ICLR

ICLR 2026

NeurIPS TS4H Spotlight

Tokenizing Single-Channel EEG with Time-Frequency Motif Learning

Jathurshan Pradeepkumar[1] Xihao Piao[2] Zheng Chen[2] Jimeng Sun[1]

[1]University of Illinois Urbana-Champaign [2]Osaka University

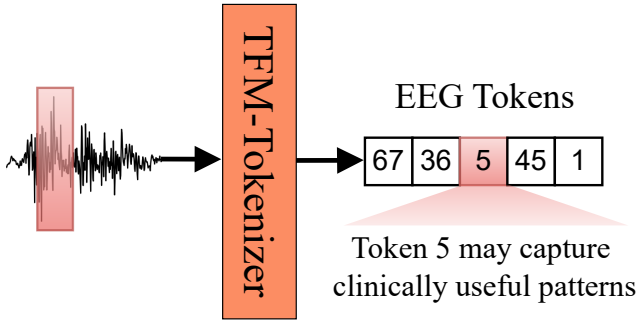
Presenter: Jathurshan Pradeepkumar

ILLINOIS

SANKEN
THE UNIVERSITY OF OSAKA

Website Paper Code

Background



EEG Foundation Modeling

- Foundation models have revolutionized natural language processing.
- This success inspired a paradigm shift in EEG analysis toward the development of task-agnostic foundation models.

Tokenization Challenge

- Tokenization, a core component in NLP, transforms raw text into meaningful tokens, which reduces data complexity and introduces a helpful inductive bias in foundation models.
- Despite progress in EEG FMs, an important open problem remains:
  - **How to design an effective tokenization method for EEG signals?**
- Existing EEG foundation models tokenize signals by directly segmenting continuous EEGs into short-duration → This fails to capture reusable, statistically grounded tokens from the signal.

Key Challenges in EEG Tokenization

1

Tokenization Target

- Real-world EEG is highly heterogeneous (devices, channels, durations)

2

Token Resolution

- Tokenization in NLP varies by granularity.
- EEG differs: signals contain oscillatory patterns (alpha, beta) and transients (spikes)

3

Tokenization Learning Objective

- EEG signals have complex temporal variations
  - Mixture of low- and high-frequency components

Key Challenges in EEG Tokenization

1

Tokenization Target

- Real-world EEG is highly heterogeneous (devices, channels, durations)

2

Token Resolution

- Tokenization in NLP varies by granularity.
- EEG differs: signals contain oscillatory patterns (alpha, beta) and transients (spikes)

3

Tokenization Learning Objective

- EEG signals have complex temporal variations
  - Mixture of low- and high-frequency components

Single Channel Level

Key Challenges in EEG Tokenization

1

Tokenization Target

- Real-world EEG is highly heterogeneous (devices, channels, durations)

2

Token Resolution

- Tokenization in NLP varies by granularity.
- EEG differs: signals contain oscillatory patterns (alpha, beta) and transients (spikes)

3

Tokenization Learning Objective

- EEG signals have complex temporal frequency variations
  - Mixture of low- and high-frequency components

Single Channel Level

Motifs

Key Challenges in EEG Tokenization

1

Tokenization Target

- Real-world EEG is highly heterogeneous (devices, channels, durations)

Single Channel Level

2

Token Resolution

- Tokenization in NLP varies by granularity.
- EEG differs: signals contain oscillatory patterns (alpha, beta) and transients (spikes)

Motifs

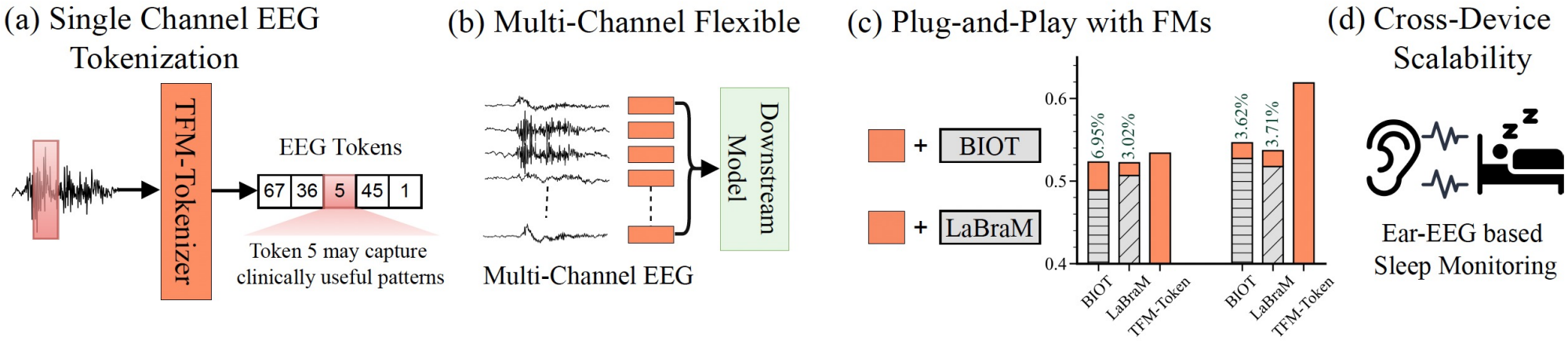
3

Tokenization Learning Objective

- EEG signals have complex temporal variations
  - Mixture of low- and high-frequency components

Time Frequency Representation

Key Contributions



- Formulating Single-Channel EEG Tokenization
  - Introduce learning of discrete token vocabulary from single-channel EEGs by capturing time-frequency motifs.
- We introduce a single-channel EEG tokenization framework:
  - **TFM-Tokenizer**: Converts EEG into a discrete token sequence.
  - A lightweight transformer model then uses EEG tokens for cross-channel and downstream modeling.
- Broad Evaluation across Datasets, Foundation Models, and Devices.

TFM-Tokenizer

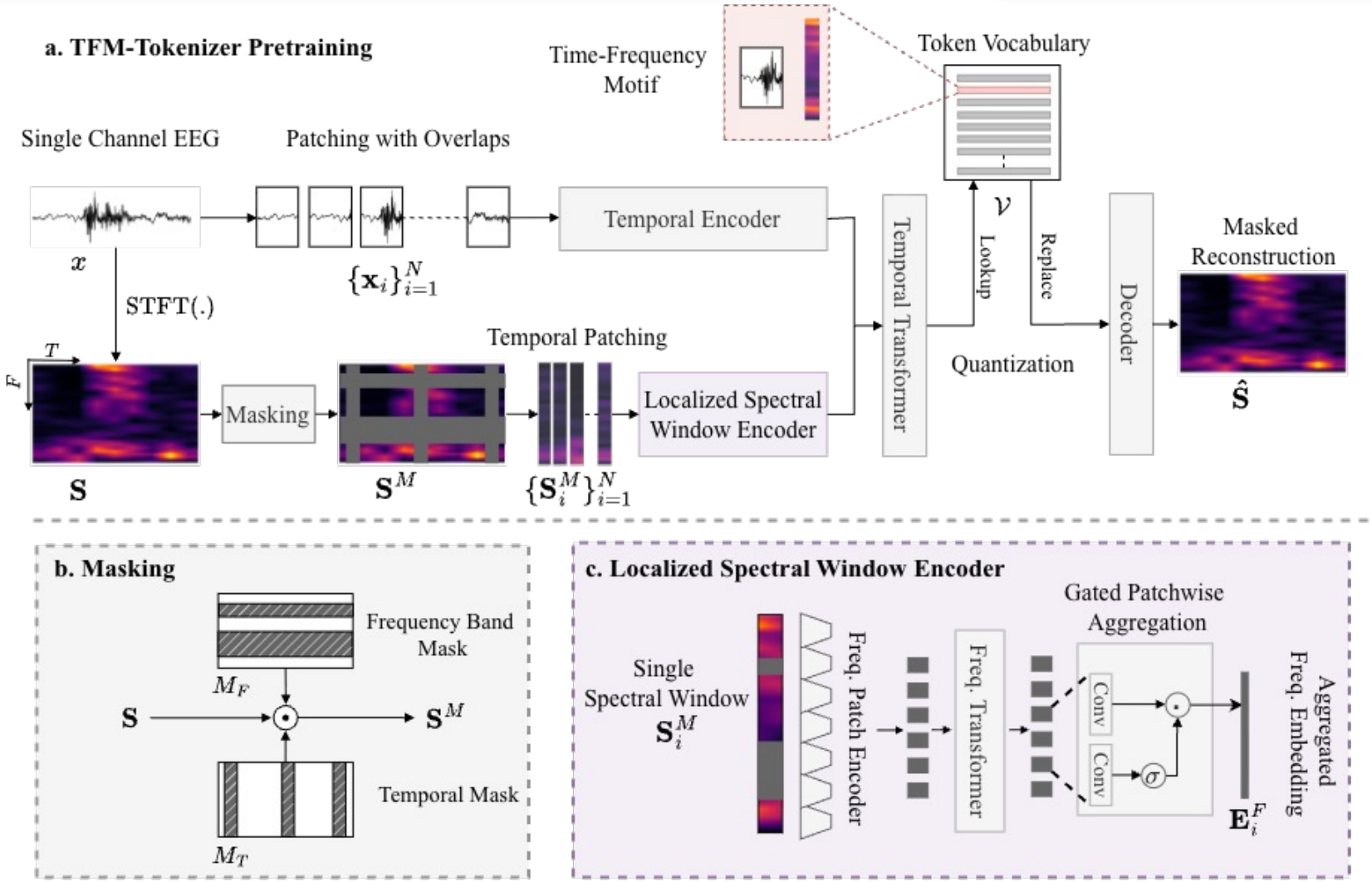
Phase 1: TFM-Tokenizer Training

Phase 2: Downstream Transformer

TFM-Tokenizer

Phase 1: TFM-Tokenizer Training

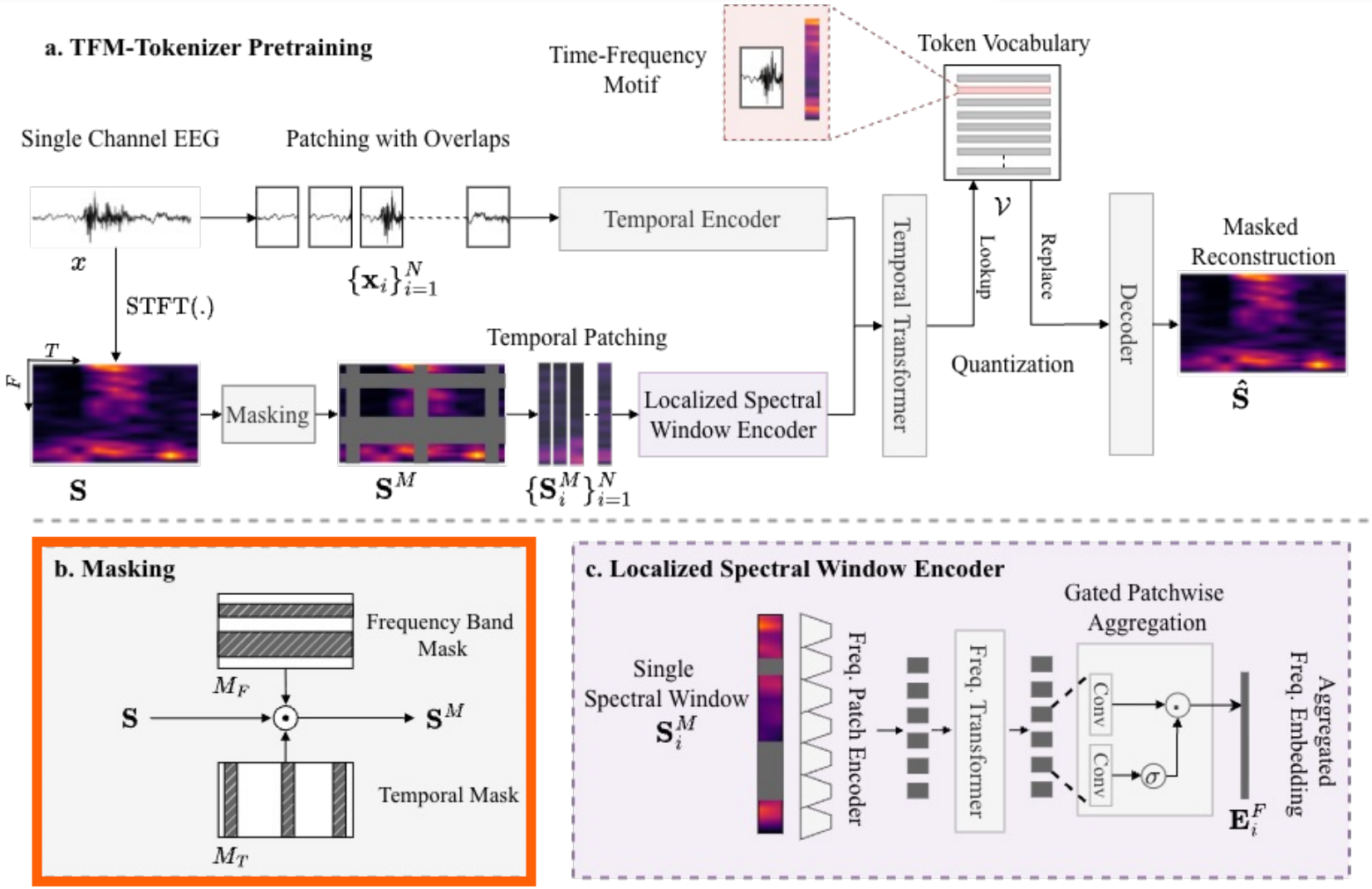
Phase 2: Downstream Transformer



TFM-Tokenizer

Phase 1: TFM-Tokenizer Training

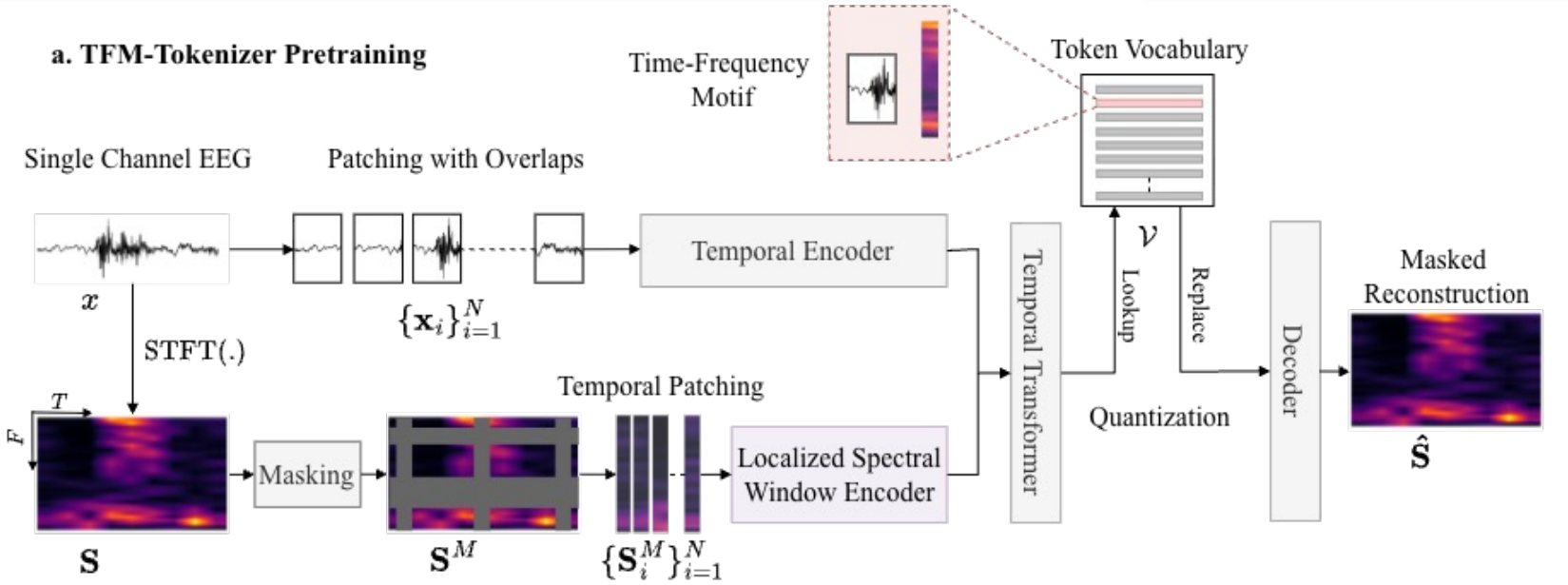
Phase 2: Downstream Transformer

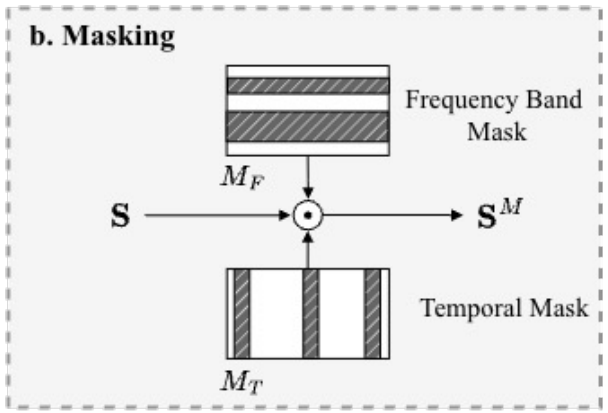


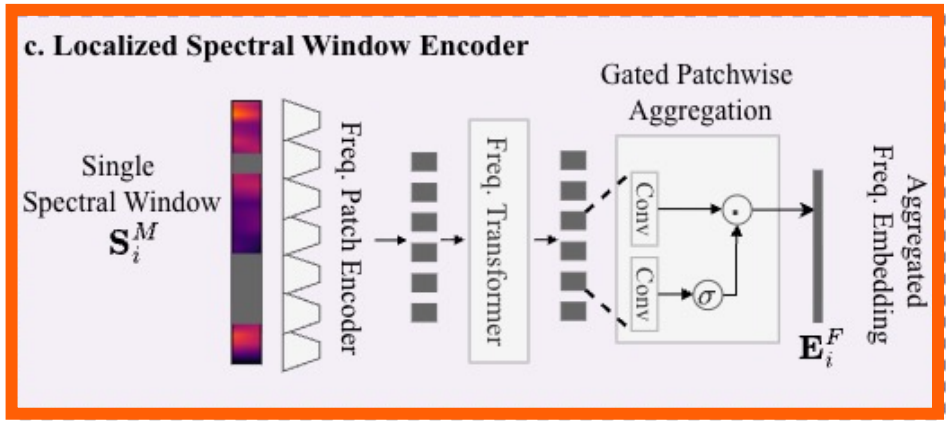
TFM-Tokenizer

Phase 1: TFM-Tokenizer Training

Phase 2: Downstream Transformer



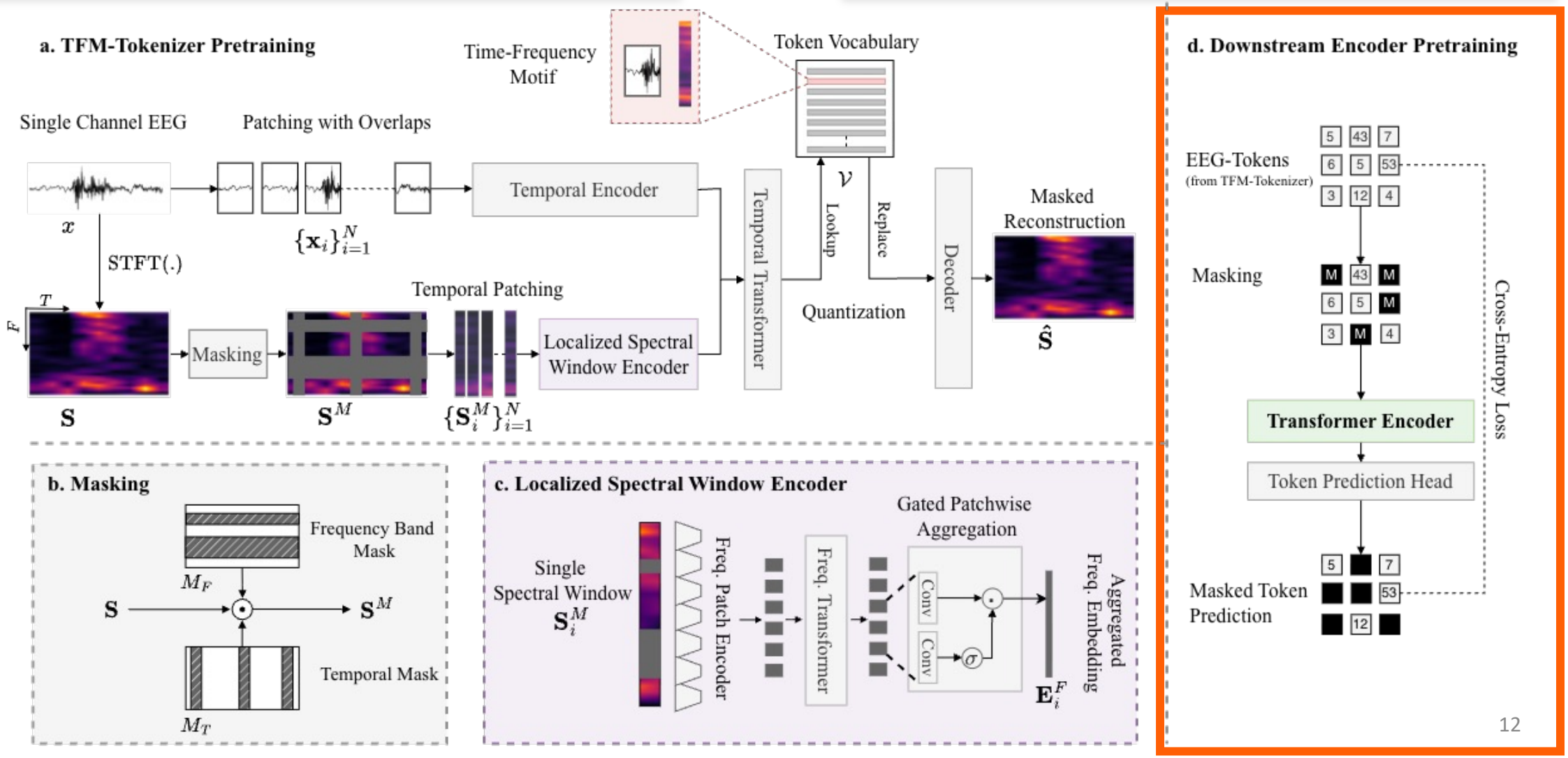




TFM-Tokenizer

Phase 1: TFM-Tokenizer Training

Phase 2: Downstream Transformer

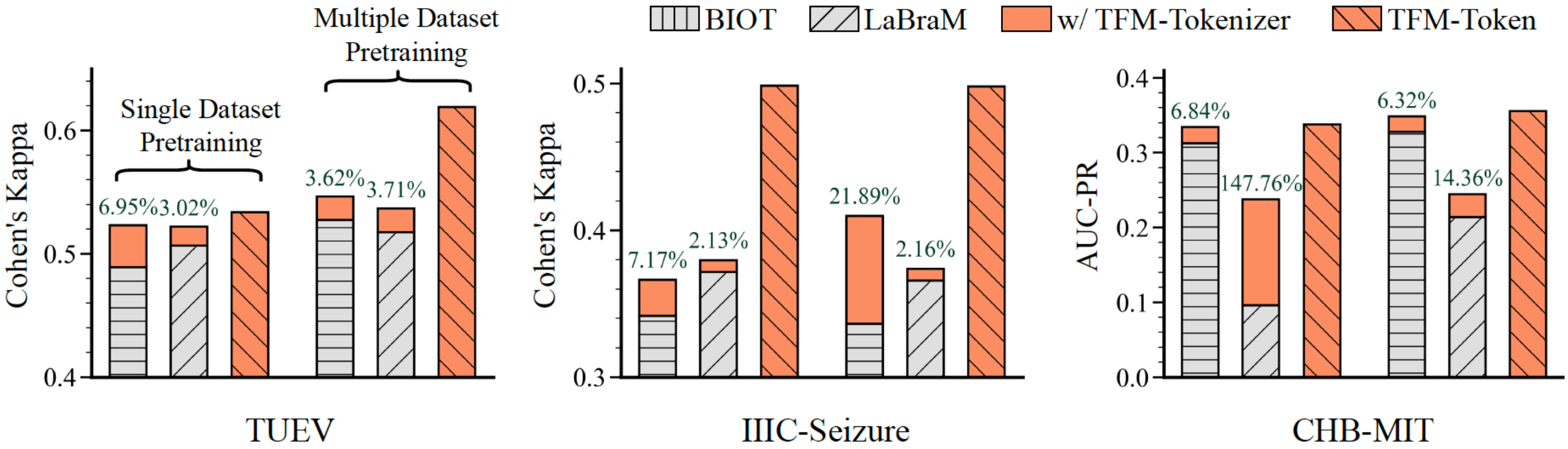


How Does TFM-Tokenizer Compare To Existing Baselines?

Table 1: Performance comparison on TUEV and TUAB datasets.

Models	Model Size	TUEV (event type classification)			TUAB (abnormal detection)		
		Balanced Acc.	Cohen's Kappa	Weighted F1	Balanced Acc.	AUC-PR	AUROC
Single Dataset Setting							
SPaRCNet (Jing et al., 2023)	0.79M	0.4161 ± 0.0262	0.4233 ± 0.0181	0.7024 ± 0.0104	0.7896 ± 0.0018	0.8414 ± 0.0018	0.8676 ± 0.0012
ContraWR (Yang et al., 2023)	1.6M	0.4384 ± 0.0349	0.3912 ± 0.0237	0.6893 ± 0.0136	0.7746 ± 0.0041	0.8421 ± 0.0104	0.8456 ± 0.0074
CNN-Transformer (Peh et al., 2022)	3.2M	0.4087 ± 0.0161	0.3815 ± 0.0134	0.6854 ± 0.0293	0.7777 ± 0.0022	0.8433 ± 0.0039	0.8461 ± 0.0013
FFCL (Li et al., 2022)	2.4M	0.3979 ± 0.0104	0.3732 ± 0.0188	0.6783 ± 0.0120	0.7848 ± 0.0038	0.8448 ± 0.0065	0.8569 ± 0.0051
ST-Transformer (Song et al., 2021)	3.5M	0.3984 ± 0.0228	0.3765 ± 0.0306	0.6823 ± 0.0190	0.7966 ± 0.0023	0.8521 ± 0.0026	0.8707 ± 0.0019
Vanilla BIOT (Yang et al., 2024)	3.2M	0.4682 ± 0.0125	0.4482 ± 0.0285	0.7085 ± 0.0184	0.7925 ± 0.0035	0.8707 ± 0.0087	0.8691 ± 0.0033
BIOT* (Yang et al., 2024)	3.2M	0.4679 ± 0.0354	0.4890 ± 0.0407	0.7352 ± 0.0236	0.7955 ± 0.0047	0.8819 ± 0.0046	0.8834 ± 0.0041
LaBraM-Base* (Jiang et al., 2024b)	5.8M	0.4682 ± 0.0856	0.5067 ± 0.0413	0.7466 ± 0.0202	0.7720 ± 0.0046	0.8498 ± 0.0036	0.8534 ± 0.0027
TFM-Tokenizer (Ours)	1.9M	**0.4943** ± 0.0516	**0.5337** ± 0.0306	**0.7570** ± 0.0163	**0.8152** ± 0.0014	**0.8946** ± 0.0008	**0.8897** ± 0.0008
With Multiple Dataset Pretraining							
BIOT (Yang et al., 2024)	3.2M	0.5281 ± 0.0225	0.5273 ± 0.0249	0.7492 ± 0.0082	0.7959 ± 0.0057	0.8792 ± 0.0023	0.8815 ± 0.0043
EEGPT (Wang et al., 2024a)	4.7M	0.5670 ± 0.0066	0.5085 ± 0.0173	0.7535 ± 0.0097	0.7959 ± 0.0021	-	0.8716 ± 0.0041
NeuroLM-B (Jiang et al., 2024a)	254M	0.4560 ± 0.0048	0.4285 ± 0.0048	0.7153 ± 0.0028	0.7826 ± 0.0065	0.6975 ± 0.0081	0.7816 ± 0.0079
LaBraM-Base† (Jiang et al., 2024b)	5.8M	0.5550 ± 0.0403	0.5175 ± 0.0339	0.7450 ± 0.0194	0.7735 ± 0.0030	0.8531 ± 0.0028	0.8557 ± 0.0027
CBraMod† (Wang et al., 2024d)	4M	0.5696 ± 0.0221	0.5588 ± 0.0273	0.7702 ± 0.0137	0.5000 ± 0.0000	0.4938 ± 0.0443	0.5281 ± 0.0409
TFM-Tokenizer (Ours) †	1.9M	**0.5974** ± 0.0079	**0.6189** ± 0.0302	**0.8010** ± 0.0161	**0.8032** ± 0.0035	**0.8886** ± 0.0032	**0.8870** ± 0.0022

Can TFM-Tokenizer Improve Existing Foundation Models?



Does TFM-Tokenizer Scale To Other Brain-signal Types / Devices?

Table 3: Scalability experiments results on EESM23.

Models	Ear-EEG (Sleep Staging)		
	Balanced Acc.	**Cohen's Kappa**	**Weighted F1**
BIOT	0.3858 ± 0.0085	0.3406 ± 0.0096	0.4888 ± 0.0124
BIOT-TFM	0.3952 ± 0.0170 ↑	0.3603 ± 0.0252 ↑	0.5033 ± 0.0165 ↑
LaBraM-Base	0.3890 ± 0.0182	0.3322 ± 0.0232	0.4827 ± 0.0157
LaBraM-TFM	0.4004 ± 0.0086 ↑	0.3475 ± 0.0128 ↑	0.4864 ± 0.0118 ↑
TFM-Tokenizer	**0.4148** ± 0.0209	**0.3883** ± 0.0233	**0.5174** ± 0.0141

Do The Learned Tokens Capture Meaningful EEG Motifs?

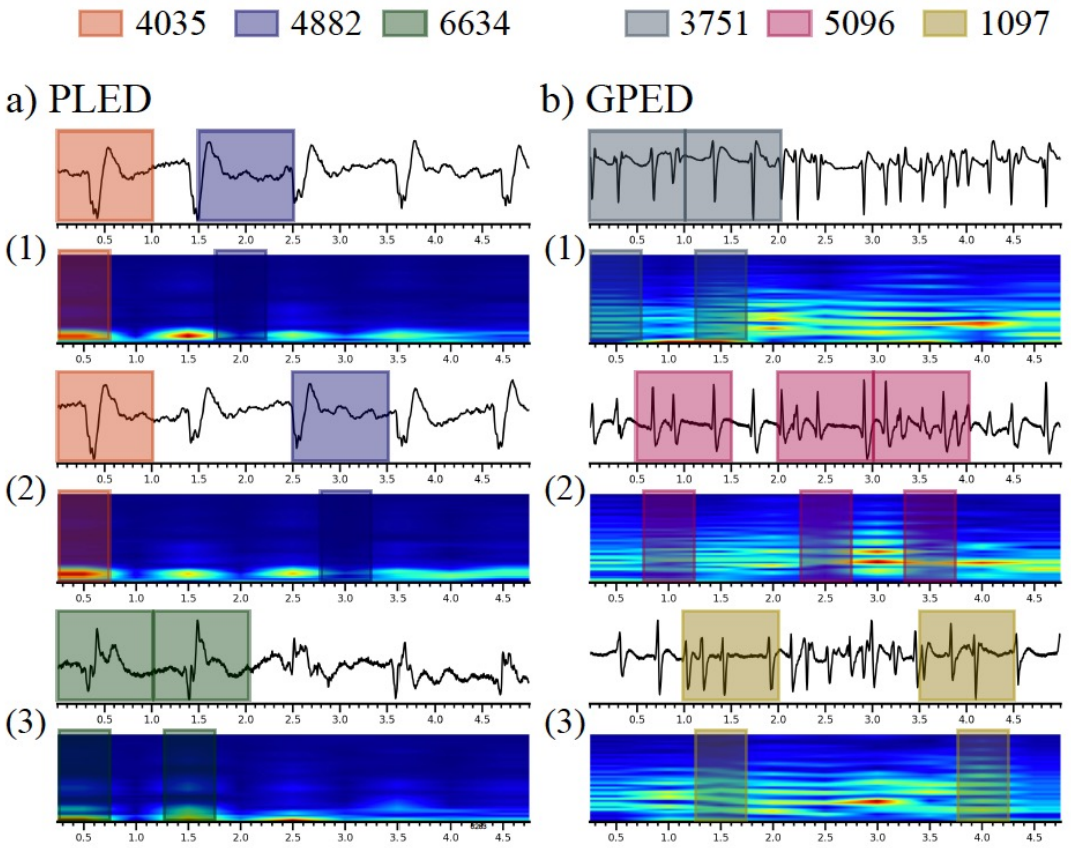


Figure 5: Overview of motifs captured by TFM-Tokenizer on TUEV: (a) three samples from the PLED class and (b) three samples from the GPED.

ICLR

ICLR 2026

NeurIPS TS4H Spotlight

Tokenizing Single-Channel EEG with Time-Frequency Motif Learning

Paper

Code

Project Page

ILLINOIS

SANKEN
THE UNIVERSITY OF OSAKA

https://www.jathurshanpradeepkumar.com/TFM-Tokenizer/