bioengineering

MDPI

Article

Deep Transfer Learning Using Real-World Image Features for Medical Image Classification, with a Case Study on Pneumonia X-ray Images

Chanhoe Gu [1] and Minhyeok Lee [1,2,*]

[1] Department of Intelligent Semiconductor Engineering, Chung-Ang University, Seoul 06974, Republic of Korea; kum0100@cau.ac.kr
[2] School of Electrical and Electronics Engineering, Chung-Ang University, Seoul 06974, Republic of Korea
* Correspondence: mlee@cau.ac.kr

Abstract: Deep learning has profoundly influenced various domains, particularly medical image analysis. Traditional transfer learning approaches in this field rely on models pretrained on domain-specific medical datasets, which limits their generalizability and accessibility. In this study, we propose a novel framework called real-world feature transfer learning, which utilizes backbone models initially trained on large-scale general-purpose datasets such as ImageNet. We evaluate the effectiveness and robustness of this approach compared to models trained from scratch, focusing on the task of classifying pneumonia in X-ray images. Our experiments, which included converting grayscale images to RGB format, demonstrate that real-world-feature transfer learning consistently outperforms conventional training approaches across various performance metrics. This advancement has the potential to accelerate deep learning applications in medical imaging by leveraging the rich feature representations learned from general-purpose pretrained models. The proposed methodology overcomes the limitations of domain-specific pretrained models, thereby enabling accelerated innovation in medical diagnostics and healthcare. From a mathematical perspective, we formalize the concept of real-world feature transfer learning and provide a rigorous mathematical formulation of the problem. Our experimental results provide empirical evidence supporting the effectiveness of this approach, laying the foundation for further theoretical analysis and exploration. This work contributes to the broader understanding of feature transferability across domains and has significant implications for the development of accurate and efficient models for medical image analysis, even in resource-constrained settings.

Keywords: deep learning; transfer learning; medical image analysis; X-ray images; pneumonia classification; convolutional neural networks; feature extraction

MSC: 68T27

Citation: Gu, C.; Lee, M. Deep Transfer Learning Using Real-World Image Features for Medical Image Classification, with a Case Study on Pneumonia X-ray Images. *Bioengineering* **2024**, *11*, 406. https://doi.org/10.3390/bioengineering11040406

Academic Editors: Andrea Cataldo, Ming Liu, Liquan Dong and Qingliang Jiao

Received: 3 April 2024
Revised: 14 April 2024
Accepted: 18 April 2024
Published: 20 April 2024

1. Introduction

Deep learning has profoundly transformed various domains, particularly in the field of medical image analysis [1–3]. Convolutional Neural Networks (CNNs) have demonstrated exceptional performance in automatically extracting hierarchical features from images, leading to significant advancements in applications such as object recognition, image segmentation, and medical diagnostics [4–8].

Transfer learning has emerged as a fundamental technique in deep learning, allowing models to leverage knowledge acquired from source tasks to improve performance on target tasks [9–11]. This approach commonly employs pretrained models initially trained on large-scale datasets such as ImageNet as feature extractors for related tasks. These pretrained models capture a wide range of features, from low-level edges and textures to high-level semantic information, which can be effectively transferred to downstream tasks.

However, the application of transfer learning in the context of medical image analysis presents unique challenges. Medical images such as X-rays, Computed Tomography (CT) scans, and Magnetic Resonance Imaging (MRI) exhibit distinct characteristics compared to natural images. These images are typically grayscale, have high resolution, and contain intricate anatomical structures and pathological patterns [12–14]. Consequently, the direct application of pretrained models from natural image domains to medical image tasks may not always yield optimal results.

To address this challenge, current approaches in medical image transfer learning often rely on domain-specific pretrained models. These models are trained on large-scale medical image datasets, which capture the unique characteristics of medical images. While these domain-specific pretrained models have shown promising results, they have limitations in terms of accessibility and generalizability. The availability of large-scale annotated medical image datasets is limited, and the development of domain-specific pretrained models requires substantial computational resources and expertise.

In this paper, we propose a novel approach to transfer learning for medical image analysis, which we term "real-world feature transfer learning". Our approach leverages pretrained models from general image domains such as ImageNet as feature extractors for medical image tasks. We hypothesize that the features learned by these models, despite being derived from natural images, can provide meaningful representations for medical images. By utilizing readily available pretrained models, our approach aims to overcome the limitations of domain-specific pretrained models and enable more accessible and efficient transfer learning for medical image analysis.

To validate our approach, we focus on the task of pneumonia detection using chest X-ray images. Pneumonia is a common respiratory infection that affects millions of people worldwide and is a leading cause of mortality, particularly in children and the elderly [15]. Accurate and timely diagnosis of pneumonia is crucial for effective treatment and patient management. Chest X-ray imaging is the primary diagnostic tool for pneumonia; however, the interpretation of these images can be challenging, even for experienced radiologists [16].

We conducted extensive experiments to evaluate the effectiveness of real-world feature transfer learning for pneumonia detection. We employed various conventional CNN architectures, including ResNet [17] and DenseNet [18], pretrained on the ImageNet dataset. To adapt these models to the grayscale nature of chest X-ray images, we propose a simple yet effective technique of replicating the grayscale channel to form a three-channel input compatible with the pretrained models.

Our experimental results demonstrate that real-world-feature transfer learning achieves superior performance compared to training models from scratch on the pneumonia detection task. The pretrained models exhibited faster convergence along with higher accuracy, precision, recall, and F1 scores. These findings suggest that the features learned from natural images can indeed be effectively transferred to medical image domains, even when the source and target domains have significant differences.

The significance of our work lies in its potential to democratize deep learning for medical image analysis. By leveraging readily available pretrained models from general image domains, our approach reduces the reliance on domain-specific pretrained models and large-scale annotated medical image datasets. This can facilitate the development of accurate and efficient models for various medical image analysis tasks even in resource-constrained settings. Furthermore, our approach opens up new avenues for exploring the transferability of features across different domains and modalities, potentially leading to more generalized and robust deep learning models.

From a mathematical perspective, our work contributes to the understanding of feature representations and their transferability across domains. We formalize the concept of real-world feature transfer learning and provide a rigorous mathematical formulation of the problem. Our experimental results provide empirical evidence supporting the effectiveness of this approach, laying the foundation for further theoretical analysis and exploration.

The main contributions of our work are as follows:

- We introduce the concept of real-world feature transfer learning, which leverages pretrained models from general image domains for medical image analysis tasks.
- We propose a simple and effective technique for adapting pretrained models to grayscale medical images by replicating the grayscale channel.
- We conduct extensive experiments on the pneumonia detection task using chest X-ray images, demonstrating the superiority of real-world feature transfer learning over training models from scratch.
- We provide empirical evidence supporting the feasibility and effectiveness of transferring knowledge from natural image domains to medical image domains, opening up new possibilities for accessible and efficient transfer learning in medical image analysis.

2. Related Work

Transfer learning has gained significant attention in the deep learning community, particularly in the field of medical image analysis. The concept of transfer learning involves leveraging knowledge acquired from a source domain to improve performance on a target domain [19,20]. In the context of deep learning, transfer learning often involves using pretrained models trained on large-scale datasets as feature extractors or performing initialization for target tasks [21–23].

Recent studies have investigated the effectiveness of transfer learning for various medical image analysis tasks. Alzubaidi et al. [24] proposed a novel transfer learning approach for skin and breast cancer classification with medical images. They trained a deep convolutional neural network (DCNN) model on large unlabeled medical image datasets and then transferred the knowledge to train the model on a small amount of labeled medical images. Their approach showed significantly improved performance on both skin and breast cancer classification tasks compared to training from scratch. In the domain of chest X-ray analysis, transfer learning has been extensively explored in recent years. Rahman et al. [25] investigated the effectiveness of transfer learning for pneumonia detection in chest X-rays. They evaluated various pretrained CNN architectures, such as AlexNet, ResNet18, DenseNet201, and SqueezeNet, and found that fine-tuning these models led to improved performance compared to training from scratch. Chouhan et al. [26] proposed a novel transfer learning approach for pneumonia detection using an ensemble of pretrained CNN models. Their approach achieved state-of-the-art performance on a public dataset, outperforming individual CNN models.

While these studies have shown promising results, they primarily focus on transfer learning within the medical domain, using pretrained models that have been trained on medical image datasets. In contrast, our work explores the feasibility of transfer learning from general image domains to medical image domains by leveraging pretrained models from datasets such as ImageNet.

The concept of transferring knowledge from general image domains to medical image domains has gained attention in recent studies. Raghu et al. [27] investigated the transferability of features from natural image datasets to medical image tasks. They found that pretrained models from ImageNet can be effectively adapted to medical image classification tasks, achieving comparable performance to models trained from scratch on medical datasets.

Building upon these findings, our work aims to provide a comprehensive analysis of the effectiveness of real-world feature transfer learning for medical image analysis, focusing on the task of pneumonia detection in chest X-rays. We extend the existing literature by investigating the transferability of features from general image domains to the specific domain of chest X-ray analysis, and propose a simple yet effective technique for adapting pretrained models to grayscale medical images.

In addition to transfer learning, our work is related to the broader field of deep learning for medical image analysis. Deep learning has revolutionized the field of medical image analysis, enabling the development of automated systems for various tasks, including classification, detection, and segmentation [2,28]. CNNs have become the dominant

architecture in medical image analysis thanks to their ability to learn hierarchical features directly from raw image data [29,30].

Several deep learning-based approaches have been proposed for pneumonia detection in chest X-rays in recent years. Ayan and Ünver [31] developed a deep learning model using a custom CNN architecture for pneumonia detection. They achieved high accuracy and sensitivity on a public dataset, demonstrating the effectiveness of deep learning for this task. Our work extends these studies by specifically focusing on the transfer learning aspect and exploring the feasibility of leveraging pretrained models from general image domains for pneumonia detection. We provide a comprehensive analysis of different CNN architectures and their performance when utilized within a real-world feature transfer learning framework.

3. Background

Transfer learning is a machine learning technique that leverages knowledge gained from solving one problem and applies it to a different but related problem [19]. The goal of transfer learning is to improve learning performance in the target domain utilizing the knowledge learned from the source domain. This section provides a mathematical formulation of transfer learning and its key concepts.

3.1. Problem Formulation

Definition 1. *A domain $\mathcal{D}$ consists of a feature space $\mathcal{X}$ and a marginal probability distribution $P(X)$, where $X = \{x_1, \ldots, x_n\} \in \mathcal{X}$.*

Definition 2. *Given a domain $\mathcal{D} = \{\mathcal{X}, P(X)\}$, a task $\mathcal{T}$ consists of a label space $\mathcal{Y}$ and a conditional probability distribution $P(Y|X)$, which is learned from the training data $\{(x_i, y_i)\}$, where $x_i \in X$ and $y_i \in \mathcal{Y}$.*

Definition 3. *Given a source domain $\mathcal{D}_S$ and corresponding source task $\mathcal{T}_S$ along with a target domain $\mathcal{D}_T$ and corresponding target task $\mathcal{T}_T$, transfer learning aims to improve the learning of the targeted conditional probability distribution $P(Y_T|X_T)$ in $\mathcal{D}_T$ using the knowledge learned from $\mathcal{D}_S$ and $\mathcal{T}_S$, where $\mathcal{D}_S \neq \mathcal{D}_T$ or $\mathcal{T}_S \neq \mathcal{T}_T$.*

3.2. Categories of Transfer Learning

Transfer learning can be categorized into three main settings based on the differences between the source and target domains and the tasks involved [19].

3.2.1. Inductive Transfer Learning

Definition 4. *In inductive transfer learning, the target task is different from the source task, regardless of whether or not the source and target domains are the same. In this setting, the target domain labeled data are available, and the goal is to improve the target task performance using the knowledge learned from the source domain.*

Given a source domain $\mathcal{D}_S$ and corresponding source task $\mathcal{T}_S$ along with a target domain $\mathcal{D}_T$ and corresponding target task $\mathcal{T}_T$, inductive transfer learning aims to improve the learning of the targeted conditional probability distribution $P(Y_T|X_T)$ using the knowledge learned from $\mathcal{D}_S$ and $\mathcal{T}_S$, where $\mathcal{T}_S \neq \mathcal{T}_T$.

3.2.2. Transductive Transfer Learning

Definition 5. *In transductive transfer learning, the source and target tasks are the same, while the source and target domains are different. In this setting, the target domain labeled data are not available, and the goal is to improve the target task performance using the knowledge learned from the source domain.*

Given a source domain $\mathcal{D}_S$ and corresponding source task $\mathcal{T}_S$ along with a target domain $\mathcal{D}_T$ and corresponding target task $\mathcal{T}_T$, transductive transfer learning aims to improve the learning of the target conditional probability distribution $P(Y_T|X_T)$ using the knowledge learned from $\mathcal{D}_S$ and $\mathcal{T}_S$, where $\mathcal{D}_S \neq \mathcal{D}_T$ and $\mathcal{T}_S = \mathcal{T}_T$.

3.2.3. Unsupervised Transfer Learning

Definition 6. *In unsupervised transfer learning, the target task is different from but related to the source task, and no labeled data are available in either the source or target domains. The goal is to improve the target task performance using the knowledge learned from the source domain in an unsupervised manner.*

Given a source domain $\mathcal{D}_S$ and corresponding source task $\mathcal{T}_S$ along with a target domain $\mathcal{D}_T$ and corresponding target task $\mathcal{T}_T$, unsupervised transfer learning aims to improve the learning of the targeted conditional probability distribution $P(Y_T|X_T)$ using the knowledge learned from $\mathcal{D}_S$ and $\mathcal{T}_S$, where $\mathcal{T}_S \neq \mathcal{T}_T$ and where no labeled data are available in either $\mathcal{D}_S$ or $\mathcal{D}_T$.

3.3. Transfer Learning in Deep Neural Networks

Deep neural networks have shown remarkable success in various machine learning tasks, particularly in computer vision and natural language processing [32,33]. Transfer learning in deep neural networks typically involves using a pretrained model trained on a large-scale source dataset as a starting point for a target task with limited labeled data.

Definition 7. *Let $f_S : \mathcal{X}_S \to \mathcal{Y}_S$ be a pretrained deep neural network model for source task $\mathcal{T}_S$, parameterized by θ_S. The goal of transfer learning in deep neural networks is to learn a target model $f_T : \mathcal{X}_T \to \mathcal{Y}_T$, parameterized by θ_T, by leveraging the knowledge learned from f_S.*

The transfer learning process in deep neural networks can be formalized as follows:

$$\theta_T^{(0)} = \theta_S^*, \tag{1}$$

$$\theta_T^* = \arg\min_{\theta_T} \mathcal{L}_T(f_T(X_T; \theta_T), Y_T), \tag{2}$$

where θ_S^* represents the optimal parameters of the pretrained source model f_S, $\theta_T^{(0)}$ represents the initial parameters of the target model f_T, $\mathcal{L}_T$ is the loss function for the target task, and θ_T^* represents the optimal parameters of the target model after fine-tuning.

Remark 1. *The pretrained source model f_S can be used in various ways for transfer learning, such as:*

- *Using f_S as a fixed feature extractor and training a new classifier on top of the extracted features for the target task.*
- *Fine-tuning the entire network f_S using the target task data, allowing the pretrained parameters to adapt to the target domain.*
- *Freezing some layers of f_S and fine-tuning the remaining layers for the target task, balancing the knowledge transfer and adaptation to the target domain.*

The choice of transfer learning strategy depends on factors such as the size of the target dataset, the similarity between the source and target domains, and the available computational resources.

3.4. Transfer Learning as a Foundation for Segmentation and Object Detection

The paradigm of transfer learning, in which pretrained models serve as the backbone for high-level vision tasks, has become a fundamental approach in the fields of segmentation

and object detection [34–39]. This section provides a rigorous mathematical formulation of how this methodology is adapted for these specific tasks.

3.4.1. Segmentation

Definition 8. *Let $\mathcal{X}$ be the input space of images and let $\mathcal{Y}$ be the output space of pixel-wise class labels. The goal of image segmentation is to learn a mapping function $f_{seg} : \mathcal{X} \to \mathcal{Y}$ that assigns a class label to each pixel in an image.*

Assumption 1. *We assume the existence of a pretrained model, represented by parameters θ^*, that has learned meaningful and generalizable representations from a large-scale dataset $\mathcal{D}_S$.*

Proposition 1. *The mapping function f_{seg} can be decomposed into two components: the pretrained backbone model, parameterized by θ^*, and additional segmentation-specific layers, parameterized by ψ.*

$$f_{seg}(x; \theta^*, \psi) = \psi \circ \theta^*(x) \quad \forall x \in \mathcal{X} \tag{3}$$

Definition 9. *Let x_i denote a pixel in an image $x \in \mathcal{X}$ and let y_i be its corresponding segmentation label. The objective of the segmentation task is to learn the optimal parameters ψ^*_{seg} that minimize the expected loss over the target dataset $\mathcal{D}_T$:*

$$\psi^*_{seg} = \arg\min_{\psi} \mathbb{E}_{x,y\sim\mathcal{D}_T} \left[\frac{1}{|\mathcal{X}|} \sum_{i=1}^{|\mathcal{X}|} L(y_i, f_{seg}(x_i; \theta^*, \psi)) \right] \tag{4}$$

where L is a suitable loss function, such as the cross-entropy loss, and $|\mathcal{X}|$ denotes the total number of pixels in the image.

3.4.2. Object Detection

Definition 10. *Let $\mathcal{B}$ be the space of bounding boxes and let $\mathcal{C}$ be the space of object classes. The goal of object detection is to learn a mapping function $f_{det} : \mathcal{X} \to \mathcal{B} \times \mathcal{C}$ that predicts a set of bounding boxes and their associated class labels for objects in an image.*

Proposition 2. *Similar to segmentation, the mapping function f_{det} can be decomposed into the pretrained backbone model, parameterized by θ^*, and additional detection-specific layers, parameterized by ψ.*

$$f_{det}(x; \theta^*, \psi) = \psi \circ \theta^*(x) \quad \forall x \in \mathcal{X} \tag{5}$$

Definition 11. *Let $y_{b,c}$ denote the ground-truth bounding boxes and their associated class labels for an image $x \in \mathcal{X}$. The objective of the object detection task is to learn the optimal parameters ψ^*_{det} that minimize the expected loss over the target dataset $\mathcal{D}_T$:*

$$\psi^*_{det} = \arg\min_{\psi} \mathbb{E}_{x,y\sim\mathcal{D}_T} [L(y_{b,c}, f_{det}(x; \theta^*, \psi))] \tag{6}$$

where L is a loss function that typically includes terms for both the bounding box coordinates and the class labels.

Remark 2. *In practice, it is common to fine-tune the entire network, both θ and ψ, on the target task dataset in order to achieve better performance. This approach differs from conventional transfer learning, where only the target task parameters ψ are fine-tuned while keeping the source task parameters θ fixed.*

Proposition 3. *The fine-tuning process for segmentation and object detection can be formulated as follows:*

$$(\theta^*_{seg}, \psi^*_{seg}) = \arg\min_{\theta,\psi} \mathbb{E}_{x,y\sim\mathcal{D}_T}\left[\frac{1}{|\mathcal{X}|}\sum_{i=1}^{|\mathcal{X}|} L(y_i, f_{seg}(x_i;\theta,\psi))\right] \quad (7)$$

$$(\theta^*_{det}, \psi^*_{det}) = \arg\min_{\theta,\psi} \mathbb{E}_{x,y\sim\mathcal{D}_T}[L(y_{b,c}, f_{det}(x;\theta,\psi))] \quad (8)$$

where θ^*_{seg} and θ^*_{det} denote the respective fine-tuned backbone parameters for segmentation and object detection.

Remark 3. *The fine-tuning process allows the pretrained backbone to adapt its learned representations to the specific characteristics of the target task, leading to improved performance compared to using fixed backbone parameters.*

3.5. Examples of Transfer Learning

Transfer learning has become a fundamental technique in deep learning, particularly in the field of computer vision. The main idea behind transfer learning is to leverage the knowledge gained from a source task and apply it to a related target task, thereby improving the performance and efficiency of learning on the target task. This is especially useful when the target task has limited labeled data, as is often the case in medical image analysis.

In the context of deep learning, transfer learning typically involves using a pretrained neural network model as a starting point for the target task. The pretrained model is usually trained on a large-scale dataset, such as ImageNet, which contains millions of labeled images spanning a wide range of object categories. During the pretraining phase, the model learns a rich set of features that capture various aspects of the images, including low-level edges and textures as well as high-level semantic information.

When applying transfer learning to a target task, such as medical image classification, the pretrained model is used as a feature extractor. The weights of the pretrained model are initialized with the learned values from the source task, and the model is then fine-tuned on the target task dataset. Fine-tuning involves updating the weights of the pretrained model using the labeled data from the target task, allowing the model to adapt its learned features to the specific characteristics of the target domain.

There are several strategies for fine-tuning a pretrained model. One common approach is to freeze the weights of the early layers of the network, which capture general low-level features, and only fine-tune the later layers, which capture more task-specific high-level features; this allows the model to retain the knowledge learned from the source task while adapting to the target task. Another approach is to fine-tune the entire network, allowing all of the weights to be updated based on the target task data. The choice of fine-tuning strategy depends on factors such as the size of the target dataset, the similarity between the source and target domains, and the available computational resources.

Transfer learning has been successfully applied to a wide range of computer vision tasks, including image classification, object detection, and semantic segmentation. For example, in image classification, transfer learning has been used to improve the performance of models trained on limited labeled data by leveraging pretrained models from large-scale datasets. In object detection, transfer learning has been used to adapt pretrained models trained on general object detection datasets to specific domains, including medical imaging, by fine-tuning the models on domain-specific data. Similarly, in semantic segmentation, transfer learning has been used to improve the performance of models by initializing them with weights learned from pretraining on large-scale datasets and then fine-tuning them on the target task data.

The effectiveness of transfer learning in deep learning can be attributed to several factors. First, deep neural networks have the ability to learn hierarchical features that capture different levels of abstraction in the data. By pretraining on a large-scale dataset, the model learns a rich set of features that can be transferred to related tasks. Second, the use of transfer learning allows for more efficient learning on the target task, as the model

does not need to learn the features from scratch. This is particularly important when the target task has limited labeled data, as the model can leverage the knowledge learned from the source task to improve its performance. Finally, transfer learning can help to reduce overfitting on the target task, as the pretrained model has already learned a robust set of features that generalize well to related tasks.

In the context of medical image analysis, transfer learning has been widely adopted to address the challenges of limited labeled data and the need for efficient and accurate models. Medical imaging datasets are often small and expensive to annotate, making it difficult to train deep learning models from scratch. By leveraging pretrained models from large-scale datasets, transfer learning allows for the development of accurate models for medical image analysis tasks, even with limited labeled data. This has led to significant improvements in the performance of deep learning models for tasks such as medical image classification, segmentation, and detection.

Despite the success of transfer learning in medical image analysis, there are still challenges and opportunities for further research. One challenge is the domain shift between the source and target tasks, which can limit the effectiveness of transfer learning. This is particularly relevant when transferring knowledge from natural image datasets to medical image datasets, as there can be significant differences in the characteristics of the images and the underlying data distributions. To address this challenge, techniques such as domain adaptation and unsupervised domain adaptation have been proposed to align the feature spaces of the source and target domains and improve the transferability of the learned features.

4. Proposition: Similarity of Learned Features in Conventional and Medical Images

This section presents a mathematical framework to demonstrate the similarity between the features learned by deep learning models from conventional images and medical images. We establish this proposition through a series of definitions, assumptions, and mathematical formulations.

Definition 12. *Let $\mathcal{X}_C$ and $\mathcal{X}_M$ denote the input spaces of conventional images and medical images, respectively. We define a feature extraction function $\phi : \mathcal{X} \to \mathcal{F}$, where $\mathcal{F}$ is the feature space and $\mathcal{X} = \mathcal{X}_C \cup \mathcal{X}_M$.*

Assumption 2. *We assume that the feature extraction function ϕ is a deep learning model parameterized by θ which has been trained on a large dataset of conventional images $\mathcal{D}_C = \{(x_i, y_i)\}_{i=1}^{N_C}$, where $x_i \in \mathcal{X}_C$ and y_i is the corresponding label.*

Definition 13. *Let $\mathcal{F}_C$ and $\mathcal{F}_M$ be the feature spaces corresponding to conventional and medical images, respectively, such that $\mathcal{F}_C = \phi(\mathcal{X}_C)$ and $\mathcal{F}_M = \phi(\mathcal{X}_M)$.*

Proposition 4. *The features learned by the deep learning model ϕ from conventional images are similar to the features of medical images, i.e., $\mathcal{F}_C \approx \mathcal{F}_M$.*

To prove this proposition, we introduce a measure of similarity between the feature spaces.

Definition 14. *Let $d : \mathcal{F} \times \mathcal{F} \to \mathbb{R}$ be a distance function that quantifies the dissimilarity between two feature vectors. We define the average distance between the feature spaces $\mathcal{F}_C$ and $\mathcal{F}_M$ as*

$$D(\mathcal{F}_C, \mathcal{F}_M) = \frac{1}{|\mathcal{X}_C||\mathcal{X}_M|} \sum_{x_C \in \mathcal{X}_C} \sum_{x_M \in \mathcal{X}_M} d(\phi(x_C), \phi(x_M)). \tag{9}$$

Assumption 3. *We assume that the distance function d satisfies the properties of a metric, i.e., non-negativity, identity of indiscernibles, symmetry, and triangle inequality.*

Lemma 1. *If the average distance between the feature spaces $\mathcal{F}_C$ and $\mathcal{F}_M$ is small, i.e., if $D(\mathcal{F}_C, \mathcal{F}_M) < \epsilon$ for some small $\epsilon > 0$, then the features learned from conventional images are similar to the features of medical images.*

Proof. Let $x_C \in \mathcal{X}_C$ and $x_M \in \mathcal{X}_M$ be arbitrary conventional and medical images, respectively. From the definition of the average distance and the assumption that $D(\mathcal{F}_C, \mathcal{F}_M) < \epsilon$, we have

$$d(\phi(x_C), \phi(x_M)) \leq \sum_{x'_C \in \mathcal{X}_C} \sum_{x'_M \in \mathcal{X}_M} d(\phi(x'_C), \phi(x'_M)), \tag{10}$$

$$= |\mathcal{X}_C||\mathcal{X}_M| D(\mathcal{F}_C, \mathcal{F}_M), \tag{11}$$

$$< |\mathcal{X}_C||\mathcal{X}_M|\epsilon. \tag{12}$$

This implies that, for any pair of conventional and medical images, the distance between their feature representations is bounded by a small value proportional to ϵ; therefore, the features learned from conventional images are similar to the features of medical images. □

Remark 4. *The choice of the distance function d is crucial for quantifying the similarity between feature spaces. Common choices include the Euclidean distance, cosine distance, and more advanced metrics such as the Fréchet Inception Distance (FID) [40].*

Conjecture 1. *The similarity between the features learned from conventional images and the features of medical images can be further improved by fine-tuning the pretrained model ϕ on a small dataset of medical images $\mathcal{D}_M = \{(x_i, y_i)\}_{i=1}^{N_M}$, where $x_i \in \mathcal{X}_M$ and y_i is the corresponding label.*

Proposition 5. *Fine-tuning the pretrained model ϕ on the medical image dataset $\mathcal{D}_M$ results in an updated feature extraction function ϕ' that minimizes the average distance between the feature spaces $\mathcal{F}_C$ and $\mathcal{F}_M$.*

Proof. The fine-tuning process can be formulated as an optimization problem:

$$\theta^* = \arg\min_{\theta} \frac{1}{N_M} \sum_{i=1}^{N_M} L(y_i, f(\phi'(x_i; \theta))) \tag{13}$$

where L is a suitable loss function and f is a task-specific function (e.g., a classifier or segmentation model) that operates on the features extracted by ϕ'.

By minimizing the loss function on the medical image dataset, the fine-tuning process adapts the parameters θ to better capture the characteristics of medical images. As a result, the updated feature extraction function ϕ' maps medical images to a feature space $\mathcal{F}'_M$ that is closer to the feature space $\mathcal{F}_C$ of conventional images, i.e., $D(\mathcal{F}_C, \mathcal{F}'_M) < D(\mathcal{F}_C, \mathcal{F}_M)$. □

5. Method

Figure 1 illustrates the overall framework of our proposed real-world feature transfer learning approach for medical image classification. Our methodology aims to reconcile the commonly held belief that transfer learning between general image datasets such as ImageNet and specialized medical image datasets such as X-ray images is challenging due to the disparity in their features. Typically, deep learning models trained on datasets such as ImageNet learn to recognize a plethora of real-world objects, such as the shape of animals, color of fruits, structure of machines, etc. In contrast, medical images are predominantly grayscale, and contain intricate patterns and details that represent health conditions and diseases. This substantial difference in features makes transfer learning a challenging task.

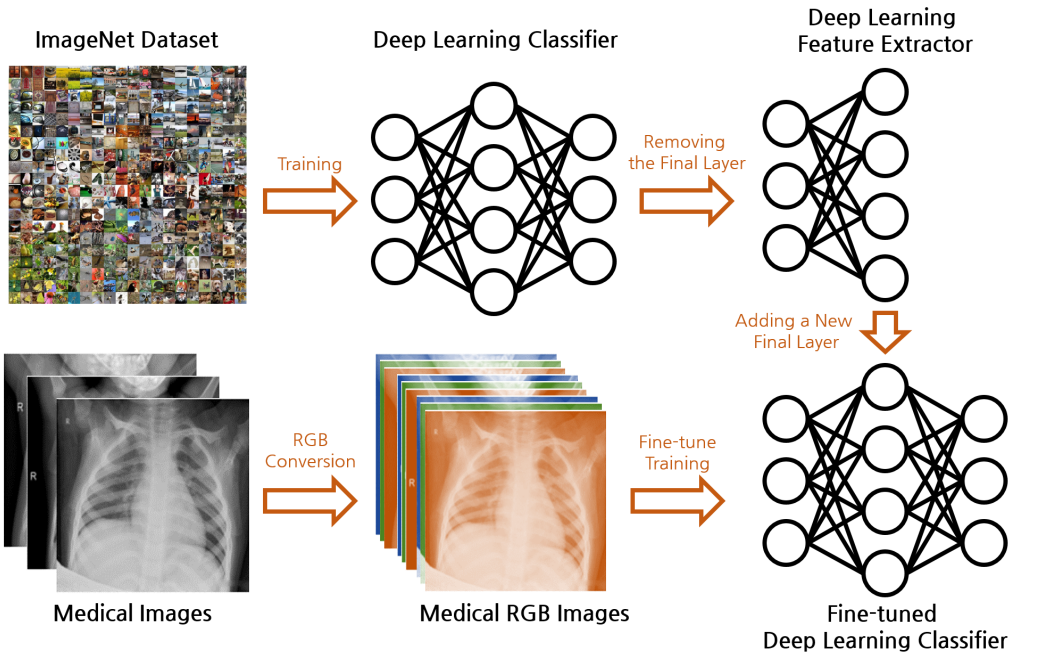

Figure 1. Overall framework of the proposed real-world feature transfer learning approach for medical image classification.

Further, a conventional deep learning model expects RGB images as input, necessitating the alteration of the model structure to accommodate grayscale images, which often complicates the transfer learning process. In our methodology, we propose an approach that circumvents these limitations and demonstrate the feasibility of transfer learning between these dissimilar domains.

5.1. Data Preparation

Medical images are generally grayscale, representing the information in a single color channel, whereas typical real-world images are composed of three color channels (RGB). In order to apply a pretrained model expecting three channels to medical images, we propose a simple yet effective approach: we replicate the grayscale channel three times to mimic the structure of an RGB image, ensuring the compatibility of the medical images with the pretrained model without requiring any modification to the model's architecture.

$$I_{RGB} = I_{gray} \otimes [1, 1, 1] \tag{14}$$

Here, I_{RGB} is the transformed RGB image, I_{gray} is the original grayscale medical image, and $[1, 1, 1]$ is the vector used to replicate the grayscale channel.

5.2. Model Architecture

Our model architecture is designed to leverage the power of transfer learning while adapting to the specific characteristics of medical image classification. The architecture consists of two primary components: a pretrained model serving as the backbone feature extractor, and an appended layer(s) for task-specific fine-tuning.

The pretrained model, denoted as $\mathcal{F}(x; \theta, \phi)$, is a variant of deep CNN that has been previously trained on a large-scale dataset such as ImageNet. This pretrained model has learned a rich set of features that capture the intricate patterns and structures present in natural images. By leveraging these learned features, the knowledge gained from the source domain can be effectively transferred to the target domain of medical image classification.

In our architecture, we utilize the pretrained model as a fixed feature extractor. This means that we freeze the weights of the pretrained model layers, represented by the parameters θ, during the fine-tuning process. By keeping these weights fixed, we preserve

the valuable feature representations learned from the source domain and prevent overfitting to the limited labeled data in the target domain.

To adapt the pretrained model to the specific task of medical image classification, we append one or more layers, represented by $\mathcal{G}(x;\theta,\psi)$, on top of the pretrained model. These appended layers are designed to learn the task-specific mapping between the extracted features and the corresponding class labels. The parameters of these appended layers, denoted as ψ, are trainable, and are fine-tuned using the labeled medical image data.

During the fine-tuning process, the weights of the appended layers are updated using the labeled medical image data. This allows the model to adapt to the specific characteristics and patterns present in the medical images. The fine-tuning process is typically performed using a smaller learning rate compared to the initial training of the pretrained model, as we want to preserve the knowledge gained from the source domain while gradually adapting to the target domain.

By leveraging the power of transfer learning through the use of a pretrained model as the backbone and fine-tuning the appended layers, our model architecture effectively captures the relevant features and patterns in medical images. This approach allows us to benefit from the knowledge gained from large-scale datasets while adapting to the specific requirements of medical image classification tasks.

5.3. Transfer Learning

We employ transfer learning, leveraging the knowledge acquired by the pretrained model from the source domain (general images) and applying it to the target domain (medical images); this technique assumes that the learned feature representations in the source task are beneficial to the learning in the target task.

As discussed in regard to related work, this is expressed mathematically as follows:

$$\psi^* = \arg\min_{\psi} \mathbb{E}_{x,y\sim D_T}[L(y,\mathcal{G}(x;\theta^*,\psi))]. \tag{15}$$

In the above equation, D_T represents the target domain data and $L(y,\mathcal{G}(x;\theta^*,\psi))$ is the loss function that we aim to minimize. The parameters ψ are then fine-tuned based on the medical image data.

5.4. Experimental Setup

Our study employed seven widely used pretrained models as the basis for transfer learning, each of which has demonstrated exceptional performance on the ImageNet dataset. We then tailored these models to our medical imaging task by freezing their convolutional layers and training a new fully connected layer that we appended at the end of each model. Below, we provide more detailed descriptions of each model used:

- **ResNet18 and ResNet50** [17]: The ResNet family of models are residual learning frameworks that provide an effective solution to the problem of training deep networks. They introduce the concept of skip connections, or shortcuts, that allow the gradient to be directly backpropagated to earlier layers. In our study, we used ResNet18 and ResNet50, which consist of 18 and 50 layers, respectively.
- **DenseNet161** [18]: DenseNet161 is a member of the DenseNet family, which connects each layer to every other layer in feed-forward fashion. This allows for feature reuse throughout the network, making the model more parameter-efficient. DenseNet161, the variant we used, consists of 161 layers.
- **ShuffleNet v2 (1.0×)** [41]: ShuffleNet is a highly efficient convolutional neural network designed for mobile devices. It introduces two novel operations, pointwise group convolution and channel shuffle, to reduce computation cost while maintaining model accuracy. We utilized version 2, with a scale factor of 1.0×, in our work.
- **MobileNet v2** [42]: Similar to ShuffleNet, MobileNet v2 is designed for mobile and embedded vision applications. It leverages inverted residuals and linear bottlenecks to create a lightweight yet performant architecture.

- **ResNeXt50 (32x4d)** [43]: ResNeXt50 is a member of the ResNeXt family, which offers a simple and scalable way to increase model capacity without a significant increase in computational cost. The variant we used, denoted as 32x4d, indicates that the network has a cardinality (number of parallel paths) of 32 and a width (number of channels) of 4.
- **Wide ResNet50-2** [44]: Wide ResNet is a variant of the ResNet models with wider convolutional layers instead of deeper ones. The version we used, Wide ResNet50-2, indicates that the network depth is 50 and the width is twice that of the original ResNet50.

Each model was fine-tuned using the Pneumonia X-ray dataset [45] with Binary Cross-Entropy (BCE) as the loss function. We adopted a training approach that preserves the learned features in the pretrained models by freezing the weights in the convolutional layers and only updating the weights in the newly appended fully connected layer.

5.5. Model Description

Each model was obtained from the model pretrained on ImageNet. We retained the entire architecture except for the last layer, which we replaced with a fully connected layer followed by a Sigmoid activation function. This modification was necessary in order to adjust the output size to match our binary classification task (Pneumonia/Normal). These models are each popular deep learning architectures that have shown excellent performance in image classification tasks.

5.6. Hyperparameters

The chosen hyperparameters for our experiments were an epoch size of 20 and a batch size of 64. These values were chosen based on previous studies and our preliminary experiments.

5.7. Data Augmentation

Data augmentation is crucial for training deep learning models, as it can reduce overfitting and improve the model's generalization capability. We applied a series of random transformations, including resizing and cropping, color jittering, horizontal flipping, and rotation. In addition, we transformed the images to grayscale, as the original X-ray images are grayscale, and normalized them to have a mean of 0.5 and a standard deviation of 0.5.

5.8. Dataset

The Pneumonia X-ray dataset, which contains a total of 5863 X-ray images categorized into Pneumonia and Normal, was used for this study. The dataset is divided into training and testing sets, with 1341 normal and 3875 pneumonia images in the training set and 234 normal and 390 pneumonia images in the testing set. Large disparities in the distribution of classes represent a common challenge in medical datasets, which we addressed during the training phase by using class weights in the loss function. The sample distribution and example images of each class are displayed in Table 1 and Figure 2.

Table 1. Distribution of images in the Pneumonia X-ray dataset.

Set	Normal	Pneumonia
Training	1341	3875
Test	234	390

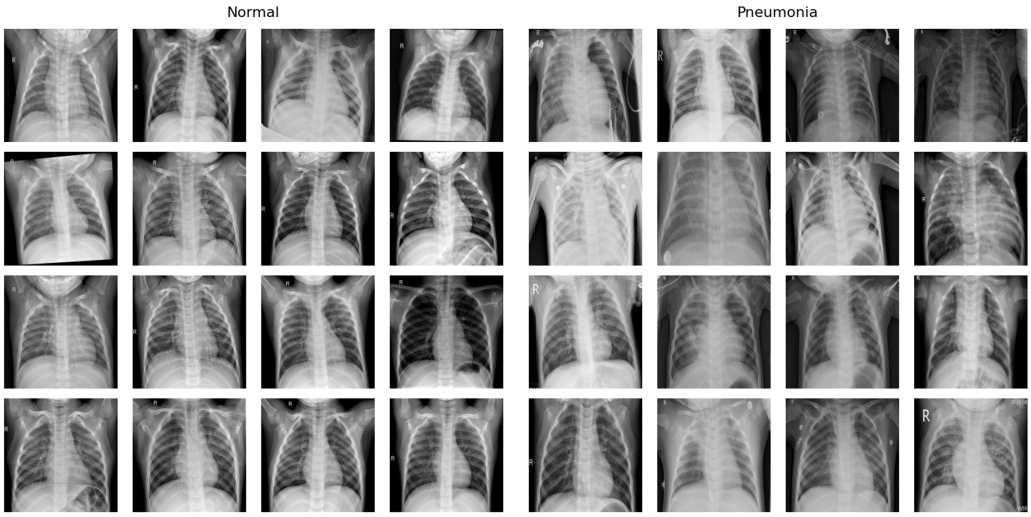

Figure 2. Example X-ray images of each class within the Pneumonia X-ray dataset. The images on the left showcase normal X-ray images and the images on the right showcase pneumonia images.

5.9. Evaluation Metrics

To assess the performance of our models, we utilized four key metrics: accuracy, precision, recall, and F1 score. These metrics provide a comprehensive understanding of model performance by considering the true positive, false positive, true negative, and false negative predictions made by the model.

The accuracy is defined as the proportion of correct predictions (both true positives and true negatives) out of all predictions.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{16}$$

where TP denotes true positives, TN denotes true negatives, FP denotes false positives, and FN denotes false negatives.

Precision (also known as positive predictive value) measures the proportion of correctly predicted positive observations out of the total predicted positives.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{17}$$

Recall (or sensitivity) measures the proportion of correctly predicted positive observations out of the total actual positives.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{18}$$

Finally, the F1 score is the weighted average (harmonic mean) of precision and recall, which helps to gauge the balance between precision and recall. It is particularly useful when the data class distribution is uneven.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{19}$$

The above metrics were computed at the end of each epoch to provide a detailed evaluation of the model's performance over time.

6. Results

6.1. Efficacy of Transfer Learning: Training Results

To validate the efficacy of our approach, we first analyze the results obtained during the training phase. Figure 3 presents a graphical representation of both the training loss and the accuracy achieved by the models over time. The rapid convergence of the models

substantiates the premise that transfer learning from real-world images to medical images is indeed feasible and highly effective.

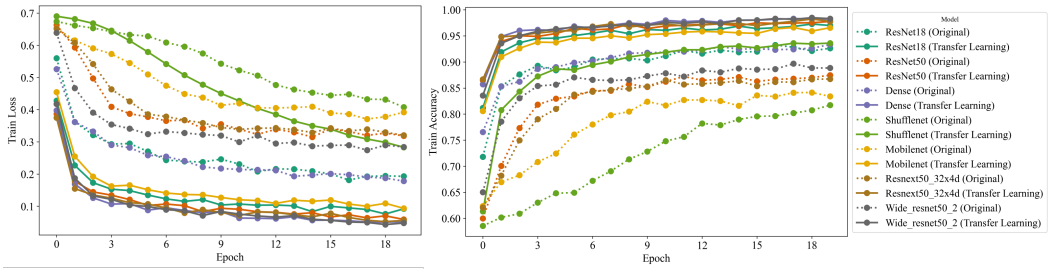

Figure 3. Graphical representation of the training loss and accuracy achieved by the models over time. The solid lines represent the real-world feature transfer learning models, while the dashed lines represent from-scratch training models. The rapid convergence of the proposed models demonstrates the feasibility and effectiveness of transferring learning from real-world images to medical images.

A comprehensive comparison of the different model architectures shows that all of the models except ShuffleNet surpassed 80% accuracy within a single epoch. This rapid ascent in accuracy reflects the efficiency of transfer learning and the ability of the pretrained models to generalize their learning from the source task (general image data) to the target task (medical image data). The adoption of features already learned by the models on a large-scale image dataset (ImageNet) evidently accelerates the learning process, leading to impressive early-stage performance.

Furthermore, we observed that all the models leveraging transfer learning significantly outperformed the regular models (without transfer learning) by reaching an accuracy above 90% after 20 epochs. In stark contrast, the regular models fell short of achieving a similar level of performance. This discrepancy underscores the considerable advantage conferred by transfer learning, particularly the ability to reuse the knowledge from general images, paving the way for high-performance models even in complex domains such as medical imaging.

Among the models we experimented with, ResNeXt50 and Wide ResNet50 performed exceptionally well in the early stages, both achieving over 95% training accuracy within the first few epochs. These models' rapid convergence can be attributed to their deep and wide architectures and unique design principles (skip connections in ResNet), which evidently facilitate feature propagation and mitigate the problem of vanishing gradients.

Conversely, the ShuffleNet model, designed with a focus on computational efficiency, demonstrated a slower convergence rate compared to the other models, only reaching 80% accuracy after several epochs. Nevertheless, it is worth noting that while ShuffleNet trailed in early-stage accuracy, its design philosophy aims to maintain a balance between accuracy and computational efficiency, which is paramount in certain scenarios, particularly mobile and embedded applications.

These experiments yielded compelling evidence supporting the effectiveness of transfer learning from general image data to medical image data. The models leveraging transfer learning exhibited rapid convergence and high accuracy, underscoring the feasibility of our approach and its potential implications for the domain of medical image analysis.

6.2. Comparative Analysis of Models: Test Results

In order to more deeply investigate the performance of our transfer learning approach, we carried out a comprehensive comparison between transfer learning models and models trained from scratch. We evaluated several metrics, including test loss, test accuracy, precision, recall, and F1 score. The results are summarized in Table 2 and Figure 4.

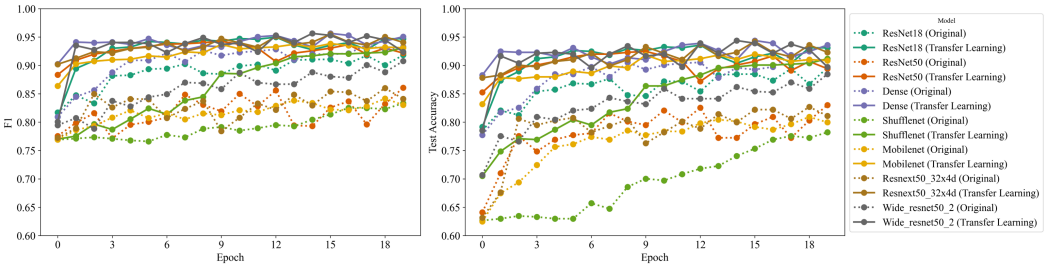

Figure 4. Comprehensive comparison between transfer learning models and models trained from scratch. The line graphs illustrate the performance metrics of test accuracy and F1 score for each model. The solid lines represent the real-world feature transfer learning models, while the dashed lines represent from-scratch training models. The comparison emphasizes the superiority of the transfer learning approach for handling medical image classification.

Table 2. Performance comparison of from-scratch training models vs. real-world feature transfer learning models. Bold text indicates the best results for each experimental setting.

	Model	Test Loss	Test Accuracy	Precision	Recall	F1	G-Mean
ResNet18	Scratch Model	0.272	89.6%	90.1%	93.6%	0.918	0.894
	Transfer Learning	**0.168**	**93.1%**	**91.2%**	**98.5%**	**0.947**	**0.904**
ResNet50	Scratch Model	0.383	83.0%	**88.4%**	83.8%	0.861	**0.858**
	Transfer Learning	**0.293**	**89.4%**	85.7%	**99.7%**	**0.922**	0.827
DenseNet161	Scratch Model	0.301	91.2%	89.2%	97.7%	0.933	0.885
	Transfer Learning	**0.162**	**93.6%**	**91.7%**	**98.7%**	**0.951**	**0.910**
ShuffleNet v2	Scratch Model	0.473	78.2%	79.8%	87.2%	0.833	0.788
	Transfer Learning	**0.354**	**91.2%**	**91.0%**	**95.4%**	**0.931**	**0.908**
MobileNet v2	Scratch Model	0.437	80.0%	**88.2%**	78.5%	0.830	0.846
	Transfer Learning	**0.245**	**90.9%**	88.1%	**98.7%**	**0.931**	**0.866**
ResNeXt50 (32x4d)	Scratch Model	0.426	81.1%	88.6%	80.0%	0.841	0.856
	Transfer Learning	**0.190**	**92.3%**	**90.9%**	**97.4%**	**0.941**	**0.907**
Wide ResNet50-2	Scratch Model	**0.290**	88.5%	**90.8%**	90.8%	0.908	**0.885**
	Transfer Learning	0.307	**89.7%**	86.1%	**99.7%**	**0.924**	0.835

First, it is noteworthy to mention that transfer learning consistently achieved a lower test loss compared to training from scratch for all evaluated models. This demonstrates that the models were able to leverage the knowledge from general images to decrease the error rate, proving the universal applicability of transfer learning to medical image data.

In terms of test accuracy, transfer learning exhibited superior performance across all models. Most notably, there was a significant 3.5% improvement in test accuracy for ResNet18, from 89.6% for the scratch model to 93.1% for the transfer learning model. Similarly, ShuffleNet v2 exhibited a remarkable rise of almost 13% in test accuracy, leaping from 78.2% to 91.2%, proving the efficacy of transfer learning even for models designed to prioritize computational efficiency.

The superiority of transfer learning was further emphasized by the precision, recall, and F1 score results. For instance, ResNet18 saw an improvement in precision from 90.1% with the scratch model to 91.2% with the transfer learning model. This increment indicates that transfer learning contributes to the reduction of false positives. More impressively, the recall jumped from 93.6% to a near-perfect 98.5%, denoting that the model was able to correctly identify almost all positive instances. Consequently, the F1 score, which is the harmonic mean of precision and recall, increased from 0.918 to 0.947, underscoring the improved balance achieved between precision and recall.

Interestingly, in the case of Wide ResNet50-2, the scratch model showed a slightly lower test loss (0.290) compared to the transfer learning model (0.307). Despite this, the

transfer learning model managed to outperform it in terms of accuracy (89.7% compared to 88.5%), recall (99.7% compared to 90.8%), and F1 score (0.924 compared to 0.908). This suggests that even though the test loss was slightly higher, transfer learning improved the model's ability to correctly classify instances and minimize false negatives, thereby achieving a better trade-off between precision and recall.

To further investigate the impact of hyperparameter tuning on the performance of transfer learning, we conducted additional experiments with a reduced learning rate. Table 3 presents the results obtained when decreasing the learning rate by a factor of 0.1 compared to the original setting. The models trained with the reduced learning rate exhibit a slight decrease in performance across most architectures, with the test accuracy ranging from 0.2 to 1.1 percentage points lower than their counterparts trained with the original learning rate. However, it is important to note that the transfer learning models still consistently outperform the models trained from scratch, even with the reduced learning rate. For instance, the ResNet18 model achieves a test accuracy of 92.6% with the reduced learning rate, which is 3.0 percentage points higher than the corresponding scratch model. Similarly, the DenseNet161 and ResNeXt50 (32x4d) models maintain their superiority over their scratch counterparts, with test accuracies of 92.9% and 91.5%, respectively. These findings demonstrate that the effectiveness of real-world-feature transfer learning is robust against variations in hyperparameters such as the learning rate. While the optimal learning rate may depend on the specific characteristics of the dataset and the model architecture, the benefits of transfer learning in terms of improved performance and faster convergence remain evident across a range of hyperparameter settings.

Table 3. Performance comparison of from-scratch training models vs. real-world feature transfer learning models with reduced learning rate. Bold text indicates the best results for each experimental setting.

	Model	Test Loss	Test Accuracy	Precision	Recall	F1	G-Mean
ResNet18	Scratch Model	0.288	89.1%	89.5%	93.3%	0.914	0.894
	Transfer Learning	**0.183**	**92.6%**	**90.8%**	**98.2%**	**0.944**	**0.901**
ResNet50	Scratch Model	0.395	82.4%	**88.1%**	83.2%	0.856	**0.855**
	Transfer Learning	**0.311**	**88.7%**	85.3%	**99.5%**	**0.919**	0.823
DenseNet161	Scratch Model	0.315	90.7%	88.8%	97.4%	0.929	0.881
	Transfer Learning	**0.176**	**92.9%**	**91.2%**	**98.4%**	**0.947**	**0.906**
ShuffleNet v2	Scratch Model	0.488	77.5%	79.2%	86.6%	0.827	0.782
	Transfer Learning	**0.372**	**90.4%**	**90.5%**	**94.9%**	**0.926**	**0.903**
MobileNet v2	Scratch Model	0.453	79.3%	**87.9%**	77.8%	0.825	0.842
	Transfer Learning	**0.261**	**90.1%**	87.6%	**98.4%**	**0.927**	**0.861**
ResNeXt50 (32x4d)	Scratch Model	0.441	80.4%	88.2%	79.3%	0.835	0.852
	Transfer Learning	**0.206**	**91.5%**	**90.3%**	**96.9%**	**0.935**	**0.901**
Wide ResNet50-2	Scratch Model	**0.305**	87.9%	**90.4%**	90.2%	0.903	**0.881**
	Transfer Learning	0.322	**89.0%**	85.6%	**99.5%**	**0.920**	0.830

Overall, these results demonstrate the clear advantages of applying transfer learning for medical image analysis, even when the source and target data domains differ greatly. Not only did transfer learning models achieve rapid convergence during training, they delivered superior performance in comparison to their from-scratch counterparts across multiple performance metrics, thereby solidifying the merit of our approach.

6.3. Comparative Analysis with InceptionNet

In order to provide a more comprehensive evaluation of the effectiveness of transfer learning, we extended our experiments to include InceptionNet, a well-established and widely used deep learning architecture in computer vision tasks. InceptionNet, introduced

by Szegedy et al. [46], is known for its deep and wide architecture, which allows for efficient computation and improved performance compared to traditional CNN architectures.

The InceptionNet architecture is characterized by its use of inception modules, which consist of multiple convolutional layers with different kernel sizes operating in parallel. This design enables the network to capture features at various scales and resolutions, enhancing its ability to represent complex patterns in the input data. The inception modules are stacked together to form a deep network, with additional pooling layers and fully connected layers added for classification purposes.

InceptionNet has been widely adopted in various computer vision tasks, including image classification, object detection, and semantic segmentation. Its success can be attributed to its ability to learn rich and discriminative feature representations from large-scale datasets such as ImageNet. As a result, InceptionNet has become a popular choice for transfer learning, where pretrained models are fine-tuned on target tasks to leverage the knowledge gained from the source domain.

To evaluate the performance of InceptionNet in the context of real-world feature transfer learning for medical image analysis, we conducted experiments using the InceptionNet-v3 variant, which has 48 layers and has been pretrained on the ImageNet dataset. We followed the same experimental setup as described in Section 5.4, where we fine-tuned the pretrained InceptionNet model on the Pneumonia X-ray dataset and compared its performance to a model trained from scratch.

Table 4 presents the results of our comparative analysis between the InceptionNet model trained from scratch and the one utilizing transfer learning. The transfer learning model achieves a lower test loss of 0.187 compared to 0.315 for the scratch model, indicating better generalization and reduced overfitting. In terms of test accuracy, the transfer learning model reaches 93.3%, outperforming the scratch model by 2.1 percentage points.

Table 4. Performance comparison of InceptionNet trained from scratch vs. real-world feature transfer learning.

Model	Test Loss	Test Accuracy	Precision	Recall	F1	G-Mean
Scratch Model	0.315	91.2%	89.7%	97.2%	0.933	0.892
Transfer Learning	**0.187**	**93.3%**	**91.4%**	**98.5%**	**0.948**	**0.907**

Furthermore, the transfer learning model exhibits improvements in precision, recall, F1 score, and G-mean compared to the scratch model. The precision increases from 89.7% to 91.4%, indicating a reduction in false positive predictions. The recall improves from 97.2% to 98.5%, suggesting that the transfer learning model is better at identifying positive instances. The F1 score, which balances precision and recall, increases from 0.933 to 0.948, demonstrating the overall superiority of the transfer learning approach. Finally, the G-mean, which measures the model's ability to handle class imbalance, improves from 0.892 to 0.907, indicating better performance on both the majority and minority classes.

These results further validate the effectiveness of real-world feature transfer learning even when applied to a large and complex architecture such as InceptionNet. The pretrained InceptionNet model, which has learned rich feature representations from the ImageNet dataset, can successfully transfer its knowledge to the medical image domain, leading to improved performance compared to training from scratch. This finding reinforces the potential of transfer learning to accelerate the development of accurate and robust models for medical image analysis tasks by leveraging the knowledge gained from general image datasets.

7. Discussion

The results presented in this study provide compelling evidence for the effectiveness of real-world- feature transfer learning in the domain of medical image analysis. By leveraging pretrained models from general image datasets such as ImageNet, we have

demonstrated that it is possible to significantly improve the performance of deep learning models on medical image classification tasks even when the source and target domains differ substantially in terms of image characteristics and features.

One of the key findings of our study is the rapid convergence and high accuracy achieved by the transfer learning models during the training phase. The majority of the models were able to surpass 80% accuracy within a single epoch, indicating that the features learned from real-world images are indeed transferable and applicable to the medical image domain. This rapid convergence can be attributed to the rich and diverse set of features captured by the pretrained models, which have been exposed to a wide range of visual patterns and concepts during their initial training on large-scale datasets such as ImageNet.

The comparative analysis between transfer learning models and models trained from scratch further reinforces the superiority of the transfer learning approach. Across all evaluated models, transfer learning consistently achieved lower test loss and higher test accuracy, precision, recall, and F1 scores compared to their from-scratch counterparts. This indicates that the transferred features not only accelerate the learning process but also contribute to improved generalization and robustness of the models.

It is worth noting that the performance gains observed in our study are not limited to a specific model architecture or dataset. The transfer learning approach proved effective across a range of popular CNN architectures, including ResNet, DenseNet, ShuffleNet, and MobileNet, demonstrating its versatility and applicability to various network designs. Moreover, the successful transfer of features from ImageNet to the Pneumonia X-ray dataset suggests that the proposed methodology can be extended to other medical image analysis tasks and datasets, potentially benefiting a wide range of clinical applications.

The inclusion of computationally efficient models such as ShuffleNet and MobileNet in our study highlights the potential of real-world feature transfer learning to enable the deployment of deep learning models on resource-constrained devices. While these models may exhibit slower convergence compared to their more complex counterparts, the transfer learning approach still yields significant performance improvements over training from scratch. This finding opens up possibilities for the development of lightweight and efficient models that can be deployed on mobile or embedded devices, expanding the reach and accessibility of medical image analysis tools.

Despite the promising results, it is important to acknowledge the limitations of our study and identify areas for future research. One potential limitation is the reliance on a single medical image dataset (Pneumonia X-ray) for evaluation. While this dataset serves as a representative example of the challenges faced in medical image analysis, further validation on a broader range of medical image modalities and clinical tasks would strengthen the generalizability of our findings. Future studies could explore the application of real-world feature transfer learning to other medical imaging domains, such as CT scans, MRIs, and histopathology images, to assess its effectiveness across different data characteristics and diagnostic goals.

Another avenue for future research is the investigation of more advanced transfer learning techniques and strategies. In our study, we employed a straightforward approach of fine-tuning the entire pretrained model on the target task. However, there are various other transfer learning strategies, such as feature extraction, partial fine-tuning, and domain adaptation, that could potentially further improve the performance and efficiency of the models. Exploring these techniques and their impact on the transferability of features from the real-world to the medical image domain could provide valuable insights and guide the development of more sophisticated transfer learning pipelines.

It is important to note that the dataset used in this study, which consists of chest X-ray images labeled as “Pneumonia” or “Normal”, may not always represent a definitive diagnosis of pneumonia. Lung opacity, a common radiological finding in pneumonia, can be caused by other conditions such as inflammation, tuberculosis, and other respiratory pathologies. While the presence of lung opacity is often associated with pneumonia, it is not a specific indicator of the disease. The dataset used in this study is more accurately

described as a collection of images exhibiting lung opacity, which may or may not be caused by pneumonia. To establish a definitive diagnosis of pneumonia, additional clinical information such as patient history, physical examination findings, and laboratory test results would be required. Future research could focus on curating datasets with confirmed pneumonia cases and exploring the use of transfer learning for differentiating between various causes of lung opacity. Despite this limitation, the dataset serves as a valuable resource for evaluating the effectiveness of transfer learning in identifying radiological abnormalities, and showcases the potential of this approach for assisting in the initial screening and triage of patients with suspected pneumonia.

8. Conclusions

This study investigated the application of transfer learning from general image datasets to the specialized domain of medical imaging, specifically focusing on chest X-rays. The research was motivated by the common belief that the significant differences in features between general and medical images could hinder the effective use of transfer learning. Through comprehensive experiments and analysis, we have provided substantial evidence to challenge this assumption and demonstrate the viability of transfer learning for medical image data.

Our experimental results consistently showed that using pretrained models, such as those trained on ImageNet, significantly accelerates the training process for medical image analysis tasks. Furthermore, the simple preprocessing step of converting grayscale medical images to RGB format proved to be effective in facilitating the transfer of learned features.

Among the evaluated architectures, most models achieved rapid convergence and high accuracy within the initial training epochs. In particular, ResNeXt50 and Wide ResNet50 demonstrated exceptional performance, reaching over 95% training accuracy in the early stages. These findings highlight the importance of network depth and design principles in enabling effective feature propagation during transfer learning.

The comparative analysis between transfer learning models and models trained from scratch further emphasized the superiority of transfer learning. Across multiple evaluation metrics, including test loss, accuracy, precision, recall, and F1 score, the transfer learning models consistently outperformed their counterparts. This underscores the consistent effectiveness of transfer learning even when the source and target domains differ significantly.

One notable observation was the performance of ShuffleNet, a computationally efficient model. Despite initially lagging behind the other models in accuracy, ShuffleNet showed a substantial improvement in test accuracy when transfer learning was applied. This demonstrates the versatility of our approach and suggests that even models designed for specific constraints can benefit from transfer learning.

The implications of our work extend beyond the immediate experimental results. By demonstrating the feasibility of leveraging pretrained models from general image datasets for medical imaging, we have provided a means to accelerate the development of deep learning applications in the medical field. The accessibility of pretrained models and the effectiveness of transfer learning could facilitate rapid advancements in medical image analysis, potentially leading to more accurate and timely diagnoses.

However, it is important to acknowledge the limitations of this study and the opportunities for future research. The experiments were conducted using only the Pneumonia X-ray dataset; further validation on other medical image datasets would enhance the generalizability of our findings. Additionally, investigating the effects of different transfer learning strategies, such as fine-tuning specific layers or employing alternative initializations, could provide deeper insights into the mechanisms that enable successful knowledge transfer.

In summary, this paper contributes to the field of deep learning by empirically validating the effectiveness of transfer learning from general image data to medical image data. Our findings challenge prevailing assumptions and open up new possibilities for applying transfer learning in specialized domains. This research highlights the adaptability

and power of deep learning and encourages further exploration and adoption of transfer learning in various fields, particularly in domains with limited or highly specialized data.

Author Contributions: Conceptualization, C.G. and M.L.; formal analysis, C.G.; investigation, C.G.; writing—original draft preparation, C.G. and M.L.; writing—review and editing, C.G. and M.L.; visualization, M.L.; supervision, M.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Chung-Ang University Research Scholarship Grants in 2024.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study.

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

References

1. Shen, D.; Wu, G.; Suk, H.I. Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.* **2017**, *19*, 221–248. [CrossRef] [PubMed]
2. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; Van Der Laak, J.A.; Van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [CrossRef] [PubMed]
3. Lee, M. Recent Advancements in Deep Learning Using Whole Slide Imaging for Cancer Prognosis. *Bioengineering* **2023**, *10*, 897. [CrossRef] [PubMed]
4. Kaur, A.; Singh, Y.; Neeru, N.; Kaur, L.; Singh, A. A survey on deep learning approaches to medical images and a systematic look up into real-time object detection. *Arch. Comput. Methods Eng.* **2022**, *29*, 2071–2111. [CrossRef]
5. Ge, Y.; Zhang, Q.; Sun, Y.; Shen, Y.; Wang, X. Grayscale medical image segmentation method based on 2D&3D object detection with deep learning. *BMC Med. Imaging* **2022**, *22*, 33.
6. Erdaş, Ç.B.; Sümer, E. A fully automated approach involving neuroimaging and deep learning for Parkinson's disease detection and severity prediction. *PeerJ Comput. Sci.* **2023**, *9*, e1485. [CrossRef] [PubMed]
7. Kim, M.; Yun, J.; Cho, Y.; Shin, K.; Jang, R.; Bae, H.j.; Kim, N. Deep learning in medical imaging. *Neurospine* **2019**, *16*, 657. [CrossRef] [PubMed]
8. Tang, W.; Zhang, M.; Xu, C.; Shao, Y.; Tang, J.; Gong, S.; Dong, H.; Sheng, M. Diagnostic efficiency of multi-modal MRI based deep learning with Sobel operator in differentiating benign and malignant breast mass lesions—a retrospective study. *PeerJ Comput. Sci.* **2023**, *9*, e1460. [CrossRef] [PubMed]
9. Kim, H.E.; Cosa-Linan, A.; Santhanam, N.; Jannesari, M.; Maros, M.E.; Ganslandt, T. Transfer learning for medical image classification: A literature review. *BMC Med. Imaging* **2022**, *22*, 69. [CrossRef]
10. Taşyürek, M.; Öztürk, C. A fine-tuned YOLOv5 deep learning approach for real-time house number detection. *PeerJ Comput. Sci.* **2023**, *9*, e1453. [CrossRef] [PubMed]
11. Yu, X.; Wang, J.; Hong, Q.Q.; Teku, R.; Wang, S.H.; Zhang, Y.D. Transfer learning for medical images analyses: A survey. *Neurocomputing* **2022**, *489*, 230–254. [CrossRef]
12. Shamshad, F.; Khan, S.; Zamir, S.W.; Khan, M.H.; Hayat, M.; Khan, F.S.; Fu, H. Transformers in medical imaging: A survey. *Med. Image Anal.* **2023**, *88*, 102802. [CrossRef] [PubMed]
13. Malhotra, P.; Gupta, S.; Koundal, D.; Zaguia, A.; Enbeyle, W. Deep neural networks for medical image segmentation. *J. Healthc. Eng.* **2022**, *2022*. [CrossRef] [PubMed]
14. Chen, Y.; Yang, X.H.; Wei, Z.; Heidari, A.A.; Zheng, N.; Li, Z.; Chen, H.; Hu, H.; Zhou, Q.; Guan, Q. Generative adversarial networks in medical image augmentation: A review. *Comput. Biol. Med.* **2022**, *144*, 105382. [CrossRef] [PubMed]
15. Rudan, I.; Boschi-Pinto, C.; Biloglav, Z.; Mulholland, K.; Campbell, H. Epidemiology and etiology of childhood pneumonia. *Bull. World Health Organ.* **2008**, *86*, 408–416B. [CrossRef]
16. Rajpurkar, P.; Irvin, J.; Zhu, K.; Yang, B.; Mehta, H.; Duan, T.; Ding, D.; Bagul, A.; Langlotz, C.; Shpanskaya, K.; et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv* **2017**, arXiv:1711.05225.
17. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity mappings in deep residual networks. In Proceedings of the Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Proceedings, Part IV 14; 2016; pp. 630–645.
18. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
19. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2009**, *22*, 1345–1359. [CrossRef]
20. Tan, C.; Sun, F.; Kong, T.; Zhang, W.; Yang, C.; Liu, C. A survey on deep transfer learning. In Proceedings of the Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, 4–7 October 2018; Proceedings, Part III 27; 2018; pp. 270–279.

21. Yosinski, J.; Clune, J.; Bengio, Y.; Lipson, H. How transferable are features in deep neural networks? *Adv. Neural Inf. Process. Syst.* **2014**, *27*.
22. Sharif Razavian, A.; Azizpour, H.; Sullivan, J.; Carlsson, S. CNN features off-the-shelf: An astounding baseline for recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, 23–28 June 2014; pp. 806–813.
23. Kornblith, S.; Shlens, J.; Le, Q.V. Do better imagenet models transfer better? In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2661–2671.
24. Alzubaidi, L.; Al-Amidie, M.; Al-Asadi, A.; Humaidi, A.J.; Al-Shamma, O.; Fadhel, M.A.; Zhang, J.; Santamaría, J.; Duan, Y. Novel transfer learning approach for medical imaging with limited labeled data. *Cancers* **2021**, *13*, 1590. [CrossRef] [PubMed]
25. Rahman, T.; Chowdhury, M.E.; Khandakar, A.; Islam, K.R.; Islam, K.F.; Mahbub, Z.B.; Kadir, M.A.; Kashem, S. Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray. *Appl. Sci.* **2020**, *10*, 3233. [CrossRef]
26. Chouhan, V.; Singh, S.K.; Khamparia, A.; Gupta, D.; Tiwari, P.; Moreira, C.; Damaševičius, R.; De Albuquerque, V.H.C. A novel transfer learning based approach for pneumonia detection in chest X-ray images. *Appl. Sci.* **2020**, *10*, 559. [CrossRef]
27. Raghu, M.; Zhang, C.; Kleinberg, J.; Bengio, S. Transfusion: Understanding transfer learning for medical imaging. *Adv. Neural Inf. Process. Syst.* **2019**, *32*.
28. Anwar, S.M.; Majid, M.; Qayyum, A.; Awais, M.; Alnowami, M.; Khan, M.K. Medical image analysis using convolutional neural networks: A review. *J. Med. Syst.* **2018**, *42*, 1–13. [CrossRef] [PubMed]
29. Yamashita, R.; Nishio, M.; Do, R.K.G.; Togashi, K. Convolutional neural networks: An overview and application in radiology. *Insights Imaging* **2018**, *9*, 611–629. [CrossRef] [PubMed]
30. Khan, A.; Sohail, A.; Zahoora, U.; Qureshi, A.S. A survey of the recent architectures of deep convolutional neural networks. *Artif. Intell. Rev.* **2020**, *53*, 5455–5516. [CrossRef]
31. Ayan, E.; Ünver, H.M. Diagnosis of pneumonia from chest X-ray images using deep learning. In Proceedings of the 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT), Istanbul, Turkey, 24–26 April 2019; pp. 1–5.
32. Yeom, T.; Gu, C.; Lee, M. DuDGAN: Improving class-conditional GANs via dual-diffusion. *IEEE Access* **2024**. [CrossRef]
33. Ko, K.; Lee, M. ZIGNeRF: Zero-shot 3D Scene Representation with Invertible Generative Neural Radiance Fields. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 4–8 January 2024; pp. 4986–4995.
34. Ywet, N.L.; Maw, A.A.; Nguyen, T.A.; Lee, J.W. YOLOTransfer-DT: An Operational Digital Twin Framework with Deep and Transfer Learning for Collision Detection and Situation Awareness in Urban Aerial Mobility. *Aerospace* **2024**, *11*, 179. [CrossRef]
35. Kim, S.; Nam, B.H.; Jung, Y.H. Comparison of Deep Transfer Learning Models for the Quantification of Photoelastic Images. *Appl. Sci.* **2024**, *14*, 758. [CrossRef]
36. Huber, F.; Inderka, A.; Steinhage, V. Leveraging Remote Sensing Data for Yield Prediction with Deep Transfer Learning. *Sensors* **2024**, *24*, 770. [CrossRef] [PubMed]
37. Mohammadi, S.; Belgiu, M.; Stein, A. Few-Shot Learning for Crop Mapping from Satellite Image Time Series. *Remote Sens.* **2024**, *16*, 1026. [CrossRef]
38. Nikezić, D.P.; Radivojević, D.S.; Lazović, I.M.; Mirkov, N.S.; Marković, Z.J. Transfer Learning with ResNet3D-101 for Global Prediction of High Aerosol Concentrations. *Mathematics* **2024**, *12*, 826. [CrossRef]
39. Lee, S.; Lee, M. MetaSwin: A unified meta vision transformer model for medical image segmentation. *PeerJ Comput. Sci.* **2024**, *10*, e1762. [CrossRef] [PubMed]
40. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Adv. Neural Inf. Process. Syst.* **2017**, *30*.
41. Ma, N.; Zhang, X.; Zheng, H.T.; Sun, J. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 116–131.
42. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
43. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1492–1500.
44. Zagoruyko, S.; Komodakis, N. Wide Residual Networks. In Proceedings of the British Machine Vision Conference 2016, British Machine Vision Association, York, UK, 19–22 September 2016.
45. Mooney, P. Chest X-ray Images (Pneumonia). 2021. Available online: https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia (accessed on 20 June 2023).
46. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.