AI Medicine
https://www.sciltp.com/journals/aim

Article

LTS-Net: A Lightweight and Efficient Deep Learning Framework for Automated Tuberculosis and Pneumonia Detection in Chest X-ray Images

Si-Yuan Lu [1,†], Ziquan Zhu [2,†], Baihua Zhang [3], Yu-Dong Zhang [2,4] and Yu-Dong Yao [5,*]

[1] School of Communications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China
[2] School of Computing and Mathematical Sciences, University of Leicester, Leicester LE1 7RH, UK
[3] Research Institute for Medical and Biological Engineering, Ningbo University, Ningbo 315211, China
[4] School of Computer Science and Engineering, Southeast University, Nanjing 210096, China
[5] Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030, USA
* Correspondence: yu-dong.yao@stevens.edu
† These authors contributed equally to this work.

How To Cite: Lu, S.-Y.; Zhu, Z.; Zhang, B.; et al. LTS-Net: A Lightweight and Efficient Deep Learning Framework for Automated Tuberculosis and Pneumonia Detection in Chest X-ray Images. *AI Medicine* **2026**, *3*(1), 3. https://doi.org/10.53941/aim.2026.100003

Received: 3 March 2026
Revised: 30 May 2026
Accepted: 3 June 2026
Published: 9 June 2026

Abstract: Chest X-rays (CXRs) are widely utilized for screening pulmonary diseases. However, distinguishing between tuberculosis (TB) and pneumonia remains a challenging task, as their radiological features often exhibit significant overlap. This challenge is particularly acute in resource-constrained settings, where rapid and low-cost automated analysis techniques would be especially beneficial. To address this issue, we propose a lightweight deep learning framework with Triple Permuter and Split Multi-Head Self-Attention (LTS-Net), specifically designed for the classification of TB and pneumonia based on CXR images. The model integrates three core components: a multi-level knowledge distillation strategy aimed at transferring useful knowledge from a large teacher model to a compact encoder; a Triple Permuter (TP) module, which enhances spatial feature interactions without significantly increasing the parameter count; and a Split Multi-Head Self-Attention (SMHSA) module, which effectively reduces the computational complexity of the attention mechanism while preserving global contextual information. We evaluated the proposed method on four publicly available datasets: PX, K-4, TB, and SC. Experimental results demonstrate that LTS-Net achieves classification accuracies ranging from 96.47% to 99.95% across these datasets, while excelling at maintaining relatively low model size and computational cost. Collectively, these findings indicate that LTS-Net strikes a practical balance between diagnostic performance and operational efficiency, positioning it as a highly promising technical solution for automated chest X-ray (CXR) analysis in clinical environments with limited computational resources.

Keywords: tuberculosis; pneumonia; chest X-ray; self-attention; computer-aided diagnosis

1. Introduction

Tuberculosis (TB) and pneumonia remain among the most severe respiratory infectious diseases globally. They account for a substantial burden of morbidity and mortality, placing immense strain on public health systems [1]. Tuberculosis is caused by Mycobacterium tuberculosis and primarily affects the lungs. The disease can exist in a latent state or progress to active disease, thereby perpetuating its continued transmission—particularly in

regions with limited healthcare resources. In contrast, pneumonia can be caused by bacterial, viral, or fungal infections [2]. It remains one of the leading causes of infection-related deaths, a trend that is particularly pronounced among young children and the elderly. Although tuberculosis and pneumonia have distinct etiologies, their radiological manifestations on chest X-rays (CXRs) often share certain similarities. Both conditions can present with pathological features such as infiltrates, nodular opacities, and cavitation [3]. This overlap in radiological characteristics makes it challenging to reliably differentiate between the two conditions in routine clinical practice.

The diagnosis of tuberculosis and pneumonia typically relies on a comprehensive approach involving clinical assessment, laboratory testing, and microbiological analysis. However, in real-world clinical settings, these diagnostic methods face numerous limitations: for instance, long turnaround times, low sensitivity in certain cases, and a lack of necessary medical infrastructure in resource-constrained regions. Chest X-ray examination, characterized by its low cost, rapid acquisition, and non-invasive nature, remains one of the most widely used tools for the initial assessment of pulmonary diseases [4]. Nevertheless, interpreting chest X-rays is not a trivial task. Lesions in the early stages of infection, atypical clinical presentations, and the presence of other pulmonary comorbidities can all introduce diagnostic uncertainty—even for experienced radiologists. This highlights an urgent need for the introduction of automated diagnostic methods to enhance the consistency and efficiency of radiological interpretation.

Computer-aided diagnosis (CAD) has emerged as a pivotal direction in advancing medical image analysis technologies [5]. Driven by the application of deep learning techniques, CAD systems have demonstrated immense potential in identifying disease-specific radiological patterns and assisting in the classification of chest images. In the diagnosis of tuberculosis and pneumonia, such models are capable of capturing subtle nuances—pertaining to lesion texture, spatial distribution, and other radiological imaging features—that are difficult for the human eye to quantify or discern. Convolutional Neural Networks (CNNs) have been widely applied in this domain, achieving excellent performance in disease recognition tasks based on chest X-rays [6]. However, CNN models rely heavily on local receptive fields, which may limit their ability to model the more macroscopic and extensive contextual relationships present across the entire image.

Vision Transformer (ViT) models offer a fundamentally different approach to representing visual information. These models leverage a self-attention mechanism to capture long-range dependencies within an image [7]. This proves highly beneficial for chest X-ray analysis, as diagnostically significant lesions are often diffusely distributed rather than confined to a specific, localized region. On the other hand, ViTs typically entail substantial computational overhead and often rely on large-scale pre-training, rendering them less practical for actual clinical deployment. In essence, superior global modeling capabilities often come at the cost of higher computational expense—a trade-off that is critically important for lightweight medical AI systems.

The Segment Anything Model (SAM) has recently garnered significant attention for its impressive zero-shot segmentation performance on natural images [8]. However, its direct application in the diagnosis of tuberculosis (TB) and pneumonia still faces numerous limitations. Lesions in chest X-ray images often exhibit blurred boundaries, overlapping morphologies, and weak local contrast, rendering the interpretation of pathological features significantly more challenging than general object segmentation tasks. Furthermore, applying SAM to high-resolution medical images typically incurs substantial computational overhead and proves difficult to seamlessly integrate into fully automated diagnostic workflows. Given these challenges, recent research has begun to explore a hybrid strategy: combining SAM-related modules with diagnostic backbone networks—such as CNNs or ViTs [9]. While these approaches point toward a promising research direction, striking a balance among diagnostic accuracy, computational efficiency, and practical clinical utility remains an unresolved challenge.

Based on these observations, we propose LTS-Net—a lightweight, SAM-based framework specifically designed to assist in the diagnosis of TB and pneumonia using chest X-ray images. The framework's architecture is designed around three core principles: first, the adoption of a multi-level knowledge distillation strategy aimed at transferring rich information from a large teacher model to a compact student encoder; second, the introduction of a Triple Permutator (TP) module to enhance interactions among spatial features while strictly controlling parameter count; and third, the development of a Splitting Multi-Head Self-Attention (SMHSA) module to reduce the computational burden of the attention mechanism without sacrificing global contextual modeling capabilities. Collectively, the synergistic interplay of these components enables LTS-Net to achieve a pragmatic balance between diagnostic accuracy and computational efficiency, thereby precisely meeting the rigorous demands of real-world chest X-ray detection applications where both metrics are of critical importance.

2. Related Work

Research on computer-aided diagnosis (CAD) for tuberculosis (TB) and pneumonia has steadily shifted from traditional machine learning pipelines toward deep learning-based models. Early studies typically combined hand-crafted image descriptors—designed to capture texture, shape, and intensity features—with classifiers such as Support Vector Machines or Random Forests to identify pulmonary abnormalities [10–12]. Although these methods laid the early groundwork for automated screening, they struggled to accommodate the immense visual variability exhibited by TB and pneumonia, particularly when these two diseases presented overlapping radiological imaging features [13].

The transition to deep learning—specifically to CNNs—introduced end-to-end feature learning into the domain of chest X-ray analysis [14]. CNN-based models have demonstrated robust capabilities in detecting imaging signs associated with TB (such as cavitary lesions and miliary nodules) as well as consolidation lesions associated with pneumonia [15]. In certain controlled evaluation experiments, their sensitivity and specificity have approached the performance levels of expert radiologists [4,16]. Subsequent studies further enhanced model performance by incorporating recurrent connections or attention mechanisms, enabling them to better capture spatial and contextual cues [17–19]. Nevertheless, CNNs inherently tend to focus on local receptive fields, which may limit their capacity to comprehend global contextual relationships across the entire image. In practical applications, this limitation becomes particularly critical when the task involves distinguishing multifocal TB lesions from diffuse pneumonia infiltrates [20].

Vision Transformers (ViTs) offer a fundamentally different architectural paradigm. By applying self-attention mechanisms across the entire image, ViTs are able to model long-range dependencies in a manner that CNN architectures cannot naturally achieve. This capability holds significant importance for the interpretation of chest X-rays, as diagnostic imaging signs in such scans are often not concentrated at a single focal point but are instead distributed across multiple distinct regions. Multiple studies have reported that in tasks requiring enhanced global image understanding—including the detection of disseminated tuberculosis lesions and the subtyping of pneumonia—ViTs consistently outperform CNNs [21]. Concurrently, however, the high computational demands of ViTs, their memory footprint—which scales quadratically with image resolution—and their reliance on large-scale pre-training continue to pose practical barriers to their deployment in resource-constrained clinical settings [22,23].

Recently, large-scale vision foundation models—most notably the SAM—have pioneered a new paradigm, enabling promptable and zero-shot image segmentation [8]. Nevertheless, directly applying these models to the field of medical imaging remains a formidable challenge. In the diagnostic contexts of tuberculosis and pneumonia, subtle radiological signs—such as ground-glass opacities, faint infiltrates, or early cavitation—often fail to align with the "object-centric" assumptions inherent in the SAM model. Recent advancements—including the SAMMed-2D model and fine-tuning strategies applied to curated chest X-ray (CXR) datasets [9,24]—have effectively enhanced model adaptability within this specific domain [25].

3. Materials

We evaluate the proposed LTS-Net on a total of four public chest X-ray datasets: the pneumonia CXR dataset (PX) [26], the Kaggle four-class CXR dataset (K-4) [27], the TB chest radiography dataset (TB) [28], and The Shenzhen CXR dataset (SC) [3].

The PX dataset originates from a pediatric study and comprises 5856 frontal (AP) chest X-rays, categorized into 1583 normal cases and 4273 pneumonia cases. These images were retrospectively collected from pediatric patients aged 1 to 5 at a hospital in Guangzhou. The resolution of individual images varies significantly, ranging approximately from 912 × 672 pixels to 2916 × 2583 pixels.

The K-4 dataset contains 7135 chest X-rays collected from multiple medical institutions across China, Qatar, and Bangladesh. This dataset covers four categories—pneumonia, tuberculosis (TB), COVID-19, and normal cases—and comes with pre-defined training, validation, and test subsets. The image resolutions within this dataset generally range from approximately 500 × 500 to 2000 × 2000 pixels.

The TB dataset consists of 4200 chest X-rays, comprising 3500 normal images and 700 images exhibiting signs of tuberculosis. This dataset was constructed by aggregating several image collections specifically focused on tuberculosis; all images have been standardized to a uniform resolution of 512 × 512 pixels.

The SC dataset was obtained through a collaboration between the Shenzhen Third People's Hospital and Guangdong Medical College. This dataset contains 662 frontal chest X-rays, comprising 326 normal cases and 336 tuberculosis cases; the image resolution is approximately 3000 × 3000 pixels.

For these four datasets, we adopted two distinct evaluation schemes. For datasets that provided original, pre-defined splits for training and testing, we directly utilized their existing partitioning methods; conversely, for

datasets lacking such pre-defined splits, we employed a 5-fold cross-validation approach. This hybrid strategy respects the original experimental designs inherent to each dataset while ensuring that evaluations across different data sources remain broadly comparable.

4. Methodology

Figure 1 gives an overview of the LTS-Net architecture. At its core lie three key modules: multi-level distillation, a Triple Permutator, and a Splitting Multi-Head Self-Attention block. The shared goal across these components is to strengthen feature representation without inflating computational overhead.

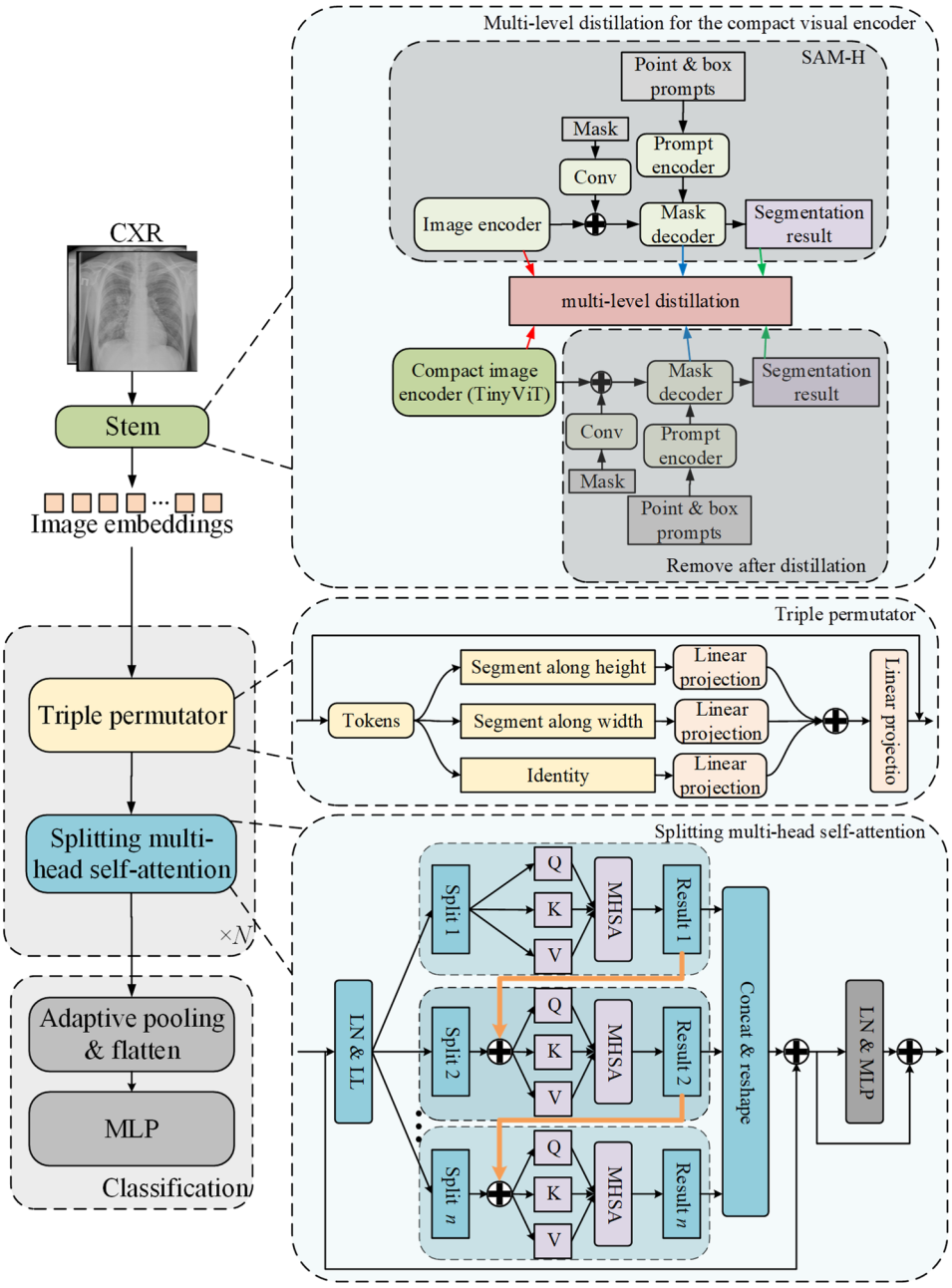


Figure 1. Structure of the proposed LTS-Net for CXR analysis.

4.1. Multi-Level Distillation for the Compact Visual Encoder

We use TinyViT-5M as the visual backbone of LTS-Net. To transfer knowledge from a larger vision model to this compact encoder, a multi-level distillation (MLD) framework is introduced. The teacher model is SAM-H, and the student model is trained with ImageNet data. Distillation is performed at three levels: image embedding level, mask decoder token level, and segmentation result level.

The first level aligns image embeddings produced by the teacher and student encoders. For an input CXR image, the embedding-level distillation loss is defined as

$$loss_{emb} = \left\| En_{Huge}(CXR) - En_{Tiny}(CXR) \right\|_1 \quad (1)$$

where En_{Huge} and En_{Tiny} denote the teacher and student image encoders, respectively, and $\|\cdot\|_1$ is the ℓ_1 loss. We use ℓ_1 loss instead of mean squared error because it provides a more stable alignment objective and is less sensitive to large deviations.

The second level is applied to the mask decoder. This stage aligns the token representations generated by the teacher and student transformers under the same prompt embedding. The token-level distillation loss is given by

$$loss_{token} = \left\| TF_{Huge}\left(En_{Huge}(CXR), query\right) - TF_{Tiny}\left(En_{Tiny}(CXR), query\right) \right\|_1 \quad (2)$$

where TF_{Huge} and TF_{Tiny} denote the transformer blocks in the teacher and student mask decoders, respectively, and $query$ is the prompt embedding. This term encourages the student decoder to match the intermediate token responses of the teacher.

The third level aligns the final segmentation outputs of the two models. The segmentation-level distillation loss is defined as

$$loss_{seg} = \left\| De_{Huge}\left(En_{Huge}(CXR), query\right) - De_{Tiny}\left(En_{Tiny}(CXR), query\right) \right\|_1 \quad (3)$$

where De_{Huge} and De_{Tiny} denote the teacher and student mask decoders, respectively. This loss constrains the student to produce segmentation results consistent with those of the teacher.

The overall distillation objective is a weighted combination of the three losses:

$$loss_{dst} = a \cdot loss_{emb} + b \cdot loss_{token} + c \cdot loss_{seg} \quad (4)$$

where a, b, and c are weighting coefficients. By combining supervision at the embedding, token, and output levels, the proposed distillation strategy transfers structural and semantic information from the teacher to the compact student encoder.

4.2. Triple Permutator

The Triple Permutator (TP) is used to model spatial interactions in a lightweight manner. Unlike standard convolutions or self-attention layers, TP processes the height, width, and channel dimensions separately, so that spatial structure can be preserved while interactions across different dimensions are still modeled.

Given a 3D token representation $I \in \mathbb{R}^{H \times W \times C}$, TP contains three parallel branches, as illustrated in Figure 1:

- Height Branch: Applies linear projections after permuting height and channel dimensions to capture vertical spatial interactions. Specifically, the input representation I is first divided into several chunks along the channel dimension. Then, each chunk is permutated in height and channel dimension, and concatenated along the channel dimension. A fully-connected layer is employed to processes the concatenated output to integrate height-specific spatial information across channels. Finally, the representation is permutated in height and channel dimension to recover the input shape.
- Width Branch: Similarly processes width and channel dimensions to model horizontal dependencies.
- Channel Branch: Directly projects channel features without spatial permutation.

After all the three permutations, the three outputs are added and fed into a fully-connected layer for fusion. A residual connection is used for effective training. In our model, the TP replaces traditional spatial convolutions or self-attention layers. Its lightweight design synergizes with the multi-level distillation framework. By preserving 2D spatial structure, the Permutator retains fine-grained positional cues essential for distinguishing overlapping pathologies.

4.3. Splitting Multi-Head Self-Attention

In a standard multi-head self-attention setup, every head operates on the full feature representation, which introduces a good deal of redundant computation across heads. To cut down on this redundancy, we propose a Splitting Multi-Head Self-Attention (SMHSA) module.

Rather than letting each head see the full embedding, SMHSA first divides the feature channels into *n* splits. Each attention head is then restricted to one split, so that different heads naturally work on different subsets of the channel space.

Information does not stay isolated within these splits. The output from one split is added to the input of the next split sequentially, creating a simple inter-split communication pathway. In this way, cross-split interaction is introduced, while the computation inside each head remains confined to a lower-dimensional subspace.

Relative to standard self-attention, SMHSA cuts both the parameter counts and the overall computational cost, simply because the input and output channels of each attention branch are reduced by a factor of n. At the same time, the sequential handoff between splits effectively increases the transformation depth of the block without adding any extra parameters.

Taken together, these design choices make SMHSA a more lightweight attention mechanism for feature interaction, one that still preserves the ability to model diverse information from the input representation.

5. Experimental Results

LTS-Net was implemented in PyTorch 2.3 (PyTorch Foundation, San Francisco, CA, USA), and all experiments were conducted on an NVIDIA 3090 graphic processing unit (GPU) (NVIDIA, Santa Clara, CA, USA). Model performance was evaluated using accuracy, sensitivity, specificity, precision, F1-score, receiver operating characteristic (ROC) analysis, and area under the curve (AUC). For the PX and K-4 datasets, the predefined training and testing splits were used. For the TB and SC datasets, five-fold cross-validation was applied.

5.1. Hyper-Parameter Settings

The hyperparameter configurations are summarized in Table 1. To leverage pre-trained weights from TinySAM's image encoder, we align the input size with its original training configuration (256 × 256 pixels) and set the embedding dimension of LTS-Net to 256 for feature compatibility. In the splitting multi-head self-attention module, we designate 8 splits—matching the number of attention heads—to maintain consistency with conventional multi-head self-attention frameworks, while stacking 3 blocks to balance computational depth and efficiency. The model is optimized using AdamW with a cosine-decayed learning rate schedule. Batch size is empirically adjusted based on GPU memory constraints, and training is capped at 50 epochs to prevent overfitting. The pre-trained weights of the large SAM-B model is from the ImageNet dataset.

Table 1. Hyper-parameters.

Parameter	Value
Input size	256 × 256
Embedding dimension	256
# of splits n	8
Block length N	3
Optimizer	AdamW
Learning rate scheduler	CosineLRScheduler
Learning rate range	2×10^{-4} to 1×10^{-5}
Batch size	72
Max epochs	50

5.2. Diagnostic Results of the Proposed LTS-Net on the Four Datasets

The evaluation results of the proposed LTS-Net on the four CXR datasets are presented in Tables 2–5, respectively.

When tested on the PX dataset, LTS-Net achieved an average accuracy of 96.47%, alongside a precision of 96.97% and a sensitivity of 95.56%. Notably, pneumonia cases were relatively easy to identify, yielding a high sensitivity of 99.23%; conversely, the reliability of identifying the "normal" category was slightly lower (with a sensitivity of 91.88%). This disparity suggests that, within this specific dataset, features associated with pneumonia are more distinctive than those found in normal radiographic images.

Turning to the multi-class K-4 dataset, the model achieved a perfect score of 100% across all reported metrics for the Tuberculosis (TB) category. Other categories, however, failed to reach this benchmark: the "normal" category recorded a sensitivity of 89.74%, while the "pneumonia" category yielded a specificity of 94.23%. These relatively lower figures indicate—as might be expected—that distinguishing between multiple disease categories is a more challenging task than performing a simple binary classification between "TB" and "normal" cases.

For LTS-Net, the Tuberculosis (TB) dataset proved to be the easiest to handle. The model achieved an average accuracy of 99.95%; during 5-fold cross-validation, the majority of metrics in four of the five folds reached the theoretical ceiling of 100%. Significantly, the folds in this cross-validation were strictly separated, ensuring that the test samples in each fold had never been "seen" by the model during the training phase. One potential explanation for this exceptional performance is that the dataset itself is highly standardized, and the visual distinctions between TB cases and normal cases are more pronounced than those found in other datasets. That said, given that the results are already approaching the theoretical upper limit of performance, a cautious approach

remains warranted; conducting additional testing on external cohorts would help validate whether the model's performance demonstrates robust generalization capabilities.

The SC dataset proved to be the most challenging. On this dataset, the model's average accuracy dropped to 90.94%, and its performance exhibited significant fluctuations during five-fold cross-validation—for instance, the sensitivity for the tuberculosis category varied substantially, ranging from 80.00% to 97.06%. This instability can likely be attributed to several factors. First, the dataset itself is relatively small in scale; it features high heterogeneity in both image appearance and disease manifestations, and may be particularly susceptible to inconsistencies in data annotation. Furthermore, although the original SC images possess high resolution, the resizing operations performed during the preprocessing stage may result in the loss of subtle lesion features—details that are often critical for distinguishing challenging cases. From a macro perspective, the overall trend is quite clear: LTS-Net demonstrates robust performance in tuberculosis-centric classification tasks, yet its reliability diminishes when confronted with datasets that are smaller in scale or exhibit higher heterogeneity. This method appears to achieve optimal efficacy in scenarios where inter-class visual distinctions are pronounced and training data is abundant. For highly challenging real-world clinical application scenarios, further validation remains necessary.

Table 2. Performance of LTS-Net on PX dataset (unit: %).

Class	Precision	Sensitivity	Specificity	F1 Score	Accuracy
Normal	98.62	91.88	99.23	95.13	96.47
Pneumonia	95.32	99.23	91.88	97.24	
Average	96.97	95.56	95.56	96.18	96.47

Table 3. Performance of LTS-Net on K-4 dataset (unit: %).

Class	Precision	Sensitivity	Specificity	F1 Score	Accuracy
COVID-19	97.22	99.06	99.55	98.13	
Normal	98.13	89.74	99.26	93.75	96.24
Pneumonia	94.61	98.97	94.23	96.74	
TB	100.00	100.00	100.00	100.00	
Average	97.32	96.58	98.68	96.81	96.24

Table 4. Performance of LTS-Net on TB dataset (unit: %).

Fold	Class	Precision	Sensitivity	Specificity	F1 Score	Accuracy
Fold-1	Normal	100.00	100.00	100.00	100.00	100.00
	TB	100.00	100.00	100.00	100.00	
Fold-2	Normal	99.86	100.00	99.31	99.93	99.88
	TB	100.00	99.31	100.00	99.65	
Fold-3	Normal	99.86	100.00	99.34	99.93	99.88
	TB	100.00	99.34	100.00	99.67	
Fold-4	Normal	100.00	100.00	100.00	100.00	100.00
	TB	100.00	100.00	100.00	100.00	
Fold-5	Normal	100.00	100.00	100.00	100.00	100.00
	TB	100.00	100.00	100.00	100.00	
Average		99.96	99.87	99.87	99.92	99.95

Table 5. Performance of LTS-Net on SC dataset (unit: %).

Fold	Class	Precision	Sensitivity	Specificity	F1 Score	Accuracy
Fold-1	Normal	91.43	91.43	90.48	91.43	90.98
	TB	90.48	90.48	91.43	90.48	
Fold-2	Normal	85.19	94.52	80.00	89.61	87.97
	TB	92.31	80.00	94.52	85.72	
Fold-3	Normal	94.92	91.80	95.77	93.33	93.94
	TB	93.15	95.77	91.80	94.44	
Fold-4	Normal	86.89	91.38	89.19	89.08	90.15
	TB	92.96	89.19	91.38	91.04	
Fold-5	Normal	96.49	85.94	97.06	90.91	91.67
	TB	88.00	97.06	85.94	92.31	
Average		91.18	90.76	90.76	90.84	90.94

5.3. Ablation Studies on the Four Datasets

The ablation results for all four datasets are summarized in Tables 6–9, respectively. For reference, we have also included direct comparative results against SAM-B—a baseline model of significantly larger scale.

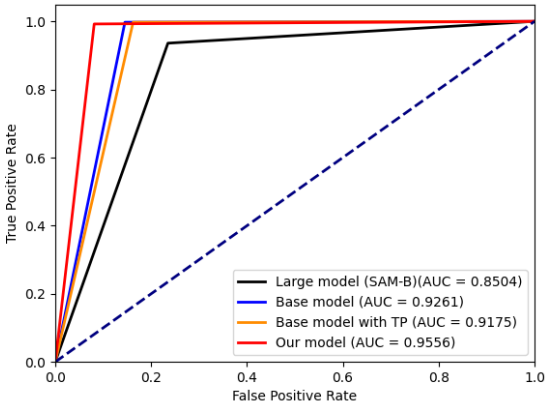
First, let us examine the PX dataset. On this dataset, SAM-B achieved an accuracy of 85.04%, whereas the full configuration of LTS-Net (i.e., incorporating the MLD, TP, and SMHSA modules simultaneously) boosted the accuracy to 96.47%. When the TP module was introduced in isolation, accuracy rose to 92.61%; the subsequent addition of the SMHSA module then brought the model's performance to its optimal level. The AUC metric exhibited a similar trend, rising from 0.8504 for SAM-B to 0.9556 for the full version of LTS-Net. In other words, the primary contribution to this performance improvement stems from the synergistic interplay between the TP and SMHSA modules.

Experimental results in the multi-class K-4 scenario further corroborate this conclusion. In this setting, SAM-B achieved an accuracy of 92.61%, while LTS-Net—equipped with both the TP and SMHSA modules—reached 96.24%. Precision and specificity metrics also benefited accordingly, with improvements in the Pneumonia and Tuberculosis (TB) categories being particularly pronounced. This indicates that, within this more complex multi-class task, the introduced auxiliary modules indeed played a constructive role.

On the TB dataset, although SAM-B had already established a robust baseline of 96.76%, LTS-Net successfully pushed this figure even higher, reaching 99.95%. Following the integration of the TP and SMHSA modules, precision, sensitivity, and specificity all saw improvements; concurrently, the AUC metric surged from 0.9154 to 0.9986—marking a substantial performance leap atop an already highly competitive baseline. Among all datasets, the SC dataset presents the highest overall difficulty; on this dataset, the accuracy of SAM-B stands at a mere 79.15%. However, the full LTS-Net configuration boosts this figure to 90.94%, while simultaneously achieving corresponding improvements across precision, sensitivity, specificity, and AUC metrics (specifically, the AUC rose from 0.7928 to 0.9094). Even when confronted with this highly heterogeneous dataset, the modules proposed in this paper maintained their effectiveness. In summary, the results of the ablation studies indicate that TP and SMHSA are the primary drivers behind the observed performance improvements. A plausible explanation is that, compared to the more massive SAM-B model, LTS-Net's compact architecture is more amenable to optimization on these relatively smaller Chest X-ray (CXR) datasets. Looking ahead, we plan to conduct a deeper investigation into the performance of SMHSA under more stringent resource constraints, with a particular focus on enhancing sensitivity for the "Normal" class—a category that consistently lagged behind the others in our experiments.

Table 6. Ablation study on PX dataset (unit: %).

MLD	TP	SMHSA	Precision	Sensitivity	Specificity	F1 Score	Accuracy
×	×	×	87.33	85.04	85.04	85.93	87.18
√	×	×	95.37	92.61	92.61	93.82	94.39
√	√	×	95.30	91.75	91.75	93.09	93.75
√	√	√	96.97	95.56	95.56	96.18	96.47



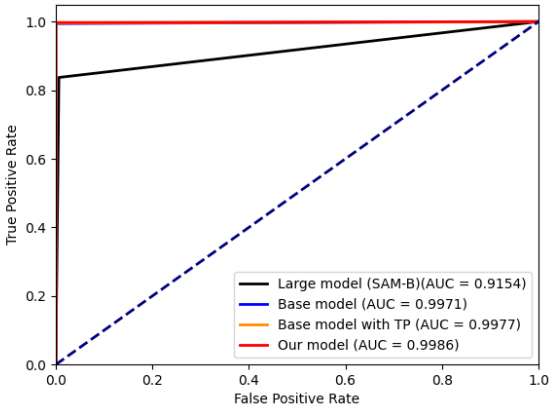
ROC curve

Table 7. Ablation study on K-4 dataset (unit: %).

MLD	TP	SMHSA	Precision	Sensitivity	Specificity	F1 Score	Accuracy
×	×	×	92.90	93.91	96.47	93.18	92.09
√	×	×	95.70	94.79	97.16	95.00	94.03
√	√	×	95.99	95.34	99.38	97.45	94.55
√	√	√	97.32	96.58	98.68	96.81	96.24

Table 8. Ablation study on TB dataset (unit: %).

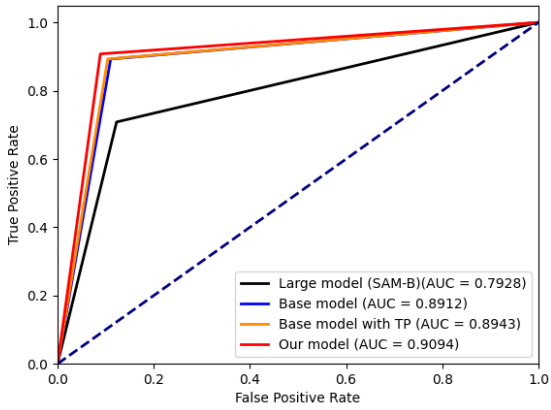
MLD	TP	SMHSA	Precision	Sensitivity	Specificity	F1 Score	Accuracy
×	×	×	96.45	91.34	91.34	95.51	96.76
√	×	×	99.94	99.72	99.72	99.83	99.91
√	√	×	99.88	99.77	99.77	99.83	99.91
√	√	√	99.96	99.87	99.87	99.92	99.95



ROC curve

Table 9. Ablation study on SC dataset (unit: %).

MLD	TP	SMHSA	Precision	Sensitivity	Specificity	F1 Score	Accuracy
×	×	×	80.36	79.47	79.47	78.95	79.15
√	×	×	89.07	89.28	89.28	89.09	89.13
√	√	×	89.48	89.45	89.45	89.36	89.43
√	√	√	91.18	90.76	90.76	90.84	90.94



ROC curve

5.4. Computational Costs

Table 10 details the computational performance profiles of LTS-Net under various configurations, using the SAM-B encoder as a baseline reference. This baseline model is substantial in scale: it comprises 94.56 M parameters, entails a computational cost (FLOPs) of 33.29 G, and achieves a throughput of 109.04 FPS.

The introduction of Multi-Level Distillation (MLD) alone significantly reduced these metrics—parameter count dropped to 11.86 M, FLOPs fell to 3.93 G, and inference speed surged to 528.94 FPS. Incorporating the TP module on top of this yielded further resource savings, resulting in the most streamlined configuration: a parameter count of just 9.10 M and 3.12 G FLOPs; coincidentally, this also proved to be the fastest configuration in terms of inference speed (598.15 FPS). Upon the subsequent integration of the SMHSA module, the complete LTS-Net model ultimately settled at 9.77 M parameters and 3.34 G FLOPs, operating at a speed of 587.23 FPS.

Compared to the baseline model, the complete LTS-Net model reduced the parameter count by approximately 89.7% and FLOPs by roughly 90.0%, while simultaneously boosting inference speed from around 109 FPS to over 587 FPS. In essence, the design proposed in this paper achieves a substantial reduction in computational overhead without sacrificing the high-speed inference capabilities essential for practical application deployment.

Table 10. Computational cost analysis.

MLD	TP	SMHSA	# of Parameters	FLOPs	FPS
×	×	×	94.56 M	33.29 G	109.04
√	×	×	11.86 M	3.93 G	528.94
√	√	×	9.10 M	3.12 G	598.15
√	√	√	9.77 M	3.34 G	587.23

5.5. Visual Explanation

To gain a more intuitive understanding of the specific features the model focuses on, we examined Gradient-weighted Class Activation Maps (Grad-CAM), as illustrated in Figure 2. A notable characteristic observed across multiple heatmaps is the presence of activation patterns exhibiting a distinct horizontal orientation.

A plausible explanation for this phenomenon lies in the Triple Permutator (TP) module. Since the TP module processes the height, width, and channel dimensions independently, it is capable of introducing directionally sensitive responses into the intermediate feature representations—responses that may subsequently manifest as horizontal bands in the final Grad-CAM visualizations.

Nevertheless, we exercise caution against overinterpreting these patterns. While Grad-CAM effectively highlights the regions that play a decisive role in the model's output, it does not directly reveal the specific underlying computational processes. The horizontal structures visible in Figure 2 may indeed reflect the influence of the TP module; however, they could also—at least to some extent—be artifacts of the visualization technique itself, or simply a consequence of inherent characteristics within the dataset images. Determining with certainty whether these activation patterns consistently correspond to clinically significant anatomical structures requires further, more targeted investigation.

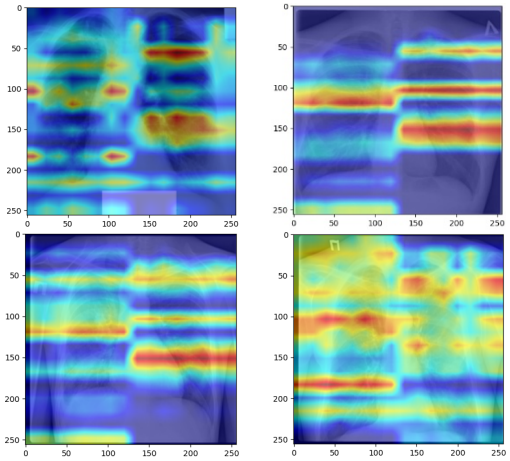


Figure 2. Grad-CAMs of the LTS-Net.

5.6. Comparison with State-of-the-Art Methods

Table 11 compares LTS-Net with several previously published methods that have been evaluated on the same four chest X-ray (CXR) datasets. To ensure completeness, we have listed the validation settings, dataset sizes, accuracy rates, and AUC values extracted from the original studies; however, it must be noted upfront that these figures were derived under varying experimental environments.

First, regarding the PX dataset, LTS-Net achieved an accuracy of 96.47% and an AUC of 95.56%. In comparison, VGG-16 yielded an accuracy of 84.5% (AUC: 87.0%), while B-CNN achieved an accuracy of 94.4% (AUC: 94.5%). Among the various methods compared on this specific dataset, LTS-Net's performance advantage is particularly pronounced.

On the K-4 dataset, our measured accuracy of 96.24% significantly surpassed the figures reported by Adaptive CNN (93.16%) and Transferred CNN (79.00%).

As for the TB dataset, our model achieved an accuracy of 99.95% and an AUC of 99.86%; when compared against the results listed for EN-CNN, GoogleNet, and TB-CNN, this performance gap appears particularly substantial. That said, conducting a direct and rigorous "one-to-one" comparison here remains somewhat challenging. Given the significant disparities in sample sizes and validation protocols across the studies being compared, these figures should be viewed primarily as a rough frame of reference rather than as precise comparative results derived under strictly controlled conditions.

In our experiments, the SC dataset served as the most challenging benchmark, on which LTS-Net achieved an accuracy of 90.94% and an AUC of 90.94%. By way of comparison, the accuracy rates for Deep CNN, TS-SVM, and TB-CNN were 83.7%, 82.5%, and 84.4%, respectively. Although LTS-Net once again topped the rankings, its relatively modest absolute performance on the SC dataset underscores the inherent difficulty of this particular dataset. In summary, the data presented in Table 11 indicates that, compared to the selected existing methods, LTS-Net not only demonstrates highly competitive performance but, in several instances, even achieves significantly superior results. However, we must explicitly clarify here precisely what conclusions can—and cannot—be drawn from these comparative results. Given the disparities in sample size, validation set partitioning strategies, and preprocessing schemes, conducting a strictly equivalent comparison is by no means an easy task. Consequently, the results presented in this paper should be regarded as informative guidelines rather than definitive conclusions derived from strictly controlled benchmarks. Conducting fair, side-by-side evaluations under completely identical training and testing conditions remains a key objective for future research.

Table 11. Comparison with state-of-the-art methods.

Dataset	Method	# Images	Validation	Accuracy	AUC
PX	VGG-16 [15]	5856	Hold-out	84.5	87.0
	B-CNN [20]	5840	Hold-out	94.4	94.5
	LTS-Net (ours)	**5856**	Hold-out	**96.47**	**95.56**
K-4	Adaptive CNN [27]	7135	Hold-out	93.16	-
	Transferred CNN [29]	7135	Hold-out	79.00	-
	LTS-Net (ours)	7135	Hold-out	**96.24**	-
TB	EN-CNN [16]	800	Hold-out	86.42	-
	GoogleNet [30]	800	Hold-out	94.90	-
	TB-CNN [4]	1104	Hold-out	86.2	92.5
	LTS-Net (ours)	**4200**	**5-fold**	**99.95**	**99.86**
SC	Deep CNN [31]	662	Hold-out	83.7	**92.6**
	TS-SVM [3]	615	Hold-out	82.5	88.0
	TB-CNN [4]	662	Hold-out	84.4	90.0
	LTS-Net (ours)	**662**	**5-fold**	**90.94**	90.94

6. Discussion

Taken together, these experiments point to a clear overarching conclusion: across the four publicly available chest X-ray datasets we tested, LTS-Net successfully strikes a practical balance between diagnostic accuracy and computational efficiency. By integrating three key techniques—multi-level distillation, the Triple Permutator (TP) module, and Split Multi-Head Self-Attention (SMHSA)—the model consistently outperforms lightweight baseline models while incurring a computational overhead significantly lower than that required by the original, large-scale model architectures. Specifically, the full LTS-Net model demonstrates exceptional performance in both binary and multi-class classification tasks; it achieves a substantial reduction in parameter count and Floating Point Operations (FLOPs) while maintaining an inference speed sufficient to meet the demands of real-time clinical workflows.

One particularly insightful finding concerns the pivotal role played by the multi-level distillation technique. This suggests that, in environments with limited computational resources, transferring the rich representational knowledge embedded within a large teacher model to a compact student model appears to be of greater importance than simply scaling up the model size. Data from ablation studies further corroborates this view: building upon an already high-performing distilled backbone network, the TP and SMHSA modules each contribute complementary performance enhancements—a result that aligns perfectly with our initial design philosophy.

While observing the Grad-CAM visualization results, we indeed identified several intriguing patterns; however, we are cautious about over-interpreting them. The activation regions exhibiting a horizontal distribution likely reflect the "direction-sensitivity" characteristic introduced by the TP module—a result consistent with the module's intended design. Yet, an alternative possibility remains: these patterns may, in reality, be amplified—or even partially created—by the Grad-CAM technique itself, or by inherent properties specific to the particular image datasets used. Consequently, for the time being, it is best to regard these heatmaps as qualitative reference cues rather than as definitive mechanistic proofs of strict causality. Establishing an unequivocal link between these activation patterns and clinically significant anatomical regions would require further rigorous and in-depth investigation.

As anticipated, the model's robustness varies across different datasets. Among them, the SC dataset stands out by a significant margin as the most challenging test case. The observed results likely stem from a confluence of factors: the dataset's limited size, its high intra-class heterogeneity, and its pronounced sensitivity to the specific manner in which cross-validation folds are partitioned. Furthermore, the necessity of uniformly resizing high-resolution SC images to a fixed input dimension—a standard preprocessing step—may lead to the loss of subtle

lesion details. Crucially, in those edge cases where the model struggles most, these discarded details often hold vital diagnostic significance. Consequently, the phenomena observed here may not arise from a single mode of failure, but rather represent the cumulative effect of multiple interacting factors: limited data volume, high variability in image features, and information loss during the preprocessing stage. Another noteworthy limitation is the model's exclusive reliance on imaging data. In actual clinical practice, diagnoses are rarely determined solely on the basis of radiological images; laboratory results, patient history, and other metadata often provide critical contextual information capable of resolving otherwise ambiguous image interpretations. Lacking access to such information, LTS-Net is naturally unable to perceive the imaging-clinical correlations that physicians routinely leverage in their daily practice. Moreover, while the metrics presented in Table 11 are encouraging, we must reiterate that cross-study comparisons in this field are frequently confounded by various variables—including dataset size, validation set partitioning strategies, preprocessing pipelines, and even category definitions—which rarely align perfectly across different studies. Therefore, these results should be regarded as providing only approximate reference value, rather than serving as definitive benchmarks derived from rigorous comparisons. To truly establish the relative performance merits of different models, controlled evaluations conducted under fully standardized training and testing conditions are indispensable.

Finally, the current framework remains constrained by its reliance on a pre-trained SAM teacher model. Although this knowledge distillation strategy proves effective in practice, it almost certainly introduces a degree of domain dependency. Should the distribution of downstream data diverge significantly from the training data upon which the teacher model was built, the quality of the transferred representations may subsequently degrade; without specific out-of-distribution testing, such performance degradation often goes undetected. In light of this, future research efforts could focus on exploring the following directions: conducting domain-specific pre-training based on large-scale medical imaging datasets; pursuing self-supervised representation learning that does not rely on a fixed teacher model; and integrating clinical metadata into the model's processing pipeline through multi-modal fusion techniques. Furthermore, if complemented by lightweight deployment techniques—such as quantization and structural pruning—the efficiency boundaries of these models in practical application scenarios will undoubtedly be further expanded.

7. Conclusions

In this study, we set out to design a lightweight framework—LTS-Net—for the automated diagnosis of tuberculosis and pneumonia based on chest X-rays (CXR). This framework is built around three complementary core concepts: Multi-level Distillation, the Triple Permutator, and Splitting Multi-Head Self-Attention. Our primary objective was to enhance the quality of feature representations while avoiding any increase in computational overhead.

Across four publicly available chest X-ray datasets, LTS-Net demonstrated consistent and competitive performance in both binary and multi-class classification tasks. Crucially, these improvements in accuracy were achieved without sacrificing efficiency. Compared to baseline models, the final model reduced the parameter count by approximately 89.7% and the floating-point operations (FLOPs) by 90.0%, while boosting inference speed by more than fivefold—soaring from approximately 109 FPS to over 587 FPS. These figures translate into a tangible practical advantage: while maintaining exceptional diagnostic performance, the model no longer requires the high-end hardware support that is often difficult to procure in resource-constrained environments.

However, a gap remains—one not yet fully bridged—between robust overall performance and real-world reliability. In scenarios involving smaller and highly heterogeneous datasets—such as the SC dataset—the model's robustness remains a subject of concern. Furthermore, as the current system relies solely on image data for diagnosis, it is currently unable to fully leverage the rich clinical contextual information that often plays a pivotal role in actual clinical decision-making. Looking ahead, we identify several directions of significant research value: enhancing the model's generalization capabilities when confronted with dataset variations; improving its diagnostic sensitivity regarding challenging or ambiguous cases; and evolving toward a multi-modal framework that more closely simulates and replicates actual clinical diagnostic workflows.

Ultimately, LTS-Net represents a solid stride forward in our quest to achieve efficient, automated chest X-ray analysis. It serves as compelling proof that exceptional performance and low computational costs are not mutually exclusive, while simultaneously charting a novel path worthy of deep exploration within the field of lightweight medical image analysis.

Author Contributions

S.-Y.L.: conceptualization, methodology; Z.Z.: writing—original draft preparation, visualization; B.Z.: methodology; Y.-D.Z.: supervision; Y.-D.Y.: supervision. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement

Ethical review and approval were waived for this study because it exclusively utilized de-identified, publicly available chest X-ray datasets. No identifiable human subject data were accessed or processed.

Informed Consent Statement

Patient informed consent was not required as all clinical images obtained from the public datasets were fully de-identified and anonymized prior to analysis, ensuring no re-identification risks.

Data Availability Statement

Datasets in this study are all public.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-Assisted Technologies

During the preparation of this work, the authors used ChatGPT to refine and polish the language. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

1. Ibrahim, D.M.; Elshennawy, N.M.; Sarhan, A.M. Deep-chest: Multi-classification deep learning model for diagnosing COVID-19, pneumonia, and lung cancer chest diseases. *Comput. Biol. Med.* **2021**, *132*, 104348. https://doi.org/10.1016/j.compbiomed.2021.104348.
2. Abdel-Basset, M.; Hawash, H.; Moustafa, N.; et al. Two-Stage Deep Learning Framework for Discrimination between COVID-19 and Community-Acquired Pneumonia from Chest CT scans. *Pattern Recognit. Lett.* **2021**, *152*, 311–319. https://doi.org/10.1016/j.patrec.2021.10.027.
3. Jaeger, S.; Karargyris, A.; Candemir, S.; et al. Automatic tuberculosis screening using chest radiographs. *IEEE Trans. Med. Imaging* **2014**, *33*, 233–245. https://doi.org/10.1109/TMI.2013.2284099.
4. Pasa, F.; Golkov, V.; Pfeiffer, F.; et al. Efficient Deep Network Architectures for Fast Chest X-ray Tuberculosis Screening and Visualization. *Sci. Rep.* **2019**, *9*, 6268. https://doi.org/10.1038/s41598-019-42557-4.
5. Lakhani, P.; Sundaram, B. Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. *Radiology* **2017**, *284*, 574–582. https://doi.org/10.1148/radiol.2017162326.
6. Ul Abideen, Z.; Ghafoor, M.; Munir, K.; et al. Uncertainty Assisted Robust Tuberculosis Identification with Bayesian Convolutional Neural Networks. *IEEE Access* **2020**, *8*, 22812–22825. https://doi.org/10.1109/ACCESS.2020.2970023.
7. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; et al. An Image Is Worth 16 × 16 Words: Transformers for Image Recognition at Scale. In Proceedings of the International Conference on Learning Representations (ICLR 2021), Vienna, Austria, 4 May 2021.
8. Kirillov, A.; Mintun, E.; Ravi, N.; et al. Segment Anything. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023.
9. Cheng, J.; Ye, J.; Deng, Z.; et al. SAM-Med2D. *arXiv* **2023**, 1–16, *preprint*, arXiv:2308.16184.
10. Afzali, A.; Mofrad, F.B.; Pouladian, M. Feature Selection for Contour-based Tuberculosis Detection from Chest X-ray Images. In Proceedings of the 2019 26th National and 4th International Iranian Conference on Biomedical Engineering (ICBME), Tehran, Iran, 27–28 November 2019.
11. Chandra, T.B.; Verma, K.; Singh, B.K.; et al. Automatic detection of tuberculosis related abnormalities in Chest X-ray images using hierarchical feature extraction scheme. *Expert Syst. Appl.* **2020**, *158*, 113514. https://doi.org/10.1016/j.eswa.2020.113514.

12. Govindarajan, S.; Swaminathan, R. Extreme Learning Machine based Differentiation of Pulmonary Tuberculosis in Chest Radiographs using Integrated Local Feature Descriptors. *Comput. Methods Programs Biomed.* **2021**, *204*, 106058. https://doi.org/10.1016/j.cmpb.2021.106058.
13. Ahmed, F.; Nuwagira, B.; Torlak, F.; et al. Topo-CXR: Chest X-ray TB and Pneumonia Screening with Topological Machine Learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023.
14. Lopes, U.K.; Valiati, J.F. Pre-trained convolutional neural networks as feature extractors for tuberculosis detection. *Comput. Biol. Med.* **2017**, *89*, 135–143. https://doi.org/10.1016/j.compbiomed.2017.08.001.
15. Ayan, E.; Unver, H.M. Diagnosis of pneumonia from chest X-ray images using deep learning. In Proceedings of the 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science, Istanbul, Turkey, 24–26 April 2019; Vol. 7, pp. 1–5.
16. Hernández, A.; Panizo, Á.; Camacho, D. An Ensemble Algorithm Based on Deep Learning for Tuberculosis Classification. In Proceedings of the Intelligent Data Engineering and Automated Learning—IDEAL 2019: 20th International Conference, Manchester, UK, 14–16 November 2019; Ch. 17, pp. 145–154.
17. Sadik, F.; Dastider, A.G.; Subah, M.R.; et al. A dual-stage deep convolutional neural network for automatic diagnosis of COVID-19 and pneumonia from chest CT images. *Comput. Biol. Med.* **2022**, *149*, 105806. https://doi.org/10.1016/j.compbiomed.2022.105806.
18. Celik, G. Detection of Covid-19 and other pneumonia cases from CT and X-ray chest images using deep learning based on feature reuse residual block and depthwise dilated convolutions neural network. *Appl. Soft Comput.* **2023**, *133*, 109906. https://doi.org/10.1016/j.asoc.2022.109906.
19. Khattab, R.; Abdelmaksoud, I.R.; Abdelrazek, S. Automated detection of COVID-19 and pneumonia diseases using data mining and transfer learning algorithms with focal loss from chest X-ray images. *Appl. Soft Comput.* **2024**, *162*, 111806. https://doi.org/10.1016/j.asoc.2024.111806.
20. Saraiva, A.A.; Santos, D.; Costa, N.; et al. Models of learning to classify X-ray images for the detection of pneumonia using neural networks. *Bioimaging* **2019**, *7*, 76–83.
21. Tan, Z.; Madzin, H.; Norafida, B.; et al. SwinUNeLCsT: Global–local spatial representation learning with hybrid CNN–transformer for efficient tuberculosis lung cavity weakly supervised semantic segmentation. *J. King Saud Univ. Comput. Inf. Sci.* **2024**, *36*, 102012. https://doi.org/10.1016/j.jksuci.2024.102012.
22. Liu, X.; Peng, H.; Zheng, N.; et al. EfficientViT: Memory Efficient Vision Transformer with Cascaded Group Attention. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 17–24 June 2023.
23. Hou, Q.; Jiang, Z.; Yuan, L.; et al. Vision permutator: A permutable MLP-like architecture for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 1328–1334.
24. Wang, F.; Zhou, Y.; Wang, S.; et al. Multi-Granularity Cross-modal Alignment for Generalized Medical Visual Representation Learning. In Proceedings of the 36th Conference on Neural Information Processing Systems (NeurIPS 2022), New Orleans, LA, USA, 28 November–9 December 2022.
25. Shu, H.; Li, W.; Tang, Y.; et al. TinySAM: Pushing the Envelope for Efficient Segment Anything Model. *arXiv* **2023**, *preprint,* arXiv:2312.13789.
26. Kermany, M.G.; Cai, W.; Valentim, C.C.S.; et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* **2018**, *172*, 1122–1131. https://doi.org/10.1016/j.cell.2018.02.010.
27. Kakani, V.; Varun, B.; Bodapati, J.D.; et al. Post-COVID Chest Disease Monitoring using self-adaptive Convolutional Neural Network. In Proceedings of the IEEE 8th International Conference for Convergence in Technology (I2CT), Pune, India, 7–9 April 2023.
28. Rahman, T.; Khandakar, A.; Kadir, M.A.; et al. Reliable Tuberculosis Detection using Chest X-ray with Deep Learning, Segmentation and Visualization. *IEEE Access* **2020**, *8*, 191586–191601. https://doi.org/10.1109/ACCESS.2020.3031384.
29. Marvin, G.; Alam, M.G.R. Explainable Augmented Intelligence and Deep Transfer Learning for Pediatric Pulmonary Health Evaluation. In Proceedings of the 2022 International Conference on Innovations in Science, Engineering and Technology (ICISET), Chittagong, Bangladesh, 26–27 February 2022.
30. Yadav, O.; Passi, K.; Jain, C.K. Using Deep Learning to Classify X-ray Images of Potential Tuberculosis Patients. In Proceedings of the 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain, 3–6 December 2018.
31. Hwang, S.; Kim, H.-E.; Jeong, J.; et al. A Novel Approach for Tuberculosis Screening Based on Deep Convolutional Neural Networks. In Proceedings of the Medical Imaging 2016: Computer-Aided Diagnosis, San Diego, CA, USA, 27 February–3 March 2016; Vol. 9785, pp. 750–757.