Test-Time Adaptation through Semantically-guided Feature Decomposition for Few-shot Chest X-ray Diagnosis

Jayant Mahawar
Indian Institute of Technology Jodhpur
India
mahawar.2@iitj.ac.in

Angshuman Paul
Indian Institute of Technology Jodhpur
India
apaul@iitj.ac.in

1. t-SNE Visualization of Feature Representation

This section presents comparative t-SNE visualizations (Figures 1, 2, and 3) of the learned feature representations from both our model's invariant and discriminative branches across the CheXpert, MIMIC-CXR, and NIH chest x-ray datasets. These visualizations serve to qualitatively demonstrate the efficacy of our proposed dual latent-space autoencoder in separating class-specific, discriminative information from shared, class-agnostic patterns.

In all three datasets, the invariant branch consistently captures highly overlapping distributions across various classes. This indicates that this branch successfully extracts general, shared features such as anatomical structures that are present irrespective of the specific abnormality. These class-agnostic features, while crucial for establishing general visual context, are inherently non-discriminative and thus insufficient for reliable disease classification.

In stark contrast, the discriminative branch consistently produces more separable and distinct clusters, each corresponding to a specific disease class. This clear separation reflects the model's ability to effectively isolate class-specific cues, which are critical for accurate diagnosis.

Delving into specific dataset observations:

- CheXpert Dataset (Figure 1): While the discriminative branch achieves commendable separation, some degree of inter-class overlap persists. This is particularly noticeable in panel (b), where conditions like Pneumonia, Edema, and Atelectasis—all of which frequently manifest as diffuse opacities with overlapping radiographic densities—exhibit close clustering. Similarly, panel (c) shows that Lung Lesion, Lung Opacity, and Consolidation, which may present with similar focal patterns or diffuse appearances, tend to cluster near one another. These findings underscore that even with our discriminative representation, some inherent inter-class similarity is naturally preserved due to genuine visual resemblance among certain chest x-ray conditions.
- MIMIC-CXR Dataset (Figure 2): A similar trend of inherent visual similarity impacting feature separation is evident here. Consolidation and Pneumonia (panel b) appear almost indistinguishable in the latent space. This is a direct consequence of their profound visual similarity in radiographs, where both often manifest as localized or diffuse airspace opacities. In panel (c), Lung Lesion, Lung Opacity, and Consolidation again exhibit overlapping feature representations, reinforcing the observation that abnormalities with similar focal appearances or textural characteristics tend to occupy nearby regions in the latent space, even within the discriminative output.
- NIH Dataset (Figure 3): This dataset further reinforces these insights. Here, Mass and Nodule in panel (b) have overlapping data and are represented in close proximity. This proximity stems from their common nature as lesion-type pathologies, which frequently share similar size, shape, boundary characteristics, or calcification patterns on chest x-rays. Even in panel (c), despite overall clearer separation for the selected classes, some degree of feature similarity remains among visually related conditions.

These observations highlight an inherent challenge and limitation for any discriminative model: when diseases exhibit strong visual overlap, achieving perfect, clean separation in feature space may be intrinsically difficult, emphasizing the importance of our semantic guidance and TTA for robustness. Collectively, these t-SNE visualizations compellingly reveal the strength of our proposed method in successfully decomposing shared (class-agnostic) and unique (class-specific) visual information, which is a key factor contributing to its superior performance in few-shot chest x-ray diagnosis.

2. More Image Results

Figures 4, 5, and 6 provide qualitative insights into the performance of our proposed method on the NIH, CheXpert, and MIMIC-CXR datasets, respectively. Each figure dis-

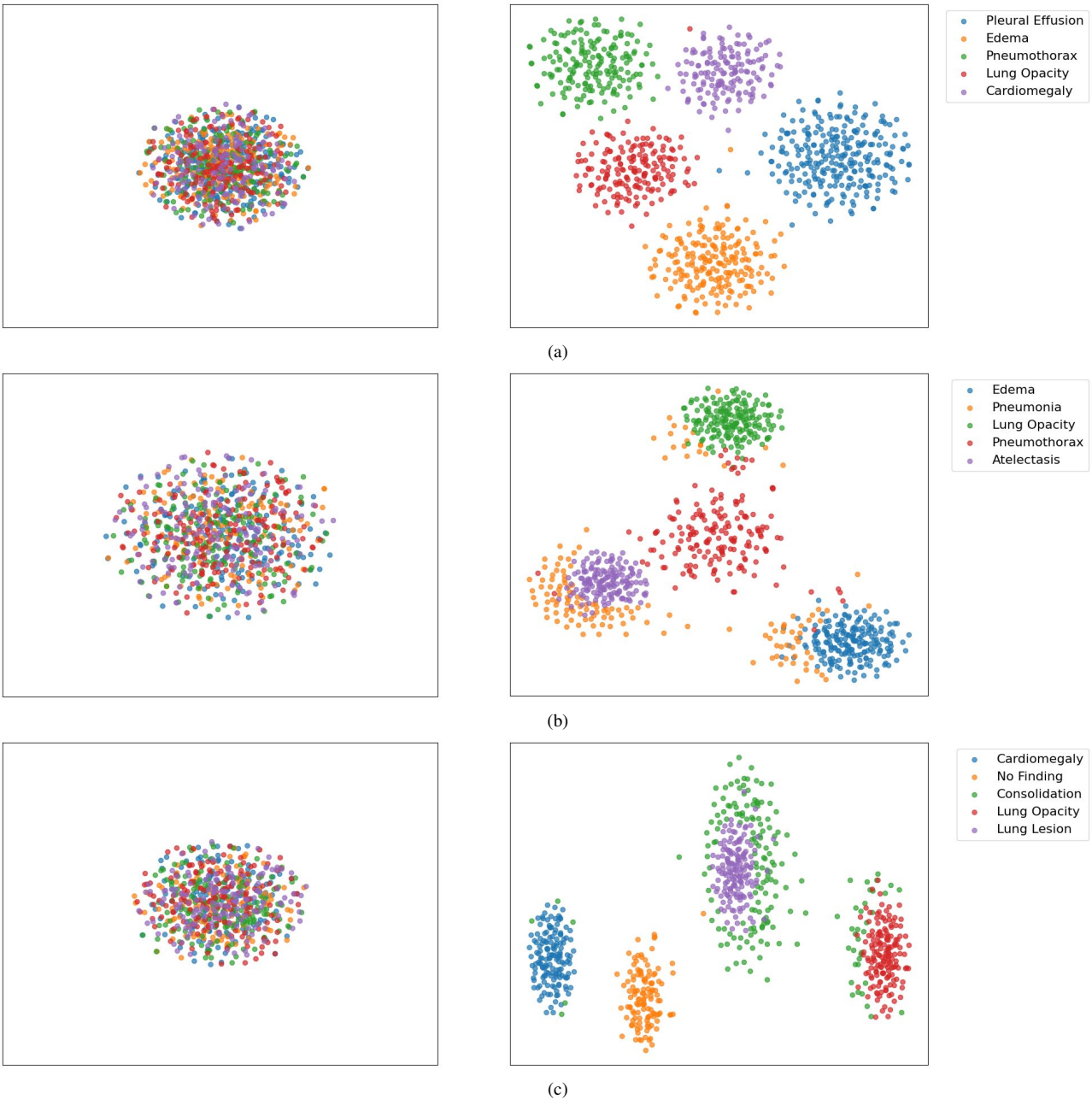
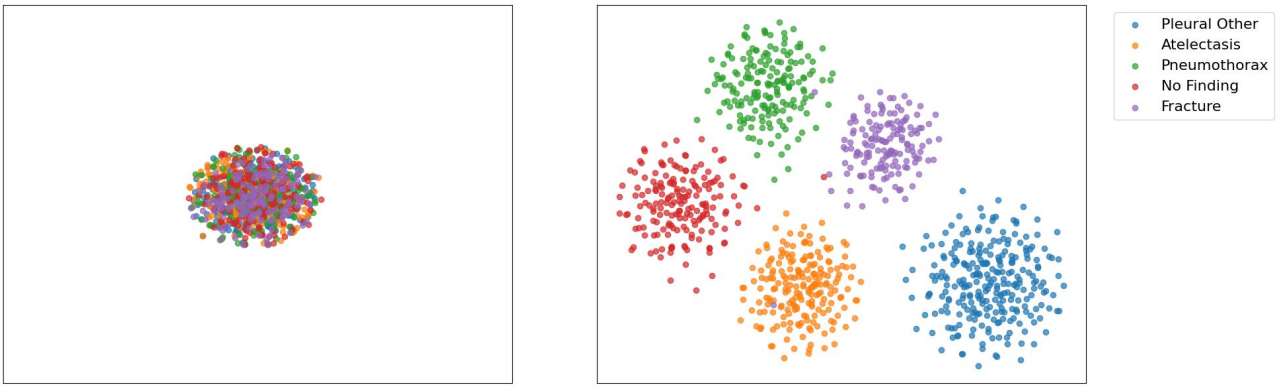
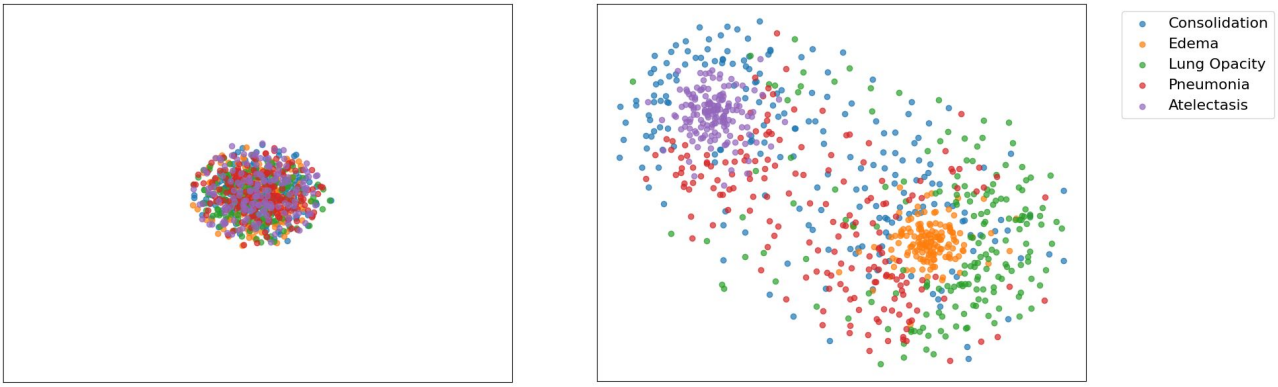


Figure 1. t-SNE visualization of feature representations, depicting the latent spaces of the discriminative (right) and invariant (left) branches for different combinations of classes on the CheXpert dataset. In the invariant branch (left), class-agnostic features form an overlapping distribution across the five classes, demonstrating shared, non-discriminative representations. In contrast, the discriminative branch (right) exhibits distinctly well-separated clusters of class-specific features, highlighting the proposed method's capability to isolate highly discriminative information. (a) All abnormalities shown are characterized by distinct traits. (b) Among these, Pneumonia, Edema, and Atelectasis are observed to be opacities with overlapping densities in the feature space. (c) Lung Lesion, Lung Opacity, and Consolidation demonstrate an overlap in their focal positioning, suggesting visual commonalities.

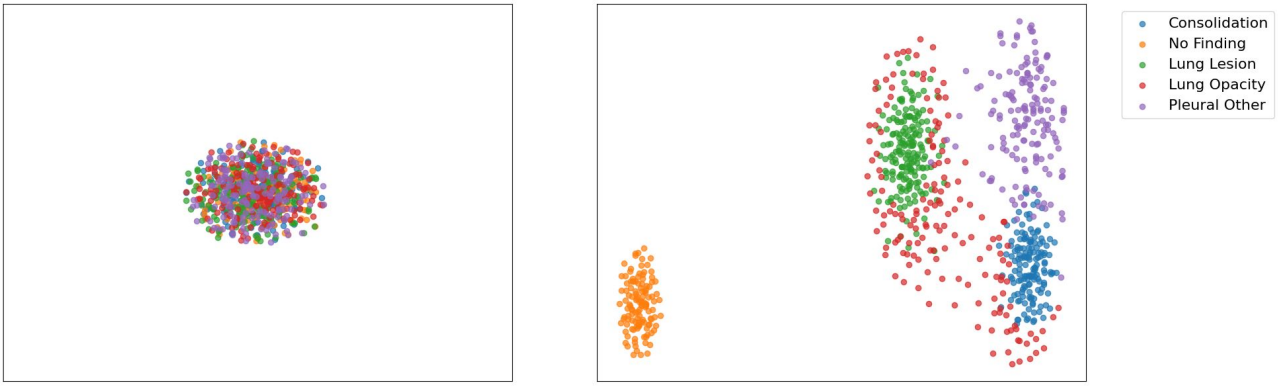
plays representative sample images alongside their Ground Truth (GT) and Predicted (D) class labels. These visualizations demonstrate the model's ability to accurately classify various chest x-ray abnormalities for the majority of cases. The correct classification of these samples underscores the efficacy of our semantically-guided feature decomposition and test-time adaptation in extracting discriminative features and adapting to real-world data variability. These vi-



(a)



(b)



(c)

Figure 2. t-SNE visualizations of feature representations from the discriminative (right) and invariant (left) branches, illustrating their respective latent spaces for various class combinations on the MIMIC-CXR dataset. The invariant branch (left) displays class-agnostic features forming an overlapping distribution across all five classes, indicating shared, non-discriminative representations. Conversely, the discriminative branch (right) clearly exhibits well-separated clusters of class-specific features, highlighting the method's ability to effectively isolate discriminative information. (a) All depicted abnormalities possess distinct radiographic traits. (b) This panel highlights five airspace diseases, demonstrating that Consolidation and Pneumonia are nearly indistinguishable in the feature space. (c) Lung Lesion, Lung Opacity, and Consolidation show an overlap in their focal positioning within the latent space, reflecting their visual similarities.

sual examples complement our quantitative results, highlighting the practical diagnostic capabilities of the framework.

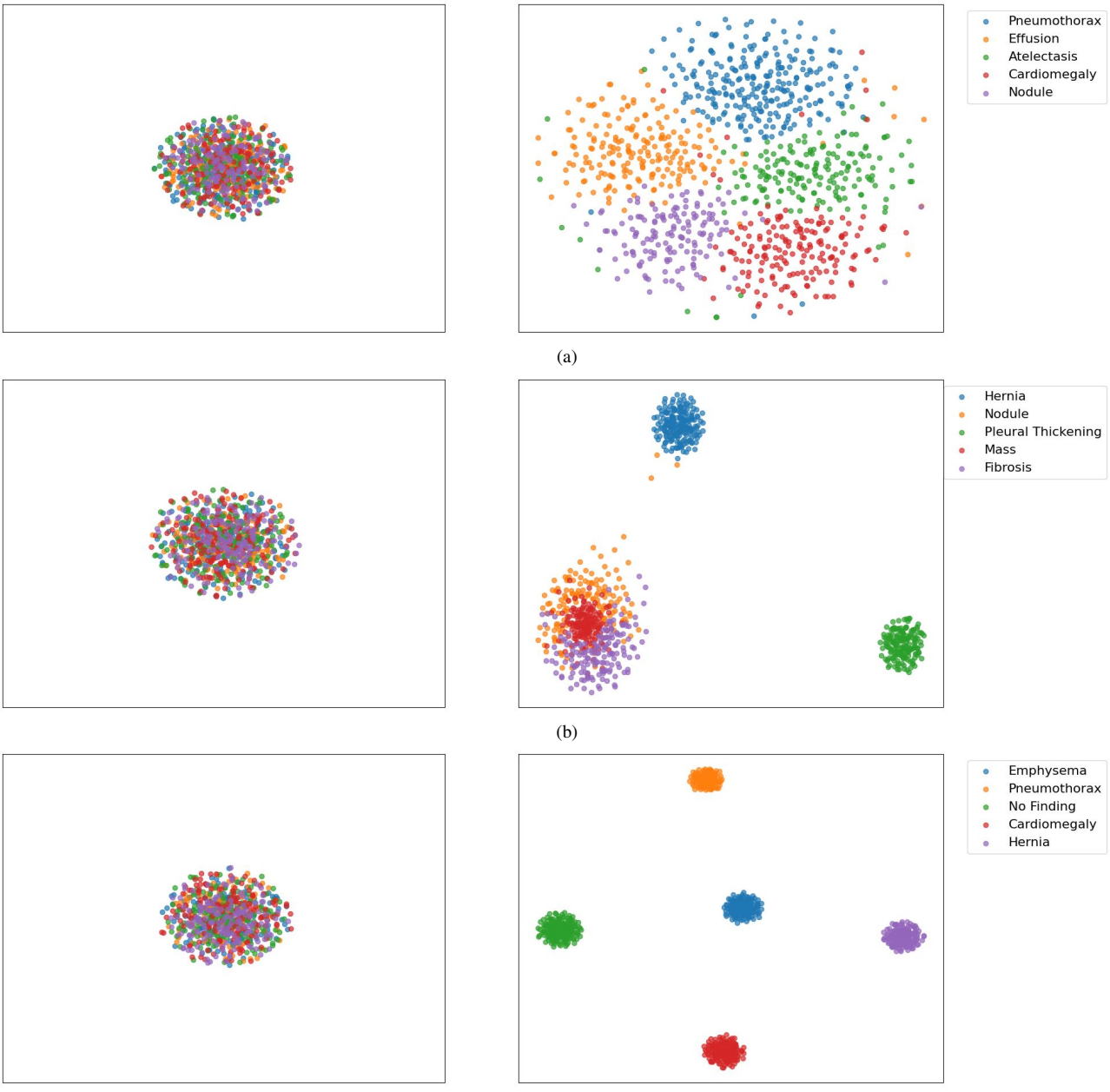


Figure 3. t-SNE visualization of feature representations, depicting the latent spaces of the discriminative (right) and invariant (left) branches for different combinations of classes on the NIH dataset. In the invariant branch (left), class-agnostic features form an overlapping distribution across the five classes, demonstrating shared, non-discriminative (class-agnostic) representations. In contrast, the discriminative branch (right) exhibits distinctly well-separated clusters of class-specific features, highlighting the proposed method's capability to isolate highly discriminative information. (a) All depicted abnormalities, while distinct, share common boundaries in the latent feature space. (b) Masses, Nodules, and Fibrosis, being similar lesion types, are represented in closer proximity within the feature space. (c) This combination of abnormalities tends to possess more unique radiographic patterns, contributing to their clearer differentiation in the discriminative feature space.

GT: Cardiomegaly
D: Cardiomegaly

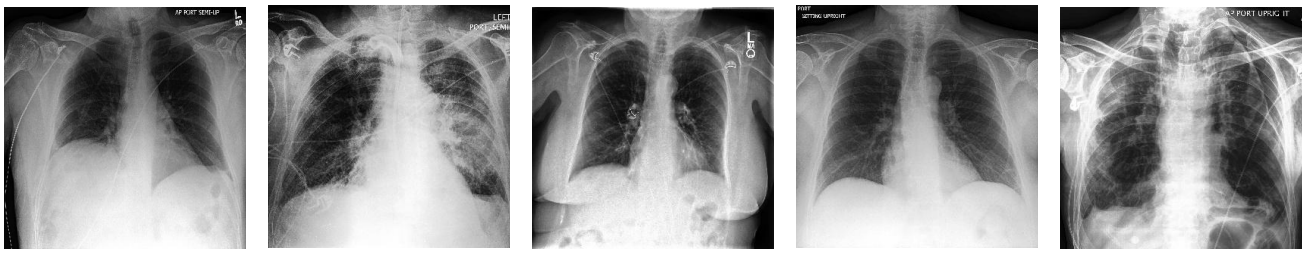
GT: Infiltration
D: Infiltration

GT: Cardiomegaly
D: Atelectasis

GT: No Finding
D: No Finding

GT: Effusion
D: Pneumothorax

Figure 4. Sample images from the NIH dataset with class predictions made by the proposed method. GT and D denote the ground truth and detected class labels for the input images, respectively. More image results have been presented in the supplementary material.



GT: Atelectasis
D: Atelectasis

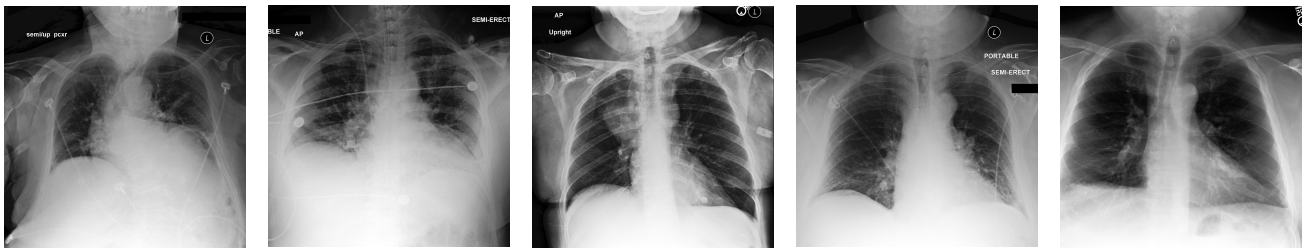
GT: Consolidation
D: Consolidation

GT: Lung Opacity
D: Lung Lesion

GT: No Finding
D: No Finding

GT: Pneumothorax
D: Pneumonia

Figure 5. Sample images from the CheXpert dataset with class predictions made by the proposed method. GT and D denote the ground truth and detected class labels for the input images, respectively. More image results have been presented in the supplementary material.



GT: Cardiomegaly
D: Cardiomegaly

GT: Edema
D: Edema

GT: Lung Lesion
D: Lung Opacity

GT: No Finding
D: No Finding

GT: Pleural Effusion
D: Pneumothorax

Figure 6. Sample images from the MIMIC-CXR dataset with class predictions made by the proposed method. GT and D denote the ground truth and detected class labels for the input images, respectively. More image results have been presented in the supplementary material.