Geometry-Aware Tabular Diffusion

Attention learns relationships. Geometry computes them.

Paper, code, and datasets
morphresearch.ai/gatd

ICML 2026 Poster

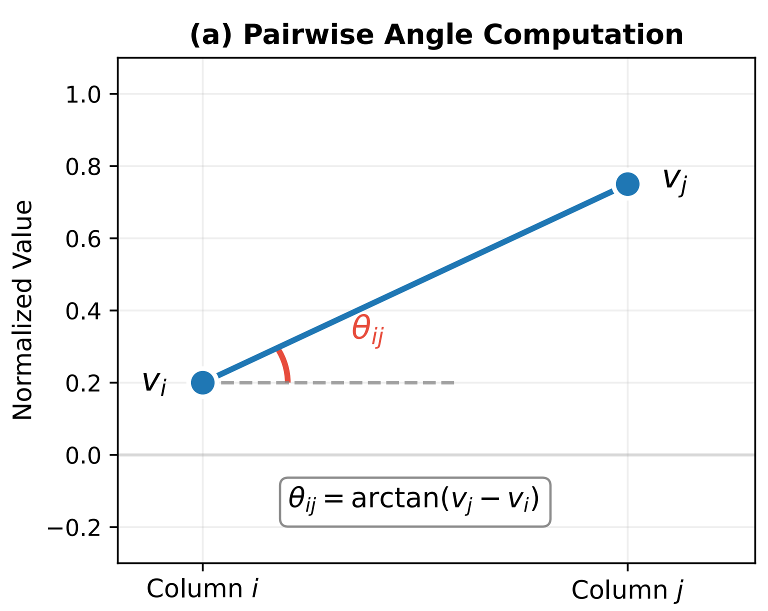
David Turtora Zagardo
Morph Research
dave@morphresearch.ai

1 MOTIVATION / PROBLEM

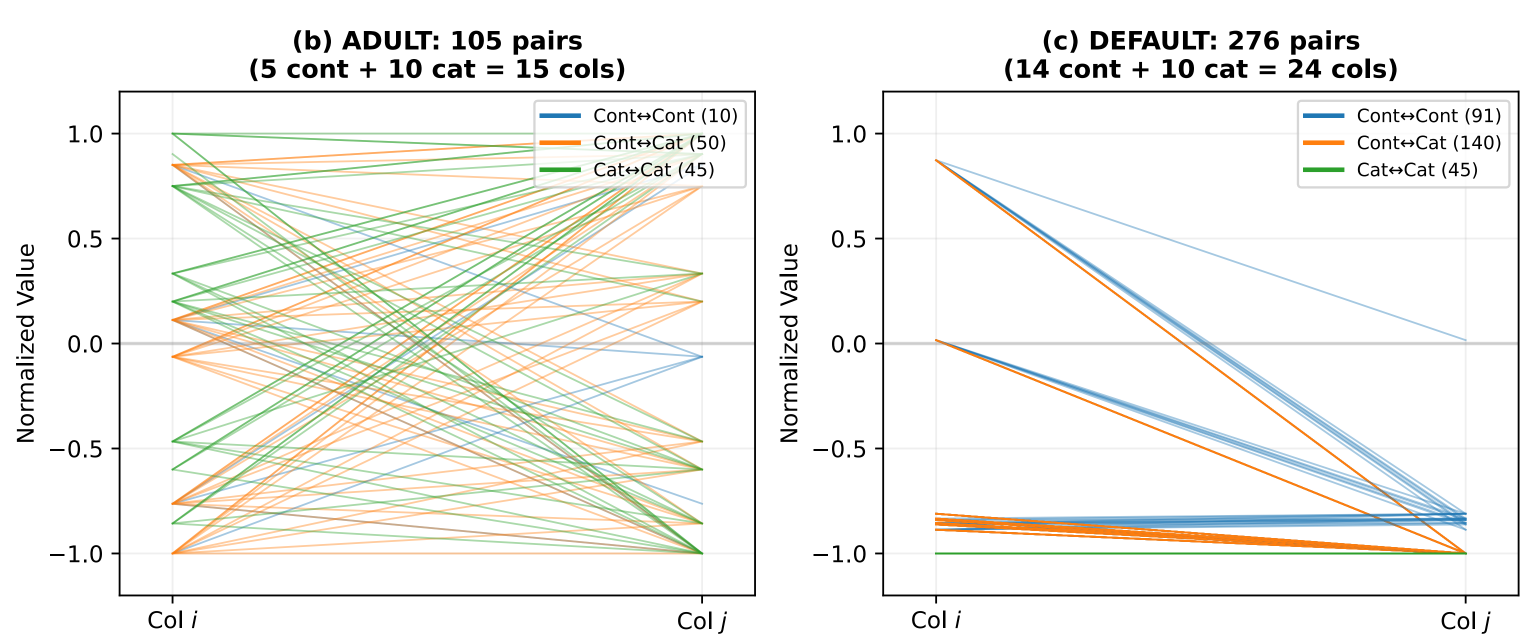
- Tabular diffusion models must capture complex inter-column relationships.
- Transformer baselines often rely on column self-attention to learn these relationships implicitly which is slow and costly.
- Can we provide explicit relational structure instead of leaving all relationships to be learned implicitly?
- Main idea: compute pairwise geometric relationships directly from column values and use them as inputs and auxiliary targets.

2 KEY IDEA: PAIRWISE GEOMETRY

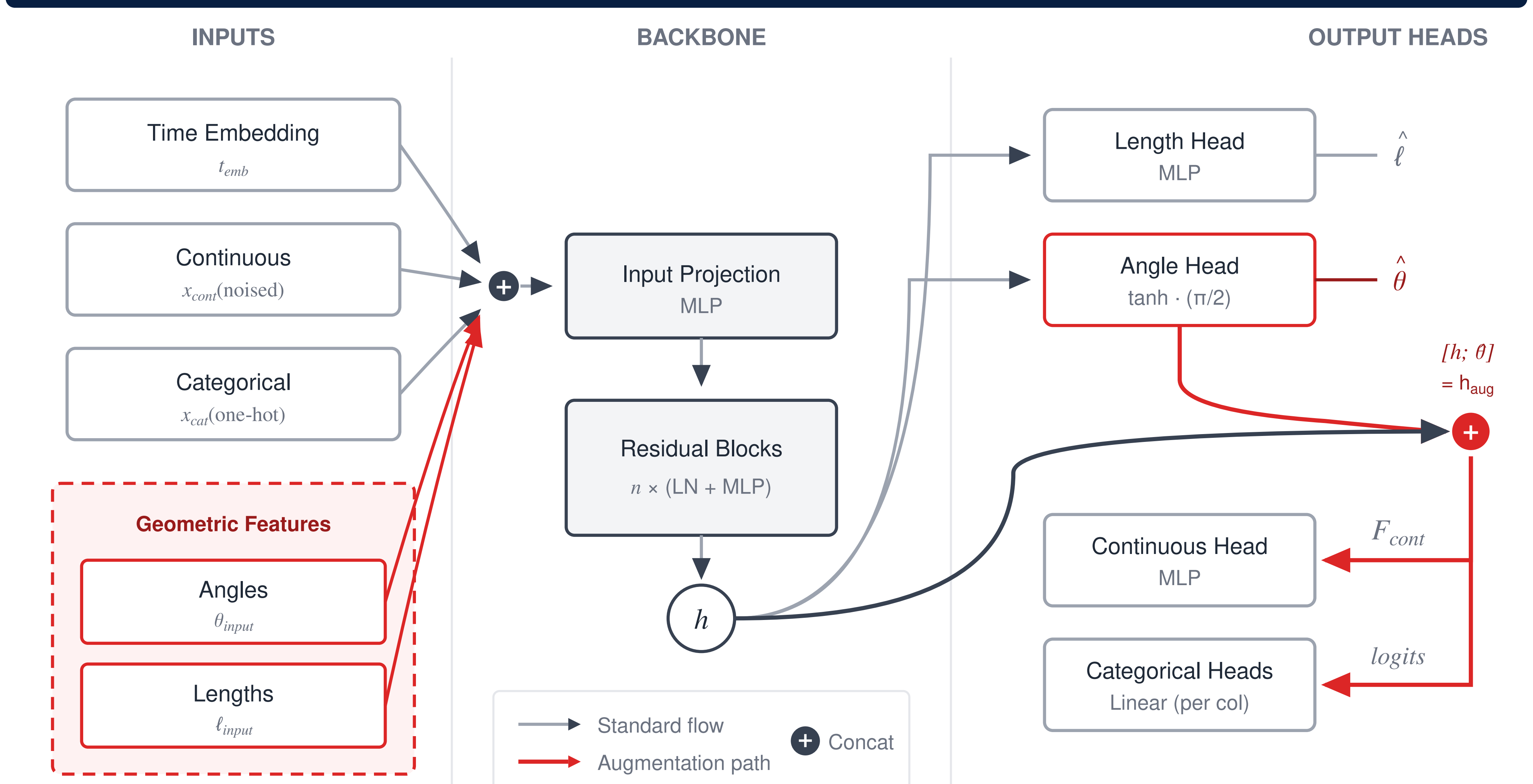
For each row, compute pairwise relationships between columns (features). Angles give us both direction and magnitude, while lengths only provide magnitude. Though $O(d^2)$ and quadratically complex, pairwise geometry creates rich, efficient and learnable features that allow for faster per-epoch and per-pipeline runtime compared to transformer alternatives.



Geometry tells a model exactly how column V_i should behave when column V_j looks like it does for a given row. **This encodes direct relational structure and allows the model to learn an exact, shift-invariant representation of the tabular data.**



3 METHOD: GEOMETRY-AWARE TABULAR DIFFUSION



Geometry is used as both inputs and auxiliary targets.

4 EXPERIMENTAL SETUP

- 10 Benchmark datasets
- 5 classification, 5 regression tasks
- Shape, Trend, R^2, AUROC, F1, RMSE
- MLP, GNN, Transformer
- 20k epochs, 3x train, 20x gen

5 MAIN RESULTS vs. TabDiff (SOTA Transformer-based Diffusion Baseline)

1.5× better utility

1.7× faster train + gen

3.5× smaller models

Shape wins vs. TabDiff	08/10
Trend wins vs. TabDiff	07/10
Downstream utility wins (AUROC, R^2, F1, RMSE) vs. TabDiff	16/20

5 MAIN RESULTS +Geom vs. NoGeom

29% better Shape

18% better Trend

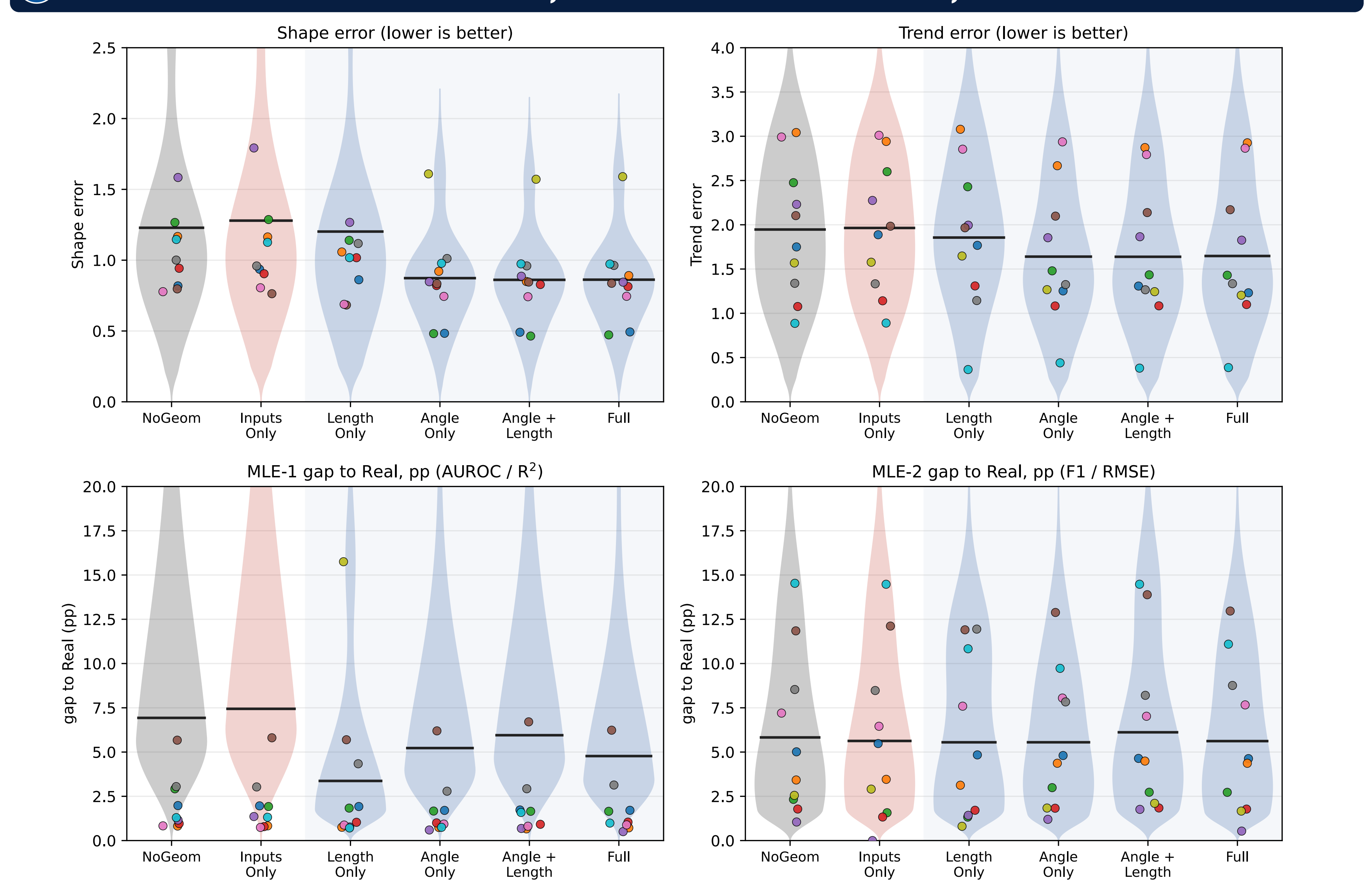
39% better AUROC/R^2

16% better F1/RMSE

Shape wins MLP, GNN, Transformer	27/30
Trend wins MLP, GNN, Transformer	25/30
Shape + Trend wins (no ties) (all architectures)	52/60
AUROC/R^2 wins, 1 tie MLP, GNN, Transformer	20/30
F1/RMSE wins, 3 ties MLP, GNN, Transformer	19/30
Shape + Trend wins, ties (all architectures)	39/60

52 / 60 Shape + Trend cells; exact sign test $p = 5.21 \times 10^{-9}$

6 ABLATION: SUPERVISION, NOT EXTRA INPUTS, DRIVES THE GAIN



Without supervised geometry losses, the geometric heads do not learn useful signal.

7 TAKEAWAYS / CONCLUSION

- **Explicit relational supervision is a portable inductive bias for tabular diffusion.**
- **The MLP instantiation of GATD matches or exceeds strong transformer-based baselines.**
- **Default geometry loss weights (15A, 15L, 8C) transfer across MLP, GNN, and Transformer denoisers.**
- **Geometry can replace learned attention as the primary mechanism for relationship modeling.**
- **Both column-wise attention and pairwise geometry are $O(d^2)$, but attention requires many computational operations per step, while geometry is pre-computed at the start of training, provides similar information, and runs meaningfully faster per epoch.**