MTL-FECAM: Bridging the stability-plasticity tradeoff in Exemplar-free Continual Learning

Sakshi Ranjan[1] Jatin Kumar[2] Sanjay Kumar Singh[1]

[1] Indian Institute of Technology (BHU), Varanasi, India-751004
[2] Vellore Institute of Technology, Vellore, India-632014

Presented by
Sakshi Ranjan
Paper Id : 150
sakshiranjan.rs.cse21@itbhu.ac.in

Introduction

- Continual Learning (CL) models suffer from Catastrophic Forgetting (CF) in class-incremental settings [1].
- Exemplar-Free Class-Incremental Learning (EFCIL) avoids storing past data due to privacy/storage concerns.
- Feature extractors are often frozen, which limits adaptability to new classes.
- Mahalanobis-based methods and pre-trained Transformers like ChemBERTa are promising for molecular property prediction [2].
- MTL-FECAM is proposed to balance the stability-plasticity trade-off in EFCIL for molecular data.

Motivation

- Drug discovery involves dynamic, high-dimensional, and privacy-sensitive data.
- Existing CL methods compromise either stability or plasticity.
- Mahalanobis distance can model non-linear boundaries better than Euclidean in heterogeneous data. [3]
- Data privacy concerns make exemplar storage infeasible in biomedical settings.
- A need exists for a scalable, adaptive, and exemplar-free approach to molecular property prediction. [4]

Related Work

- oEWC and LwF-MC use regularization/distillation but retain some CF.
- Prototype-based methods like NCM use Euclidean distance but fail with evolving features [5].
- FeTrIL and SSRE handle exemplar-free learning, but perform poorly in high-variance domains [9].
- Graph-based SSL and Transformer models lack mechanisms for CL in bioinformatics [6].
- Existing approaches ignore covariance modeling and Mahalanobis-based classification in molecular CL.

Challenges and Objectives

Challenges

- Balancing feature stability with adaptability to new classes (stability-plasticity trade-off) [2].
- Mitigating semantic drift without storing data from previous tasks.
- Learning optimal class boundaries in heterogeneous molecular feature spaces.

Objectives

- Introduce the Mahalanobis metric and feature covariance modeling for robust classification.
- Design an exemplar-free multitask model using ChemBERTa, EWC, and FECAM [1] [7].

Methodology

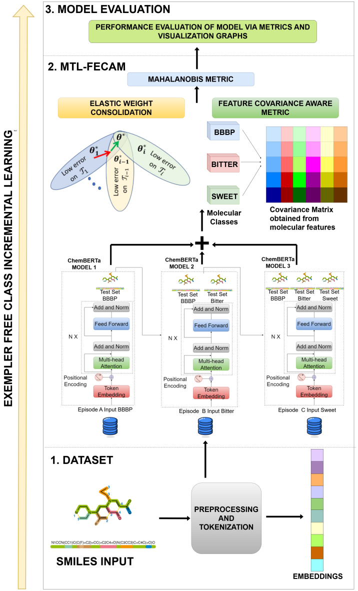
Dataset and Featurization

- Used BBBP, Bitter, and Sweet datasets from MolecularNet [4].
- Pre-processing: Canonicalization, encoding, padding, and truncation of SMILES.

MTL-FECAM

- Extracts SMILES embeddings via ChemBERTa and models class-wise and task-wise covariance matrices.
- Applies Tukey's transformation to reduce skewness, and then uses Mahalanobis distance for classification with Bayes-optimal boundaries [3].
- Employs Elastic Weight Consolidation (EWC) for parameter regularization [8].
- Supports both FSCIL and MSCIL setups.

Methodology



EXEMPLER FREE CLASS INCREMENTAL LEARNING

3. MODEL EVALUATION

PERFORMANCE EVALUATION OF MODEL VIA METRICS AND VISUALIZATION GRAPHS

2. MTL-FECAM

MAHALANOBIS METRIC

ELASTIC WEIGHT CONSOLIDATION

FEATURE COVARIANCE AWARE METRIC

Molecular Classes

Covariance Matrix obtained from molecular features

1. DATASET

PREPROCESSING AND TOKENIZATION

SMILES INPUT

EMBEDDINGS

Empirical Analysis

Table: Anytime average accuracy in exemplar-free MSCIL with different incremental tasks.

CIL METHODS	BBBP			BITTER			SWEET		
	T=5	T=10	T=20	T=5	T=10	T=20	T=5	T=10	T=20
oEWC	86.88	85.85	87.67	85.17	86.81	88.13	84.86	88.57	90.08
LwF-MC	89.80	90.34	90.88	85.34	85.79	87.21	84.84	86.34	87.91
DeeSIL	74.05	74.49	74.73	49.35	51.06	50.52	40.22	40.39	42.60
MUC	70.93	66.12	57.46	61.60	61.74	62.32	58.01	60.45	59.21
SDC	76.29	69.71	66.62	55.05	55.11	55.35	63.83	60.43	60.68
PASS	89.51	89.08	89.12	84.19	85.24	85.53	89.29	83.59	84.26
IL2A	89.00	88.63	88.36	87.30	87..62	88.18	84.00	86.07	86.41
SSRE	63.27	58.68	56.47	62.45	62.20	62.44	66.88	61.84	63.65
FeTrIL	55.26	54.71	45.20	61.14	64.85	61.12	54.26	58.66	56.14
Eucl-NCM	64.88	64.88	64.88	72.24	72.24	72.24	73.68	73.68	73.68
MTL-FECAM	90.17	90.67	91.11	89.41	90.22	90.93	88.61	89.57	90.41

Model Performance Comparison

- MTL-FECAM achieves highest accuracy and 0.0 FM across BBBP, Bitter, and Sweet datasets.
- Efficiently handles Catastrophic Forgetting while learning new tasks incrementally.
- Multitask learning promotes shared feature representation, improving task generalization, also covariance modeling and Mahalanobis distance ensure better class separation.
- Competing methods like oEWC and LwF-MC show partial retention; others like
- MTL-FECAM excels in memory efficiency, task adaptability, and stability-plasticity tradeoff.

Empirical Analysis : Baseline and Ablation Comparison

DISTANCE	Covariance Matrix	Turkey Eqn.	Shrinkage Eqn.	Normalization Eqn.	BBBP	BITTER	SWEET
Eucledian	-	×	-	-	55.77	58.18	59.47
Eucledian	-	✓	-	-	62.00	59.09	50.27
Mahalanobis	Full	×	×	×	60.00	57.99	65.75
Mahalanobis	Full	✓	×	×	63.12	64.31	50.69
Mahalanobis	Full	×	×	✓	50.61	61.93	52.67
Mahalanobis	Full	×	✓	×	62.54	62.77	52.48
Mahalanobis	Full	✓	✓	×	66.44	63.75	64.27
Mahalanobis	Full	×	✓	✓	71.80	62.64	64.52
Mahalanobis	Full	✓	✓	✓	73.85	68.04	67.00

- MTL-FECAM outperforms baseline models across all datasets.
- Ablation study confirms best performance with full Mahalanobis + Tukey + Shrinkage + Normalization.

Graphical Analysis

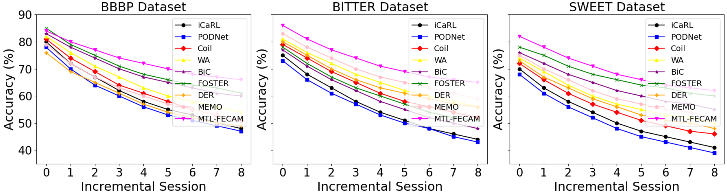


Figure: FSCIL methods accuracy of each incremental task for molecular datasets

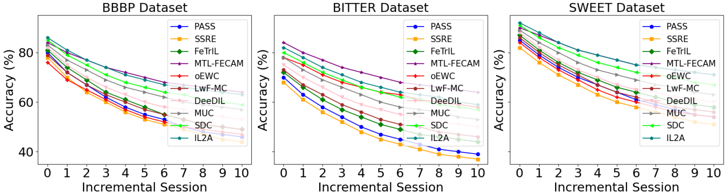


Figure: Accuracy of each incremental task for molecular datasets and multiple MSCIL methods.

Graphical Analysis

- MTL-FECAM shows consistent accuracy across incremental sessions.
- Demonstrates strong resistance to CF and outperforms others in long-term stability and task adaptation.
- BBBP shows sharpest decline — harder to retain; Bitter/Sweet degrade slowly.
- Achieves optimal decision boundaries via feature distribution modeling.
- IL2A and FeTrIL are competitive but less stable than MTL-FECAM, whereas PASS and SSRE fail due to reliance on distillation.

Conclusion and Future Scope

- The proposed MTL-FECAM integrates FECAM, the Mahalanobis metric, and EWC in ChemBERTa.
- Effectively addresses Catastrophic Forgetting and semantic drift in molecular CL.
- Balances stability-plasticity tradeoff, a key challenge in bio-informatics.
- Leverages covariance adjusted features and the Bayes-optimal classifier for robust predictions.
- Operates in an exemplar-free manner, ideal for privacy-sensitive biomedical applications.
- In future, we can extend MTL-FECAM to multi-modal biomedical data (e.g., genomics + SMILES).
- Adapt for real-time CL systems in drug discovery pipelines.
- Explore integration with explainable AI for interpretable molecular predictions.

References I

1. Robins, Anthony. "Catastrophic forgetting, rehearsal and pseudorehearsal." Connection Science 7.2 (1995): 123-146.
2. Tao, Xiaoyu, et al. "Few-shot class-incremental learning." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.
3. Mensink, Thomas, et al. "Distance-based image classification: Generalizing to new classes at near-zero cost." IEEE transactions on pattern analysis and machine intelligence 35.11 (2013): 2624-2637.
4. Hinton, Geoffrey. "Distilling the Knowledge in a Neural Network." arXiv preprint arXiv:1503.02531 (2015).

References II

5. Janson, Paul, et al. "A simple baseline that questions the use of pretrained-models in continual learning." arXiv preprint arXiv:2210.04428 (2022).
6. Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is all you need." Advances in neural information processing systems 30 (2017).
7. Catalfamo, Alessio, et al. "Machine Learning Workflows in the Computing Continuum for Environmental Monitoring." International Conference on Computational Science. Cham: Springer Nature Switzerland, 2024.

References III

8. Kirkpatrick, James, et al. "Overcoming catastrophic forgetting in neural networks."Proceedings of the national academy of sciences 114.13 (2017): 3521-3526.
9. Petit, Grégoire, et al. "Fetril: Feature translation for exemplar-free class-incremental learning." Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2023.