# A Graph Neural Network Approach for Localized and High-Resolution Temperature Forecasting

Joud El-Shawa[1,2] Elham Bagheri[1,2] Sedef Akinli Kocak[1] Yalda Mohsenzadeh[1,2*]

[1]Vector Institute for Artificial Intelligence [2]Western University

## Introduction

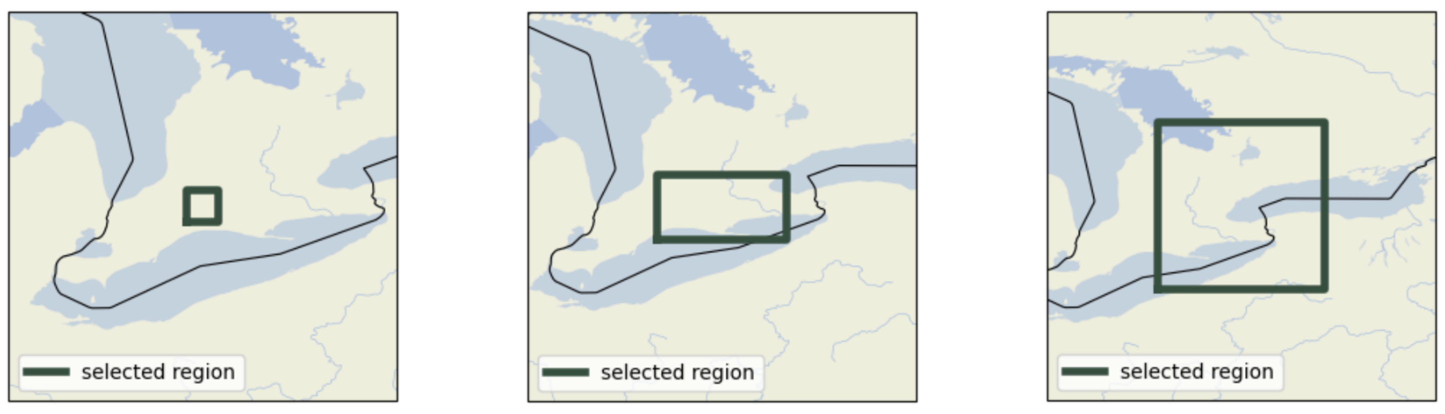
Heatwaves are among the deadliest climate hazards (∼**489k** deaths/year, 2000–2019 [3]) and disproportionately harm low-income, racialized, and Global South communities [5, 2].

Most operational forecasts run at 10–30 km, smoothing over urban heat islands and neighborhood "hot spots," which leads to systematic underestimation exactly where targeted action is needed.

We introduce a high-resolution (2.5 km) Graph Neural Network framework that produces localized temperature forecasts 1–48 hours ahead, which can be post-processed to match local heatwave definitions, supporting equitable early-warning workflows.

## Data

We use NOAA's URMA dataset (2.5 km, hourly). We predict 2-meter air temperature as a low-level signal, and focus on Southwestern Ontario across three nested domains, seen in Figure 2, which cover mixed land types (urban, farmland, forest, water).




(a) Region A (b) Region B (c) Region C

Figure 1: Bounding boxes around Regions A–C.

## Embeddings Framework

Data from resource-limited regions is often sparse, inconsistent, and difficult to align with information-rich datasets.

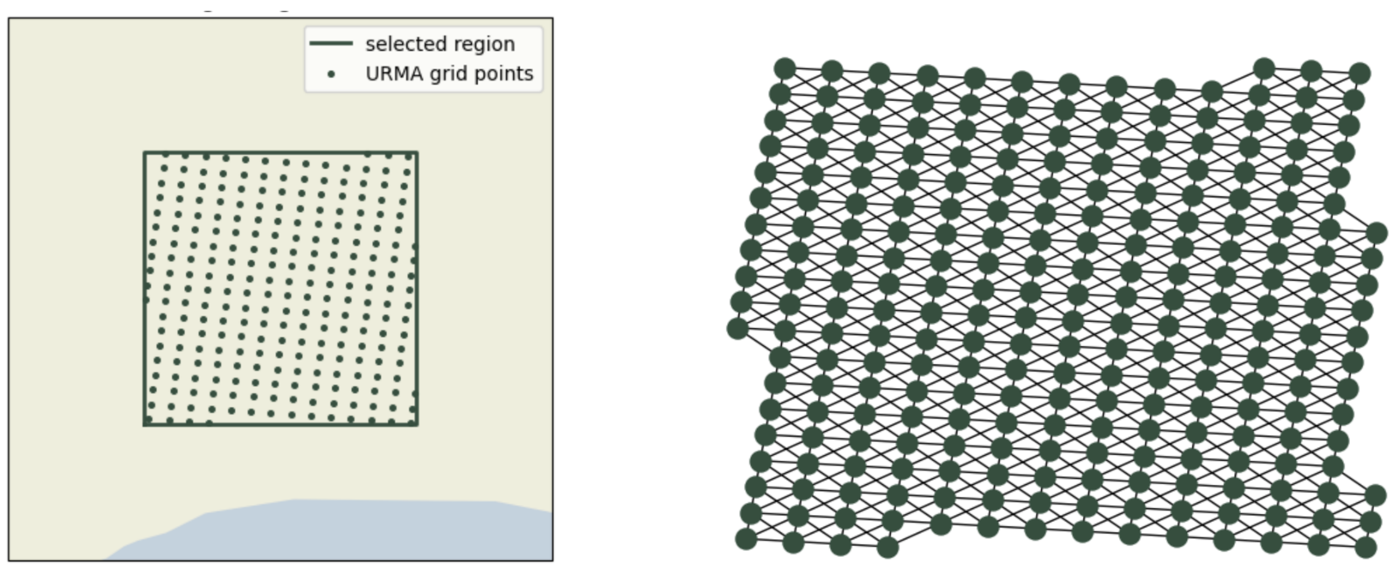
In a separate setup, we also explore language-model embeddings as an intermediate representation. Each Region A observation is transformed into a short paragraph, for example:

```
temperature is 291.6 K, dew point is 283.7 K, u wind component
is 4.0 m/s, v wind component is -2.1 m/s, surface pressure is
99209 Pa, ...  , elevation is 172.0 meters.
```

These descriptions are encoded using ClimateBERT [6], which become node features in the pipeline.

## GNN-Based Framework

A hybrid Graph Convolutional Network (GCN) with a Gated Recurrent Unit (GRU) was trained for each region.




(a) Sample selected region. (b) Corresponding graph.

Figure 2: Graph setup. Each grid point in the region is represented as a graph node with meteorological features, connected by edges to capture spatial interactions.

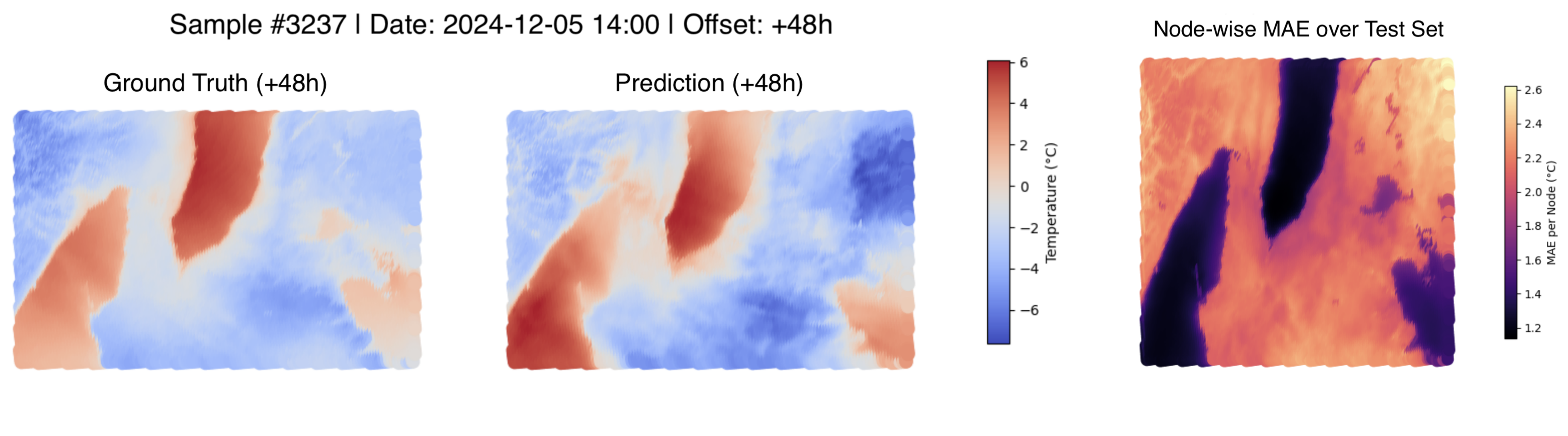
Graph convolution layers model neighborhood effects [4], while GRUs capture temporal dependencies [1]. The objective was to predict temperature at 1, 6, 12, 18, 24, 36, and 48h ahead from the current time.

## Results

Table 1: Per-region performance. Mean across horizons; 48h at the farthest forecast horizon.

| Region | Mean MAE (°C) | MAE@48h (°C) | RMSE@48h (°C) |
|---|---|---|---|
| A | 2.55 | 3.78 | 4.84 |
| B | 2.48 | 3.73 | 4.84 |
| C | 1.93 | 2.93 | 3.90 |

Training on Region C strains memory, so we also sampled every 6h. The 6h model reached mean MAE 2.39, MAE@48h 3.15, and RMSE@48h 4.16—i.e., modest degradations (+0.46, +0.22, +0.26) for substantially lower compute.




(a) Randomly sampled test timestamp showing ground truth and model predictions 48 hours ahead in Region C.

(b) Average node-wise MAE across the test set in Region C.

Figure 3: Example results from Region C.

Table 2: Region A performance comparisons. Mean across horizons; 48h at the farthest horizon.

| Model | Mean MAE (°C) | MAE@48h (°C) | RMSE@48h (°C) |
|---|---|---|---|
| Baseline (tabular) | 2.55 | 3.78 | 4.84 |
| Embeddings (ClimateBERT) | 3.34 | 4.34 | 5.54 |
| Control (random weights) | 9.11 | 8.89 | 10.49 |

## Key Insights & Impact

- Performance improved as the spatial window expanded (Region A → Region C).
- On Region A, performance with embeddings shows a modest decrease in mean MAE relative to the tabular baseline, while a control model performs noticeably worse, suggesting that embedding features carry meaningful signals.

Localized forecasting is an equity issue: coarse models miss neighbourhood-level risks that disproportionately affect marginalized communities. Our approach offers a transferrable solution: models trained in data-rich regions can be adapted to under-monitored contexts to strengthen early warnings and resource allocation.

## Conclusions & Next Steps

We present a 2.5 km GCN–GRU framework for 2-meter temperature forecasting across 3 regions, with performance improving as spatial context increases; a 6-hour sampling variant preserves most skill, and an embedding approach standardizes heterogeneous inputs.

Next steps include:

- Building matched-resolution baseline models.
- Broadening geographic coverage to additional regions.
- Extending the framework to other climate extremes (wildfire, floods, drought).

## References

[1] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation, 2014. URL https://arxiv.org/abs/1406.1078.

[2] T. A. Deivanayagam, S. English, J. Hickel, J. Bonifacio, R. R. Guinto, K. X. Hill, M. Hug, and R. Issa. Envisioning environmental equity: climate change, health, and racial justice. *The Lancet*, 2023. doi: https://doi.org/10.1016/S0140-6736(23)00919-4.

[3] Q. Z. et al. Global, regional, and national burden of mortality associated with non-optimal ambient temperatures from 2000 to 2019. *The Lancet Planetary Health*, 5(7):e415–e425, 2021.

[4] T. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations (ICLR)*, 2017.

[5] E. S. Parsons, A. Jowell, E. Veidis, M. Barry, and S. T. Israni. Climate change and inequality. *Pediatric Research*, 2024.

[6] N. Webersinke, M. Kraus, J. Bingler, and M. Leippold. ClimateBERT: A Pretrained Language Model for Climate-Related Text. In *Proceedings of AAAI 2022 Fall Symposium: The Role of AI in Responding to Climate Challenges*, 2022. doi: https://doi.org/10.48550/arXiv.2212.13631.