Optimizing Pediatric Pneumonia Diagnosis with Transfer Learning

Machine Learning II - Group 22

Hugo Beltman
Student number: s2385937

Daan Schram
Student number: s2692759

Pascal Vriend
Student number: s2739046

Wester Bottema
Student number: s2246740

Stefan De Laat
Student number: s2465086

January 31, 2025

1 Abstract

Deep learning has gained significant attention in the field of medical diagnostics, particularly in imaging-based applications. However, pediatric diagnostics using convolutional neural networks (CNNs) is challenged by the scarcity of pediatric datasets. This study investigates whether transfer learning can optimize the performance of pediatric pneumonia diagnosis by leveraging a CNN model pre-trained on adult chest X-ray data.

A comparative analysis was conducted using three approaches: training a CNN only on pediatric data, applying a pre-trained adult model without any further training, and employing transfer learning with pediatric data to refine the pre-trained model.

The results show that transfer learning improves the accuracy of pneumonia classification from 80.75% to 83.50%, while using the adult-trained model only provides an accuracy of 67.7%, highlighting the efficacy of adapting adult-trained models for pediatric applications.

This research underscores the potential of transfer learning to mitigate data limitations in pediatric imaging and suggests future directions for improving CNN-based medical diagnostics.

AI Usage Disclaimer

This report was written with the help of AI as a writing assistant (such as ChatGPT and Overleaf's Writefull AI). It was used solely to check spelling and grammar errors and improve structure. It was not used to find facts or generate content beyond language refinement. The work remains our own, and we are fully responsible for any mistakes, omissions, or factual inaccuracies in the text.

2 Introduction

Deep learning has emerged as a promising field, allowing researchers to better understand complex physical and biological phenomena [1]. In recent years, its application in the medical domain has become increasingly relevant, with many diagnostic tools being developed based on deep learning models [2]. While deep learning has experienced rapid growth in the context of adult medical imaging, its application in pediatric imaging remains limited. Although several studies have proposed the use of convolutional neural networks (CNNs) for pediatric imaging [3, 4, 5] the scarcity of pediatric data presents a significant challenge. This lack of sufficient data hinders the ability to effectively train neural networks, limiting the broader adoption of deep learning techniques in this area.

Furthermore, while multiple models have been proposed for pneumonia diagnosis in both adult and pediatric populations [6, 7, 8], a common issue is the use of the same dataset for both training and validation. This practice can lead to overfitting and inflated performance metrics, as the models are evaluated on data they have already seen during training. Although these models often demonstrate promising results, their true efficacy is limited by the lack of external validation on independent datasets.

The limitation posed by the scarcity of pediatric data raises the question of whether CNNs trained on adult images can be effectively applied to pediatric populations. Since model validation typically requires a smaller dataset compared to the large amounts of data needed for training, it may be more efficient to use the limited available pediatric images to validate well-established networks originally developed for adult populations. This approach could shift the focus towards validating adult-trained models for pediatric diagnoses, rather than training new CNNs from scratch with the limited pediatric data. By doing so, the applicability and performance of these models in pediatric imaging can potentially be enhanced, particularly for tasks such as pneumonia diagnosis.

The results of an earlier study [9] already showed that adult-trained CNNs can be effectively applied to the pediatric population. This network was trained with adult-data only, and achieved a pneumonia discrimination AUC (Area Under the Curve) of 0.95 in the adult data set. When tested in a pediatric dataset, the AUC of pneumonia discrimination reached 0.82.

While this demonstrates that adult-trained CNNs can be applied to pediatric data, there is still significant potential for improving the AUC of pneumonia discrimination. Thus, the goal of this research is to improve the performance of pneumonia detection in the pediatric population, through transfer learning with models trained on the adult population to pediatric data. This will be done by training a CNN on adult data. This is a form of Deep Transfer learning and results in the research question below:

Research question: Can we optimize the performance of pediatric pneumonia diagnoses using a CNN by applying transfer learning with a small set of pediatric x-rays on an adult-trained model?

To answer this research question, this paper will first review related work and establish the methodology. Subsequently, the results will be presented and discussed, leading to the final conclusion.

3 Related Work

3.1 Type of Models for Transfer Learning

The concept of transfer learning is derived from the generalization theory of transfer. It was proposed by psychologist C.H Judd, and it states that learning to transfer is the result of generalization of experience. Regarding machines, transfer learning embodies the process of the machines ability to transfer their knowledge of a domain it is trained on (the source domain) to another domain were that is scarce [10] is being trained.
The definition of transfer learning can be stated as: "*Given some source domains D_{s_i} that contain a number of observations o^s, and the corresponding learning tasks τ_{s_i} (that is, $\{(D_{s_i}, \tau_{s_i})|i = 1, 2, \ldots, o^s\}$). Towards certain observations, o^t, and learning tasks τ_t in the target domain D_t, the TL method aims to improve the learned predictive function $f_t(*)$ in the domain D_t based on the knowledge it learned from the source domain D_{s_i} and the selected learning task τ_{s_i}, where $D_{s_i} \neq D_t$ or $\tau_{s_i} \neq \tau_t$.*"[11].

Categorization of transfer learning: Regarding the discrepancy between domain, transfer learning can be roughly divided in to two classes: homogeneous and heterogeneous transfer learning. They highlight the difference between the source and their target domain in terms of feature space and data distribution. Homologous transfer learning applies in situations where the domains are of the same feature space. It is often used to avoid time-consuming and other cost of recollecting new data when big data is available. Heterogeneous transfer learning refers to transfer learning process where the domains have different feature spaces [12]. Another categorization method is based on the data and tasks of the transfer learning, and can be split in to three sections: *transductive transfer learning*, *inductive transfer learning*, and *unsupervised transfer learning*.

1. Transductive transfer learning: is applicable when the labeled data are available only in the source domain. In this situation, the difference lies in the domains of the source and the target but they relate to each other to provide similar outer knowledge supporting the learning. The learning process consists of the pre-trained model acquire knowledge from the source domain D_s and the source task τ_s. The available knowledge from the source and task domain is then used to improve the learning process of the target predictive function [11].

2. Inductive learning: this method enables the model to transfer knowledge within the same domain. In general, data within the source domain can be labeled as well as unlabeled, while date from the target domain often is labeled, to induce an objective predictive model. Inductive learning can be subdivided in to two more classes based on the data situations in the source domain: *multi-task learning* and *self-taught learning*. The first sub-category happens when abundant labeled data is available in the source domain, but only a small amount of labeled data exists within the target-domain. Self-taught learning is applied in situation where only labeled data are available in the source domain. Inductive transfer works by using the knowledge acquired from the source domain ans task to enhance the learning of the predictive function [11].

3. Unsupervised transfer learning: This method is used when no labeled data is available in either the source or target domain. Two key characteristics often associated with unsupervised transfer learning are the pre-training and fine-tuning of a model.

3.2 Deep Transfer learning

Deep learning has shown it self as the solution for many problems in the field of machine learning in the past decades. However, the use has come with two significant constraints: the dependency on labeled data and the high training costs. Deep transfer learning addresses these constraints by reducing reliance on large labeled datasets and lowering costs. It achieves this by reusing knowledge obtained from a source domain or task to assist in training on a target domain or task. Despite the restriction its use has gain popularity in several fields like image processing, natural language processing, and numerical data [13]. To asses the quality of the transfer of the model, each models classification capabilities are assessed by evaluating the classification error rate of the model. In the past decades, several methods have been developed in order to conduct this specific evaluation. A widely used method is k-fold cross-validation.

K-fold cross-validation is not a inherent part of deep transfer learning. It is a technique used for model selection, ensuring a robust validation process.It achieves this by estimating the classification error, thereby assessing the model's overall accuracy. In K-fold cross-validation, the data is structured into a file containing k equally sized segments, each with the same dimensions. Next, the first segment n/k will be set aside and the remaining (n-n/k)[14] are used for the parametrization of the model. After the parametrization has been conducted, the remaining segments will be tested against the first segment for the rates of classification error. Subsequently, from the file a second segment will be drawn. This will undergo the same process to asses the rates of classification error. This process will repeat it self till all the segments from the file are used for the evaluation of the classification error. With these results, the likelihood that the models results are generated by chance can be proven/disproven [15].

4 Methodology

In order to answer the main research question, this study adhered to a methodology based on studied literature. This section contains a detailed description of the data, the machine learning techniques used, and the evaluation principles to draw conclusions.

4.1 Data Selection

Both adult and pediatric x-ray datasets were essential in answering the research questions. These datasets were chosen and filtered carefully, ensuring diversity in the demographics of patients as well as a balance between pneumonia and healthy x-rays.

The **adult dataset** was created from the NIH Chest X-ray dataset [16] [17]. This dataset consists of around 100000 chest X-rays from 30000 patients aged between 1 and 84. The labels are text-mined by using natural language processing techniques on the reports, but are expected to be over 90% accurate. Next to pneumonia and no finding chest x-rays, this dataset contains 13 other classes. The dataset contains a CSV file with labels as well as file names. It was investigated for 1323 chest x-rays, the patient was an adult and had pneumonia. The names of these images were extracted, as well as the filenames of 1323 x-rays labeled as no finding. The entire dataset was too large to download and filter on a laptop, so this was done on the universities EEMCS Hadoop Cluster. The labels were uploaded to this server cluster, and the dataset was downloaded on it. From there, the correct pneumonia images and no finding images could be retrieved. These were zipped and send back to a laptop. All images have a size of 1024 x 1024. Out of the 2646 images, 75% was used for training and 25% was used for testing.

The **pediatric dataset** utilized in this research is Chest X-Ray Images (Pneumonia) [18]. This dataset comprises 5863 chest X-rays, divided into two categories (Pneumonia/Normal). They were selected from retrospective cohorts of patients between one and five years old. They were gathered at the Guangzhou Women and Children's Medical Center, Guangzhou. This dataset was already categorized into training, validation and testing sets. Among all of the images, 4273 x-rays contained pneumonia, and 1583 x-rays showed no sign of pneumonia. Since this study is about transfer learning, with the idea that the amount of pediatric data is limited, only 100 images of this set were used for training and 300 for testing.

A summary of the datasets used in this study, alongside the amount of images retrieved from it, is displayed in Table 1.

Type	**Dataset**	**Filtered Dataset size**	**Train Size**	**Test Size**
Adult	NIH Chest X-Ray	2646	1984	662
Pediatric	Chest X-Ray Images (Pneumonia)	5863	100	300

Table 1: Overview of Adult and Pediatric Chest X-Ray Datasets

4.2 Data Preprocessing

The preprocessing of chest X-ray images is a critical step to ensure the input data is consistent and optimized for training the machine learning model. Combining the steps outlined in [19] and [20], our approach integrates image resizing, adaptive masking, contrast enhancement, noise reduction, and format transformation. Below, we describe each step and its importants.

4.2.1 Image Resizing

The original chest X-ray images in the dataset were resized to 224×224 pixels. This step ensures uniform input dimensions, which are important for compatibility with CNN's.

4.2.2 Adaptive Masking

To remove the influence of the diaphragm region and other artifacts such as text, which contains high-intensity pixels, adaptive masking was applied. This technique, adapted from [19], involved identifying and eliminating the top 10% of the brightest pixels in the image. The threshold method $\min + 0.9 \cdot (\max - \min)$ was used, effectively isolating unimportant regions, while increasing the contrast of meaningful lung features, by using an wider range of pixel values for these features.

4.2.3 Histogram Equalization

After masking, histogram equalization was applied to enhance image contrast. This step redistributes pixel intensity values, making lung patterns and tissue characteristics more detailed.

4.2.4 Noise Reduction Using Gaussian Blur

Gaussian blurring with a kernel size of 3×3 and a sigma of 0 was performed to reduce noise while preserving essential edges. This technique minimizes the influence of minor variations and artifacts that might confuse the CNN. [20] highlights that Gaussian blur is particularly effective in conjunction with histogram equalization, balancing noise reduction and feature preservation.

4.2.5 Conversion to 3-Channel Format

The preprocessed grayscale images were stacked into three identical channels to form pseudo-RGB images. This conversion aligns with the input requirements of pre-trained CNNs, which expect 3-channel RGB images.

4.3 Model Training

To evaluate the performance of transfer learning in the context of this study, several models were trained. For all models, one custom CNN architecture was used. The custom architecture can be seen in Figure 1. It consists of three convolutional layers combined with ReLU activation and max-pooling layers. These extract hierarchal features from the input data. After, a sequence of three fully connected layers is used to classify whether an x-ray contains pneumonia or not. The following sections will describe the several models that were trained.

4.3.1 On Adult Data

Several models were trained using adult chest x-ray data. In several iterations, different models, hyperparameters, preprocessing techniques, and optimization techniques were tested. Building on these models, the final adult model used the correct preprocessed images like described in section 4.2. In addition, optimization techniques were used in this model to improve performance. Weight decay was utilized to prevent overfitting by applying a regularization term to the loss function. In addition, cosine annealing learning rate scheduling was used to dynamically adjust the learning rate, allowing for finer convergence at the end of the training process. Because of the large input size of the adult dataset, the UTwente Jupyter GPU servers were used to train these models. In the end, the model state was saved to a file to be further used by pediatric models.

4.3.2 On Pediatric Data

In order to compare the performance of the transfer step, two models were trained as a baseline. The first model was trained using a new empty model. It was trained using only pediatric data. Next to this, a model was trained utilizing the pre-trained adult dataset but without any additional training.

To explore transfer learning in this context, the adult pre-trained model was used as basis for pediatric data to be trained on. Similar optimization techniques to those used in adult models were retained, however only the fully connected layers were trainable. By freezing the initial convolutional layers, the features learned from adult chest x-rays were kept. To improve the confidence in performance evaluation,

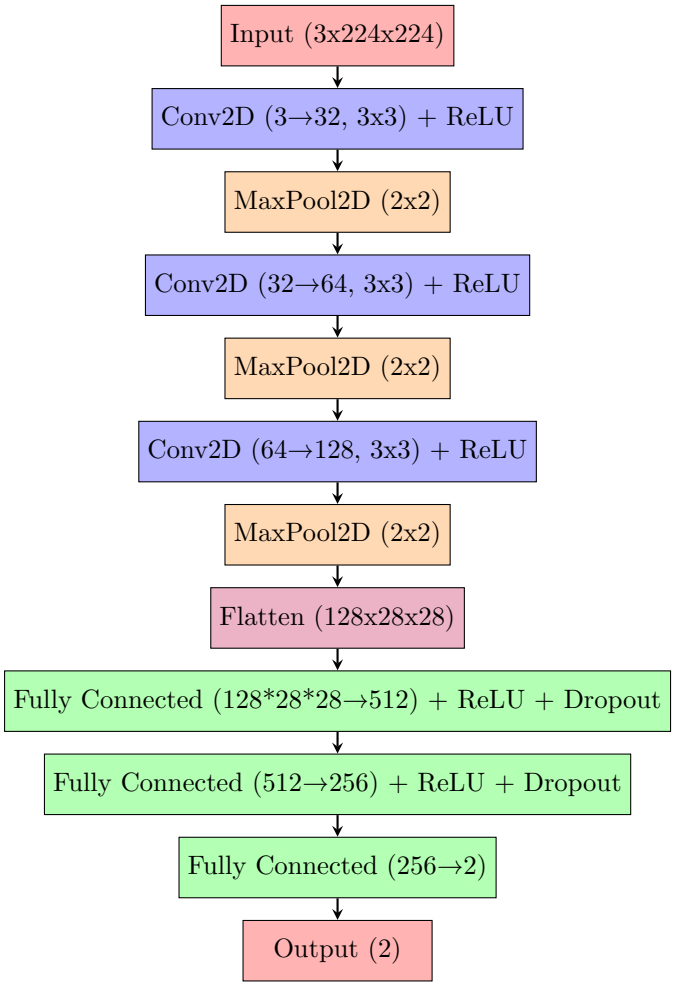

Figure 1: Custom CNN Model Architecture

k-fold cross validation was applied to this model. This ensured a more robust assessment of the transfer learning step.

4.3.3 On Mixed Data

In order to leverage both adult and pediatric data, a mixed data approach was implemented as well. A dynamic sampler was utilized to graduately increase the weight of pediatric data on the training process. This made sure that the model could progressively adapt to the pediatric data, while benefiting from knowledge gained from adult x-rays as well.

4.4 Evaluation

The performance of the model in this research was evaluated using standard classification metrics. These include:

- **Accuracy**: Proportion of correctly classified images
- **Precision**: Proportion of correctly identified positive cases among all predicted positive cases
- **Precision**: Proportion of actual positive cases correctly identified
- **F1-score**: Harmonic mean of precision and recall

5 Results

5.1 Adult Model

The best performing adult model was named **Adult Model V5**. It achieved an accuracy of 71.7%, with a precision of 71.8%, recall of 71.7%, and an F1-score of 71.6%. These are visible in Table 2.

Model	Accuracy	Precision	Recall	F1-Score
Adult Model V5	71.7%	71.8%	71.7%	71.6%

Table 2: Evaluation Metrics for the Adult Model

5.2 Pediatric Models

The first model used the adult pre-trained model and applied it directly onto the pediatric data. This model was called **Pediatric Model Without Transfer**. All metrics stayed at 50%, so the model was randomly guessing. Therefore it is not included in Table 3.

The second model trained on pediatric data was without a pre-trained model, training purely on the pediatric data available. This model was called **Pediatric Model Without Initialization K Fold**. It used k-fold cross-validation, using 4 folds. For each iteration, 3 were used for testing and 1 for training. This means each fold trained a model on 100 images, and tested it against 300 images. Therefore, a 95% confidence interval could be established. The results can be found in the row named 'Empty + K-Fold' in Table 3.

The last model was trained using a transfer learning approach. It used the adult pre-trained model, and continued training the fully connected layers of the model on pediatric training data. This model was called **Pediatric Model K Fold**. It also used k-fold cross-validation, with the same setup as described in the previous paragraph. The results, including 95% confidence intervals, can be found in the row named 'Transfer + K-Fold' in Table 3.

Model	Accuracy	Precision	Recall	F1-Score
Empty + K-Fold	80.75 [78.40, 83.10]	81.64 [79.12, 84.15]	80.75 [78.40, 83.10]	80.62 [78.24, 82.99]
Transfer + K-Fold	83.50 [81.51, 85.49]	83.55 [81.50, 85.60]	83.50 [81.51, 85.49]	83.49 [81.51, 85.48]

Table 3: Evaluation Metrics for Pediatric Models including 95% Confidence Intervals

5.3 Mixed Model

The **Mixed Model Adult Pediatric**, trained with dynamic sampling, resulted in an accuracy of 67.7%, with a precision of 74.9%, recall of 67.7%, and an F1-score of 65.1%, like seen in table Table 4.

Model	Accuracy	Precision	Recall	F1-Score
Mixed Model Adult Pediatric	67.7%	74.9%	67.7%	65.1%

Table 4: Evaluation Metrics for the Mixed Model

6 Discussion and Critical Reflection

The first goal of this project was to build an adult pneumonia classification model similar to the one described in Rollan-Martinez-Herrera et Al [9], which serves as the basis for our work. That study achieved a pneumonia classification accuracy of 0.95 on an adult dataset. However, our model never reached this level of performance. Several factors may explain this discrepancy, one of which is the size of our dataset. Our adult dataset contained a total of 2,646 X-ray images (1,984 for training and 662 for testing), whereas the dataset used in Rollan-Martinez-Herrera et Al [9] comprised 32,088 X-rays. A larger dataset generally improves model performance by providing more training examples. Additionally, Rollan-Martinez-Herrera et Al. [9] excluded x-ray images from their dataset containing two or more external devices (e.g. cables, leads, tubes, pacemakers) over the lung field. They also omitted images with extreme angular rotations, improper positioning, or incorrect pulmonary field capture (e.g., image truncation). In contrast, our dataset was not filtered for these artifacts, which may have led the model to learn features unrelated to pneumonia, such as the presence of external devices. Since the primary objective of this project was to explore transfer learning rather than to maximize classification accuracy, we stopped refining the model once it achieved a discrimination score of at least 0.70. At this threshold, the model demonstrated learning beyond random guessing, which was sufficient for our research.

Initially, ResNet and Xception were tested as base models. However, after training and testing with an equal number of pneumonia and normal X-ray images, both models achieved a discrimination score between 0.50 and 0.60—indicating random guessing rather than learning. This was largely due to the dataset size. Pre-trained models such as ResNet and Xception require large amounts of data to meaningfully adjust their weights. Since they are already trained on extensive datasets, our limited number of adult pneumonia X-rays had minimal impact on their learning, resulting in poor performance. To address this, a custom CNN model was developed, which successfully learned from the X-ray images and achieved a discrimination score of 0.71. These findings suggest that future research would benefit from a significantly larger dataset.

As shown in Section 5.2, the model 'Pediatric Model Without Initialization K Fold', trained solely on pediatric data (without a pretrained model). This model achieved an accuracy of 80.75%— 9% higher than the model trained on adult data. However, this result should be interpreted with caution, as the pediatric model was trained on only 400 X-ray images, a much smaller dataset compared to the adult model. This difference in dataset size likely influenced performance, as deep learning models generally require larger datasets to generalize effectively. The accuracy difference may also be influenced by variations in image quality between the pediatric and adult datasets. The pediatric dataset may have fewer imaging artifacts, better contrast, or more consistent labeling, but no formal analysis was conducted to confirm this. Additionally, differences in disease presentation between children and adults could contribute to the model's performance gap. Future research with larger, more balanced datasets and a detailed analysis of image quality would be needed to draw stronger conclusions.

Applying transfer learning from a model pre-trained on adult chest X-rays improved pediatric pneumonia classification accuracy from 80.75% to 83.50%. This gain suggests that the use of knowledge from a larger dataset enhances performance, even with limited pediatric data. The improvement can be caused by the retained feature representations in the convolutional layers, which capture general lung structures. Additional training of only the fully connected layers allowed the model to adapt to pediatric pneumonia without overfitting. Additionally, transfer learning provided a better starting point, reducing the amount of pediatric training data needed for stable performance.

The mixed data model performed 15.8% worse compared with the Pediatric Model K Fold. A reasons that the mixed model might have performed worse could be the use of a dynamic sampler. The dynamic sampler progressively increases the weight of pediatric data suggests an attempt to address potential class imbalance between the adult and pediatric datasets. However, if the adult data was still dominating the training process, the model might have struggled to learn the patterns specific to the pediatric data, leading to a lower performance when tested across both groups.

A potential enhancement to the models is the automatic tuning of hyperparameters to identify optimal values. This is a process that can improve a models performance while reducing the manual effort needed. Traditional approaches often use grid search or random search, but these methods can be computationally expensive and inefficient. Instead, a more efficient approach such as Mango or Optuna can be applied[21][22]. These frameworks are designed to efficiently explore and exploit the hyperparameter search space using techniques like Bayesian optimization. These methods aim to minimize computational cost by balancing exploration (trying new hyperparameter values) and exploitation (refining promising regions of the search space).

As mentioned in Section 4.1, the availability of pediatric chest X-ray images is limited. To address this, we aim to minimize reliance on these images by creating a model pre-trained on adult data as a baseline, followed by a transfer learning step. However, additional pediatric data may still be necessary, for example to ensure the learning during the transfer is sufficient. Image processing techniques during training can help mitigate this limitation by generating augmented data. By randomly rotating, scaling, or adjusting contrast, new variations of images can be created. This increases the size of the dataset and improves model robustness if needed.

7 Conclusion

This study explored the use of transfer learning to optimize pediatric pneumonia diagnosis using a CNN model pre-trained on adult chest X-ray data. The results demonstrated that a model trained exclusively on a small pediatric dataset performed reasonably well, achieving an accuracy of 80.75%. However, transferring the adult-trained model using pediatric images further improved performance to 83.50%, suggesting that transfer learning can enhance diagnostic accuracy when pediatric data is limited, thus providing a positive answer to the research question: *Can we optimize the performance of pediatric pneumonia diagnoses using a CNN by applying transfer learning with a small set of pediatric X-rays on an adult-trained model?*

Several factors influenced the model's performance, including the size of the dataset and the quality of the images. The adult dataset was significantly smaller than those typically used for training models, which likely contributed to performance differences. Additionally, the limited number of pediatric x-ray images also introduced uncertainties.

Additionally, the lack of standardized image filtering in the adult dataset may have introduced noise, impacting classification accuracy. Future research should focus on expanding datasets, filter the data sets, and optimizing hyperparameters through advanced tuning techniques.

Overall, this study underscores the potential of transfer learning in medical imaging, demonstrating its ability to improve pediatric pneumonia classification with limited data. Further work in this area could enhance model generalizability, aiding in the development of more reliable AI-assisted diagnostic tools.

References

[1] A. Akay and H. Hess. Deep learning: Current and emerging applications in medicine and technology. *IEEE Journal of Biomedical and Health Informatics*, 23(3):906–920, 2019.

[2] F. Wang, L. P. Casalino, and D. Khullar. Deep learning in medicine—promise, progress, and challenges. *JAMA Internal Medicine*, 179(3):293–294, 2019.

[3] Classification of bacterial and viral childhood pneumonia using deep learning in chest radiography. In *Proceedings of the 3rd International Conference on Multimedia and Image Processing*, 2022. Accessed September 28, 2022.

[4] G. Labhane, R. Pansare, S. Maheshwari, R. Tiwari, and A. Shukla. Detection of pediatric pneumonia from chest x-ray images using cnn and transfer learning. In *2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE)*, pages 85–92, 2020.

[5] G. Liang and L. Zheng. A transfer learning method with deep residual network for pediatric pneumonia diagnosis. *Computer Methods and Programs in Biomedicine*, 187:104964, 2020.

[6] E. Ayan, B. Karabulut, and H. M. Ünver. Diagnosis of pediatric pneumonia with ensemble of deep convolutional neural networks in chest x-ray images. *Arab J Sci Eng*, 47(2):2123–2139, 2022.

[7] V. Ravi, H. Narasimhan, and T. D. Pham. A cost-sensitive deep learning-based meta-classifier for pediatric pneumonia classification using chest x-rays. *Expert Systems*, 39(7):e12966, 2022.

[8] R. Alsharif, Y. Al-Issa, A. M. Alqudah, I. A. Qasmieh, W. A. Mustafa, and H. Alquran. Pneumonianet: Automated detection and classification of pediatric pneumonia using chest x-ray images and cnn approach. *Electronics*, 10(23):2949, 2021.

[9] Maria Rollan-Martinez-Herrera, Alejandro A. Díaz, Rubén San José Estépar, Gonzalo Vegas Sanchez-Ferrero, James C. Ross, Raúl San José Estépar, and Pietro Nardelli. CNNs trained with adult data are useful in pediatrics. a pneumonia classification example. 19(7):e0306703.

[10] Chiyuan Zhang Behnam Neyshabur, Hanie Sedghi. What is being transferred in transfer learning?

[11] Jinglan Zhang Ye Duan Yuantong Gu Zehui Zhao, Laith Alzubaidi. A comparison review of transfer learning and self-supervised learning: Definitions, applications, advantages and limitations.

[12] Keyu Duan Dongbo Xi Yongchun Zhu Hengshu Zhu Fuzhen Zhuang, Zhiyuan Qi. A comprehensive survey on transfer learning.

[13] Khaled Rasheed Mohammadreza Iman, Hamid Reza Arabnia. Review of deep transfer learning and recent advancements. *MDPI*, 2023.

[14] Anca M. Hanea Bruce G. Marcot. What is an optimal value of k in k-fold cross-validation in discrete bayesian network analysis? *Springer*, 36, 2020.

[15] F Segovia J Ramirez A Ortiz J. Suckling Juan M Gorriz, R. Martin Clemente. Is k-fold cross validation the best model selection method for machine learning. *Cornell University*, 2024.

[16] CXR8 | Ontwikkeld door Box.

[17] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M. Summers. ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3462–3471, July 2017. ISSN: 1063-6919.

[18] Chest X-Ray Images (Pneumonia).

[19] Morteza Heidari, Seyedehnafiseh Mirniaharikandehei, Abolfazl Zargari Khuzani, Gopichandh Danala, Yuchen Qiu, and Bin Zheng. Improving the performance of cnn to predict the likelihood of covid-19 using chest x-ray images with preprocessing algorithms. *International Journal of Medical Informatics*, 144:104284, 2020.

[20] Agata Giełczyk, Anna Marciniak, Martyna Tarczewska, and Zbigniew Lutowski. Pre-processing methods in chest x-ray image classification. *PLOS ONE*, 17(4):e0265949, 2022.

[21] Sandeep Singh Sandha, Mohit Aggarwal, Igor Fedorov, and Mani Srivastava. Mango: A python library for parallel hyperparameter tuning. pages 3987–3991, 2020.

[22] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework. page 2623–2631, 2019.