WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Detection of Pneumonia Using Deep Learning and Machine Learning Techniques

Addala Anisha Vidyani[1], Kondapally Saicharan[2], Laasya Gowthami Thota[3], Pon Vignesh G[4], Harshini Pothireddy[5], Ayush Makarand Deo[6], Nishant Prakashchandra Patel[7], Kartikeya Reddy Tunga[8]

Gitam School of Pharmacy, Vizag[1], SRMIST[2], Vignan University, Vadlamudi[3], Thiagarajar College of Engineering, Madurai[4], Independent Researcher[5], Shri Ramdeobaba College of Engineering and Management[6], Devang Patel Institute of Advance Technology and Research[7], R.M.K. Engineering College[8]

ABSTRACT :

Pneumonia continues to be a leading cause of mortality, especially in children under five and the elderly population. Its diagnosis typically relies on the visual inspection of chest radiographs (CXRs) by experienced radiologists, which is time-consuming, subjective, and often inconsistent due to overlapping visual indicators with other thoracic conditions such as pulmonary edema, atelectasis, and lung malignancies. In this capstone project, we explore the application of deep learning techniques for the automated detection and localization of lung opacities in chest X-ray images, which are one of the primary visual cues of pneumonia.

The project utilizes a two-fold approach. First, image classification models based on transfer learning were developed using pre-trained architectures—VGG16 and ResNet50—to classify CXR images into three categories: Normal, No Lung Opacity/Not Normal, and Lung Opacity. The models were trained using enhanced images processed through histogram equalization to improve contrast. Extensive exploratory data analysis (EDA) revealed class imbalance and multiple bounding box annotations for single patients, which were carefully addressed during preprocessing.

Second, Faster Region-based Convolutional Neural Network (Faster R-CNN) was implemented for object detection to accurately localize lung opacity regions within the radiographs. This model employed ResNet50 as its backbone and was trained exclusively on images labeled with lung opacities (Target=1), due to annotation limitations in the other categories. Performance metrics such as classification accuracy, Intersection over Union (IoU), and mean Average Precision (mAP) were used to evaluate the effectiveness of the models. The ResNet50-based Faster R-CNN achieved a classification accuracy of approximately 86.57% and a mAP of 0.383, indicating promising results in opacity localization.

Despite the performance, limitations such as computational constraints, model overfitting, and lack of generalizability to diverse demographics were encountered. The report also discusses the ethical, practical, and regulatory implications of deploying AI in medical diagnostics and highlights the need for further validation with clinical data and multi-institutional datasets.

This study demonstrates the feasibility of using deep learning to support radiologists in detecting pneumonia by automating the identification of lung opacities. It lays the groundwork for future integration of AI in real-world clinical workflows, aiming to enhance diagnostic speed, accuracy, and accessibility in resource-limited settings.

Introduction

Pneumonia is a potentially life-threatening infection that affects the lungs, leading to the inflammation of the air sacs known as alveoli. These air sacs may fill with fluid or pus, causing symptoms such as cough with phlegm or pus, fever, chills, and difficulty breathing. It can be caused by a variety of organisms including bacteria, viruses, and fungi. Globally, pneumonia remains one of the leading causes of death in children under the age of five, accounting for over 15% of all deaths in this age group. In 2017 alone, an estimated 920,000 children died from pneumonia worldwide. Despite being a preventable and treatable disease, the burden remains high due to delayed or incorrect diagnoses, especially in low-resource settings.

A reliable diagnosis of pneumonia typically involves a combination of clinical examination, laboratory tests, and radiographic imaging—most commonly, a chest X-ray (CXR). However, interpreting CXRs requires specialized radiological expertise, and the manual process is prone to inter-observer variability, fatigue, and human error. Furthermore, several lung conditions, such as atelectasis, pulmonary edema, lung malignancies, and pleural effusions, can mimic pneumonia on radiographs by presenting similar patterns like increased lung opacity. This overlap often complicates accurate diagnosis, making it both time-sensitive and skill-intensive.

The growing availability of digital medical images and advances in deep learning techniques have opened new avenues in medical diagnostics. Particularly, convolutional neural networks (CNNs) have demonstrated remarkable success in image recognition tasks across various domains, including

healthcare. Their ability to automatically extract and learn hierarchical image features makes them highly suitable for analyzing complex medical images such as CXRs.

This capstone project aims to leverage the power of deep learning to assist in the diagnosis of pneumonia through two primary objectives:

1. **Classification Task:** To develop a robust deep learning model that can classify chest X-ray images into categories based on the presence or absence of lung opacity—an essential radiographic indicator of pneumonia.
2. **Localization Task:** To design an object detection system capable of accurately identifying and drawing bounding boxes around lung opacity regions, thereby pinpointing areas of concern on the CXR images.

The study employs pre-trained CNN models, specifically VGG16 and ResNet50, for classification tasks, and integrates the Faster R-CNN object detection framework for spatial localization of lung opacity. Extensive preprocessing, including histogram equalization, was conducted to enhance image contrast, improving the performance of learning algorithms. Additionally, comprehensive exploratory data analysis (EDA) was performed on the metadata and image datasets to understand underlying patterns, class distributions, and potential biases.

Through this research, we not only seek to improve the speed and accuracy of pneumonia detection but also aim to demonstrate how artificial intelligence (AI) can support healthcare professionals in making more informed clinical decisions. The integration of such models has the potential to significantly reduce diagnostic delays, especially in resource-constrained environments where access to expert radiologists is limited.

Ultimately, this work contributes to the ongoing efforts in medical AI to bridge the gap between radiological expertise and scalable diagnostic tools by creating a model that can function as a second opinion, triage assistant, or even an autonomous diagnostic system in the future.

Related Work

In recent years, the application of deep learning in the medical imaging domain has witnessed exponential growth, particularly in radiology. Several studies have demonstrated the efficacy of convolutional neural networks (CNNs) in detecting anomalies in chest radiographs, such as pneumonia, tuberculosis, lung cancer, and other pulmonary diseases.

One of the earliest and most cited works is **CheXNet** by Rajpurkar et al. (2017), where a 121-layer DenseNet was trained on the ChestX-ray14 dataset to detect pneumonia. The model not only outperformed traditional machine learning classifiers but also showed performance comparable to expert radiologists. CheXNet set a foundation for deep learning-based diagnosis and was a benchmark for several subsequent studies.

Another notable study is by **Stephen Oakden-Rayner et al. (2019)**, which explored the generalizability issues of CNN-based models across different hospital systems. Their research highlighted the challenge of dataset bias — models trained on one hospital's data often performed poorly on data from another due to demographic and equipment variations. This finding emphasizes the need for robust and diverse datasets in clinical AI research.

The **RSNA Pneumonia Detection Challenge** on Kaggle (2018) spurred significant research activity. Participants attempted to identify and localize pneumonia in chest X-rays using DICOM images and corresponding bounding box annotations. Many top-performing models used ensemble methods combining architectures like ResNet, VGG, and DenseNet with region-based detection algorithms like **Faster R-CNN**, **YOLO**, and **RetinaNet**.

In the object detection space, **Ren et al. (2015)** introduced **Faster R-CNN**, which replaced the traditional selective search with a Region Proposal Network (RPN). This drastically improved the speed and accuracy of object detection, making it highly suitable for medical imaging tasks where real-time and precise localization is critical.

Further improvements were seen in **RetinaNet** (Lin et al., 2017), which introduced **Focal Loss** to handle the extreme class imbalance problem in object detection — a common scenario in medical datasets. RetinaNet's one-stage detector architecture provided a balance between speed and accuracy, making it a preferred choice for many healthcare AI solutions.

From a real-world clinical integration standpoint, **Google Health's AI for lung cancer detection** (Ardila et al., 2019) and **NYU's breast cancer detection system** (McKinney et al., 2020) showed that deep learning models could achieve or exceed radiologist-level accuracy in specialized tasks. However, these models also highlighted critical issues like lack of interpretability, generalizability, and regulatory barriers.

Recent advancements have also focused on **explainable AI (XAI)** in radiology, to improve clinician trust. Techniques such as Grad-CAM and integrated gradients are being used to visualize the decision-making process of CNNs.

Model Design and Implementation Overview

The system pipeline is designed to handle both classification and object detection tasks on chest X-ray images for pneumonia detection. The workflow begins with preprocessing steps, where DICOM-format images are converted to PNG format, followed by histogram equalization to enhance contrast. The images are uniformly resized to 300×300 pixels and normalized for input to deep learning models.

For the classification task, transfer learning was utilized with two pretrained CNN models: **VGG16** and **ResNet50**. These models were modified by replacing the top layers with custom fully connected layers suitable for binary (Target = 0/1) and multiclass classification (Normal, No Opacity, Lung Opacity). The initial layers of the models were kept frozen during early training to retain low-level visual features from the original ImageNet training, while the deeper layers were fine-tuned using the chest X-ray dataset. To improve generalization and address overfitting, the dataset was augmented using horizontal and vertical flips, random rotations, and zoom transformations via ImageDataGenerator.

In the second stage, object detection was performed using the **Faster R-CNN** architecture with **ResNet50** as the backbone. Only images with labeled lung opacities (Target = 1) were used for this phase, as they included bounding box annotations. The Region Proposal Network (RPN) was configured with anchor boxes of different scales and aspect ratios to capture varying opacity sizes. RoI pooling was applied to standardize region sizes, followed by classification and bounding box regression heads to refine predictions.

The model was trained for 40 epochs with the Adam optimizer and a learning rate of 1e-5. Evaluation metrics included accuracy, IoU, mAP, and loss values. The system demonstrated robust performance in localizing lung opacities, with good alignment between predicted and ground truth boxes.

Training Strategy and Evaluation Metrics

The training process was carried out in two phases: image classification and object detection. For classification, both VGG16 and ResNet50 were initialized with pretrained ImageNet weights. The top fully connected layers were replaced with task-specific dense layers including ReLU activations and dropout regularization to prevent overfitting. During training, the lower layers were initially frozen, and only the custom classifier layers were trained. In the later epochs, selective unfreezing of deeper convolutional blocks was performed for fine-tuning.

The models were compiled using the **categorical cross-entropy loss** function for multiclass classification and **binary cross-entropy** for the binary target prediction. Two optimizers—**Adam** and **RMSprop**—were tested across learning rates ranging from 0.0001 to 2e-5. Batch size was kept relatively low (16 or 32) due to hardware limitations on Google Colab.

To ensure robust evaluation, the dataset was split into 70% training and 30% validation sets. Class imbalance (Target = 0 dominating Target = 1) was mitigated using class weights and data augmentation strategies including random horizontal/vertical flips, zooming, and rotations.

For object detection, **Faster R-CNN** was trained using only the images labeled with Target = 1. The loss function for the detection model was a **multi-task loss**, combining:

- **RPN classification loss** (object vs background),
- **RPN regression loss** (bounding box adjustment),
- **ROI classification loss** (classifying proposal region),
- **ROI regression loss** (refining box coordinates).

The following evaluation metrics were used to quantify model performance:

- **Accuracy**: Percentage of correctly predicted classes.
- **Precision and Recall**: To assess relevance and completeness of classification, especially in imbalanced classes.
- **F1-Score**: Harmonic mean of precision and recall.
- **Intersection over Union (IoU)**: Measures the overlap between predicted and ground-truth bounding boxes.
- **Mean Average Precision (mAP)**: Primary metric for object detection tasks, summarizing detection precision across multiple thresholds.
- **Loss Curves**: Monitored across epochs to track convergence and detect overfitting.

During object detection training, the model achieved steady improvements in mAP and IoU over 40 epochs, validating its capability to accurately localize pneumonia-affected regions.

Objectives

The objective of this project is to develop an AI-based diagnostic system for the automated detection and localization of pneumonia in chest X-ray images by identifying regions of lung opacity. The study aims to build a robust image classification model using convolutional neural networks (CNNs), specifically VGG16 and ResNet50, to distinguish between normal cases, no lung opacity/not normal cases, and cases with evident lung opacity. Additionally, an object detection model based on the Faster R-CNN architecture is implemented to accurately identify and localize affected regions using bounding boxes. To ensure consistent image quality and model performance, the project incorporates preprocessing techniques such as histogram equalization and image resizing.

The research further involves extensive exploratory data analysis (EDA) of the metadata and image attributes to understand class distribution, patient demographics (age and sex), and view positions. Transfer learning is applied by fine-tuning pretrained models originally trained on large-scale datasets like ImageNet, allowing the reuse of low-level visual features for medical image analysis. The models are evaluated using standard performance metrics including accuracy, precision, recall, F1-score, intersection over union (IoU), and mean average precision (mAP). The project also addresses common challenges such as data imbalance, model overfitting, and limited computational resources by applying augmentation strategies and hyperparameter optimization.

Finally, the study reflects on the ethical and practical implications of deploying AI in clinical environments, especially in radiology. It emphasizes the importance of generalizability across different patient populations and imaging conditions, and proposes strategies to bridge the gap between academic research and real-world clinical adoption through better data collection, validation, and regulatory alignment.

Methodology

This study adopts a hybrid methodology combining image classification and object detection to identify and localize pneumonia-related lung opacities in chest X-ray images. The approach is structured in multiple stages: data preprocessing, exploratory data analysis, image enhancement, model training for classification, and object detection using deep learning architectures.

Initially, the dataset comprising over 26,000 chest radiograph images in DICOM (.dcm) format was obtained from the RSNA Pneumonia Detection Challenge on Kaggle. The images were supplemented by two CSV files containing metadata and bounding box annotations. Data preprocessing began with converting DICOM files to a suitable image format (PNG), followed by histogram equalization to enhance image contrast. All images were then resized uniformly to 300x300 pixels to standardize input dimensions for the deep learning models.

Exploratory Data Analysis (EDA) was performed on both image metadata and class labels to understand the distribution of data across various parameters including patient age, gender, view position, and class imbalance. It was observed that a significant portion of the dataset was skewed towards non-opacity classes, necessitating balancing techniques during model training.

For the classification task, pre-trained convolutional neural networks—VGG16 and ResNet50—were employed using a transfer learning approach. The final classification layers were replaced with custom dense layers to adapt the networks for binary (Target: 0/1) and multiclass (Normal, No Opacity,

Lung Opacity) classification. Two training methodologies were evaluated: the first involved direct training on enhanced images, and the second incorporated data augmentation using the Keras ImageDataGenerator to improve model generalization. Loss functions used included categorical cross-entropy, and optimizers like Adam and RMSprop were tested with various learning rates. Despite high training accuracy, models exhibited overfitting due to class imbalance and limited augmentation.

For object detection, the Faster R-CNN architecture with ResNet50 as the backbone was implemented to localize lung opacities through bounding boxes. Since only images labeled with lung opacity (Target = 1) had annotation data, the object detection model was trained exclusively on that subset. The Region Proposal Network (RPN) was used to generate candidate regions, which were refined through RoI (Region of Interest) pooling and passed through classification and regression heads. Bounding box coordinates were adjusted based on image resizing, and augmentation was applied through random flips and rotations to increase training variability.

Model performance was assessed using metrics such as accuracy, precision, recall, F1-score for classification, and Intersection over Union (IoU) along with mean Average Precision (mAP) for object detection. Training progress and loss curves were logged across epochs to monitor learning trends and prevent overfitting.

The proposed methodology successfully integrates classification and detection to provide a comprehensive diagnostic aid, capable of identifying both the presence and precise location of pneumonia-related lung opacities from chest X-ray images.

Experiments and Results

To evaluate the effectiveness of the proposed pneumonia detection framework, a series of experiments were conducted for both classification and object detection tasks. All experiments were executed using Python and Keras with TensorFlow backend, and due to resource constraints, training was performed on Google Colab with GPU acceleration.

For the **classification task**, both VGG16 and ResNet50 models were fine-tuned using transfer learning. The last few convolutional layers of the pre-trained networks were unfrozen, and new fully connected layers were added for binary classification (Target: 0 - No Opacity, 1 - Opacity) as well as multiclass classification (Normal, No Lung Opacity/Not Normal, Lung Opacity). The input images were resized to 300×300 and normalized, while histogram equalization was applied to enhance contrast. Two optimizers—Adam and RMSprop—were tested with learning rates of 0.0001, 0.001, and 2e-5. Despite achieving high training accuracy (>98%), the validation accuracy plateaued around 72–74% for binary classification and 33–35% for multiclass classification, indicating overfitting due to class imbalance.

The **ResNet50 model** demonstrated slightly better performance than VGG16, with a validation accuracy of 73.47%, precision of 36.73%, recall of 50%, and F1-score of 42.35% for the binary classification task. The multiclass setup yielded lower metrics due to overlapping visual features and label imbalance. Confusion matrices further revealed a high false-negative rate, especially for lung opacity cases, suggesting a need for better data augmentation and balancing.

For the **object detection task**, the Faster R-CNN architecture was implemented with ResNet50 as the backbone. Due to the absence of bounding box annotations for Target = 0 images, the model was trained exclusively on 6,012 lung opacity images (Target = 1). The Region Proposal Network generated anchors at different scales (64, 128, 256) and aspect ratios (1:1, 1:2, 2:1), and data augmentation techniques such as horizontal/vertical flips and 90° rotations were employed. The input image size was standardized to 300×300, and bounding boxes were rescaled accordingly.

The Faster R-CNN model was trained for 40 epochs. Key evaluation metrics included Mean Intersection over Union (IoU), classification accuracy, and Mean Average Precision (mAP). The IoU steadily improved across epochs, reaching over 5.33, and the final mAP achieved was **0.383**, indicating a reasonable detection capability given the dataset size and hardware constraints. The classification accuracy for object detection reached **86.57%**, while overall loss reduced significantly from 5.12 to 1.63 during training.

Qualitative results further validated the model's performance. In sample predictions, bounding boxes closely matched the ground truth, and the model performed robustly even on augmented (flipped/rotated) images. However, certain false positives were noted, especially in regions with overlapping anatomical structures.

These experiments demonstrate that the proposed framework, especially with Faster R-CNN for localization, holds promise for clinical screening assistance. However, further improvements in data balance, augmentation, and training infrastructure are needed to enhance generalizability and real-world performance.

Discussion and Limitations

The results of this study highlight both the potential and challenges of using deep learning for pneumonia detection in chest radiographs. While classification models such as ResNet50 and VGG16 showed promising training performance, their relatively lower validation accuracy indicates overfitting, primarily due to class imbalance and limited variability in the dataset. Additionally, the multiclass classification task faced difficulty distinguishing between "No Opacity/Not Normal" and "Normal" classes, likely due to overlapping visual features and insufficient contextual data.

The object detection approach using Faster R-CNN yielded better generalization, achieving a mean average precision (mAP) of 0.383 and classification accuracy of 86.57%. The model demonstrated good localization of lung opacities even in augmented scenarios, validating its robustness. However, computational limitations on Google Colab restricted the ability to run larger batch sizes, deeper networks (like EfficientDet or RetinaNet), or extended epochs, which could further improve results.

From a real-world deployment perspective, challenges remain. The current model is trained on a specific dataset and may not generalize well to different hospitals or demographic groups. Furthermore, lung opacity alone is not a definitive indicator of pneumonia, and clinical symptoms must also be considered. Future work should integrate NLP-based symptom analysis, diverse datasets, and potentially multi-modal diagnostics for higher clinical relevance.

Conclusion

This study presents a deep learning-based framework for the classification and localization of pneumonia-related lung opacities from chest X-ray images. By leveraging transfer learning on pre-trained CNN architectures such as VGG16 and ResNet50, and integrating Faster R-CNN for object detection, the proposed system effectively identifies both the presence and spatial location of abnormalities. The classification models demonstrated high training accuracy but encountered overfitting during validation, primarily due to class imbalance and limited augmentation. Nonetheless, the object detection model offered better generalization, achieving a mean average precision (mAP) of 0.383 and providing reasonably accurate bounding box predictions on lung opacity cases.

The outcomes of this research confirm the potential of AI-powered tools to assist in radiological diagnostics, especially in environments with limited access to trained radiologists. The combination of classification and object detection in a unified pipeline adds clinical relevance by not only flagging abnormal cases but also visually guiding radiologists to the regions of concern. Furthermore, the use of image preprocessing techniques like histogram equalization, and the adoption of data augmentation strategies, contributed to enhanced model robustness despite hardware constraints.

However, for real-world deployment, the framework must address several challenges. The dataset used is restricted in diversity, and the models may not perform equally well across varying demographics, imaging devices, or clinical conditions. Moreover, lung opacity, although indicative of pneumonia, is not a definitive marker in isolation; other conditions such as COVID-19, edema, or fibrosis may present similar radiographic patterns. Thus, to truly serve as a clinical decision support tool, integration with patient symptoms, history, and additional diagnostic tests is essential.

Future Work

Building upon the findings of this study, several directions can be explored to further enhance the performance, reliability, and clinical applicability of the proposed framework. One of the primary areas of improvement lies in the expansion and diversification of the dataset. Incorporating chest X-ray data from multiple healthcare institutions across various geographic regions, age groups, and patient demographics will help reduce model bias and improve generalizability. Additionally, the inclusion of varied imaging modalities and devices can enable the model to adapt to real-world heterogeneity in medical imaging practices.

Another promising direction involves the integration of clinical metadata and symptom-based patient information through natural language processing (NLP). Since radiographic findings alone may not be sufficient for conclusive diagnosis, fusing imaging data with clinical symptoms such as fever, cough, oxygen saturation, or blood reports could significantly enhance diagnostic accuracy. The development of a multi-modal AI system that considers both visual and textual information can provide a more holistic and context-aware assessment.

From a technical standpoint, future research can explore the adoption of more advanced deep learning architectures. Models such as **RetinaNet**, **EfficientDet**, **YOLOv8**, and transformer-based vision models like **Swin Transformer** or **Vision Transformers (ViT)** offer improved performance and faster inference times. These architectures also come with features such as focal loss and multi-scale detection, which are well-suited for detecting subtle abnormalities like lung opacities.

Moreover, the current model's deployment is limited by the computational resources available during experimentation. Future implementations can leverage cloud-based GPU clusters or high-performance computing platforms to support larger batch sizes, longer training schedules, and hyperparameter optimization through AutoML frameworks.

Lastly, transitioning from research to real-world clinical deployment will require rigorous validation in collaboration with radiologists, as well as alignment with medical regulations such as FDA or CE approvals. User interface design, explainability of predictions (e.g., using Grad-CAM), and integration into hospital information systems (HIS) are also critical steps toward operationalizing AI in medical imaging.

In essence, while the current study lays a solid foundation, future work must focus on data diversity, clinical context integration, architectural innovation, and real-world validation to build a trustworthy and scalable AI solution for pneumonia detection and beyond.

REFERENCES

[1] P. Rajpurkar, J. Irvin, K. Zhu, et al., "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning," *arXiv preprint arXiv:1711.05225*, 2017.

[2] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2017.

[3] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988.

[4] D. Ardila, C. Kiraly, S. Bharadwaj, et al., "End-to-End Lung Cancer Screening with Three-Dimensional Deep Learning on Low-Dose Chest CT," *Nature Medicine*, vol. 25, no. 6, pp. 954–961, 2019.

[5] S. McKinney, M. Sieniek, V. Godbole, et al., "International Evaluation of an AI System for Breast Cancer Screening," *Nature*, vol. 577, pp. 89–94, 2020.

[6] Kaggle RSNA Pneumonia Detection Challenge, [Online]. Available: https://www.kaggle.com/c/rsna-pneumonia-detection-challenge

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[8] Z. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 936–944.

[9] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 6517–6525.

[10] A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in *Proc. Int.*

Conf. Learn. Representations (ICLR), 2021.

[11] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and Efficient Object Detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 10781–10790.

[12] J. Deng, W. Dong, R. Socher, et al., "ImageNet: A Large-Scale Hierarchical Image Database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 248–255.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[14] B. Zhou, A. Khosla, A. Lapedriza, et al., "Learning Deep Features for Discriminative Localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2921–2929.

[15] TensorFlow DICOM Image Processing Tutorial, [Online]. Available: https://www.tensorflow.org/tutorials/load_data/images

[16] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A Simple Framework for Contrastive Learning of Visual Representations," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 1597–1607.

[17] R. R. Selvaraju, M. Cogswell, A. Das, et al., "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 618–626.

[18] B. Litjens, T. Kooi, B. E. Bejnordi, et al., "A Survey on Deep Learning in Medical Image Analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[19] S. Oakden-Rayner, "Exploring Large-scale Public Medical Image Datasets," *Academic Radiology*, vol. 27, no. 1, pp. 106–112, 2020.

[20] A. M. O'Leary, K. A. Towbin, and D. T. Maeder, "Radiologic Diagnosis of Pneumonia: A Practical Guide," *Radiologic Clinics of North America*, vol. 59, no. 5, pp. 703–716, 2021.