ALLEVIATING OVERCONFIDENCE IN LONG-TAILED TIME SERIES CLASSIFICATION WITH HIERARCHICAL REVERSE DISTILLATION

1. EXPERIMENT

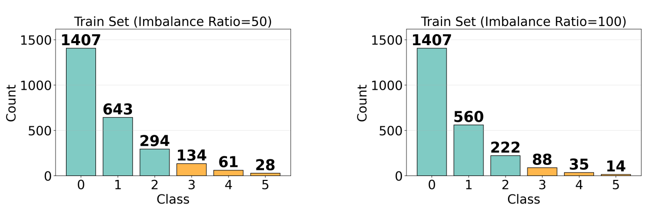
1.1. Datasets

The effectiveness of HRD-LT was evaluated on three long-tailed datasets: HAR, ISRUC, and Epilepsy. The following section presents detailed introductions to these datasets.

Human Activity Recognition (HAR). The UCI HAR dataset (Anguita et al., 2013) comprises sensor data collected from 30 participants aged between 19 and 48. Each subject performed six distinct activities, and the recordings were obtained using a Samsung Galaxy S2 smartphone with a sampling frequency of 50 Hz. Figure 1 illustrates the class distributions of the long-tailed HAR datasets under imbalance ratios of 50 and 100.

Sleep Stage Classification (ISRUC-S3). The ISRUC-S3 dataset (Khalighi et al., 2016) contains recordings from 10 healthy subjects, including electrophysiological and respiratory signals as well as background information. All data were visually examined by two experts. Figure 2 shows the class distributions of the long-tailed ISRUC datasets under imbalance ratios of 50 and 100.

Epileptic Seizure Prediction (Epilepsy). The Epileptic Seizure Recognition dataset (Andrzejak et al., 2001) consists of EEG recordings from 500 subjects, with each sample covering 23.6 seconds of brain activity. Figure 3 presents the class distributions of the long-tailed Epilepsy datasets generated under imbalance ratios of 50 and 100.

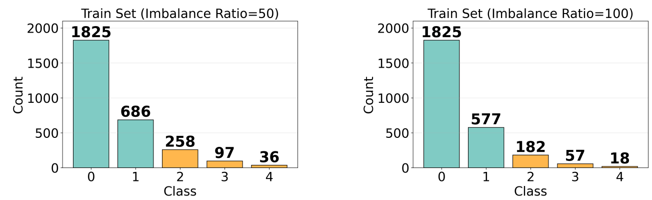


(a) HAR-LT-50 (b) HAR-LT-100

Fig. 1: Class distributions of HAR-LT datasets.

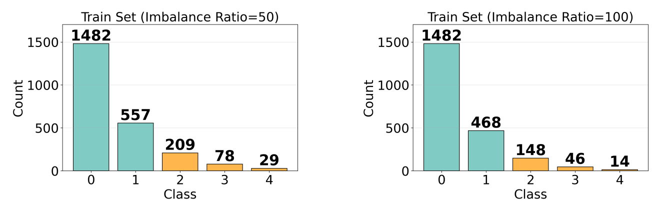
1.2. Evaluation Metrics

Model performance is evaluated using accuracy (ACC), F1-score (F1), and Matthews correlation coefficient (MCC). The



(a) ISRUC-LT-50 (b) ISRUC-LT-100

Fig. 2: Class distributions of ISRUC-LT datasets.



(a) Epilepsy-LT-50 (b) Epilepsy-LT-100

Fig. 3: Class distributions of Epilepsy-LT datasets.

definitions of these metrics are given as follows:

$$\mathrm{ACC} = \frac{TP + TN}{TP + FP + FN + TN},$$

$$\mathrm{F1} = \frac{2 \cdot P \cdot R}{P + R},$$

$$\mathrm{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}. \quad (1)$$

where P and R represent the precision and recall, respectively. TP, FN, TN, and FP denote true positives, false negatives, true negatives, and false positives, respectively.

2. BASELINES

We conduct comparisons with 7 widely recognized state-of-the-art methods:

PatchTST (Nie et al. 2022) This method partitions time series into equal-length segments (patches) and employs a vanilla Transformer encoder to model inter-patch dependencies. It operates in a channel-independent manner, where each variable sequence is encoded separately, thereby improving the flexibility of feature representation.

iTransformer (Liu et al. 2023) This model embeds temporal observations of each variable into tokens, upon which attention and feed-forward networks are applied to capture multivariate dependencies. Without altering the core Transformer architecture, it substantially enhances both predictive accuracy and generalization in time series forecasting.

Timemixer++(Wang et al. 2024) It extends TimeMixer by incorporating hybrid modules that further enhance multi-scale temporal decomposition. Through refined MLP-based operators and residual integration, it achieves stronger representation of both global and local temporal dependencies.

DisMS-TS (Liu et al. 2025)It tackles long-tailed time series classification with a disentangled multi-scale framework. By separating global trends from local variations and introducing collaborative distillation between scales, it improves feature discriminability under severe imbalance.

Hybrid-PSC (Wang et al. 2024)It combines prototype-based supervision with contrastive learning for imbalanced time series. By dynamically updating class prototypes and aligning tail samples with both head prototypes and contrastive signals, it ensures balanced performance across classes.

IWB (Dang et al. 2024)It introduces instance-wise balancing for long-tailed time series data. By re-weighting samples at the instance level rather than class level, it adaptively emphasizes informative tail examples, mitigating overconfidence and improving generalization.

DGMSCL (Qian et al. 2025)It addresses imbalanced time series classification via dynamic graph reconstruction and a mixed contrast loss. The loss combines weight-augmented inter-graph supervision and minority class-aware contrast to enhance representation.

3. REFERENCES

Chongsheng Zhang, George Almpanidis, Gaojuan Fan, Binquan Deng, Yanbo Zhang, Ji Liu, Aouaidjia Kamel, Paolo Soda, and Jo ˜ao Gama, "A systematic review on long-tailed learning," TNNLS, 2025.

Yifan Zhang, Bingyi Kang, Bryan Hooi, Shuicheng Yan, and Jiashi Feng, "Deep long-tailed learning: A survey," TPAMI, 2023.

Shiyu Wang, Jiawei Li, Xiaoming Shi, Zhou Ye, Baichuan Mo, Wenze Lin, Shengtong Ju, Zhixuan Chu, and Ming Jin, "Timemixer++: A general time series pattern machine for universal predictive analysis," ICLR, 2025.

Yaxuan Kong, Zepu Wang, Yuqi Nie, Tian Zhou, Stefan Zohren, Yuxuan Liang, Peng Sun, and Qingsong Wen, "Unlocking the power of lstm for long term time series forecasting," CoRR, 2024.

Yucheng Wang, Yuecong Xu, Jianfei Yang, Min Wu, Xiaoli Li, Lihua Xie, and Zhenghua Chen, "Graph-aware contrasting for multivariate time-series classification," AAAI, 2024.

Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam, "Atime series is worth 64 words: Long-term forecasting with transformers," ICLR, 2023.

Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jian- min Wang, and Mingsheng Long, "Timesnet: Temporal 2d-variation modeling for general time series analysis," ICLR, 2023.

Peng Wang, Kai Han, Xiu-Shen Wei, Lei Zhang, and Lei Wang, "Contrastive learning based hybrid networks for long-tailed image classification," CVPR, 2021.

Wenqi Dang, Zhou Yang, Weisheng Dong, Xin Li, and Guangming Shi, "Inverse weight-balancing for deep long-tailed learning," AAAI, 2024.

Liuyu Xiang, Guiguang Ding, and Jungong Han, "Learn- ing from multiple experts: Self-paced knowledge distil- lation for long-tailed classification," ECCV, 2020.

Harsh Rangwani, Pradipto Mondal, Mayank Mishra, Ashish Ramayee Asokan, and R Venkatesh Babu, "Deit- lt: Distillation strikes back for vision transformer training on long-tailed datasets," CVPR, 2024.

Yin-Yin He, Peizhen Zhang, Xiu-Shen Wei, Xiangyu Zhang, and Jian Sun, "Relieving long-tailed instance segmentation via pairwise class balance," CVPR, 2022.

Pengkun Wang, Xu Wang, Binwu Wang, Yudong Zhang, Lei Bai, and Yang Wang, "Long-tailed time series classi- fication via feature space rebalancing," DASFAA, 2023.

Lipeng Qian, Qiong Zuo, Dahu Li, and Hong Zhu, "Dgm-scl: A dynamic graph mixed supervised contrastive learning approach for class imbalanced multivariate time se- ries classification," Neural Networks, 2025.

Chen Huang, Yining Li, Chen Change Loy, and Xiaoou Tang, "Learning deep representation for imbalanced classification," CVPR, 2016.

Chih-Hui Ho, Kuan-Chuan Peng, and Nuno Vasconce- los, "Long-tailed anomaly detection with learnable class names," CVPR, 2024.

Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, Chee Keong Kwoh, Xiaoli Li, and Cuntai Guan, "Time-series representation learning via temporal and contextual contrasting," arXiv, 2021.

Rohan Anil, Gabriel Pereyra, Alexandre Passos, Robert Ormandi, George E Dahl, and Geoffrey E Hinton, "Large scale distributed neural network training through online distillation," arXiv, 2018.

Qiushan Guo, Xinjiang Wang, Yichao Wu, Zhipeng Yu, Ding Liang, Xiaolin Hu, and Ping Luo, "Online knowl- edge distillation via collaborative learning," CVPR, 2020.

Zhipeng Liu, Peibo Duan, Binwu Wang, Xuan Tang, Qi Chu, Changsheng Zhang, Yongsheng Huang, and Bin Zhang, "Disms-ts: Eliminating redundant multi-scale features for time series classification," arXiv, 2025.

Chih-Hui Ho, Kuan-Chuan Peng, and Nuno Vasconce- los, "Long-tailed anomaly detection with learnable class names," CVPR, 2024.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long, "itransformer: Inverted transformers are effective for time series fore- casting," arXiv, 2023.

Shiyu Wang, Jiawei Li, Xiaoming Shi, Zhou Ye, Baichuan Mo, Wenze Lin, Shengtong Ju, Zhixuan Chu, and Ming Jin, "Timemixer++: A general time series pattern machine for universal predictive analysis," arXiv, 2024.