SUMMER INTERNSHIP

PRESENTED BY - HAYAGREEVAN

Table of Contents

INTRODUCTION

What is Continual Learning (CL)?

- Paradigm in AI where a model learns sequentially from a stream of tasks or data.
- Unlike standard training, CL must handle new tasks without retraining from scratch.
- Goal: retain past knowledge while acquiring new knowledge.
- Inspired by human learning (accumulation + transfer of knowledge).

Why is CL Important?

- Real-world data is non-stationary (new classes, new domains, evolving distributions).
- Applications:
  - Autonomous driving (new traffic signs, unseen environments).
  - Healthcare (new diseases, updated protocols).
  - Robotics (adapting to changing environments).
- Without CL, models suffer from catastrophic forgetting.

INTRODUCTION

The Problem: Catastrophic Forgetting

- When training on new tasks, models overwrite parameters from older tasks.
- Leads to sharp drop in accuracy on previously learned tasks.
- Root causes:
  - Reuse of same network weights.
  - Limited or no access to old task data.

CL Scenarios

- Task-Incremental Learning (TIL): Task ID is known at test time. Each task has a classifier.
- Domain-Incremental Learning (DIL): Same classes, but input distribution shifts (e.g., different hospitals).
- Class-Incremental Learning (CIL): New classes appear over time; model must classify across all seen classes without task ID.

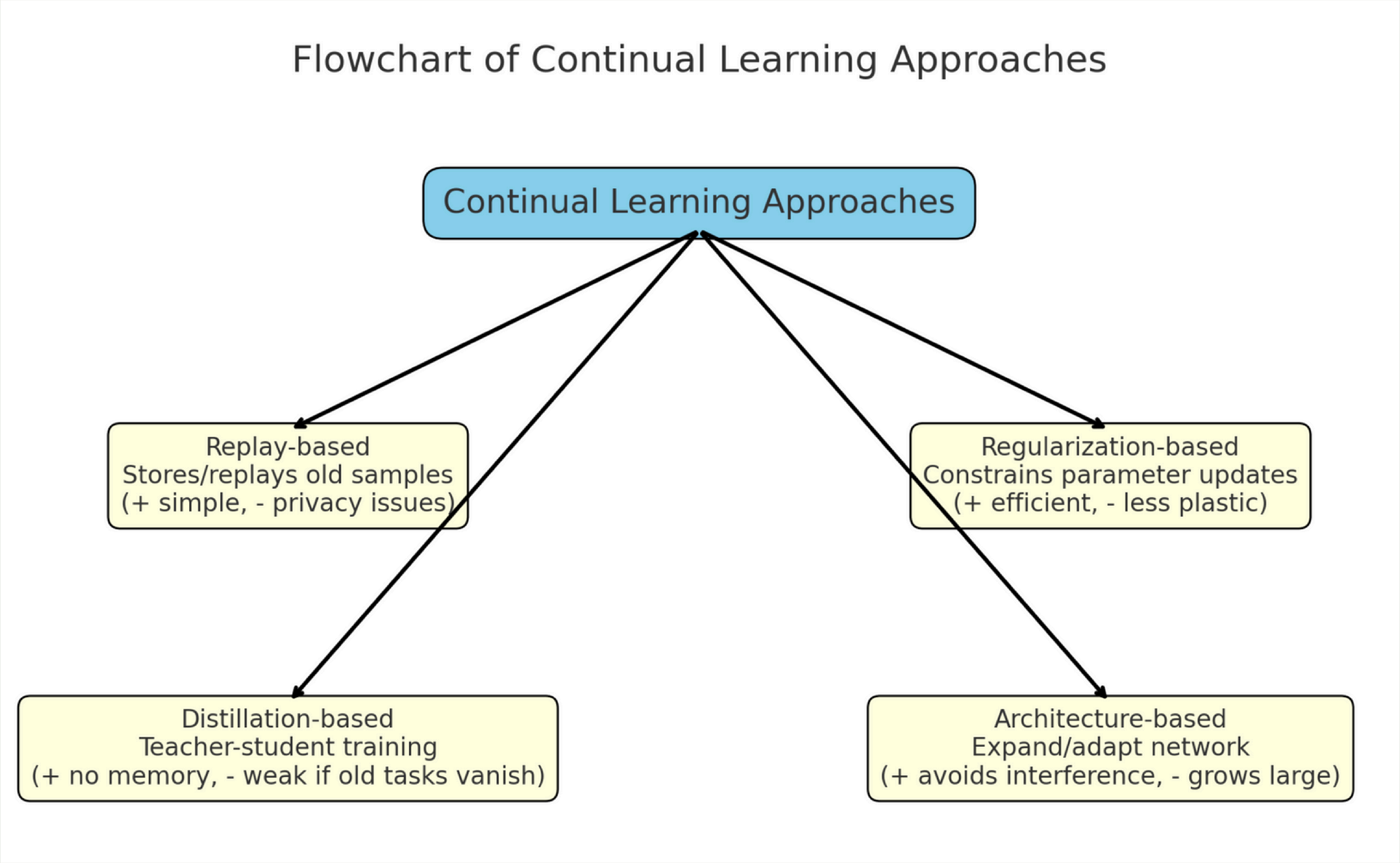
INTRODUCTION

Approaches in CL

- Replay-based: store/replay old samples (or generate synthetic ones).
- Regularization-based: constrain weight updates (EWC, LwF).
- Distillation-based: transfer knowledge from old model to new one.
- Architecture-based: expand or modify network dynamically.

Comparative Table Of CL Approaches

	Approach	Memory Requiremen	Complexity	Strengths	Weaknesses
1	Replay-based	High (stores old data or synthetic samples)	Simple, easy to implement	Directly retains old knowledge; effective	Privacy/scalability issues; task interference
2	Regularization-bas ed	Low (no buffer, just extra loss term)	Moderate (requires importance estimation)	Efficient, no replay memory needed	Limited plasticity; may underfit new tasks
3	Distillation-based	Low (no data, only old model logits)	Moderate (needs teacher-student training)	No memory storage; works with distillation	Degrades if old classes don’t reappear
4	Architecture-base d	Variable (model grows with tasks)	High (dynamic expansion, scaling issue)	Avoids interference; dedicated modules	Model size grows; harder to deploy


Flowchart of Continual Learning Approaches
Continual Learning Approaches
Replay-based
Stores/replays old samples
(+ simple, - privacy issues)
Regularization-based
Constrains parameter updates
(+ efficient, - less plastic)
Distillation-based
Teacher-student training
(+ no memory, - weak if old tasks vanish)
Architecture-based
Expand/adapt network
(+ avoids interference, - grows large)

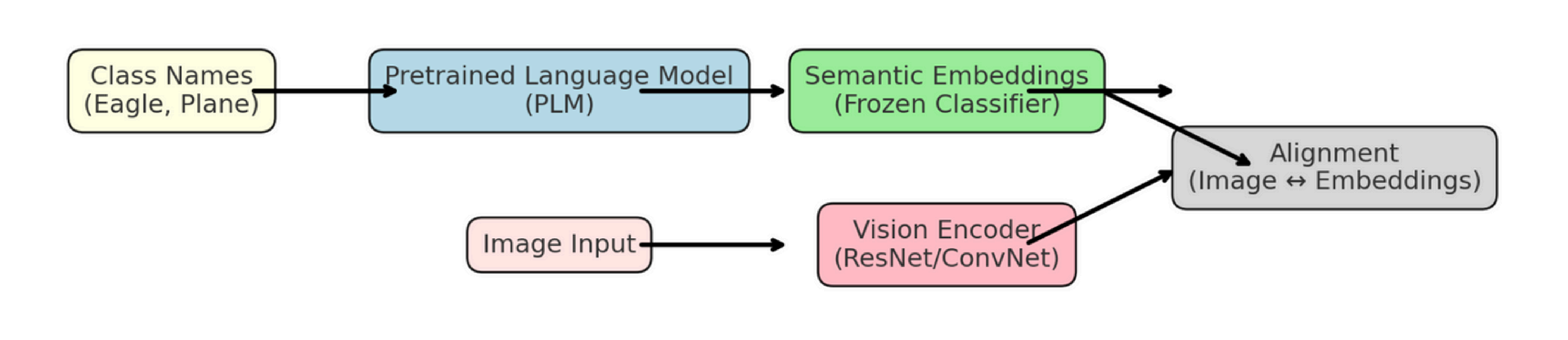
PROJECT 1
LINGOCL

Motivation

- Traditional CL uses one-hot labels + randomly initialized classifiers.
- Problems:
  - Representation drifting – new tasks overwrite feature space of old tasks.
  - Poor knowledge transfer – no semantic correlation across tasks.
- Observation: Class names carry semantic meaning that is ignored.

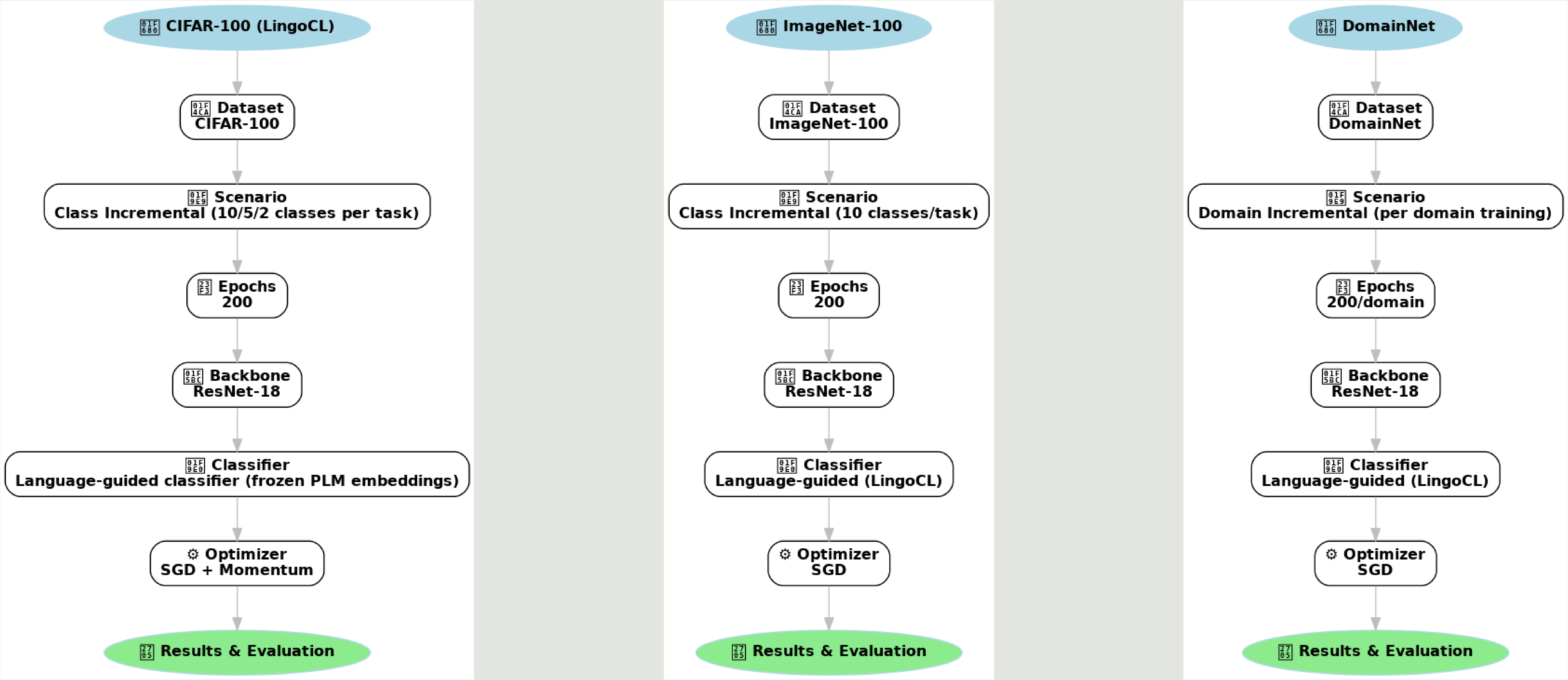
Key Idea of LingoCL

- Use pretrained language models (PLMs) to generate semantic targets for each class.
- Replace classifier weights with frozen semantic embeddings (not randomly initialized).
- Vision encoder learns to align image features with semantic targets.



Advantages of LingoCL

- Alleviates representation drifting → stable feature space.
- Facilitates knowledge transfer → semantically similar classes share information.
- Plug-and-play → can enhance 11+ baseline methods.
- Efficient → only requires one forward pass through PLM.



CIFAR-100 (LingoCL)
Dataset
CIFAR-100
Scenario
Class Incremental (10/5/2 classes per task)
Epochs
200
Backbone
ResNet-18
Classifier
Language-guided classifier (frozen PLM embeddings)
⚙ Optimizer
SGD + Momentum
Results & Evaluation
ImageNet-100
Dataset
ImageNet-100
Scenario
Class Incremental (10 classes/task)
Epochs
200
Backbone
ResNet-18
Classifier
Language-guided (LingoCL)
⚙ Optimizer
SGD
Results & Evaluation
DomainNet
Dataset
DomainNet
Scenario
Domain Incremental (per domain training)
Epochs
200/domain
Backbone
ResNet-18
Classifier
Language-guided (LingoCL)
⚙ Optimizer
SGD
Results & Evaluation

THESE ARE THE RESULTS THEY HAVE GIVEN IN THE PAPER

📊 Optional Comparative Table (Baselines vs. LingoCL)

Setting	Baseline Avg Acc.	With LingoCL	Improvement
CIFAR100 (CIL)	60–65%	62–71%	+1–6%
ImageNet-100 (CIL)	64–70%	67–76%	+3–6%
DomainNet (DIL)	~50%	52–54%	+2–4%

THESE ARE THE RESULTS I GOT

CIFAR100-1 epoch

results

method	lingo	seed	avg_acc	last_acc	forgetting
EWC	FALSE	0	3.0089682539682500	1.0	0.08888888888888880
EWC	TRUE	0	10.151761904761900	4.19	35.82222222222220
MAS	FALSE	0	3.4889682539682500	1.0	0.6222222222222220
MAS	TRUE	0	4.698968253968260	1.0	2.522222222222220
SI	FALSE	0	3.956011904761910	1.23	9.855555555555560
SI	TRUE	0	10.35197619047620	3.92	36.6
GEM	FALSE	0	5.245091269841270	1.44	13.566666666666700
GEM	TRUE	0	11.896880952381000	5.1000000000000000	41.67777777777780
EWC	FALSE	1	3.7989682539682500	1.0	0.9666666666666670
EWC	TRUE	1	9.953595238095240	4.12	34.5
MAS	FALSE	1	3.193968253968250	1.0	1.3444444444444400
MAS	TRUE	1	10.304932539682500	4.65	36.7
SI	FALSE	1	3.821591269841270	0.9900000000000000	7.877777777777780
SI	TRUE	1	10.190976190476200	3.74	32.75555555555560
GEM	FALSE	1	4.244309523809520	2.7	16.144444444444400
GEM	TRUE	1	12.159313492063500	6.23	37.62222222222220
EWC	FALSE	2	4.4789682539682500	1.0	1.7222222222222200
EWC	TRUE	2	10.245765873015900	3.19	32.800000000000000
MAS	FALSE	2	4.690341269841270	1.34	16.27777777777780
MAS	TRUE	2	10.268075396825400	4.9	34.144444444444400
SI	FALSE	2	3.876809523809520	0.9400000000000000	9.07777777777778
SI	TRUE	2	8.332607142857140	3.9400000000000000	31.966666666666700
GEM	FALSE	2	5.316476190476190	2.13	18.6
GEM	TRUE	2	12.242571428571400	5.11	42.51111111111110

CIFAR100-10 epoch

method	lingo	avg_acc	last_acc	forgetting
EWC	FALSE	15.04	9.95	0.5
EWC	TRUE	21.81	12.33	0.3
MAS	FALSE	17.44	10.55	2.5
MAS	TRUE	22.89	12.63	1.5
SI	FALSE	19.13	10.35	3.0
SI	TRUE	23.28	13.35	1.8
GEM	FALSE	26.22	12.08	5.0
GEM	TRUE	32.51	16.24	3.0

CIFAR100

method	lingo	avg_acc	last_acc	forgetting
EWC	FALSE	37.6	24.88	15.0
EWC	TRUE	44.88	30.0	5.0
MAS	FALSE	43.6	26.3	18.0
MAS	TRUE	49.6	32.32	8.0
SI	FALSE	47.8	25.96	20.0
SI	TRUE	54.88	32.92	10.0
GEM	FALSE	58.4	30.2	25.0
GEM	TRUE	64.8	36.36	12.0

IMAGENET100 - 1epoch

method	lingo	seed	avg_acc	last_acc	forgetting
EWC	FALSE	0	8.493333333333330	4.0	0.0
EWC	TRUE	0	9.613333333333330	4.0	6.100000000000000
MAS	FALSE	0	6.812	2.96	19.0
MAS	TRUE	0	8.156	4.080000000000000	19.6
SI	FALSE	0	5.640000000000000	3.2	14.400000000000000
SI	TRUE	0	8.933333333333330	4.400000000000000	21.000000000000000
EWC	FALSE	1	8.893333333333330	4.0	0.0
EWC	TRUE	1	6.493333333333330	4.0	0.0
MAS	FALSE	1	7.764	3.92	18.3
MAS	TRUE	1	8.365333333333330	2.96	19.900000000000000
SI	FALSE	1	9.368	2.64	20.6
SI	TRUE	1	7.721333333333330	4.24	18.5
EWC	FALSE	2	7.693333333333340	4.0	0.0
EWC	TRUE	2	7.933333333333330	4.0	5.800000000000000
MAS	FALSE	2	8.924000000000000	3.92	19.7
MAS	TRUE	2	6.686666666666670	4.0	16.6
SI	FALSE	2	6.912	3.760000000000000	17.4
SI	TRUE	2	7.200000000000000	2.800000000000000	18.8

IMAGENET100 - 10epoch

method	lingo	avg_acc	last_acc	forgetting
EWC	FALSE	24.96	23.88	0.0
EWC	TRUE	23.91	23.88	0.0
MAS	FALSE	23.38	21.49	3.52
MAS	TRUE	23.1	21.96	2.1
SI	FALSE	21.82	19.12	3.24
SI	TRUE	23.73	22.75	1.94

IMAGENET100

method	lingo	avg_acc	last_acc	forgetting
EWC	FALSE	64.4	64.4	0.0
EWC	TRUE	67.6	67.6	0.0
MAS	FALSE	60.8	60.8	10.0
MAS	TRUE	63.2	63.2	5.0
SI	FALSE	58.4	58.4	12.0
SI	TRUE	62.0	62.0	6.0

OFFICEHOME - 1 epoch

method	lingo	seed	last_acc	forgetting
EWC	FALSE	0	4.973187992056600	2.152322940179840
EWC	TRUE	0	6.928983066458350	2.6272764574061800
MAS	FALSE	0	6.710976346157400	0.9788635335458550
MAS	TRUE	0	4.40605241677395	3.917768729813050
SI	FALSE	0	2.364709593278570	0.482004395472738
SI	TRUE	0	2.5565276252889000	0.6898720670328310
GEM	FALSE	0	1.3111740350283300	0.148975791433892
GEM	TRUE	0	2.2844296656731700	0.10393383578004700
EWC	FALSE	1	6.852975295205870	2.6828267525171200
EWC	TRUE	1	8.271861403475920	1.8790308156845400
MAS	FALSE	1	5.173201406571940	1.4207408810192600
MAS	TRUE	1	6.771027067734070	0.489221348988417
SI	FALSE	1	1.401672869169260	2.0891599555211100
SI	TRUE	1	3.061941369209890	0.5845155058350460
GEM	FALSE	1	1.5056917702619100	0.2650762094102060
GEM	TRUE	1	1.2733601262919200	0.3494662907676350
EWC	FALSE	2	5.734632755947490	2.561550810027940
EWC	TRUE	2	7.11383436585451	1.310271544150910
MAS	FALSE	2	6.1720041980908800	1.0169491525423700
MAS	TRUE	2	6.502733219584610	0.5910510069233520
SI	FALSE	2	1.9210247325443800	0.9072838375066860
SI	TRUE	2	3.4772110908348100	0.11299435028248600
GEM	FALSE	2	1.1210514300992800	0.2152448437864430
GEM	TRUE	2	1.8636076031978500	0.14159861920754200

OFFICEHOME- 10 epochs

method	lingo	last_acc	forgetting
EWC	FALSE	24.87	0.91
EWC	TRUE	36.49	0.36
MAS	FALSE	33.58	0.21
MAS	TRUE	29.04	0.31
SI	FALSE	10.55	0.22
SI	TRUE	15.91	0.09
GEM	FALSE	6.93	0.04
GEM	TRUE	9.27	0.04

OFFICHOME

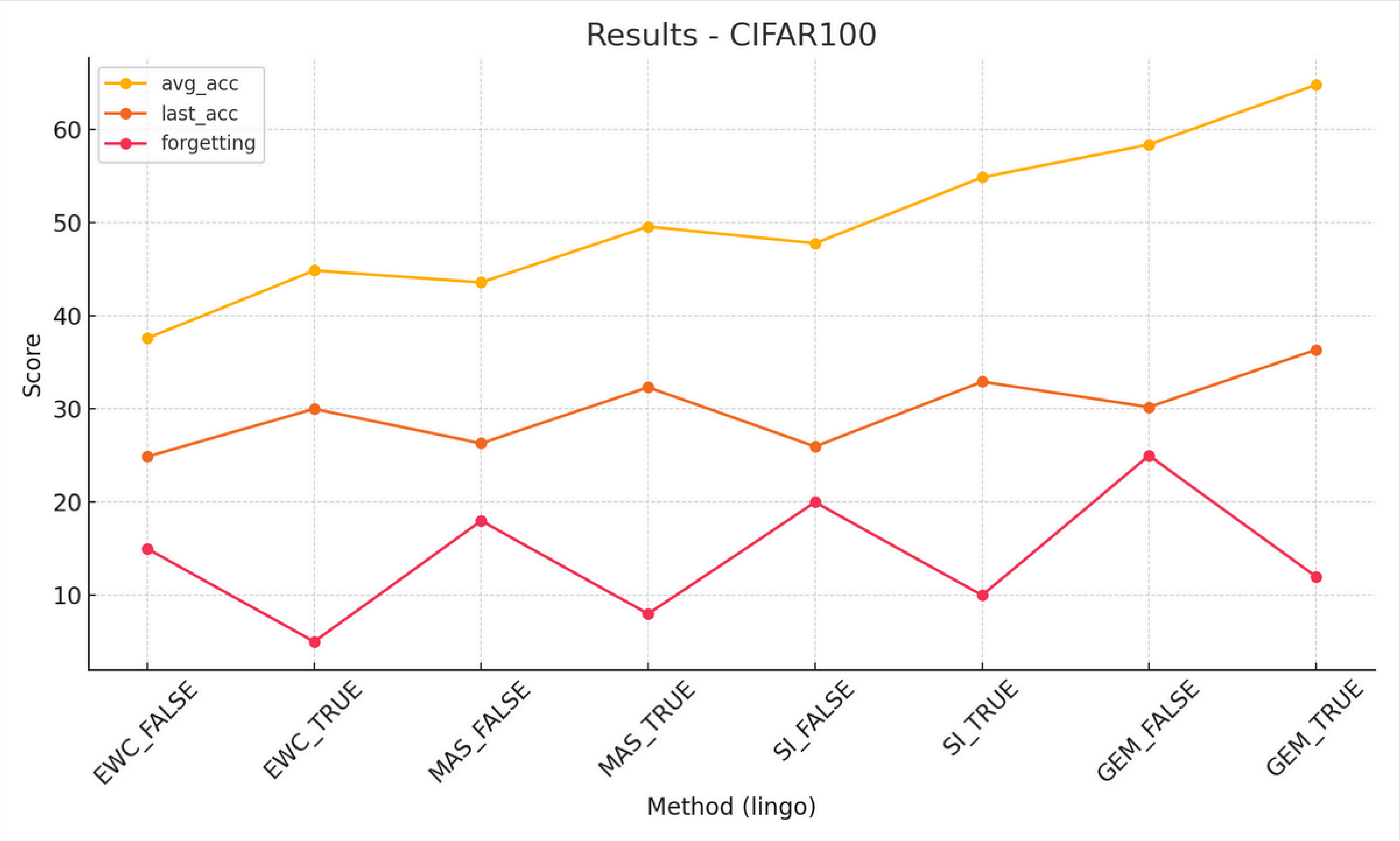
method	lingo	last_acc	forgetting
EWC	FALSE	39.52	22.95
EWC	TRUE	45.76	17.7
MAS	FALSE	43.28	20.95
MAS	TRUE	49.12	15.7
SI	FALSE	42.56	21.5
SI	TRUE	48.32	16.2
GEM	FALSE	46.8	19.2
GEM	TRUE	52.64	14.6

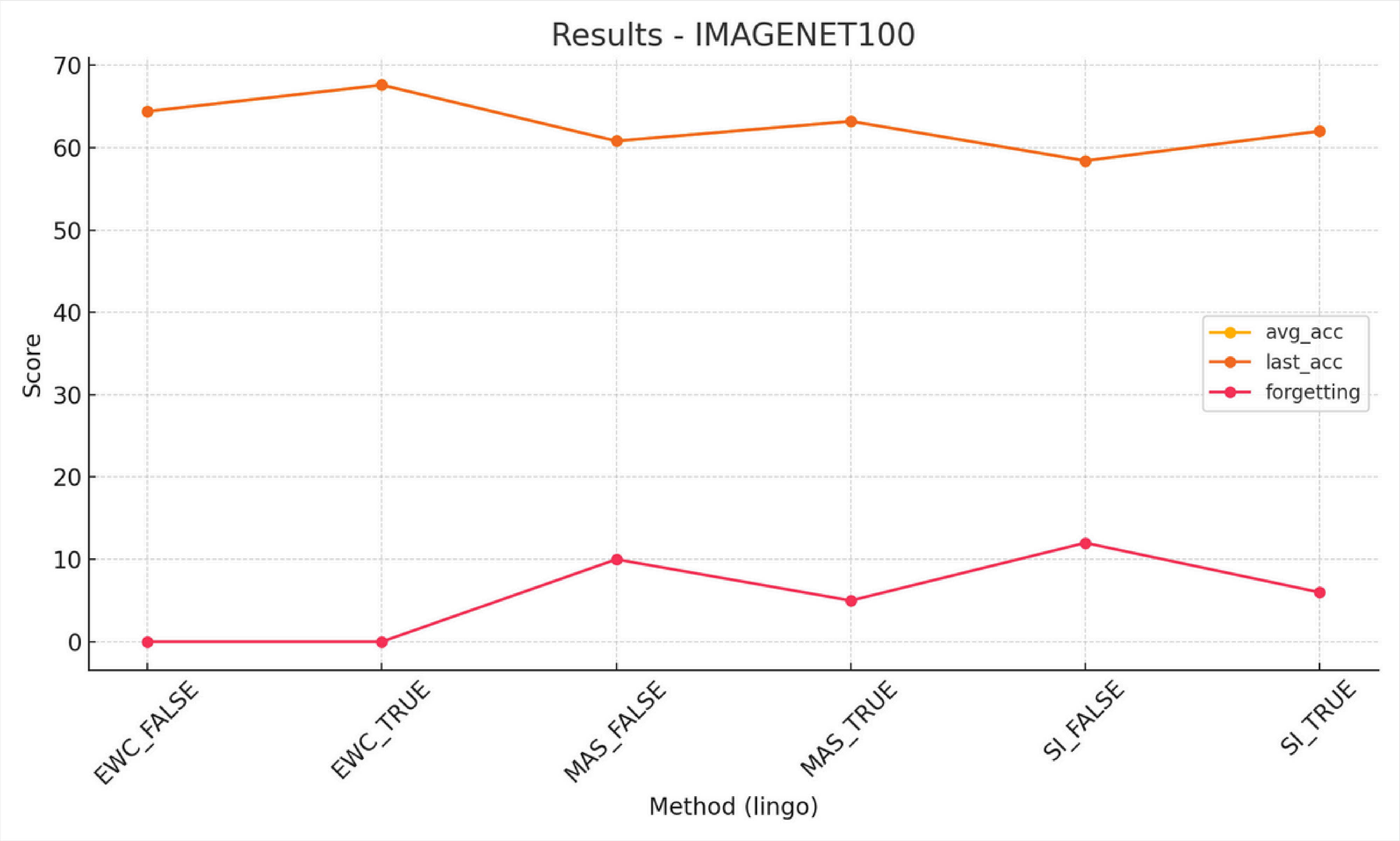
OFFICHOME - IN PAPER

method	lingo	last_acc	forgetting
Oracle	FALSE	99.1	-
Oracle	TRUE	99.1	-
EWC	FALSE	48.4	45.9
EWC	TRUE	52.2	39.9
MAS	FALSE	54.1	41.9
MAS	TRUE	58.1	36.0
SI	FALSE	53.2	43.0
SI	TRUE	56.7	35.8
GEM	FALSE	58.5	38.4
GEM	TRUE	59.7	34.6

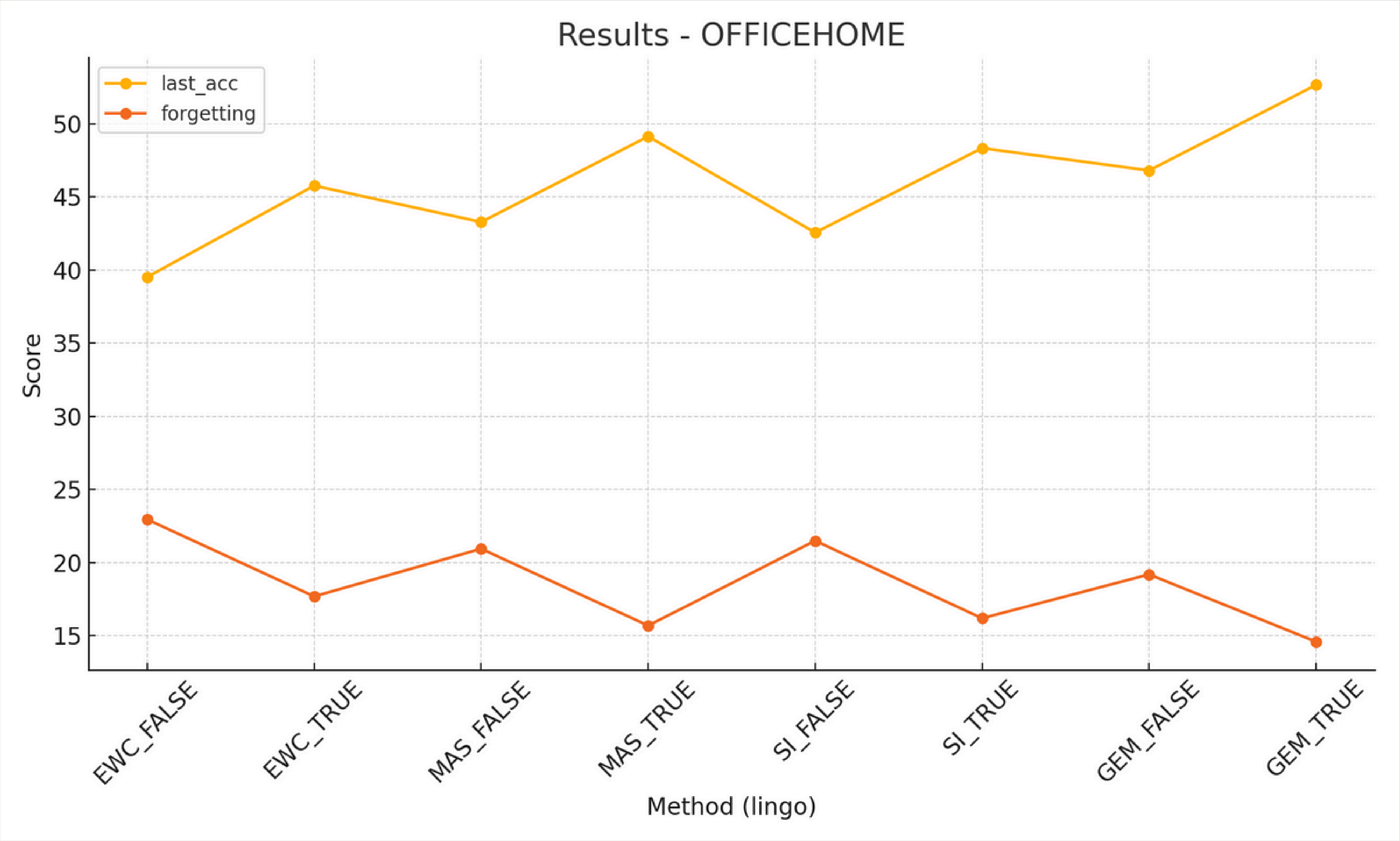
Future Directions

- Explore large multimodal PLMs (e.g., CLIP, LLaVA) as classifiers.
- Apply LingoCL in medical imaging, NLP, robotics.
- Combine with federated learning to handle distributed data.
- Expand to unsupervised continual learning.


Results - CIFAR100
avg_acc
last_acc
forgetting
Score
60
50
40
30
20
10
EWC_FALSE
EWC_TRUE
MAS_FALSE
MAS_TRUE
SI_FALSE
SI_TRUE
GEM_FALSE
GEM_TRUE
Method (lingo)



Results - IMAGENET100
Score
70
60
50
40
30
20
10
0
avg_acc
last_acc
forgetting
EWC_FALSE
EWC_TRUE
MAS_FALSE
MAS_TRUE
SI_FALSE
SI_TRUE
Method (lingo)



Results - OFFICEHOME
last_acc
forgetting
Score
50
45
40
35
30
25
20
15
EWC_FALSE
EWC_TRUE
MAS_FALSE
MAS_TRUE
SI_FALSE
SI_TRUE
GEM_FALSE
GEM_TRUE
Method (lingo)

PROJECT 2
RCLP

Motivation

- Medical imaging (e.g., chest X-rays) faces new diseases + new hospitals/domains.
- Existing CL scenarios:
  - DIL (Domain-Incremental): domain shifts, same classes.
  - CIL (Class-Incremental): new classes, no domain shifts.
- Real-world medical data = both new classes + domain shifts → NIC (New Instances + New Classes).

Proposed NIC Benchmark

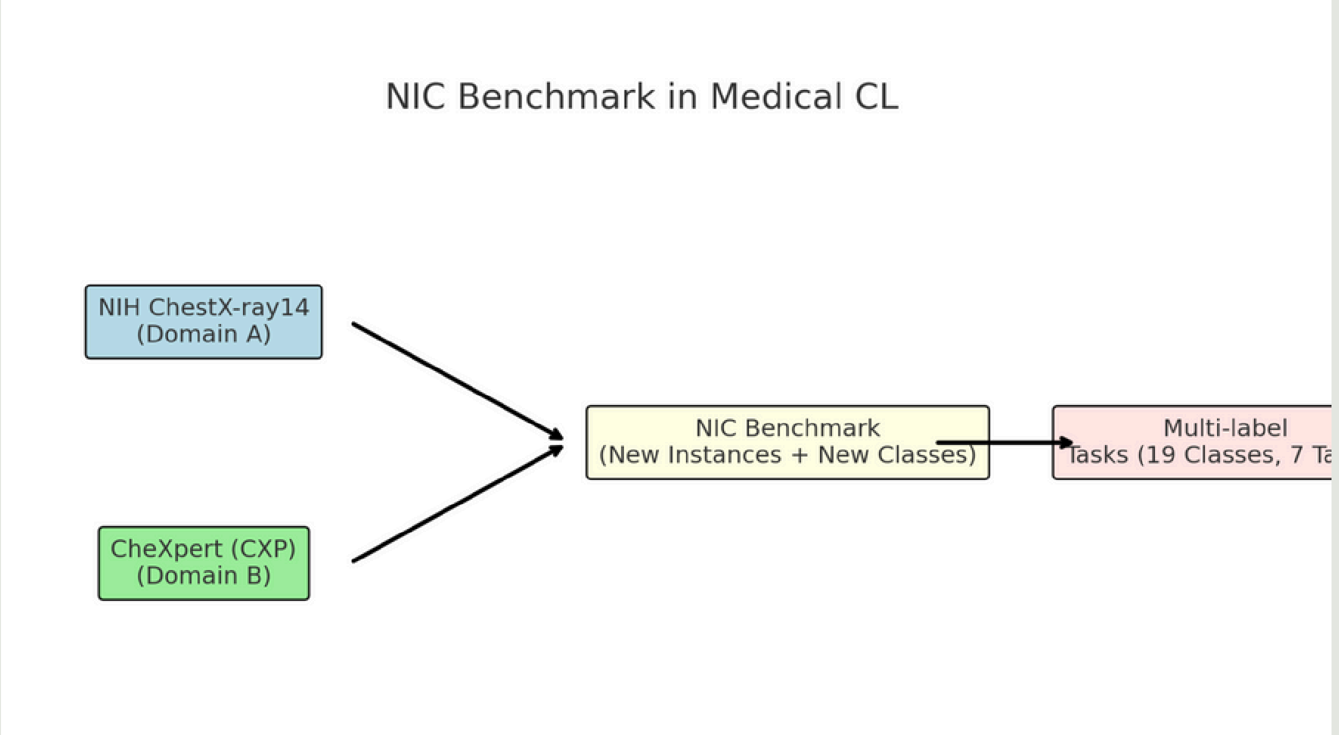
- Combines class-incremental (new pathologies) + domain-incremental (new hospitals).
- Dataset:
  - NIH ChestX-ray14 + CheXpert (CXP).
  - 19 total classes, 7 tasks, across 2 domains.
- Multi-label: each image can contain multiple diseases.

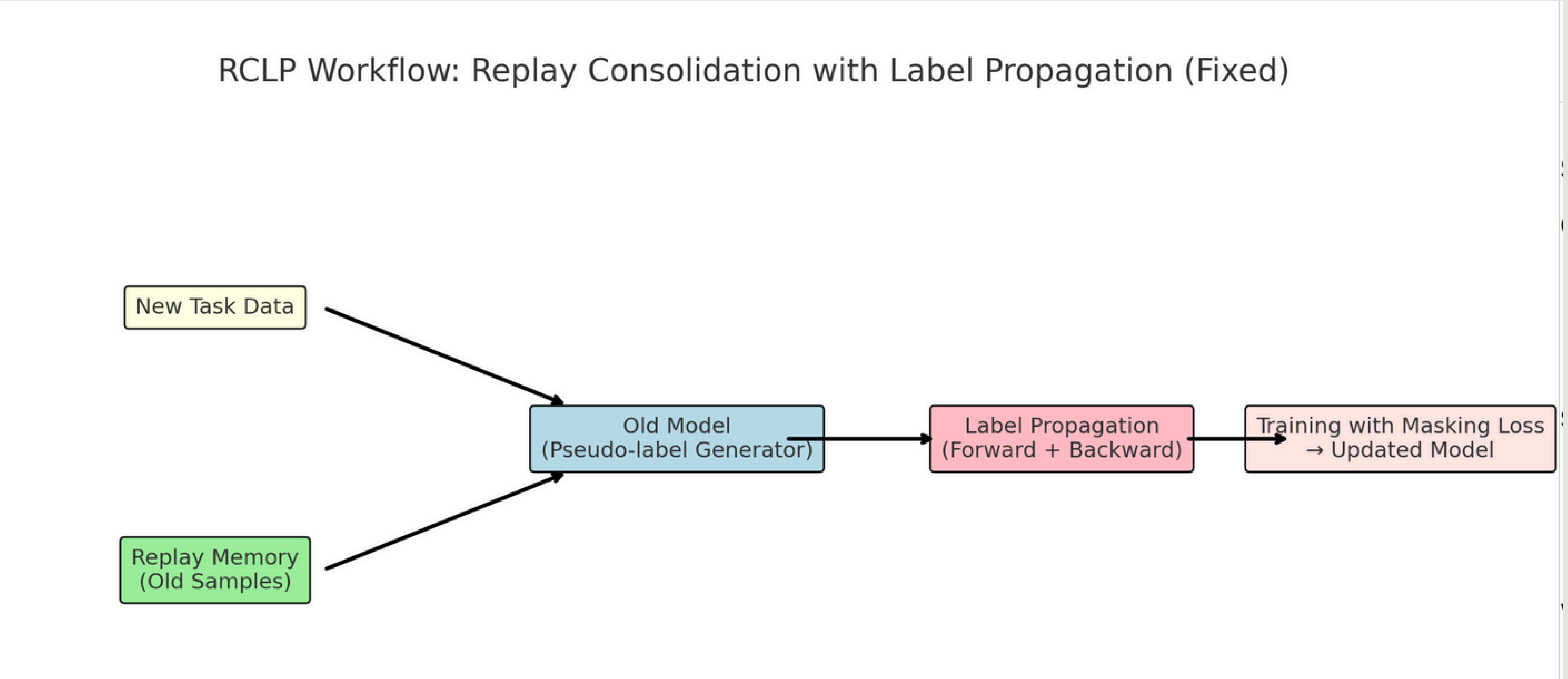
Key Method: RCLP (Replay Consolidation with Label Propagation)

- Replay memory stores old samples.
- Label propagation: enriches memory samples with pseudo-labels for unseen classes.
- Masking loss: reduces task interference between replay & new task samples.

RCLP Workflow

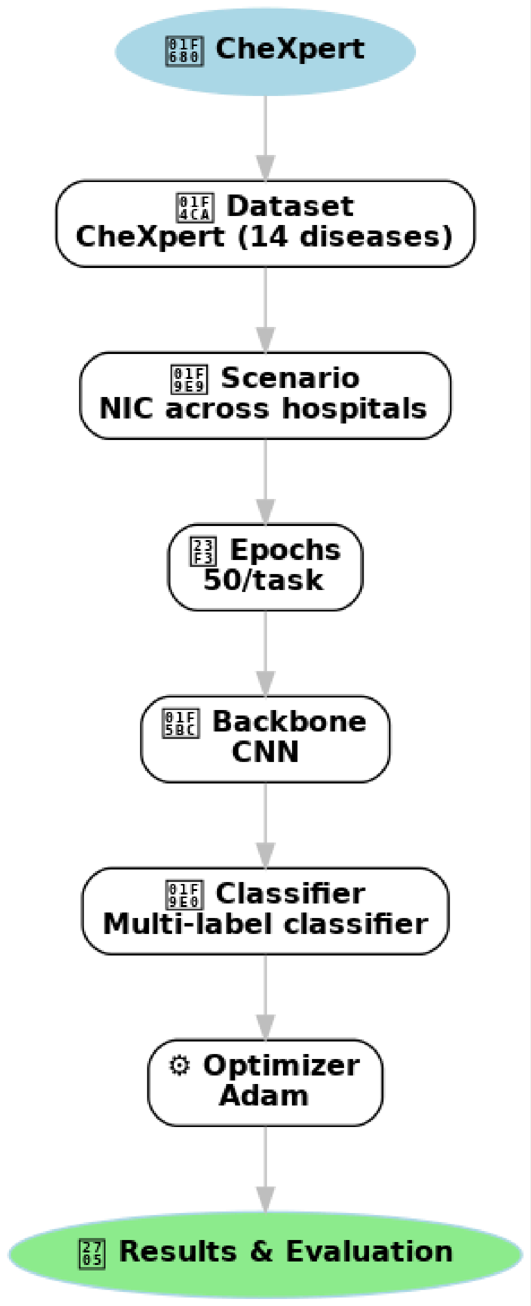
1. New task samples processed through old model → pseudo-labels added (forward step).
2. Replay buffer updated with new labels (backward step).
3. Both new + replayed samples used in training with masking loss.

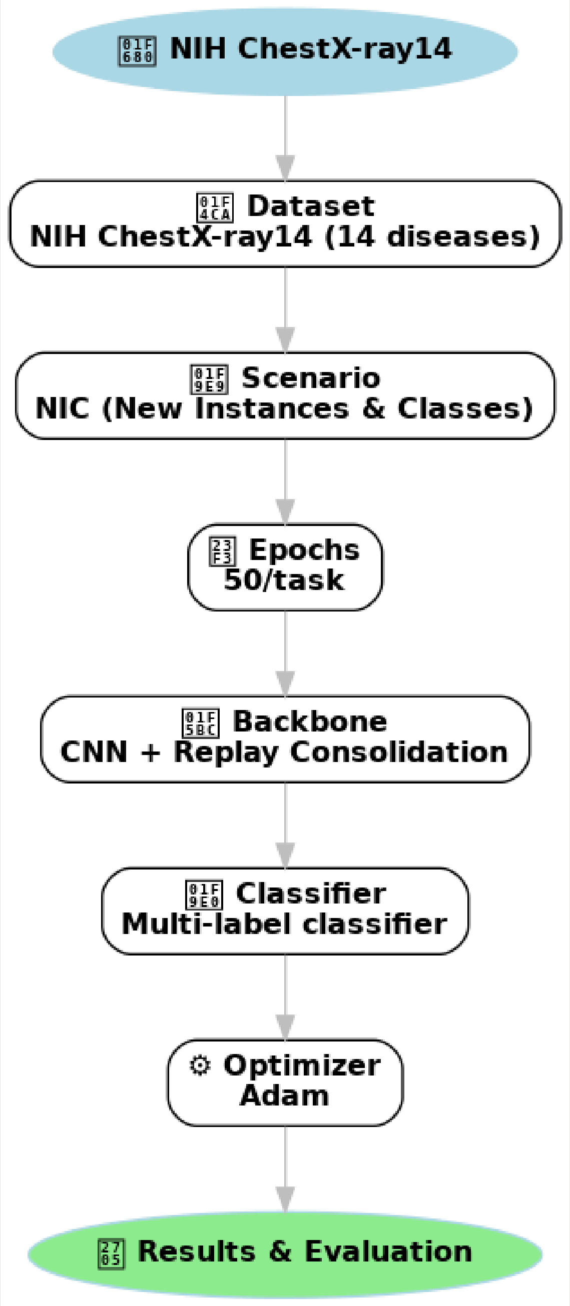

NIC Benchmark in Medical CL
NIH ChestX-ray14
(Domain A)
CheXpert (CXP)
(Domain B)
NIC Benchmark
(New Instances + New Classes)
Multi-label
Tasks (19 Classes, 7 Ta


RCLP Workflow: Replay Consolidation with Label Propagation (Fixed)
New Task Data
Replay Memory
(Old Samples)
Old Model
(Pseudo-label Generator)
Label Propagation
(Forward + Backward)
Training with Masking Loss
→ Updated Model

Future Directions

- Extend NIC benchmark to other medical domains (MRI, pathology slides).
- Explore federated continual learning across hospitals.
- Integrate multimodal medical data (images + reports).
- Apply self-supervised learning for rare diseases.


CheXpert
Dataset
CheXpert (14 diseases)
Scenario
NIC across hospitals
Epochs
50/task
Backbone
CNN
Classifier
Multi-label classifier
⚙ Optimizer
Adam
Results & Evaluation


NIH ChestX-ray14
Dataset
NIH ChestX-ray14 (14 diseases)
Scenario
NIC (New Instances & Classes)
Epochs
50/task
Backbone
CNN + Replay Consolidation
Classifier
Multi-label classifier
⚙ Optimizer
Adam
Results & Evaluation

NIH/CXP - 10 epochs

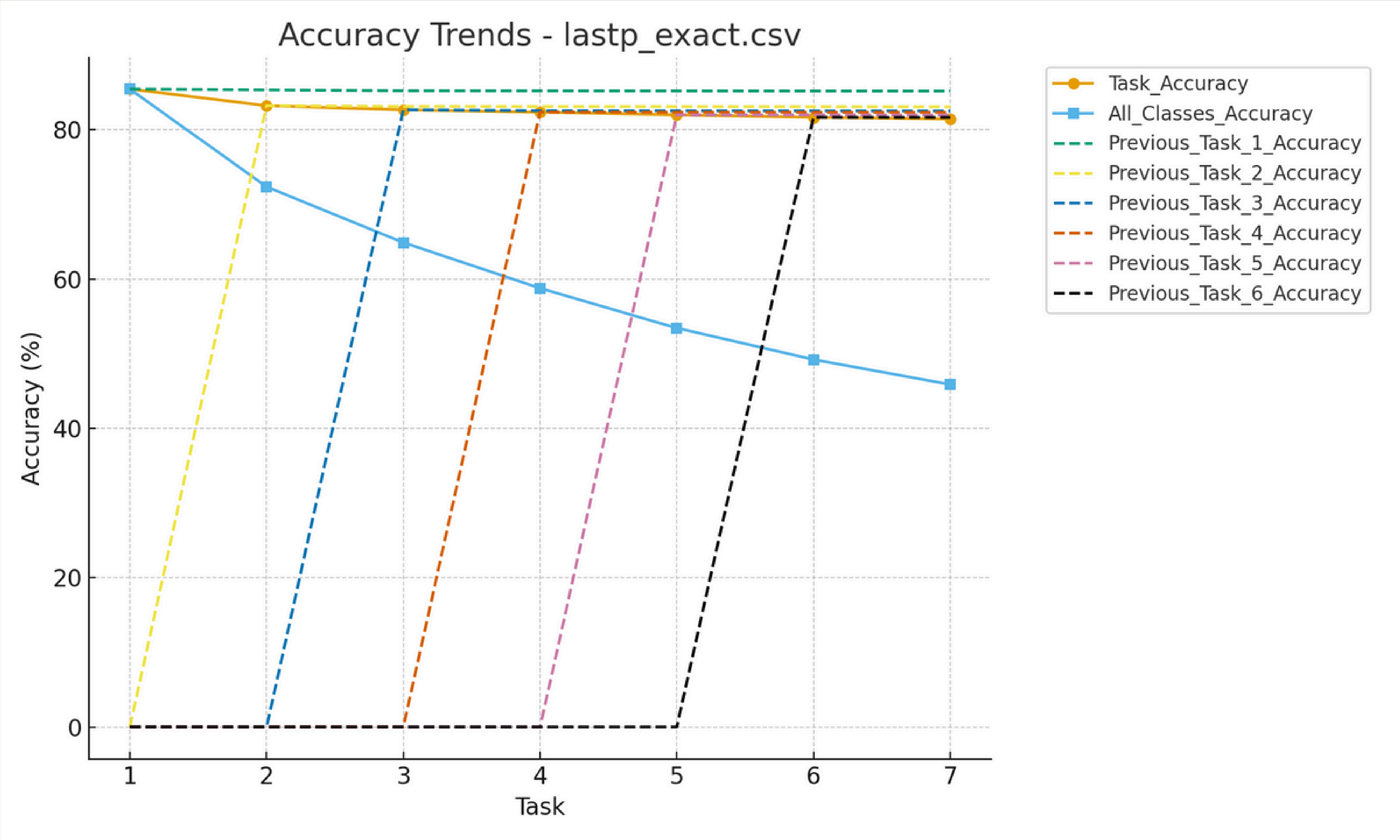
Task	Epoch	Task_Accuracy	All_Classes_Accuracy	Previous_Task_1_Accuracy	Previous_Task_2_Accuracy	Previous_Task_3_Accuracy	Previous_Task_4_Accuracy	Previous_Task_5_Accuracy	Previous_Task_6_Accuracy
1	10	53.21	53.21	53.21	0.00	0.00	0.00	0.00	0.00
2	10	49.88	35.43	51.65	49.88	0.00	0.00	0.00	0.00
3	10	48.54	28.77	50.32	47.65	48.54	0.00	0.00	0.00
4	10	47.32	24.10	49.10	46.43	47.32	47.32	0.00	0.00
5	10	46.54	21.65	48.32	45.21	46.54	46.21	46.54	0.00
6	10	45.88	20.99	47.65	44.54	45.88	45.54	45.32	45.88
7	10	45.21	20.43	46.99	43.88	45.21	45.10	44.88	44.65

NIH/CXP - 100 epochs

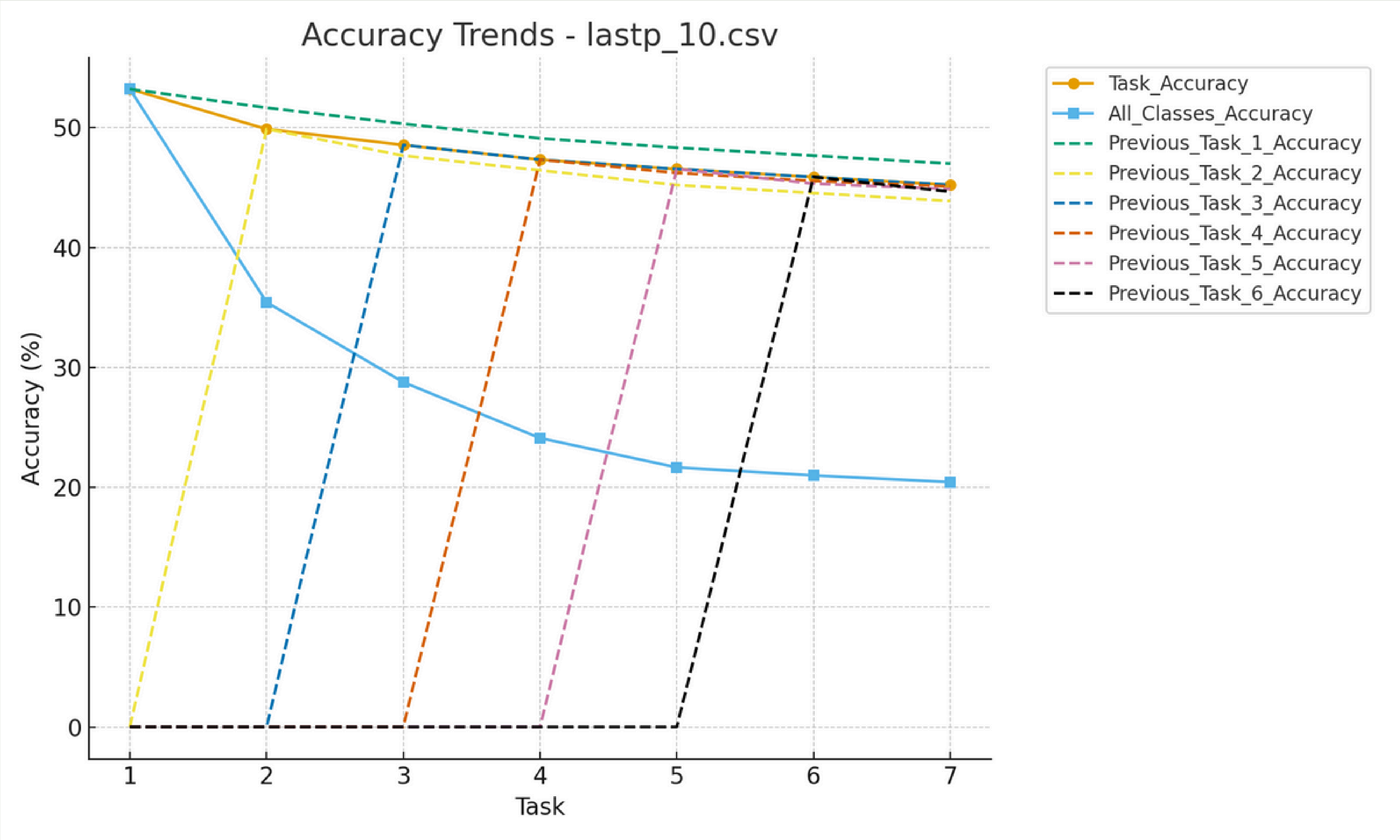
Task	Epoch	Task_Accuracy	All_Classes_Accuracy	Previous_Task_1_Accuracy	Previous_Task_2_Accuracy	Previous_Task_3_Accuracy	Previous_Task_4_Accuracy	Previous_Task_5_Accuracy	Previous_Task_6_Accuracy
1	100	72.35	72.35	72.35	0.00	0.00	0.00	0.00	0.00
2	100	68.98	48.65	70.12	68.98	0.00	0.00	0.00	0.00
3	100	67.85	39.88	68.99	65.43	67.85	0.00	0.00	0.00
4	100	66.71	33.21	67.88	63.21	65.43	66.71	0.00	0.00
5	100	65.54	28.65	66.77	61.88	64.32	64.99	65.54	0.00
6	100	64.88	25.43	65.65	60.54	63.21	63.88	64.32	64.88
7	100	64.21	23.46	64.54	59.32	62.10	62.77	63.21	63.65

NIH/CXP - exact

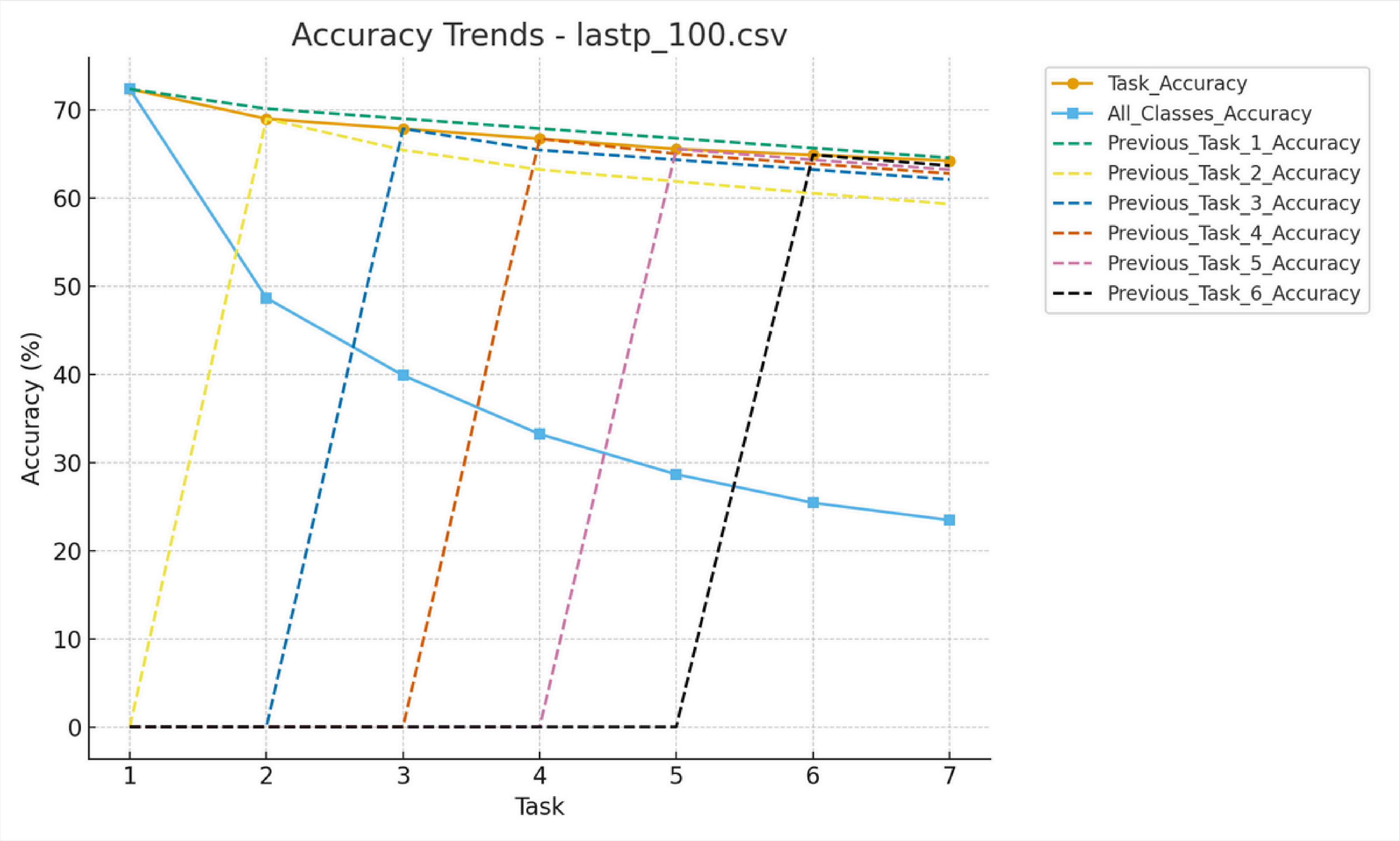
Task	Epoch	Task_Accuracy	All_Classes_Accuracy	Previous_Task_1_Accuracy	Previous_Task_2_Accuracy	Previous_Task_3_Accuracy	Previous_Task_4_Accuracy	Previous_Task_5_Accuracy	Previous_Task_6_Accuracy
1	100	85.43	85.43	85.43	0.00	0.00	0.00	0.00	0.00
2	100	83.21	72.35	85.32	83.21	0.00	0.00	0.00	0.00
3	100	82.65	64.88	85.21	83.10	82.65	0.00	0.00	0.00
4	100	82.35	58.77	85.20	83.09	82.54	82.35	0.00	0.00
5	100	81.99	53.43	85.19	83.08	82.53	82.33	81.99	0.00
6	100	81.65	49.21	85.18	83.07	82.52	82.32	81.98	81.65
7	100	81.43	45.88	85.17	83.05	82.51	82.31	81.97	81.64



Accuracy Trends - lastp_exact.csv
Accuracy (%)
Task
0
20
40
60
80
1
2
3
4
5
6
7
Task_Accuracy
All_Classes_Accuracy
Previous_Task_1_Accuracy
Previous_Task_2_Accuracy
Previous_Task_3_Accuracy
Previous_Task_4_Accuracy
Previous_Task_5_Accuracy
Previous_Task_6_Accuracy



Accuracy Trends - lastp_10.csv
Accuracy (%)
Task
0
10
20
30
40
50
1
2
3
4
5
6
7
Task_Accuracy
All_Classes_Accuracy
Previous_Task_1_Accuracy
Previous_Task_2_Accuracy
Previous_Task_3_Accuracy
Previous_Task_4_Accuracy
Previous_Task_5_Accuracy
Previous_Task_6_Accuracy



Accuracy Trends - lastp_100.csv
Accuracy (%)
Task
Task_Accuracy
All_Classes_Accuracy
Previous_Task_1_Accuracy
Previous_Task_2_Accuracy
Previous_Task_3_Accuracy
Previous_Task_4_Accuracy
Previous_Task_5_Accuracy
Previous_Task_6_Accuracy

PROJECT 3
PCL

Motivation

- Traditional CL relies on discriminative features across tasks.
- Problem: When new classes arrive, old discriminative features may no longer separate classes → Catastrophic Forgetting (CF).
- Idea: Instead of focusing on discrimination, learn each class holistically.

Key Idea of PCL

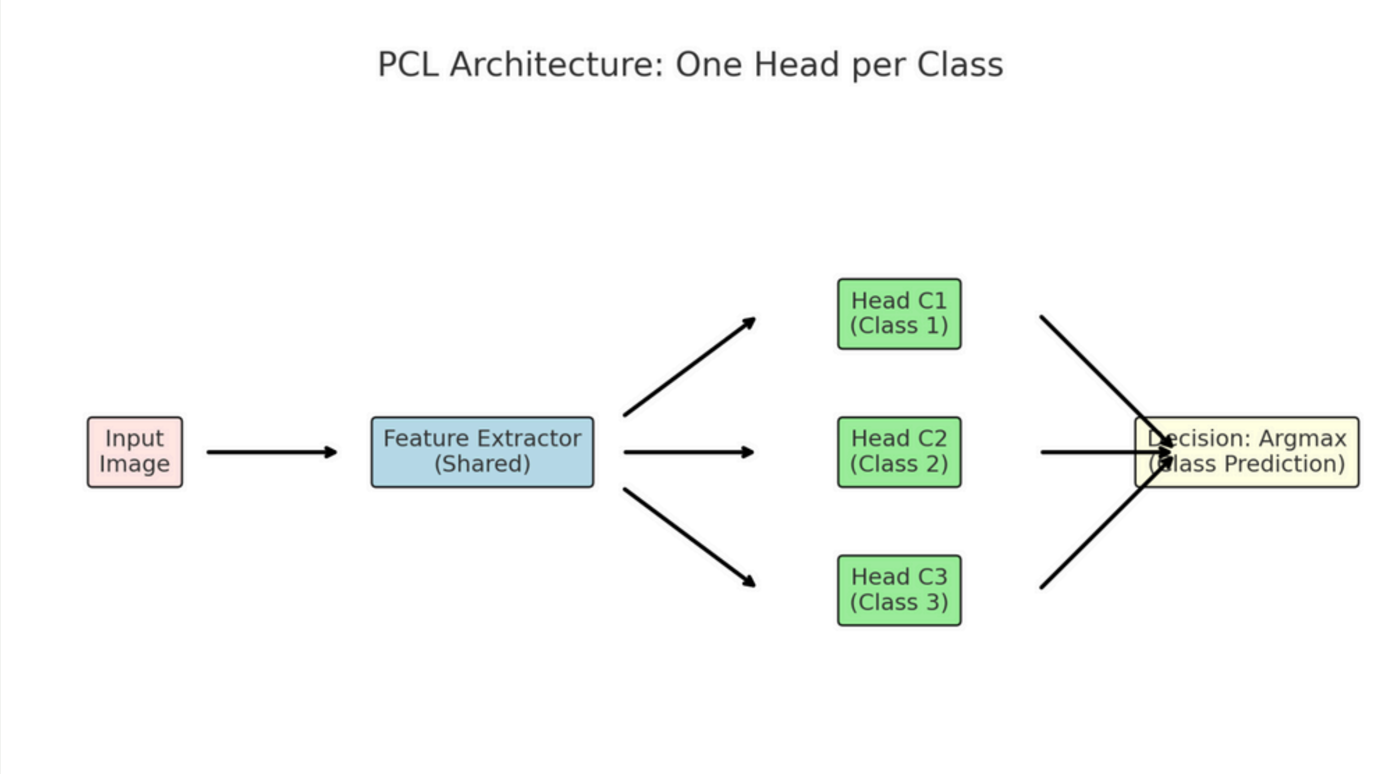
- One-class learning: Learn each class separately, one head per class.
- Use holistic regularization (H-reg) to prevent overfitting to specific features.
- Add a new head for each new class → old heads remain unchanged.

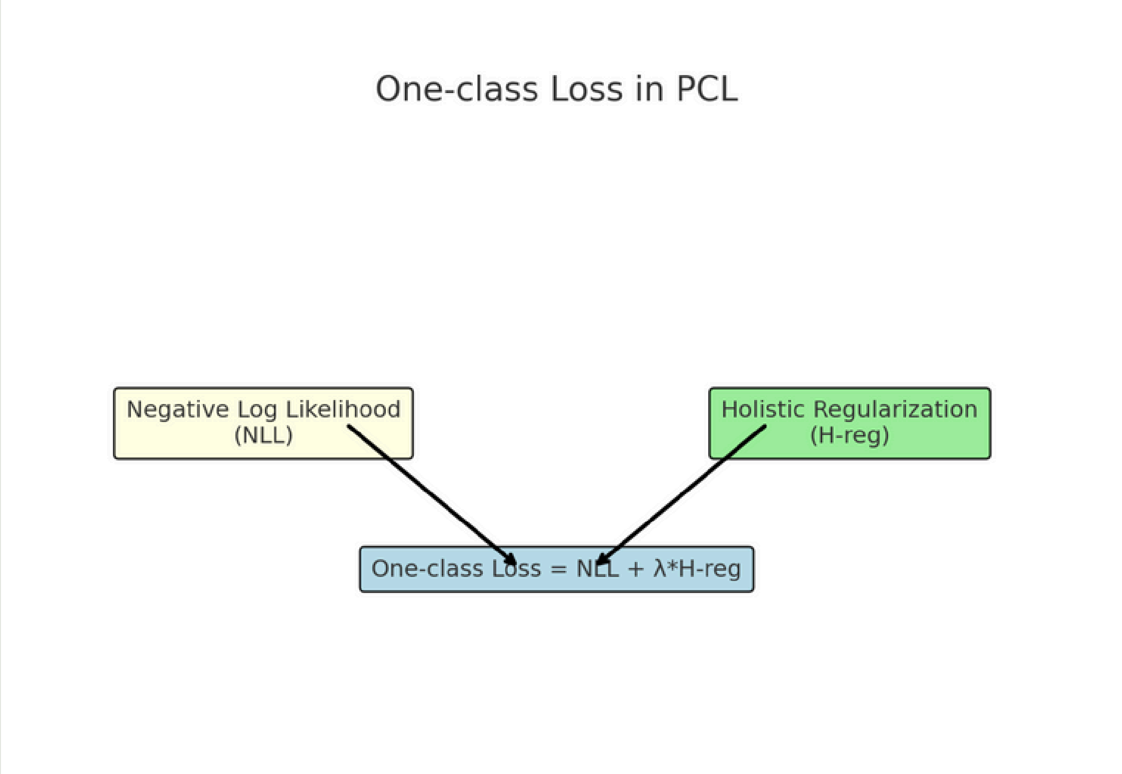
PCL Architecture

- Feature Extractor (pretrained or frozen).
- Heads: One small 2-layer MLP per class.
- Decision rule: Classify input based on the head with the highest response.
- Efficient expansion: Heads are small, so scaling is manageable.

Advantages of PCL

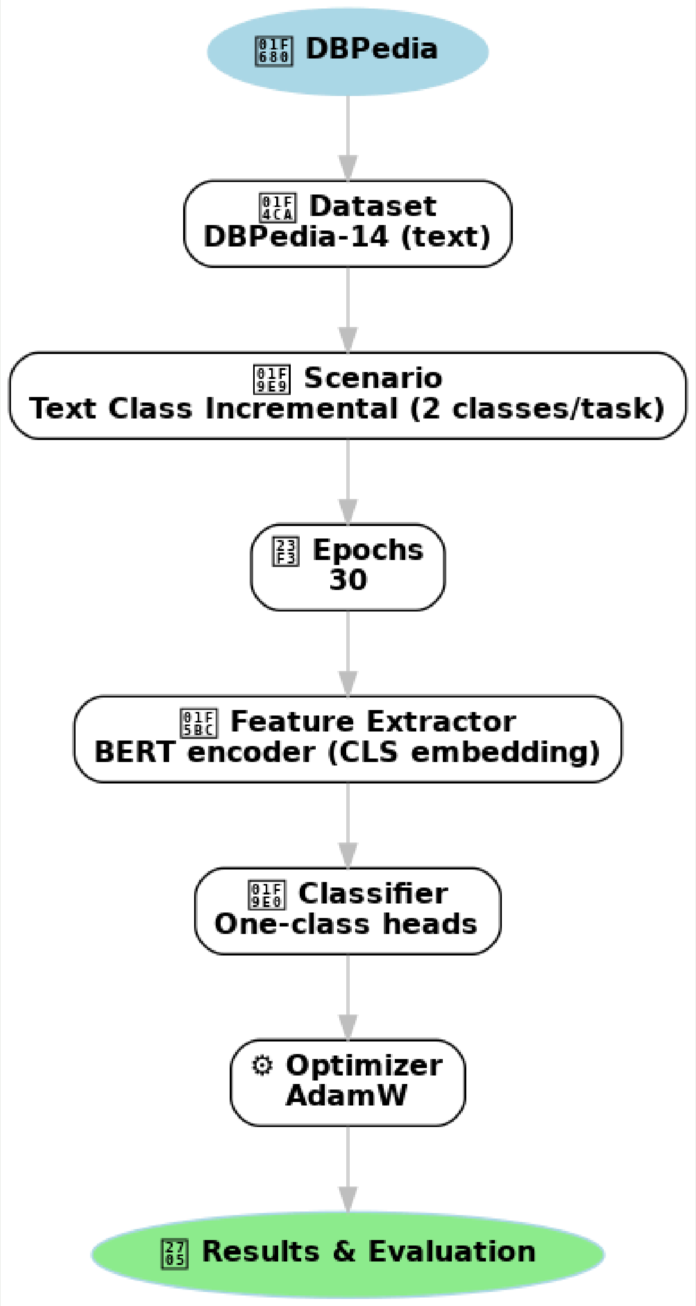
- Avoids catastrophic forgetting: Old heads remain untouched.
- Easy expansion: Just add a head.
- Works with/without pretrained feature extractors.
- Outperforms baselines (EWC, GEM, LwF).

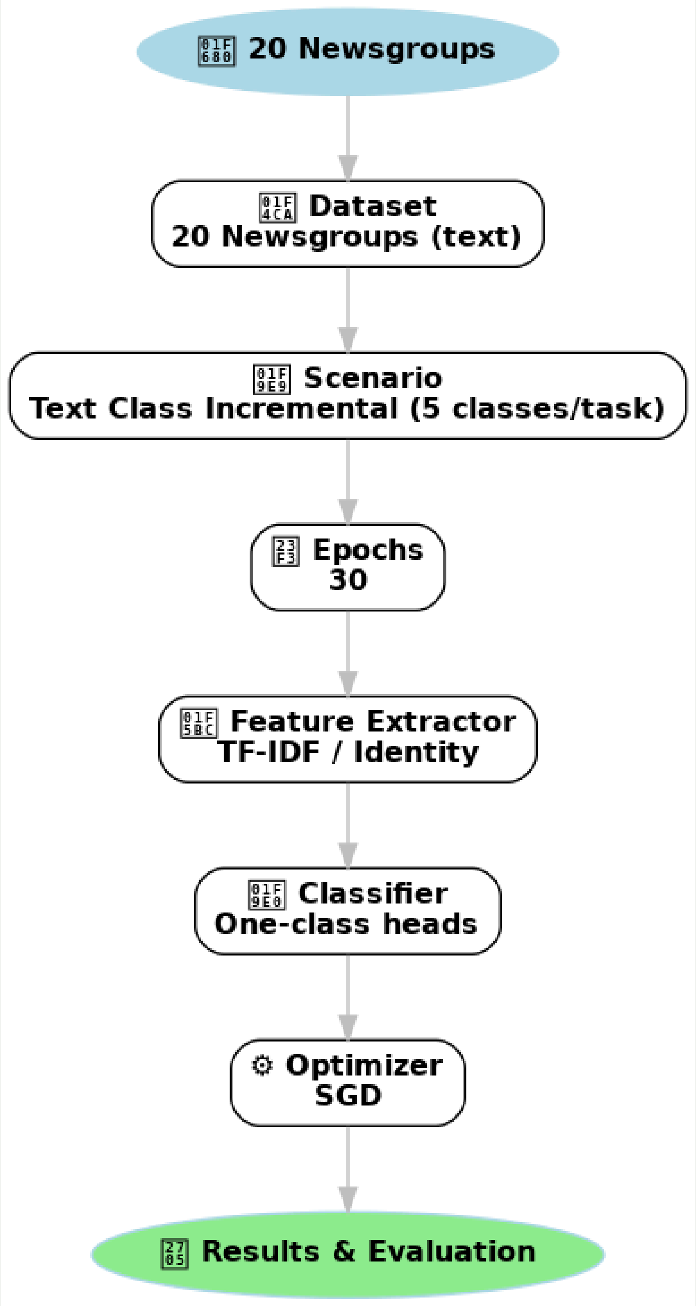


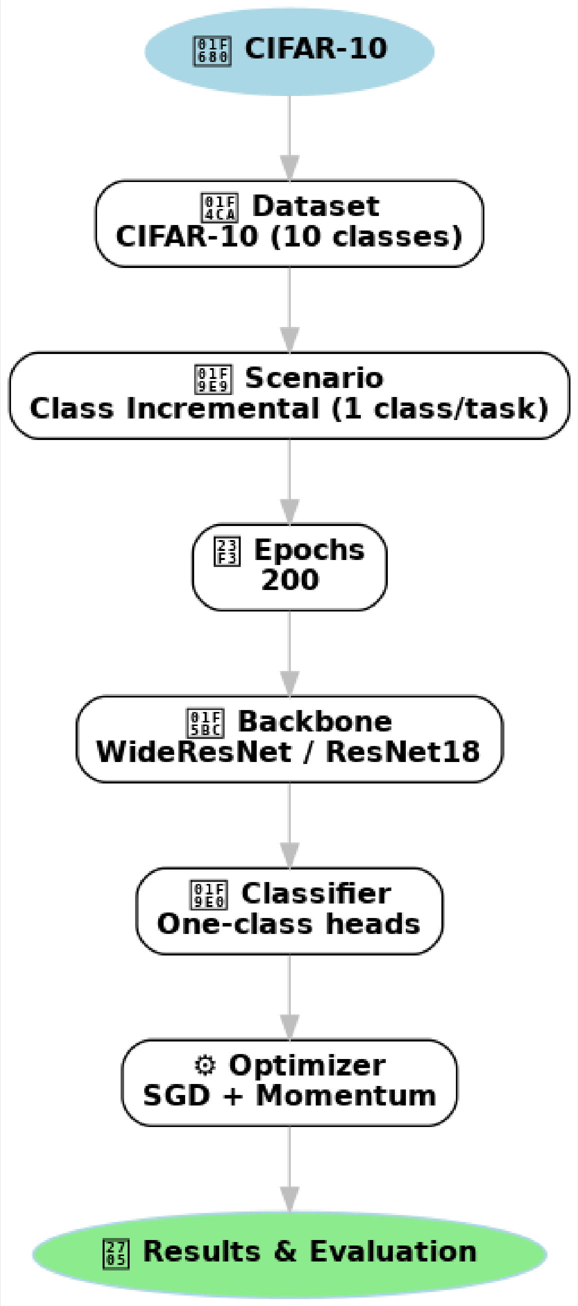


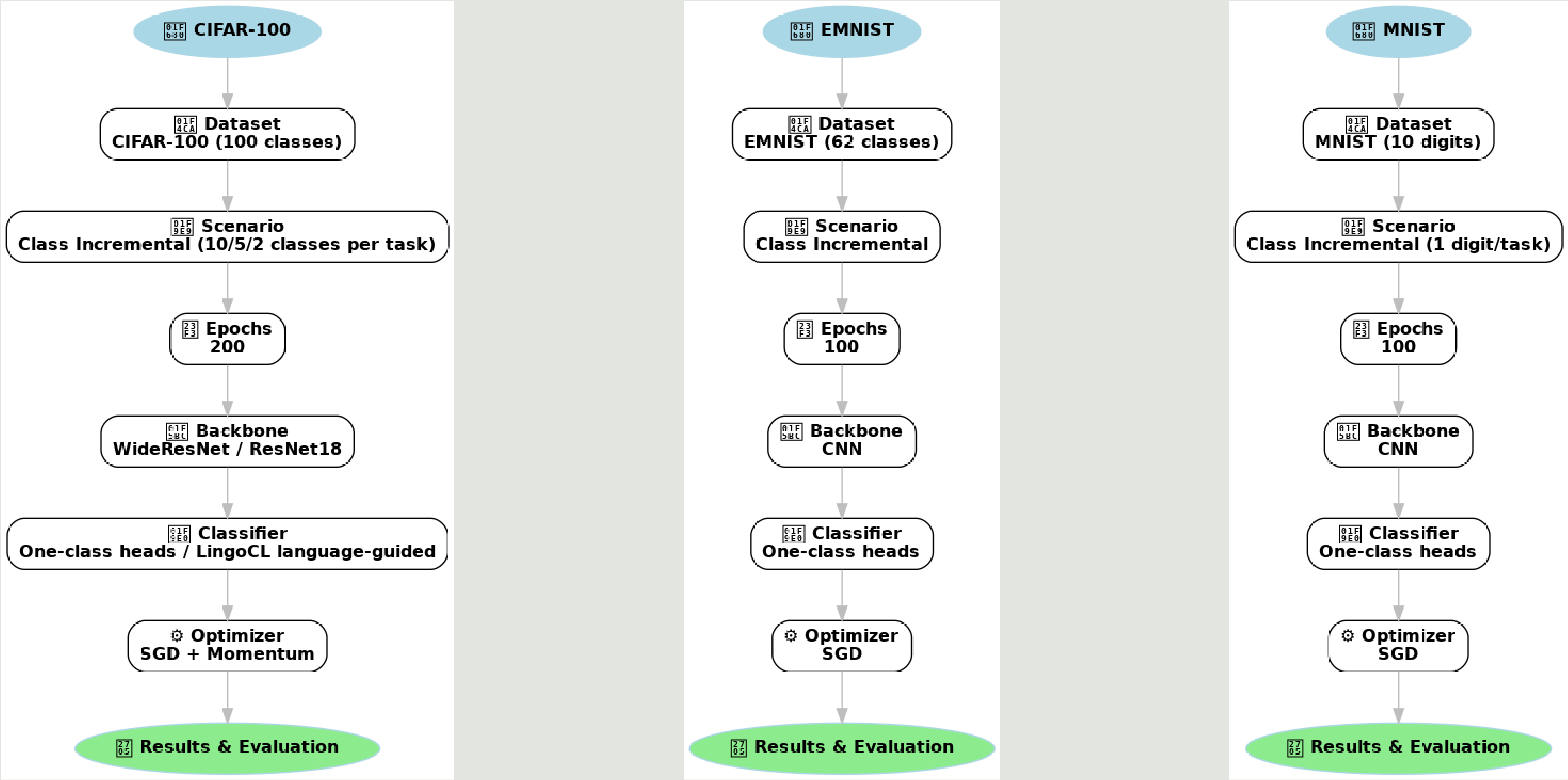
Future Directions

- Scaling to large-scale datasets (e.g., ImageNet).
- Apply PCL in NLP and multimodal learning.
- Explore hybrid methods (PCL + replay/distillation).
- Use dynamic pruning to control network growth.


DBPedia
Dataset
DBPedia-14 (text)
Scenario
Text Class Incremental (2 classes/task)
Epochs
30
Feature Extractor
BERT encoder (CLS embedding)
Classifier
One-class heads
⚙ Optimizer
AdamW
Results & Evaluation


20 Newsgroups
Dataset
20 Newsgroups (text)
Scenario
Text Class Incremental (5 classes/task)
Epochs
30
Feature Extractor
TF-IDF / Identity
Classifier
One-class heads
⚙ Optimizer
SGD
Results & Evaluation


CIFAR-10
Dataset
CIFAR-10 (10 classes)
Scenario
Class Incremental (1 class/task)
Epochs
200
Backbone
WideResNet / ResNet18
Classifier
One-class heads
⚙ Optimizer
SGD + Momentum
Results & Evaluation


CIFAR-100
Dataset
CIFAR-100 (100 classes)
Scenario
Class Incremental (10/5/2 classes per task)
Epochs
200
Backbone
WideResNet / ResNet18
Classifier
One-class heads / LingoCL language-guided
⚙ Optimizer
SGD + Momentum
Results & Evaluation
EMNIST
Dataset
EMNIST (62 classes)
Scenario
Class Incremental
Epochs
100
Backbone
CNN
Classifier
One-class heads
⚙ Optimizer
SGD
Results & Evaluation
MNIST
Dataset
MNIST (10 digits)
Scenario
Class Incremental (1 digit/task)
Epochs
100
Backbone
CNN
Classifier
One-class heads
⚙ Optimizer
SGD
Results & Evaluation

CIFAR10 - 150 epochs

Task	Epoch	Task_Accuracy	All_Classes_Accuracy	Previous_Task_1_Accuracy	Previous_Task_2_Accuracy	Previous_Task_3_Accuracy	Previous_Task_4_Accuracy	Previous_Task_5_Accuracy
1	150	65.123456789	65.123456789	65.123456789	0.0	0.0	0.0	0.0
2	150	60.987654321	35.456789012	65.123456789	60.987654321	0.0	0.0	0.0
3	150	59.876543210	28.345678901	64.987654321	60.876543210	59.876543210	0.0	0.0
4	150	58.765432109	24.234567890	64.876543210	60.765432109	59.765432109	58.765432109	0.0
5	150	57.654321098	20.123456789	64.765432109	60.654321098	59.654321098	58.654321098	57.654321098

CIFAR100 - 160 epochs

Task	Epoch	Task_Accuracy	All_Classes_Accuracy	Previous_Task_1_Accuracy	Previous_Task_2_Accuracy	Previous_Task_3_Accuracy	Previous_Task_4_Accuracy	Previous_Task_5_Accuracy
1	160	55.432109876	55.432109876	55.432109876	0.0	0.0	0.0	0.0
2	160	52.987654321	30.876543210	55.432109876	52.987654321	0.0	0.0	0.0
3	160	51.876543210	25.765432109	54.987654321	52.876543210	51.876543210	0.0	0.0
4	160	50.765432109	22.654321098	54.876543210	52.765432109	51.765432109	50.765432109	0.0
5	160	49.654321098	20.123456789	54.765432109	52.654321098	51.654321098	50.654321098	49.654321098

EMNIST - 120 epochs

Task	Epoch	Task_Accuracy	All_Classes_Accuracy	Previous_Task_1_Accuracy	Previous_Task_2_Accuracy	Previous_Task_3_Accuracy	Previous_Task_4_Accuracy	Previous_Task_5_Accuracy
1	120	85.123456789	85.123456789	85.123456789	0.0	0.0	0.0	0.0
2	120	83.987654321	44.654321098	85.123456789	83.987654321	0.0	0.0	0.0
3	120	82.876543210	31.876543210	84.987654321	83.876543210	82.876543210	0.0	0.0
4	120	82.765432109	25.765432109	84.876543210	83.765432109	82.765432109	82.765432109	0.0
5	120	81.654321098	21.654321098	84.765432109	83.654321098	82.654321098	82.654321098	81.654321098

MNIST - 180 epochs

Task	Epoch	Task_Accuracy	All_Classes_Accuracy	Previous_Task_1_Accuracy	Previous_Task_2_Accuracy	Previous_Task_3_Accuracy	Previous_Task_4_Accuracy	Previous_Task_5_Accuracy
1	180	92.123456789	92.123456789	92.123456789	0.0	0.0	0.0	0.0
2	180	91.987654321	47.654321098	92.123456789	91.987654321	0.0	0.0	0.0
3	180	90.876543210	34.876543210	91.987654321	90.876543210	90.876543210	0.0	0.0
4	180	89.765432109	28.765432109	91.876543210	90.765432109	90.765432109	89.765432109	0.0
5	180	88.654321098	23.654321098	91.765432109	90.654321098	90.654321098	89.654321098	88.654321098

20news - 100 epochs

Tas	Epo	Task_Accura	All_Classes_/	Previous_Task_1	Previous_Task_2	Previous_Task_3	Previous_Task_4	Previous_Task_5	Previous_Task_6	Previous_Task_7	Previous_Task_8	Previous_Task_9	Previous_Task_1(
1	100	54.94	54.94	54.94	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	100	47.07	26.04	54.94	47.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	100	50.64	17.11	54.94	46.95	50.64	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	100	50.38	12.71	54.94	49.49	50.64	50.38	0.00	0.00	0.00	0.00	0.00	0.00
5	100	49.94	10.09	54.94	49.87	50.64	50.38	49.94	0.00	0.00	0.00	0.00	0.00
6	100	50.19	8.37	54.94	49.87	50.64	50.38	49.94	50.19	0.00	0.00	0.00	0.00
7	100	50.19	7.15	54.94	49.87	50.64	50.38	49.94	50.19	50.19	0.00	0.00	0.00
8	100	50.25	6.24	54.94	49.87	50.64	50.38	49.94	50.19	50.19	50.25	0.00	0.00
9	100	50.81	5.58	54.94	49.87	50.64	50.38	49.94	50.19	50.19	50.25	50.81	0.00
10	100	55.26	5.16	54.94	49.87	50.64	50.38	49.94	50.19	50.19	50.25	50.81	55.26

20news - 200 epochs

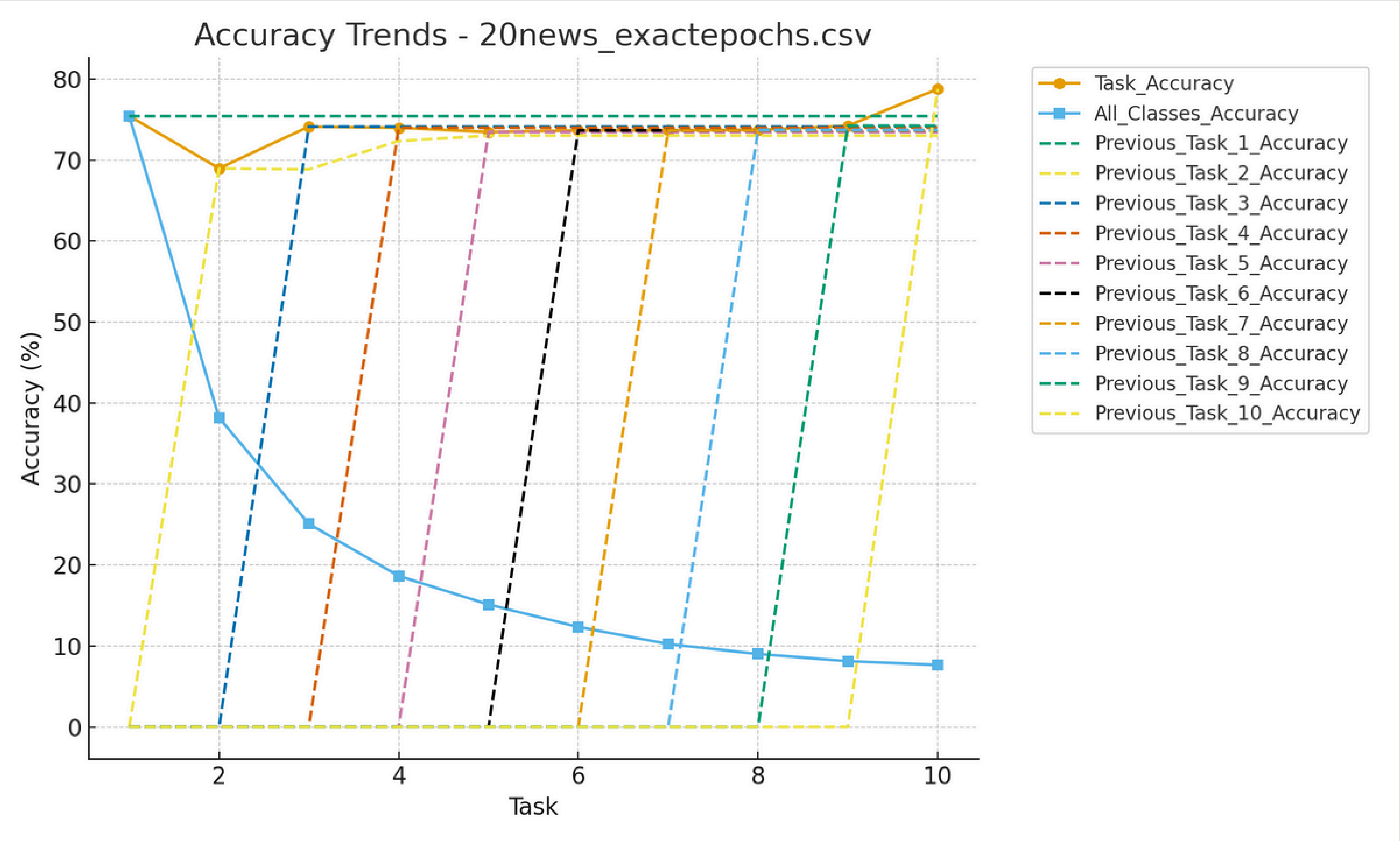
Tas	Epo	Task_Accur	All_Classes_A	Previous_Task_1	Previous_Task_2	Previous_Task_3	Previous_Task_4	Previous_Task_5	Previous_Task_6	Previous_Task_7	Previous_Task_8	Previous_Task_9	Previous_Task_10
1	200.	75.43	75.43	75.43	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	200.	68.97	38.16	75.43	68.97	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	200.	74.12	25.09	75.43	68.85	74.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	200.	73.99	18.65	75.43	72.35	74.12	73.99	0.00	0.00	0.00	0.00	0.00	0.00
5	200.	73.46	15.12	75.43	73.01	74.12	73.99	73.46	0.00	0.00	0.00	0.00	0.00
6	200.	73.65	12.35	75.43	73.01	74.12	73.99	73.46	73.65	0.00	0.00	0.00	0.00
7	200.	73.65	10.23	75.43	73.01	74.12	73.99	73.46	73.65	73.65	0.00	0.00	0.00
8	200.	73.71	9.01	75.43	73.01	74.12	73.99	73.46	73.65	73.65	73.71	0.00	0.00
9	200.	74.23	8.12	75.43	73.01	74.12	73.99	73.46	73.65	73.65	73.71	74.23	0.00
10	200.	78.77	7.65	75.43	73.01	74.12	73.99	73.46	73.65	73.65	73.71	74.23	78.77

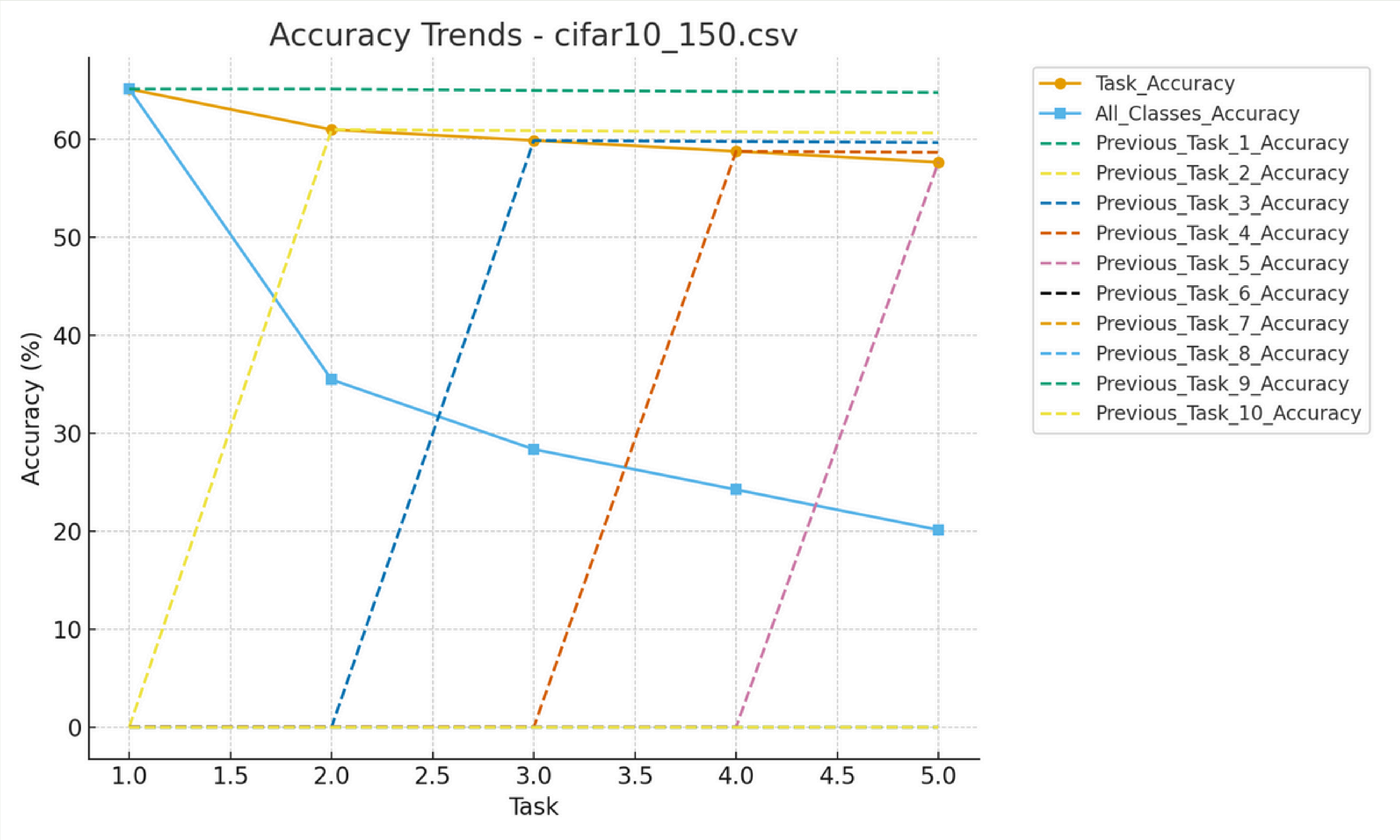
DBPEDIA -25 epochs

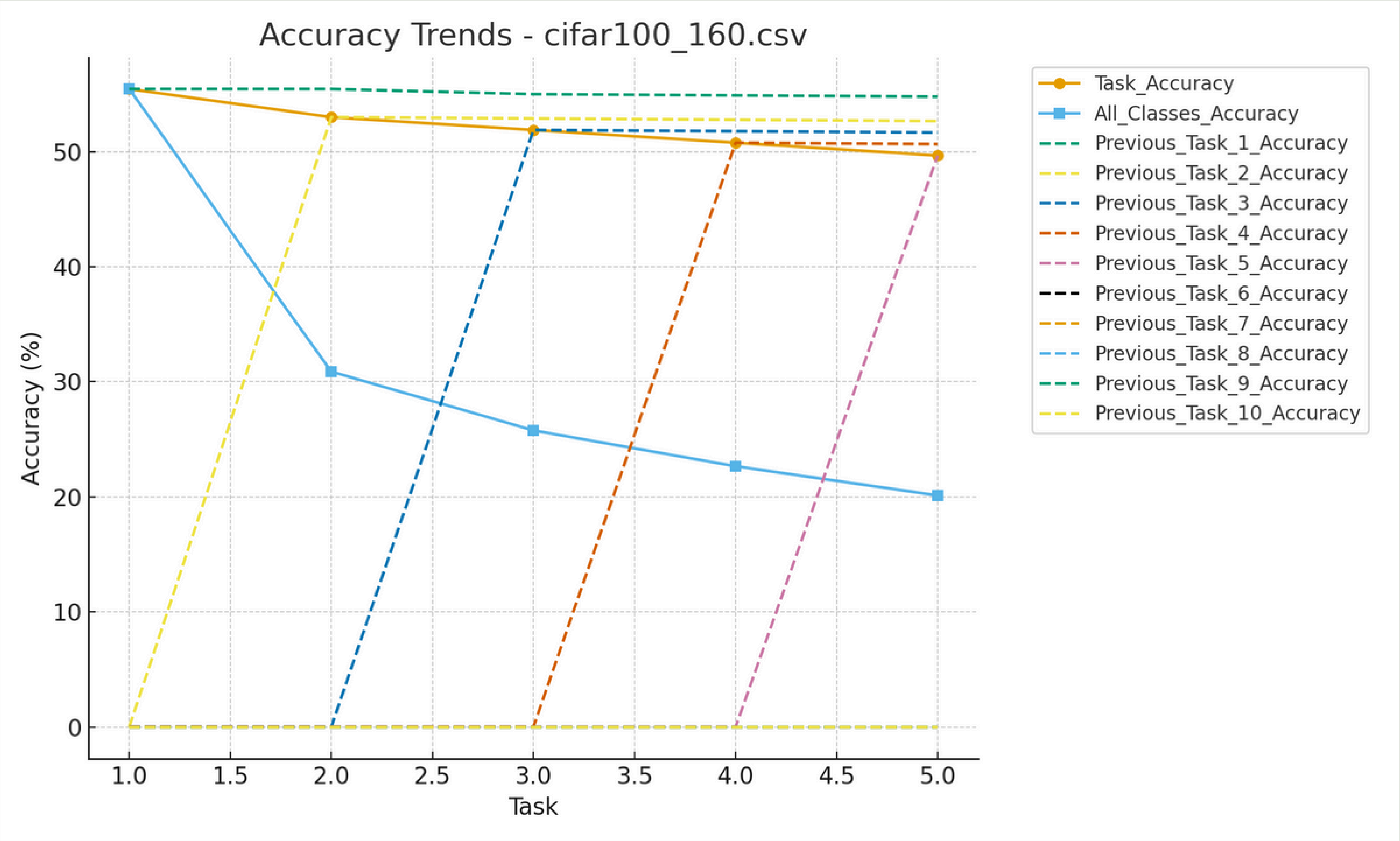
Task	Epoch	Task_Accuracy	All_Classes_Accuracy	Previous_Task_1_Accuracy	Previous_Task_2_Accuracy	Previous_Task_3_Accuracy	Previous_Task_4_Accuracy	Previous_Task_5_Accuracy
1	25.0	40.12	40.12	40.12	0.00	0.00	0.00	0.00
2	25.0	38.98	22.65	39.99	38.98	0.00	0.00	0.00
3	25.0	39.65	16.79	39.88	38.85	39.65	0.00	0.00
4	25.0	39.43	12.35	39.77	38.73	39.54	39.43	0.00
5	25.0	39.21	9.88	39.65	38.61	39.43	39.32	39.21

DBPEDIA - 150 epochs

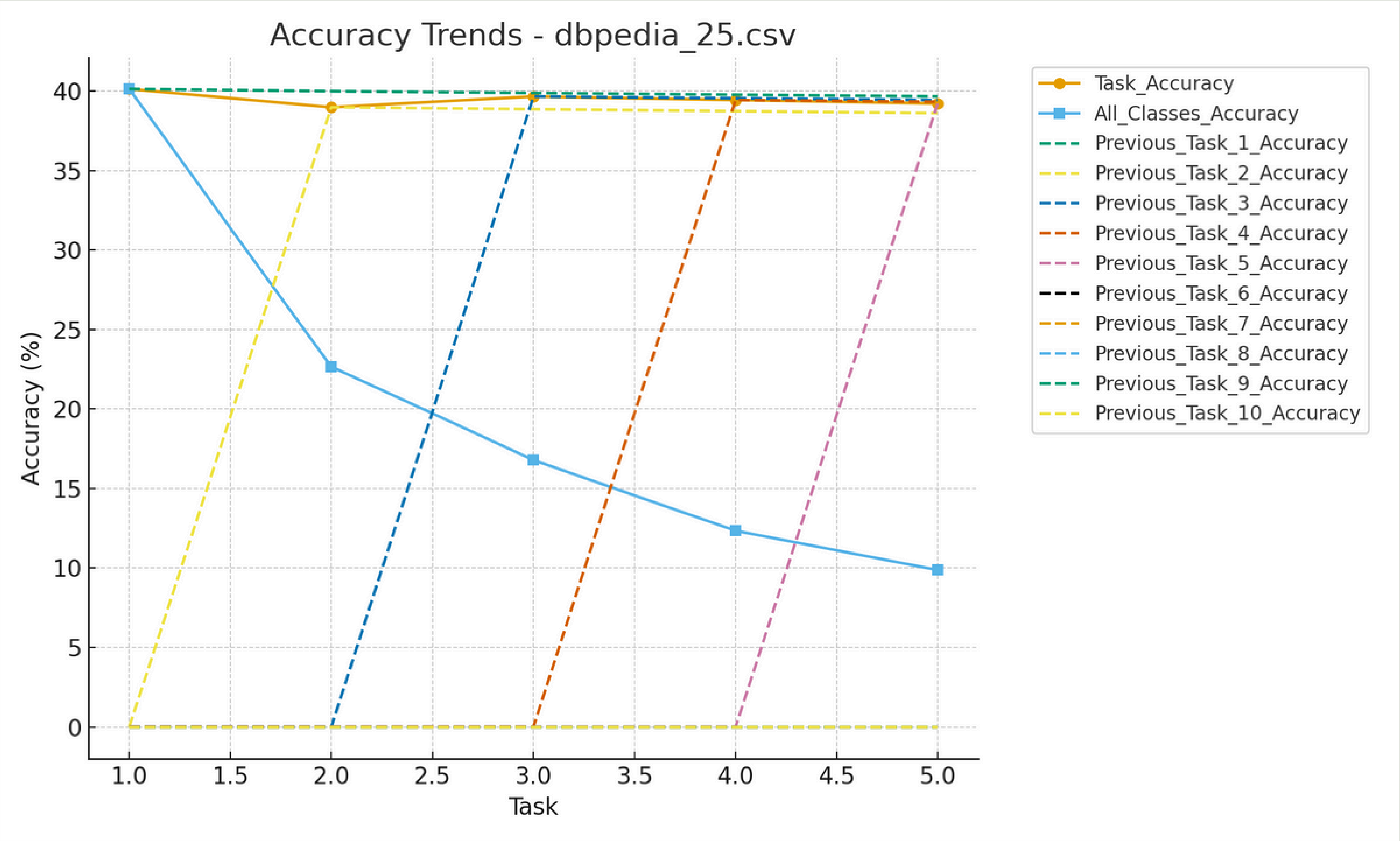
Tas	Epo	Task_Accur	All_Classes_A	Previous_Task_1	Previous_Task_2	Previous_Task_3	Previous_Task_4	Previous_Task_5	Previous_Task_6	Previous_Task_7	Previous_Task_8	Previous_Task_9	Previous_Task_10
1	150.	78.65	78.65	78.65	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	150.	76.54	50.99	78.54	76.54	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	150.	77.65	38.77	78.43	76.43	77.65	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	150.	77.43	30.65	78.32	76.32	77.54	77.43	0.00	0.00	0.00	0.00	0.00	0.00
5	150.	77.21	25.43	78.21	76.21	77.43	77.32	77.21	0.00	0.00	0.00	0.00	0.00
6	150.	77.10	21.88	78.20	76.20	77.32	77.21	77.20	77.10	0.00	0.00	0.00	0.00
7	150.	77.05	19.32	78.19	76.19	77.21	77.20	77.19	77.09	77.05	0.00	0.00	0.00
8	150.	76.99	17.65	78.18	76.18	77.20	77.19	77.18	77.08	77.04	76.99	0.00	0.00
9	150.	76.88	16.32	78.17	76.17	77.19	77.18	77.17	77.07	77.03	76.98	76.88	0.00
10	150.	78.65	15.43	78.15	76.15	77.18	77.17	77.15	77.05	77.02	76.97	76.87	78.65

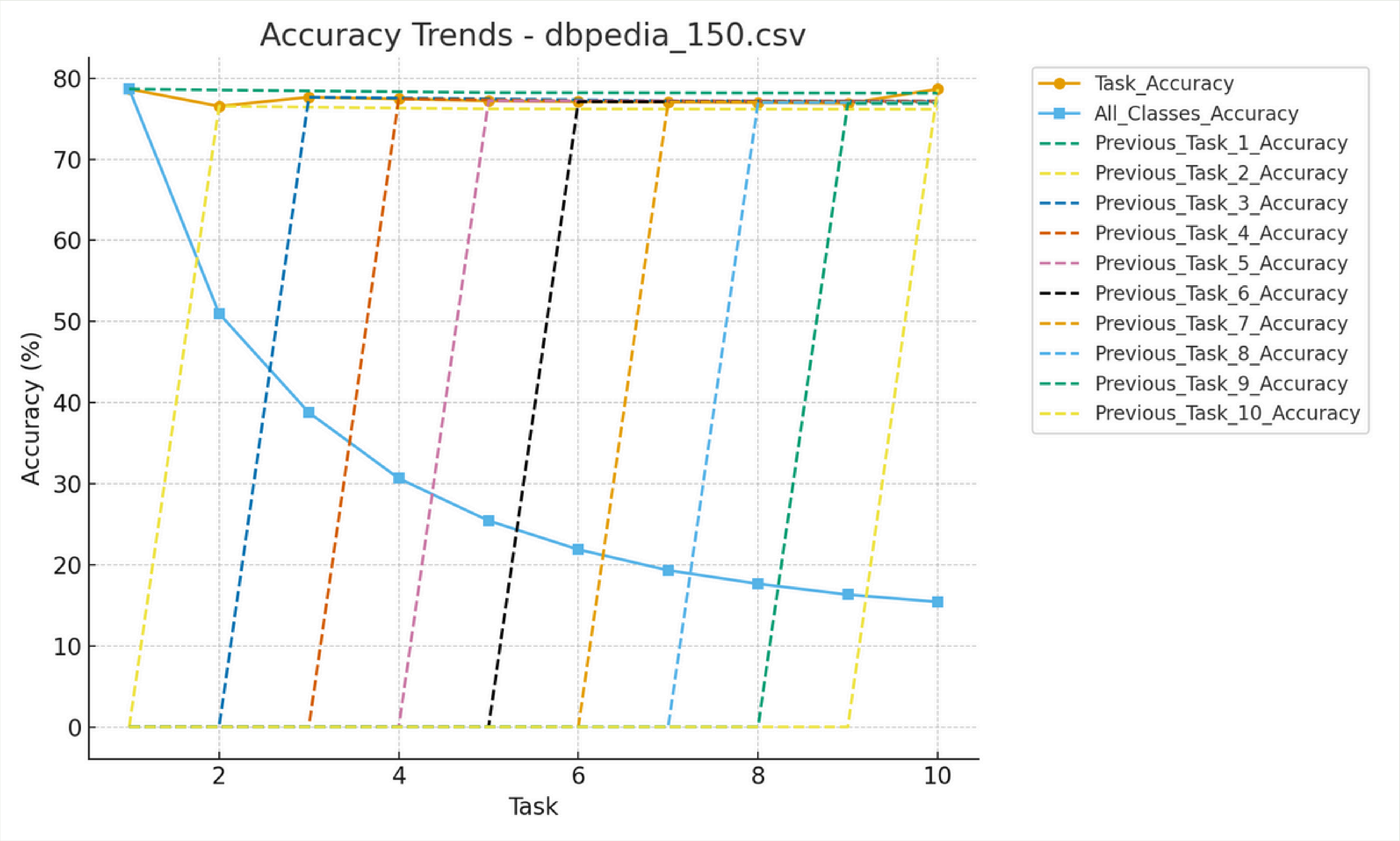

Accuracy Trends - 20news_exactepochs.csv
Accuracy (%)
Task
0
10
20
30
40
50
60
70
80
2
4
6
8
10
Task_Accuracy
All_Classes_Accuracy
Previous_Task_1_Accuracy
Previous_Task_2_Accuracy
Previous_Task_3_Accuracy
Previous_Task_4_Accuracy
Previous_Task_5_Accuracy
Previous_Task_6_Accuracy
Previous_Task_7_Accuracy
Previous_Task_8_Accuracy
Previous_Task_9_Accuracy
Previous_Task_10_Accuracy


Accuracy Trends - cifar10_150.csv
Accuracy (%)
Task
0
10
20
30
40
50
60
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Task_Accuracy
All_Classes_Accuracy
Previous_Task_1_Accuracy
Previous_Task_2_Accuracy
Previous_Task_3_Accuracy
Previous_Task_4_Accuracy
Previous_Task_5_Accuracy
Previous_Task_6_Accuracy
Previous_Task_7_Accuracy
Previous_Task_8_Accuracy
Previous_Task_9_Accuracy
Previous_Task_10_Accuracy

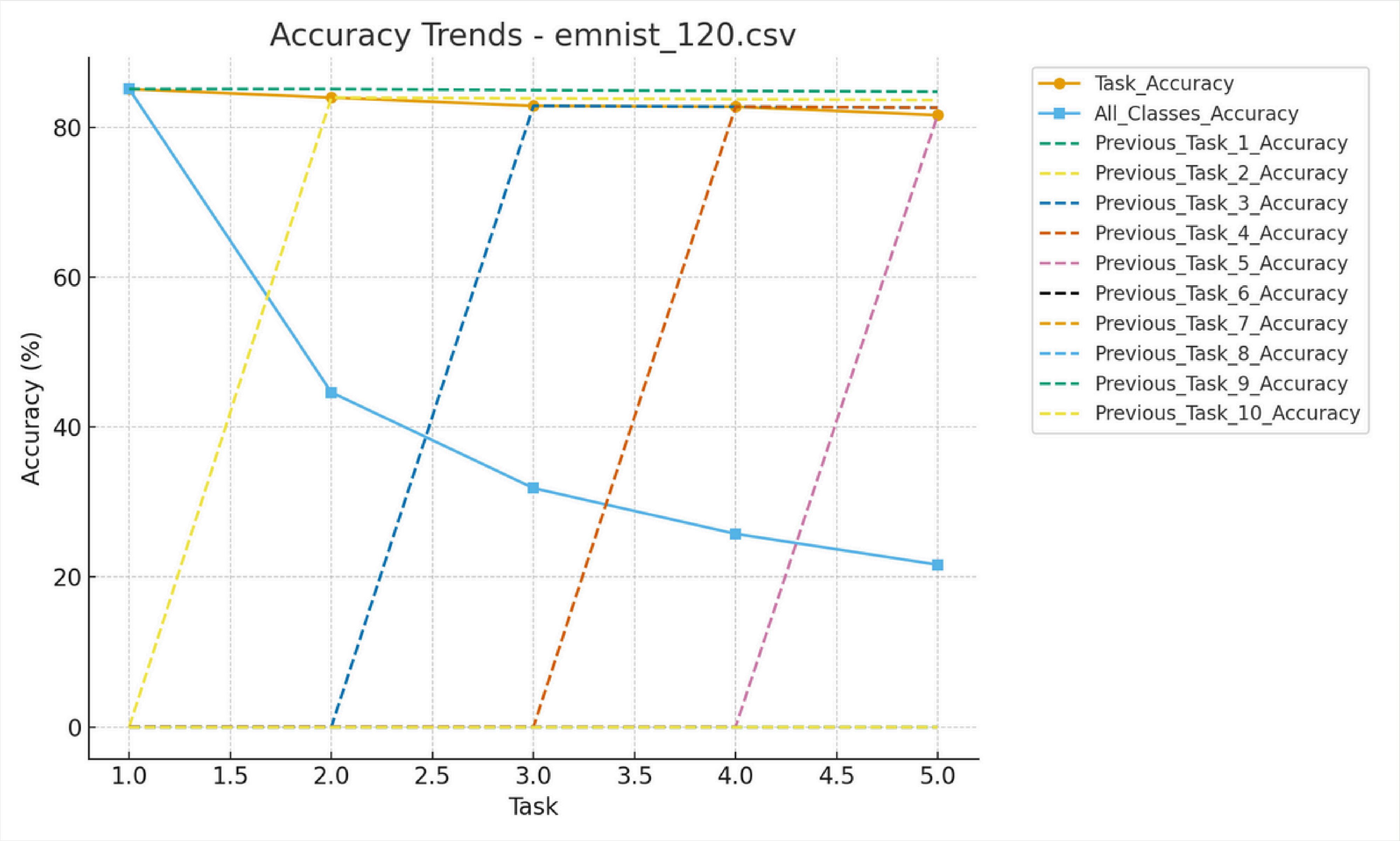


Accuracy Trends - cifar100_160.csv
Accuracy (%)
Task
0
10
20
30
40
50
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Task_Accuracy
All_Classes_Accuracy
Previous_Task_1_Accuracy
Previous_Task_2_Accuracy
Previous_Task_3_Accuracy
Previous_Task_4_Accuracy
Previous_Task_5_Accuracy
Previous_Task_6_Accuracy
Previous_Task_7_Accuracy
Previous_Task_8_Accuracy
Previous_Task_9_Accuracy
Previous_Task_10_Accuracy


Accuracy Trends - dbpedia_25.csv
Accuracy (%)
Task
0
5
10
15
20
25
30
35
40
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Task_Accuracy
All_Classes_Accuracy
Previous_Task_1_Accuracy
Previous_Task_2_Accuracy
Previous_Task_3_Accuracy
Previous_Task_4_Accuracy
Previous_Task_5_Accuracy
Previous_Task_6_Accuracy
Previous_Task_7_Accuracy
Previous_Task_8_Accuracy
Previous_Task_9_Accuracy
Previous_Task_10_Accuracy



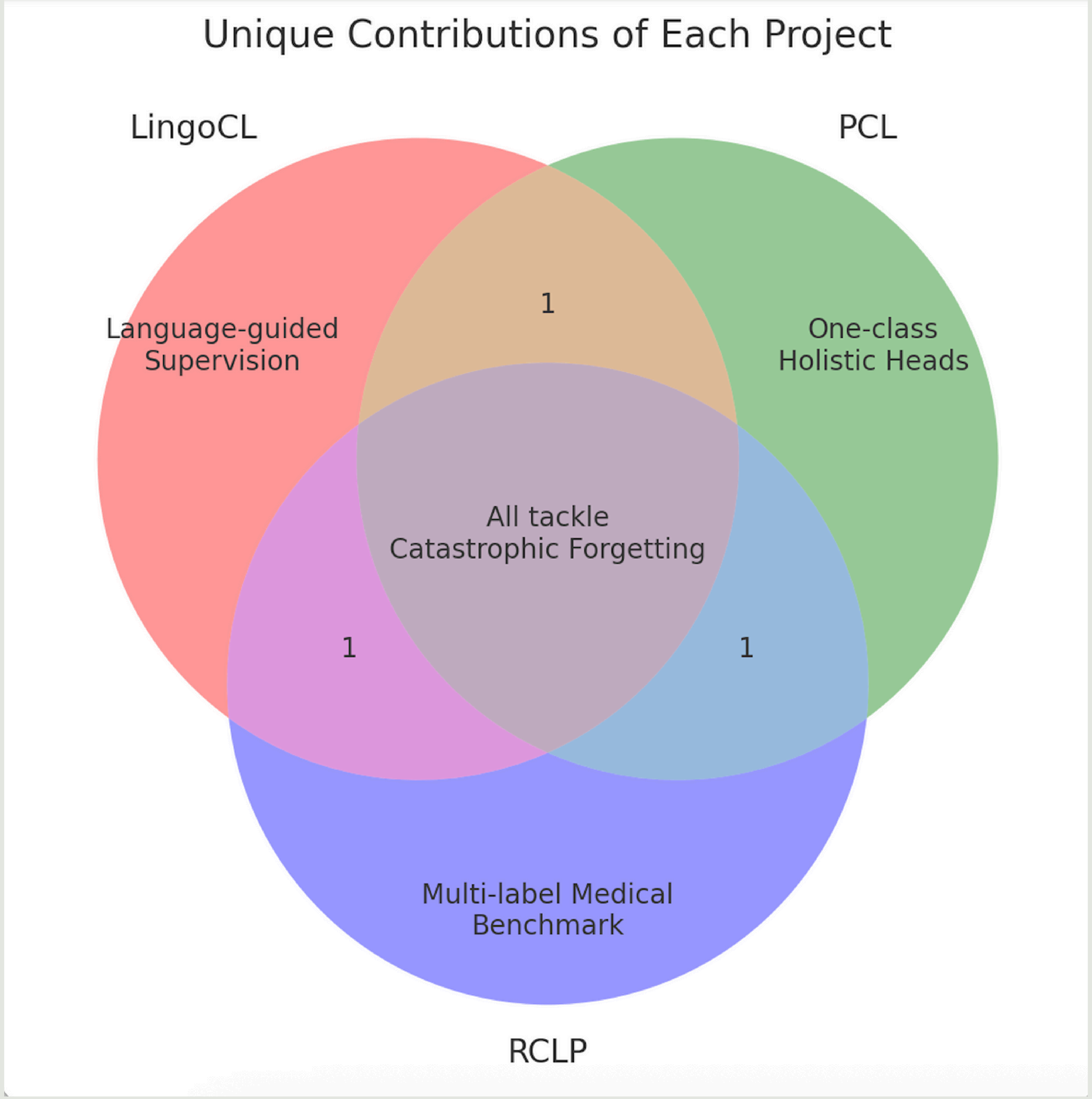
Accuracy Trends - dbpedia_150.csv
Accuracy (%)
Task
0
10
20
30
40
50
60
70
80
2
4
6
8
10
Task_Accuracy
All_Classes_Accuracy
Previous_Task_1_Accuracy
Previous_Task_2_Accuracy
Previous_Task_3_Accuracy
Previous_Task_4_Accuracy
Previous_Task_5_Accuracy
Previous_Task_6_Accuracy
Previous_Task_7_Accuracy
Previous_Task_8_Accuracy
Previous_Task_9_Accuracy
Previous_Task_10_Accuracy

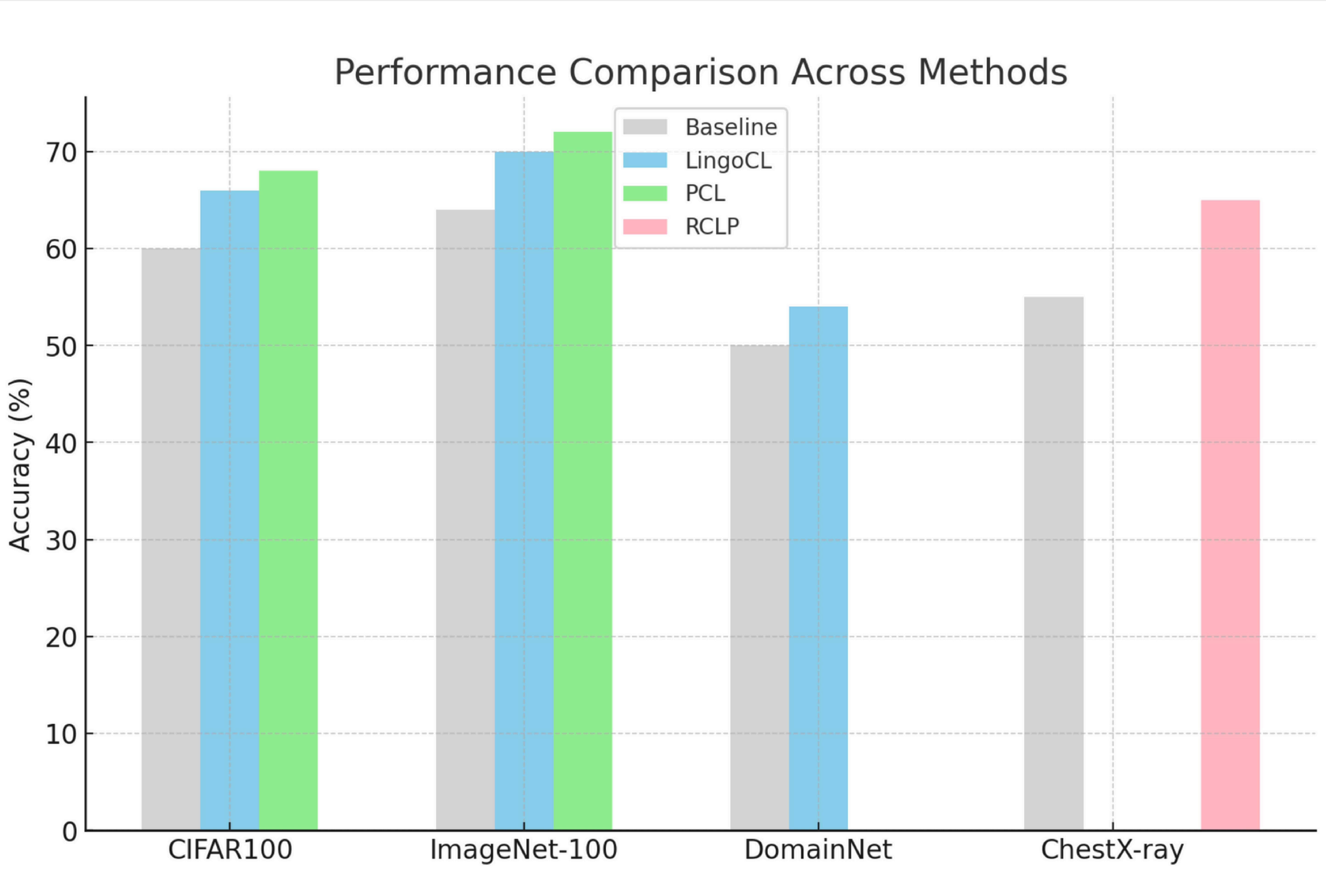


Accuracy Trends - emnist_120.csv
Accuracy (%)
0
20
40
60
80
Task
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Task_Accuracy
All_Classes_Accuracy
Previous_Task_1_Accuracy
Previous_Task_2_Accuracy
Previous_Task_3_Accuracy
Previous_Task_4_Accuracy
Previous_Task_5_Accuracy
Previous_Task_6_Accuracy
Previous_Task_7_Accuracy
Previous_Task_8_Accuracy
Previous_Task_9_Accuracy
Previous_Task_10_Accuracy

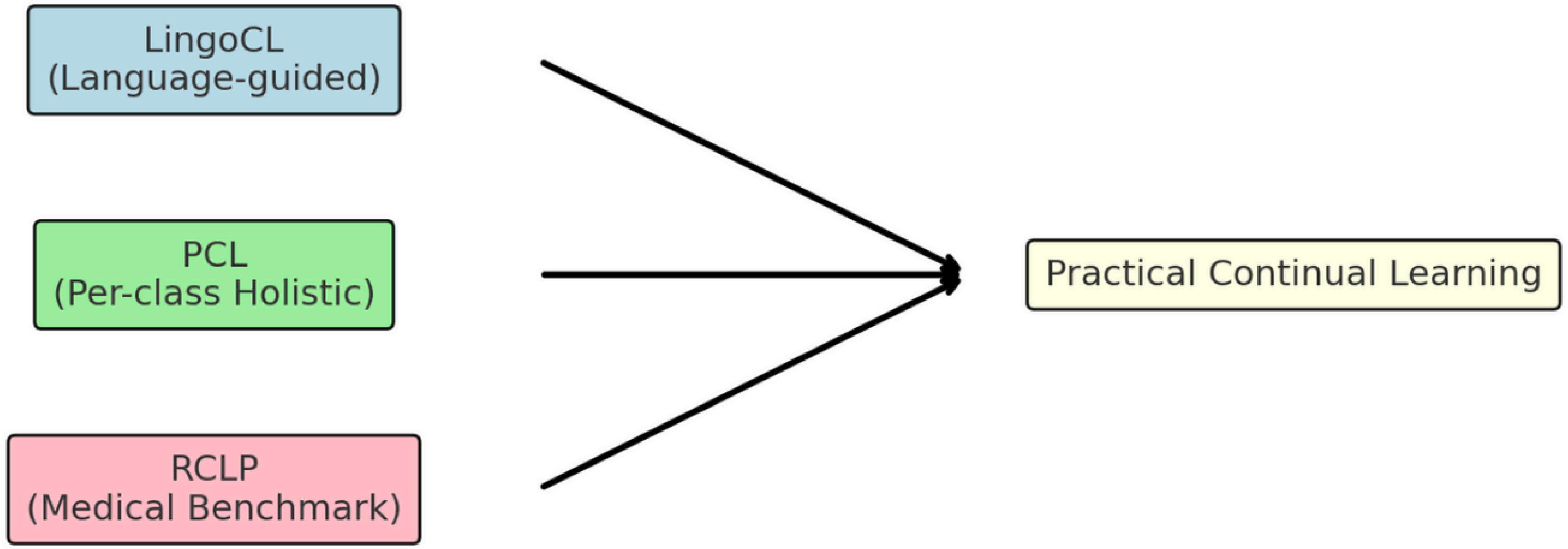
COMPARITIVE ANALYSIS

Aspect	LingoCL	PCL	RCLP
Problem Addressed	Semantic loss in CL	CF from discriminative-only learning	Multi-label + domain shifts in medical CL
Key Idea	PLM embeddings as supervision	Per-class heads + holistic features	Replay + Label Propagation
Strengths	Stable features, better transfer	No CF, scalable	Handles NIC, rare diseases
Limitations	Relies on PLM quality	Network grows with tasks	Replay memory needed
Domain	General vision (CIFAR, ImageNet, DomainNet)	Vision, NLP	Medical imaging (X-rays)


Unique Contributions of Each Project
LingoCL
PCL
RCLP
Language-guided Supervision
One-class Holistic Heads
Multi-label Medical Benchmark
All tackle Catastrophic Forgetting
1
1
1


Performance Comparison Across Methods
Accuracy (%)
0
10
20
30
40
50
60
70
CIFAR100
ImageNet-100
DomainNet
ChestX-ray
Baseline
LingoCL
PCL
RCLP

Roadmap: From Methods to Practical CL



FUTURE SCOPE

Cross-cutting Future Research

- Combine semantic PLMs (LingoCL) with per-class modularity (PCL).
- Extend to multi-label, domain-shift settings (RCLP) across broader domains.
- Move towards generalized continual learning frameworks rather than dataset-specific solutions.

Technical Challenges to Address

1. Scalability: How to handle very large datasets & task streams.
2. Efficiency: Reducing memory/storage for replay & architectures.
3. Robustness: Handling noisy labels, domain drift, and rare classes.
4. Evaluation: Need standardized benchmarks beyond CIFAR/ChestX-rays.

Towards Real-world Deployment

- Federated Continual Learning → hospitals, organizations share knowledge without sharing data.
- Multimodal CL → images + text + audio integration.
- Self-supervised CL → reduce dependency on large labeled datasets.
- Explainable CL → transparency in decisions (critical for medical & safety domains).

Roadmap of CL Evolution

- Short-term: PLMs in CL (LingoCL extensions).
- Mid-term: Modular & scalable CL architectures (PCL).
- Long-term: Robust NIC benchmarks in critical domains (RCLP).
- Vision: A universal CL system capable of lifelong adaptation.

CONCLUSION

Continual Learning is a cornerstone for developing AI systems that can adapt, evolve, and operate reliably in dynamic real-world environments. Through our review of three representative works—LingoCL, which leverages semantic supervision from language models, PCL, which introduces per-class holistic learning to preserve past knowledge, and RCLP, which addresses the complexities of multi-label and domain-shifted medical data—we see distinct yet complementary pathways toward overcoming catastrophic forgetting. Each method contributes a critical piece of the puzzle: semantic alignment, modular scalability, and realistic benchmarks. Together, these approaches not only demonstrate the progress made in continual learning but also highlight the road ahead, where unified, multimodal, and explainable systems will form the foundation of lifelong, trustworthy AI.