

RESEARCH

Open Access



Enhancing medical diagnosis on chest X-rays: knowledge distillation from self-supervised based model to compressed student model

Jaydeep Kishore¹, Akshita Jain¹, K. Krishna Koushika¹, Pawan Kumar Mishra², Shekhar Karanwal³ and Surendra Solanki^{1*}

*Correspondence:
Surendra Solanki

surendra.solanki@jaipur.manipal.edu

¹Department of Artificial Intelligence and Machine Learning, Manipal University Jaipur, Jaipur 303007, Rajasthan, India

²Department of CSE, Graphic Era Deemed to be University, Dehradun 248002, Uttarakhand, India

³Department of Computer Engineering and Applications, GLA University, Mathura 281406, Uttar Pradesh, India

Abstract

Deep learning and self-supervised learning techniques have advanced, making it possible to diagnose medical images more accurately. Our goal in this work is to increase the accuracy of medical diagnosis using chest X-rays by utilising information distillation and model compression techniques. By extracting knowledge from a self-supervised model which is a teacher model (SWAV Model with ResNet-50 as backbone), we enhance inference speed and reduce the computational resources needed for correct diagnosis. Our strategy entails translating the knowledge acquired by self-supervised models to smaller, more efficient models without sacrificing accuracy. In addition, we look at how model compression and distillation affect the diagnosis's interpretation. The results of this study may enhance medical diagnosis procedures and increase their accessibility in environments with limited resources. Our extensive experiments prove the efficacy with 97.34% accuracy for the student model with knowledge distillation.

Keywords Knowledge distillation, Deep neural networks, Student–teacher learning (S–T), Chest X-ray

1 Introduction

Chest X-Rays (CXR) are a frequent diagnostic technique in medicine. They provide information on many lung and cardiovascular disorders, such as pneumonia, lung cancer, and tuberculosis. Since deep learning has been developed, there have been more attempts to automate the interpretation of CXR images in an effort to assist radiologists with diagnosis and therapy. However, there are several obstacles that must be overcome before deep learning models can be used for CXR diagnosis. These include the requirement for sizable annotated datasets, computing power, and model interpretability. For a number of illnesses, including those that cause coughing and chest pain, a chest X-ray can be quite helpful or dyspnea. For example, a chest X-ray can be used to distinguish between heart failure, which may show signs of fluid buildup in the lungs, and pneumonia, which usually shows signs of pulmonary infiltration.



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

In contrast to more sophisticated imaging methods like CT or MRI scans, lung X-rays are much more affordable and accessible. This makes them an invaluable tool, particularly in environments with limited resources where access to pricey diagnostic tools may be restricted. Furthermore, many patients—including children and pregnant women—find chest X-rays to be a safer alternative due to their comparatively low radiation exposure when compared to other imaging modalities. Moving from predicting a single illness to predicting multiple diseases can be utilized to categorize recent studies on CXR image classification. These tags show an illness affecting the heart, lungs, or other frequent chest ailments. Developments in NLP techniques now allow for the automated analysis of radiology reports, which can pinpoint specific labels related to each chest x-ray image. Recent advancements have enabled the development and distribution of various CXR datasets, fostering the potential for the creation of more sophisticated deep learning models or complex model structures [1]. Having access to extensive, marked datasets is advantageous for the data-focused method of deep learning, and the number of CXR imaging datasets that are publicly accessible is growing, enabling the development of highly precise models [2]. Improved performance of deep learning makes it an attractive option for medical image analysis [3–7]. DenseNet [8] is an advanced deep neural network (DNN) that provides state-of-the-art precision in medical imaging tasks. Nevertheless, due to the significant computational expenses associated with processing dense features, smaller devices may encounter challenges in implementing them.

Most researchers, such as [9–13], have focused on classifying chest X-ray images for Pneumonia/Normal detection using various deep learning architectures. However, their primary emphasis has been on improving model performance, with limited attention given to reducing the model's complexity—specifically the number of parameters and GFLOPs. One of the less-explored yet significant challenges in deep learning is the high computational overhead associated with large models. Reducing parameters and GFLOPs is a promising strategy to alleviate this burden, especially during the training and deployment of deep learning models. To tackle this issue, knowledge from large, pre-trained models (with high parameter count and GFLOPs) can be distilled into smaller, lightweight architectures tailored for specific tasks or datasets. Knowledge distillation is a technique that enables this process by transferring the learned representations from a high-capacity teacher model to a more compact student model [14–16]. This not only supports effective model compression but also facilitates the deployment of deep learning solutions on resource-constrained devices. Despite its advantages, traditional knowledge distillation methods still offer room for improvement, particularly in enhancing the performance of student models without compromising efficiency. Recently, KD has been utilized in various applications for categorizing CXR images. Examples of such applications can be seen in studies related to CXR image categorization [17–19].

In recent years, self-supervised learning (SSL) models have emerged as a promising alternative to supervised learning for computer vision tasks, primarily due to their ability to learn from data with minimal or no annotations in the target domain. SSL involves training a model to extract meaningful features from unlabeled data. These features can later be fine-tuned using a smaller set of labeled data to perform downstream tasks, for instance, predicting diseases from chest X-ray datasets.

In our previous works [20–23], we observed that SSL models consistently outperformed traditional supervised learning approaches. This motivated us to adopt SSL

models, particularly SWAV [24], as the teacher model in our knowledge distillation framework.

This paper presents a knowledge distillation approach that transfers knowledge from a self-supervised teacher model (SWAV) to a student model based on ResNet-18, aimed at disease prediction using the chest X-rays dataset [25]. The overall framework of the proposed method is illustrated in Figure 1. The main contributions of this work can be summarized as follows:

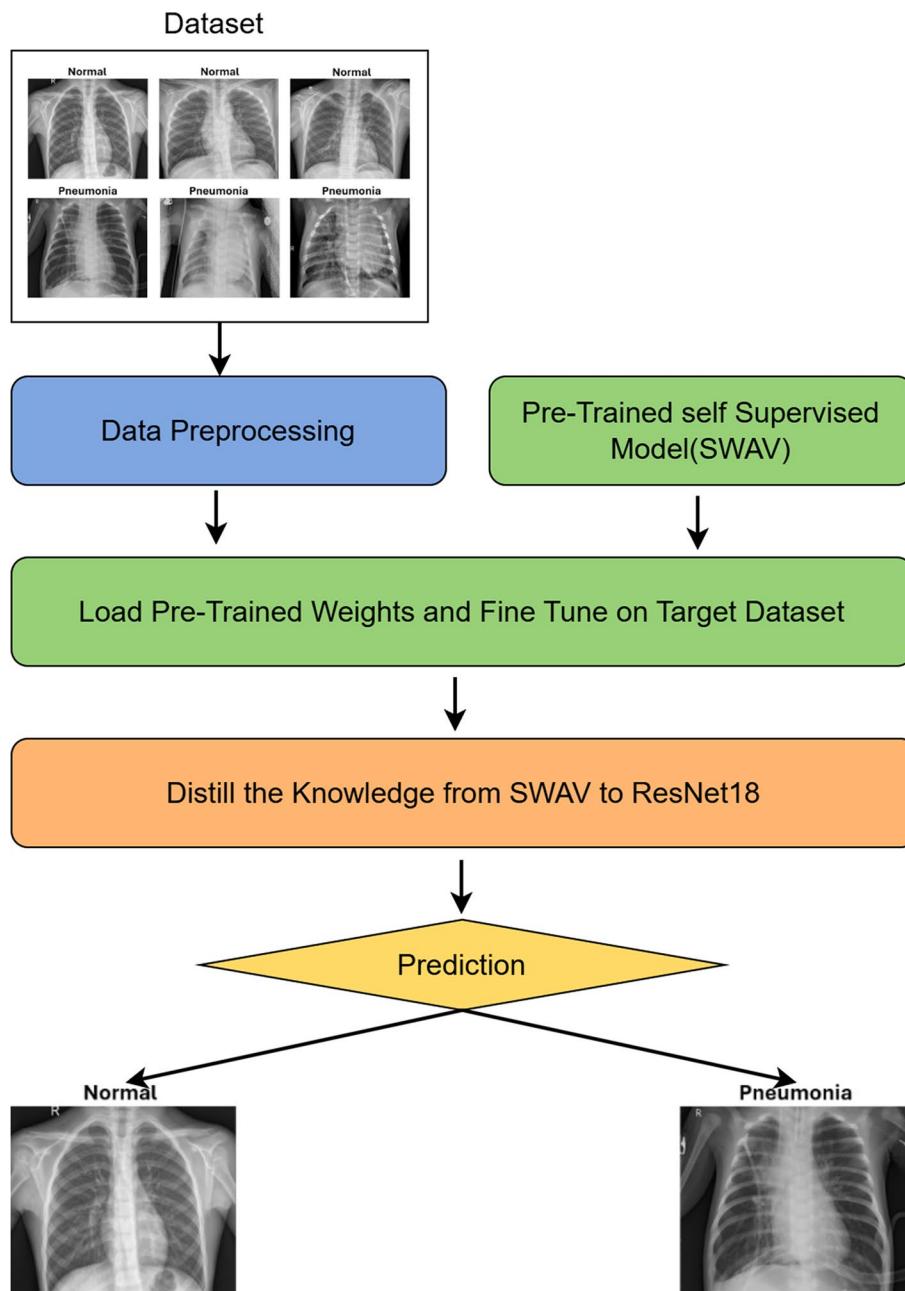


Fig. 1 Flow of our methodology: Here we take the pretrained weights of SWAV [24] self supervised model which contain ResNet-50 as backbone. The weights SWAV models [24] is tuned using transfer learning (TL) against the target dataset. Finally the knowledge of SWAV model is transfer to the student model (ResNet-18) using Knowledge distillation techniques

- This paper employ a pretrained self-supervised (SWAV) model and fine-tune it on the Chest X-ray Pneumonia dataset, showing the potential of self-supervised learning representations in medical image classification.
- We propose a knowledge distillation framework in which the SWAV-based teacher model transfers its knowledge to a lightweight ResNet-18 student model, achieving high diagnostic accuracy with reduced computational cost in term of parameters and GFLOPs.
- Our student model achieves an accuracy of 97.34%, outperforming several state-of-the-art models while being optimized for resource-constrained environments.
- We perform detailed analysis of model complexity and show the student model's suitability for real-time deployment in clinical and mobile settings.
- We provide a comparative evaluation using standard metrics (accuracy, precision, recall, and F1-score) and benchmark the model against other deep learning architectures.

2 Related work

This section reviews related works in key areas relevant to our proposed method. We begin with an overview of deep learning applications in medical imaging, followed by a focused discussion on chest X-ray classification using deep neural networks. We then explore an overview of studies on self-supervised learning in medical imaging. Subsequently, we examine existing approaches to model compression specifically knowledge distillation. Finally, we highlight previous efforts that specifically focus on applying knowledge distillation techniques in the context of medical imaging.

2.1 Deep learning applications in medical imaging

Deep learning has made significant contributions in medical imaging, providing solutions for disease classification, lesion detection, segmentation, and image enhancement across multiple modalities, including X-ray, MRI, CT, and histopathology. Convolutional Neural Networks, transformers, and hybrid deep learning architectures have shown promising performance in diagnostic tasks such as breast cancer classification [26], lung cancer detection [27], and pathology image enhancement using super-resolution techniques [28]. Additionally, weakly supervised and self-supervised frameworks have been utilized to reduce the dependency on large labeled datasets while maintaining competitive accuracy [29, 30]. Even with these advancements, using deep learning models in real hospitals and clinics is still difficult. This is because they often need a lot of computing power, large amounts of labeled data, and their decisions can be hard to understand. These challenges have led researchers to focus on building models that are faster, and easier to use. In our work, we tackle these challenges by utilizing a self-supervised learning-based pretrained SWAV model [24], and employing knowledge distillation to train a lightweight and efficient student model (ResNet-18) for pneumonia vs. normal classification using a Chest X-ray dataset. This approach enables real-time deployment in resource-constrained settings without compromising model performance. In the following subsection, we examine prior work on Chest X-ray diagnosis using deep learning, which is closely related to our problem statement.

2.2 Chest X-ray diagnosis using deep learning

Deep learning approaches have shown better results in chest X-ray (CXR) analysis and other medical imaging applications. One of the models is CheXNet, introduced by Rajpurkar et al. [31], which was trained on a large-scale CXR dataset labeled with multiple diseases. CheXNet provide state-of-the-art performance and even surpassed radiologists in diagnosing conditions such as pneumonia, cardiomegaly, and pneumothorax. Despite being a core diagnostic skill for radiologists, interpreting CXR images remains a complex and cognitively demanding task, as highlighted by Wang et al. [32], which makes automated deep learning solutions particularly valuable in clinical practice.

Similarly, Irvin and colleagues developed an automated system for recognising and classifying thoracic diseases from chest X-ray pictures, and they were able to achieve excellent levels of accuracy in identifying anomalies such as atelectasis, consolidation, and pleural effusion [33]. The study demonstrated how deep learning algorithms may help radiologists diagnose patients quickly and correctly, particularly in resource-constrained settings.

In our proposed model, we utilize the Chest X-ray dataset [25] for classifying images as either pneumonia or normal. Several prior studies [12, 13, 34–41] have also employed the same dataset for pneumonia detection. Kundu et al. [34] proposed an ensemble model combining GoogLeNet, ResNet-18, and DenseNet-121 to boost classification performance, although their work did not focus on optimizing model parameters or reducing GFLOPs. Similarly, Mujahid et al. [35] followed an ensemble strategy using Inception-V3, VGG-16, and ResNet-50 on the same dataset. Qiuyu et al. [13] introduced an attention-based ensemble method to further enhance performance, while Arun et al. [39] applied channel attention within deep CNN architectures for improved pneumonia diagnosis. Other notable studies include the use of EfficientNetV2L [40], weak-label-based learning [41], models based on VGG16 and neural networks [12, 36], and many more. While these approaches demonstrate strong performance, they often rely on large and complex architectures, which are computationally expensive.

Despite the remarkable accuracy of deep learning models in identifying chest X-rays, their use in clinical practice may be limited due to their size and computational requirements. Model compression techniques address this issue by simplifying deep learning models, enabling their deployment on edge devices and in resource-constrained environments without compromising performance.

In our proposed method, we employ a pretrained self-supervised learning (SSL) model, SWAV [24], as the teacher model, which is fine-tuned on our target Chest X-ray dataset. Therefore, in the following subsection, we review some SSL approaches that have been applied in the field of medical imaging.

2.3 Self-supervised learning in medical imaging

The growing popularity of self-supervised learning (SSL) models [42] over traditional supervised methods is primarily due to their ability to learn from unlabeled datasets. SSL enables the extraction of meaningful features through pretext tasks [43], which are then used for downstream tasks such as image classification-using the chest X-ray dataset in our case. Numerous researchers have explored this area, either by training models in a self-supervised manner or by employing pretrained SSL models for downstream

tasks. Zhang et al. [44] provide a review of SSL in the medical imaging domain, analyzing its effectiveness across various tasks, including classification and segmentation.

Cho et al. [45] proposed a methodology in which a MoCo v2 self-supervised model was trained on 4.8 million chest X-ray (CXR) images to learn robust feature representations. These learned features were then fine-tuned on a CXR dataset containing six disease classes for the downstream task of classification. Extending this concept to the veterinary domain, Celniak et al. [46] employed SSL strategies such on a diverse set of inter-species and inter-pathology X-ray datasets. In contrast, Imagawa et al. [47] proposed the feasibility of reducing the number of labeled images by utilizing a limited set of unlabeled medical images. The chest X-ray (CXR) images were first used to pretrain a model using the SSL framework, and the learned representations were subsequently fine-tuned through supervised learning for the target task.

Inspired by recent research on fine-tuning pretrained self-supervised models for downstream tasks in medical imaging, we extend our previous work [20–23] to address pneumonia classification using chest X-ray images. While our earlier studies focused on exploring pretrained SSL models on natural image datasets, this work shifts the focus toward medical domain-specific adaptation and evaluation.

As outlined in the introduction, our proposed methodology employs knowledge distillation to transfer knowledge from a large, fine-tuned self-supervised model (SWAV) to a smaller student model (ResNet-18) over the target chest X-ray dataset. Accordingly, the following subsection reviews relevant prior work on knowledge distillation in medical imaging.

2.4 Model compression and knowledge distillation

Model compression strategies are designed to maintain the accuracy and efficiency of deep learning models while substantially reducing their computational and memory requirements. Common techniques include quantization, pruning, and knowledge distillation (KD), each offering distinct advantages for optimizing model deployment—particularly in resource-constrained environments. Among these, knowledge distillation has emerged as one of the most widely adopted approaches. Kim et al. [48] provided strong experimental evidence supporting the superiority of KD over other compression methods, highlighting its effectiveness in preserving model performance even under aggressive compression settings.

Hinton et. al. [49] initially presented knowledge distillation as a model compression strategy, and they showed that it works well for combining knowledge from multiple models into one [49]. Research in the following years has concentrated on the potential applications and expansions of knowledge distillation in a range of domains, such as speech recognition, computer vision, and natural language processing.

Knowledge distillation has been widely applied in medical imaging to compress deep learning models across various modalities, including MRI, X-ray, and histopathological images, with the goal of improving disease identification. Tajbakhsh et al. [50] introduced a technique using convolutional neural networks (CNNs) for histopathology image analysis, which significantly reduced model size and computational complexity while maintaining high accuracy. Similarly, Liu et al. [51] demonstrated the effectiveness of knowledge distillation in MRI-based brain tumor segmentation, achieving reduced

memory consumption and faster inference times without compromising diagnostic performance.

In the study by Zeng et al. [52], a strategy is proposed for graph classification tasks that combines contrastive self-supervised learning (SSL) with knowledge distillation. The authors design a structured framework involving both teacher and student models, where the teacher is trained using contrastive SSL to learn meaningful representations of graph data. In this self-supervised setting, the teacher generates soft labels (also known as pseudo-labels) for the unlabeled data, which are then used to train the student model. The student learns to replicate the teacher's output using a contrastive SSL-based loss function [53].

Termritthikun et al. [54] developed a knowledge distillation method to detect diseases from chest X-rays on mobile devices. Their approach used powerful teacher models like CNNs and Transformers to guide a smaller student model. They focused on detecting multiple diseases using large datasets such as ChestX-ray14 and CheXpert, and added Grad-CAM to explain the student model's decisions. However, their method still relies on large supervised models and doesn't make use of self-supervised learning.

Two other studies by Asham et al. [55] and Kabir et al. [56] have also utilized the Chest X-ray Pneumonia dataset. Asham et al. proposed a highly compact student model combining MobileNet and ConvMixer, while Kabir et al. implemented a teacher-student framework using a shallow student network. Although both studies effectively demonstrate the application of knowledge distillation on this dataset, our method distinguishes itself by incorporating self-supervised learning in the teacher model. This integration enables our student model to maintain high performance while remaining lightweight and suitable for deployment in real-world, resource-constrained environments.

Next we discuss the proposed method.

3 Proposed methodology

Figure 2 illustrates the overall workflow of the proposed framework for chest X-ray classification using knowledge distillation from a self-supervised teacher model. The approach consists of the following main steps:

- Dataset preparation: The process begins with the collection of labeled chest X-ray images from the publicly available dataset [25]. The data undergoes preprocessing and is then split into training and testing sets (70:30 split) for evaluation.
- Teacher Model Fine-Tuning: A pretrained SWAV model with a ResNet-50 backbone, originally trained in a self-supervised manner on a large-scale image dataset (e.g., ImageNet), is fine-tuned on the labeled chest X-ray dataset. This enables the teacher model to learn representations specific to pneumonia classification without the need for training from scratch.
- Knowledge distillation: After fine-tuning, the knowledge captured by the teacher model is transferred to a lightweight student model (ResNet-18) through knowledge distillation. This step involves training the student to mimic the teacher's softened output probabilities (soft labels), as well as the ground truth labels (hard labels).
- Student model architecture: The student network used in this study is ResNet-18, selected for its compact architecture and strong performance on image classification tasks. The student is trained using a hybrid loss function composed of cross-entropy

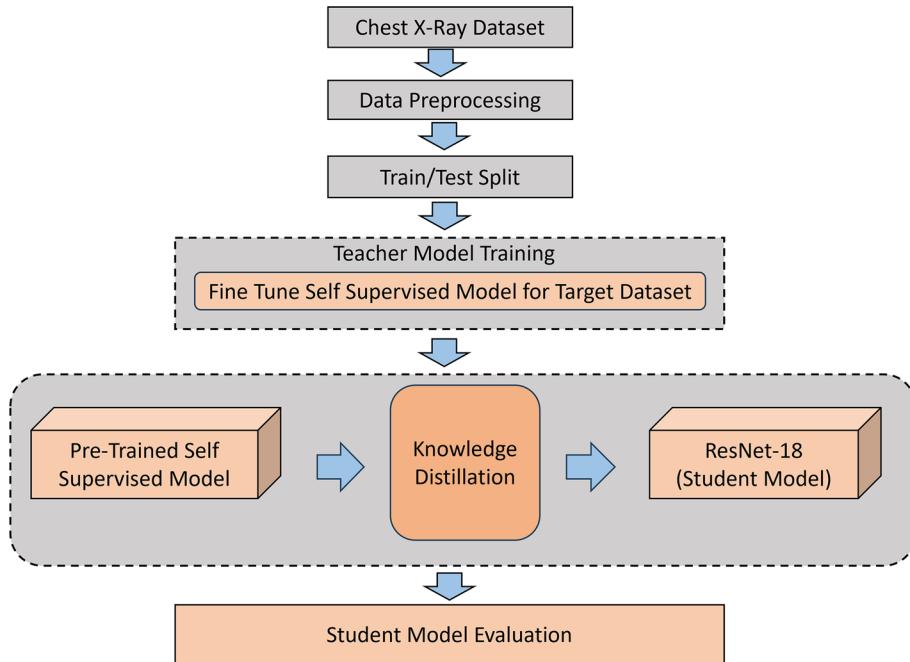


Fig. 2 The overall idea of the proposed algorithm

loss and Kullback–Leibler (KL) divergence to balance the influence of soft and hard targets.

- Evaluation: The performance of the student model is evaluated using standard metrics such as accuracy, precision, recall, and F1-score. Additionally, the number of parameters and GFLOPs are reported to demonstrate the model's suitability for resource-constrained environments.

3.1 Dataset and their description

The Chest X-ray Pneumonia dataset [25] is originally structured into three predefined directories: train, test, and val, each containing subfolders for the two image classes—Pneumonia and Normal—as illustrated in Fig. 3. However, to maintain consistency and control across our experimental pipeline, particularly during the fine-tuning of the teacher model and training of the student model, we reorganized the dataset. All 5,863 images were combined and then randomly split into training and validation sets using a 70:30 ratio. This custom split ensured a balanced distribution of classes across both subsets and provided a uniform setup for model training and validation under identical conditions.

The chest X-ray images in this dataset were collected at the Guangzhou Women's and Children's Medical Center, from a historical cohort of pediatric patients aged between one and five years. As part of routine clinical protocols, each patient underwent a chest radiograph. Prior to analysis, all images underwent a quality assurance step to exclude any low-quality or unreadable scans. Furthermore, two board-certified medical professionals assessed each radiograph, and a third expert confirmed the evaluations to ensure the reliability and consistency of the image labels before using them for AI model training.

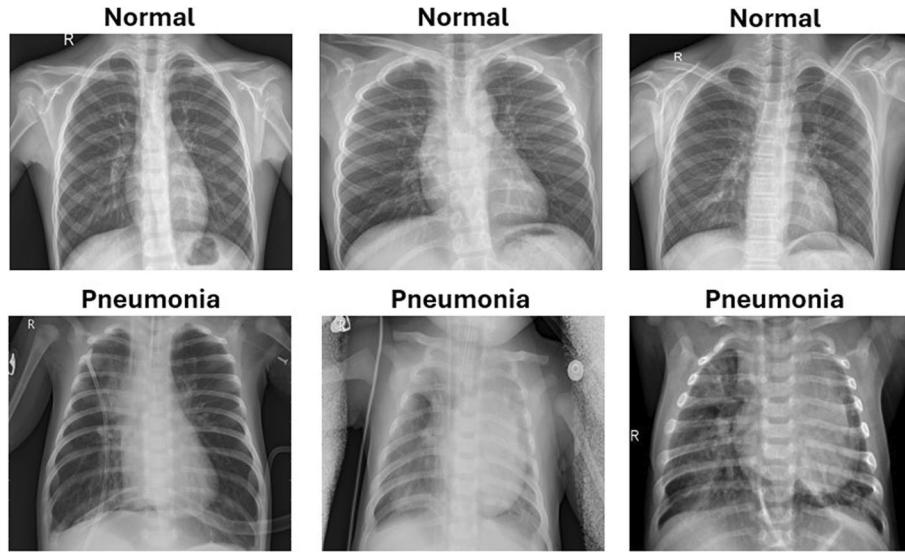


Fig. 3 Images of normal vs pneumonia

A collection of chest X-ray images is chosen, and training and testing validation groups are split up at random. The use of data augmentation is not part of the data transformation. After training on a dataset of medical images, weight for the pretrained ResNet50 model shall be used. In order to increase network performance the ResNet50 model was improved by using pretrained weights. The training and evaluation of these ResNet50 models is then carried out. Lastly, in the proposed model, images of chest x-ray are taken using a binary classification to separate pneumonia from healthy results.

3.2 Knowledge distillation and loss functions

In the knowledge distillation framework, the total loss is a weighted combination of two components: the soft target loss (L_{Dist}) and the hard target loss (L_{CE}). The hard target loss corresponds to the conventional cross-entropy loss and is mathematically expressed as:

$$L_{CE} = - \sum_{x \in D} \sum_{i=1}^C y^i \log(p_s^i(x; \tau)), \quad (1)$$

where y^i denotes the ground truth label for class i , and $\tau = 1$ indicates the use of standard cross-entropy. The soft target loss, based on Kullback–Leibler divergence, is defined as:

$$L_{Dist} = -\tau^2 \sum_{x \in D} \sum_{i=1}^C p_T^i(x; \tau) \log(p_S^i(x; \tau)), \quad (2)$$

where τ is the temperature parameter used for smoothing the output distributions, and the subscripts T and S refer to the teacher and student models, respectively. Here, D is the training dataset, C is the number of classes, and x denotes a given input sample. The softened probability distribution $p^i(x; \tau)$ is computed as:

$$p^i(x; \tau) = \frac{e^{S_i(x)/\tau}}{\sum_k e^{S_k(x)/\tau}}, \quad (3)$$

where $S_i(x)$ represents the logit (raw score) for class i . The final combined loss for knowledge distillation (L_{AT_KD}) is calculated using:

$$L_{KD} = \alpha L_{Dist} + (1 - \alpha)L_{CE}, \quad (4)$$

where α is an automatically tuned coefficient that balances the contribution of each loss component. In our approach, this loss formulation is employed to transfer knowledge from the SWAV-based teacher model to a lightweight ResNet-18 student model.

3.3 Proposed architecture

The proposed method consists of three main stages: selecting a self-supervised pre-trained model (SWAV), fine-tuning the SWAV model on the target chest X-ray dataset, and performing knowledge distillation from the fine-tuned SWAV model to a lightweight ResNet-18 student model. The architecture of the SWAV model (with ResNet-50 as the backbone) serving as the teacher model, and the ResNet-18 student model, is illustrated in Fig. 4.

We select the SWAV model as a teacher model inspired from our previous study [20–22] due its better performance compared to other SSL models like SimClr, MOCO, CPC, and other. The SWAV (Swapping Assignments between Views) model [24], which uses ResNet-50 as its backbone, is based on the core idea as SimCLR, learning representations through different views of the same image, but it introduces a key improvement: a scalable online clustering loss instead of contrastive loss. In SWAV, the process begins by extracting features from multiple augmented views of an image. These features are then grouped into clusters, with each cluster represented by a prototype or centroid. The model then swaps the cluster assignments between different views of the same image—this means that the representation from one view is matched to the cluster assignments of another. This swapping helps the model learn to associate similar features across different augmentations. The online clustering loss is calculated to bring features from the

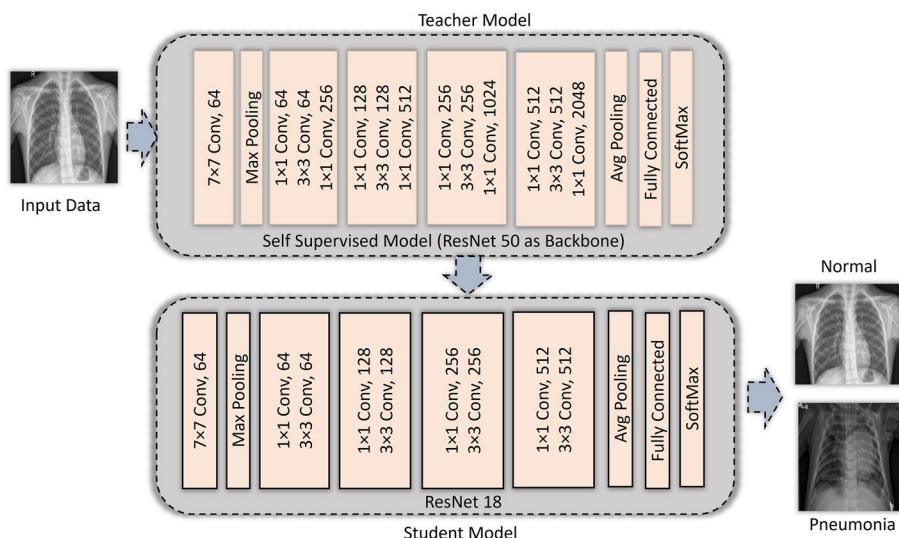


Fig. 4 Architecture of teacher model (ResNet50)–student model (ResNet18)

same cluster closer together and push features from different clusters further apart in the representation space. Unlike contrastive learning, this loss doesn't require comparing every pair of images, making it more efficient-especially with small or large batch sizes. Additionally, SWAV uses a multi-crop strategy, where multiple small and large crops (views) of each image are generated during training. This increases the diversity of views, helping the model learn more robust and generalizable features, as demonstrated in the results presented in [24].

Pre-existing weights from the SWAV model, which uses ResNet-50 [57] as its backbone, which is a well-known model that excels in a variety of computer vision applications. Among the others are GoogLeNet , MobileNet , DenseNet , Inception v3 , AlexNet , and VGG . Numerous datasets including a range of picture classifications are used to train these algorithms. Training Objective Function: A normal cross-entropy loss between the actual labels and the predictions made by the student model is usually included in the training goal, in addition to an additional loss term that computes the variance between the predictions made by the student and teacher models.

During knowledge distillation, the student model (ResNet-18) is trained with both the original chest X-ray images and the intermediate feature maps and representations learned by the teacher model. This can be accomplished by employing techniques like feature distillation, attention transfer, or using the outputs of the teacher model as soft objectives for the student model.

Data from the ground truth labels for the chest and the teacher's model X-ray images are used in the student model's training phase to reduce the loss function. This is to make sure that in order for the student model to successfully recognise images, it duplicates the teacher model's decision-making process.

After being trained, the small and effective student model (ResNet-18) may be used for real-world chest X-ray diagnosis applications. It may operate on devices with limited resources or in situations where real-time performance is essential, including in clinical settings or mobile apps. By using the information gained by bigger, more complicated models, the knowledge distillation technique allows for the production of accurate and efficient models without compromising performance.

3.4 Evaluation metrics

The accuracy, precision, recall, and F1-score are commonly used evaluation measures for classification models. A thorough summary of the model's performance is given by the Confusion Matrix, as its name implies. False Positive (FP), True Negative (TN), True Positive (TP), and False Negative (FN) are its four primary constituents. This Table 1, provides a clear guide to understanding and calculating these commonly used evaluation

Table 1 Model evaluation metrics

Metric name	Equation
Accuracy	$\frac{TP+TN}{FP+FN+TP+TN}$
Precision	$\frac{TP}{TP+FP}$
Recall or sensitivity	$\frac{TP}{TP+FN}$
F1-Score	$2 \times \frac{precision \times recall}{precision + recall}$

measures, which are essential for evaluating and contrasting the calibre of binary classification models across a range of applications.

3.5 Environmental setup

Pytorch is the deep learning library, and the programming language utilised is Python 3.10. The processor is a 16-core 2.90 GHz Intel® Core™ i7-10700. x86_64 processor architecture with four CPU cores powers the GPU, a GeForce GTX 1060. 64-bit architecture is used throughout the system.

4 Results and analysis

Deep learning models have significantly advanced medical image analysis, particularly in pneumonia detection using chest X-ray images. However, deploying these models on resource-constrained devices remains challenging due to their high computational demands. Knowledge distillation offers a viable solution by transferring knowledge from a complex teacher model to a lightweight student model, achieving a balance between accuracy and efficiency. This section presents the results and analysis of our KD-based approach, focusing on hyperparameter selection, model complexity, classification performance, and comparisons with existing studies.

The hyperparameters used in our experiments are presented in Table 2. The models were trained with an image size of 224×224 using the Stochastic Gradient Descent optimizer with a momentum of 0.9. The training process consisted of two phases: fine-tuning the teacher model for 30 epochs and training the student model using KD for 25 epochs. The learning rate was set to 0.001, and a dropout rate of 0.6 was applied. The KD process utilized a temperature value (τ) of 10 and a balancing factor (α) of 0.7, ensuring effective knowledge transfer from the teacher to the student model. The selection of the temperature (τ) and weighting factor (α) is guided by our previous study [23], where these hyperparameters were autotuned using Bayesian optimization techniques.

Table 3 presents a comparative analysis of the performance, computational complexity, and memory requirements between the teacher model (SWAV with ResNet-50 backbone) and the student model (ResNet-18) on the Chest X-ray dataset. The teacher model achieves a slightly higher accuracy (97.52%) than the student model (97.34%), but at the cost of significantly greater resource usage. The ResNet-50-based teacher contains 23.5 million parameters, while the ResNet-18 student model reduces this to 11.1 million, resulting in a 51.06% reduction. Similarly, GFLOPs are reduced from 4.11 to 1.81, marking a 55.9% decrease in computational operations.

Table 2 Experimental hyperparameters

Hyperparameters	Name/value
Optimizer	Stochastic gradient descent with momentum (0.9)
Batch size	64
Training epochs (fine tuning teacher model for target datasets)	30
Training epochs (during KD process)	25
Learning rate	0.001
Dropout	0.6
Temperature (τ)	10
Balancing factor (α)	0.7

Table 3 Performance and % decrease of parameters, GFLOPs, and size for minimal student model compared to teacher model on chest X-ray dataset

	Teacher model SWAV (ResNet-50 as backbone)	Student mdel ResNet-18
Accuracy	97.52	97.34
Parameters in millions (M)	23.5	11.1
% decrease (parameters)		51.06
GFLOPs	4.11	1.81
% decrease (GFLOPs)		55.9
Input size (MB)	0.57	0.57
Forward/backward pass size (MB)	286.55	62.79
% decrease in size		78.08
Params size (MB)	89.69	42.64
% decrease in params size		52.4
Estimated total size (MB)	376.82	106.00
% decrease in total size		71.8

Table 4 Comparison of classification performance of various existing methods on Chest X-ray or Pneumonia datasets

S.No.	References	Year	Dataset	Methods/models	Accu-racy (%)	Preci-sion (%)	Re-call (%)	F-1 Score (%)
1	Kermany et al. [25]	2018	Chest	Transfer learning	92.8	–	–	–
2	Stephen et al. [58]	2019	X-ray	CNN	93.73	–	–	–
3	Toungacar et al. [59]	2019	pneumo-nia	AlexNet	95.80	–	–	–
4	Gulgul et al. [60]	2020	[25]	CNN	93.4	–	–	–
5	Zubair et al. [61]	2020		Transfer learning	96.6	–	–	–
6	Liang et al. [62]	2020		Transfer learning	90.50	–	–	–
7	Rahimzadeh et al. [63]	2020		ResNet50 + Xception	91.40	–	–	–
8	Masud et al. [10]	2021		Ensemble	86.30	–	–	–
9	GM et al. [11]	2021		CNN	90.78	–	–	–
10	Demir et al. [64]	2021		CNN	96.31	–	–	–
11	Ieracitano et al. [65]	2022		CNN	81.00	–	–	–
12	Mobrouk et al. [66]	2022		Ensemble	93.91	93.96	92.99	93.43
13	Wang et al. [67]	2022		DenseNet + attention	92.80	92.60	96.20	94.30
14	Sharma et al. [12]	2023		VGG16	92.15	–	–	–
15	Demir et al. [37]	2023		Modified AlexNet	96.31	–	–	–
16	Bhatt et al. [68]	2023		Ensemble	84.12	80.04	99.23	88.56
17	Goyal et al. [69]	2023		Ensemble	94.31	88.89	95.41	92.03
18	An et al. [13]	2024		CNN with attention	95.19	98.38	93.84	96.06
19	Ali et al. [13]	2024		EfficientNetV2L	94.02	–	–	–
20	Kabir et al. [56]	2024		Knowledge distillation	96.08	–	–	–
21	Ours	2025		Knowledge distillation	97.34	97.35	97.24	97.29

In terms of memory consumption, the total estimated size drops from 376.82 MB in the teacher model to 106.00 MB in the student model, demonstrating a 71.8% overall reduction. This includes an 78.08% decrease in forward/backward pass memory and a 52.46% decrease in parameter size. These results confirm that the student model offers a highly efficient alternative for real-time and edge deployment, with minimal compromise in classification performance.

Table 4 presents the classification performance of various models applied to Chest X-ray or Pneumonia dataset. The comparison includes methods based on transfer learning, CNN variants, ensembles, and knowledge distillation. While several approaches

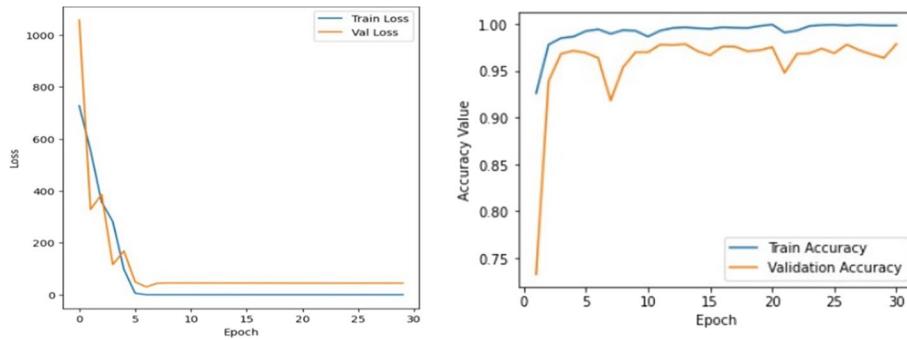


Fig. 5 The graphs of training and validation sets on loss and accuracy are displayed for the teachers model

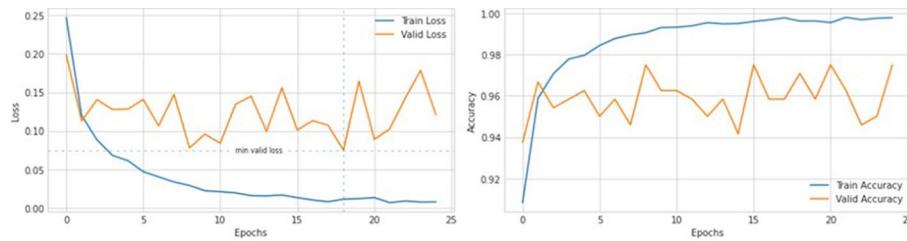


Fig. 6 The graphs of training and validation sets on loss and accuracy are displayed for the students model

have achieved high accuracy, such as EfficientNetV2L (96.08%) and CNN with attention mechanisms (95.19%), our proposed method achieves the highest accuracy of 97.34%, along with strong precision (97.35%), recall (97.24%), and F1-score (97.29%). This demonstrates the effectiveness of our knowledge distillation framework using a self-supervised SWAV teacher and a lightweight ResNet-18 student model.

Figures 5 and 6 illustrate the training and validation loss and accuracy curves for the teacher and student models, respectively. These plots provide insights into model convergence and generalization. The teacher model (ResNet-50) exhibits smooth convergence with minimal overfitting, indicating effective learning. The student model (ResNet-18) demonstrates consistent improvement, with validation accuracy closely following training accuracy, suggesting successful knowledge transfer.

The confusion matrix shown in Fig. 7 represents the performance of a binary classification model trained to distinguish between Normal and Pneumonia cases in a dataset of X-ray images. The matrix provides a breakdown of the model's predictions compared to the actual labels, highlighting the number of true positives (correctly predicted Pneumonia cases), true negatives (correctly predicted Normal cases), false positives (Normal cases misclassified as Pneumonia), and false negatives (Pneumonia cases misclassified as Normal). With an overall accuracy of 97.34%, the matrix offers insights into the model's strengths and potential weaknesses, such as the tendency to misclassify certain cases.

The experimental results highlight the effectiveness of knowledge distillation in transferring knowledge from a complex teacher model (SWAV with ResNet-50 as backbone) to a lightweight student model (ResNet-18). The proposed approach achieves superior classification accuracy while significantly reducing computational complexity, making it suitable for real-time medical image analysis applications. Future research will explore additional optimizations, including quantization and pruning, to further enhance the efficiency of the student model without compromising performance.

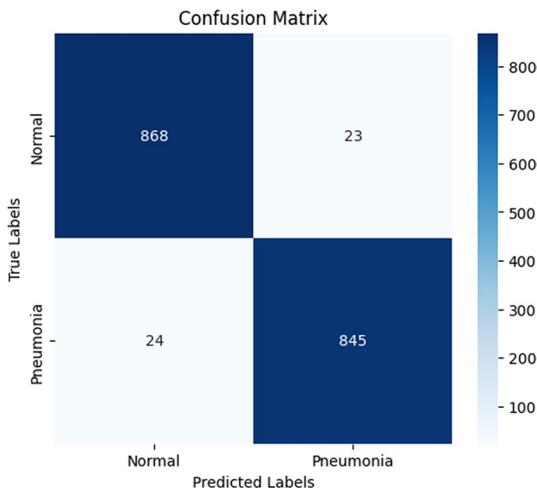


Fig. 7 Confusion matrix of the student model (ResNet-18) evaluated on the Chest X-ray Pneumonia test dataset. The model correctly classified 868 normal and 845 pneumonia cases, with only 23 and 24 misclassifications, respectively, demonstrating strong discriminative performance across both classes

5 Conclusion and future work

In this work, we proposed an efficient knowledge distillation framework for chest X-ray image classification. A self-supervised SWAV model with a ResNet-50 backbone was fine-tuned on pneumonia dataset and employed as the teacher model. Knowledge was then distilled to a compact ResNet-18 student model using a hybrid loss function combining cross-entropy and KL divergence. Our approach achieved a high classification accuracy of 97.34%, with precision, recall, and F1-score values of 97.35%, 97.24%, and 97.29%, respectively. Notably, the proposed student model reduced the number of parameters by 51.06%, GFLOPs by 55.9%, and the overall memory footprint by 71.8%, demonstrating its suitability for deployment in real-time and resource-constrained environments.

The results confirm that combining self-supervised learning with knowledge distillation can effectively retain high-level semantic features while significantly reducing computational costs. The comparative analysis with existing methods further validates the strength of our approach in balancing accuracy and efficiency.

Furthermore, although our experiments focus on chest X-ray pneumonia classification, the proposed framework is generalizable to other medical imaging tasks. By retraining the self-supervised teacher model and the student model on different datasets, this approach can be adapted for applications such as CT scan analysis, MRI interpretation, and histopathology image classification. In future work, we plan to extend and validate our framework across multiple imaging modalities to further demonstrate its versatility.

Acknowledgements

Not applicable.

Author contributions

A.J and K. K.K were involved in experimentation, coding, and writing the initial draft. J.K, S. K and P. K. M, S. S were involved in ideation, Analysis, supervision, and preparing the final draft.

Funding

Open access funding provided by Manipal University Jaipur.

Data availability

The datasets analysed during the current study are available in the Kaggle repository: <https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia/data>.

Code availability

The entire code for this study will be released at [github](#) after acceptance.

Declarations**Ethics approval and consent to participate**

The authors have no competing interest to disclose. Further, the authors certify that the research presented in this article does not involve any human participants or animals.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests

Received: 17 February 2025 / Accepted: 30 May 2025

Published online: 13 June 2025

References

1. Çalli E, Sogancioglu E, van Ginneken B, van Leeuwen KG, Murphy K. Deep learning for chest X-ray analysis: a survey. *Med Image Anal.* 2021;72: 102125.
2. Rehman A, Butt MA, Zaman M. A survey of medical image analysis using deep learning approaches. In: 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), 2021;1334–1342. IEEE.
3. Wang Y, Sun L, Jin Q. Enhanced diagnosis of pneumothorax with an improved real-time augmentation for imbalanced chest X-rays data based on dcnn. *IEEE/ACM Trans Comput Biol Bioinf.* 2019;18(3):951–62.
4. Zhou J, Jing B, Wang Z, Xin H, Tong H. Soda: detecting covid-19 in chest X-rays with semi-supervised open set domain adaptation. *IEEE/ACM Trans Comput Biol Bioinf.* 2021;19(5):2605–12.
5. Song Y, Zheng S, Li L, Zhang X, Zhang X, Huang Z, Chen J, Wang R, Zhao H, Chong Y, et al. Deep learning enables accurate diagnosis of novel coronavirus (covid-19) with ct images. *IEEE/ACM Trans Comput Biol Bioinf.* 2021;18(6):2775–80.
6. Pathak Y, Shukla PK, Arya K. Deep bidirectional classification model for covid-19 disease infected patients. *IEEE/ACM Trans Comput Biol Bioinf.* 2020;18(4):1234–41.
7. Zhang W, Zhou T, Lu Q, Wang X, Zhu C, Sun H, Wang Z, Lo SK, Wang F-Y. Dynamic-fusion-based federated learning for covid-19 detection. *IEEE Internet Things J.* 2021;8(21):15884–91.
8. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017;4700–4708.
9. Ferreira JR, Cardenas DAC, Moreno RA, de Sá Rebelo MDF, Krieger JE, Gutierrez MA. Multi-view ensemble convolutional neural network to improve classification of pneumonia in low contrast chest x-ray images. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2020;1238–1241. IEEE.
10. Masud M, Bairagi AK, Nahid A-A, Sikder N, Rubaiee S, Ahmed A, Anand D. A pneumonia diagnosis scheme based on hybrid features extracted from chest radiographs using an ensemble learning algorithm. *J Health Eng.* 2021;2021(1):8862089.
11. Gm H, Gourisaria MK, Rautaray SS, Pandey M. Pneumonia detection using cnn through chest x-ray. *J Eng Sci Technol (JESTEC).* 2021;16(1):861–76.
12. Sharma S, Guleria K. A deep learning based model for the detection of pneumonia from chest X-ray images using vgg-16 and neural networks. *Proc Comput Sci.* 2023;218:357–66.
13. An Q, Chen W, Shao W. A deep convolutional neural network for pneumonia detection in x-ray images with attention ensemble. *Diagnostics.* 2024;14(4):390.
14. Gou J, Yu B, Maybank SJ, Tao D. Knowledge distillation: a survey. *Int J Comput Vision.* 2021;129:1789–819.
15. Alkhulaifi A, Alsahlı F, Ahmad I. Knowledge distillation in deep learning and its applications. *PeerJ Comput Sci.* 2021;7:474.
16. Wang L, Yoon K-J. Knowledge distillation and student–teacher learning for visual intelligence: a review and new outlooks. *IEEE Trans Pattern Anal Mach Intell.* 2021.
17. Ho TKK, Gwak J. Utilizing knowledge distillation in deep learning for classification of chest x-ray abnormalities. *IEEE Access.* 2020;8:160749–61.
18. Van Sonsbeek T, Zhen X, Worring M, Shao L. Variational knowledge distillation for disease classification in chest x-rays. In: Information Processing in Medical Imaging: 27th International Conference, IPMI 2021, Virtual Event, June 28–June 30, 2021, Proceedings 27, 2021;334–345. Springer.
19. Chen B, Zhang Z, Li Y, Lu G, Zhang D. Multi-label chest x-ray image classification via semantic similarity graph embedding. *IEEE Trans Circuits Syst Video Technol.* 2021;32(4):2455–68.
20. Kishore J, Mukherjee S. Impact of autotuned fully connected layers on performance of self-supervised models for image classification. *Mach Intell Res.* 2024;1–13.
21. Kishore J, Mukherjee S. Auto cnn classifier based on knowledge transferred from self-supervised model. *Appl Intell.* 2023;53(19):22086–104.
22. Kishore J, Mukherjee S. Minimizing parameter overhead in self-supervised models for target task. *IEEE Trans Artif Intell.* 2023;5(4):1635–46.
23. Kishore J, Mukherjee S. An ensemble of self-supervised teachers for minimal student model with auto-tuned hyperparameters via improved bayesian optimization. *Progr Artif Intell.* 2024;13(3):201–15.
24. Caron M, Misra I, Mairal J, Goyal P, Bojanowski P, Joulin A. Unsupervised learning of visual features by contrasting cluster assignments. *Adv Neural Inf Process Syst.* 2020;33:9912–24.
25. Kermany DS, Goldbaum M, Cai W, Valentim CC, Liang H, Baxter SL, McKeown A, Yang G, Wu X, Yan F, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell.* 2018;172(5):1122–31.

26. Hayat M, Ahmad N, Nasir A, Tariq ZA. Hybrid deep learning efficientnetv2 and vision transformer (effnetv2-vit) model for breast cancer histopathological image classification. *IEEE Access* 2024.
27. Ren Z, Zhang Y, Wang S. A hybrid framework for lung cancer classification. *Electronics*. 2022;11(10):1614.
28. Hayat M. Squeeze & excitation joint with combined channel and spatial attention for pathology image super-resolution. *Franklin Open*. 2024;8: 100170.
29. Ren Z, Wang S, Zhang Y. Weakly supervised machine learning. *CAAI Trans Intell Technol*. 2023;8(3):549–80.
30. Ren Z, Lan Q, Zhang Y, Wang S. Exploring simple triplet representation learning. *Comput Struct Biotechnol J*. 2024;23:1510–21.
31. 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 preprint arXiv:1711.05225* 2017.
32. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. Chest x-ray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017;2097–2106.
33. Irvin J, Rajpurkar P, Ko M, Yu Y, Ciurea-Ilcus S, Chute C, Marklund H, Haghighi B, Ball R, Shpanskaya K, et al. Chexpert: a large chest radiograph dataset with uncertainty labels and expert comparison. *Proc AAAI Conf Artif Intell*. 2019;33:590–7.
34. Kundu R, Das R, Geem ZW, Han G-T, Sarkar R. Pneumonia detection in chest x-ray images using an ensemble of deep learning models. *PLoS ONE*. 2021;16(9):0256630.
35. Mujahid M, Rustam F, Álvarez R, Luis Vidal Mazón J, Díez IDIT, Ashraf I. Pneumonia classification from x-ray images with inception-v3 and convolutional neural network. *Diagnostics*. 2022;12(5):1280.
36. Sharma S, Guleria K. A deep learning model for early prediction of pneumonia using vgg19 and neural networks. In: Mobile Radio Communications and 5G Networks: Proceedings of Third MRCN 2022, pp. 597–612. Springer 2023.
37. Demir Y, Bingöl Ö. Detection of pneumonia from pediatric chest x-ray images by transfer learning. *Sigma J Eng Nat Sci*. 2023;41(6):1264–71.
38. Rifai AM, Harjaro S, Utami E, Ariyatmanto D. Analysis for diagnosis of pneumonia symptoms using chest x-ray based on mobilenetv2 models with image enhancement using white balance and contrast limited adaptive histogram equalization (clahe). *Biomed Signal Process Control*. 2024;90: 105857.
39. Asswin C, Ks DK, Dora A, Ravi V, Sowmya V, Gopalakrishnan E, Soman K, et al. Transfer learning approach for pediatric pneumonia diagnosis using channel attention deep cnn architectures. *Eng Appl Artif Intell*. 2023;123: 106416.
40. Ali M, Shahroz M, Akram U, Mushtaq MF, Altamiranda SC, Obregon SA, Díez IDLT, Ashraf I. Pneumonia detection using chest radiographs with novel efficientnetv2l model. *IEEE Access* 2024.
41. Gui K, Cheng J, Li K, Wang L, Lv Y, Cao D. Diagnosis and detection of pneumonia using weak-label based on x-ray images: a multi-center study. *BMC Med Imaging*. 2023;23(1):209.
42. Gui J, Chen T, Zhang J, Cao Q, Sun Z, Luo H, Tao D. A survey on self-supervised learning: algorithms, applications, and future trends. *IEEE Trans Pattern Anal Mach Intell*. 2024.
43. Liu X, Zhang F, Hou Z, Mian L, Wang Z, Zhang J, Tang J. Self-supervised learning: generative or contrastive. *IEEE Trans Knowl Data Eng*. 2021.
44. Zhang C, Zheng H, Gu Y. Dive into the details of self-supervised learning for medical image analysis. *Med Image Anal*. 2023;89: 102879.
45. Cho K, Kim KD, Nam Y, Jeong J, Kim J, Choi C, Lee S, Lee JS, Woo S, Hong G-S, et al. Chess: chest x-ray pre-trained model via self-supervised contrastive learning. *J Digit Imaging*. 2023;36(3):902–10.
46. Celniak W, Wodziński M, Jurgas A, Burti S, Zottl A, Atzori M, Müller H, Banzato T. Improving the classification of veterinary thoracic radiographs through inter-species and inter-pathology self-supervised pre-training of deep learning models. *Sci Rep*. 2023;13(1):19518.
47. Imagawa K, Shiomoto K. Evaluation of effectiveness of self-supervised learning in chest x-ray imaging to reduce annotated images. *J Imag Inf Med*. 2024;37(4):1618–24.
48. Kim J, Chang S, Kwak N. Pqk: model compression via pruning, quantization, and knowledge distillation. *arXiv preprint arXiv:2106.14681* 2021.
49. Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531* 2015.
50. Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, Liang J. Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Trans Med Imaging*. 2016;35(5):1299–312.
51. Li R, Yang H, Yang C. Parallel multilevel restricted schwarz preconditioners for implicit simulation of subsurface flows with peng-robinson equation of state. *J Comput Phys*. 2020;422: 109745.
52. Zeng J, Xie P. Contrastive self-supervised learning for graph classification. *Proc AAAI Conf Artif Intell*. 2021;35:10824–32.
53. Chen B, Li J, Lu G, Yu H, Zhang D. Label co-occurrence learning with graph convolutional networks for multi-label chest x-ray image classification. *IEEE J Biomed Health Inform*. 2020;24(8):2292–302.
54. Termirthikun C, Umer A, Suwanwimolkul S, Xia F, Lee I. Explainable knowledge distillation for on-device chest x-ray classification. In: IEEE/ACM Transactions on Computational Biology and Bioinformatics 2023.
55. Asham MA, Al-Shargabi AA, Al-Sabri R, Meftah I. A lightweight deep learning model with knowledge distillation for pulmonary diseases detection in chest x-rays. *Multimedia Tools Appl*. 2024;1–29.
56. Kabir MM, Mridha M, Rahman A, Hamid MA, Monowar MM. Detection of covid-19, pneumonia, and tuberculosis from radiographs using ai-driven knowledge distillation. *Heliyon*. 2024;10(5).
57. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016;770–8.
58. Stephen O, Sain M, Maduh UJ, Jeong D-U. An efficient deep learning approach to pneumonia classification in healthcare. *J Healthcare Eng*. 2019;2019(1):4180949.
59. Toğacıar M, Ergen B, Sertkaya ME. Zatürre hastalığının derin öğrenme modeli ile tespiti. *Fırat Univ J Eng Sci*. 2019;31(1).
60. Gülgün OD, Erol H. Classification performance comparisons of deep learning models in pneumonia diagnosis using chest x-ray images. *Turkish J Eng*. 2020;4(3):129–41.
61. Shah U, Abd-Alrazeq A, Alam T, Househ M, Shah Z. An efficient method to predict pneumonia from chest x-rays using deep learning approach. In: The Importance of Health Informatics in Public Health During a Pandemic, pp. 457–460. IOS Press 2020.

62. Liang G, Zheng L. A transfer learning method with deep residual network for pediatric pneumonia diagnosis. *Comput Methods Progr Biomed.* 2020;187: 104964.
63. Rahimzadeh M, Attar A. A modified deep convolutional neural network for detecting covid-19 and pneumonia from chest x-ray images based on the concatenation of xception and resnet50v2. *Inf Med Unlocked.* 2020;19: 100360.
64. Demir Y, Bingöl Ö. Pneumonia detection from pediatric lung x-ray images with convolutional neural network method. In: 2021 29th Signal Processing and Communications Applications Conference (SIU), 2021;1–4. IEEE.
65. Ieracitano C, Mammone N, Versaci M, Varone G, Ali A-R, Armentano A, Calabrese G, Ferrarelli A, Turano L, Tebala C, et al. A fuzzy-enhanced deep learning approach for early detection of covid-19 pneumonia from portable chest x-ray images. *Neurocomputing.* 2022;481:202–15.
66. Mabrouk A, Diaz Redondo RP, Dahou A, Abd Elaziz M, Kayed M. Pneumonia detection on chest x-ray images using ensemble of deep convolutional neural networks. *Appl Sci.* 2022;12(13):6448.
67. Wang K, Jiang P, Meng J, Jiang X. Attention-based densenet for pneumonia classification. *Irbm.* 2022;43(5):479–85.
68. Bhatt H, Shah M. A convolutional neural network ensemble model for pneumonia detection using chest x-ray images. *Healthcare Anal.* 2023;3: 100176.
69. Goyal S, Singh R. Detection and classification of lung diseases for pneumonia and covid-19 using machine and deep learning techniques. *J Ambient Intell Humaniz Comput.* 2023;14(4):3239–59.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.