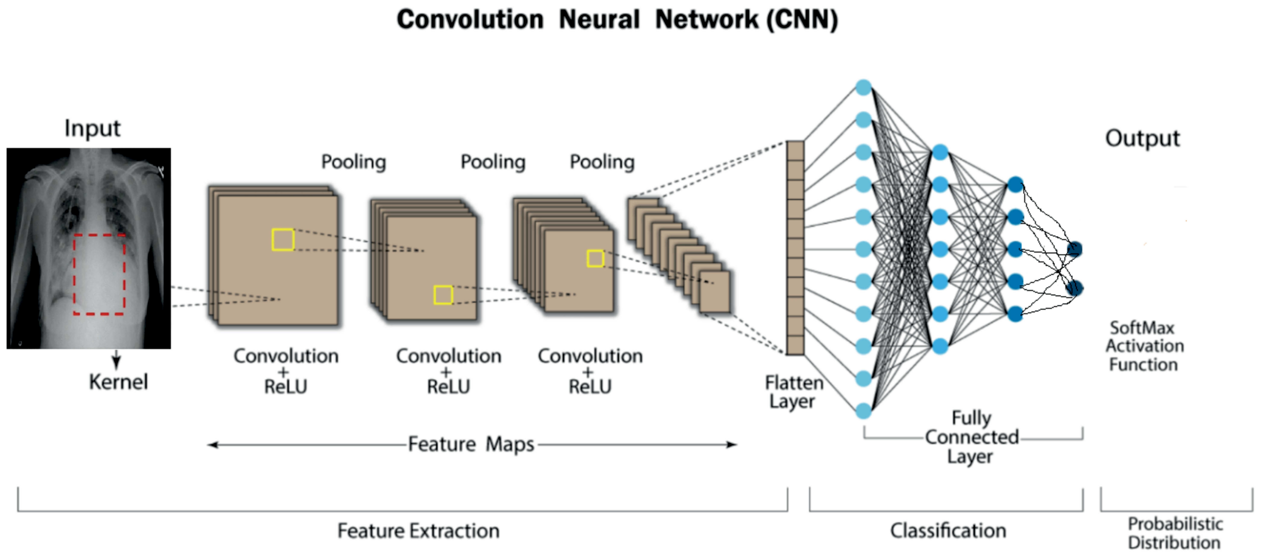


Fig. 4. CNN Architecture.

The first layer is a 2D convolutional layer with 32 filters, a kernel size of 3×3, and employs the 'ReLU' (Rectified Linear Unit) activation function, designed to process input images of shape $128 \times 128 \times 1$. Following this layer, a max-pooling layer is applied to down-sample the feature maps. The next segment of the network consists of another 2D convolutional layer, this time with 16 filters and a 3×3 kernel size, again utilizing the ReLU activation function. This is followed by another max-pooling layer to further reduce the spatial dimensions. Continuing, the network adds yet another convolutional layer with 12 filters and a 3×3 kernel size, paired with the ReLU activation function, succeeded by its respective max-pooling layer.

After processing through the convolutional and pooling layers, the data is flattened to a single dimension, preparing it for the fully connected layers. The network then includes a dense layer with 144 neurons, using the ReLU activation function for non-linearity. This is succeeded by another dense layer comprising 78 neurons, again leveraging the ReLU activation.

The final layer in the architecture is a dense layer with 2 neurons, intended for binary classification. This layer employs the softmax activation function to output probabilities for the two potential classes.

The cost function used in our work is the 'binary-cross entropy' function defined, for N examples, as

$$L = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)\right], \tag{1}$$

where y_i is the true label (0 or 1) for the example i, and $\hat{y}_i$ is the predicted probability of the label being 1, for example i, calculated using weights and biases.

It is the error function that we aim to minimize using gradient descent in binary classification tasks. When training a model with gradient descent (or its variants like the Adam optimizer used in our CNN model), the goal is to find the model parameters (weights and biases) that minimize this error function on the training data. Backpropagation is used to calculate the gradient of the error function with respect to the model parameters, and gradient descent utilizes these gradients to update the parameters in the direction that reduces the error function [14].

The cost function of binary cross-entropy measures the gap between two probability distributions. The actual distribution (the labels) and the distribution predicted by the model (the predicted probabilities). When using gradient descent, the goal is to minimize this loss by adjusting the model weights. At each step, the gradient of the loss function with respect to each weight is calculated. The explicit calculation of gradients for cost functions such as binary cross entropy in a CNN model is performed by the backpropagation process.

3.3. ResNet-50 architecture

In this section, the ResNet-50 architecture is used. It is a popular Convolutional Neural Network (CNN) architecture primarily used for image classification [15]. The main goal of ResNet is to address the vanishing gradient problem in deep networks by introducing residual connections (or skip connections) [16]. These connections allow the gradient to bypass layers directly, thereby facilitating the training of very deep networks. Since ResNet offers higher accuracy on image classification tasks, particularly on datasets such as ImageNet [17], then it used the 'safe duplicate gray channel' function which takes a grayscale image (with a single channel dimension, $(128, 128, 1)$) and duplicates it to obtain an image with three identical channels, i.e. a false color image $(128, 128, 3)$, where the three channels (R,G,B) have the same values as the original grayscale channel, see Figure 5.

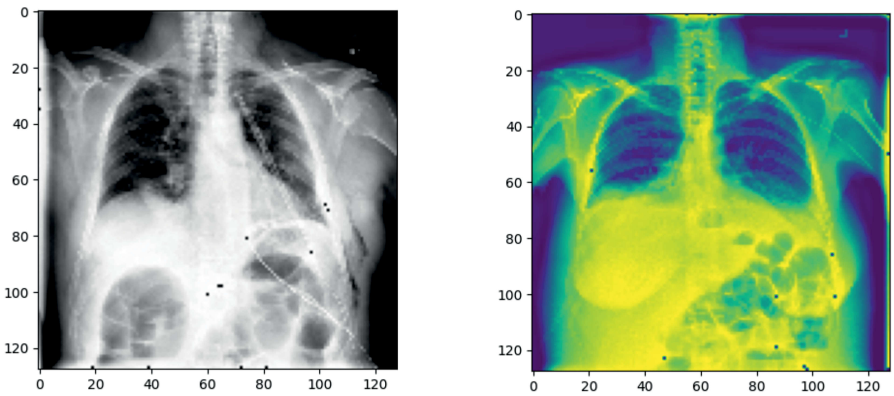


Fig. 5. Safe duplicate gray channel.

It is necessary to use this function to adapt the grayscale images so that they can be used with ResNet (or any other CNN model pretrained on RGB images). The model sets up a neural network model using TensorFlow with the architecture of ResNet-50, which is pre-trained on the ImageNet dataset. This model is tailored for input images of shape $128 \times 128 \times 3$. After utilizing the feature extraction capabilities of the ResNet-50 base, additional custom classification layers are appended: a flattening layer followed by three dense layers with 'relu' activations, ending with a softmax activation for binary classification. The model is compiled using the Adam optimizer with a learning rate of 0.0001, and is set up to train for binary classification. Finally, the structure of the entire model is displayed.

Remark 1. The challenge of vanishing gradients is a prevalent issue encountered in the training of deep neural networks. This phenomenon occurs when the gradients used to update the network weights diminish significantly as they propagate from the output layer to the hidden layers. The consequence of this diminution is a deceleration or even a stagnation in the learning process, given that the weights of the deeper layers undergo only marginal updates. This stagnation impedes the deep layers from effectively learning valuable information and hampers their substantial contribution to the overall network performance. The vanishing gradient problem is particularly pronounced in deep architectures, affecting the network ability to capture and leverage intricate patterns within the data. Strategies such as careful weight initialization, activation function choices, and the use of normalization techniques like batch normalization are often employed to mitigate the impact of vanishing gradients and facilitate more effective training of deep neural networks. Addressing this challenge is crucial for unlocking the full potential of deep learning models and ensuring their capacity to learn intricate representations from complex data.

3.4. MobileNet architecture

We have also used this pre-trained MobileNet model, it uses depth-wise separable convolutions, which break down a standard convolution into a depth-wise convolution (filtering each input channel separately) followed by a point-wise convolution (combining the input channels), see [18]. This significantly reduces the number of required multiplications. It takes as input an image of size $(128, 128, 3)$ because just like other popular architectures, MobileNet is often available with weights pre-trained on the ImageNet dataset, which facilitates transfer learning for new tasks [19].

The code establishes a neural network model utilizing TensorFlow MobileNet architecture, pre-trained on the ImageNet dataset. MobileNet is adapted for input images of shape $128 \times 128 \times 3$ and is configured without its final layer, leveraging average pooling. After the base MobileNet model, custom layers are appended, including two dense layers with 'relu' activations. The final output layer uses a softmax activation, designed for binary classification. The entire architecture is then consolidated into a single model, which is compiled using the Adam optimizer with a learning rate of 0.0001 for binary classification. Lastly, the structure of the model is displayed for review.

4. Results and discussion

This section summarizes the results of the suggested models, which were implemented with Python software.

4.1. CNN model

After compiling the CNN model code using Python, we achieved an accuracy of 0.7325 and a loss value of 0.5294. Notice that the precision is not high (73.25%), but it is acceptable, see Figure 6.

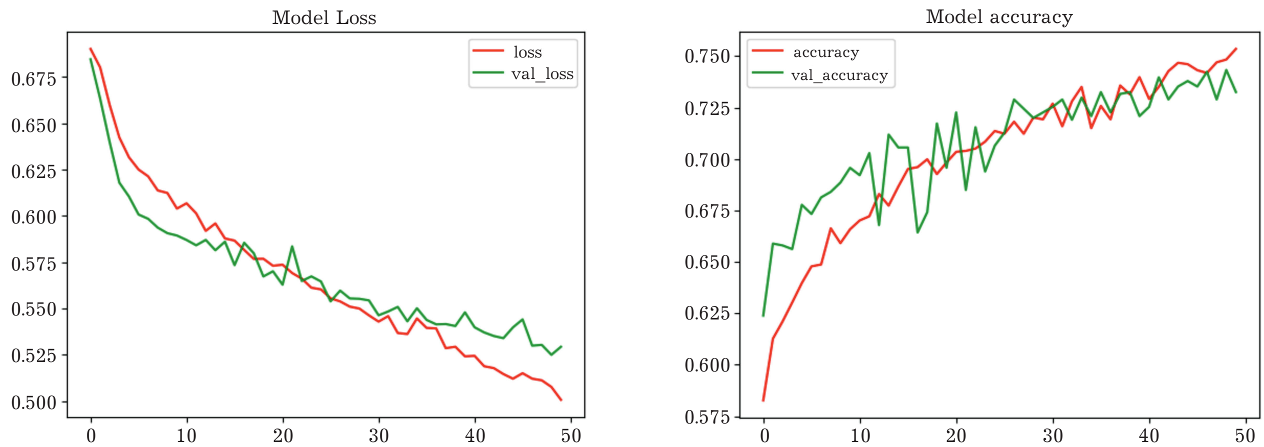


Fig. 6. Model LOSS and accuracy.

To test the performance of our CNN model, we use the confusion matrix. It is a useful tool for assessing the performance of a classification algorithm. It displays the number of true positives, false positives, true negatives, and false negatives that your model has produced, allowing to gain a clearer understanding of where the model is making mistakes, see Figure 7.

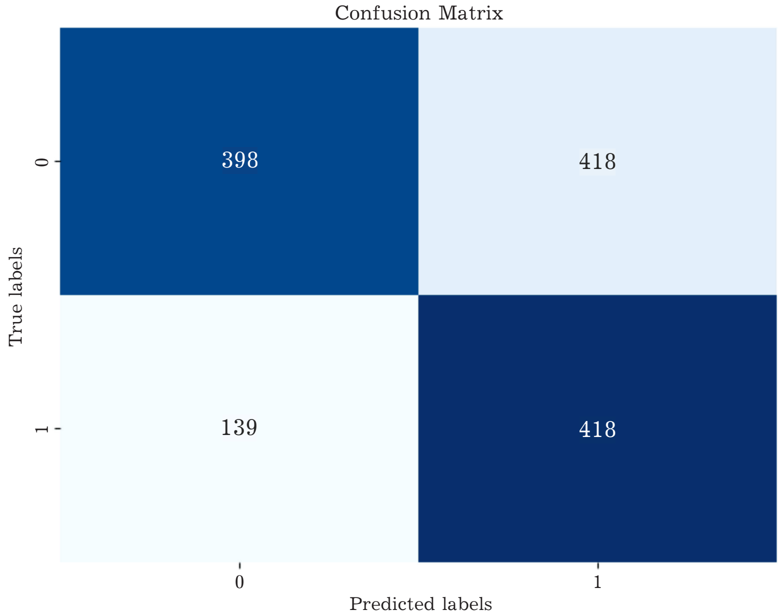


Fig. 7. Confusion Matrix.

Let us break down the information presented in this matrix:

- The vertical axis (y-axis) represents the true labels of the data, while the horizontal axis (x-axis) represents the predicted labels from the model.
- The two classes in this binary classification task are labeled as "0" and "1".
- Here is what each cell in the matrix indicates:
  1. Top-left (398): this is the True Positive (TP) count. It means 398 instances were correctly predicted as class "0" by the model.
  2. Top-right (159): this is the False Negative (FN) count. It means 159 instances were incorrectly predicted as class "1" by the model when they actually belong to class "0".
  3. Bottom-left (139): this is the False Positive (FP) count. It means 139 instances were incorrectly predicted as class "0" by the model when they actually belong to class "1".
  4. Bottom-right (418): this is the True Negative (TN) count. It means 418 instances were correctly predicted as class "1" by the model.

  To further clarify in the context of the disease:
  1. Top-left (398) – True Positives (TP): This means that 398 instances were correctly identified by the model as not having the disease (0 signifies absence of disease).
  2. Top-right (159) – False Negatives (FN): This means that 159 instances were incorrectly identified by the model as having the disease when they do not have the disease.
  3. Bottom-left (139) – False Positives (FP): This means that 139 instances were incorrectly identified by the model as not having the disease when they actually have the disease.
  4. Bottom-right (418) – True Negatives (TN): This means that 418 instances were correctly identified by the model as having the disease (1 signifies presence of disease).

  In this medical context:
  1. True Positives (TP) are cases where the model accurately predicts the absence of the disease.
  2. True Negatives (TN) are cases where the model accurately predicts the presence of the disease.
  3. False Positives (FP) are cases where the model mistakenly predicts the absence of the disease when the person is ill.

4. False Negatives (FN) are cases where the model mistakenly predicts the presence of the disease when the person is not ill.

From this matrix, we can derive several performance metrics like accuracy, precision, recall, and F1-score:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{2}$$

$$\text{Precision(class “0”)} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{3}$$

$$\text{Recall(class “0”)} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{4}$$

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{5}$$

The F1-Score is a harmonic mean of Precision and Recall. It is a single metric that combines both precision and recall into a single value.

4.2. ResNet-50 model

For this model, we obtained an accuracy of 0.8016.

```
35/35 [=====================] - 39s 1s/step - loss: 1.3665 - accuracy: 0.8016
Test accuracy: 80.16%
```

Fig. 8. Test accuracy.

Typically, ResNet tends to deliver superior performance, meaning higher accuracy, on image classification tasks compared to MobileNet. This is especially evident when dealing with datasets like ImageNet [20]. The architecture of ResNet, with its deep layers and residual connections, enables it to capture intricate patterns in the data. While MobileNet is designed for efficiency and speed, especially on devices with limited computational resources, ResNet focuses on achieving state-of-the-art accuracy, even if it requires more computational power and memory. This makes ResNet a preferred choice when accuracy is a top priority and computational resources are abundant.

4.3. MobileNet model

This pre-trained model is designed to be compact with fewer parameters than models such as ResNet, while maintaining acceptable performance. After training this model we obtain an accuracy of 0.7935.

```
35/35 [=====================] - 8s 222ms/step - loss: 0.8608 - accuracy: 0.7935
Test accuracy: 79.35%
```

Fig. 9. Test accuracy.

Although generally less accurate than ResNet for standard image classification tasks, MobileNet is faster in terms of inference time, especially on devices with limited resources.

MobileNet stands out as an appealing choice for developers and researchers in need of a lightweight and fast CNN architecture, yet without sacrificing too much on performance. Designed for efficiency, MobileNet is optimized for mobile and other low-resource devices, making it especially suitable for applications where computational resources or bandwidth are constrained. Its streamlined architecture achieves a good trade-off between computational demand and accuracy, allowing for real-time processing or deployment in edge devices, while still maintaining competitive accuracy on various tasks. Its flexibility in adapting to different resource constraints makes it a versatile tool for a wide range of machine learning applications.

5. Conclusion

Chest radiography is essential for large-scale screening and early detection of lung and heart issues like cardiomegaly, an enlarged heart condition. Automated techniques can be applied in CXR-based tools. With the advancement of deep neural network models, especially CNNs, they have become prominent in image processing. This study explores the potential of using transfer learning to detect cardiomegaly from X-ray images. In this study, we propose a pre-trained supervised neural network model for detecting Cardiomegaly from X-Ray images. The dataset was utilized to evaluate the application of Python on X-ray images. ResNet, when compared to MobileNet, consistently demonstrated higher accuracy in our evaluations. This suggests that the architecture of ResNet, with its unique residual connections, might be better suited for certain image classification tasks. While both models have their own advantages and are designed with different goals in mind, in terms of sheer precision, ResNet seems to outperform MobileNet in our experiments. This observation could be particularly valuable for those looking to prioritize accuracy over computational efficiency in specific use cases.

[1] Candemir S., Rajaraman S., Thoma G., Antani S. Deep learning for grading cardiomegaly severity in chest X-rays: An investigation. 2018 IEEE Life Sciences Conference (LSC). 109–113 (2018).

[2] Sogancioglu E., Murphy K., Calli E., Scholten E. T., Schalekamp S., Van Ginneken B. Cardiomegaly detection on chest radiographs: Segmentation versus classification. IEEE Access. **8**, 94631–94642 (2020).

[3] Zhou S., Zhang X., Zhang R. Identifying cardiomegaly in ChestX-ray8 using transfer learning. MEDINFO 2019: Health and Wellbeing e-Networks for All. **264**, 482–486 (2019).

[4] Yanar E., Hardalaç F. Clinical Decision Support for Early Diagnosis of Cardiomegaly by Using Deep Learning Techniques on Chest X-rays. Computing in Cardiology. 42 (2023).

[5] Hamedian M. Automatic classification of cardiomegaly using deep convolutional neural network (2022).

[6] Grant D., Papież B. W., Parsons G., Tarassenko L., Mahdi A. Deep learning classification of cardiomegaly using combined imaging and non-imaging ICU data. Medical Image Understanding and Analysis: 25th Annual Conference, MIUA 2021, Oxford, United Kingdom, July 12–14, 2021. 547–558 (2021).

[7] Lee M. S., Kim Y. S., Kim M., Usman M., Byon S. S., Kim S. H., Lee B. I., Lee B. D. Evaluation of the feasibility of explainable computer-aided detection of cardiomegaly on chest radiographs using deep learning. Scientific Reports. **11** (1), 16885 (2021).

[8] Kim M.-J., Kim J.-H. Development of convolutional neural network model for classification of cardiomegaly X-ray images. Journal of Mechanics in Medicine and Biology. **22** (08), 2240020 (2022).

[9] Bar Y., Diamant I., Wolf L., Greenspan H. Deep learning with non-medical training used for chest pathology identification. Medical Imaging 2015: Computer-Aided Diagnosis. **9414**, 215–221 (2015).

[10] Rubin J., Sanghavi D., Zhao C., Lee K., Qadir A., Xu-Wilson M. Large scale automated reading of frontal and lateral chest X-rays using dual convolutional neural networks. Preprint arXiv:1804.07839 (2018).

[11] Chen L., Mao T., Zhang Q. Identifying cardiomegaly in chest X-rays using dual attention network. Applied Intelligence. **52** (10), 11058–11067 (2022).

[12] Rajaraman S., Kim I., Antani S. K. Detection and visualization of abnormality in chest radiographs using modality-specific convolutional neural network ensembles. PeerJ. **8**, e8693 (2020).

[13] Bar Y., Diamant I., Wolf L., Lieberman S., Konen E., Greenspan H. Chest pathology identification using deep feature selection with non-medical training. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization. **6** (3), 259–263 (2018).

[14] Calin O. Deep Learning Architectures – A Mathematical Approach. Springer Series in the Data Sciences, Springer (2020).

[15] Wu J. X., Pai C. C., Kan C. D., Chen P. Y., Chen W. L., Lin C. H. Chest X-ray image analysis with combining 2D and 1D convolutional neural network based classifier for rapid cardiomegaly screening. IEEE Access. **10**, 47824–47836 (2022).

[16] Kim M. J., Kim J. H. Proposal of a Convolutional Neural Network Model for the Classification of Cardiomegaly in Chest X-ray Images. Journal of the Korean Society of Radiology. **15** (5), 613–620 (2021).

[17] Yoo H., Han S., Chung K. Diagnosis support model of cardiomegaly based on CNN using ResNet and explainable feature map. IEEE Access. **9**, 55802–55813 (2021).

[18] Junior J. R. F., Cardenas D. A. C., Moreno R. A., de Sá Rebelo M. d. F., Krieger J. E., Gutierrez M. A. A general fully automated deep-learning method to detect cardiomegaly in chest X-rays. Medical Imaging 2021: Computer-Aided Diagnosis. **11597**, 537–542 (2021).

[19] Jeong W. Y., Kim J. H. A Study on the Improvement of Accuracy of Cardiomegaly Classification Based on InceptionV3. Journal of Biomedical Engineering Research. **43** (1), 45–51 (2022).

[20] Bougias H., Georgiadou E., Malamateniou C., Stogiannos N. Identifying cardiomegaly in chest X-rays: a cross-sectional study of evaluation and comparison between different transfer learning methods. Acta Radiologica. **62** (12), 1601–1609 (2021).

Порівняння деяких архітектур CNN для виявлення кардіомегалії на знімках грудної клітки в X-променях

Реграгуі Х.[1], Зіані М.[1], Муссаїд Н.[2]

[1] *Лабораторія LMSA, кафедра математики, факультет природничих наук, Університет Мухаммеда V у Рабаті, Марокко*

[2] *LMCSA, Університет Хасана II Касабланки, п.с. 146, Мохаммедіа, Марокко*

В аналізі медичних зображень глибоке навчання та згорткові нейронні мережі (CNN) широко використовуються, зокрема в таких завданнях, як класифікація та сегментація. У цьому дослідженні конкретно розглядається їх застосування в галузі охорони здоров'я для виявлення кардіомегалії – стану, що характеризується збільшенням серця, часто пов'язаного з такими факторами, як гіпертонія або захворювання коронарних артерій. Основною метою є розробка алгоритму, здатного ідентифікувати кардіомегалію на рентгенівських знімках грудної клітки, що становить проблему бінарної класифікації (незалежно від того, чи є на зображенні кардіомегалія чи ні). Використовуючи набір даних CXR8 Клінічного центру Національного інституту охорони здоров'я, який містить 2 776 зображень кардіомегалії та 60 361 зображення без її виявлення, вхідні дані є зображеннями з мітками, а вихідні дані є відповідними мітками (кардіомегалія або відсутність її виявлення). Використовуючи бібліотеки Keras та TensorFlow Python, виконана спроба побудувати модель CNN, яка відмінно підходить для бінарної класифікації, розрізняючи кардіомегалію та її відсутність знімках грудної клітки в X-променях.

Ключові слова: *глибоке навчання; згорткові нейронні мережі; кардіомегалія; бінарна класифікація.*