Text Is All You Need

Predicting Conflict Escalation Using Global News Content

Diana Kolusheva
dik431@g.harvard.edu
Harvard University

Chad Stoughton
ch0793@g.harvard.edu
Harvard University

Evan Wheeler
evw217@g.harvard.edu
Harvard University

Introduction

- Forecasting conflict escalation is of interest to policy makers, executives, and social scientists.
- In our paper, we introduce:
  - A novel binary classification metric for conflict escalation
  - Three methods to extract generalizable conflict signals directly from document embeddings
    - Text with country characteristics
    - Text with country labels
    - Text alone

Challenge of global models

- Most contemporary work focuses on country or region-specific models.
  - This introduces a “small-data problem”
- We chose to persue a global model, a relatively new area of research in this domain
  - This introduces a new problem of accounting for local conditions

Defining Conflict Escalation

- Modeling conflict escalation requires a precise definition of what conflict escalation means.
  - Many definitions in the literature.
- A global model requires a metric that is consistent acoss local conditions.
  - The presence or absence of conflict is not sufficient.

- We used data from from the Armed Conflict Location & Event Data Project (ACLED) for our raw event and fatality counts

Our “Spike” Conflict Metric

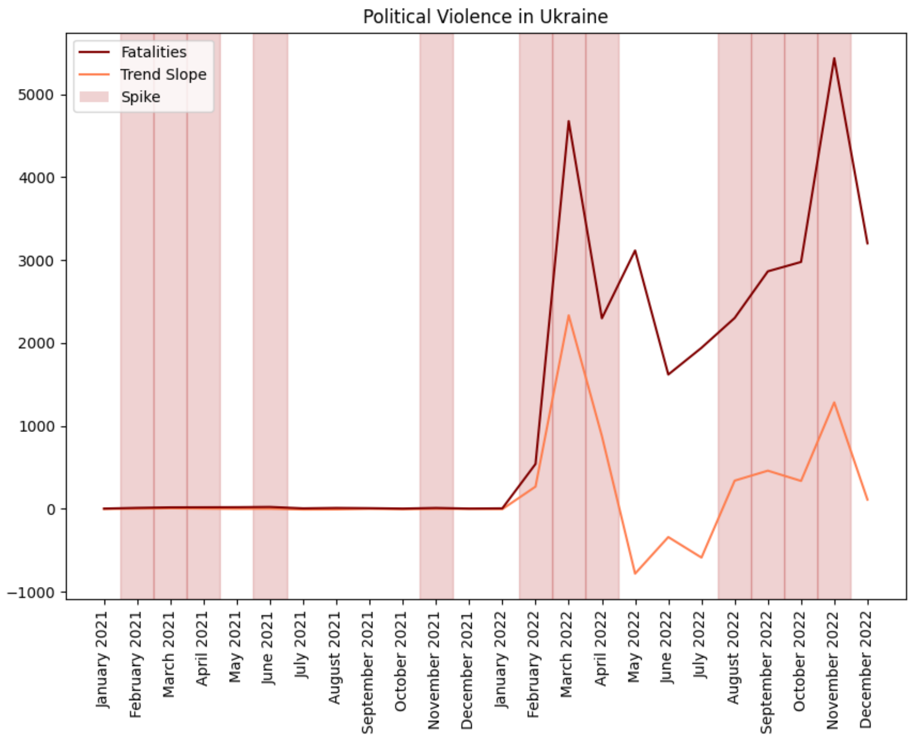


Figure 1: The Spike metric applied to Ukraine

- Highlights relative “spikes” in political violence.
- Sensitive to small increases in political violence in peaceful countries/periods.
- Filters out small fluctuations in long-standing conflicts.

Training Dataset

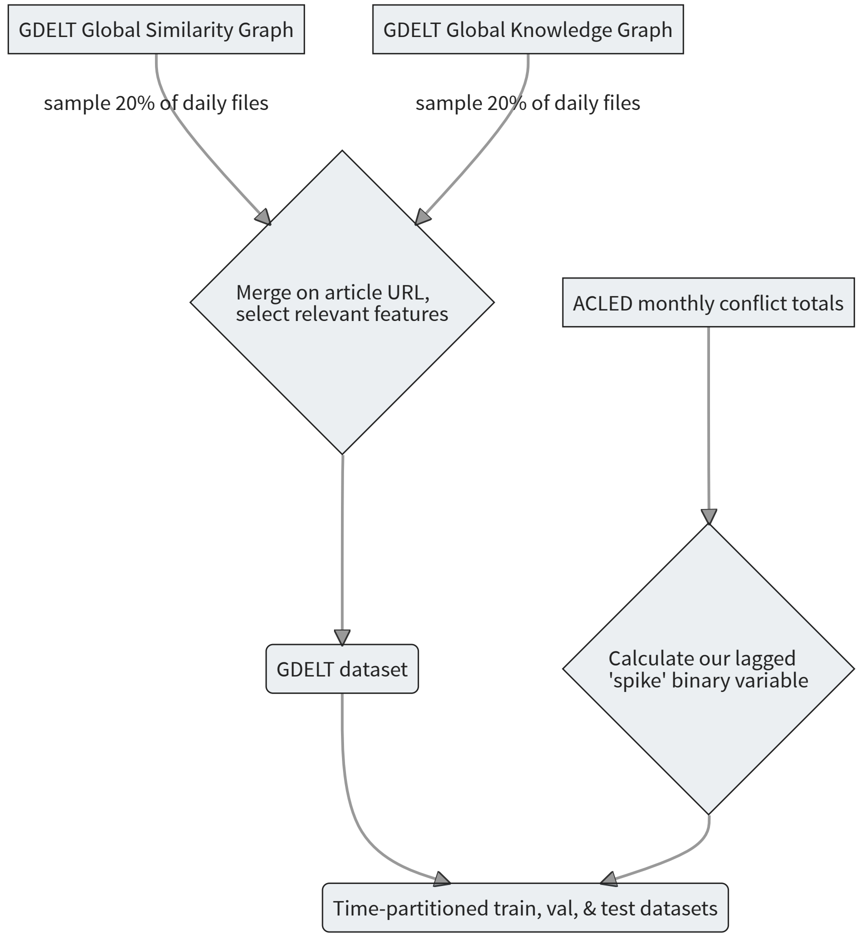


Figure 2: Data generation process

Conflict Prediction Using Text and Country Features

- Our first modeling approach employed a feed forward neural network model.
- Uniform manifold approximation and projection (UMAP) was used to reduce dimensionality of GDELT article embeddings in a supervised manner.

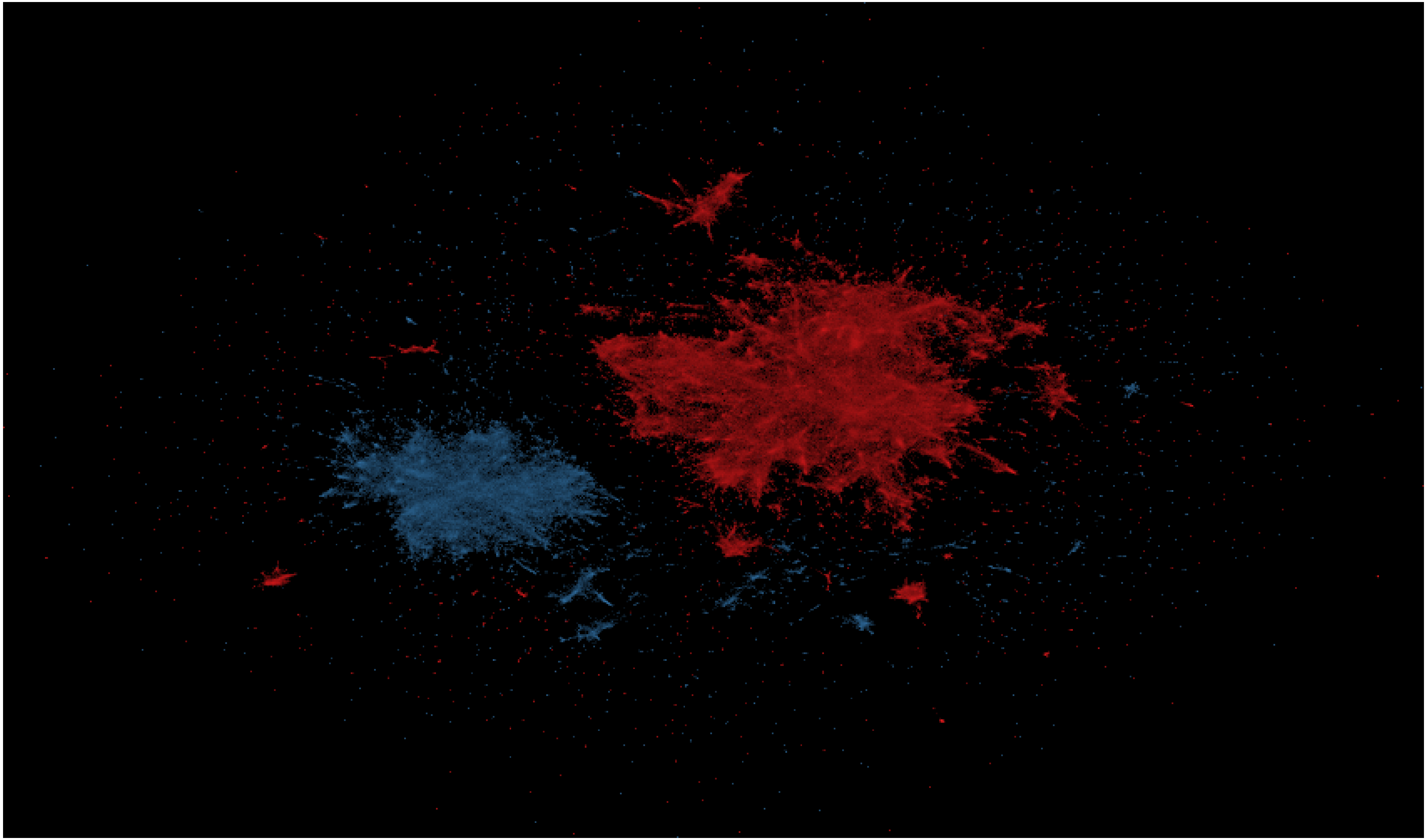


Figure 3: Scatterplot of 2-dimension UMAP-reduced articles with spike labels

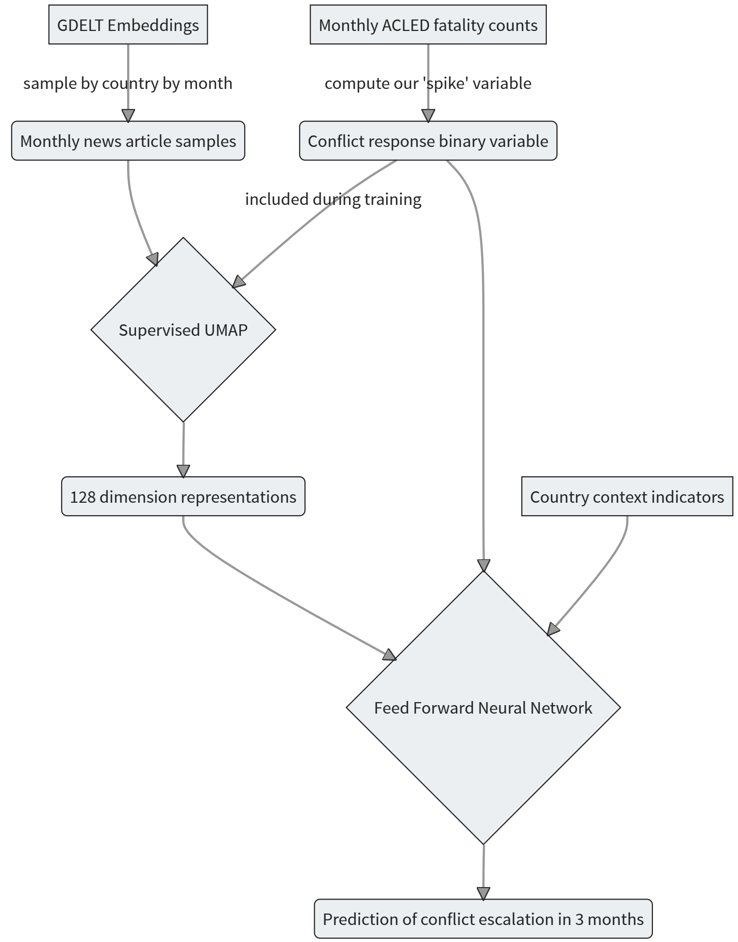


Figure 4: Feed forward neural network model architecture

Conflict Prediction Using Text and Country Labels

Conflict Prediction Using Text Alone

- Our third approach used a pair of encoder-only transformers
- The “encoder” and the “classifier”.
- The encoder distils the embedding corpus by analyzing a subset of article corpus and reducing it’s dimensionality.
- The classifier makes final predictions on the distilled dataset.

The Encoder

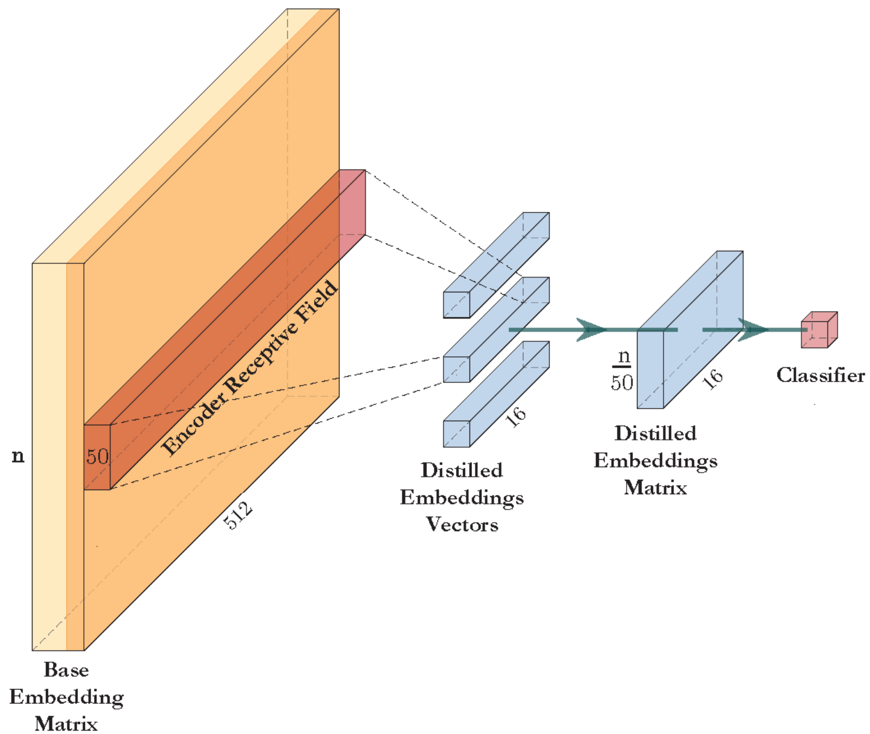


Figure 6: Transformer Model

- Trained using a dataset that drew multiple samples per country per month of 50 articles each
- This created multiple unique training examples for each country and month
- Once trained, the Encoder's weights were frozen and penultimate layer was exposed

The Classifier

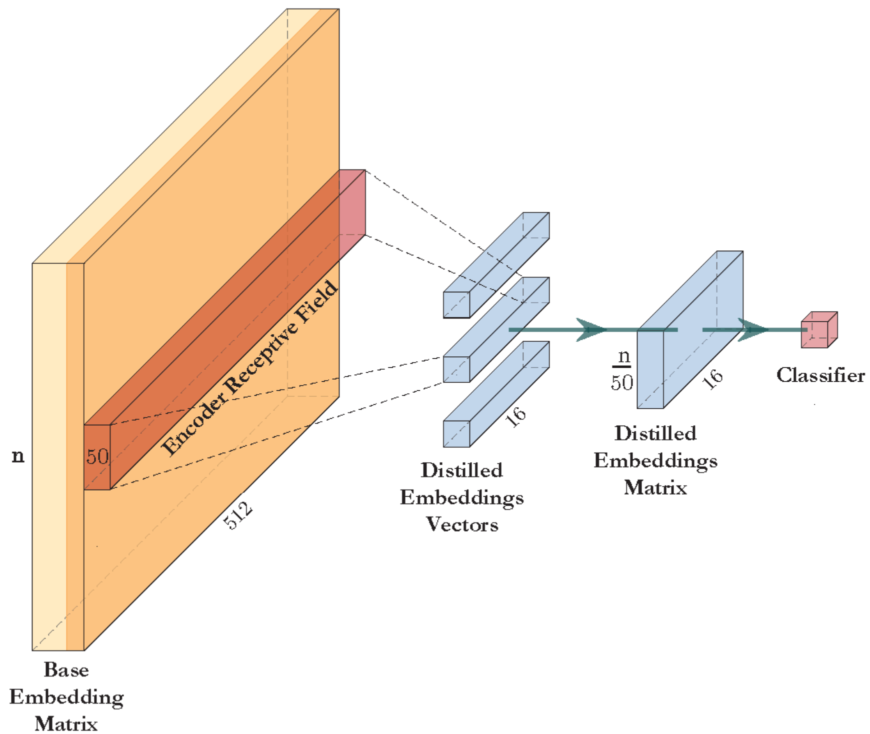


Figure 7: Transformer Model

- Each country/month embedding matrix was passed into the Encoder in batches of 50 articles
- This produced a single distilled training example for each country and month
- This final, undistorted data set was then used for classification

Ethical Considerations

- Bias in the training data
- Adversarial attacks
- Self-fulfilling prophesies

Mitigating Risk

- Models, not oracles
- Not a substitute for careful deliberation and dialogue
- At their best, these models can provide an early warning for conflict, and if the right actions are taken, then they can be made to be wrong

Results: metrics

metric	FFNN	XGBoost	Transformer
Cohen Kappa Score	0.18	0.09	0.11
F1 Score	0.33	0.28	0.28
Precision Score	0.27	0.17	0.20
Recall Score	0.41	0.72	0.47
Roc Auc Score	0.67	0.69	0.64
True Negative Rate	0.81	0.49	0.70

Figure 8: Model metrics (large values are better)

Results: comparison to prior work

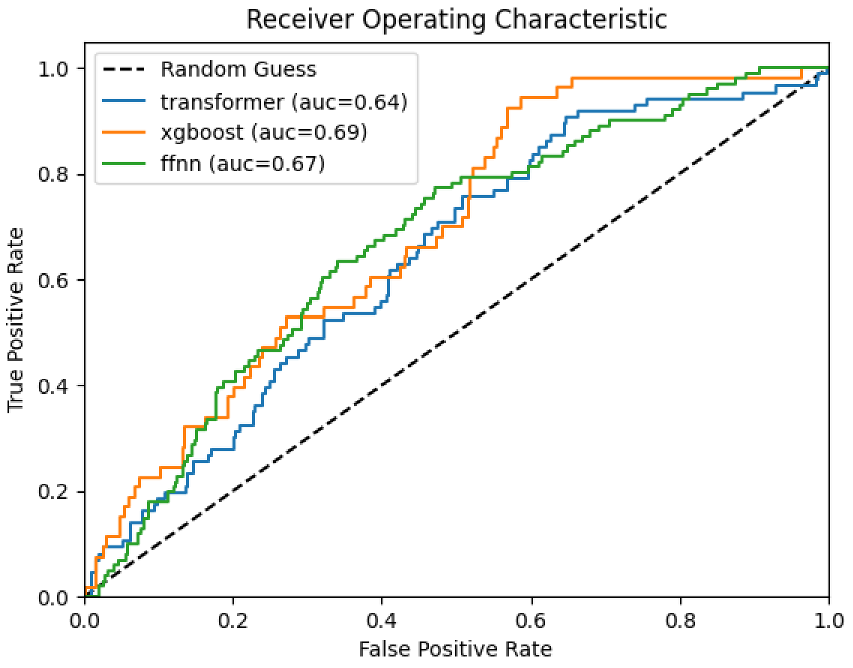
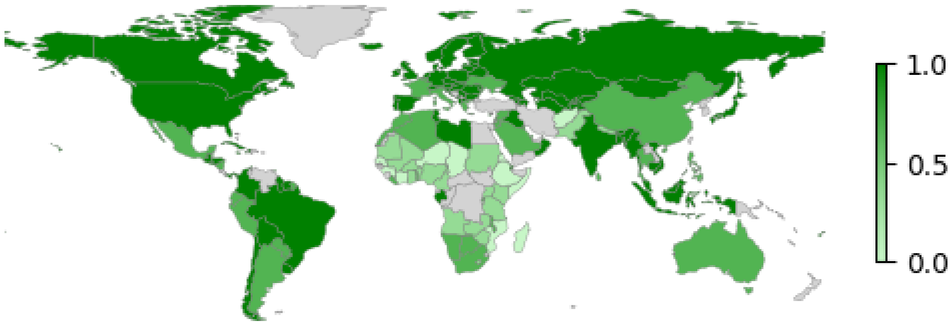


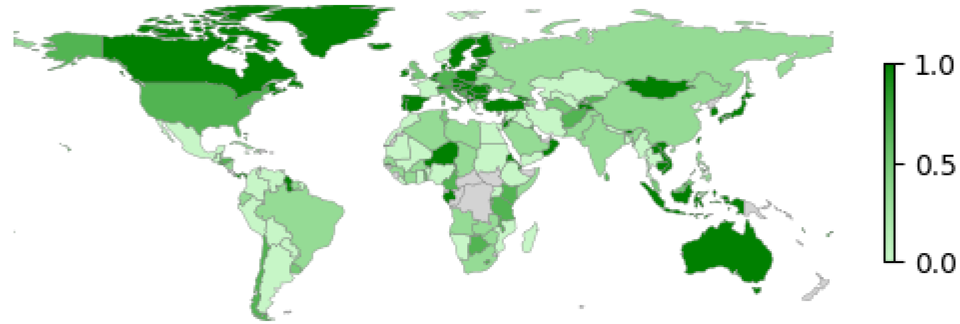
Figure 9: Receiver Operator Characteristics

- Comparison is not very meaningful because our target is different
- Comparing to 0.83 overall and 0.75 for hard cases in paper by Mueller and Rauh 2022 (another binary classification problem)
- Our best ROC_AUC is 0.69

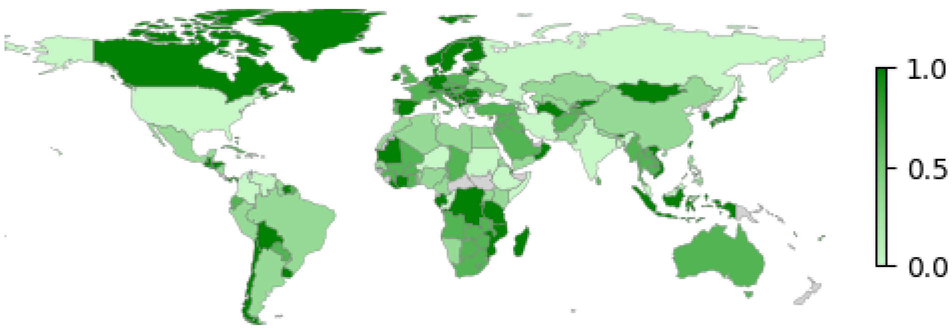
Results: generalizability



(a) FFNN



(b) XGBoost



(c) Transformer

Results: practicality

- Practical goal – a tool that can assist UNICEF
- Cost & complexity are important
- Average monthly embeddings dataset & corresponding statistical model fit in memory and can be trained on CPU
- Article-level embeddings in our other methods require additional data engineering steps to be processed in batches and need GPU to train

So which model is best?

- All three approaches produced promising results
- The model choice depends on:
  - user’s goals
  - priority performance metrics
  - available resources

Future work

- Increase size of training set (we only used 20%)
- Continue experimentation with data sampling and aggregation approaches and model tuning
- Research regional disparities in model performance
- Build an application with user interface
- Build a more robust and automated data pipeline

Conclusions

- Text-only data vs data enriched with country labels and social indicators
- Conflict escalation definition
- Text embeddings provide a strong enough signal for predictions
- Comparable results across different data sampling and modeling approaches
- Neural network-based models – better generalization across geographies
- Statistical model – smaller size/cost of deployment

So what?

- We hope this work will contribute to both UNICEF operations and political science research
- The performance results can serve as a benchmark for future research projects in this field

Thank you