ZAYAN: Disentangled Contrastive Transformer for Tabular Remote Sensing Data

Al Zadid Sultan Bin Habib[1], Tanpia Tasnim[2], Md. Ekramul Islam[3], Muntasir Tabasum[1]

[1]West Virginia University, USA [2]Green University of Bangladesh

[3]Stamford University Bangladesh

ICPR 2026 Oral Presentation · Paper #649

International Conference on Pattern Recognition

What is ZAYAN?

Zero-Anchor dYnamic feAture eNcoding
A feature-centric self-supervised framework for tabular remote sensing data.

Self-Supervised

Learns useful representations from tabular data **without labels during pretraining.**

Feature-Level

Instead of contrasting samples,
ZAYAN contrasts **features and their augmented views.**

Disentangled

Encourages feature embeddings that are **robust, diverse, and redundancy-reduced.**

Core idea: move contrastive learning from the **sample level** to the **feature level**.

Roadmap

1
Problem

Remote sensing tables are noisy, redundant, heterogeneous, and often label-scarce.

2
Idea

Move contrastive learning from samples to features, without explicit anchors.

3
Method

ZAYAN-CL learns feature embeddings; ZAYAN-T performs downstream classification.

4
Evidence

Eight datasets, 31 baselines, ablations, robustness, calibration, and OOD analysis.

Main message: feature-level self-supervision is a practical recipe for tabular remote sensing and environmental data.

Why Tabular Remote Sensing Is Hard

Feature heterogeneity

Satellite, GIS, socio-economic, and environmental features often live on different scales and semantics.

Redundancy and noise

Many derived features are correlated, noisy, or only weakly informative for the final task.

Scarce labels and shift

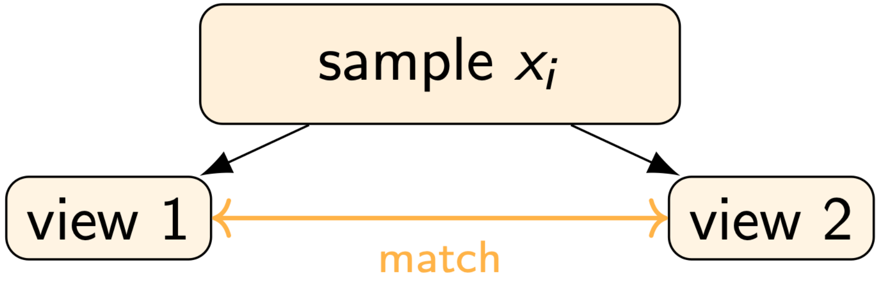
Ground-truth labels are costly, and deployment may face distribution shift across regions or events.

Most contrastive remote sensing SSL is image patch/instance-centric, while tabular models often treat features as uniform tokens.

ZAYAN asks: can we learn directly at the feature level?

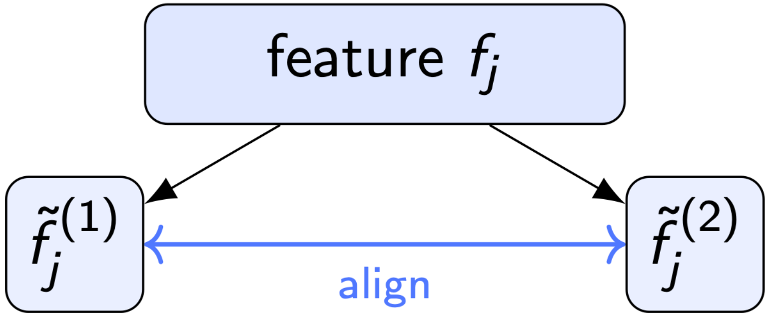
From Instance-Level Contrast to Feature-Level Contrast

Conventional contrastive learning



The anchor is usually a sample or patch. This may miss column-wise redundancy and feature semantics in tables.

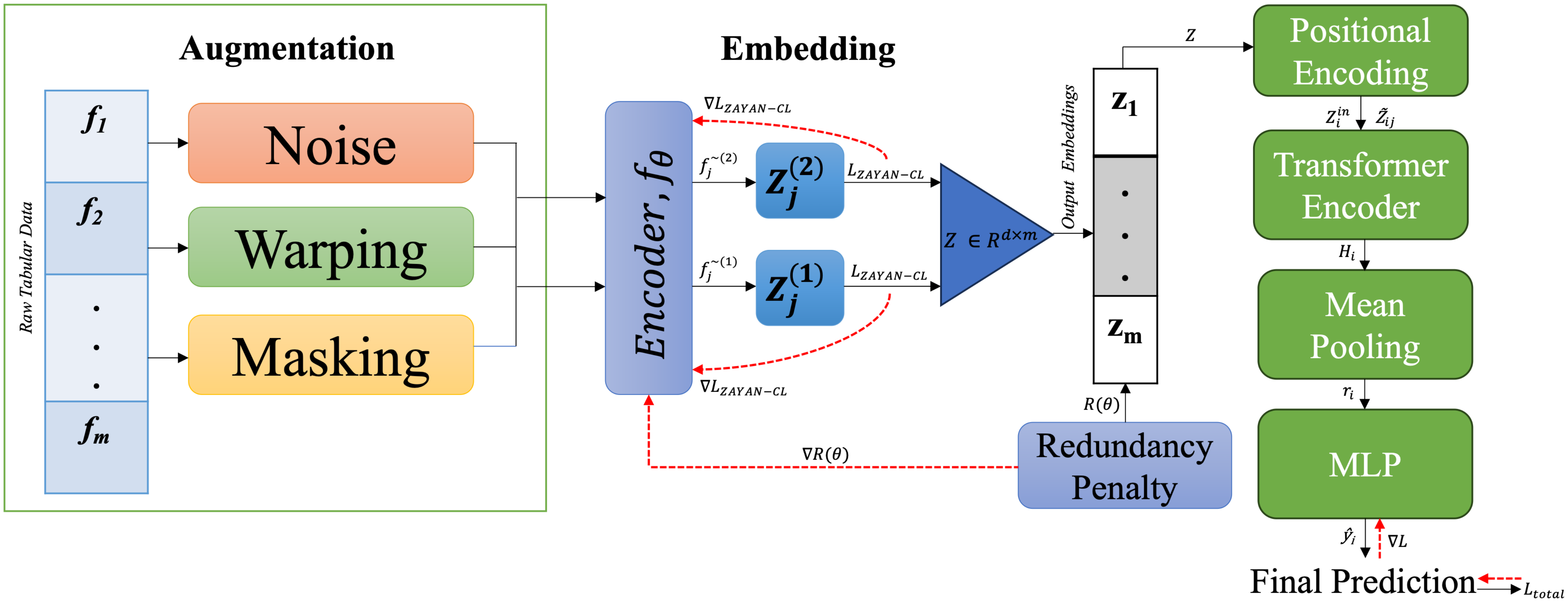
ZAYAN feature-level contrast



The positive pair is two augmented views of the same feature column. Other features act as negatives.

Design principle: separate feature representation learning from downstream supervised classification.

ZAYAN Architecture



ZAYAN-CL: Feature-Level Pretraining

1
Augment

Create two views of each feature using

Noise
Warping
Masking

2
Contrast

Same feature views **closer**

Different features **farther apart**

3
Disentangle

Redundancy penalty encourages

diverse feature embeddings

Output: robust, disentangled feature representations learned **without labels**.

ZAYAN-CL: Zero-Anchor Feature-Level Pretraining

Input and augmented views

Given $X \in \mathbb{R}^{N \times m}$, each feature column $f_j \in \mathbb{R}^N$ is treated as a learning unit.

For each feature:

$$\tilde{f}_j^{(1)}, \tilde{f}_j^{(2)} \leftarrow \text{Augment}(f_j)$$

Augmentations:

- Gaussian noise
- quantile warping with jitter
- random masking

Feature-level InfoNCE

$$\mathcal{L}_{\text{ZAYAN-CL}} = -\sum_{j=1}^{m} \log \frac{\exp\left(\frac{z_j^{(1)\top} z_j^{(2)}}{\tau \|z_j^{(1)}\| \|z_j^{(2)}\|}\right)}{\sum_{k \neq j} \exp\left(\frac{z_j^{(1)\top} z_k^{(2)}}{\tau \|z_j^{(1)}\| \|z_k^{(2)}\|}\right)}$$

- positive: two views of the same feature
- negatives: views of other features
- no explicit anchor sample or label is required

Redundancy Reduction in ZAYAN-CL

Redundancy Penalty

ZAYAN discourages different feature embeddings from pointing in similar directions.

$$\mathcal{R}(\theta) = \sum_{i \neq j} \left(\frac{\mathbf{z}_i^\top \mathbf{z}_j}{\|\mathbf{z}_i\| \, \|\mathbf{z}_j\|} \right)^2$$

Equivalent matrix form

$$\mathcal{R}(\theta) = \left\| \mathbf{Z}^\top \mathbf{Z} - \mathbf{I} \right\|_F^2$$

Goal: reduce similarity and redundancy among feature embeddings.

ZAYAN-CL objective: $\min_{\theta} \; \mathcal{L}_{\text{ZAYAN-CL}}(\theta) + \lambda \, \mathcal{R}(\theta)$

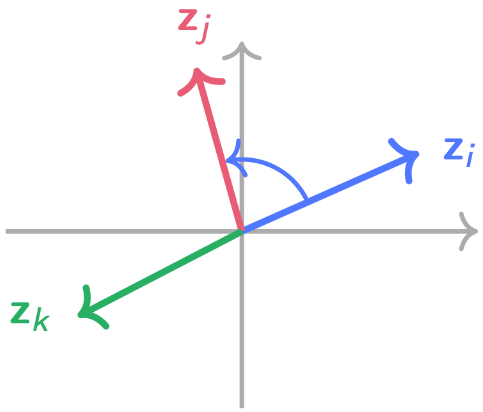
Geometric Interpretation

Same feature views
Pulled **closer together**.

Different features
Pushed toward **angular separation**.

Redundant directions
Penalized by cosine similarity.

Result
More **diverse, disentangled embeddings**.



Lower cosine similarity
$\Downarrow$
Greater angular diversity

Takeaway: ZAYAN separates feature directions to obtain **less redundant, more disentangled representations**.

ZAYAN-T: From Feature Embeddings to Tokens

After ZAYAN-CL Pretraining

For each sample,

$$\mathbf{Z}_i = [\mathbf{z}_{i1}, \ldots, \mathbf{z}_{im}]$$

Each learned feature embedding

$$\mathbf{z}_{ij} \in \mathbb{R}^d$$

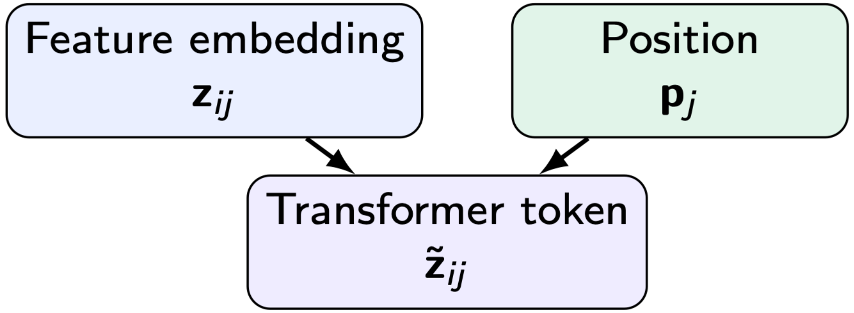
acts as a **feature token**.

Key point:
the Transformer receives the feature geometry learned by ZAYAN-CL.

ZAYAN-Aware Token Input

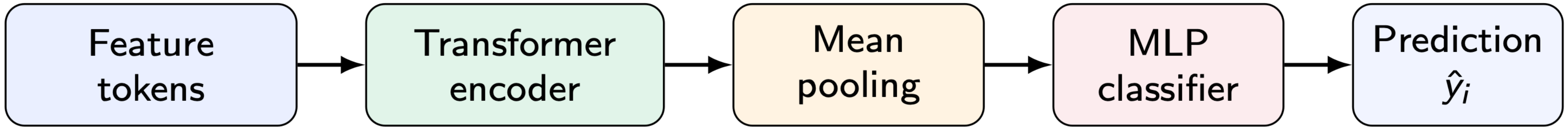
Positional information is added to each feature embedding:

$$\tilde{\mathbf{z}}_{ij} = \mathbf{z}_{ij} + \mathbf{p}_j$$



Output: a sequence of structured feature tokens.

ZAYAN-T: Structure-Preserving Classification



The Transformer models **interactions among learned feature tokens**, then mean pooling produces the representation used for classification.

Supervised objective

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{CE}} + \gamma \mathcal{L}_{\text{preserve}}$$

Structure preservation

$$\mathcal{L}_{\text{preserve}} = \sum_{i,j} \|\mathbf{z}_{ij} - \mathbf{h}_{ij}\|^2$$

Keeps downstream tokens aligned with the **pretrained feature geometry**.

Experimental Setup

8 datasets

Six remote sensing tabular benchmarks and two flood-prediction datasets.

31 baselines

Classical ML, tree ensembles, deep tabular models, Transformers, and TabPFN-style methods.

5-fold CV

Mean accuracy and standard deviation, with tunable models optimized by Optuna.

Benchmarks include noisy, subsampled, balanced, PCA-compressed, and high-dimensional environmental settings.

Key Results Against Top Baselines

5-fold CV accuracy (%) $\pm$ std.

Model	Urban	Wilt	Crop*	Forest	Tree+	RSI#	Pluvial$	Indian	Avg. Rank
TabR	83.21±1.66	98.78±0.17	98.62±0.68	90.25±1.94	64.78±0.42	99.32±0.18	89.88±0.62	49.59±1.88	9.25
TabM	82.07±0.98	98.80±0.20	98.92±0.96	88.72±0.98	65.34±0.24	99.36±0.10	89.64±0.98	51.49±0.92	6.94
TANDEM	83.29±1.53	98.37±0.39	99.18±0.12	90.05±2.07	66.96±0.12	99.72±0.14	90.72±0.32	50.26±1.55	6.06
TabICL	83.30±0.98	98.76±0.17	99.46±0.10	90.05±2.35	66.98±0.20	99.74±0.16	90.75±0.74	51.27±0.66	3.38
ZAYAN	**84.80±7.10**	**99.69±0.40**	**99.66±0.34**	**97.21±0.45**	**67.00±0.12**	**99.77±0.15**	**93.61±0.47**	**51.55±0.43**	**1.06**

ZAYAN ranks first overall and remains consistently strong across remote sensing and flood-related datasets.

Takeaway: the gain is not confined to one dataset or regime; **ZAYAN achieves the best overall cross-dataset ranking**.

Ablation: What Makes ZAYAN Work?

Component-level evidence

Removing key components hurts performance.

No redundancy reduction → lower accuracy

No contrastive learning → weaker embeddings

No augmentations → lower robustness

Anchor-based variants → weaker than full ZAYAN

Message from ablations

The model is not only a Transformer classifier. Its gains come from the combination of:

1. feature-level contrastive alignment,
2. redundancy-aware angular separation,
3. dynamic perturbation and masking,
4. structure-preserving downstream learning.

Inference-Time Diagnostics

Robustness

Accuracy is nearly stable up to 25% shuffled features; feature dropping is more damaging because it removes information.

Selective prediction

Higher confidence thresholds improve retained accuracy, trading coverage for reliability.

OOD behavior

Noisy and permuted inputs lower confidence and increase predictive entropy relative to in-distribution data.

Deployment-style triage on Urban Land Cover reaches AUC $= 1.000$ for a one-vs-rest binary task, with efficient batch inference.

Conclusion

ZAYAN separates feature representation learning from downstream classification.

Feature-level SSL

Learns from columns, not only rows or image patches.

Zero-anchor objective

Avoids explicit sample anchors and label dependence during pretraining.

Remote sensing gains

Improves accuracy, robustness, and diagnostics across eight benchmarks.

Future direction: sparse or low-rank attention for very high-dimensional feature spaces.

Resources and Acknowledgment

Project links

Paper: `https://link.springer.com/chapter/10.1007/978-3-032-31397-3_1`

arXiv: `https://arxiv.org/abs/2604.27606`

GitHub: `https://github.com/zadid6pretam/ZAYAN`

Try or ask

Install: `pip install zayan`

Issues: `https://github.com/zadid6pretam/ZAYAN/issues`

Contact: `ah00069@mix.wvu.edu`

Acknowledgment: This work was supported in part by the International Association for Pattern Recognition (IAPR) through ICPR 2026 registration support.

Thank You!

Questions?

github.com/zadid6pretam/ZAYAN/issues

ah00069@mix.wvu.edu