Deep Learning for Time Series

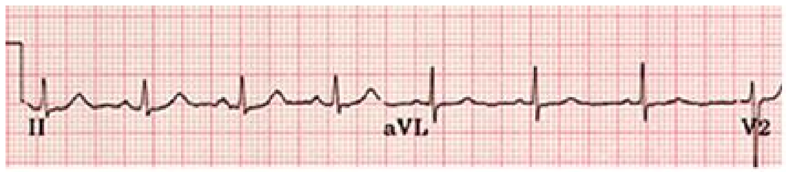
Session 4b: Time Series Foundation Models

Romain Tavenard

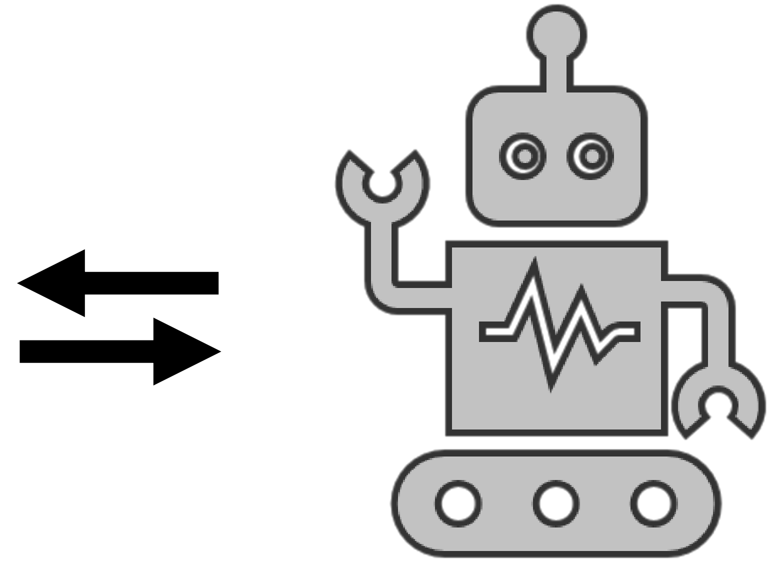
NB: Some figures in this slide deck are borrowed from Ievgen Redko's (great) course on time series forecasting and classification

Traditional Machine Learning

Normal ECG



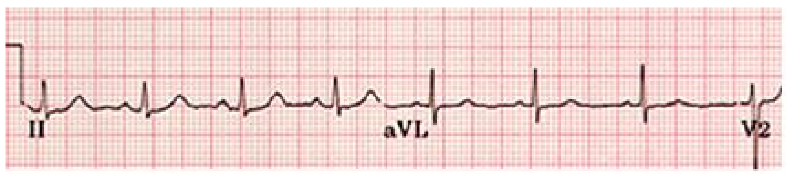
Abnormal ECG



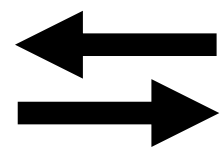
- Traditional ML pipeline
  1. Collect training data
  2. Train a new model
  3. Deploy it

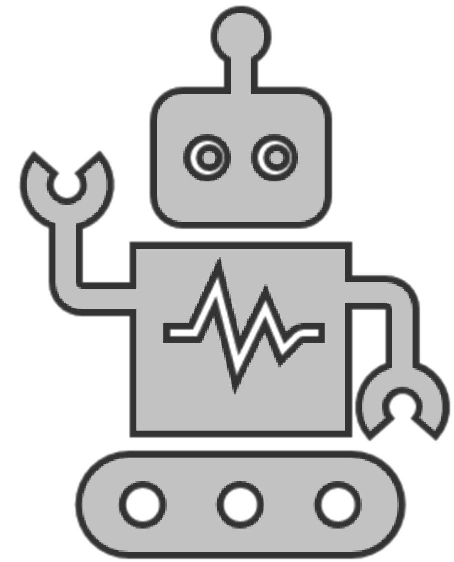
Traditional Machine Learning

Normal ECG



Abnormal ECG



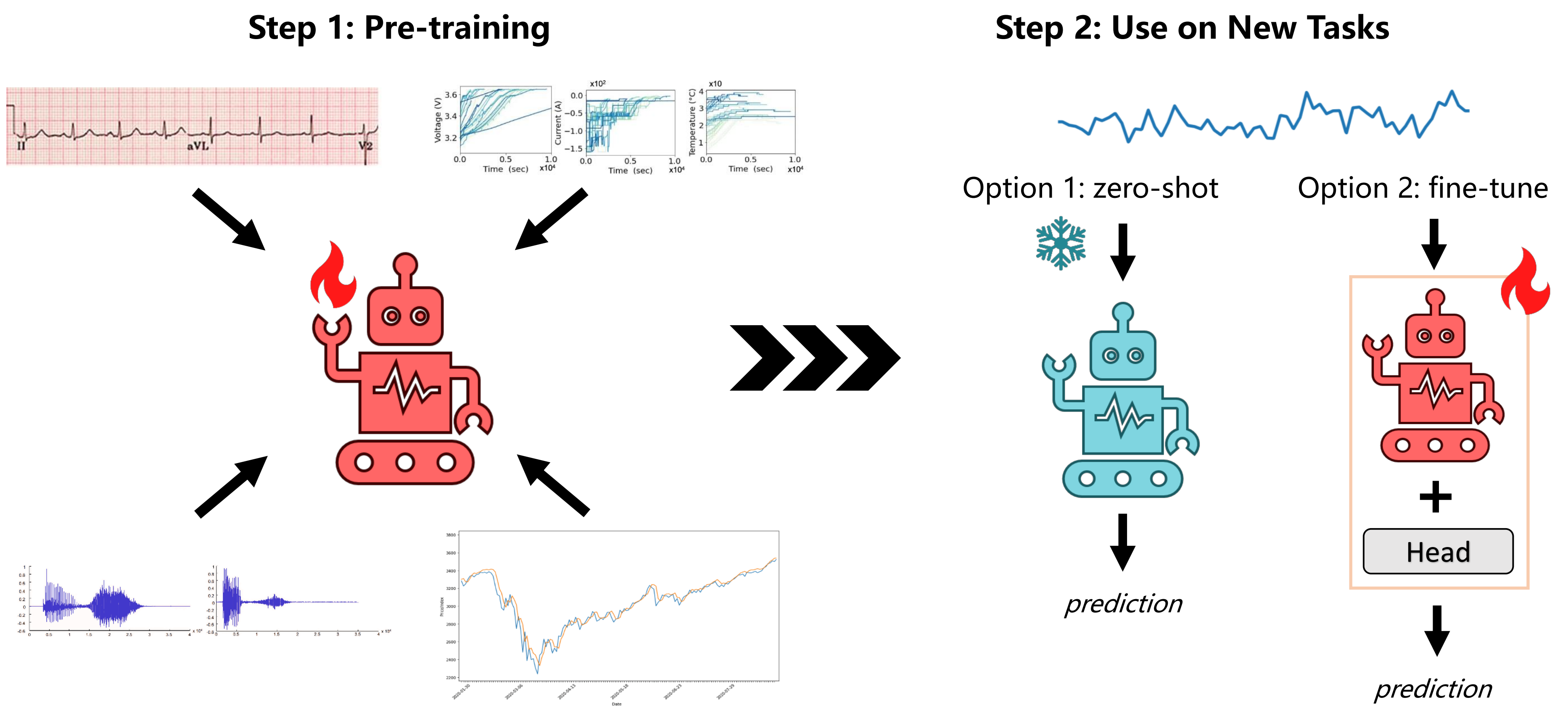


- Traditional ML pipeline
  1. Collect training data $\leftarrow$ Issue 1: requires large training set
  2. Train a new model $\leftarrow$ Issue 2: a new model for each task
  3. Deploy it

Success of pre-training

- In Computer Vision: ImageNet pre-training is everywhere
- In text: multi-task LLMs
- In time series
  - Is there an ImageNet?
  - Do we have competitive task-adaptive models?

Time Series Foundation Models (TSFMs)



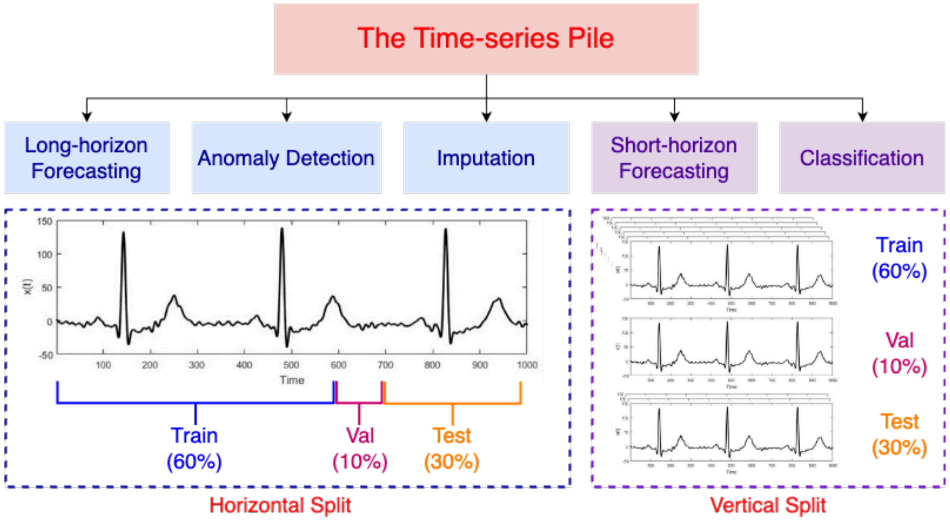
Step 1: Pre-training

Pre-training datasets

- LOTSA (MOIRAI)
  - 27.7B observations
  - 4M time series
  - LOTSAMini exists: < 1% of LOTSA
- TimeSeries-PILE (MOMENT)
  - 1.1B observations
  - 13M time series
- Time300B
  - 300B observations
  - 48M time series
- TimePFN's synthetic data generators
  - used 300-600M observations for TimePFN pre-training

Pre-training strategies

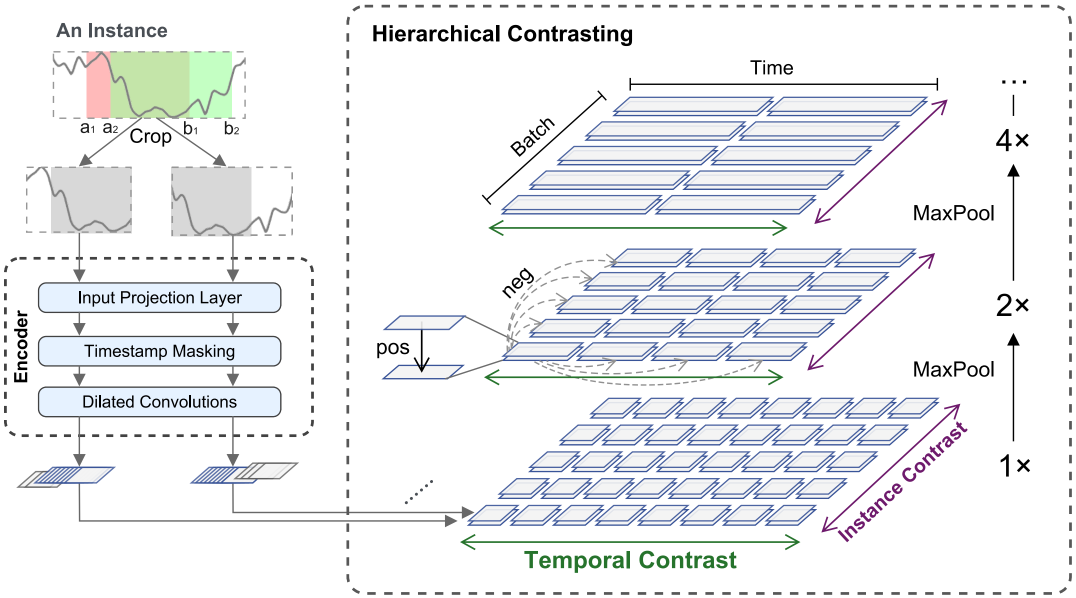
- Pre-training across tasks



Source: "MOMENT: A Family of Open Time-series Foundation Models", ICML'24

- Semi-Supervised Learning
  - eg. MOMENT (masking), TS2Vec (contrastive)
- Use pre-trained LLM
  - eg. GPT4TS

- Self-supervised contrastive learning for time series
  - Contrast positives: same series, different segments/scales
  - Contrast negatives: different time series
- Learns multi-scale (hierarchical) representations
- Temporal encoder (CNN-based), no decoder
- Outputs fixed-length embeddings for downstream tasks



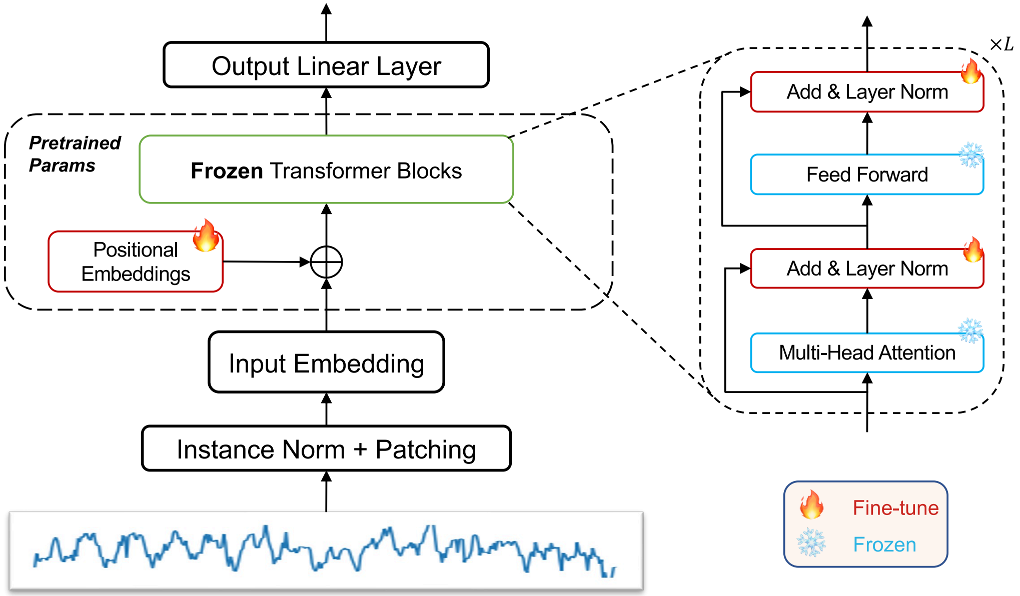
Source: “TS2Vec: Towards Universal Representation of Time Series”, AAAI’22

TS2Vec

- First strong task-agnostic TS representation method
- Validated pretrain once $\rightarrow$ reuse everywhere paradigm
- Shifted SSL from reconstruction to semantic contrastive learning
- Strong zero-shot / linear-probe baseline

General idea: make use of LLM backbone

1. Normalization + Patching (*à la* PatchTST)
2. Projection to LLM embedding dimension
3. Fine-tuning of positional embeddings + layer normalization, keeping others frozen

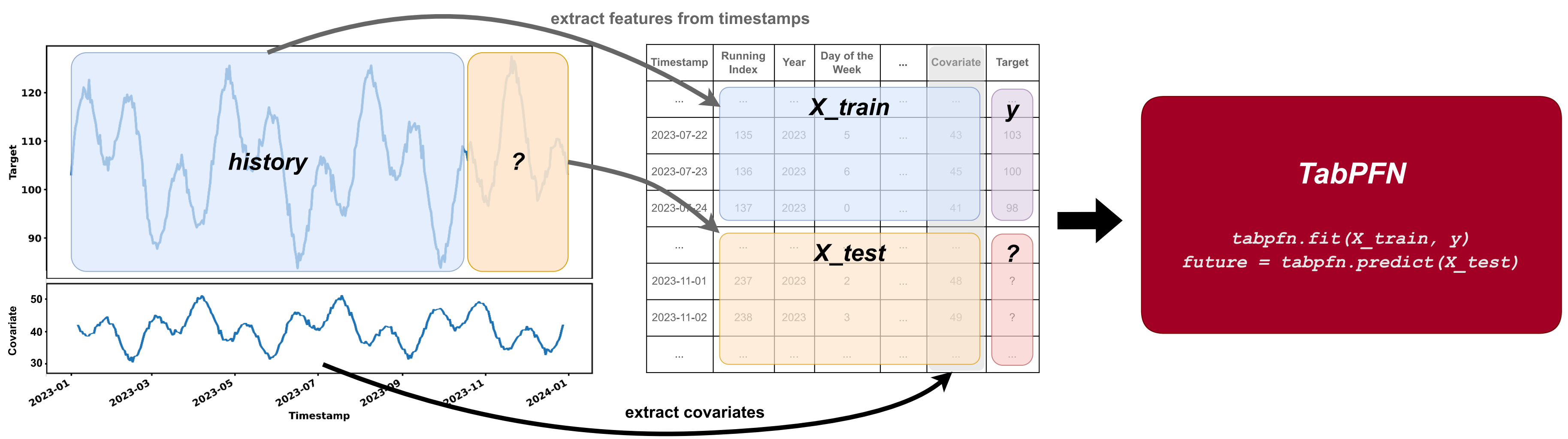


Source: “One Fits All: Power General Time Series Analysis by Pretrained LM”, NeurIPS’23

Step 2: Adapting

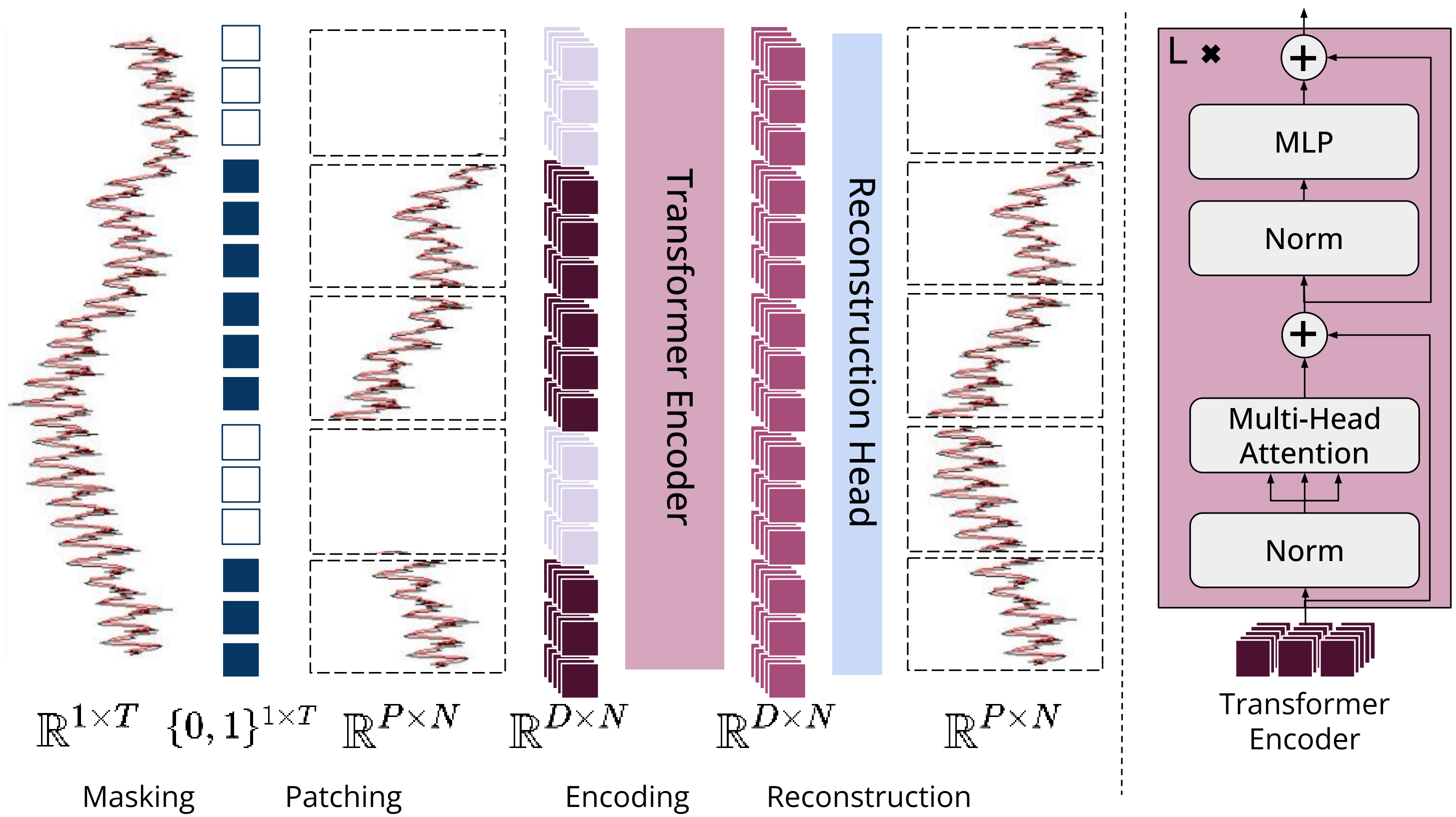
TabPFN-TS: In-Context learning

- TabPFN is pre-trained on tabular data alone
- Builds basic TS features (calendar, seasonal, ...)
- No TS prior at all!



Source: “From Tables to Time: Extending TabPFN-v2 to Time Series Forecasting”, ArXiV’26

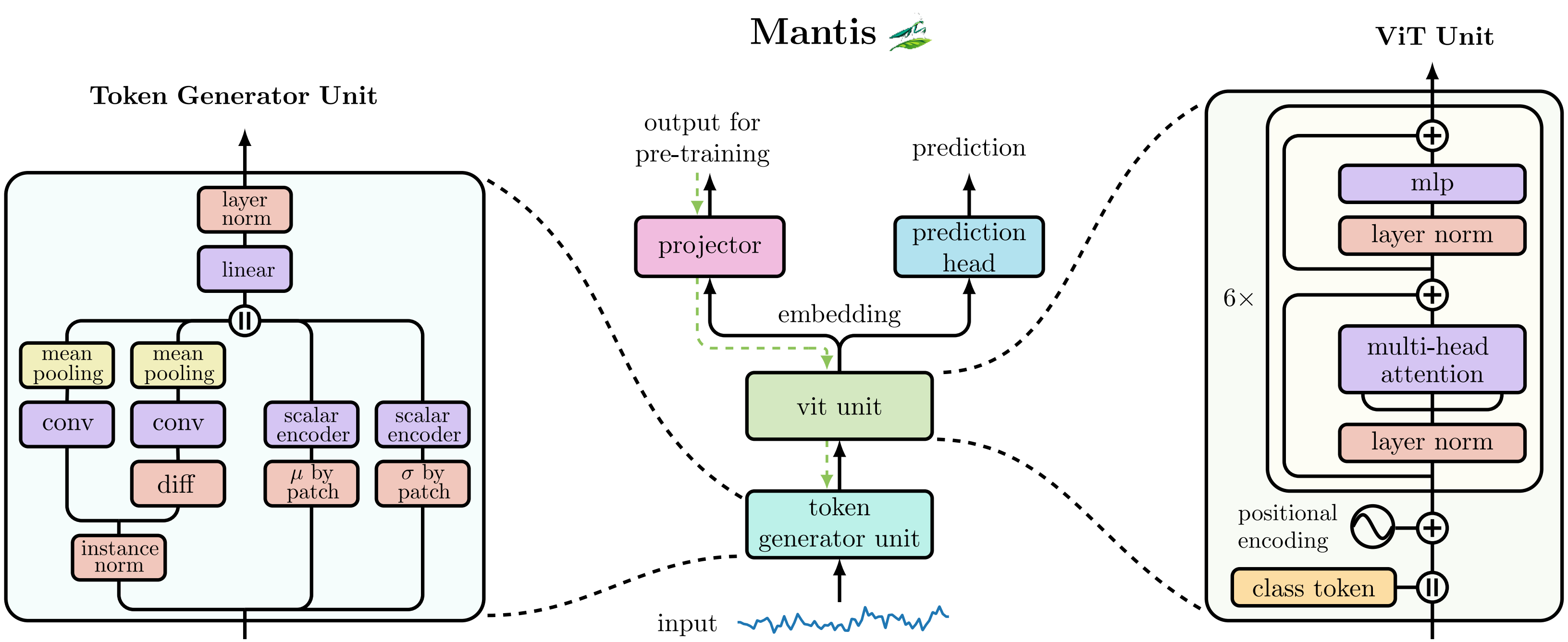
Encoder-only model: can be used for any downstream task (classification, forecasting, imputation etc)



Source: “MOMENT: A Family of Open Time-series Foundation Models”, ICML’24

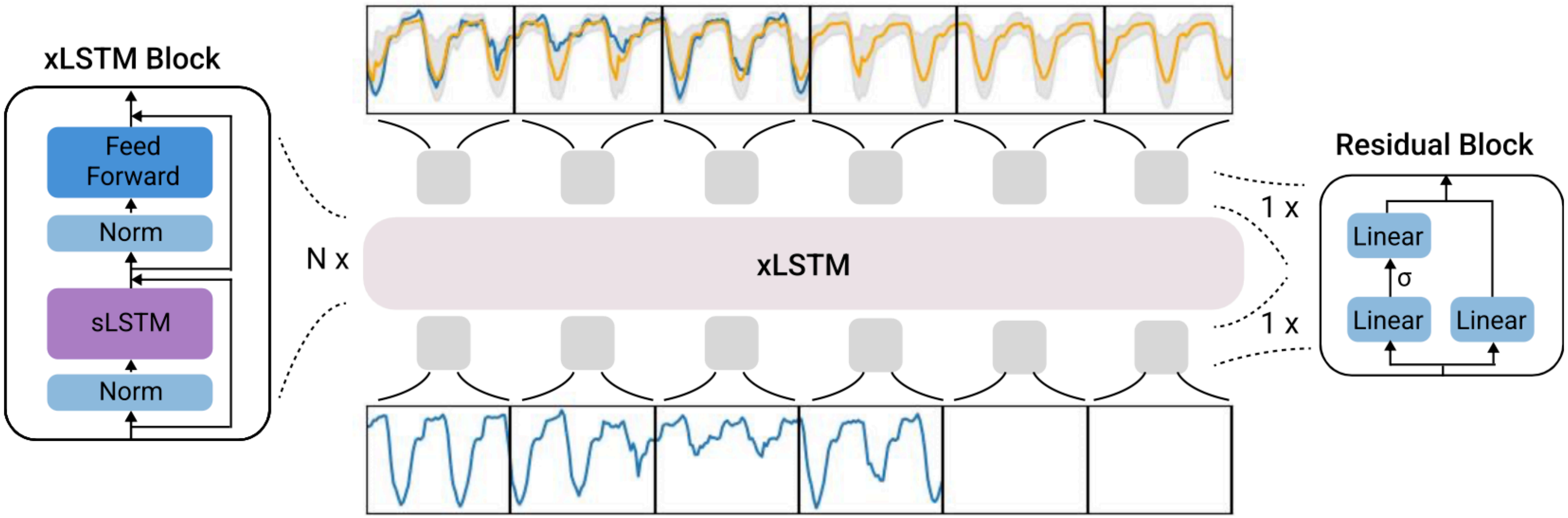
MANTIS: a FM for TSC

- Contrastive pre-training
- Scalar embedding module for preserving basic statistics
- Strong zero-shot and fine-tuning performance for TSC



Source: "Mantis: Lightweight Calibrated Foundation Model for User-Friendly Time Series Classification", ICML workshops'25

- Masking-based pre-training
- Simple xLSTM architecture (only sLSTM blocks)



Source: “TiRex: Zero-Shot Forecasting Across Long and Short Horizons with Enhanced In-Context Learning”, NeurIPS’25

Step 3: Evaluating

GIFTEval — Benchmarking Time-Series Foundation Models

- Unified evaluation benchmark for TSFMs
- Focuses on forecasting
- Diverse datasets across domains and time scales
  - short / long term
  - univariate / multivariate
- https://huggingface.co/spaces/Salesforce/GIFT-Eval