DISTRIBUTION-GUIDED GENERATIVE REPLAY WITH SEMANTIC PROMPTS FOR CLASS-INCREMENTAL CHEST X-RAY DIAGNOSIS

Jayant Mahawar, Devi Prasad Maharathy, and Angshuman Paul

Indian Institute of Technology Jodhpur

BMVC 2025

Paper ID: 122

Motivation and Objective

- Evolving respiratory diseases require adaptable AI, but traditional CIL is limited:
  - Privacy Risk: Requires storing raw patient data (exemplars).
  - Instability: Compromised discriminative power.
  - Resource Intensive: Needs huge pretrained models
- We aim to develop a privacy-preserving, stable, and accurate CIL framework for multi-class CXR diagnosis.
- Our framework integrates Prompt Tuning and Distribution-Guided Generative Replay using a VAE-based network.
- Our approach balances stability and plasticity by combining semantic prompts with privacy-preserving feature replay.

Methodology

- Core Architecture
  - Image Encoder (E_I): Frozen after initial session for feature stability.
  - DPL (VAE-based): Learns latent feature distribution (μ, Σ) and refines prompts.
  - Prompt Pool ($\mathcal{P}$): Stores semantically-initialized, class-specific prompts.
- Feature Replay: Past knowledge is retained by generating synthetic features using frozen latent distribution parameters and stored class prompts, eliminating raw data storage.
- Training Strategy:
  - Session 0 (Base): Train E_I and DPL. Freeze E_I and LEmNet (distribution alignment). $\mathcal{L}_{\text{base}}$ loss.
  - $\rightarrow$ Sessions 1 to T (Incremental): Train on New Real features + Old Synthetic features. $\mathcal{L}_{\text{new}}$ loss.

Results

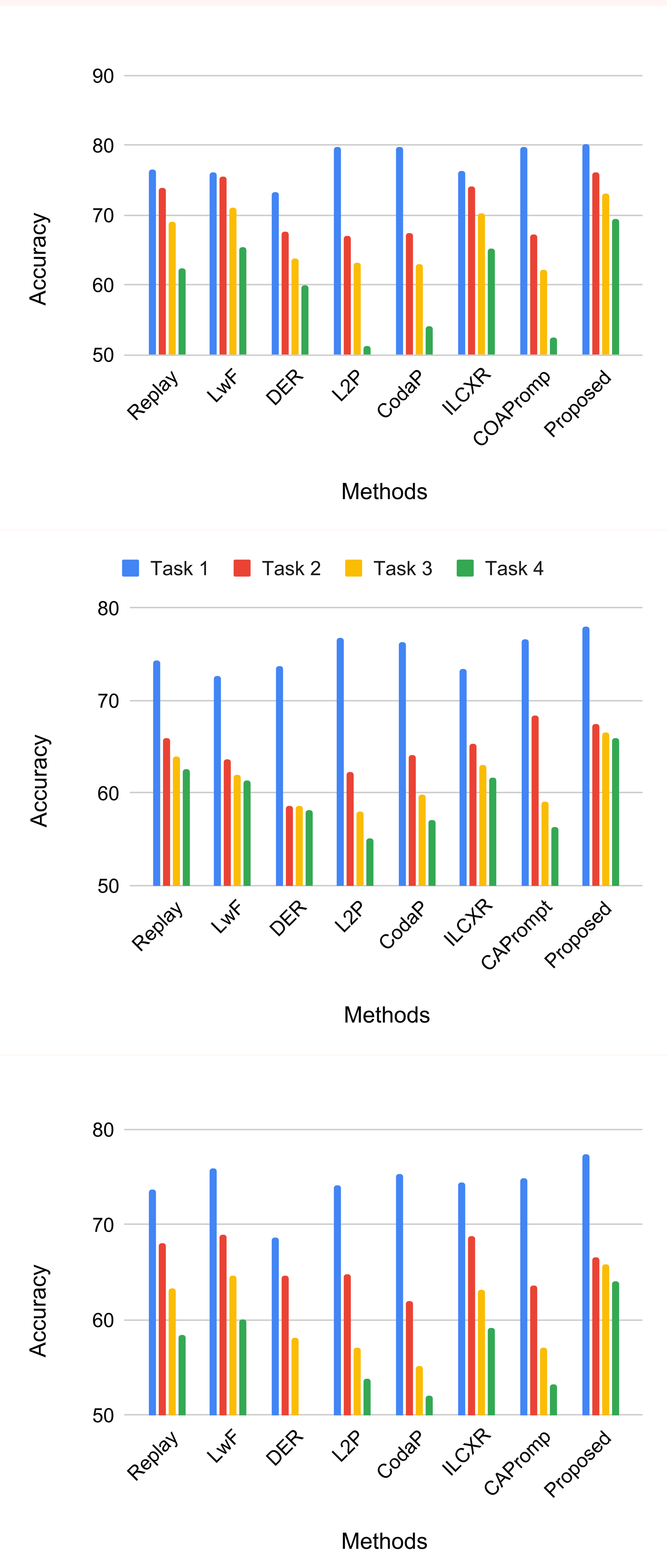

Figure 1. Comparison of task-wise accuracy across sessions for different methods on chest x-ray images from the NIH (top), CXP (middle), and MCXR (bottom) datasets.

Block Diagram

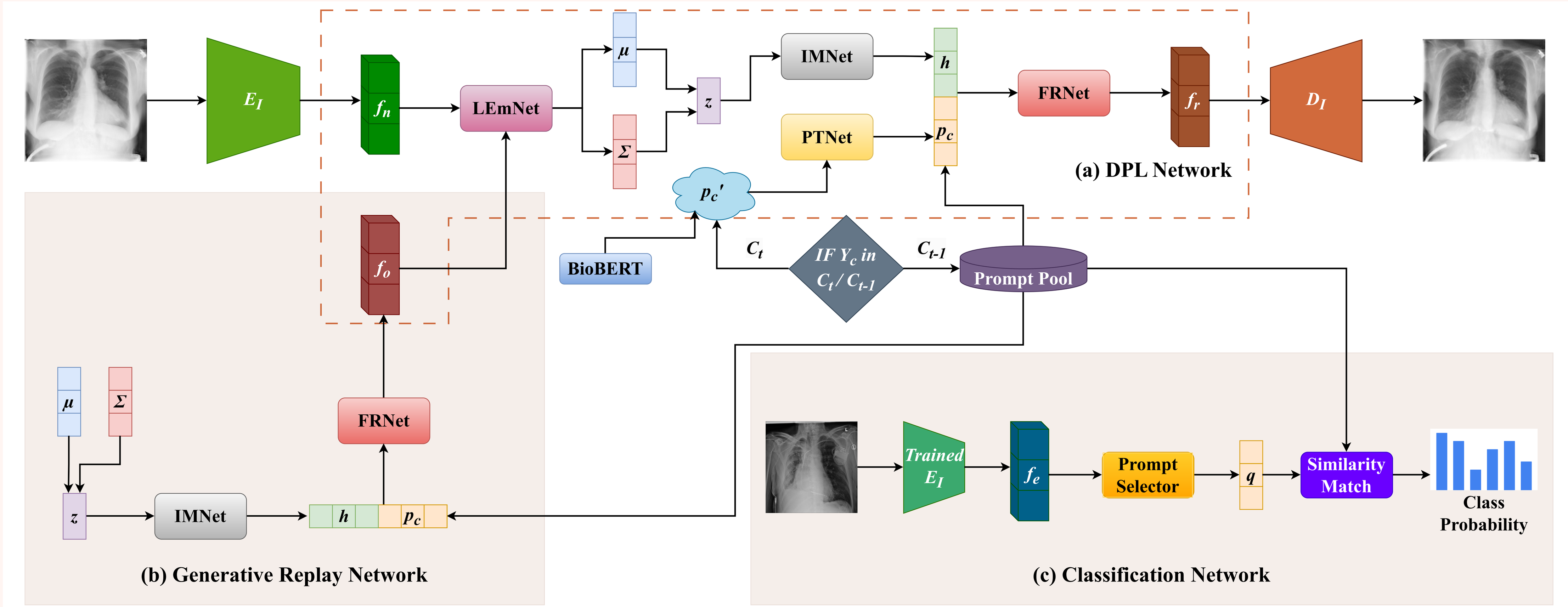

Figure 2. The block diagram of our method with an image encoder E_I, and decoder D_I. (a) The Distribution & Prompt Learner (DPL) network is built on VAE. It learns class-specific prompts (p_c) and captures the data distribution using LEmNet (latent distribution μ,Σ), IMNet (intermediate space projection), PTNet (semantic embedding to prompt), and FRNet (feature reconstruction via intermediate features and prompts). After initial training, the image encoder (E_I) and LEmNet are frozen to retain past knowledge, while other components adapt to new classes. (b) To preserve past knowledge, synthetic features are generated using stored prompts (from Prompt Pool $\mathcal{P}$) and the frozen latent distribution. (c) Classification involves computing the cosine similarity between a query vector q and learned prompts p_c.

Image Results

	Ground Truth Pleural Effusion	Task 1 Top-4 Prediction Score **Pleural Effusion:0.59** Edema: 0.15 Lung Opacity: 0.15 No Finding: 0.01	Task 2 Top-4 Prediction Score **Pleural Effusion:0.56** Cardiomegaly: 0.17 Edema: 0.15 Enlarged_Cardio: 0.12	Task 3 Top-4 Prediction Score **Pleural Effusion:0.52** Consolidation: 0.18 Atelectasis:0.17 Cardiomegaly:0.13	Task 4 Top-4 Prediction Score **Pleural Effusion:0.64** Pneumonia:0.23 Consolidation: 0.07 Edema: 0.06

Figure 3. Example of prior knowledge retention in our methodology. An image from the CheXpert (CXP) dataset with the ground truth label Pleural Effusion is shown. The model, trained in a CIL setting, predicts based on the highest confidence score. The top-1 predicted labels after Task 1 to Task 4 are highlighted in bold, demonstrating consistent correct predictions across all tasks.

Performance Comparison with SOTA

Table 1. Comparative performance in terms of average accuracy (A_{avg}, mean ± sd for five runs) and Performance Drop (PD, mean ± sd for five runs) for different approaches on the NIH, CXR, and MCXR datasets .

Method	NIH		CXP		MCXR	
	$A_{\text{avg}}\uparrow$	$PD\downarrow$	$A_{\text{avg}}\uparrow$	$PD\downarrow$	$A_{\text{avg}}\uparrow$	$PD\downarrow$
Replay	70.48 ± 2.5	14.2 ± 1.2	66.70 ± 1.9	11.7 ± 1.1	65.88 ± 1.7	15.2 ± 0.9
LwF	72.03 ± 2.9	10.8 ± 1.5	64.93 ± 1.7	**11.4 ± 1.1**	67.40 ± 1.9	15.8 ± 0.8
DER	66.15 ± 2.8	13.4 ± 1.2	62.35 ± 2.1	15.6 ± 1.9	60.25 ± 2.5	19.3 ± 2.1
L2P	65.30 ± 4.1	28.6 ± 2.3	63.05 ± 3.8	21.7 ± 1.7	62.48 ± 4.1	20.2 ± 2.7
CodaP	66.05 ± 4.6	25.7 ± 2.6	64.35 ± 3.2	19.3 ± 1.6	61.10 ± 4.0	23.3 ± 2.5
ILCXR	71.13 ± 2.2	11.2 ± 1.5	65.73 ± 1.8	11.9 ± 1.0	67.87 ± 1.8	15.6 ± 1.3
CAPrompt	65.75 ± 4.3	27.3 ± 2.4	65.15 ± 3.6	20.1 ± 1.2	62.91 ± 4.2	22.1 ± 2.6
Proposed	**74.70** ± 2.3	**10.8** ± 0.7	**69.48** ± 1.2	12.1 ± 1.0	**68.50** ± 2.1	**13.3** ± 1.3

Ablation Studies

Table 2. Performance comparison of the proposed method under three configurations: (1) without generative replay (WGR), (2) without freezing encoder weights (WFEW), and (3) with prompts initialized with random prompts, evaluated on the NIH, CXR, and MCXR datasets for incremental learning .

Configuration	NIH		CXP		MCXR	
	$A_{\text{avg}}\uparrow$	$PD\downarrow$	$A_{\text{avg}}\uparrow$	$PD\downarrow$	$A_{\text{avg}}\uparrow$	$PD\downarrow$
WGR	72.92 ± 2.2	11.7 ± 1.5	66.73 ± 2.1	12.7 ± 1.1	66.42 ± 1.6	13.9 ± 1.6
WFEW	71.24 ± 4.1	12.7 ± 2.1	67.28 ± 1.3	13.1 ± 1.1	65.19 ± 1.9	13.9 ± 1.4
WRPI	73.42 ± 5.3	12.1 ± 1.7	68.17 ± 4.1	12.6 ± 1.2	67.15 ± 3.3	15.3 ± 2.3
Proposed	**74.70** ± 2.3	**10.8** ± 0.7	**69.48** ± 1.2	**12.1** ± 1.0	**68.50** ± 2.1	**13.3** ± 1.3

Conclusion

- Our method is a robust CIL foundation for privacy-sensitive medical AI.
- We outperform SOTA CIL methods on NIH, CXP, and MCXR datasets.
- Our method significantly improves average accuracy (A_{avg}) and minimizes forgetting (PD).
- Future Work: Explore dynamic prompt growth; investigate conditional encoder updates for better adaptability; and use Vision Transformer backbones.

Lab Website:

Contact: