

Article

Comparative Analysis of Conventional CNN v's ImageNet Pretrained ResNet in Medical Image Classification

Christos Raptis ^{1,2}, Efstratios Karavasilis ¹, George Anastasopoulos ³  and Adam Adamopoulos ^{1,2,*} 

¹ Medical Physics Laboratory, Department of Medicine, Democritus University of Thrace, 681 00 Alexandroupolis, Greece; craptis@med.duth.gr (C.R.); ekaravas@med.duth.gr (E.K.)

² School of Science and Technology, Hellenic Open University, 263 31 Patra, Greece

³ Medical Informatics Laboratory, Department of Medicine, Democritus University of Thrace, 681 00 Alexandroupolis, Greece; anasta@med.duth.gr

* Correspondence: adam@med.duth.gr

Abstract: Convolutional Neural Networks (CNNs) are the prevalent technology in computer vision and have become increasingly popular for medical imaging data classification and analysis. In this field, due to the scarcity of medical data, pretrained ResNets on ImageNet can be considered a suitable first approach. This paper examines the medical imaging classification accuracy of conventional basic custom CNNs compared to ImageNet pretrained ResNets on various medical datasets in an effort to give more information about the importance of medical data and its preprocessing techniques for disease studies. Microscope-extracted cytological images were examined along with chest X-rays, MRI brain scans, and melanoma photographs. The medical images were examined in various sets, class combinations, and resolutions. Augmented image datasets and asymmetrical training and validation splits among the classes were also examined. Models were developed after they were tested and fine-tuned with respect to their network size, parameter values and network methods, image resolution, size of dataset, multitude, and genre of class. Overfitting was also examined, and comparative studies regarding the computational cost of different models were performed. The models achieved high accuracy in image classification that varies depending on the dataset and can be easily incorporated in future over-the-internet medical decision-supporting (telemedicine) environments. In addition, it appeared that conventional basic custom CNN overperformed ImageNet pretrained ResNets. The obtained results indicate the importance of utilizing medical image data as a testbed for improvements in CNN classification performance and the possibility of using CNNs and data preprocessing techniques for disease studies.

Keywords: medical image classification; CNN; pretrained ResNet



Citation: Raptis, C.; Karavasilis, E.; Anastasopoulos, G.; Adamopoulos, A. Comparative Analysis of Conventional CNN v's ImageNet Pretrained ResNet in Medical Image Classification. *Information* **2024**, *15*, 806. <https://doi.org/10.3390/info15120806>

Academic Editors: Francesco Beritelli and Xiaoshuang Shi

Received: 20 August 2024

Revised: 4 November 2024

Accepted: 10 December 2024

Published: 14 December 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Medical imaging data are crucial for identifying health issues and monitoring therapy progress. Radiologists, cytologists, and other medical professionals are responsible for interpreting these images. However, time constraints and inherent human limitations [1,2] present significant challenges. To address these issues, the emerging field of deep learning in computer vision is developing models for medical image classification and segmentation [3–6].

Previous research has demonstrated the effectiveness of deep learning in this area, with one study using ResNet50 on histopathology images for breast cancer classification, achieving a high accuracy of 99.24% [7]. Similarly, another study reported a high accuracy of 99.17% using ResNet50 trained on chest X-rays for COVID-19 classification [8]. This paper also investigates the accuracy performance of various data types using basic custom CNNs, comparing their performance with that of pretrained ResNet models (with fixed weights). Specifically, we evaluate a straight-structure residual network (ResNet50), an inverted residual structure (MobileNetV2), and conventional custom CNNs.

An emerging question is how the different categories of medical images affect the classification accuracy of CNNs and ImageNet pretrained ResNets. Also, the number of images and variety of classes are examined. Different optimizers and learning rates are also tested, along with various layer configurations and pooling methods. Additionally, we evaluate data preprocessing techniques to assess their impact on the model's classification performance and, more importantly, to uncover insights into the nature of the disease as depicted in medical images. Finally, we analyze the computational costs associated with using two different processors: CPU and GPU.

2. Materials and Methods

2.1. Materials

To compare the different methods, data from four medical image categories are examined: microscope-extracted cytological images, chest X-rays (adult and pediatric), MRI brain scans, and melanoma photographs (Table 1).

2.1.1. Cytological Images

The cytological images that were used to train the models were low-resolution (100×120 p) images extracted from pap smear tests [9]. Each of the images contains a single cell from the following five categories:

Koilocytotic cells indicate an HPV infection and can be found alone or in small groups. They have big and empty perinuclear space and dense basophilic or eosinophilic cytoplasm. The nucleus is double or triple the size of the ordinary nucleus.

Dyskeratotic cells have undergone an abnormal process of keratinization. They are small cells with irregular shapes and big irregular nuclei.

Metaplastic cells are round and polygonic cells with dense double-color cytoplasm and round-centered nucleus.

Parabasal are the smallest epithelial cells in a test pap. They are round, and they have a big nucleus-to-cytoplasm ratio (N:C ratio). Cytoplasm is dense, and the nucleus is wave-like with thin, granular chromatin.

Intermediate cells are two to three times bigger than Parabasal cells but with nuclei of the same size. Small intermediate cells are almost cyclical with big nuclei. Large intermediate cells or superficial intermediate cells have a polygonic shape and a small nucleus-to-cytoplasm ratio (N:C ratio). Superficial (dead cells) are the largest cells in a pap smear test. They have a polygonic shape and a small dark nucleus or no nucleus at all. Cytoplasm can have vacuoles as the cell ages [10].

Trained cytologists can classify the classes with very high accuracy, but the process is laborious and time-consuming. The basic custom CNN model that was developed can classify these five classes with high accuracy (~94%) and with a very fast speed (~5 ms/image). Also, the model seems to perform well on the unseen data we tested, like the ones in Figure 1, classifying the images correctly with high certainty (>95%).

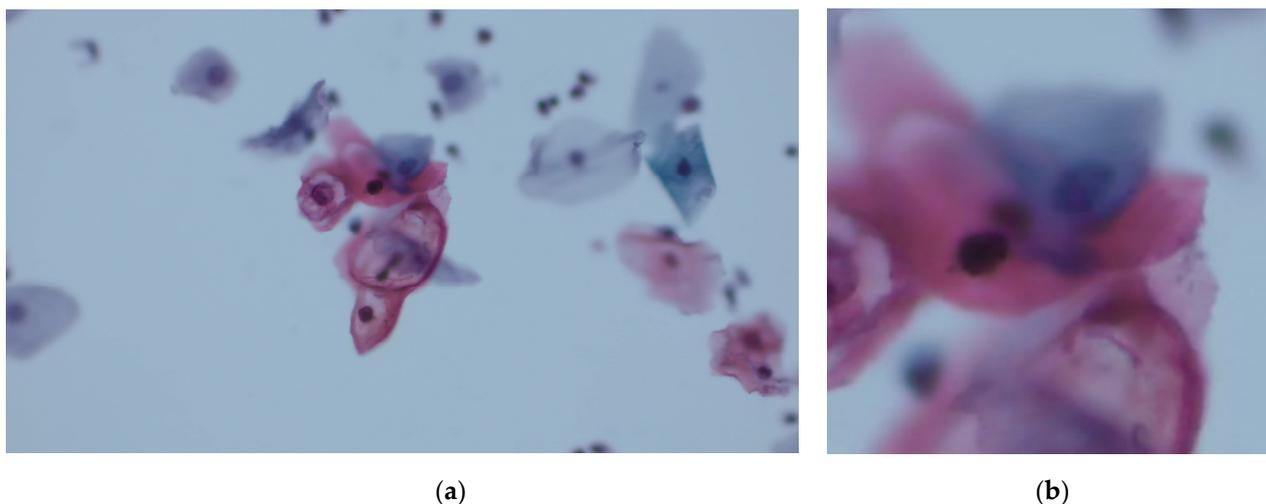


Figure 1. Pap smear test: (a) koilocytotic cells (original size); (b) koilocytotic cells (zoomed and cropped). Images from Axioscope 5 microscope camera AxioCam 208 Color.

2.1.2. Chest X-Rays

Pneumonia is a general term used to describe an infection that inflames the air sacs of the lungs. It is usually caused by the fluid or purulent material that fills the air sacs. Bacterial pneumonia is a type of serious pneumonia caused by the development of bacteria in the lungs. The most common are *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Staphylococcus aureus*. These infections often happen after a cold or the flu when the immune system is weakened. Viruses that infect the respiratory system are influenza A, B, rhinoviruses, and the respiratory syncytial virus RSV [11,12].

On X-rays, dark areas represent areas full of air, and gray areas suggest subcutaneous tissue or fat. Light gray represents soft tissues like the heart and the blood vessels. The white areas show bones, and the bright white areas suggest metallic objects like pacemakers or defibrillators.

Viral pneumonias show seasonal distribution and occur mostly in the winter and spring. In most cases, chest X-rays, as interpreted by medical experts, cannot classify viral to bacterial pneumonia with absolute certainty [13].

An example that shows the difficulty (even for the trained professionals) of classification between viral and bacterial pneumonia is given in Figure 2.

To many professionals, these chest X-rays indicate viral infections, but the top center and bottom left X-rays are, in fact, from adults that have bacterial pneumonia. In this case, the developed custom CNN models can correctly capture the difference.

To test the model on X-rays, four different datasets were used. Two datasets with 768×1024 p image resolution and two classes (pneumonia and healthy), one with adult X-rays [14], one with pediatric X-rays [15], and two with three (bacterial pneumonia, viral pneumonia, healthy) and four classes (bacterial pneumonia, viral pneumonia, COVID19, healthy) of low to medium resolution (300×400 p) [16].

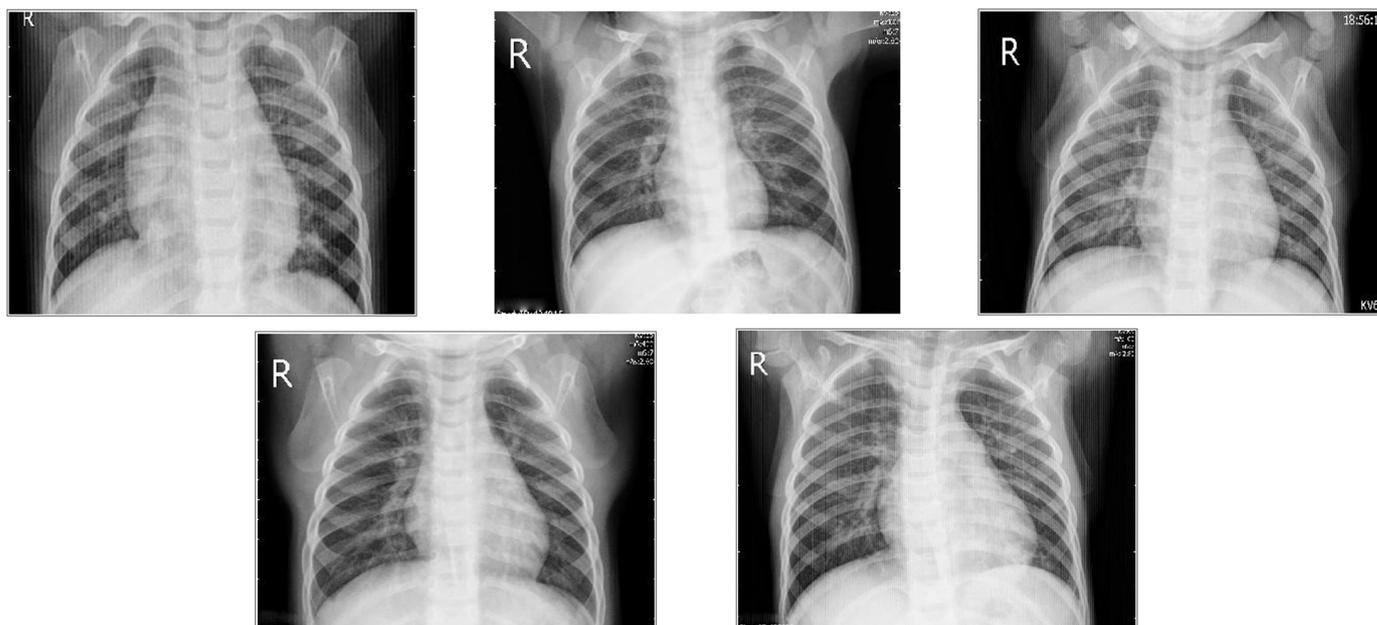


Figure 2. Adult chest X-rays. From top to bottom and left to right: 1,3,5 viral pneumonia; 2,4 bacterial pneumonia. Images from the adult chest X-ray dataset that was used to train the models.

2.1.3. MRI Brain Scans

Magnetic Resonance Imaging (MRI) is used to examine almost any part of the body like the brain, the spinal cord, bones, breasts, internal organs, heart, and blood vessels. MRI provides better contrast in images of soft tissues, and for that reason, it is commonly used to scan the brain. MRI is a safe method for brain imaging, diagnosing tumors, aneurysms, epilepsy, dementia, encephalitis, etc. [17].

The pituitary tumor occurs in the anterior body of the pituitary gland [18].

Meningioma is a tumor of leptomeningeal origin that develops on the brain's meninges, the three layers of membrane that cover the brain and spinal cord [19].

Gliomas (Figure 3) are tumors out of various types of glial cells and form in the central nervous system (brain or spinal cord) and peripheral nervous system [20].

To test the MRI scans, the dataset of medium resolution (300×300 p) images [21] was tested in 4-class ('Normal', 'glioma_tumor', 'meningioma_tumor', 'pituitary_tumor') and 2-class ('Tumor', 'Normal') combinations.

The augmentation technique was also tested with a pre-augmented dataset. The augmented dataset of the 4-class MRI brain scans related to tumors ('Normal', 'glioma_tumor', 'meningioma_tumor', 'pituitary_tumor') were created by the following techniques: (a) random noise insertion (salt and pepper noise) setting pixels to 0 or 255; (b) histogram equalization to enhance the contrast and image details; (c) clockwise and counterclockwise rotation on a certain angle; (d) brightness adjustment (adding or removing intensity from pixels); (e) horizontal and vertical flipping of the image.

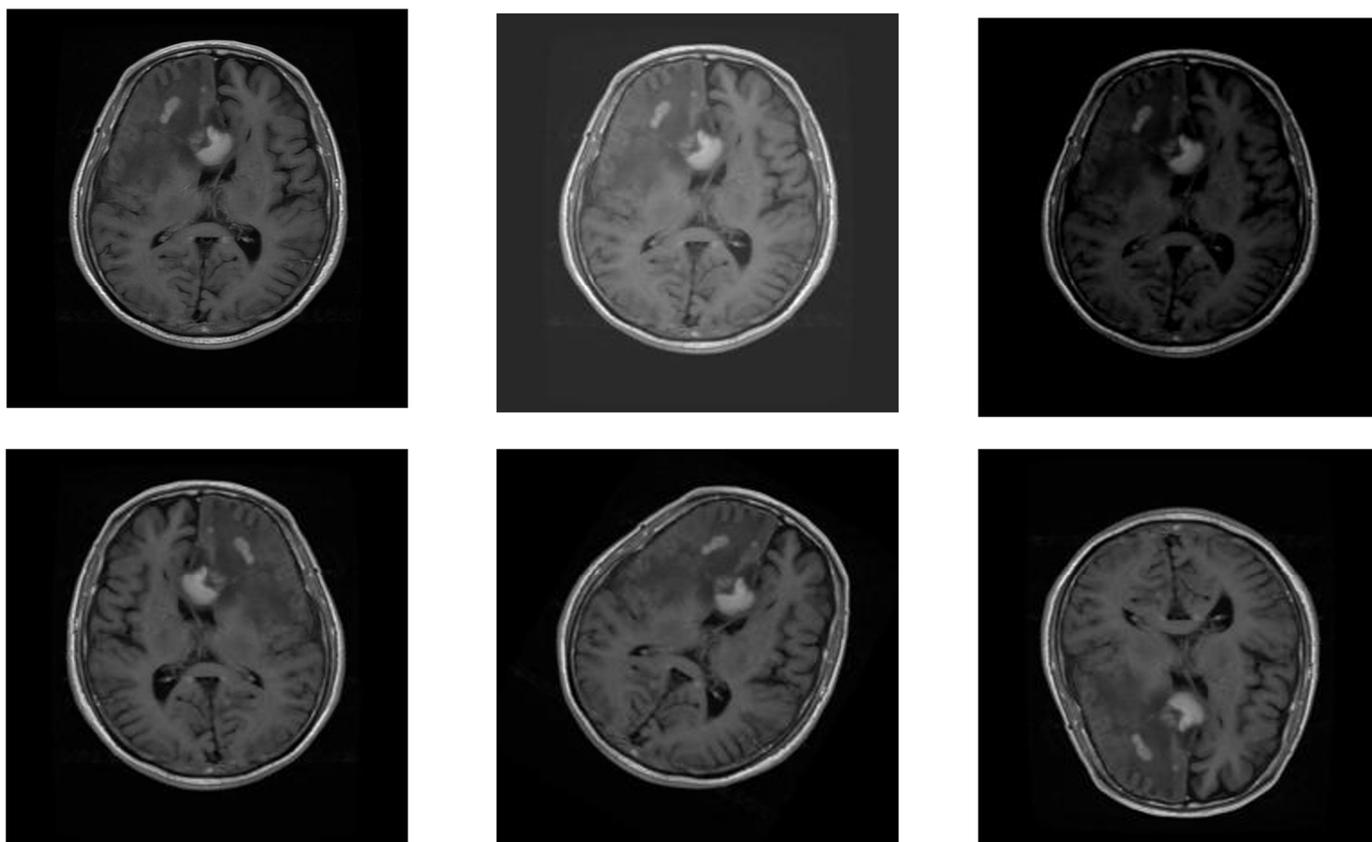


Figure 3. From top to bottom and left to right: original image; brighter; darker; horizontal flip; rotated; vertical flip (Glioma). Glioma tumor image from the dataset used to train the models.

2.1.4. Melanoma Photographs

Melanoma (malignant melanoma) is a kind of skin cancer that starts in the melanocytes. Melanoma's early symptoms are usually a change in the mole or the development of an unusual pigment with irregular shapes and different colors. The melanoma features that medical professionals look for when diagnosing and classifying melanomas are known as ABCDE [22,23]. ABCDE stands for asymmetry, border, color, diameter, and evolving:

Asymmetry: Refers to the uniformity of its shape. Benign melanomas (moles) are typically uniform and symmetrical.

Border: Refers to the irregularity of its borders. Benign melanomas usually have well-defined borders.

Color: Malignant melanoma lesions are often more than one color or shade. Benign melanomas are typically one color.

Diameter: Malignant melanoma growths are mostly larger than 6 mm in diameter.

Evolving: Malignant melanoma tends to change over time in size, shape, or color.

The fact that melanoma data (Figure 4) can be extracted by simpler means drove researchers recently to examine them for telemedicine applications [24]. Telemedicine is being used in health centers and medical offices in small towns, villages, and rural areas as a means for people to have access to more specialized healthcare when visiting these primary care units. Also, some clinics use telemedicine to monitor patients via virtual visits. During the COVID-19 pandemic, telemedicine saw a significant rise, and many people still use it [25]. Telemedicine is expected to be more popular as more people become familiar with new technologies and as the hardware becomes better. Internet of Things devices and wearable devices nowadays can automatically record and send data, such as blood pressure, blood sugar and oxygen levels, heart rate, physical activity, or sleep.

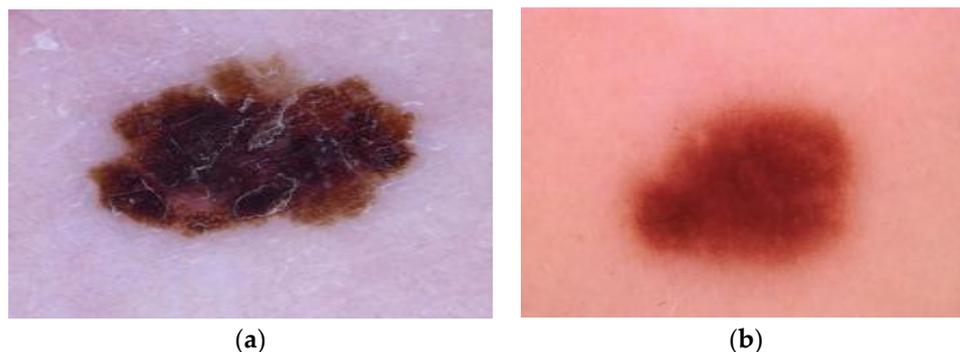


Figure 4. (a) Malignant melanoma; (b) benign melanoma. Melanoma photographs from the dataset used to train the models.

To train the melanoma model, a set of 2 classes (Benign and Malignant) of low resolution (224×224 p) was used [26].

Datasets are summarized in the following table:

Table 1. Datasets.

Dataset	Modalities	Number of Images	Number of Classes	Image Resolution
Melanoma	Photographs	13,869	2	224×224
Pap smear test	Microscope Photographs	1652	2	100×120
		4039	5	100×120
Pneumonia	Chest X-rays	5850	2	$768 \times 1024, 500 \times 700$
		5847	2	$768 \times 1024, 500 \times 700$
		9203	4	300×400
		7924	3	300×400
Brain tumors	MRI Brain scans	7019	4	300×300
		7019	2	300×300

2.2. Methods

2.2.1. Custom CNN

Digital images usually have 8-bit depth on each channel (color/RGB). That means the pixels range between 0 (no light) and 255 (maximum brightness) on each channel. For example, (0,0,0) is black, (255,255,255) is white, (255,0,0) is red, (0,255,0) is green, (0,0,255) is blue, (255,255,0) is yellow, and so on. Another important factor is image resolution, which is the number of pixels in each dimension. This parameter is usually changed by downsizing/downscaling the image. Another preprocessing technique is applying a filter, for example, a Gaussian filter that removes noise [27].

Basic CNN architecture (15 to 19 total layers)

The basic architecture was developed on the tensorflow platform as a keras sequential model that groups a linear stack of layers into a model. On input, the 8-bit depth images were rescaled by dividing by 255 so that numbers fall between 0 and 1, which is more ‘natural’ for the neural networks. After the rescaling layer, 5 to 7 convolutional layers were added with a number of size- 3×3 filters ranging progressively from 16 to 128 to extract the features. The padding technique was applied to retain the image size. After filtering, a ReLU activation function was applied to the feature map. Each convolutional layer was followed by a corresponding pooling layer that reduces the dimensionality of the feature maps. The pooling layers used the max pooling method, except for the last pooling

layer, which used average pooling. After the average pooling layer, a dropout layer, which randomly sets inputs to 0, was added to reduce the overfitting phenomenon, and then a flattened layer was added to spread the data in one dimension. One-dimensional data were connected to the fully connected (dense) layer of 128 nodes, followed by a dense layer with as many nodes as the number of classes.

Loss functions calculate how far the predicted class is from the actual class.

$$Loss = |Y_{predicted} - Y_{actual}|$$

The model was then improved by calculating and trying to minimize that loss. The validation-loss curve, along with the validation accuracy, provided information about the performance of the model.

The two most important loss functions and the ones we tested are as follows:

Binary Cross Entropy

Binary Cross Entropy is used for two-class data and is mathematically expressed with the following equation:

$$h_p(q) = -\frac{1}{N} \sum_{i=1}^N y_i \log(p(y_i)) + (1 - y_i) \log(1 - p(y_i)) \quad (1)$$

where y is the class (1 or 0), $p(y_i)$ is the possibility that the data belong in class 1, $(1 - p(y_i))$ is the possibility of class 0. The possibilities of the two classes always add to 1.

Binary Cross Entropy has a natural synergy with the sigmoid activation function. This happens because the sigmoid gives the output values strictly between 0 and 1.

(Sparse) Categorical Cross entropy

Categorical Cross entropy (or Softmax loss) is a Softmax function along with the cross entropy loss function and is used for multiclass models.

Cross entropy is expressed as follows:

$$Cross\ Entropy = -\sum_{i=1}^C t_i \log(s_i) \quad (2)$$

Using this loss function, we can obtain a vector that contains possibilities for all the classes. These possibilities do not add to 1 like in Binary Cross Entropy.

Sparse categorical cross entropy is the same as categorical cross entropy, with the only difference being the way that class labels are expressed. For example, on categorical cross entropy, we have (for three classes) labels [1,0,0], [0,1,0], [0,0,1], while on sparse categorical cross entropy, we have [1–3].

If the network makes a wrong prediction with high certainty, sigmoid (or Softmax) gives a number close to 0 (or to 1), and thus Cross entropy becomes high.

Optimizers are algorithms for calculating and updating network parameters like weights and the learning rate, with the purpose of reducing the cost.

Two often used optimizers, which are also the ones we tested, are as follows:

Adam and AdamW

Adam (Adaptive Momentum) [28] (see Appendix A) combines ideas from RMSProp and Gradient Descent with Momentum algorithms.

Adam converges fast and works well with big datasets and noisy data but needs more parameter tuning than other algorithms [28]. AdamW is a variant of Adam that uses a method for decaying weights that decouples the weight decaying from the learning rate to increase the generality of the models [29].

The models were compiled with Adam (and its variant AdamW) optimizer with learning rates ranging between 8×10^{-5} and 3×10^{-4} . The loss function for the multiclass models was sparse categorical cross entropy, and the final activation function was Softmax.

For the two-class models, Binary Cross Entropy with Sigmoid was also tested. The metric used to evaluate the models was accuracy. Accuracy is expressed as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

where TP = true positives; TN = true negatives; FP = false positives; FN = false negatives, e.g., for the melanoma:

Melanoma	Predicted Value	
	Malignant	Benign
Actual Value		
Malignant	TP	FN
Benign	FP	TN

An example of a basic custom CNN with 19 layers (total) is the following:
CNN19

```
model = tf.keras.Sequential([
    tf.keras.layers.Rescaling(1./255),
    tf.keras.layers.Conv2D(16, 3, activation = 'relu', padding='same'),
    tf.keras.layers.MaxPooling2D(),
    tf.keras.layers.Conv2D(32, 3, activation = 'relu', padding='same'),
    tf.keras.layers.MaxPooling2D(),
    tf.keras.layers.Conv2D(32, 3, activation = 'relu', padding='same'),
    tf.keras.layers.MaxPooling2D(),
    tf.keras.layers.Conv2D(32, 3, activation = 'relu', padding='same'),
    tf.keras.layers.MaxPooling2D(),
    tf.keras.layers.Conv2D(64, 3, activation = 'relu', padding='same'),
    tf.keras.layers.MaxPooling2D(),
    tf.keras.layers.Conv2D(64, 3, activation = 'relu', padding='same'),
    tf.keras.layers.MaxPooling2D(),
    tf.keras.layers.Conv2D(64, 3, activation = 'relu', padding='same'),
    tf.keras.layers.AveragePooling2D(),
    tf.keras.layers.Dropout(0.2),
    tf.keras.layers.Flatten(),
    tf.keras.layers.Dense(128, activation='relu'),
    tf.keras.layers.Dense(num_classes)
])
```

2.2.2. Pretrained ResNets

Other researchers have previously shown the potential of ResNet50 in medical imaging [7,8,30–33]. Behar et al. reached 99.24% on a 2-class breast cancer model. However, their model was a fully retrained ResNet50 and had unbalanced classes of histopathological images (2480 benign and 5429 malignant). Hossain et al. also reached 99.17% accuracy with retrained ResNet50 on chest X-rays for COVID-19 classification.

In this paper, the ImageNet pretrained ResNets had their layers frozen (fixed weights), and a new dense layer was added as a classifier and trained on the new medical imaging data. This is the basic strategy of transfer learning. They were also tested with pre-augmented data produced by a random horizontal flip rotation.

Computational costs were examined for each model on CPU (Windows 11) and GPU (Ubuntu). The training speeds of the models were measured on a modern conventional desktop computer (see Appendix B) in order to provide a benchmark for comparison.

3. Results

3.1. Model and Data Setup

The models demonstrated high accuracy, achieving 94.3% on the 5-class classification, even when tested with external data sourced from ultra-high-resolution images that were subsequently downsized to a lower resolution. Specifically, we used test images of pap plaques captured with a microscope camera at a resolution of 3840×2160 . These plaques consisted of two types: those with superficial intermediate cells and those with koilocytotic cells. The images were then resized to 120×100 pixels, yet the CNN model classified them correctly with high confidence. Notably, the model accurately identified koilocytotic cells even when they were located in a small region of the image, demonstrating the strength of CNNs in local detection.

For the high-resolution dataset of chest X-rays, the models achieved over 96% accuracy in distinguishing between pneumonia and normal cases. Similar results were observed in pediatric X-rays, which involved the conditions of Pediatric Pneumonia and Pediatric Normal, also at a resolution of 768×1024 pixels. This indicates that variations in child anatomy do not significantly impact performance, as the CNN models maintained high validation accuracy across both datasets.

In the 4-class chest X-ray dataset (comprising COVID-19, bacterial pneumonia, viral pneumonia, and normal), validation accuracy reached 86.6% at a resolution of 300×400 pixels. The same dataset was also tested using various 3-class combinations. The subset consisting of bacterial pneumonia, viral pneumonia, and normal cases showed the same validation accuracy as the original dataset, indicating that the models struggle more with distinguishing between the two types of pneumonia. For the other 3-class combinations, validation accuracy ranged from 96% to 98%. Finally, in the 2-class classification of COVID-19 versus normal, the model achieved an impressive validation accuracy of 99.5%. To gain deeper insight into the challenges associated with the two types of pneumonia, we further divided the dataset into two sub-datasets:

- A. Healthy and viral pneumonia, for which we achieved 97.24% validation accuracy;
- B. Healthy and bacterial pneumonia, for which we achieved 97.99% validation accuracy.

We then preprocessed the images of both sub-datasets with the following two techniques, which we imposed separately only on the training sets:

1. We imposed a random rotation on each training set with angles ranging from -40% of 2π to 40% of 2π with equal chance. Accuracy dropped to 94.18% and 95.28%, respectively.
2. We imposed a random—with a 50% chance—horizontal flip on each training set. Accuracies dropped to 96.28% and 97.53%, respectively.

From these results, one can suggest the importance of the lungs and the patterns that appear on them for these diseases (which, of course, is something to be expected). Also, viral pneumonia images may possibly contain more “global” patterns. Another possible indication is that viral pneumonia images contain more local inbetween/interlobal patterns, which get lost when a horizontal flip is applied (bigger val_accuracy % loss on dataset A after technique number 2) and/or bacterial pneumonia images contain more local lobal patterns.

These are possible explanations because, by core design, CNNs with their convolutional filters are decisive mainly because of local patterns. This characteristic has been previously shown: “a CNN’s decision is based mainly on recognizing local-patterns without taking into account their relative positions” [34] In order to evaluate these possibilities, further research on more datasets is needed. In cases of specific diseases that are not fully understood by humans from their medical images, CNN models and data preprocessing techniques could be examined as a possible means for more insight into these diseases.

On the 4-class MRI brain scan data that are tumor-related, on a small to medium dataset size with 300×300 -pixel resolution, we obtained a 97.29% validation accuracy. On the same dataset, after the (premade) augmentation preprocessing was applied, the same vali-

dation accuracy was achieved in only 10 epochs and a lesser resolution of 224×224 pixels. Padding on these images showed no significant effect on the validation accuracy of the models.

Adam optimizer and its variant AdamW had the best validation accuracy and best performance overall in all models.

On all CNN models, Adam and AdamW performed about the same, with AdamW being a bit smoother on loss convergence while being a bit slower. The successful learning rates were close to 10^{-4} , and the total number of layers was close to 17, depending on the dataset. Average pooling on the last convolutional layer gave better results than maxpooling. Binary Cross Entropy with sigmoid function gave about the same results on two-class classification with sparse categorical cross entropy/Softmax.

Fixed-weight pretrained ResNets performed very well overall but were slightly behind the basic custom CNNs. Random horizontal flip did not have a significant effect on the validation accuracy of these datasets. By reading the curves, the deeper residual networks had better convergence, while the custom CNNs had to be stopped early to reduce overfitting.

3.2. Model Architecture and Curve Results

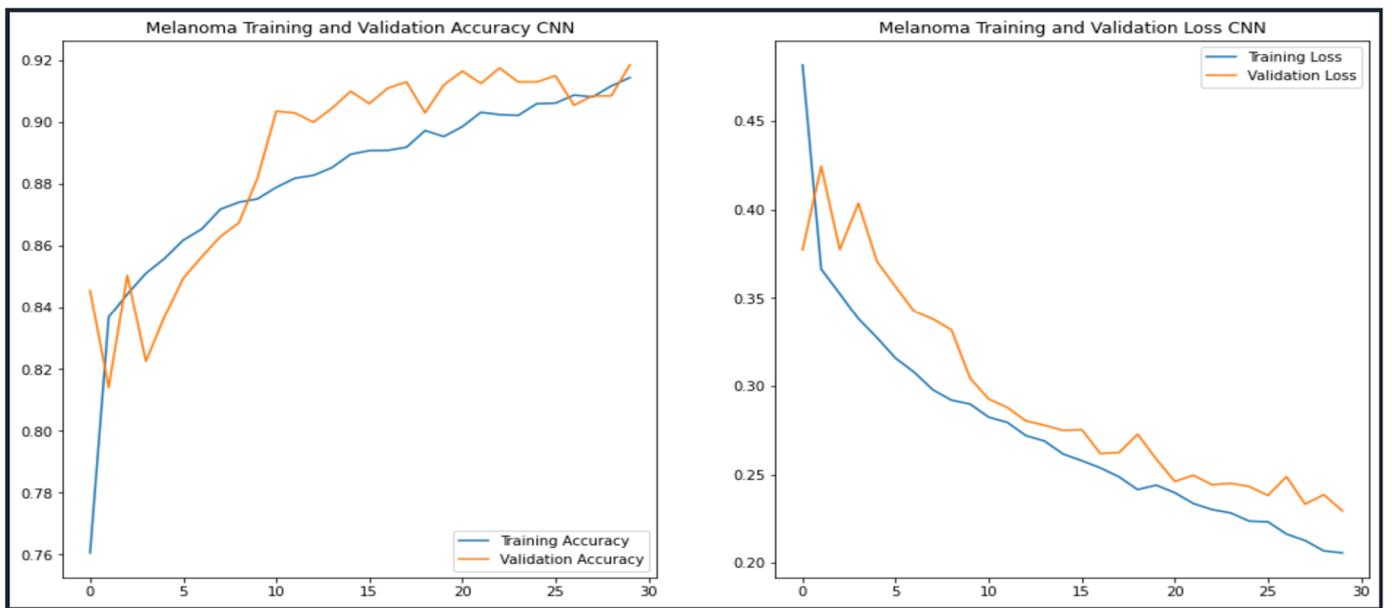
3.2.1. Melanoma Classification Model

For melanoma classification, a basic model with 15 layers (total) was developed. Various optimizers and learning rates were tested (Table 2). Adam, with a 3×10^{-4} learning rate, had the best results with 93.05% classification accuracy (Figure 5). Binary Cross Entropy and sigmoid had roughly the same results as sparse categorical cross entropy/Softmax.

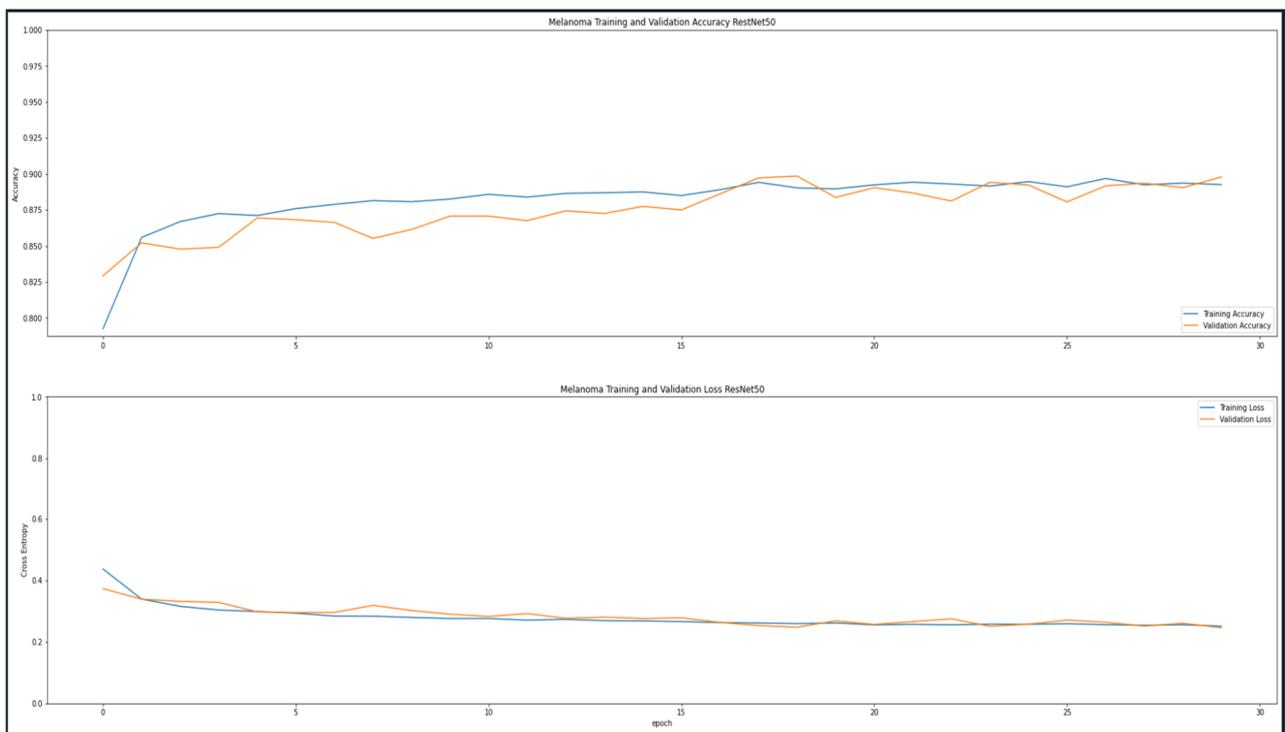
Table 2. Melanoma photographs: 15-layer conventional CNN model architecture (SparseCategorical-CrossEntropy/no padding).

Layer (Type)	Output Shape	Parameters #
Rescaling	(None, 224, 224, 3)	0
Conv2d	(None, 222, 222, 16)	448
Max Pooling	(None, 111, 111, 16)	0
Conv2d	(None, 109, 109, 32)	4640
Max Pooling	(None, 54, 54, 32)	0
Conv2d	(None, 52, 52, 32)	9248
Max Pooling	(None, 26, 26, 32)	0
Conv2d	(None, 24, 24, 64)	18,496
Max Pooling	(None, 12, 12, 64)	0
Conv2d	(None, 10, 10, 64)	36,928
Average Pooling	(None, 5, 5, 64)	0
Dropout	(None, 5, 5, 64)	0
Flatten	(None, 1600)	0
Dense	(None, 128)	204,928
Dense	(None, 2)	258
Total parameters	274,946 (1.05 MB)	
Trainable parameters	274,946 (1.05 MB)	
Non-trainable parameters	0 (0.00 B)	

ResNets had worse results: 89.79% (ResNet50) and 86.39% (MobileNetV2). ResNets converged faster than the custom CNN model.

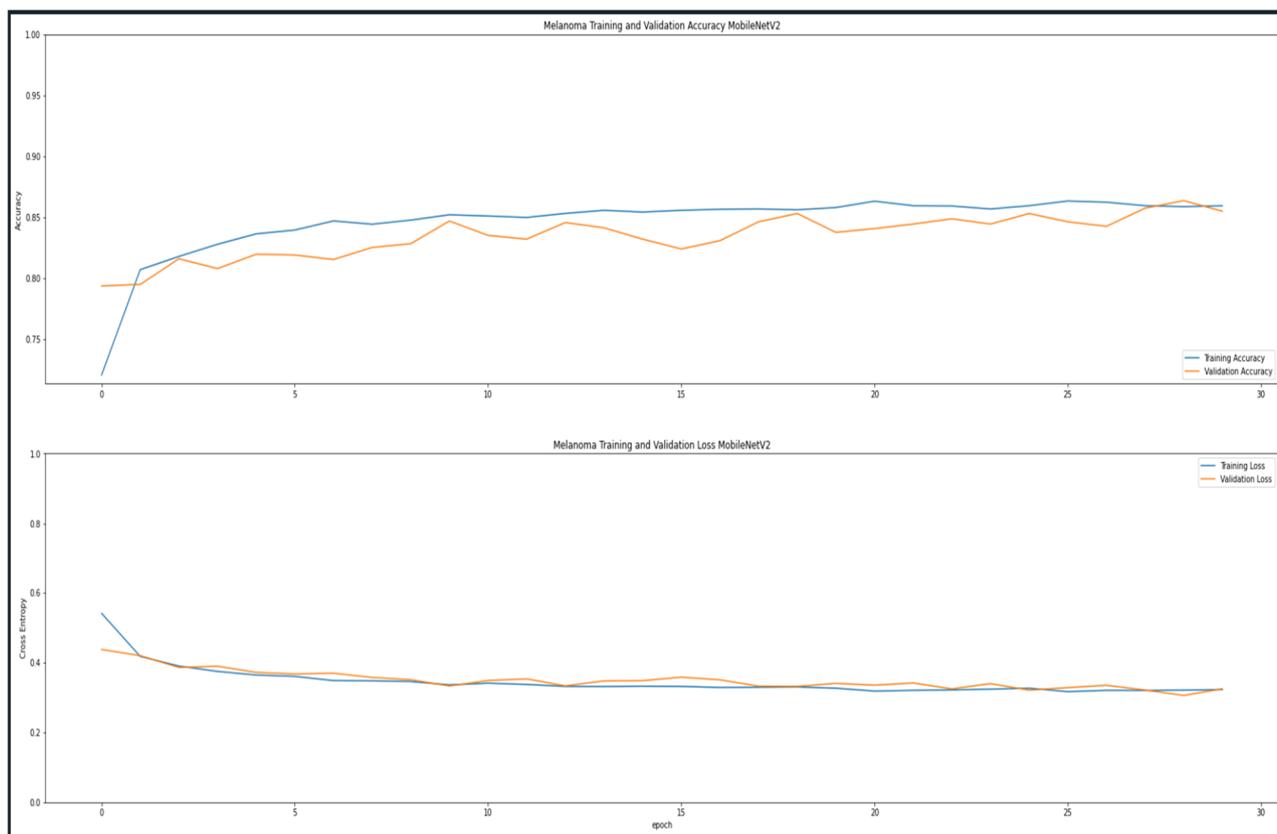


(a)



(b)

Figure 5. Cont.



(c)

Figure 5. Validation (orange)/training accuracy (blue) and validation (orange)/training error (blue) of the melanoma classification models: (a) 15-layer conventional CNN model trained on melanoma photographs; (b) ResNet50 ImageNet pretrained model; (c) MobileNetV2 ImageNet pretrained model.

3.2.2. Adult X-Rays Pneumonia Model

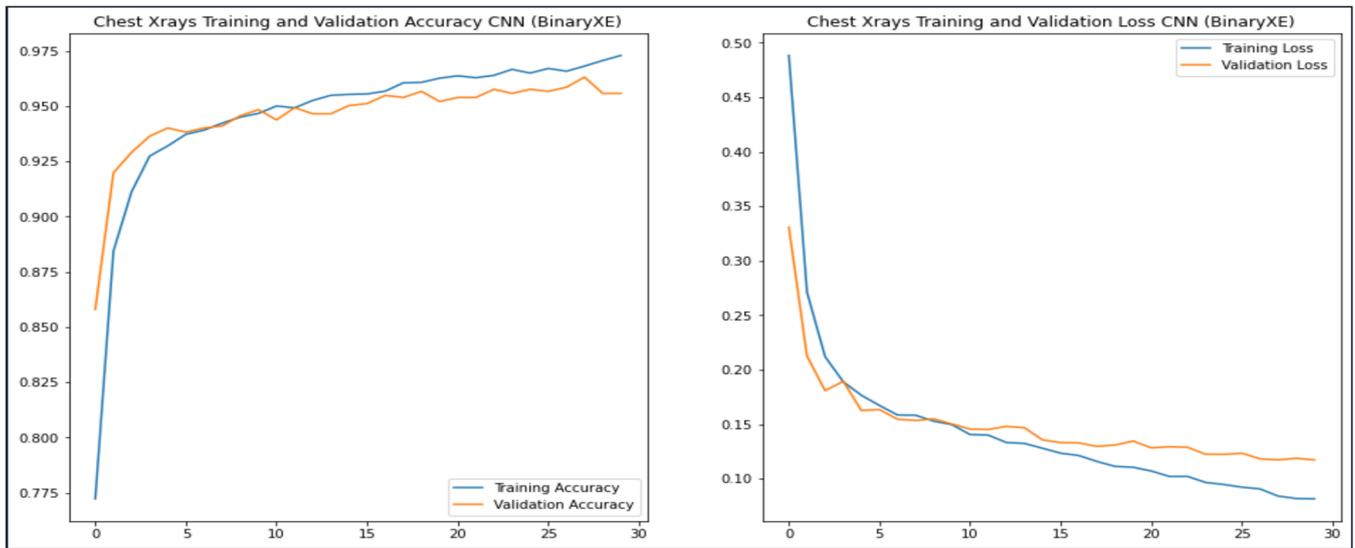
For the adult X-ray classification, a basic model with 19 layers (total) was developed. Various optimizers and learning rates were tested (Table 3). Adam, with an 8×10^{-5} learning rate, had the best results with 96.31% classification accuracy (Figure 6). Binary Cross Entropy and sigmoid had roughly the same results as sparse categorical cross entropy/Softmax.

Table 3. Adult chest X-rays: 19-layer conventional CNN model architecture (with zero padding).

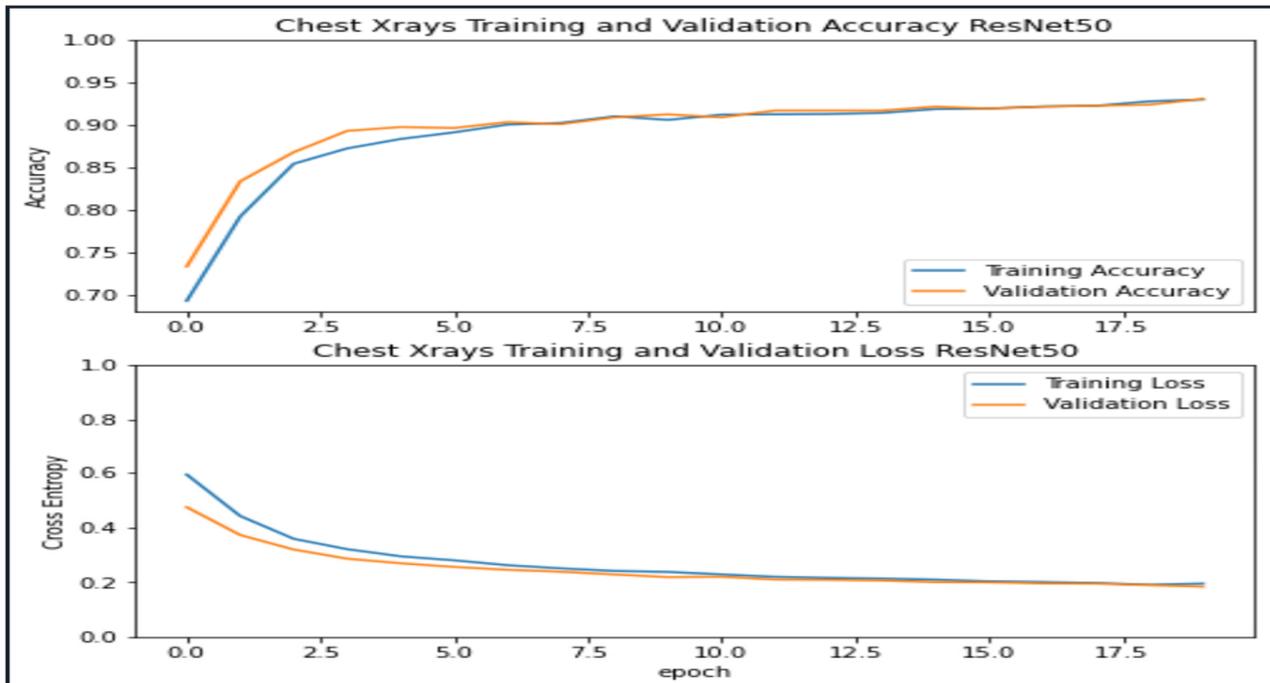
Layer (Type)	Output Shape	Parameters #
Rescaling	(None, 500, 700, 3)	0
Conv2d	(None, 500, 700, 16)	448
Max Pooling	(None, 250, 350, 16)	0
Conv2d	(None, 250, 350, 32)	4640
Max Pooling	(None, 125, 175, 32)	0
Conv2d	(None, 125, 175, 32)	9248
Max Pooling	(None, 62, 87, 32)	0
Conv2d	(None, 62, 87, 32)	9248
Max Pooling	(None, 31, 43, 32)	0
Conv2d	(None, 31, 43, 64)	18,496
Max Pooling	(None, 15, 21, 64)	0
Conv2d	(None, 15, 21, 64)	36,928
Max Pooling	(None, 7, 10, 64)	0
Conv2d	(None, 7, 10, 64)	36,928
Average Pooling	(None, 3, 5, 64)	0

Table 3. Cont.

Layer (Type)	Output Shape	Parameters #
Dropout	(None, 3, 5, 64)	0
Flatten	(None, 960)	0
Dense	(None, 128)	123,008
Dense	(None, 1)	129
Total parameters		239,073 (933.88 KB)
Trainable parameters		239,073 (933.88 KB)
Non-trainable parameters		0 (0.00 B)

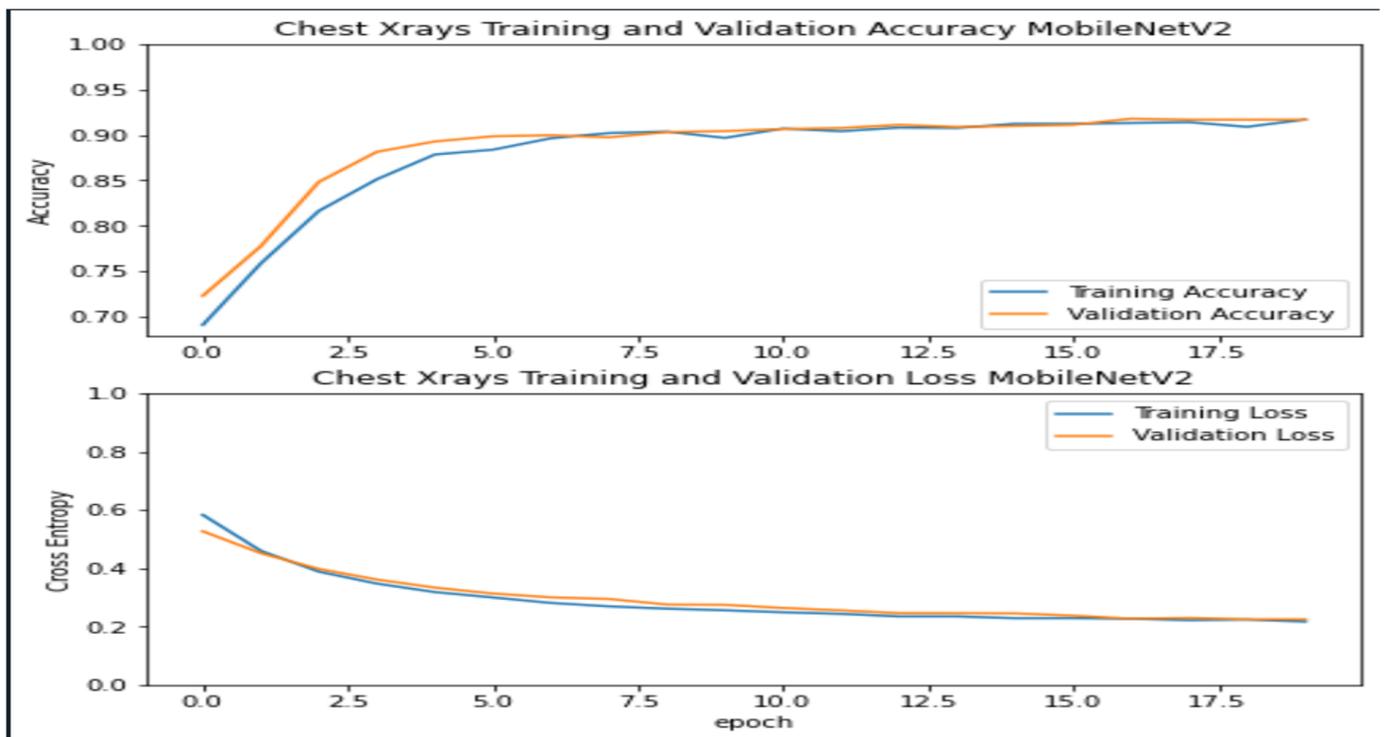


(a)



(b)

Figure 6. Cont.



(c)

Figure 6. Validation (orange)/training accuracy (blue) and validation (orange)/training error (blue) of the adult chest X-ray classification models: (a) 15-layer conventional CNN model trained on chest X-rays; (b) ResNet50 ImageNet pretrained model; (c) MobileNetV2 ImageNet pretrained model.

ResNets had worse results: 93.04% (ResNet50) and 91.79% (MobileNetV2). ResNets converged faster than the custom CNN model.

3.2.3. Pediatric X-Rays Pneumonia Model

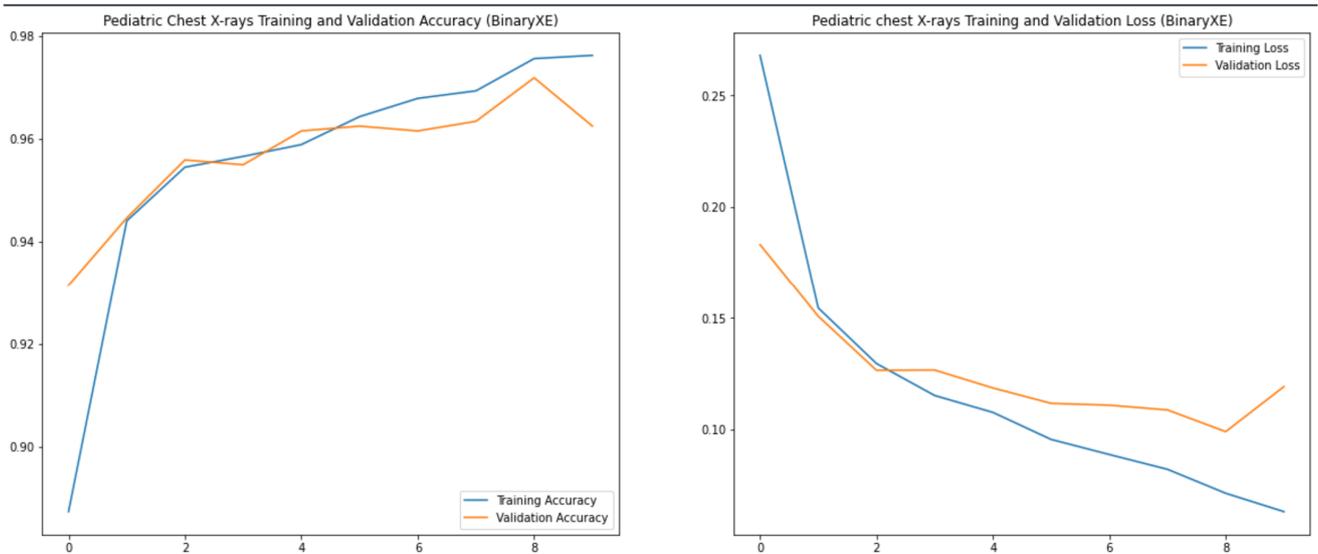
For the pediatric X-ray classification, a basic model with 17 layers (total) was developed. Various optimizers and learning rates were tested (Table 4). Adam, with a 3×10^{-4} learning rate, had the best results with 97.18% classification accuracy (Figure 7). Binary Cross Entropy and sigmoid had roughly the same results as sparse categorical cross entropy/Softmax.

Table 4. Pediatric chest X-ray: 17-layer conventional CNN model architecture (with zero padding).

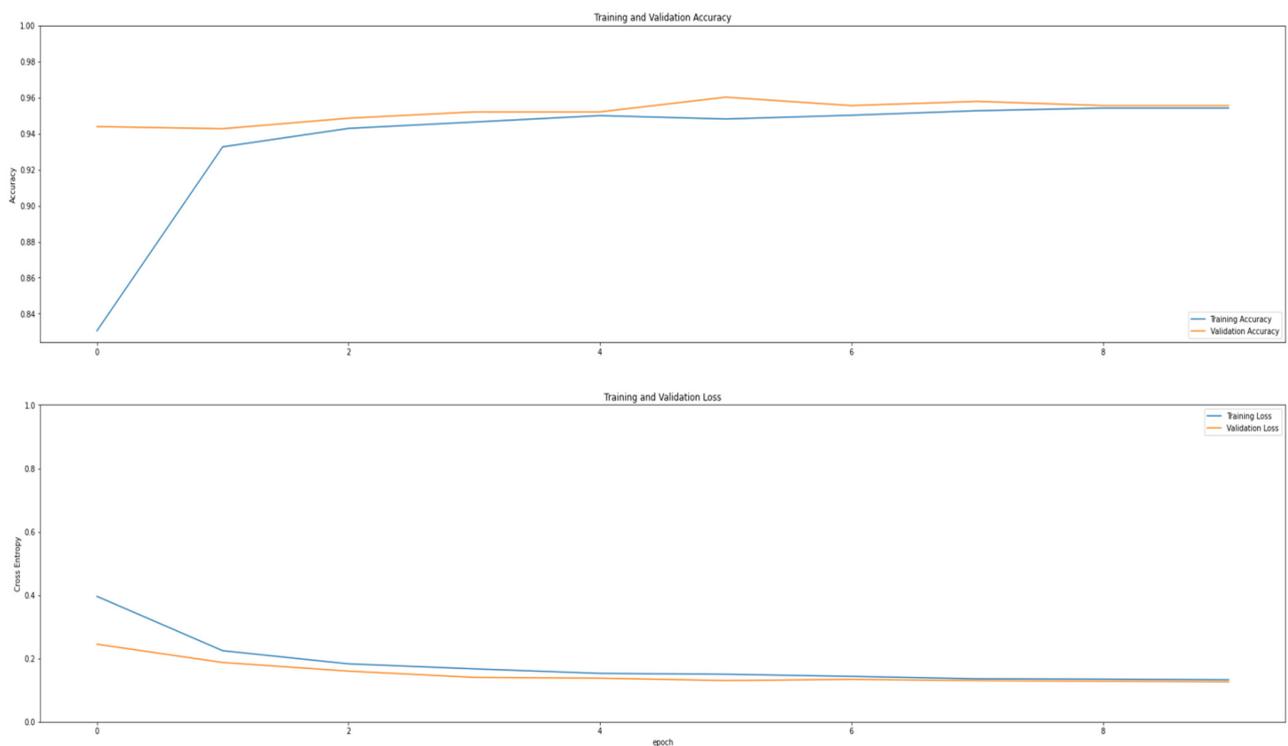
Layer (Type)	Output Shape	Parameters #
Rescaling	(None, 768, 1024, 3)	0
Conv2d	(None, 768, 1024, 16)	448
Max Pooling	(None, 384, 512, 16)	0
Conv2d	(None, 384, 512, 32)	4640
Max Pooling	(None, 192, 256, 32)	0
Conv2d	(None, 192, 256, 32)	9248
Max Pooling	(None, 96, 128, 32)	0
Conv2d	(None, 96, 128, 64)	18,496
Max Pooling	(None, 48, 64, 64)	0
Conv2d	(None, 48, 64, 64)	36,928
Max Pooling	(None, 24, 32, 64)	0
Conv2d	(None, 24, 32, 64)	36,928
Average Pooling	(None, 12, 16, 64)	0
Dropout	(None, 12, 16, 64)	0

Table 4. Cont.

Layer (Type)	Output Shape	Parameters #
Flatten	(None, 12,288)	0
Dense	(None, 128)	1,572,992
Dense	(None, 1)	129
Total parameters	1,679,809 (6.41 MB)	
Trainable parameters	1,679,809 (6.41 MB)	
Non-trainable parameters	0 (0.00 B)	

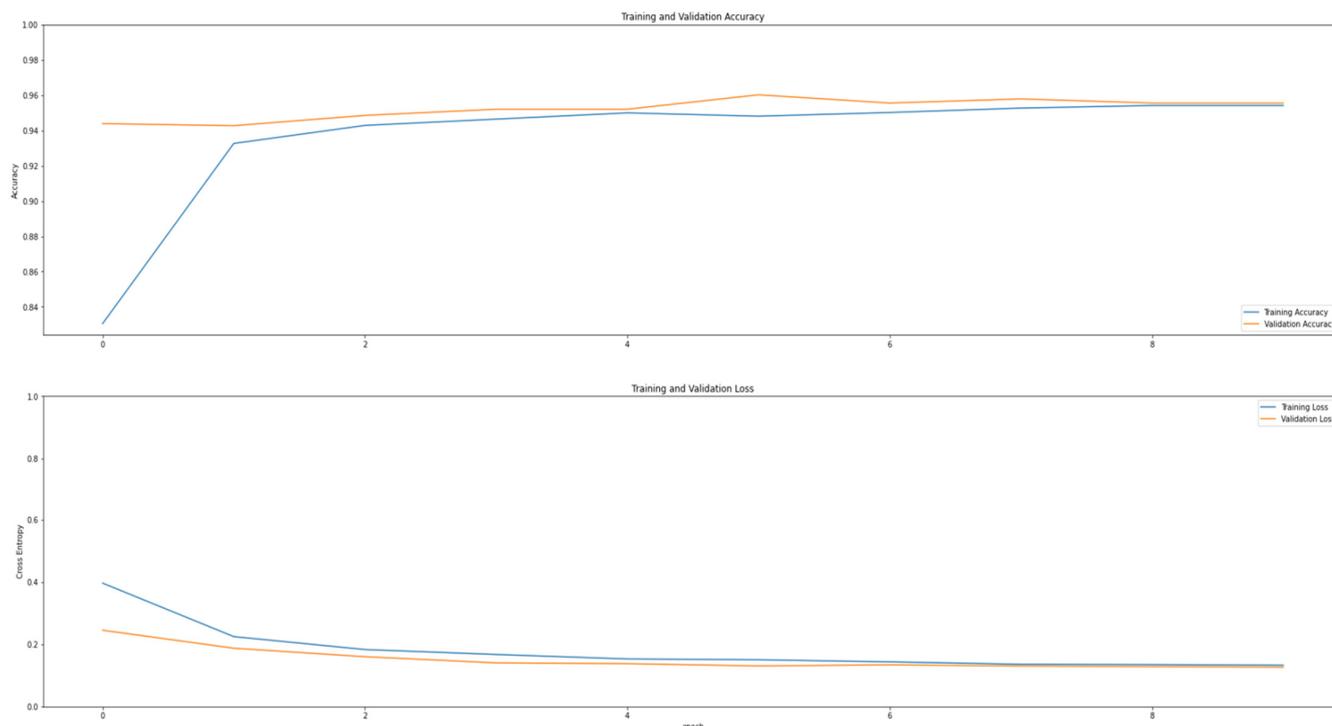


(a)



(b)

Figure 7. Cont.



(c)

Figure 7. Validation (orange)/training accuracy (blue) and validation (orange)/training error (blue) of the pediatric X-ray classification models: (a) 17-layer conventional CNN model trained on pediatric chest X-rays; (b) ResNet50 ImageNet pretrained model; (c) MobileNetV2 ImageNet pretrained model.

ResNets had worse results: 93.69% (ResNet50) and 92.76% (MobileNetV2). ResNets converged faster than the custom CNN model.

3.2.4. MRI Brain Scan Tumor Classification Model

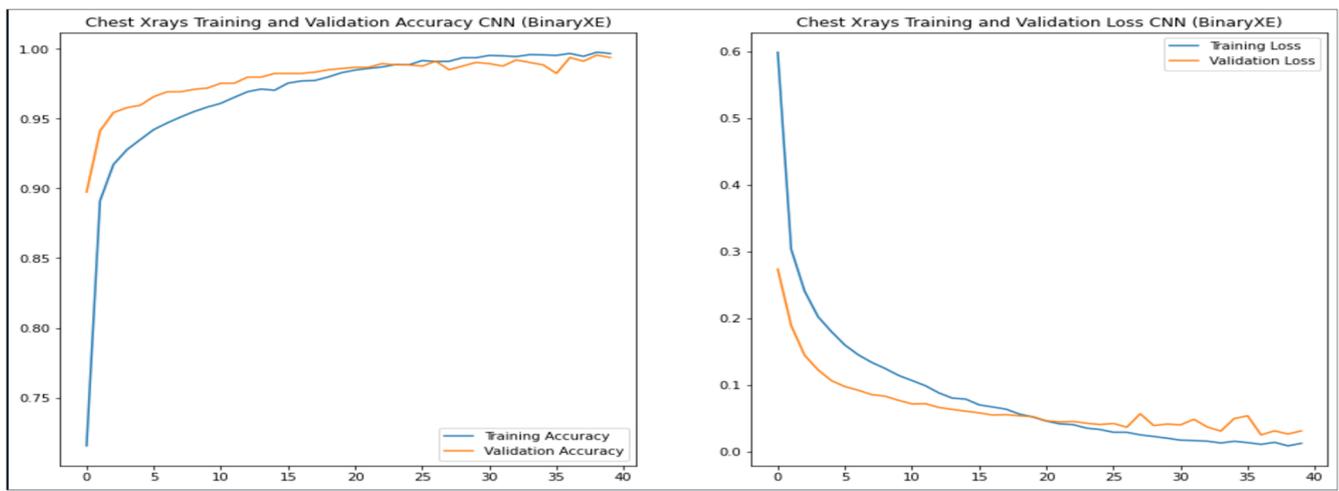
For the MRI brain scan classification, a basic model with 19 layers (total) was developed. Various optimizers and learning rates were tested (Table 5). Adam, with 3×10^{-4} learning rate, had the best results with 99.56% classification accuracy (Figure 8). Binary Cross Entropy and sigmoid had roughly the same results as sparse categorical cross entropy/Softmax.

Table 5. MRI brain scans (tumors): 19-layer conventional CNN model architecture (with zero padding).

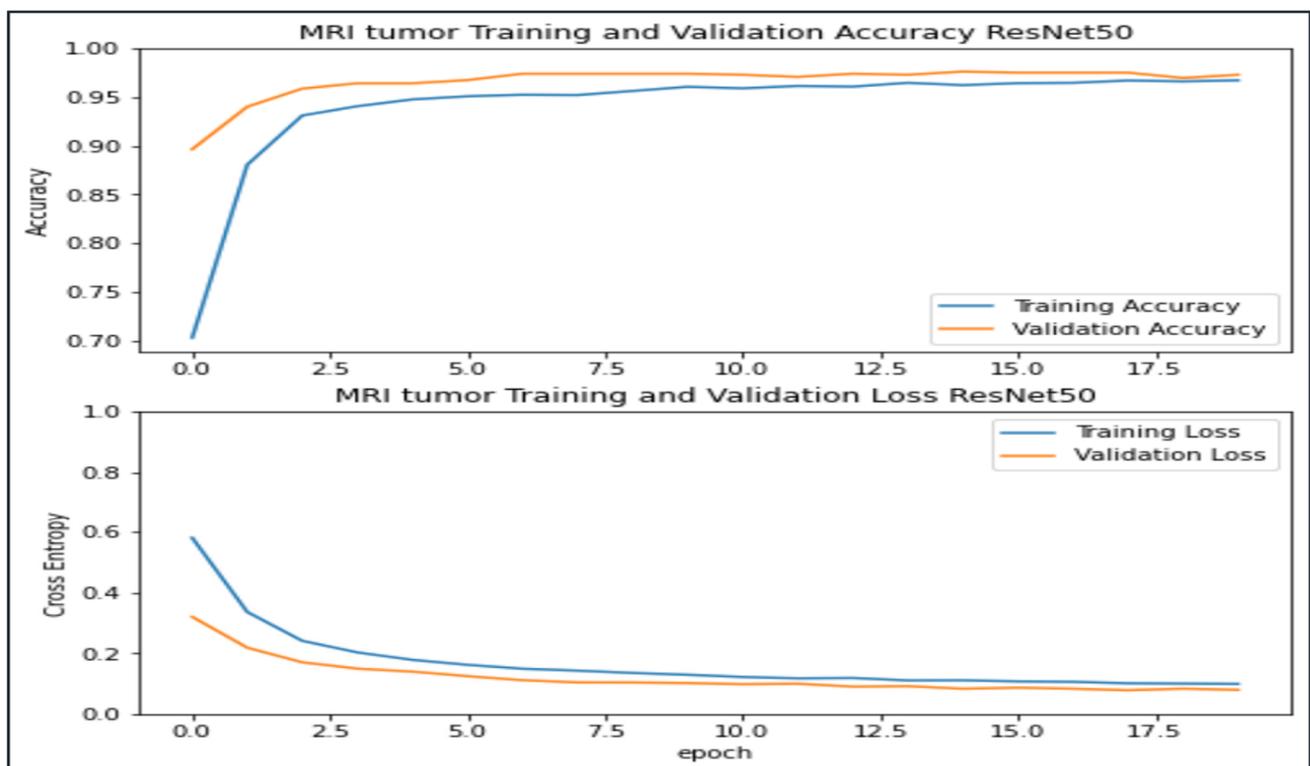
Layer (Type)	Output Shape	Parameters #
Rescaling	(None, 300, 300, 3)	0
Conv2d	(None, 300, 300, 16)	448
Max Pooling	(None, 150, 150, 16)	0
Conv2d	(None, 150, 150, 32)	4640
Max Pooling	(None, 75, 75, 32)	0
Conv2d	(None, 75, 75, 32)	9248
Max Pooling	(None, 37, 37, 32)	0
Conv2d	(None, 37, 37, 32)	9248
Max Pooling	(None, 18, 18, 32)	0
Conv2d	(None, 18, 18, 64)	18,496
Max Pooling	(None, 9, 9, 64)	0
Conv2d	(None, 9, 9, 64)	36,928
Max Pooling	(None, 4, 4, 64)	0
Conv2d	(None, 4, 4, 64)	36,928

Table 5. Cont.

Layer (Type)	Output Shape	Parameters #
Average Pooling	(None, 2, 2, 64)	0
Dropout	(None, 2, 2, 64)	0
Flatten	(None, 256)	0
Dense	(None, 128)	32,896
Dense	(None, 1)	129
Total parameters	148,961 (581.88 KB)	
Trainable parameters	148,961 (581.88 KB)	
Non-trainable parameters	0 (0.00 B)	

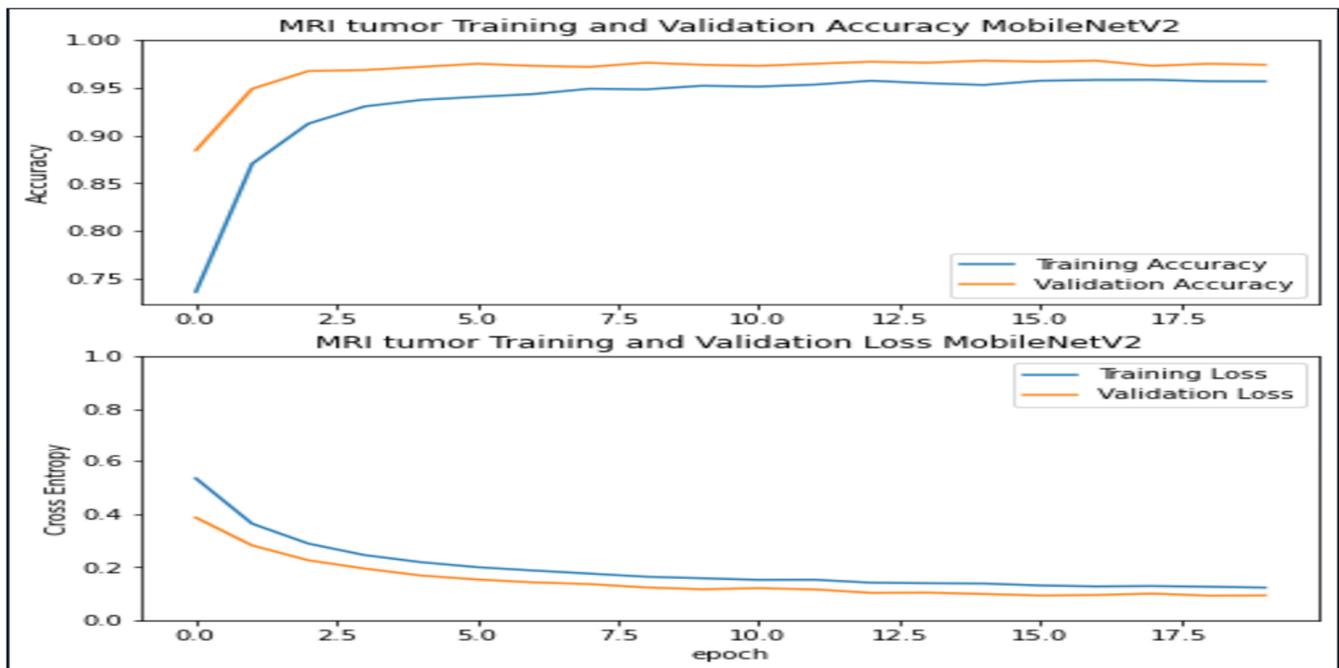


(a)



(b)

Figure 8. Cont.



(c)

Figure 8. Validation (orange)/training accuracy (blue) and validation (orange)/training error (blue) of the MRI brain scans classification models: (a) 19-layer conventional CNN model trained on MRI brain scans; (b) ResNet50 ImageNet pretrained model; (c) MobileNetV2 ImageNet pretrained model.

ResNets had worse results: 97.60% (ResNet50) and 97.82% (MobileNetV2). ResNets converged faster than the custom CNN model.

3.2.5. Pap Smear Test Model

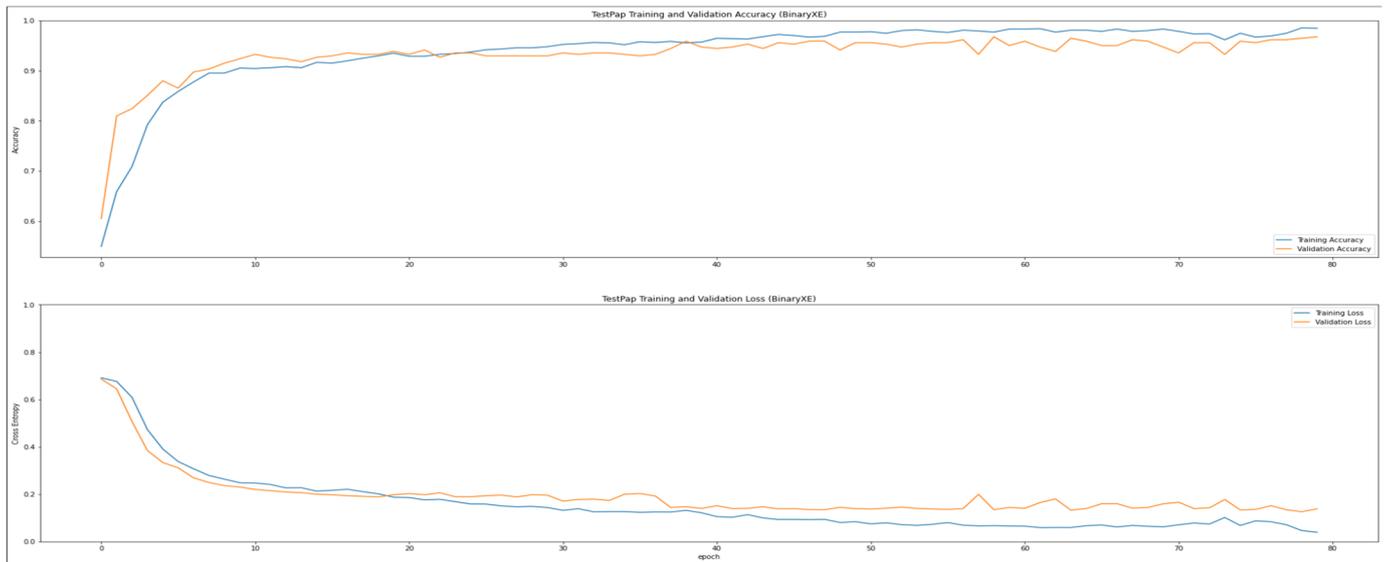
For the pap smear test classification, a basic model with 17 layers (total) was developed. Various optimizers and learning rates were tested (Table 6). Adam, with a 3×10^{-4} learning rate, had 96.78% classification accuracy (Figure 9). Binary Cross Entropy and sigmoid had roughly the same results as sparse categorical cross entropy/Softmax.

Table 6. Pap smear test images 17-layer conventional CNN model architecture (Adam 3×10^{-4} , Binary Cross Entropy/sigmoid) (with zero padding).

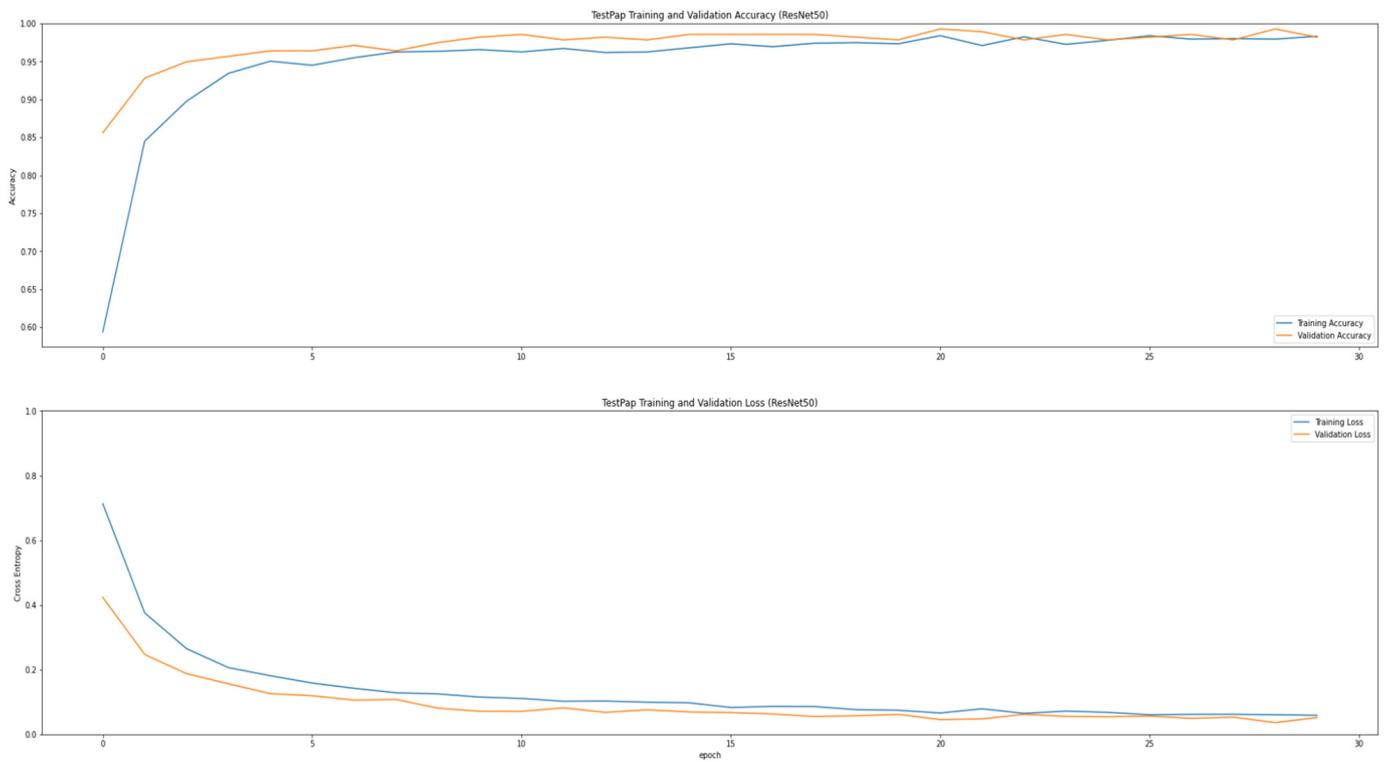
Layer (Type)	Output Shape	Parameters #
Rescaling	(None, 100, 120, 3)	0
Conv2d	(None, 100, 120, 16)	448
Max Pooling	(None, 50, 60, 16)	0
Conv2d	(None, 50, 60, 32)	4640
Max Pooling	(None, 25, 30, 32)	0
Conv2d	(None, 25, 30, 32)	9248
Max Pooling	(None, 12, 15, 32)	0
Conv2d	(None, 12, 15, 32)	9248
Max Pooling	(None, 6, 7, 64)	0
Conv2d	(None, 6, 6, 64)	18,496
Max Pooling	(None, 3, 3, 64)	0
Conv2d	(None, 3, 3, 64)	36,928
Average Pooling	(None, 1, 1, 64)	0
Dropout	(None, 1, 1, 64)	0
Flatten	(None, 64)	0
Dense	(None, 128)	8320

Table 6. Cont.

Layer (Type)	Output Shape	Parameters #
Dense	(None, 1)	129
Total parameters	87,457 (341.63 KB)	
Trainable parameters	87,457 (341.63 KB)	
Non-trainable parameters	0 (0.00 B)	



(a)



(b)

Figure 9. Cont.

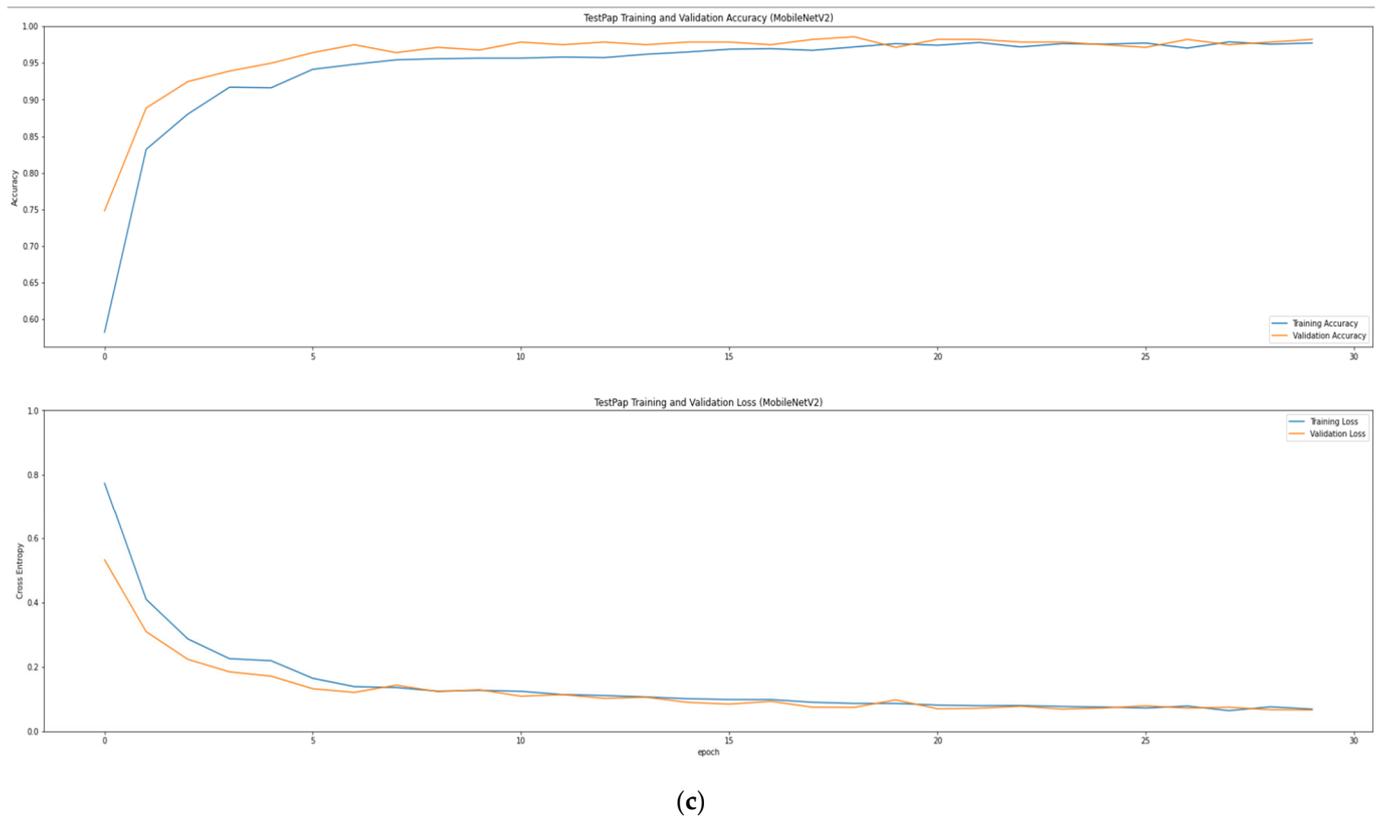


Figure 9. Validation (orange)/training accuracy (blue) and validation (orange)/training (blue) error of the pap smear test classification models: (a) 17-layer conventional CNN model trained on pap smear tests; (b) ResNet50 ImageNet pretrained model; (c) MobileNetV2 ImageNet pretrained model.

ResNets had better results: 99.28% (ResNet50) and 99.20% (MobileNetV2). ResNets converged faster than the custom CNN model.

3.2.6. Fine-Tuning on the Adult Chest X-Rays Dataset

For reference, we added a ResNet50 model that we fine-tuned for the 2-class adult chest X-rays dataset. Fine-tuning was applied with the following technique:

The model was trained for 20 epochs with all the weights fixed (as they were trained on the ImageNet dataset). Then, after epoch 20, all the layers after layer 130 were unfrozen and were retrained on the adult chest X-rays dataset. At the same time, the learning rate was reduced to 1/10 of its previous value to avoid overfitting. (See Figure 10).

The model started to improve and reached 97.83% validation accuracy, surpassing both the pretrained ResNets (93.04%) and the custom CNNs (96.31%).

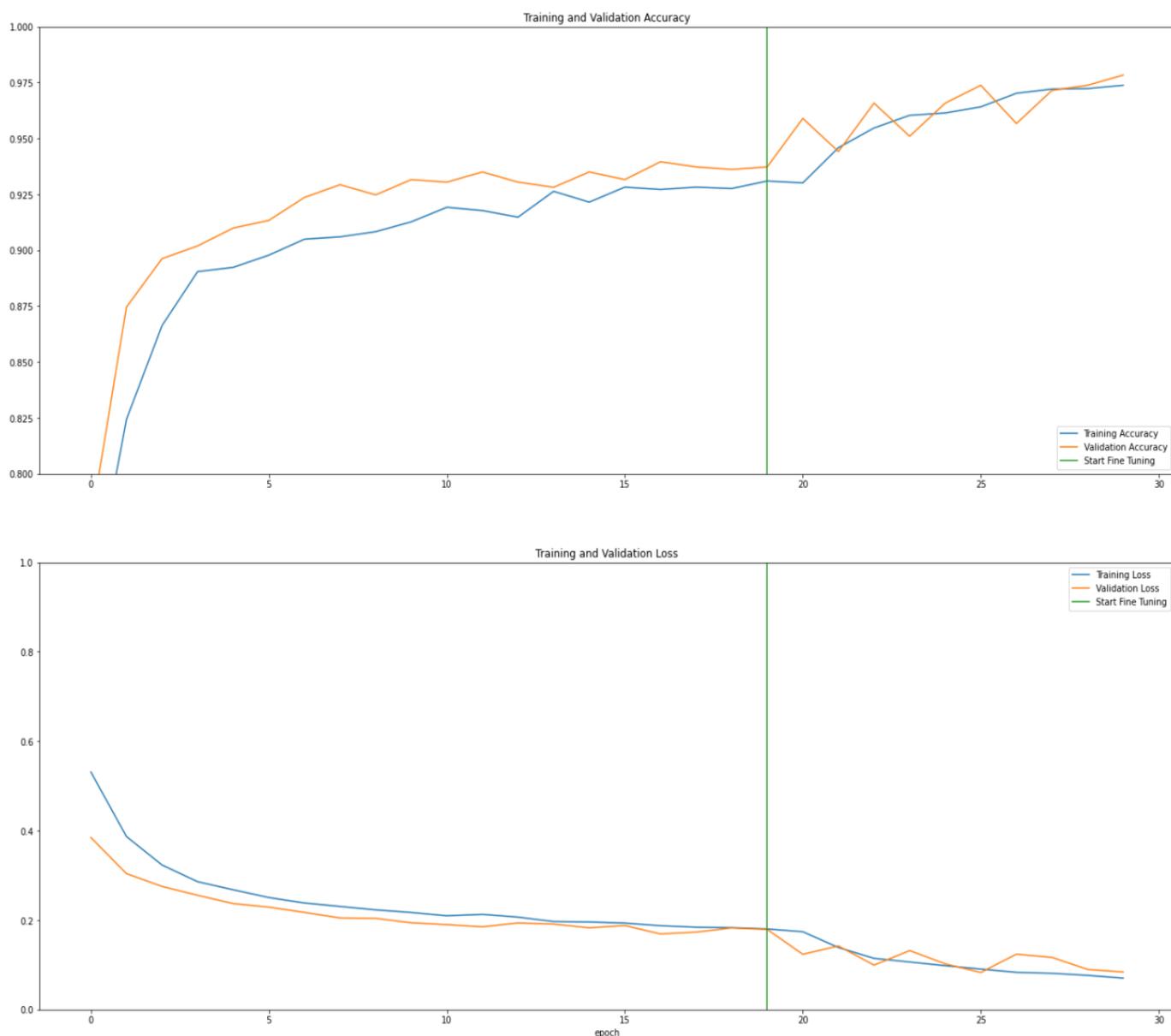


Figure 10. Validation (orange)/training accuracy (blue) and validation (orange)/training (blue) error of the fine-tuned ResNet50 ImageNet pretrained model for the adult Chest X-rays dataset. The green line shows the point (epoch 20) when fine-tuning starts.

3.3. Model Performance

Overall, for most datasets, conventional CNNs outperformed pretrained residual networks in terms of validation accuracy. Among the residual networks, ResNet50 with the standard architecture consistently achieved higher validation accuracy than the inverted-structure MobileNetV2. Notably, the fine-tuned ResNet50 for classifying pneumonia in chest X-rays demonstrated particularly promising results, outperforming all other models.

The results for melanoma classification were promising, especially considering that the data consisted of simple photographs with relatively low resolution.

The classification models for pap smear tests demonstrated high accuracy in 2-class classification across all architectures and high accuracy in 5-class classification.

Similarly, the 2-class X-ray models exhibited strong accuracy across all architectures, with comparable results for adult and pediatric data. However, further research is warranted due to the significantly varied validation accuracy percentages among different combinations of chest X-rays. While the models achieved high accuracy in distinguishing

between normal, COVID-19, and pneumonia cases, they showed lower accuracy (86.6%) when classifying normal, bacterial, and viral pneumonia. This result is still notable, given the challenges that chest X-rays pose in allowing medical professionals to definitively diagnose bacterial or viral pneumonia.

MRI brain scans demonstrated impressive classification power, particularly with custom CNNs, especially in the 2-class models. The use of a premade augmentation technique proved beneficial, and the pretrained models also performed exceptionally well.

Model performance for all datasets and all Neural Network architectures are summarized in Table 7.

Table 7. Models’ performance (validation accuracy).

Dataset	Conventional CNN	Pretrained ResNet50	Pretrained MobileNetV2
Melanoma	93.05%	89.79%	86.39
Chest X-rays	96.31%	93.04% (97.83 fine-tuned)	91.79%
Pediatric X-rays	97.18%	93.69%	92.76%
MRI brain scans	99.56%	97.60%	97.82%
Pap smear test	96.78%	99.28%	99.20%

3.4. Computational Needs

Low-resolution and medium-resolution models trained quickly, with each epoch completed in just a matter of seconds. In contrast, training times for higher resolutions (500 × 700 pixels and above) were significantly longer, and computational challenges began to arise. Implementing autotuning with pre-fetch improved RAM utilization, resulting in a 5% to 10% increase in training speed.

Pretrained models were slower than basic CNN models, which is expected given their more complex architecture. Training on a GPU was approximately 3 to 10 times faster than on a CPU, depending on the dataset and model used. A summary of the models’ training speeds can be found in Table 8.

Table 8. Models’ training speed (seconds/epoch).

Processor/OS	Conventional CNN	Pretrained ResNet50	Pretrained MobileNetV2
Ryzen 7 7700 32 GB CPU/Windows	Pap smear 100 × 120 p: 1 s/epoch	Pap smear 100 × 120 p: 29 s/epoch	Pap smear 100 × 120 p: 10 s/epoch
	MRI 300 × 300 p: 21 s/epoch	MRI 300 × 300 p: 190 s/epoch	MRI 300 × 300 p: 51 s/epoch
	Melanoma 224 × 224 p: 25 s/epoch	Melanoma 224 × 224 p: 200 s/epoch	Melanoma 224 × 224 p: 55 s/epoch
	chest X-rays 500 × 700 p: 195 s/epoch	chest X-rays 500 × 700 p: 1050 s/epoch	chest X-rays 500 × 700 p: 368 s/epoch
RTX 4060TI 16 GB GPU/Linux	Pap smear 100 × 120 p: 0.16 s/epoch	Pap smear 100 × 120 p: 2 s/epoch	Pap smear 100 × 120 p: 0.3 s/epoch
	MRI 300 × 300 p: 5 s/epoch	MRI 300 × 300 p: 24 s/epoch	MRI 300 × 300 p: 10 s/epoch
	Melanoma 224 × 224 p: 6 s/epoch	Melanoma 224 × 224 p: 25 s/epoch	Melanoma 224 × 224 p: 11 s/epoch
	chest X-rays 500 × 700 p: 18 s/epoch	chest X-rays 500 × 700 p: 75 s/epoch	chest X-rays 500 × 700 p: 35 s/epoch

4. Discussion

Conventional CNN models fully trained on medical data generally showed promising results despite their relatively shallow architecture.

The fine-tuned ResNet50 model achieved a very promising accuracy of 97.83% for pneumonia classification, which approaches the 99.24% accuracy reported by Behar et al. [7] for fully retrained ResNet50 model in breast cancer classification and the 99.17% accuracy achieved by the ResNet50 model on COVID-19 chest X-rays [8]. The fine-tuned ResNet50 model appears promising and may be suitable for broader applications in medical imaging.

All CNN models showed strong performance on cytological data, suggesting that further data collection in this area could be beneficial.

Data preprocessing for viral and bacterial pneumonia chest X-rays suggests that convolutional models, combined with data preprocessing techniques, may provide additional insights into disease expression in medical images.

These results highlight the importance of accumulating medical imaging data to enable further advancements in this field.

CNN models continue to demonstrate high performance, suggesting a future possibility—once more extensive and higher-quality data are available—of internet applications that allow users to “self-test” for melanoma via webcams, keep records of their mole, and receive guidance from medical professionals as needed.

A melanoma detection model, using an architecture similar to ours and trained on high-resolution images, can capture simple photographs of moles and classify them with high accuracy. This model could be integrated into an internet application that serves as a preliminary telemedicine tool, eliminating the immediate need for a medical professional. Both web and smartphone cameras can be used for mole screening and early melanoma detection. Additionally, a web application could track the history of moles, detecting any changes in prediction certainty, which may indicate malignancy and signal the need for a dermatologist consultation.

Overall, these developed models can be compatible with internet-based applications, making them suitable for inclusion in telemedicine environments that support medical decision-making.

It is also essential to address the potential ethical implications of medical research in the field. Medical data use is sensitive and must always ensure the preservation of data anonymity. Another critical aspect of research in this area is data transparency, which helps to avoid or account for low-quality data and ensures only accurate information is utilized. Additionally, the “black box” issue inherent in neural networks poses a challenge since their operations are not fully visible or interpretable to doctors or patients. In this paper, we approached CNN models not merely as “black boxes” that outperform human classification (e.g., distinguishing viral from bacterial pneumonia) but also as tools for deeper study of disease characteristics.

In future work, we aim to fine-tune deep residual networks further to enhance classification performance and to develop more specialized, tailor-made preprocessing and augmentation techniques for bacterial and viral pneumonia. This will allow for a more detailed study of the distinct characteristics of these diseases as represented in medical images.

Author Contributions: Conceptualization, C.R. and A.A.; methodology, C.R. and A.A.; software, C.R.; validation, C.R., E.K., G.A. and A.A.; formal analysis, A.A.; investigation, C.R. and A.A.; resources, C.R.; data curation, C.R.; writing—original draft preparation, C.R.; writing—review and editing, C.R., E.K., G.A. and A.A.; visualization, C.R.; supervision, A.A.; project administration, A.A.; funding acquisition, none. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of Democritus University of Thrace, Greece, protocol code ΔΠΘ/ΕΗΔΕ/43087/302, 24 April 2024.

Data Availability Statement: For the Pap Test models, data used were extracted from (Marina E. Plisiti 7–10 October 2018) [9]: <https://www.kaggle.com/datasets/mohaliy2016/papsinglecell> (accessed on 1 November 2024), <https://www.cs.uoi.gr/~marina/sipakmed.html> (accessed on 1 November

2024). For the 2Xray models, data used were extracted from (D. S. Kermany 2018) [14]: <https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia?datasetId=17810> (accessed on 1 November 2024). For the 43Xrays and 4XRays models, data used were extracted from (Sait, et al. 2020): <https://www.kaggle.com/datasets/unaissait/curated-chest-xray-image-dataset-for-covid19> (accessed on 1 November 2024). For the Pediatric_Xrays models, data used were extracted from (Kermany, Zhang and Goldbaum 2018) [15]: <https://www.kaggle.com/datasets/andrewmvd/pediatric-pneumonia-chest-xray/> (accessed on 1 November 2024). For the 4BrainAugmented models, data used were extracted from (Hashemi 2023) [21], <https://www.kaggle.com/datasets/mohammadhossein77/brain-tumors-dataset> (accessed on 1 November 2024). For the melanoma models, data used were extracted from <https://www.kaggle.com/datasets/bhaveshmittal/melanoma-cancer-dataset> (accessed on 1 November 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Residual Network Model Architectures

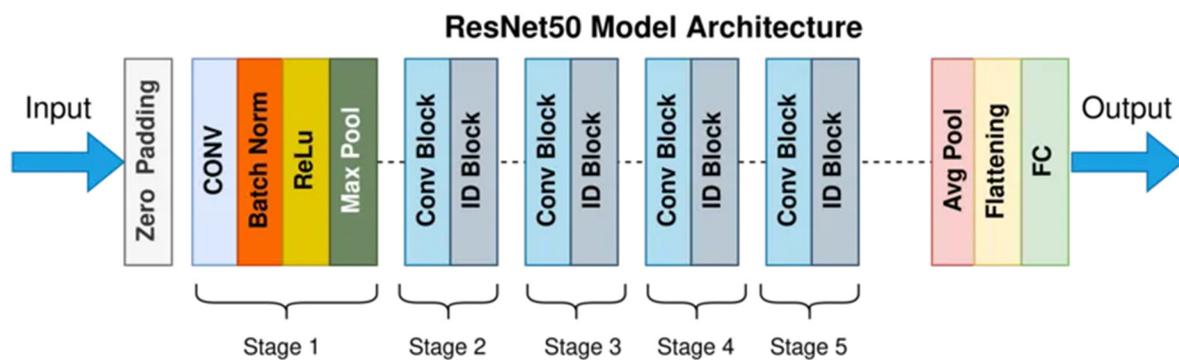


Figure A1. ResNet50 model architecture (residual network).

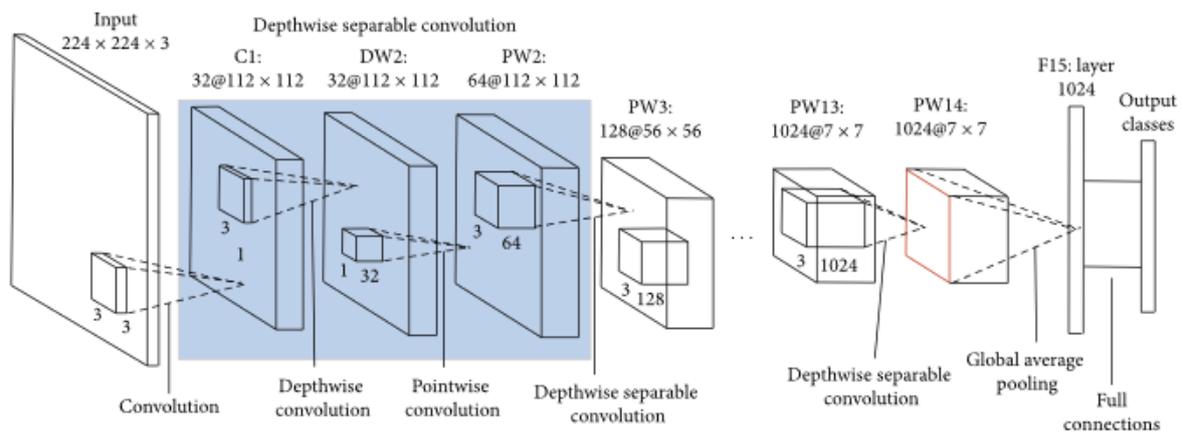


Figure A2. MobileNet model architecture (inverted residual network).

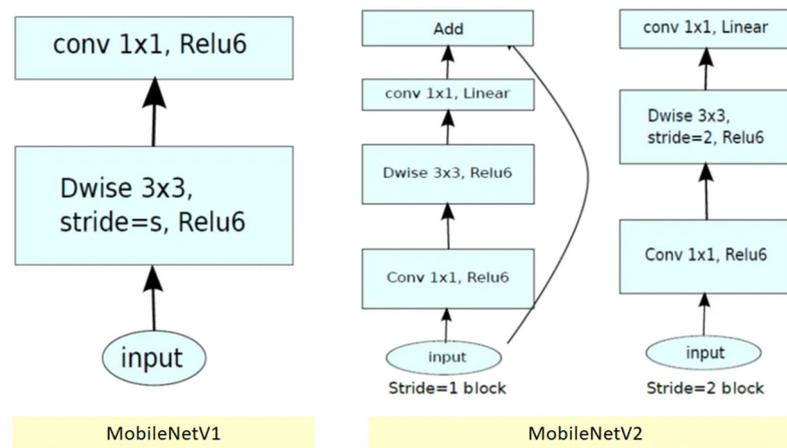


Figure A3. MobileNetV1 and MobileNetV2.

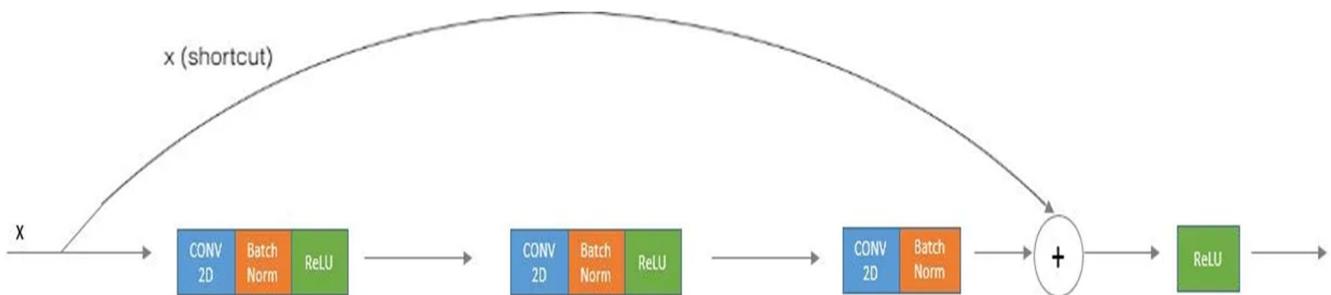


Figure A4. Identity block.

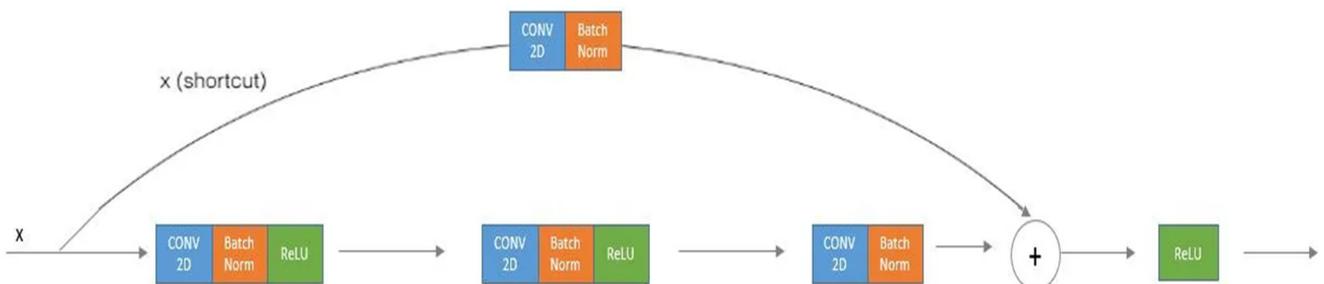


Figure A5. Convolutional block.

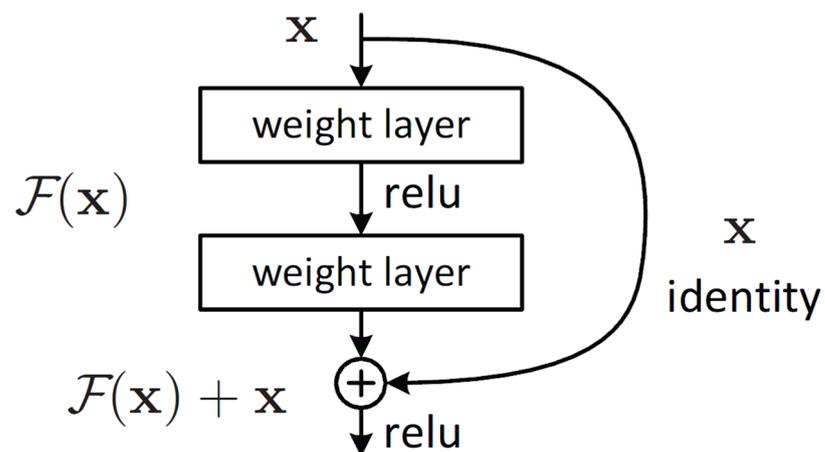


Figure A6. Residual network connection.

Adam optimizer algorithm**(as presented in its original paper [28])** g_t^2 indicates the elementwise square $g_t \odot g_t$.Good default settings for the tested machine learning problems are $a = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-8}$. All operations on vectors are element-wise.With β_1^t and β_2^t , we denote β_1 and β_2 to the power t .**Require:** a : Stepsize**Require:** $\beta_1, \beta_2 \in [0, 1)$: Exponential decay rates for the moment estimates**Require:** $f(\theta)$: Stochastic objective function with parameters θ **Require:** θ_0 : Initial parameter vector $m_0 \leftarrow 0$ (Initialize 1st moment vector) $v_0 \leftarrow 0$ (Initialize 2nd moment vector) $t \leftarrow 0$ (Initialize timestep)**while** θ_t not converged **do** $t \leftarrow t + 1$ $g_t \leftarrow \nabla_{\theta} f_{\Theta}(\theta_t - 1)$ (Get gradients w.r.t. stochastic objective at timestep t) $m_t \leftarrow \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$ (Update biased first-moment estimate) $v_t \leftarrow \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2$ (Update biased second raw moment estimate) $\hat{m}_t \leftarrow m_t / (1 - \beta_1^t)$ (Compute bias-corrected first-moment estimate) $\hat{v}_t \leftarrow v_t / (1 - \beta_2^t)$ (Compute bias-corrected second raw moment estimate) $\theta_t \leftarrow \theta_{t-1} - a \cdot \hat{m}_t / (\sqrt{\hat{v}_t} + \epsilon)$ (Update parameters)**end while****return** θ_t (Resulting parameters).**Appendix B. Computer Specifications**

The models were trained on a modern conventional computer with the following specifications:

CPU: AMD Ryzen 7700 8 cores/16 threads up to 5.3 Ghz (PBO off);

GPU: Nvidia RTX 4060Ti 16 GB VRAM;

RAM: 2×16 GB dual channel 6000MT/s DDR5 (XMP off/4800 MT/s);

ROM: 1TB PCIe 4.0 NVME M.2 SSD 7000 MB/s read, 6000 MB/s write (Windows 11);

ROM: 1TB PCIe 4.0 NVME M.2 SSD 7000 MB/s read, 6000 MB/s write (Ubuntu 22.04).

References

- Alexander, R.G.; Yazdanie, F.; Waite, S.; Chaudhry, Z.A.; Kolla, S.; Macknik, S.L.; Martinez-Conde, S. Visual Illusions in Radiology: Untrue Perceptions in Medical Images and Their Implications for Diagnostic Accuracy. *Front. Neurosci.* **2021**, *15*, 629469. [CrossRef] [PubMed]
- Waite, S.; Scott, J.; Gale, B.; Fuchs, T.; Kolla, S.; Reede, D. Interpretive Error in Radiology. *AJR Am. J. Roentgenol.* **2017**, *208*, 739–749. [CrossRef] [PubMed]
- Mall, P.K.; Singh, P.K.; Srivastav, S.; Narayan, V.; Paprzycki, M.; Jaworska, T.; Ganzha, M. A comprehensive review of deep neural networks for medical image processing: Recent developments and future opportunities. *Healthc. Anal.* **2023**, *4*, 100216. [CrossRef]
- Nallakaruppan, M.K.; Ramalingam, S.; Somayaji, S.R.K.; Prathiba, S.B. Comparative Analysis of Deep Learning Models Used in Impact Analysis of Coronavirus Chest X-ray Imaging. *Biomedicines* **2022**, *10*, 2791. [CrossRef] [PubMed] [PubMed Central]
- Chan, H.P.; Samala, R.K.; Hadjiiski, L.M.; Zhou, C. Deep Learning in Medical Image Analysis. *Adv. Exp. Med. Biol.* **2020**, *1213*, 3–21. [CrossRef] [PubMed] [PubMed Central]
- Cai, L.; Gao, J.; Zhao, D. A review of the application of deep learning in medical image classification and segmentation. *Ann. Transl. Med.* **2020**, *8*, 713. [CrossRef]
- Behar, N.; Shrivastava, M. ResNet50-Based Effective Model for Breast Cancer Classification Using Histopathology Images. *Comput. Model. Eng. Sci.* **2022**, *130*, 823–839. [CrossRef]
- Hossain, M.B.; Iqbal, S.M.H.S.; Islam, M.M.; Akhtar, M.N.; Sarker, I.H. Transfer learning with fine-tuned deep CNN ResNet50 model for classifying COVID-19 from chest X-ray images. *Inform. Med. Unlocked.* **2022**, *30*, 100916. [CrossRef] [PubMed] [PubMed Central]
- Plissiti, M.E.; Dimitrakopoulos, P.; Sfikas, G.; Nikou, C.; Krikoni, O.; Charchanti, A. Sipakmed: A New Dataset for Feature and Image Based Classification of Normal and Pathological Cervical Cells in Pap Smear Images. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 3144–3148. [CrossRef]
- Histopathology of the Uterine Cervix—Digital Atlas. Available online: <https://screening.iarc.fr/atlasglossdef.php?key=Koilocyte&lang=1> (accessed on 1 November 2024).

11. Mayoclinic.org. Available online: <https://www.mayoclinic.org/diseases-conditions/pneumonia/symptoms-causes/syc-20354204> (accessed on 1 November 2024).
12. Yale Medicine. Available online: <https://www.yalemedicine.org/conditions/rsv-respiratory-syncytial-virus> (accessed on 1 November 2024).
13. Stefanidis, K.; Konstantelou, E.; Yusuf, G.T.; Oikonomou, A.; Tavernaraki, K.; Karakitsos, D.; Loukides, S.; Vlahos, I. Radiological, epidemiological and clinical patterns of pulmonary viral infections. *Eur. J. Radiol.* **2021**, *136*, 109548. [[CrossRef](#)]
14. Kermany, D.S. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell* **2018**, *172*, 1122–1131. [[CrossRef](#)] [[PubMed](#)]
15. Kermany, D.; Zhang, K.; Goldbaum, M. Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification. Mendeley Data 2018, Version 2. Available online: <https://data.mendeley.com/datasets/rscbjbr9sj/2> (accessed on 19 August 2024).
16. Sait, U.; Lal KV, G.; Prakash Prajapati, S.; Bhaumik, R.; Kumar, T.; Shivakumar, S.; Bhalla, K. Curated Dataset for COVID-19 Posterior-Anterior Chest Radiography Images (X-Rays). Mendeley Data 2020, Version 1. Available online: <https://data.mendeley.com/datasets/9xkhgts2s6/1> (accessed on 19 August 2024).
17. National Institute of Biomedical Imaging and Bioengineering. Available online: <https://www.nibib.nih.gov/science-education/science-topics/magnetic-resonance-imaging-mri> (accessed on 19 August 2024).
18. Russ, S.; Anastasopoulou, C.; Shafiq, I. Pituitary Adenoma. In *StatPearls [Internet]*; StatPearls Publishing: Treasure Island, FL, USA, 2024. Available online: <https://www.ncbi.nlm.nih.gov/books/NBK554451/> (accessed on 27 March 2023).
19. Alruwaili, A.A.; De Jesus, O. Meningioma. In *StatPearls [Internet]*; StatPearls Publishing: Treasure Island, FL, USA, 2024. Available online: <https://www.ncbi.nlm.nih.gov/books/NBK560538/> (accessed on 23 August 2023).
20. Mesfin, F.B.; Al-Dhahir, M.A. Gliomas. In *StatPearls [Internet]*; StatPearls Publishing: Treasure Island, FL, USA, 2024. Available online: <https://www.ncbi.nlm.nih.gov/books/NBK441874/> (accessed on 20 May 2023).
21. Hashemi, S.M.H. *Crystal Clean: Brain Tumors MRI Dataset [Data Set]*; Kaggle: San Francisco, CA, USA, 2023. [[CrossRef](#)]
22. American Academy of Dermatology. Available online: <https://www.aad.org/public/diseases/skin-cancer/find/at-risk/abcdes> (accessed on 19 August 2024).
23. Duarte, A.F.; Sousa-Pinto, B.; Azevedo, L.F.; Barros, A.M.; Puig, S.; Malveyh, J.; Haneke, E.; Correia, O. Clinical ABCDE rule for early melanoma detection. *Eur. J. Dermatol.* **2021**, *31*, 771–778. [[CrossRef](#)] [[PubMed](#)]
24. Tan, A.; Greenwald, E.; Bajaj, S.; Belen, D.; Sheridan, T.; Stein, J.A.; Liebman, T.N.; Bowling, A.; Polsky, D. Melanoma surveillance for high-risk patients via telemedicine: Examination of real-world data from an integrated store-and-forward total body photography and dermoscopy service. *J. Am. Acad. Dermatol.* **2022**, *86*, 191–192. [[CrossRef](#)]
25. Shaver, J. The State of Telehealth Before and After the COVID-19 Pandemic. *Prim Care* **2022**, *49*, 517–530. [[CrossRef](#)] [[PubMed](#)] [[PubMed Central](#)]
26. Bhavesh Mittal. 2023. Available online: <https://www.kaggle.com/datasets/bhaveshmittal/melanoma-cancer-dataset/data> (accessed on 19 August 2024).
27. Mafi, M.; Martin, H.; Cabrerizo, M.; Andrian, J.; Barreto, A.; Adjouadi, M. A comprehensive survey on impulse and Gaussian denoising filters for digital images. *Signal Process.* **2019**, *157*, 236–260. [[CrossRef](#)]
28. Kingma, D.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the International Conference on Learning Representations, Banff, AB, Canada, 14–16 April 2014.
29. Loshchilov, I.; Hutter, F. Decoupled Weight Decay Regularization. In Proceedings of the International Conference on Learning Representations, Toulon, France, 24–26 April 2017.
30. Sheng, M.; Li, J.; Bhatti, U.A.; Liu, J.; Huang, M.; Chen, Y. Zero Watermarking Algorithm for Medical Image Based on Resnet50-DCT. *Comput. Mater. Contin.* **2023**, *75*, 293–309. [[CrossRef](#)]
31. Zhang, Y.; Liu, Y.L.; Nie, K.; Zhou, J.; Chen, Z.; Chen, J.H.; Wang, X.; Kim, B.; Parajuli, R.; Mehta, R.S.; et al. Deep Learning-based Automatic Diagnosis of Breast Cancer on MRI Using Mask R-CNN for Detection Followed by ResNet50 for Classification. *Acad. Radiol.* **2023**, *30* (Suppl. 2), S161–S171. [[CrossRef](#)] [[PubMed](#)] [[PubMed Central](#)]
32. Shah, N.U.H.; Mahum, R.; Nisar, D.e.M.; Aman, N.U.; Azim, T. Breast Cancer Identification Using Improved DarkNet53 Model. In *Innovations in Bio-Inspired Computing and Applications. IBICA 2022*; Lecture Notes in Networks and Systems; Abraham, A., Bajaj, A., Gandhi, N., Madureira, A.M., Kahraman, C., Eds.; Springer: Cham, Switzerland, 2023; Volume 649. [[CrossRef](#)]
33. Xu, W.; Fu, Y.L.; Zhu, D. ResNet and its application to medical image processing: Research progress and challenges. *Comput. Methods Programs Biomed.* **2023**, *240*, 107660. [[CrossRef](#)] [[PubMed](#)]
34. Matzinger, H.; Allgeier, A. CNN Image Recognition is Mainly Based on Local Features. In Proceedings of the 2022 4th International Conference on Robotics and Computer Vision (ICRCV), Wuhan, China, 25–27 September 2022; pp. 90–95. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.