Annals of Pure and Applied Mathematics
Vol. 29, No. 2, 2024, 119-132
ISSN: 2279-087X (P), 2279-0888(online)
Published on 27 June 2024
www.researchmathsci.org
DOI: http://dx.doi.org/10.22457/apam.v29n2a04938

Annals of Pure and Applied Mathematics

Innovating Graph Representation for Dynamic Weather Forecasting

***Harish Chandra Bhandari*[1,*], *Yagya Raj Pandeya*[2] and *Kanhaiya Jha*[1]**

[1]Department of Mathematics, School of Science, Kathmandu University, Nepal
E-mail: harish.bhandari@ku.edu.np, jhakn@ku.edu.np
[2]Department of Computer Science and Engineering, School of Engineering
Kathmandu University, Nepal
E-mail: yagya.pandeya@ku.edu.np
[*]Corresponding Author: harish.bhandari@ku.edu.np

Received 3 June 2024; accepted 25 June 2024

Abstract. Extreme weather conditions pose significant global challenges, making precise and reliable forecasting methods crucial. Current graph neural network (GNN) models, although successful in implementations, often struggle with accurately capturing the complexities of geographical landscapes, particularly in diverse regions like Nepal. The objective of this paper is to address these limitations by introducing a novel approach to graph representation for weather forecasting. Our method involves developing a domain-guided knowledge graph specifically tailored for large, geographically diverse regions. By employing spatio-temporal graph neural networks, we forecast multiple weather attributes over extended periods, effectively leveraging spatial dependencies within the constructed graph. This approach demonstrates remarkable improvements in forecasting accuracy, robustness under computing resource constraints, and scalability. Also, the analysis confirms the coherence between forecasted results and graph structures, providing insights into the reliability and predictive power of our method.

Keywords: Graph representation, dynamic weather attribute, spatio-temporal graph, weather forecasting

AMS Mathematics Subject Classification (2010): 05C62, 70H45

1. Introduction

Extreme weather conditions can inflict substantial harm on individuals, communities, nations, and the global community. The EMDAT database documents [1] a total of 11,360 natural disasters occurred between 1995 and 2022, averaging 398 per year. Notably, hydrological, meteorological, and climatological disasters constitute 81.2% of these incidents. In Nepal, natural disasters caused by extreme meteorological events, such as floods, erratic rainfall, landslides, droughts, thunderstorms, hail storms, heat waves, and cold spells, result in significant economic losses and human casualties each year [2]. These events adversely impact agriculture, water resources, biodiversity, and ecosystems. Given their direct correlation with weather patterns, precise weather forecasting holds the potential to alleviate risks to both human life and property damage. Weather forecasting involves predicting forthcoming atmospheric conditions like temperature, humidity, wind speed, and precipitation through scientific analysis of meteorological data and computer models. While numerical weather prediction (NWP) methods [3-6] are prevalent, they can be susceptible to inaccuracies due to incomplete understanding of atmospheric processes, limited spatio-temporal resolution, and inherent uncertainties in weather prediction. These challenges are particularly pronounced in regions with diverse landscapes like Nepal, highlighting the limitations of conventional approaches. The reliance on Euclidean structures within NWP models exacerbates these issues, as they struggle to account for the complexities of geographical features.

Many existing methods in weather forecasting primarily concentrate on short-term predictions, often forecasting only a limited number of weather attributes. Additionally, they frequently rely on datasets spanning short time-frames and data sourced from a restricted number of weather stations. Relative to these shortcomings, our study aims to address these gaps and has demonstrated notable improvements. Emerging technologies, such as graph neural network-based methods, offer a promising avenue for enhancement. By leveraging the non-Euclidean nature of geographical data, these advanced techniques can better capture the intricate relationships between weather parameters [7]. This capability holds the potential to enhance the precision and region-specificity of forecasts, crucial for mitigating risks associated with extreme weather conditions.

Existing weather forecasting methods face significant limitations, such as using datasets from short time spans, relying on data from a limited number of weather stations, and predicting only a few weather attributes with restricted input variables. Additionally, there is a lack of domain-guided graph representations that can simultaneously address diverse geographical locations. These constraints highlight the need for more comprehensive and sophisticated approaches capable of handling the complexities of varied geographic landscapes to improve the accuracy and reliability of weather forecasting.

The objective of this paper is to develop a meticulously crafted domain-guided knowledge graph representation tailored for diverse geographic landscapes like Nepal. We utilize spatio-temporal graph neural network techniques to forecast multiple weather attributes over a 24-day period, leveraging the spatial and temporal dependencies encoded within the constructed knowledge graph. Our method demonstrates robust performance even under computing resource constraints, highlighting the scalability and efficiency of the proposed methodology. Additionally, we analyze statistical behaviors to validate the coherence between the forecasted results and the underlying structure of the constructed graph, providing insights into the reliability and predictive power of our approach. Embracing such advancements signifies a significant stride toward more effective weather forecasting, vital for protecting lives and property in vulnerable regions.

2. Literature review

The field of weather forecasting employing spatio-temporal graph neural networks encompasses a variety of approaches for graph construction. For short-term wind speed forecasting [8], wind farms are represented as nodes, with edges denoting mutual correlations in wind speed and direction values, where edge weights are determined by the exponential decay of their mutual correlation distance ($e^{-d_{ij}}$). Similarly, in PM2.5 forecasting [9], cities serve as vertices with correlation-based edges. Regarding air quality prediction [10], two types of graph formations are prominent: station-level graphs, where weather stations constitute nodes and edges are established based on the inverse of Euclidean distance $\left(\frac{1}{d_{ij}}\right)$; and city-level graphs, wherein intra-city connections are fully realized. Additionally, for city air quality forecasting [11], enhancements are made to the city-level graph through a differentiable grouping network.

In the realm of multi-adversarial spatio-temporal networks [12], an unweighted graph is employed, featuring weather stations as vertices and edges determined by geographical proximity via spherical distance. Multi-stream graph attention networks [13] for windspeed forecasting adopt a graph structure with stations as nodes and utilize multi-head graph attention mechanisms (GAT) to establish edges. Frost forecasting [14] employs a similar node-based approach with weather sensors, with distance-based edges and incorporation of temporal information via temporally directed graphs.

For Point of Interest (POI) category prediction using spatio-temporal adaptive attention graph convolution [15], three graph types are considered: weather stations as vertices with edges based on spatial and Euclidean distances, as well as temporal corresponding patterns. Similarly, HistGNN [16] utilizes hierarchical, local, and global graphs with learnable multi-graph concepts. Moreover, WeatherGNN [17], a structural graph neural network [18], and CloudNine [19] employ grid graphs where grid points serve as nodes and edges connect neighboring points in cardinal directions. Table 1 gives the detailed summary of the existing works. Table 1 below shows a detailed summary of current graph representation techniques and various parameters utilized in weather forecasting through different Graph Neural Network (GNN) architectures.

Table 1: The summary of existing graph-based forecasting methods

Task	Description of parameters	
Wind speed forecasting [8]	Dataset	Eastern Wind Integration Dataset
	Time span	2007 – 2012
	Stations	145
	Input features	2
	Output features	1
$PM_{2.5}$ forecasting [9]	Dataset	$PM_{2.5}$, MEE, ERA5
	Time span	01/01/2015 – 31/12/2018
	Stations	184
	Input features	7
	Output features	2
Air quality forecasting [10]	Dataset	Air Quality, POI, Weather
	Time span	01/2018 – 12/2018
	Stations	-
	Input features	-

	Output features	1
Air quality forecasting [11]	Dataset	Chinese City Air Quality
	Time span	01/01/2017 – 30/04/2019
	Stations	209
	Input features	-
	Output features	1
Air quality forecasting [12]	Dataset	Air Quality, POI, Weather
	Time span	-
	Stations	53/87
	Input features	14
	Output features	4
Wind forecasting [13]	Dataset	Netherland Weather
	Time span	01/01/2011 – 29/03/2020
	Stations	6
	Input features	6
	Output features	1
Wind forecasting [13]	Dataset	Denmark Weather
	Time span	2000 – 2010
	Stations	4
	Input features	4
	Output features	1
Frost forecasting [14]	Dataset	-
	Time span	09/04/2020 – 04/05/2021
	Stations	11
	Input features	4
	Output features	1
Air quality prediction [15]	Dataset	Air Quality, POI, Weather
	Time span	01/2016 – 01/2018
	Stations	35
	Input features	13
	Output features	12
Air quality prediction [15]	Dataset	Air Quality, POI, Weather
	Time span	01/2014 – 04/2015
	Stations	26
	Input features	14
	Output features	20
Air quality prediction [15]	Dataset	Air Quality, POI, Weather
	Time span	01/2017 – 03/2018
	Stations	26
	Input features	11
	Output features	20
Weather forecasting [16]	Dataset	WD_BJ
	Time span	01/03/2015 – 11/03/2018
	Stations	10
	Input features	9
	Output features	3

Weather forecasting [16]	Dataset Time span Stations Input features Output features	WD_ISR 02/02/2012 – 28/10/2017 6 4 4
Weather forecasting [16]	Dataset Time span Stations Input features Output features	WD_USA 02/10/2012 – 28/10/2017 13 4 4
Weather forecasting [17]	Dataset Time span Stations Input features Output features	Ningbo 01/01/2021 – 04/01/2021 2726 (grids) 10 5
Weather forecasting [17]	Dataset Time span Stations Input features Output features	Ningxia 01/01/2021 – 01/01/2022 1200 (grids) 8 5
Weather prediction [19]	Dataset Time span Stations Input features Output features	CloudeNine 01/01/2021 – 01/01/2022 - Multiple 5

Table 1 also highlights several limitations of existing methods, including the utilization of datasets from short spans of time and the collection of weather attributes from a relatively small number of weather stations. Moreover, these methods often predict output for only a few weather attributes and consider a limited number of input variables. Additionally, there is a notable absence of domain-guided graph representations capable of addressing diverse geographical locations simultaneously. These constraints underscore the need for more comprehensive and sophisticated approaches in weather forecasting, particularly in handling the complexities of diverse geographic landscapes.

3. Mathematical formulation

In this section, we discuss some mathematical notions necessary for our simulation.

A ***weather data*** is a multivariate time series data

$$X = \{W_{t_1}, W_{t_2}, \dots, W_{t_T}\} \in \mathbf{R}^{n \times d \times T} \text{ where } W_{t_i} = \{w_1, w_2, \dots, w_n\} \in \mathbf{R}^{n \times d}$$

is the weather variables for particular time, n is the number of weather stations and

$$w_j = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix} \in \mathbf{R}^d$$

is the weather attributes from the weather station j. Also, the ***Weather forecasting*** in our context, means the prediction of major weather attributes for all the weather stations n for nest consecutive F time steps

$$\hat{Y} = \{W_{t_{T+1}}{}', W_{t_{T+2}}{}' \dots, W_{t_{T+F}}{}'\} \in \mathbf{R}^{n \times d\prime \times F}$$

where $0 < d' \leq d$ i.e. for given input $X \in \mathbf{R}^{n \times d \times T}$, we have to predict the output $\hat{Y}$ such that

$$\hat{Y} = \mathcal{M}(X)$$

where $\mathcal{M}$ is the model we aim to learn.

A graph $G = (V, \mathcal{E}, W, X)$ is the ***weighted attributed graph***[20], where V is the set of vertices, $\mathcal{E}$ is the set of edges indicating the connection between the nodes, W is the weighted adjacency matrix of size $|V| \times |V|$ whose entries indicates the connection strength between the nodes, and X is $|V| \times d$ matrix indicating the features on nodes. The collection of graphs $G = \{G_{t_1}, G_{t_2}, \dots, G_{t_T}\}$ where $G_t = (V, \mathcal{E}_t, W_t, X_t)$ is called ***temporal graph.*** If $\mathcal{E}_t = \mathcal{E}$ and $W_t = W$ for every t, then the temporal graph is known as a ***static graph with the temporal signal.***

In addition to several existing methods to the graph construction process relative to weather forecasting, we have defined a novel approach for graph construction. Our graph is a ***static graph*** with temporal signals, meaning the graph topology remains the same for all temporal frames. Thus, for each frame, we define the graph as $G = (V, \mathcal{E}, W)$, where V represents the weather stations located in 753 local governments of Nepal. The edges $\mathcal{E} = \mathcal{E}_g \cup \mathcal{E}_a$ indicate the spatial connections, which are based on geodesic proximity ($\mathcal{E}_g$) and altitude similarity ($\mathcal{E}_a$).

For spatial connections based on geodesic proximity ($\mathcal{E}_g$), we calculate the geodesic distance (d_{ij}) between weather stations i and j, and then apply the Laplacian kernel [21] to measure the connection strength:

$$w_{ij}^g = \begin{cases} \exp\left(-\frac{d_{ij}}{\sigma_g}\right) & \text{if} \exp\left(-\frac{d_{ij}}{\sigma_g}\right) \geq \tau_g \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, for spatial connections based on altitude similarity ($\mathcal{E}_a$), we calculate the difference between the altitudes of the weather stations and apply the Laplacian kernel to measure the connection strength:

$$w_{ij}^a = \begin{cases} \exp\left(-\frac{|a_i - a_j|}{\sigma_a}\right) & \text{if} \exp\left(-\frac{|a_i - a_j|}{\sigma_a}\right) \geq \tau_a \\ 0 & \text{otherwise.} \end{cases}$$

Here, a_i represents the altitude of weather station i. The parameters $\sigma_g, \sigma_a, \tau_g$, and τ_a are hyperparameters that control the connection density. Finally, $W = (w_{ij})$ denotes the weight of the edges, where:

$$w_{ij} = \begin{cases} w_{ij}^g \ \textbf{or} \ w_{ij}^a & \text{if}(i,j) \in \mathcal{E}_g \ \textbf{or} \ \mathcal{E}_a \\ \max\{w_{ij}^g, w_{ij}^a\} & \text{if}(i,j) \in \mathcal{E}_g \cap \mathcal{E}_a \end{cases}$$

This approach ensures that our graph construction effectively captures both geodesic and altitude-based spatial relationships, enhancing the accuracy and robustness of weather forecasting models in complex environments.

Now, we discuss about the model architecture and data acquisition techniques as follows:

Traditional convolutions designed for grid structure face challenges when applied to graphs. To overcome this limitation, spectral convolution, leveraging graph Fourier transform is introduced to extend convolution to graph data. For graph signal $x \in \mathbf{R}^n$ and a kernel Φ, the graph convolution is defined as

$$\Phi *_G x = U\Phi(\Lambda)U^T x \tag{1}$$

where $U \in \mathbf{R}^{n\times n}$ represents the matrix of eigenvectors of normalized Laplacian

$$L = I - D^{-\frac{1}{2}} W D^{-\frac{1}{2}} = U\Lambda U^T \tag{2}$$

$D \in \mathbf{R}^{n\times n}$ is the diagonal degree matrix with $D_{ij} = \Sigma_j W_{ij}$, $\Lambda \in \mathbf{R}^{n\times n}$ is the diagonal matrix of L and $\Phi(\Lambda)$ is a diagonal matrix. Also, to address the computational complexity of graph kernel, approximation method is employed using Chebyshev polynomial[22] defined as

$$\begin{aligned} T_0(\lambda) &= 1 \\ T_1(\lambda) &= \lambda \\ T_i(\lambda) &= 2\lambda T_{i-1}(\lambda) - T_{i-2}(\lambda), i \geq 2 \end{aligned} \tag{3}$$

The graph kernel $\Phi(\lambda)$ of $(K-1)$ order [23] is defined as

$$\Phi(\Lambda) = \Sigma_{i=0}^{K-1} \beta_i T_i(\widetilde{\Lambda}) \tag{4}$$

where $\beta \in \mathbf{R}^K$ is a vector of polynomial coefficients, and

$$\widetilde{\Lambda} = \frac{2\Lambda}{\lambda_{\max}} - 1 \tag{5}$$

is the rescaled graph Laplacian matrix transforming its eigenvalues from $[0, \lambda_{max}]$ to $[-1,1]$. Thus, it is stable for deep graph neural network. Then the graph convolution can be written as

$$\begin{aligned} \Phi *_G x &= \Phi(L)x \\ &= \Sigma_{i=0}^{K-1} \beta_i T_i(\tilde{L}) \end{aligned}$$

where $T_i(\tilde{L}) \in \boldsymbol{R}^{n\times n}$ is the Chebyshev polynomial of order K, where

$$\tilde{L} = \frac{2L}{\lambda_{max}} - 1$$

The concept for better computational efficiency, is further particularized[24] by taking $K = 2$,

$$\begin{aligned} \Phi *_G x &\approx \beta_0 x + \beta_1 \left(\frac{2L}{\lambda_{max}} - I\right) x \\ &= \beta_0 x - \beta_1 \left(D^{-\frac{1}{2}} W D^{-\frac{1}{2}}\right) x \\ &= \beta \left(\tilde{D}^{-\frac{1}{2}} \tilde{W} \tilde{D}^{-\frac{1}{2}}\right) x \end{aligned}$$

where $\beta_0 = \beta_1 = \beta; \tilde{D} = D + I$ and $\tilde{W} = W + I$.

Recurrent neural network (RNN) models have long been utilized for analyzing temporal data and have shown considerable success. However, despite their effectiveness in handling time series data, RNN models suffer from certain drawbacks such as being time-consuming, having complex gate mechanisms, and responding slowly to dynamic changes. Therefore, for weather forecasting in our scenario, we opted for spatio-temporal forecasting models [25] instead. In this approach, convolutional neural networks (CNNs) are employed to capture the dynamic behavior of weather patterns. This strategy enables a parallel and manageable training process through a multi-layer convolutional structure, facilitating hierarchical representation.

The input for each node in temporal convolution can be seen as a temporal sequence of length T with C_i channels, denoted as $Y \in \mathbf{R}^{T \times C_i}$. The convolution kernel $\Gamma \in \mathbf{R}^{K_t \times C_i \times 2C_0}$ is crafted to transform the aforementioned input into a desired output $[PQ] \in \mathbf{R}^{(T-K_t+1) \times 2C_0}$, where K_t represents the number of neighbors for each node in a graph. Here, P and Q are split evenly with the same channel size, and consequently, temporal gated convolution can be defined as:

$$\Gamma *_\tau Y = P \odot \sigma(Q) \in \mathbf{R}^{(T-K_t+1) \times C_0} \tag{6}$$

where P and Q serve as inputs for gates in the Gated Linear Unit (GLU), $\odot$ denotes the Hadamard product, and σ represents the sigmoid gate, determining the relevance of the input P for the current states. The other mechanisms remain identical to those of the model [25].

4. Results and discussion

The Figure 1 illustrates the detailed process of our method from inception to termination. In the first step, we collect daily weather data from all geographic locations in Nepal. Next, we construct a graph using the approach discussed previously. We then train a model in a supervised manner, using 60 graph snapshots (one for each day over a 60-day period) as the input to the model and predicting 24 days of weather information as the output.

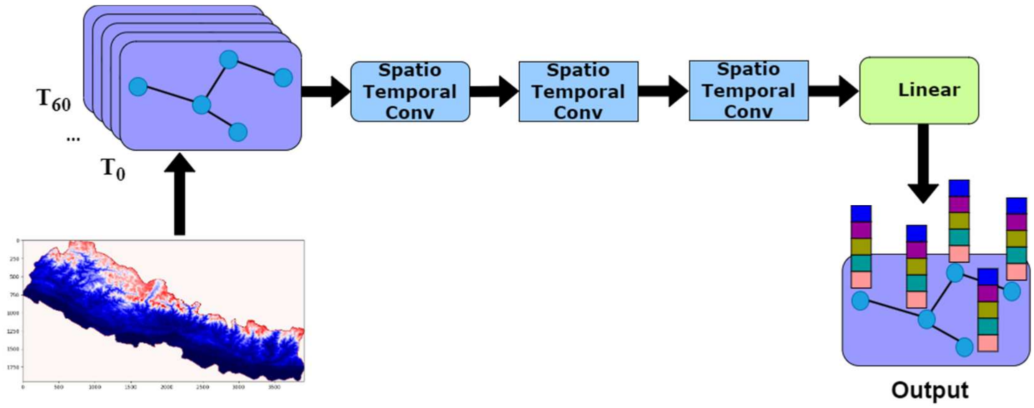

Figure 1: Architecture of Spatio-Temporal Graph Convolutional Networks (STGCN) is composed of three Spatio-Temporal Graph (STConv) blocks followed by a linear layer. Within each STConv block, the model integrates graph convolution to capture spatial information, which is interleaved with temporal gated convolution to handle temporal dynamics.

In our endeavor to gather comprehensive data for weather analysis, our focus was on the geographical diversity of Nepal's 753 local governments, ranging from the Terai lowlands to the towering Himalayan peaks. Upon inspection, three locations were identified with corrupted data, leading to their exclusion from our dataset, leaving us with a total of 750 locations.

We meticulously constructed a graphical representation incorporating these 750 stations, each serving as a node in the graph. The temporal dimension of our dataset spans from January 4, 1981, to February 24, 2024, capturing daily granularity across 18 distinct weather parameters. In this work, we have incorporated these 18 attributes namely PRECTOT: Precipitation (mm/day), PS: Surface Pressure (kPa), QV2M: Specific Humidity at 2 Meters (g/kg), RH2M: Relative Humidity at 2 Meters (%), T2M: Temperature at 2 Meters (C), T2MWET: Wet Bulb Temperature at 2 Meters (C), T2M_MAX: Maximum Temperature at 2 Meters (C), T2M_MIN: Minimum Temperature at 2 Meters (C), T2M_RANGE: Temperature Range at 2 Meters (C), TS: Earth Skin Temperature (C), WS10M: Wind Speed at 10 Meters (m/s), WS10M_MAX: Maximum Wind Speed at 10 Meters (m/s), WS10M_MIN: Minimum Wind Speed at 10 Meters (m/s), WS10M_RANGE: Wind Speed Range at 10 Meters (m/s), WS50M: Wind Speed at 50 Meters (m/s), WS50M_MAX: Maximum Wind Speed at 50 Meters (m/s), WS50M_MIN: Minimum Wind Speed at 50 Meters (m/s), WS50M_RANGE: Wind Speed Range at 50 Meters (m/s). as features for each node in the graph.

To ensure the comprehensiveness of our data collection efforts, we strategically identified landmarks within our target regions. Leveraging Application Programming Interfaces (APIs) and geographical coordinates, we properly and systematically gathered weather parameters for each landmark over a span of more than four decades. This exhaustive approach encompassed all 750 landmarks, providing a holistic understanding of the country's climatic dynamics.

To prepare this extensive dataset for analysis, a major step of standardization was employed. We normalized the features across the graph using mean and variance, ensuring equitable treatment of each weather parameter by the network. This meticulous process laid the foundation for accurate and unbiased analysis and predict.

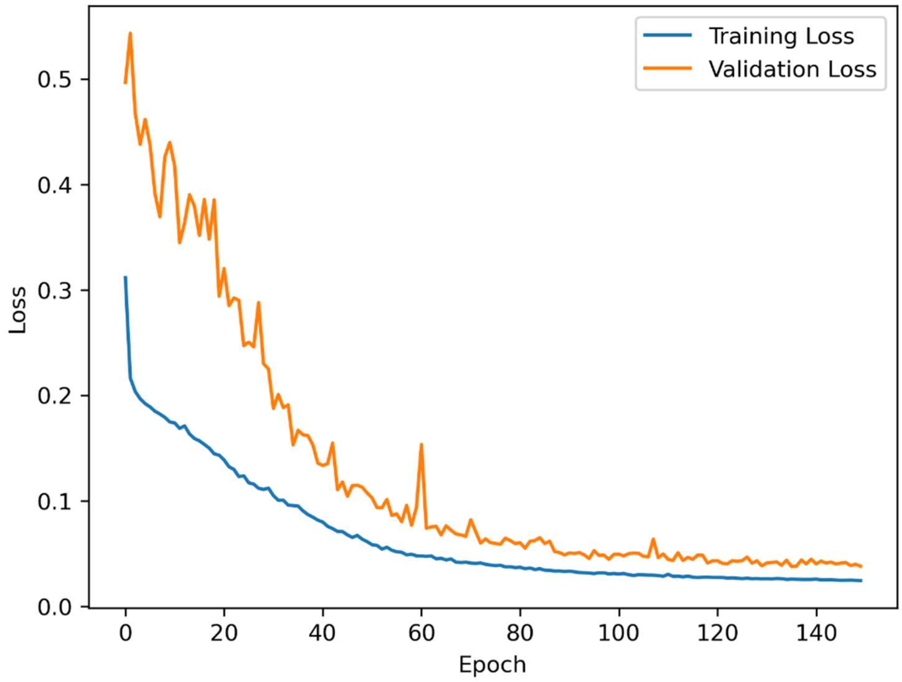

Figure 2: Both curves show a decreasing trend, which indicates that the model is learning effectively without over fitting, as there is no significant divergence between the training and validation losses

Recreating each available weather forecasting methods, we have posed significant challenges due to unavailability of datasets, public code and limited access to computational lab. Our study has reported the best average evaluation scores from existing papers for different datasets. The figure 2 illustrates the mean square error (MSE) loss for both training and validation datasets indicating effective learning capabilities of our model in weather forecasting.

Table 2: Comparison of statistical performance with existing methods

Methods	MAE	RMSE	MAPE
CloudeNine [19]	0.1900	0.1500	–
Spatio-Temporal Graph Deep Neural Networks [8]	0.3010	0.4310	–
Multi-Adversarial Spatio-Temporal Networks [12]	0.6100	–	–
Multistream Graph Attention Networks [13]	1.2530	2.6040	–

Methods	**MAE**	**RMSE**	**MAPE**
WeatherGNN (Ningbo) [17]	1.6100	2.1300	–
WeatherGNN (Ningxia) [17]	1.7100	2.2100	–
HiSTGNN (WD_ISR) [16]	2.9446	4.5606	14.1500
Graph Neural Networks with Spatio-Temporal Attention [14]	3.0800	5.4200	5.3900
HiSTGNN (WD_USA) [16]	3.5419	4.7910	19.2800
Spatio-Temporal Adaptive Attention Graph Convolution Networks (London) [15]	4.0700	5.7300	–
HiSTGNN (WD_BJ) [11]	4.2098	5.9705	36.2300
Group-aware Graph Neural Networks [13]	5.5600	10.8100	–
Multistream Graph Attention Networks	7.8800	110.3900	–
Highair [10]	8.2700	11.1800	–
$PM_{2.5}$-GNN [9]	–	19.9300	–
Spatio-Temporal Adaptive Attention Graph Convolution Networks (Beijing) [15]	12.7000	20.0800	–
Spatio-Temporal Adaptive Attention Graph Convolution Networks (Tianjin) [15]	15.0500	22.7800	–
Our Method	3.6569	5.6621	–

Our method exhibits a mixed performance when compared to existing methods, as depicted in Table 1. Notably, our method achieves competitive results in terms of Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), outperforming several existing methods such as Graph Neural Networks with Spatio-Temporal Attention and Spatio-Temporal Adaptive Attention Graph Convolution Networks on these metrics. However, our method lags behind on certain metrics when compared to some of the state-of-the-art approaches. Despite this, our method demonstrates promising potential, showcasing its ability to offer reliable predictions in certain scenarios. Additionally, our method stands out for several reasons: it incorporates a larger dataset spanning 44 years, effectively handles a large number of stations (750), includes a comprehensive set of weather attributes to forecast seven weather variables—more than almost all existing methods—and forecasts weather for diverse geographic locations, ranging from the lowlands of the Terai to the high-altitude lands of mountainous regions, simultaneously. The inclusion of these features underscores the robustness and versatility of our method in tackling complex spatio-temporal forecasting tasks. The variability in performance across different metrics underscores the complexity of the forecasting task and highlights the need for further investigation into enhancing the method's robustness and addressing its limitations. The Table 2 gives the details that those which are better than our method work only locally (less than 15 stations) and one with 87 stations. But our method works on 750 stations simultaneously. Thus we can say that our method is more robust. Overall, our method contributes valuable insights to the field, laying the groundwork for future advancements in spatio-temporal forecasting methodologies.

We present two scenarios for the ablation study. First, we explore the impact of removing independent attributes. Initially, we used 18 weather attributes to forecast only seven variables. In this ablation study, we discard the additional variables and consider only the seven attributes exactly same as the forecasted variables, allowing us to test the effectiveness of including the other weather variables in the forecasting process. Second, we examine the effect of reducing the size of the graph. Originally, we considered 25% of the fully connected edges in the spatial graph, amounting to approximately 70,000 edges with 750 nodes. Concerned that even these 25% edges might cause over-smoothing when applying graph neural networks, we reduced the edge count to 10%, resulting in about 28,000 edges, to ensure the graph remains sufficiently sparse. Both of these effects have been tested experimentally and compared with the original results, as shown in Table 2.

Table 3: Comparison of original results under multiple conditions.

Methods	MAE	RMSE	R^2
Original	3.6569	**5.6621**	**0.8624**
Reduced features	3.9600	6.4853	0.8197
Reduced features & graph size	**3.3215**	5.9283	0.8448

Table 2 shows that considering other related attributes for forecasting certain variables is beneficial. However, too much sparsifying edge connection might not be advantageous.

5. Conclusion

In summary, our method offers a robust and versatile approach to spatio-temporal weather forecasting, leveraging a comprehensive dataset spanning over four decades and encompassing 750 weather stations across Nepal's diverse landscape. Achieving competitive performance against existing methods, our approach excels in capturing complex spatio-temporal relationships while forecasting seven weather variables simultaneously. Notably, our method outperforms several existing methods on metrics like Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), showcasing its scalability and applicability to diverse geographic locations. The conducted ablation study provides valuable insights, emphasizing the importance of incorporating additional weather attributes and maintaining an optimal graph structure for enhancing forecasting accuracy. Overall, our method provides notable advancements in spatio-temporal forecasting, establishing a foundation for more precise predictions in real-world scenarios. However, developing a systematic approach for selecting weather attributes as independent variables to predict a dependent variable remains an unresolved issue.

Acknowledgments. We would like to thank the Artificial Intelligence and Smart System Research (AISSR) lab at Kathmandu University, Panchkhal, Kavre, Nepal, for their support and resources during this research. We also extend our gratitude to the reviewers for their valuable feedback and insights, which significantly contributed to the improvement of this work.

Conflicts of Interest: There is no conflict of interest among the authors.

Author's Contributions: All authors contributed equally.

REFERENCES

[1] D.Tin, L.Cheng, D.Le, R.Hata and G.Ciottone, Natural disasters: A comprehensive study using EMDAT database 1995–2022, *Public Health*, 226 (2024) 255-260.

[2] R. P. Awasthi and J. S. Owen, Observed climate extreme in Nepal, *The Geographic Base*, 7 (2020) 1-14.

[3] S. Choi and E.-S. Jung, Optimizing Numerical Weather Prediction Model Performance using Machine Learning Techniques, *IEEE Access*, (2023).

[4] S. Pereira, P. Canhoto, and R. Salgado, Development and assessment of artificial neural network models for direct normal solar irradiance forecasting using operational numerical weather prediction data, *Energy and AI*, 15 (2024) 100314.

[5] A. Rojas-Campos, M. Langguth, M. Wittenbrink, and G. Pipa, Deep learning models for generation of precipitation maps based on numerical weather prediction, *Geoscientific Model Development,* 16 (2023) 1467-1480.

[6] C. Brester, V. Kallio-Myers, A. V. Lindfors, M. Kolehmainen, and H. Niska, Evaluating neural network models in site-specific solar PV forecasting using numerical weather prediction data and weather observations, *Renewable Energy*, 207 (2023) 266-274.

[7] H. Farahmand, Y. Xu, and A. Mostafavi, A spatial–temporal graph deep learning model for urban flood nowcasting leveraging heterogeneous community features, *Scientific Reports*, 13 (2023) 6768.

[8] M. Khodayar and J. Wang, Spatio-temporal graph deep neural network for short-term wind speed forecasting, *IEEE Transactions on Sustainable Energy*, 10 (2018) 670-681.

[9] S. Wang, Y. Li, J. Zhang, Q. Meng, L. Meng, and F. Gao, Pm2. 5-gnn: A domain knowledge enhanced graph neural network for pm2. 5 forecasting, in Proceedings of the 28th International Conference on *Advances in Geographic Information Systems*, 2020, pp. 163-166.

[10] J. Xu, L. Chen, M. Lv, C. Zhan, S. Chen, and J. Chang, HighAir: A hierarchical graph neural network-based air quality forecasting method, *arXiv preprint* arXiv:2101.04264, (2021).

[11] L. Chen, J. Xu, B. Wu, and J. Huang, Group-aware graph neural network for nationwide city air quality forecasting, *ACM Transactions on Knowledge Discovery from Data*, 18 (2023) 1-20.

[12] J. Han, H. Liu, H. Zhu, H. Xiong, and D. Dou, Joint air quality and weather prediction based on multi-adversarial spatiotemporal networks, in Proceedings of the AAAI Conference on *Artificial Intelligence*, 2021, pp. 4081-4089.

[13] D. Aykas and S. Mehrkanoon, Multistream graph attention networks for wind speed forecasting, in 2021 *IEEE Symposium Series on Computational Intelligence* (SSCI), IEEE, 2021, pp. 1-8.

[14] H. Lira, L. Martí, and N. Sanchez-Pi, A graph neural network with spatio-temporal attention for multi-sources time series data: An application to frost forecast, *Sensors*, 22 (2022) 1486.

[15] H. Liu, Q. Han, H. Sun, J. Sheng, and Z. Yang, Spatiotemporal adaptive attention graph convolution network for city-level air quality prediction, *Scientific Reports*, 13 (2023) 13335.

[16] M. Ma et al., HiSTGNN: Hierarchical spatio-temporal graph neural network for weather forecasting, *Information Sciences*, 648 (2023) 119580.

[17] B. Wu, W. Chen, W. Wang, B. Peng, L. Sun, and L. Chen, Weathergnn: Exploiting complicated relationships in numerical weather prediction bias correction, *arXiv preprint arXiv*:2310.05517, (2023).

[18] X. Peng, Q. Li, L. Chen, X. Ning, H. Chu, and J. Liu, A structured graph neural network for improving the numerical weather prediction of rainfall, *Journal of Geophysical Research: Atmospheres*, 128 (2023) e2023JD039011.

[19] H.-J. Jeon, J.-H. Kang, I.-H. Kwon, and O.-J. Lee, CloudNine: Analyzing Meteorological Observation Impact on Weather Prediction Using Explainable Graph Neural Networks, *arXiv*, (2024).

[20] R. Li, Brief Note Sufficient Conditions for Hamiltonian Fuzzy Graphs, *Annals of Pure and Applied Mathematics*, 26 (2022) 15-17.

[21] F. B. Bağci and Ö. Karal, Exploring Efficient Kernel Functions for Support Vector Clustering, *Mugla Journal of Science and Technology*, 6 (2020) 36-42.

[22] M. M. Rahman, M. A. Hakim, M. K. Hasan, M. K. Alam, and L. N. Ali, Numerical solutions of Volterra integral equations of second kind with the help of Chebyshev polynomials, *Annals of Pure and Applied Mathematics*, 1 (2012) 158-167.

[23] M. Defferrard, X. Bresson, and P. Vandergheynst, Convolutional neural networks on graphs with fast localized spectral filtering, *Advances in Neural Information Processing Systems*, 29 (2016).

[24] T. N. Kipf and M. Welling, Semi-Supervised Classification with Graph Convolutional Networks, *arXiv*, (2017).

[25] B. Yu, H. Yin, and Z. Zhu, Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting, in Proceedings of the Twenty-Seventh International Joint Conference on *Artificial Intelligence*, (2018) 3634-3640.