

# Graph Representation Learning for Dengue Forecasting

1<sup>st</sup> Jiaqi Weng

*Computer Science Department  
Rensselaer Polytechnic Institute  
Troy, USA  
wengj3@rpi.edu*

2<sup>nd</sup> David Qiu

*Computer Science Department  
Rensselaer Polytechnic Institute  
Troy, USA  
qiudavid512@gmail.com*

3<sup>rd</sup> Ethan Cruz

*Computer Science Department  
Rensselaer Polytechnic Institute  
Troy, USA  
cruzethan@outlook.com*

4<sup>th</sup> Malik Magdon-Ismael

*Computer Science Department  
Rensselaer Polytechnic Institute  
Troy, USA  
magdon@cs.rpi.edu*

5<sup>th</sup> Thilanka Munasinghe

*School of Science  
Rensselaer Polytechnic Institute  
Troy, USA  
munast@rpi.edu*

6<sup>th</sup> Jennifer.C.Wei

*NASA Goddard Space Flight Center  
Greenbelt, USA  
jennifer.c.wei@nasa.gov*

7<sup>th</sup> Ashan Pathirana

*Health Promotion Bureau  
Sri Lanka  
ashanpathirana@gmail.com*

**Abstract**—In 2017, the largest recorded dengue outbreak in Sri Lanka’s history occurred. Since then, dengue has continued to threaten national health across Sri Lanka. The development of an effective Early Warning System (EWS) for dengue outbreaks is essential for Sri Lanka’s Ministry of Health to take preventative measures. We propose the use of Graph Neural Networks as EWS. Using earth observational data from NASA’s global satellites and dengue incidence data from Sri Lanka’s Ministry of Health, we developed a series of traditional and graph representation EWS to forecast Dengue cases across Sri Lanka’s 25 districts between 2013 and 2022. We demonstrate empirically that Graph Neural Networks which incorporate spatiotemporal relations significantly outperform traditional EWS such as Autoregressive Integrated Moving Average (ARIMA), Random Forest, and Long Short-Term Memory (LSTM). Our source code is available on GitHub and will be provided in the final submission.

**Index Terms**—Disease Forecasting, Bioinformatics, Graph Neural Networks

## I. INTRODUCTION

In 2017, the largest recorded dengue outbreak in Sri Lanka’s history occurred. Over 186,000 dengue cases and 440 deaths were reported that year (1). Dengue spread throughout Sri Lanka using mosquitos as their vector, despite the government’s attempts to spray mosquito breeding sites and use chemical fogging to kill mosquitoes (1). This was not a one-off event either. Historical dengue case patterns in Sri Lanka display alarming growth from the 2000s to the 2010s (2).

As dengue continues to worsen in Sri Lanka, additional preventative measures are required. An Early Warning System (EWS) which seeks to predict future dengue cases using historical data could help the government increase vector mitigation weeks before outbreak events. A review of EWS which

monitored many mosquito-borne diseases such as malaria and dengue showed promising success rates (3). The review showed almost all EWS used weather patterns as a predictor for dengue outbreaks (3). Analysis of Sri Lanka’s historical records displayed a positive correlation between dengue cases and rainfall patterns on the island (2; 4).

Ground cover can also be used for improved prediction. The use of landscape features and vegetation indices have also correlated with dengue cases (5). Factors such as Normalized Difference Vegetation Index (NDVI) had a negative correlation with dengue outbreaks, suggesting heavy urban areas with little vegetation create a dangerous environment suited for dengue outbreaks (6). Additional research of dengue across the globe suggests socioeconomic factors such as population density and human mobility affect dengue outbreaks in cities and across municipality borders (5; 7). Previous EWS for dengue which included human mobility resulted in increased outbreak prediction (8).

**Problem Novelty.** Existing works mostly view EWS as a time series prediction problem where the features (prior rainfall, ground cover, etc.) are used to predict dengue outbreaks in the next time period within a single region. Prior work has not explicitly modeled the spatial aspects of disease spread, which for vector borne diseases can be a significant factor. One reason is that data for application of spatial models is not readily available. We propose to add the spatial dimension to the disease modeling problem within a Graph Neural Network framework, with the graph being defined by the spatial connectivity of the region. That is, our work frames the traditional singular time series with proxy variables forecasting task as a multiple correlated time series graph

problem. With this change in problem definition, one can expect to see changes in the most suitable models.

**Contributions.** The contributions of this study are summarized as follows.

- We gathered meteorological, epidemiological, and socio-economic data over multiple spatial regions (GIS-polygons with a spatial resolution of approximately 500 square miles per polygon). Hence, we develop (and make available to the research community) a novel disease data set that has aligned meteorological, epidemiological, socio-economic and spatial features.
- We develop a robust EWS for disease that we apply to dengue in Sri Lanka. The model can be applied to any disease setting where the data is available.
- We compare traditional machine learning models to state-of-the-art graph network models that monitor human mobility between Sri Lankan counties.
- One conclusion from our results is that models which incorporate spatial correlations between the polygonal regions (as with our Graph Neural Network models) outperform models which predict independently on each GIS-polygon. That is, it is important to use meteorological, epidemiological, and socio-economic data in one GIS-polygon to influence the prediction in neighboring polygons. This spatial graph structure can easily be accommodated within a Graph Neural Network model, which motivates our choice of model.

This paper will flow step-by-step in the process of developing these models. We will begin by discussing research relating to statistical and machine learning models used in EWS, and our proposed graph network models. Next, we will mathematically define the EWS problem statement. We then discuss the methodology used to create our organic data, what data we collected, and where the data was gathered from. Analysis on our collected data will be discussed. Afterward, each of our ML model's architectures will be explained alongside our performance evaluation criteria. Finally, we will show our model's results and discuss what they suggest with regards to the future of EWS for dengue and disease forecasting.

## II. RELATED WORKS

**Traditional EWS.** Traditional models used for time series forecasting such as Autoregressive Integrated Moving Average (ARIMA) (9; 10; 11; 12; 13), Support Vector Regression (SVR) (14; 15), Random Forest (16), Extreme Gradient Boosting (XGBoost), LASSO Regression (14), Artificial Neural Networks (ANN) (15; 17), Auto-Regressive Neural Networks (ARNN) (18), and Long Short-Term Memory (LSTM) (16; 19) have been widely applied across both disease forecasting and similar problems. They are used both standalone as explored in (20) and (21), and as hybrid methods. In particular, (22) uses a joint Attention and LSTM method for forecasting traffic speeds, (23) uses a joint ARIMA and LSTM model, (24) uses LASSO Regression, (25) uses ARIMA, (19) uses SVM and LSTMs, (16) uses Random Forests and LSTM, (26) uses a Poisson multivariate regression model, and (27) uses

LSTMs. ARIMA in particular is the most commonly used model for dengue forecasting problems as it allows lagged meteorological attributes to serve as covariates to historical dengue data (28).

**Traditional EWS Limitations.** These traditional approaches are often used with respect to data that inadequately account for spatial relations that are inherently present in disease forecasting problems. Another common issue is the lack of high resolution disease data, which results in many papers adapting the forecasting problem to a more trivial outbreak classification task rather than a regression task. (29) uses raster data that accounts for the temporal nature of the data, but only in a single location, and is framed as a classification task. (30) gathers data from 5 districts of Malaysia, but independently creates an outbreak variable for each district without accounting for inter-district causality and is also a classification task. Likewise, (28) gathers dengue data across Colombia over a 4 year period but pools all dengue cases in their country-wide model and evaluates each district individually in their district models.

**Graph Representation Learning.** In recent years, an emerging direction for multiple-correlated time series forecasting solutions are the applications of Graph Neural Networks (GNNs). GNNs and their derivatives have been remarkably successful for problems similar in nature to disease forecasting such as traffic speed forecasting (31), water demand forecasting (32), and west Nile virus forecasting (33), where spatial attributes are a huge factor of future trends alongside temporal attributes.

Some novel graph-based approaches include Graph Attention Networks (GAT) (34), Diffusion Convolutional Recurrent Neural Networks (DCRNN) (35), Diffusion Convolutional Networks (DCNN) (36), Gated Attention Networks (GAAN) (37), Graph Embeddings (8), Spatio-Temporal Graph Neural Networks (STGNN) (38), Spatio-Temporal Graph Attention Networks (STGAT) (39), Attention-Based Spatio-Temporal Graph Convolution Networks (ASTGCN) (40), Two-Stream Adaptive Graph Convolutional Networks (AAGCN) (41), and Temporal Graph Convolutional Networks (A3TGCN) (42). Enhancements to these GNNs include optimizations to learning spatial convolutions (43), better adaptive paths (44), and edge featured graphs (45).

We believe that the graphical and temporal structure of dengue data will be better captured through graph representation learning and explore the effectiveness of applying these spatio-temporal GNNs as an EWS in disease forecasting in comparison to traditional models.

### A. Problem Statement

**Correlated Time-Series Data.** Consider a problem where we are provided with spatio-temporal data across a country with  $N$  regions. The state of the  $i$ -th region can be represented as a time series

$$TS^{(i)} = \langle x_1^{(i)}, x_2^{(i)}, \dots, x_t^{(i)} \rangle$$

across  $t$  time-steps, where  $x_j^{(i)}$  is a  $K$ -dimensional vector representing  $K$  features (e.g. Disease Cases, Precipitation) at time  $j$ . The relationship between regions is represented as an adjacency matrix  $A \in \mathbb{R}^{N \times N}$ .

The goal represented as a regression task is to predict the future status of disease cases across all  $N$  regions given the historical statuses of all regions. Formally, we use a window of historical data covering  $[t_{j+1}, t_{j+l}]$  consisting of  $l$  time-steps to predict the future disease cases across the window  $[t_{j+l+1}, t_{j+l+p}]$  with  $p$  time-steps, that is,  $p$ -step ahead forecasting.

### B. Models

We considered standard time-series models as well as spatio-temporal graph neural network models.

### C. Baseline Time Series Models

We selected five Baseline EWS Models to benchmark our GNNs against: ARIMAX, Random Forest, XGBoost, Auto-Regressive Neural Networks (ARNN), and LSTMs, which are commonly used for disease forecasting (46; 47; 48; 49). The first model we explored was ARIMAX, which is then applied individually to each of the 25 districts to account for regional variability. We determined the optimal parameter configurations as ARIMA(1, 2, 3) through iterative testing.

Our findings indicate that ARIMAX consistently outperformed SARIMAX, the seasonal variation of ARIMAX. This superiority is likely due to Sri Lanka's equatorial climate, which does not exhibit extreme seasonal changes. The second model we attempted was LSTM. We configured our model with a single hidden layer consisting of 50 units and with a window size of three. The model was then trained on normalized features using MinMaxScaler on the first 70% of the data, with the remaining 30% reserved for testing. The last baseline model we attempted was Random Forest. Our model was instantiated with 100 decision trees and trained on the 70% of the data as usual.

The result of MAE and RMSE are displayed in (table) where the ARIMAX approach outperforms LSTM and Random Forest. We will delve into more sophisticated GNNs to further improve our dengue forecasting capabilities.

### D. Spatio-Temporal Graph Neural Networks

**Model Selection.** We selected five spatio-temporal Graph Neural Networks including ASTGCN (2019) (40), A3TGCN (2020) (42), STGAT (2019) (39), DCRNN (2018) (35), and AAGCN (2019) (41). While these models are slightly dated, it is important to note that they still achieve near state-of-the-art results on either the METR-LA (35) traffic forecasting dataset, the PEMS-BAY traffic forecasting dataset, or in their own domains, i.e. AAGCN in action forecasting, while being generic enough to adapt them for our disease forecasting dataset. Due to the similar structure of the two datasets, we believe that successful traffic speed prediction models can achieve similar results on disease data.

**Training Details.** We train Random Forest and XGBoost out-of-the-box from scikit-learn. ARIMA is trained using

statsmodel, and ARNN is trained out-of-the-box using Neural Prophet (50). LSTM was trained with a hidden dimension of 64 and a dense layer. We implement ASTGCN, A3TGCN, STGAT, DCRNN, and AAGCN using PyTorch Geometric Temporal (51). Our GNNs are trained on an AMD Ryzen 5 PRO 4650U CPU using the Adam optimizer with a learning rate of 0.00001 and a weight decay of 0.000001. ASTGCN is trained with 2 blocks and a filter of 64, DCRNN, A3TGCN and STGAT are trained with a hidden dimension of 64. AAGCN is trained out-of-the-box from PyTorch Geometric. We use a batch size of 1 to prevent over-fitting, a window of 3 past time-steps, and predict a window of 3 future time-steps. Our graphs plot the prediction from the first time-step of the 3 predicted time-steps in each window. We use the standard 70/30 split for train and test data.

## III. DATA COLLECTION

**Data Source Aggregation.** We benchmark our models on a disease dataset from (52) consisting of weekly disease cases measured weekly across each of Sri Lanka's 25 districts. We also incorporate satellite data consisting of precipitation, humidity, land surface temperature, temperature, and land canopy cover from NASA EarthData's meteorological data. Using a district demarcated shapefile of Sri Lanka from the GADM database of Global Administrative Areas (53), we spatially join the two datasets based on latitude and longitude, then temporally aggregate the data to match the weekly resolution of the disease data. We save the mean, minimum, and maximum values of the satellite data over the aggregation period as features. Due to a lack of travel data between districts, our adjacency matrix is organically built independently by each member of the team based on district distances and cross-validated to produce a more accurate matrix.

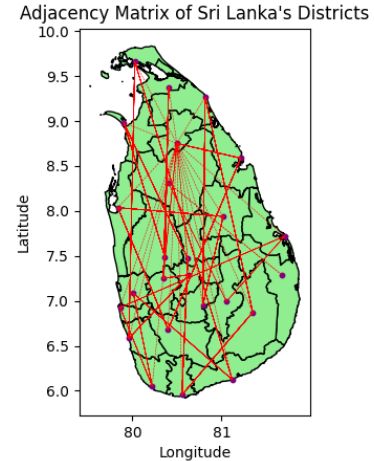


Fig. 1. Our organic adjacency matrix plotted across a GIS polygon of Sri Lanka's districts.

## IV. DATA ANALYSIS

Before the ML models were trained, exploratory data analysis was performed on the weather features and dengue case

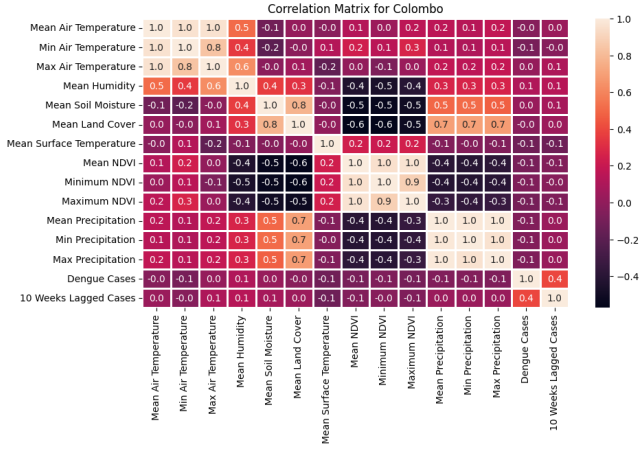


Fig. 2. Correlation matrix of meteorological features with dengue cases 10 weeks in the future. The data used is from Colombo.

data. The weather data was abstracted to weekly minimums, maximums, and averages for each variable available. The weather variables included rainfall, temperature, humidity, NDVI, and other metrics related to dengue and mosquito populations. Additionally, some social variables were added, including district population.

Correlation between dengue cases and the features were taken. Additionally, correlation when cases are 10+ weeks out was taken. This is because long-term temperature and rainfall changes have been shown to affect dengue (54). To see if this was the same in Sri Lanka, the correlation for cases 10 weeks in the future was included.

The dengue case to weather correlation increases when out 10 weeks, which is an indicator that the features should be delayed by several weeks to improve prediction. The number of weeks to delay each feature was analyzed by checking the correlation with dengue 1 to 25 weeks into the future. The rainfall features were found to have the biggest change in correlation. All precipitation was shifted back 12 weeks. Minimum NDVI was shifted 17 weeks back and Mean canopy surface water was shifted back 12 weeks.

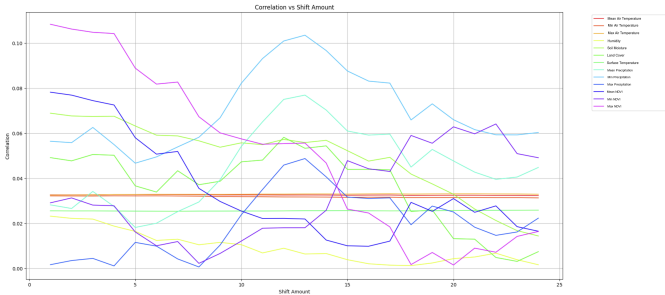


Fig. 3. Correlation of variables with all dengue cases across all districts. The absolute value of the correlation was plotted.

## V. EXPERIMENTS

### A. Dataset

We conduct experiments on our Sri Lanka Dengue dataset from (52), which consists of Sri Lanka's 25 districts, each with weekly precipitation, land surface temperature, air temperature, vegetation, humidity, and land cover data aggregated across hourly, three hourly, and daily time resolutions between 2013 and 2022. We additionally average measurements taken across various points within a district due to the lack of high resolution disease case data. Our goal is to predict the disease cases count in all 25 nodes across the next 3 time-steps given the past 3 time-steps.

Additionally, we repeat the procedure above using a time-shifted dataset based on peak feature importance metrics.

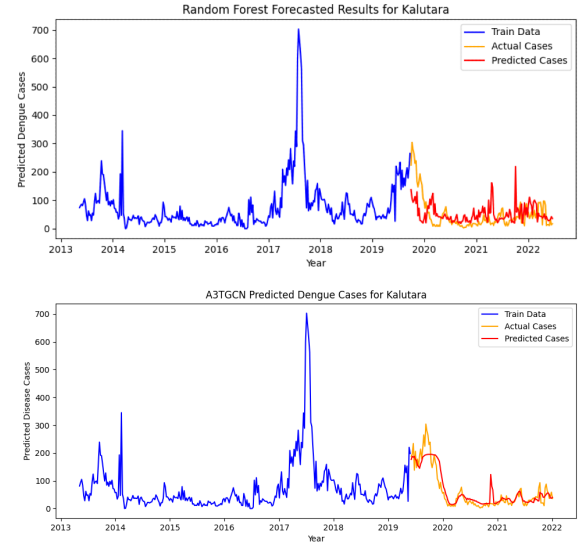


Fig. 4. Forecasting results of the best performing baseline and GNN in Kalutara. One concludes that using the spatio-temporal approach produces impressive predictions of future dengue cases.

## VI. STATISTICAL SIGNIFICANCE

We evaluate our GNNs using Time-Series (Rolling) Cross Validation as an alternative to K-Fold Cross Validation to avoid K-Fold Cross Validation's forward-looking problem with time-series data. Our Rolling Cross Validation is rolled across 5 segments consisting of 60%, 70%, 80%, 90%, and 100% of the data. Every GNN is reinitialized at each segment, and each segment is split into train and test sets following a 70/30 split. Since rolling cross validation does not incorporate the entire dataset in each of its splits, we also measure the performance of the models on the full dataset.

## VII. RESULTS

**Baselines and Metrics.** We evaluate our models using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), where a lower error is indicative of better performance. From the table, we see that GNNs consistently outperform the baseline in this dataset in both cross validation and no

TABLE I  
MODEL ERROR COMPARISON WITH THE BEST RESULTS IN BOLD.

	Cross Validated		Full Dataset	
Model	MAE	RMSE	MAE	RMSE
<b>Shifted Dataset</b>				
ARIMA	168.37 $\pm$ 187.86	189.22 $\pm$ 181.52	133.41	142.22
Random Forest	48.91 $\pm$ 11.95	84.66 $\pm$ 19.22	34.52	64.42
XGBoost	53.89 $\pm$ 10.76	95.22 $\pm$ 17.22	40.77	75.44
ARNN	82.46 $\pm$ 16.67	106.83 $\pm$ 26.14	56.38	70.16
LSTM	81.19 $\pm$ 17.95	131.36 $\pm$ 30.81	49.19	81.88
DCRNN	45.98 $\pm$ 2.34	71.61 $\pm$ 2.56	26.20	36.30
STGAT	<b>25.38 <math>\pm</math> 1.37</b>	<b>44.78 <math>\pm</math> 2.26</b>	22.28	42.65
AAGCN	41.93 $\pm$ 1.11	55.83 $\pm$ 1.98	39.27	49.20
A3TGCN	34.03 $\pm$ 1.10	58.52 $\pm$ 1.87	<b>18.65</b>	<b>30.55</b>
ASTGCN	33.68 $\pm$ 19.85	47.72 $\pm$ 18.12	20.62	32.41
<b>Unshifted Dataset</b>				
ARIMA	105.11 $\pm$ 86.63	133.35 $\pm$ 82.70	275.70	284.18
Random Forest	65.58 $\pm$ 12.44	108.32 $\pm$ 29.04	49.44	67.02
XGBoost	86.91 $\pm$ 39.76	114.06 $\pm$ 43.78	161.29	169.34
ARNN	111.87 $\pm$ 63.15	159.28 $\pm$ 51.62	123.72	134.49
LSTM	68.95 $\pm$ 25.91	92.57 $\pm$ 48.35	57.93	69.02
DCRNN	45.72 $\pm$ 2.78	70.82 $\pm$ 2.85	43.54	65.32
STGAT	<b>25.20 <math>\pm</math> 1.04</b>	<b>44.35 <math>\pm</math> 2.04</b>	<b>24.68</b>	42.85
AAGCN	40.44 $\pm$ 0.69	53.90 $\pm$ 1.07	40.39	52.90
A3TGCN	33.53 $\pm$ 1.23	58.13 $\pm$ 2.15	31.28	54.04
ASTGCN	33.08 $\pm$ 15.58	46.54 $\pm$ 13.70	27.52	<b>39.99</b>

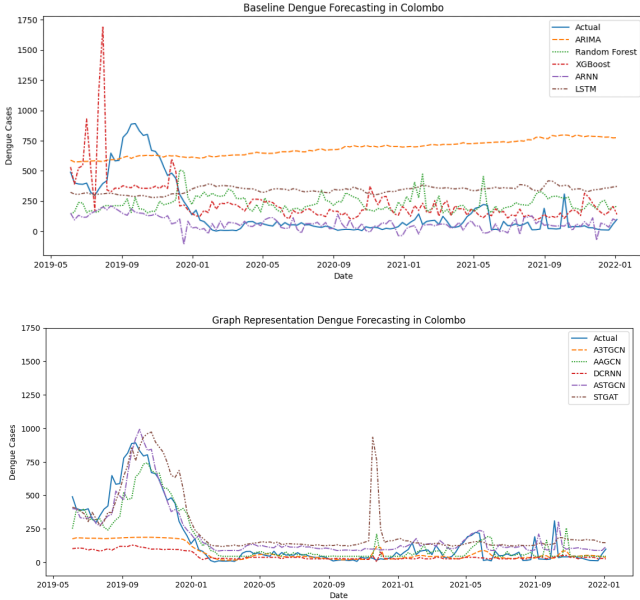


Fig. 5. Plots of the baseline model predictions and the graph model predictions in the top and bottom plots respectively. Predictions are for the test data of Colombo.

validation on both the shifted and unshifted datasets, with STGAT performing the best in cross validation and A3TGCN performing the best without. ARIMA receives a significant boost in performance after the shifting of data, which is expected given that ARIMA excels at capturing auto-correlation,

which is boosted by lagging the data. Notably, the shifting of the data on average improves the full dataset prediction results and keeps the cross validated results the same, with ARIMA and LSTM experiencing an improvement in Cross Validated results but an increase in full dataset errors.

**Analysis** The results show that on average, the baselines achieve an average RMSE of 70.65 on the shifted dataset and 144.81 on the unshifted dataset. In comparison, spatio-temporal GNNs achieve an average RMSE of 38.22 on the shifted dataset and 50.62 on the unshifted dataset. This means that on average, GNNs were 32 and 94 cases closer to the actual dengue cases than the baselines on the shifted and unshifted datasets respectively. One can observe that there is a consistent improvement from using spatio-temporal GNNs, supporting the main idea that GNNs that incorporate both spatial and temporal relations are better suited for EWS.

**Modeling Challenges.** At some predicted time-steps, the predictions in one district is greatly influenced by the cases of neighboring districts, even if the disease was not spread in actuality, causing abnormal spikes in some of the test predictions. This is one inherent flaw of using GNNs. In an ablation study, the models performed better without surface temperature data, and some models such as DCRNN performed better when using only the disease case data, which we attribute to DCRNN being designed for METR-LA, which is a single-featured dataset, while our dataset has 20 features at each time-step.



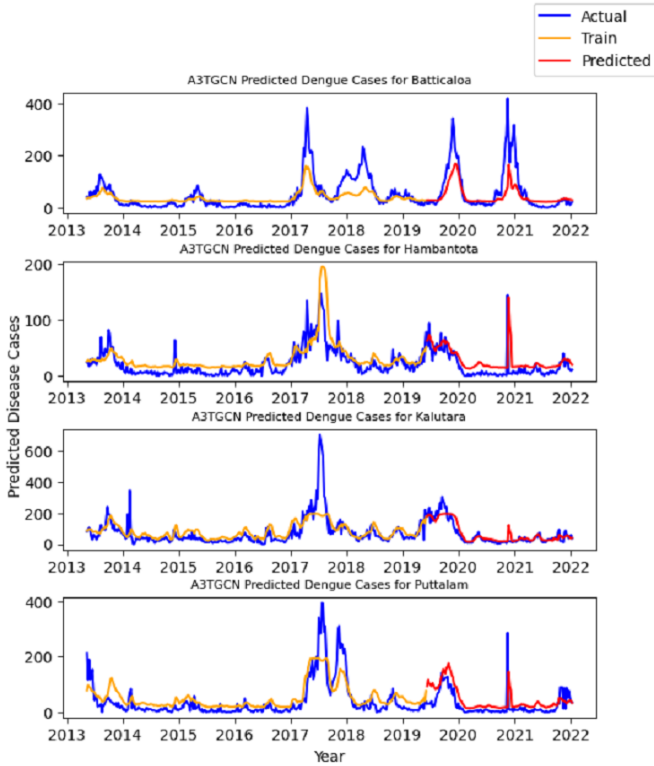


Fig. 6. Focusing on the spatio-temporal model A3TGCN, we show prediction results for four districts in Sri Lanka. Each district is one GIS-polygon in our model. One observes that the EWS model gives impressive prediction curves.

### VIII. LIMITATIONS

**Locality.** For our research, we consulted with a Sri-Lanka based epidemiologist to make the assumption that the average time-frame between Dengue infection and diagnosis is a 3 week period, hence we use a window size of 3 and 3-step ahead forecasting. We additionally measure feature importance across multiple time-steps to determine the optimal shift for each feature. We also conduct an ablation study to determine which features should be included in each model. However, these shifts and feature importances will vary across regions, and transfer learning on these problems may not be able to capture these changes.

**Data Resolution.** Another limitation is the scarcity of high resolution disease case data, both temporally and spatially. Due to disease case data being tallied weekly in Sri Lanka, the temporal resolution of our dataset is limited to weekly aggregated features. Consequently, we only have slightly over 6000 time-steps in our Sri Lanka dataset, and some models may have trouble learning from such a small dataset. Additionally, disease case data is only reported per district, and as a result, spatio-temporal features needed to be aggregated geographically as well. Consequently, our graph contains only 25 nodes while METR-LA, for example, contains 207, which can negatively impact GNNs that rely on geospatial features or graph convolutions. Looking forwards, we emphasize the importance of collecting high resolution, secure, and accurate

data that will bolster the training of GNNs.

**Data Availability.** As an organically produced dataset of its own kind, we were unable to find similar spatio-temporal datasets to benchmark the robustness of our results on other datasets. This is an aspect that should be explored in future research, but is a clear limitation that cannot be addressed at this time. Supposing that there was such data, we believe that the major differences between that data and our data would be in the graph structure, seasonality, and environmental features, all of which can be captured by spatiotemporal GNNs. Hence, we believe that our work effectively demonstrates the possibilities that spatio-temporal GNNs bring to EWS.

Another limitation with data availability is that dengue incidence rates are influenced by a wide range of environmental factors, some of which are rarely measured. The economic landscape and poverty level of the region, for example, is a big influence that we were unable to find a suitable data source for. We also had to use a trivially built adjacency matrix instead of a highly detailed and accurate representation of the relationships between Sri Lanka's districts due to a lack of transportation data between districts. We believe that with the incorporation of such economic data, geographical relationships, and more time-steps, GNNs will be able to achieve state-of-the-art results as a robust EWS for dengue.

**Compute** While GNNs have been shown to perform better within the problem definition of the paper, it is important to note that GNNs are much more computationally intensive compared to traditional methods. This may be a limitation when there is extremely high resolution spatiotemporal data, i.e. hourly cases within the United States.

### IX. CONCLUSION

In this paper, we approached the development of an EWS using both traditional machine learning models and innovative graph representation techniques. Our paper seeks not to introduce novel methods for disease forecasting, but evaluates the effectiveness of a novel representation of dengue forecasting as a correlated time series problem that enables us to leverage spatio-temporal graph representation learning used to achieve state of the art results in problems of similar nature such as traffic forecasting. We demonstrate that spatio-temporal Graph Neural Networks consistently outperformed their traditional counterparts as measured by MAE and RMSE. We also show that shifting features based on feature importance can have a nontrivial effect on forecasting results. The best-performing model was A3TGCN with an MAE of 18.65 and RMSE of 30.55 on our shifted dataset. Looking forward, we plan to apply this approach to analyze dengue data from Brazil, using the same dataset as referenced in our earlier work (52), with a more in-depth ablation study. The preprocessing and training code developed during this project has been designed for easy adaptation to other similar datasets on dengue, which could facilitate broader applications and enhance predictive accuracy in future implementations.

# REFERENCES

- [1] J. B. D. W. R. R. J. S. M. D. T. Y. S. P. C. L. P. G. A. R. C. S. A. M. d. . F. S. D. Tissera, H. A., "Severe dengue epidemic, sri lanka, 2017," *Emerging Infectious Diseases journal*, vol. 26, no. 1, 2020.
- [2] G. N. Malavige, C. Jeewandara, A. Ghouse, G. Somathilake, and H. Tissera, "Changing epidemiology of dengue in Sri Lanka—Challenges for the future," *PLOS Neglected Tropical Diseases*, vol. 15, p. e0009624, Aug. 2021. Publisher: Public Library of Science.
- [3] L. Hussain-Alkhateeb, T. Rivera Ramírez, A. Kroeger, E. Gozzer, and S. Runge-Ranzinger, "Early warning systems (EWSs) for chikungunya, dengue, malaria, yellow fever, and Zika outbreaks: What is the evidence? A scoping review," *PLoS neglected tropical diseases*, vol. 15, p. e0009686, Sept. 2021.
- [4] P. Sirisena, F. Noordeen, H. Kurukulasuriya, T. A. Romesh, and L. Fernando, "Effect of Climatic Factors and Population Density on the Distribution of Dengue in Sri Lanka: A GIS Based Evaluation for Prediction of Outbreaks," *PLOS ONE*, vol. 12, p. e0166806, Jan. 2017. Publisher: Public Library of Science.
- [5] Z. Li and J. Dong, "Big Geospatial Data and Data-Driven Methods for Urban Dengue Risk Forecasting: A Review," *Remote Sensing*, vol. 14, p. 5052, Jan. 2022. Number: 19 Publisher: Multidisciplinary Digital Publishing Institute.
- [6] M. H. F. M. I. D. S. P. R. F. R. P. B. L. L. L. D. M. S. O. L. S. D. A. R. W. T. C. Maria da Consolação Magalhaes ~ Cunha, Yang Ju, "Disentangling associations between vegetation greenness and dengue in a latin american city: Findings and challenges," *Landscape and Urban Planning*, vol. 216, no. 1, 2021.
- [7] B. Adams and D. D. Kapan, "Man Bites Mosquito: Understanding the Contribution of Human Movement to Vector-Borne Disease Dynamics," *PLOS ONE*, vol. 4, p. e6763, Aug. 2009. Publisher: Public Library of Science.
- [8] K. Liu, M. Zhang, G. Xi, A. Deng, T. Song, Q. Li, M. Kang, and L. Yin, "Enhancing fine-grained intra-urban dengue forecasting by integrating spatial interactions of human movements between urban regions," *PLOS Neglected Tropical Diseases*, vol. 14, p. e0008924, Dec. 2020. Publisher: Public Library of Science.
- [9] N. C. Dom, A. A. Hassan, Z. A. Latif, and R. Ismail, "Generating temporal model using climate variables for the prediction of dengue cases in subang jaya, malaysia," *Asian Pac. J. Trop. Dis.*, vol. 3, pp. 352–361, Oct. 2013.
- [10] F. Cortes, C. M. Turchi Martelli, R. Arraes de Alencar Ximenes, U. R. Montarroyos, J. B. Siqueira Junior, O. Gonçalves Cruz, N. Alexander, and W. Vieira de Souza, "Time series analysis of dengue surveillance data in two brazilian cities," *Acta Trop.*, vol. 182, pp. 190–197, June 2018.
- [11] M. A. Johansson, N. G. Reich, A. Hota, J. S. Brownstein, and M. Santillana, "Evaluating the performance of infectious disease forecasts: A comparison of climate-driven and seasonal dengue forecasts for mexico," *Sci. Rep.*, vol. 6, p. 33707, Sept. 2016.
- [12] S. Naish, P. Dale, J. S. Mackenzie, J. McBride, K. Mengersen, and S. Tong, "Climate change and dengue: a critical and systematic review of quantitative modelling approaches," *BMC Infect. Dis.*, vol. 14, p. 167, Mar. 2014.
- [13] M. Gharbi, P. Quenel, J. Gustave, S. Cassadou, G. La Ruche, L. Girdary, and L. Marrama, "Time series analysis of dengue incidence in guadeloupe, french west indies: forecasting models using climate variables as predictors," *BMC Infect. Dis.*, vol. 11, p. 166, June 2011.
- [14] P. Guo, T. Liu, Q. Zhang, L. Wang, J. Xiao, Q. Zhang, G. Luo, Z. Li, J. He, Y. Zhang, and W. Ma, "Developing a dengue forecast model using machine learning: A case study in china," *PLoS Negl. Trop. Dis.*, vol. 11, p. e0005973, Oct. 2017.
- [15] J. M. Scavuzzo, F. Trucco, M. Espinosa, C. B. Tauro, M. Abril, C. M. Scavuzzo, and A. C. Frery, "Modeling dengue vector population using remotely sensed data and machine learning," *Acta Trop.*, vol. 185, pp. 167–175, Sept. 2018.
- [16] E. Mussumeci and F. Codeço Coelho, "Large-scale multivariate forecasting models for Dengue - LSTM versus random forest regression," *Spatial and Spatio-temporal Epidemiology*, vol. 35, p. 100372, Nov. 2020.
- [17] A. E. Laureano-Rosario, A. P. Duncan, P. A. Mendez-Lazaro, J. E. Garcia-Rejon, S. Gomez-Carro, J. Farfan-Ale, D. A. Savic, and F. E. Muller-Karger, "Application of artificial neural networks for dengue fever outbreak predictions in the northwest coast of yucatan, mexico and san juan, puerto rico," *Trop. Med. Infect. Dis.*, vol. 3, p. 5, Jan. 2018.
- [18] O. Triebe, N. Laptev, and R. Rajagopal, "AR-Net: A simple Auto-Regressive neural network for time-series," Nov. 2019.
- [19] A. Sebastianelli, D. Spiller, R. Carmo, J. Wheeler, A. Nowakowski, L. V. Jacobson, D. Kim, H. Barlevi, Z. E. R. Cordero, F. J. Colón-González, R. Lowe, S. L. Ullo, and R. Schneider, "A reproducible ensemble machine learning approach to forecast dengue outbreaks," *Scientific Reports*, vol. 14, p. 3807, Feb. 2024. Number: 1 Publisher: Nature Publishing Group.
- [20] X. Zhang, T. Zhang, A. A. Young, and X. Li, "Applications and Comparisons of Four Time Series Models in Epidemiological Surveillance Data," *PLOS ONE*, vol. 9, p. e88075, Feb. 2014. Publisher: Public Library of Science.
- [21] P. Guo, T. Liu, Q. Zhang, L. Wang, J. Xiao, Q. Zhang, G. Luo, Z. Li, J. He, Y. Zhang, and W. Ma, "Developing a dengue forecast model using machine learning: A case study in China," *PLOS Neglected Tropical Diseases*, vol. 11, p. e0005973, Oct. 2017. Publisher: Public Library of Science.
- [22] P. Wu, Z. Huang, Y. Pian, L. Xu, J. Li, and K. Chen, "A

- Combined Deep Learning Method with Attention-Based LSTM Model for Short-Term Traffic Speed Forecasting,” *Journal of Advanced Transportation*, vol. 2020, pp. 1–15, Nov. 2020.
- [23] S. Lu, Q. Zhang, G. Chen, and D. Seng, “A combined method for short-term traffic flow prediction based on recurrent neural network,” *Alexandria Engineering Journal*, vol. 60, July 2020.
- [24] Y. Shi, X. Liu, S.-Y. Kok, J. Rajarethinam, S. Liang, G. Yap, C.-S. Chong, K.-S. Lee, S. S. Tan, C. K. Y. Chin, A. Lo, W. Kong, L. C. Ng, and A. R. Cook, “Three-Month Real-Time Dengue Forecast Models: An Early Warning System for Outbreak Alerts and Policy Decision Support in Singapore,” *Environmental Health Perspectives*, vol. 124, pp. 1369–1375, Sept. 2016. Publisher: Environmental Health Perspectives.
- [25] P.-C. Wu, H.-R. Guo, S.-C. Lung, C.-Y. Lin, and H.-J. Su, “Weather as an effective predictor for occurrence of dengue fever in Taiwan,” *Acta Tropica*, vol. 103, pp. 50–57, July 2007.
- [26] S. Sang, S. Gu, P. Bi, W. Yang, Z. Yang, L. Xu, J. Yang, X. Liu, T. Jiang, H. Wu, C. Chu, and Q. Liu, “Predicting Unprecedented Dengue Outbreak Using Imported Cases and Climatic Factors in Guangzhou, 2014,” *PLOS Neglected Tropical Diseases*, vol. 9, p. e0003808, May 2015. Publisher: Public Library of Science.
- [27] N. Laptev, J. Yosinski, L. E. Li, and S. Smyl, “Time-series Extreme Event Forecasting with Neural Networks at Uber,”
- [28] N. Zhao, K. Charland, M. Carabali, E. O. Nsoesie, M. Maheu-Giroux, E. Rees, M. Yuan, C. Garcia Balaguera, G. Jaramillo Ramirez, and K. Zinszer, “Machine learning and dengue forecasting: Comparing random forests and artificial neural networks for predicting dengue burden at national and sub-national scales in colombia,” *PLoS Negl. Trop. Dis.*, vol. 14, p. e0008056, Sept. 2020.
- [29] S. Vincenzi, A. Porrello, P. Buzzega, A. Conte, C. Ippoliti, L. Candeloro, A. Di Lorenzo, A. C. Dondona, and S. Calderara, “Spotting insects from satellites: modeling the presence of culicoides imicola through deep CNNs,” 2019.
- [30] N. A. M. Salim, Y. B. Wah, C. Reeves, M. Smith, W. F. W. Yaacob, R. N. Mudin, R. Dapari, N. N. F. F. Sapri, and U. Haque, “Prediction of dengue outbreak in selangor malaysia using machine learning techniques,” *Sci. Rep.*, vol. 11, p. 939, Jan. 2021.
- [31] R.-G. Cirstea, C. Guo, and B. Yang, “Graph Attention Recurrent Neural Networks for Correlated Time Series Forecasting – Full version,” Mar. 2021.
- [32] A. Zanfei, B. M. Brentan, A. Menapace, M. Righetti, and M. Herrera, “Graph Convolutional Recurrent Neural Networks for Water Demand Forecasting,” *Water Resources Research*, vol. 58, no. 7, p. e2022WR032299, 2022. [\\_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2022WR032299](https://onlinelibrary.wiley.com/doi/pdf/10.1029/2022WR032299).
- [33] L. Bonicelli, A. Porrello, S. Vincenzi, C. Ippoliti, F. Iapalo, A. Conte, and S. Calderara, “Spotting Virus From Satellites: Modeling the Circulation of West Nile Virus Through Graph Neural Networks,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–12, 2023.
- [34] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, “Graph Attention Networks,” Oct. 2017.
- [35] Y. Li, R. Yu, C. Shahabi, and Y. Liu, “Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting,” Feb. 2018.
- [36] J. Atwood and D. Towsley, “Diffusion-Convolutional Neural Networks,” in *Advances in Neural Information Processing Systems*, vol. 29, Curran Associates, Inc., 2016.
- [37] J. Zhang, X. Shi, J. Xie, H. Ma, I. King, and D.-Y. Yeung, “GaAN: Gated Attention Networks for Learning on Large and Spatiotemporal Graphs,” Mar. 2018. arXiv:1803.07294 [cs].
- [38] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, “A Comprehensive Survey on Graph Neural Networks,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, pp. 4–24, Jan. 2021. Conference Name: IEEE Transactions on Neural Networks and Learning Systems.
- [39] Y. Huang, H. Bi, Z. Li, T. Mao, and Z. Wang, “Stgat: Modeling spatial-temporal interactions for human trajectory prediction,” in *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.
- [40] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, “Attention based spatial-temporal graph convolutional networks for traffic flow forecasting,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 922–929, 2019.
- [41] L. Shi, Y. Zhang, J. Cheng, and H. Lu, “Two-stream adaptive graph convolutional networks for skeleton-based action recognition,” in *CVPR*, 2019.
- [42] J. Zhu, Y. Song, L. Zhao, and H. Li, “A3t-gcn: Attention temporal graph convolutional network for traffic forecasting,” 2020.
- [43] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, “Spectral Networks and Locally Connected Networks on Graphs,” May 2014. arXiv:1312.6203 [cs].
- [44] Z. Liu, C. Chen, L. Li, J. Zhou, X. Li, L. Song, and Y. Qi, “GeniePath: Graph Neural Networks with Adaptive Receptive Paths,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 4424–4431, July 2019. Number: 01.
- [45] J. Chen and H. Chen, “Edge-Featured Graph Attention Network,” Jan. 2021.
- [46] J. Herzen, F. Lässig, S. G. Piazzetta, T. Neuer, L. Tafti, G. Raille, T. Van Pottelbergh, M. Pasička, A. Skrodzki, N. Huguenin, M. Dumonal, J. Kościsz, D. Bader, F. Gusset, M. Benheddi, C. Williamson, M. Kosinski, M. Petrik, and G. Grosch, “Darts: User-friendly modern machine learning for time series,” Oct. 2021.



- [47] I. Ghosh, S. Gupta, and S. Rana, “Anticipating dengue outbreaks using a novel hybrid arima-arnn model with exogenous variables,” 2023.
- [48] J. Rob and G. Hyndman, *Forecasting: principles and practice*. OTexts, 2018.
- [49] C. E. Tedford, J. G. Phillips, R. Gregory, G. P. Pawlowski, L. Fadnis, M. A. Khan, S. M. Ali, M. K. Handley, and S. L. Yates, “Development of trans-2-[1h-imidazol-4-yl] cyclopropane derivatives as new high-affinity histamine H3 receptor ligands,” *J. Pharmacol. Exp. Ther.*, vol. 289, pp. 1160–1168, May 1999.
- [50] O. Triebe, H. Hewamalage, P. Pilyugina, N. Laptev, C. Bergmeir, and R. Rajagopal, “Neuralprophet: Explainable forecasting at scale,” 2021.
- [51] B. Rozemberczki, P. Scherer, Y. He, G. Panagopoulos, A. Riedel, M. Astefanoaei, O. Kiss, F. Beres, G. Lopez, N. Collignon, and R. Sarkar, “PyTorch Geometric Temporal: Spatiotemporal Signal Processing with Neural Machine Learning Models,” in *Proceedings of the 30th ACM International Conference on Information and Knowledge Management*, p. 4564–4573, 2021.
- [52] Cruz, Kim, Pasarkar, O’Gara, Qiu, Weng, and Zhang, “Extracting disease data using natural language processing,” 2024.
- [53] GADM, “Database of global administrative boundaries (gadm),” 2018.
- [54] T.-W. Chuang, L. F. Chaves, and P.-J. Chen, “Effects of local and regional climatic fluctuations on dengue outbreaks in southern taiwan,” *PLOS ONE*, vol. 12, pp. 1–20, 06 2017.