

Integrating Interpolated Weather Data into National Clinical Data Warehouse to Support Health Services

Muhammad Tanvir Rahman

mtanvir14.bd@gmail.com

eSRD-Lab, Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology

Sadia Sharmin

Bangladesh University of Engineering and Technology

Md Raihan Mia

Marquette University

Sheikh Iqbal Ahamed

Marquette University

Abu Sayed Md Latiful Hoque

Bangladesh University of Engineering and Technology

Research Article

Keywords: Clinical Data, Weather Data, Interpolation, Machine Learning, Data Warehousing, Analytics, Decision-support system

Posted Date: October 21st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9672553/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Integrating Interpolated Weather Data into National Clinical Data Warehouse to Support Health Services

Muhammad Tanvir Rahman^{1*}, Sadia Sharmin², Md Raihan Mia³,
Sheikh Iqbal Ahamed³, Abu Sayed Md Latiful Hoque^{1, 2, 4}

^{1*}eSRD-Lab, BUET, Dhaka, Bangladesh.

²Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh.

³Department of Computer Science, Marquette University, Milwaukee, WI, USA.

⁴Department of Electrical and Computer Engineering, North South University, Dhaka, Bangladesh.

*Corresponding author(s). E-mail(s): mtanvir14.bd@gmail.com;

Contributing authors: sadia@teacher.cse.buet.ac.bd; mdraihan.mia@marquette.edu;
sheikh.ahamed@marquette.edu; asmlatifulhoque@cse.buet.ac.bd;

Abstract

Bangladesh is increasingly vulnerable to climate-sensitive infectious diseases, including dengue, cholera, malaria, and other diarrheal diseases, influenced by dynamic transmission patterns driven by meteorological variables such as temperature, humidity, and precipitation. Spatially unevenly distributed ground observation weather stations across the country create a fundamental barrier to using meteorological data for public health preparedness. Moreover, the absence of a systematic framework incorporating clinical and weather data, as well as a disease-specific data mart, points to a significant challenge in supporting preventive public health decision-making. In response, this study integrates interpolated weather data with the National Clinical Data Warehouse (NCDW) to facilitate preparedness for epidemic and pandemic outbreaks. We applied machine learning techniques to interpolate weather data across all 516 upazilla (sub-district)s in Bangladesh from 42 weather stations and found that Random Forest outperformed other models for interpolating min temperature ($R^2=0.9658$, $MSE=0.9931$), max temperature ($R^2=0.9173$, $MSE=1.2525$), rainfall ($R^2=0.3929$, $MSE=184.3544$), and humidity ($R^2=0.8109$, $MSE=13.7712$). A snowflake-schema-based architecture, introduced as Climate-NCDW, integrated the interpolated data to map clinical data in real time and established a nationwide disease-specific data mart to support advanced analytical infrastructure. Through interpolation, 82,727 records of weather data provide a comprehensive view of the weather patterns during the 2019-2024 time period. We found rainfall has the highest association with dengue disease involvement ($Corr=0.87$) during the nationwide post-monsoon period (August-October) in 2019 and 2021. Overall, our findings demonstrate that the Climate-NCDW framework can provide a robust, multidimensional big data infrastructure for real-time assessments of disease risk to support public health policy and decision-making and optimize resource allocation, marking a significant step toward integrating environmental and health informatics for proactive disease control.

Keywords: Clinical Data, Weather Data, Interpolation, Machine Learning, Data Warehousing, Analytics, Decision-support system

1 Introduction

Certain parameters of meteorological factors, such as temperature, humidity, rainfall, and atmospheric pressure, facilitate the outbreak of several communicable and non-communicable diseases [1–4]. In the context of Bangladesh, several studies have investigated the influence of weather conditions on diseases such as dengue, malaria, and diarrhea prevalence [5–8]. The high incidence and mortality from communicable diseases, i.e., Dengue, Diarrhea, Malaria, and Pneumonia, which are highly dependent on climate features, raise grave concerns about public health preparedness and response capabilities (Table 1). The World Health Organization (WHO) developed a logistic regression-based early warning system for dengue outbreaks in association with UNICEF, UNDP, and the World Bank in 2017 [9]. However, this system faces challenges in prediction and generates false alarms due to inconsistent data. In fact, the concept of a disease surveillance system, such as IEDCR [10], has been found effective for zoonotic disease outbreaks, a COVID-19 data-driven intelligent surveillance system [11]; however, none of the existing systems integrate meteorological factors into disease outbreak information or pandemic prediction, which creates a significant research gap.

Table 1 Global and National Burden of Communicable Diseases

Disease	Global		Bangladesh		Sources
	Cases	Deaths	Cases	Deaths	
Dengue	14M (2024)	9000 (2024)	321,179 (2023)	1,705 (2023)	Haider et al.[12], Hossain et al.[13]
Diarrhea	4.67B (2021)	1.2M (2021)	47,568 (2023)	39 (2023)	GBD Report[14], Saloni et al.[15], Khan et al.[16]
Malaria	263M (2023)	597,000 (2023)	16,567 (2023)	6 (2023)	WHO Report[17], Gavi/DGHS [18]
Pneumonia	455M (est.) (2023)	2.5M (2023)	4M (est.) (2023)	24,000 (children, est.) (2023)	GBD [19], icddr [20]

It is very difficult to identify the effects of weather on disease transmission dynamics and trend determination without accurate climatic data. Ground weather stations typically yield more accurate weather data than other sources, which are sparsely and unevenly distributed across Bangladesh [21]. There are 42 functioning weather stations and 516 upazillas (sub-districts) comprising 64 districts. As illustrated in Fig. 1, upazillas (sub-districts) within districts, for instance, Kishoreganj, Gazipur, Narshigdi, Brahmanbaria, Pirojpur, and Bandarban, are not in proximity to the weather stations. Therefore, it is inaccurate to estimate weather parameters in these regions, which indicates a challenge in mapping and understanding disease-sensitive areas. In response, interpolated weather data will provide precise weather parameters based on ground stations [22], which could be useful for region-wise pandemic detection and support preparedness and response. A key challenge in using these accurate ground-station observations is interpolating the meteorological variables across the upazillas (sub-districts) in Bangladesh. While geostatistical methods exist for interpolating meteorological variables, they do not account for temporal co-variation.

In this paper, we propose a big data analytical platform, Climate-NCDW (National Clinical Data Warehouse), that integrates interpolated weather data and clinical data from Bangladesh, decomposed into 512 regions, within a central data ecosystem to support early outbreak detection and public health preparedness. Previously, we built a centralized clinical data warehouse NCDW, where infrastructure is designed to facilitate the real-time clinical information of patients’ medical histories, enable comprehensive disease analysis, support population-level disease research, enhance treatment decision-making, and contribute to the improvement of healthcare policies and services in Bangladesh [23–27]. To strengthen this initiative, integrating comprehensive nationwide weather data with NCDW would represent a highly effective and innovative strategy that advances: 1) the prediction and projection of region-wide climate-sensitive infectious disease spread to facilitate preparedness and response; 2) enables researchers to obtain real-time insights regarding disease dynamics and weather variations, as well as enhancing analytical capabilities. However, several challenges need to be addressed, including creating a robust interpolation system, a scalable system architecture, and the seamless integration of clinical data with interpolated climate data, which have motivated our contribution in the following direction.

- I. Develop an interpolation system that is capable of spatially interpolating weather data across all upazillas (sub-districts) in Bangladesh (defined by latitude and longitude) using inputs from 42 weather stations using machine learning techniques.

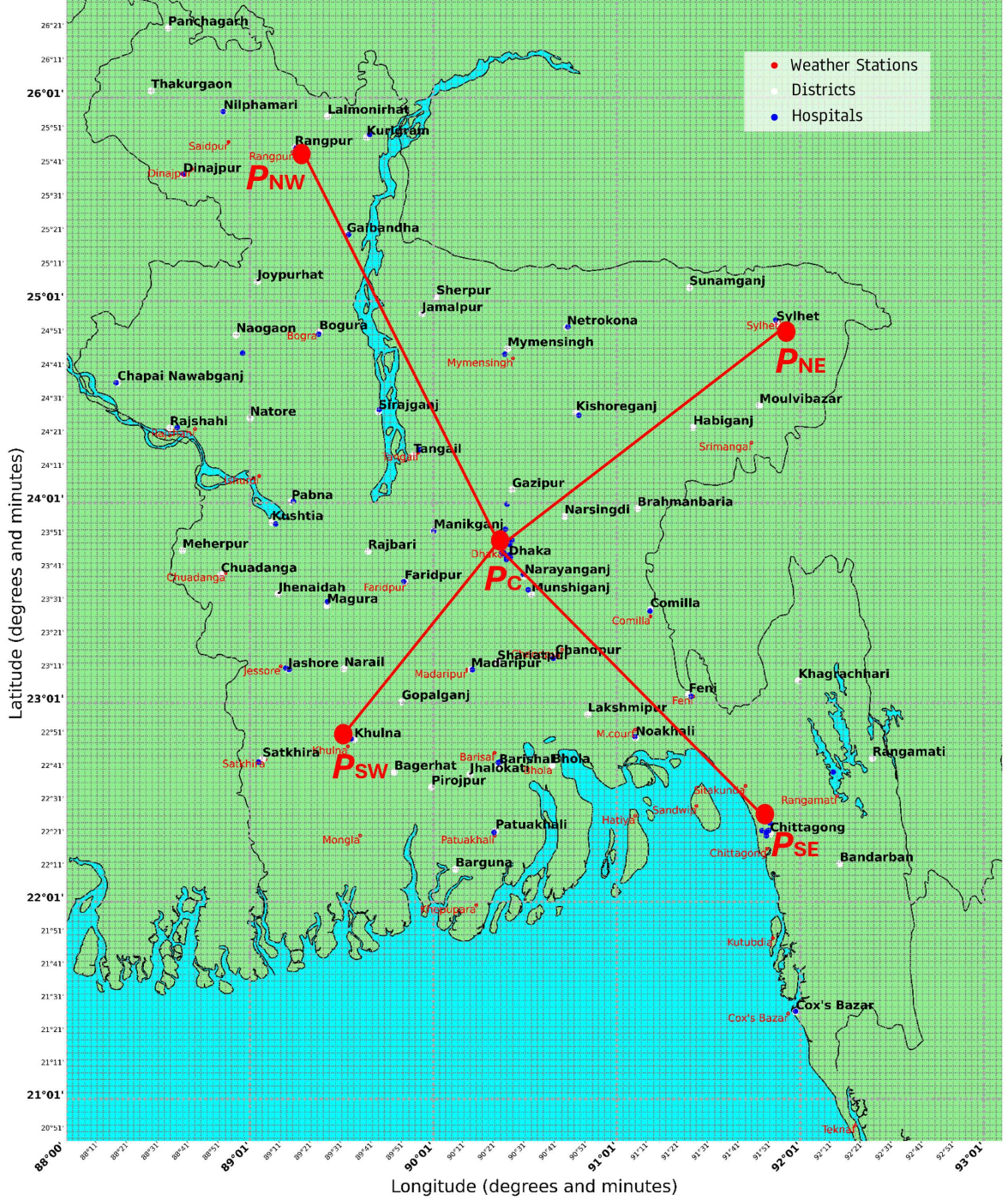


Fig. 1 Geographic distribution of weather stations, districts, and hospitals in Bangladesh.

- II. Integrate the interpolated weather data with the existing NCDW central data repository [24] to establish the Climate-NCDW interface featuring big-data storage systems and advanced analytics.
- III. Derive the ICD-11 disease classes specific datamarts from the Climate-NCDW data system to provide an online analytical processing (OLAP) interface for disease-specific research and pandemic preparedness.
- IV. Evaluate the system performance with Dengue and Diarrhea datamarts that integrate four years (2019-2024) clinical and interpolated climate data of 516 regional segments.

This paper is structured as follows. We formulated the problem statement in Section 2. Prior research on interpolation algorithms, methodological approaches, and weather-dependent diseases is reviewed and discussed in Section 3. The methodology of the Climate-NCDW system, following the constellation schema, is thoroughly described in 4. In Section 5, the climate-NCDW architecture is described with four interconnected components: Weather Data Interpolation, Clinical Data Stream,

Analytical Infrastructure, and Analytical Dashboard. Performance evaluation of the Machine Learning Models for weather data interpolation described in 6. Data sourcing and overview are discussed in Section 7. Exploratory analysis and validation of interpolated data, along with the findings and corresponding discussions, are presented in Section 8. And finally, Section 9 concludes the paper by outlining future research directions and recommendations.

2 Problem Formulation

As Bangladesh’s national health-care policies and heterogeneous data-management practices restrict access to patient-level geolocation, directly linking individual clinical records to local meteorological observations is infeasible. To enable weather–clinical integration under these constraints, we adopt a service-point proxy for location. Specifically, we consider five representative healthcare service points (Fig. 1) and assume comparable service quality across these points, as well as that each patient’s effective location corresponds to the service region of the facility they attend. We denote these targeted regions as P_C (Center: Dhaka), P_{NE} (North-East: Sylhet), P_{NW} (North-West: Rangpur), P_{SE} (South-East: Chattogram), and P_{SW} (South-West: Khulna), and define the targeted-location set as $z \in \{\text{Dhaka, Sylhet, Rangpur, Chattogram, Khulna}\}$. For each targeted location z , we define the meteorological variable set $D_z = \{T_{\min}, T_{\max}, T_{\text{dry}}, R, H\}$, where $T_{\max}$, $T_{\min}$, T_{dry} , R and H denote daily maximum and minimum temperatures, dry-bulb temperature, rainfall, and humidity, respectively.

The anonymous patients’ clinical data associated with location z is represented by set $P_z = \{P_{\text{pik}}, D_{\text{FD}}, D_{\text{VD}}, F\}$, where P_{pik} is a unique patient identification key, D_{FD} denotes fixed-dimension (time-invariant) attributes, D_{VD} denotes variable-dimension (time-varying) attributes, and F is the clinical fact/outcome of interest. At the national level, meteorological observations are recorded by (42) weather stations; represented as a full collection $D_{\text{all}} = [\tau \times \text{Location}_{s_i} \mid 1 \leq i \leq 42] \cdot \mathcal{D}$, where τ is the time index (e.g., day identifier), s_i indexes weather stations and $\mathcal{D}$ denotes the set of meteorological variables recorded per station per time.

2.1 Problem 1: Targeted-location weather extraction

Given the station-level collection D_{all} , we want to obtain the meteorological time series for each targeted location z . This requires a mapping from the set of stations to each targeted service-point region (e.g., nearest-station assignment or region-based aggregation). We define it as an operator Φ , such that $\Phi(D_{\text{all}}, z) \Rightarrow D_z$, where D_z contains the meteorological variables associated with the target location z during the specific study period.

2.2 Problem 2: Weather–clinical integration

Lets define n as number of days in the study window and m as number of healthcare data sources contributing records for location z . $D_{i,z}$ represent the meteorological variables for location z on day i , and $P_{j,z}$ is the patient records contributed by healthcare source j for location z (for the corresponding day). This problem focus on to integrate weather observations and clinical records using a spatiotemporal join over location and time as formulated in Equation 1.

$$\left(\sum_{i=1}^n D_{i,z} \right) \bowtie_{\theta} \left(\sum_{i=1}^n \sum_{j=1}^m P_{j,z} \right) \quad (1)$$

Where the join predicate is $\theta : (D_{\text{location}} = P_{\text{location}}) \wedge (D_{\text{time}} = P_{\text{time}})$, and $z \in \{\text{Dhaka, Sylhet, Rangpur, Chattogram, Khulna}\}$. This formulation supports analysis at the level of the five targeted service-point regions and can be extended to finer administrative units (e.g., upazilla/sub-district) when reliable patient location linkage becomes available.

3 Related Works

Numerous approaches have recently been used to estimate meteorological data for unknown locations. Basic methods, such as masking or using constant values, and statistical techniques such as mean, median, and mode, are considered benchmarks for more advanced models [28]. Regression-based models, such as linear regression, multivariate linear regression, ridge Regression, Lasso, etc.,

Table 2 Taxonomy of Weather Impacts on Diseases in Bangladesh

Category	Weather Features	Disease Type	Key Studies and References
Vector-Borne Diseases			
Dengue Fever	Tem_max, Tem_min, Daily_total_rainfall_in_mm, Daily_humidity	Communicable	• Generalized Linear Regression Model for Dengue Risk. [2]. • Urban environmental factors in dengue spread [36]. • Climatic factors affecting dengue in Dhaka [1]. • Studies on the potential impacts of climate change on dengue during this century [37].
Malaria	Tem_max, Tem_min, Daily_humidity, Daily_total_rainfall_in_mm	Communicable	• Climate variability and malaria [38]. • Quantitative modeling of malaria in Bangladesh [39]. • Temperature and rainfall effects on malaria [40]. • Liverpool Malaria Model for malaria spread [41].
Waterborne Diseases			
Cholera	Daily_total_rainfall_in_mm, Aver- age_Humidity_in_Month	Communicable	• Rainfall effects on cholera outbreaks [42]. • Overview of cholera transmission [43]. • Cholera prevention strategies in Bangladesh [44].
Typhoid Fever	Tem_max, Tem_min, Daily_total_rainfall_in_mm	Communicable	• Effects of temperature and rainfall on typhoid [45]. • Randomized-cluster approach for typhoid prevention [46].
Gastrointestinal Diseases			
Diarrheal Diseases	Tem_max, Tem_min, Daily_total_rainfall_in_mm, Daily_humidity	Communicable	• Temperature and precipitation effects on childhood diarrhea [4]. • Time-series analysis of 3 million diarrhea cases in Dhaka [47]. • Time-series regression led analysis between diurnal temperature range (DTR) and daily hospitalization counts for all-cause diarrhea in Dhaka, Bangladesh[48].

are exceptionally useful for trend analysis in climate data, as stated in [29]. Nonlinear approaches, such as tree-based methods for instance, random forest (RF) and decision trees (DT), are discussed in [29–31]. Widely used methods like K-Nearest Neighbors (KNN) identify the nearest data points, as demonstrated in [28, 29, 32]. Similarly, inverse distance weighting (IDW)- a technique that computes the values using the proximity and the influence of the location neighbors, has been demonstrated in [28, 32].

Neural Network-based methods such as U-NET, LSTM, and BiLSTM [28, 29, 33] and geo-statistical methods such as Gaussian Process Regression (GPR) quantify the uncertainty of the predictions discussed [31]. Multiple methodologies integrated with advanced hybrid models to ensure robustness with higher accuracy. While statistical and geo-statistical methods provide satisfactory spatial interpolation, hybrid models, such as optimal interpolation (OI), are discussed in [34] and kernel-based Lagrangian methods (DARTS) demonstrated in [35].

The evolution of weather interpolation methodologies has transformed weather forecasting from a hypothesis-based technology to a data-based refinement. The methodologies’ advantages rely on the features of the data condition, the availability of computing power, and the level of accuracy desired. This taxonomy is an all-encompassing guide to the appropriate methodologies applicable to weather interpolation problems.

With consideration to communicable and non-communicable diseases in Bangladesh, the impact of climate variability on diseases has been widely studied, as shown in Table 2. The country’s tropical climate, characterized by seasonal variations in temperature, humidity, and rainfall, plays a significant role in influencing disease outbreaks and severity.

Vector-borne viral diseases, such as dengue fever, have seen an alarming rise in Bangladesh. Studies indicate that a specific range of temperatures provides optimal conditions for the breeding of Aedes mosquitoes, which transmit the dengue virus [2]. High humidity association on dengue transmission, a particular range of rainfall involvement in creating seasonal outbreaks has been indicated in [1, 36]. Another vector-borne bacterial disease, such as cholera, thrives in Bangladesh’s fluctuating climate, with outbreaks peaking during the monsoon and post-monsoon seasons. Heavy rainfall has been linked to increased drinking from polluted water sources, increasing the risk of outbreaks [42]. Elevated

ambient temperatures further promote bacterial proliferation, exacerbating the spread of cholera [43]. Parasitic diseases like malaria, transmitted by Anopheles mosquitoes, remain a major public health concern in Bangladesh. Research highlights that malaria transmission intensifies when specific temperature, rainfall, and humidity levels enhance mosquito survival, prolonging their lifespan and increasing the likelihood of disease transmission [38, 40, 41].

Typhoid fever, a bacterial infection caused by Salmonella typhi, exhibits a strong correlation with rising temperatures, which accelerate bacterial multiplication in water sources [45]. Heavy monsoon rainfall has been linked to water contamination, particularly in urban slums [40]. Likewise, diarrheal diseases have shown a significant association with climate variability, with hospitalization risks increasing by 4.6% for every 1°C rise in temperature [47]. Heavy rainfall during the monsoon season has been found to escalate the prevalence of diarrheal infections by 29% [3].

The interaction between climate change and disease prevalence in Bangladesh highlights the critical need for proactive mitigation initiatives. The country experiences an increasing burden of climate-sensitive diseases as temperature, rainfall, and humidity patterns shift. Therefore, Bangladesh must implement vector-controlled measures and sustainable environmental policies to improve public health outcomes. These proactive mitigation efforts will enhance health sustainability and mitigate the expanding health crises.

4 Climate-NCDW Integration Methods

The NCDW schema consists of eight dimension tables and one fact table designed using star schema model, as shown in Fig. 2. The fact table 'TESTRESULT' is considered across all eight dimensions' primary keys: PATIENT, TIME, HEALTHCARE, LABINFO, TESTGROUP, TESTITEM, ATTRIBUTE, and REFRANGE, along with three measures: test_result as a non-additive fact, analyzer_name, and remarks as a factless fact. The value of measured attributes for diagnosing and screening diseases is contained in the results of clinical laboratory tests, as described in [24]. To

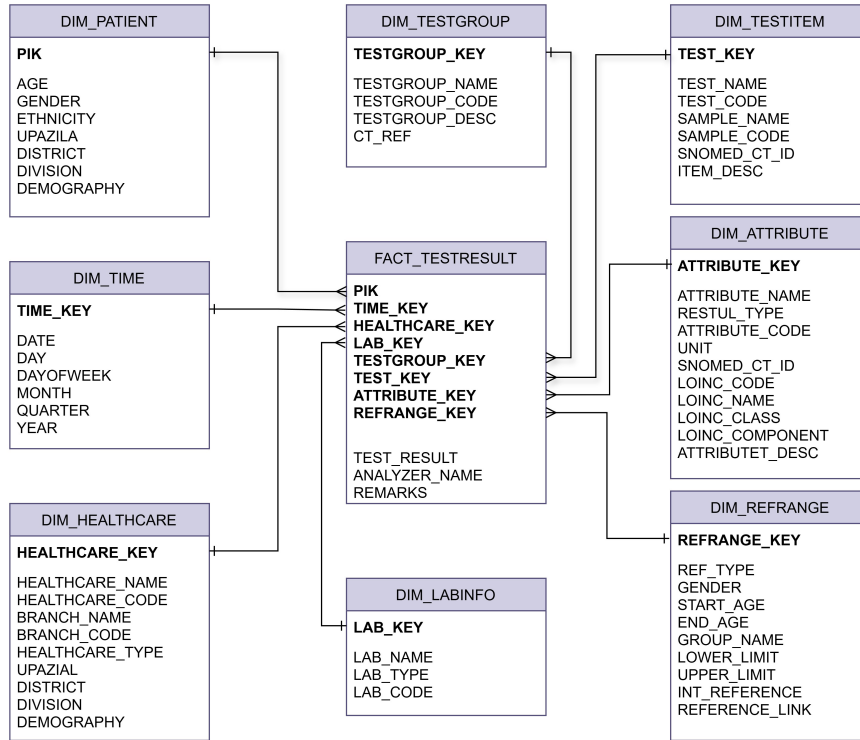


Fig. 2 Central star schema of NCDW that store anonymous clinical data

establish the state-of-the-art Climate-NCDW integration, the relational algebra representations are shown in Equation 2, 3, and 4.

$$\text{dpdl} = \left(\sigma_{\text{dp.LOC_KEY}=\text{dl.LOC_KEY}} \left(\rho_{\text{dp}}(\text{DIM_PATIENT_NEW}) \times \rho_{\text{dl}}(\text{DIM_LOCATION}) \right) \right) \quad (2)$$

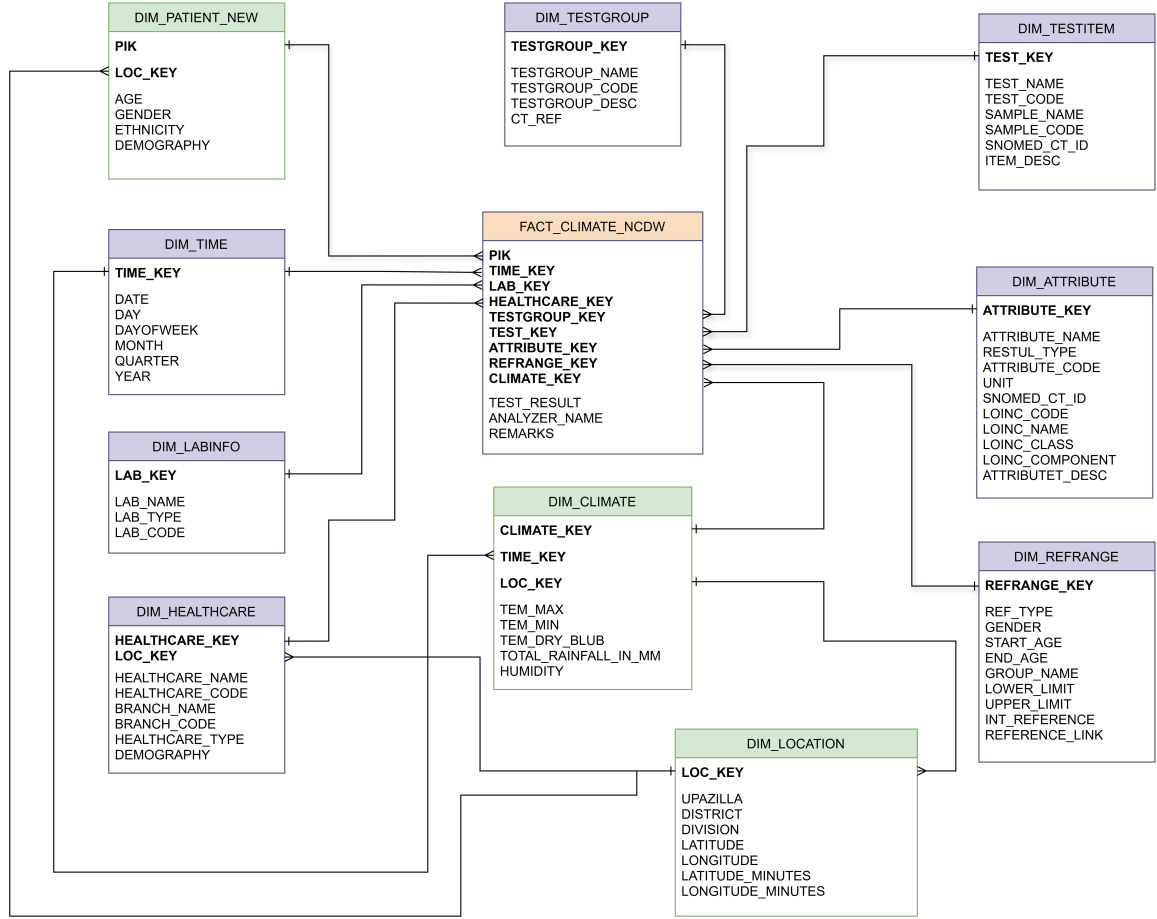


Fig. 3 Central Schema of Climate-NCDW to store clinical and interpolated weather data

$$\text{dpdl_NCDW} = \left(\sigma_{\text{dp.PIK}=\text{fn.PIK}}(\text{dpdl} \times \rho_{\text{fn}}(\text{FACT_CLIMATE_NCDW})) \right) \quad (3)$$

$$\text{Integrated Climate_NCDW} = \left(\sigma_{\text{dpdl_NCDW.CLIMATE_KEY}=\text{dc.CLIMATE_KEY}} \left(\text{dpdl_NCDW} \times \rho_{\text{dc}}(\text{DIM_CLIMATE}) \right) \right) \quad (4)$$

Therefore, to enable effective and advanced analytics, the projected Climate-NCDW representation is computed as:

$$\Pi \left(\begin{array}{l} \text{PIK, AGE, GENDER, ETHNICITY, TIME_KEY, HEALTHCARE_KEY,} \\ \text{LAB_KEY, TESTGROUP_KEY, TEST_KEY, ATTRIBUTE_KEY,} \\ \text{REFRANGE_KEY, CLIMATE_KEY, TEST_RESULT, ANALYZER_NAME,} \\ \text{LOC_KEY, TEM_MAX, TEM_MIN, TEM_DRY_BLUB,} \\ \text{TOTAL_RAINFALL_IN_MM, HUMIDITY} \end{array} \right) \left(\text{Integrated Climate_NCDW} \right)$$

This schema, as shown in Fig. 3 aims to integrate clinical factors with weather for comprehensive analysis and generating reports. By linking patient demographics and laboratory tests with weather factors, this schema enables exploring potential correlations between weather conditions and clinical outcomes. The time dimension allows for trend analysis over different periods, while referencing ranges ensures accurate interpretation of test results. This design supports medical research, disease prediction, public health monitoring, and operational efficiency in healthcare facilities. Optimized for querying and reporting, it provides a structured foundation for data-driven decision-making in healthcare and environmental research.

5 System Architecture

A data-driven, bottom-up approach has been followed to integrate weather and clinical data into the National Clinical Data Warehouse (NCDW) to provide analytical insights that support nationwide disease prevention, preparedness, and response. The architecture of the Climate-NCDW model is illustrated in Fig. 4 of four interconnected components: (i) Weather Data Interpolation, (ii) Clinical Data Stream, (iii) Analytical Infrastructure, and (iv) Analytical Dashboard. Each component serves

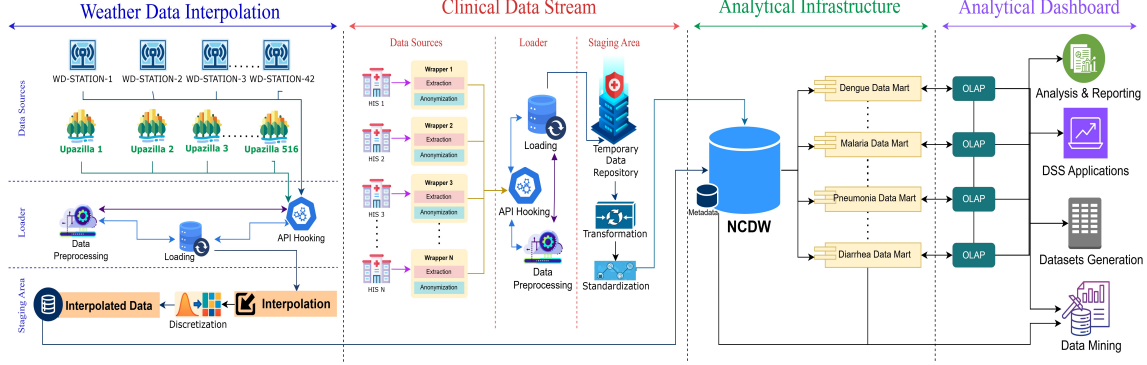


Fig. 4 Four- Layer Architecture of Climate-NCDW System 1) Interpolation layer that transfers climate data to the central repository, 2) Clinical data streaming from healthcare source to central repository using ETL process, 3) Climate-NCDW data infrastructure for data storing and deriving data marts, 4) online analytical process interface

a distinct complementary function, ensuring efficient data processing, secure storage, and advanced analytics.

5.1 Weather Data Interpolation

Due to the spatial sparsity of weather stations, ground-based weather observations are processed using a developed interpolation model to estimate weather conditions for all upazilla (sub-district) at the 'Loading' stage after completing the data pre-processing procedure. Once interpolated, the refined weather data, as 'Interpolated Data', is prepared for integration with the NCDW. The Climate-NCDW weather data interpolation methods are described below.

5.1.1 Data Modeling

For each station $s \in S$ and for each upazilla (sub-district) $u \in U$, we consider, $\text{Location}_s = (\phi_s, \lambda_s)$ and $\text{Location}_u = (\phi_u, \lambda_u)$. Where ϕ_s represents the latitude of stations, λ_s represents the longitude of stations, ϕ_u represents the latitude of upazilla (sub-district) u , λ_u represents the longitude of upazilla (sub-district) u . For each location, let $\mathcal{D}$ represent our meteorological variable set: $\mathcal{D} = \{T_{\min}, T_{\max}, T_{\text{dry}}, R, H\}$. Where, $T_{\max}$ represents maximum temperature, $T_{\min}$ represents minimum temperature, R represents rainfall, and H represents humidity. This mapping ensures each station and upazilla (sub-district) is uniquely identified by its geographical coordinates in Euclidean space, enabling spatial analysis and interpolation.

5.1.2 Generating time_id

The NCDW time_key [24] has been enhanced through the following modifications, and this methodology has also been extended to incorporate weather and upazilla (sub-district) data. To record weather information based on a specific day, a unique time_id computed as:

$$\text{time_id} = \tau = f(y, m, d) : \mathbb{N} \rightarrow \mathbb{N}[1, 12][1, 31]$$

Where, day denoted as d that values in the range $[1, 31]$, month denoted as m that values in the range $[1, 12]$, and year denoted as y that any natural number since $\mathbb{N} \rightarrow \text{Set of natural numbers } \{1, 2, 3, \dots, n\}$. $f(y, m, d)$ is a function that maps (y, m, d) to a unique time identifier. An example of time_id calculation for 5th January 2023:

$$f(2023, 1, 5) = 2023 \times 10000 + 1 \times 100 + 5 = 20230105$$

5.1.3 Coordinate System Transformation and Interpolation Algorithm

The geographical coordinates were transformed using:

$$\begin{aligned}\text{Minutes}_\phi &= 60\phi_{\text{deg}} + 60\phi_{\text{min}} \\ \text{Minutes}_\lambda &= 60\lambda_{\text{deg}} + 60\lambda_{\text{min}}\end{aligned}$$

Where, latitude in location denoted as ϕ , longitude in location denoted as λ , Minutes_ϕ represents converted latitude in minutes, Minutes_λ represents converted longitude in minutes, ϕ_{deg} represents original degree component of latitude, λ_{deg} represents original degree component of longitude, ϕ_{min} represents original minute component of latitude, λ_{min} represents original minute component of longitude. An example of coordinate transformation is shown below.

- If $\phi_s = 23.780$, then:

$$\text{Minutes}_\phi = (23 \times 60) + (0.780 \times 60) = 1426.8$$

- If $\lambda_s = 90.412$, then:

$$\text{Minutes}_\lambda = (90 \times 60) + (0.412 \times 60) = 5424.72$$

Followed by, the final weather interpolation equations are formulated as follows:
The weather station data representation:

$$\begin{aligned}[\tau \times \{Location_{s_i} | 1 \leq i \leq 42\}].\mathcal{D} \\ \Rightarrow [\{f(y, m, d)\} \times \{\phi_{s1}\lambda_{s1}, \phi_{s2}\lambda_{s2}, \phi_{s3}\lambda_{s3} \dots \phi_{s42}\lambda_{s42}\}].\{T_{\min}, T_{\max}, T_{dry}, R, H\}\end{aligned}\tag{5}$$

The upazilla (sub-district) geographical data representation:

$$\begin{aligned}[\tau \times \{Location_{u_i} | 1 \leq i \leq 516\}].\mathcal{D} \\ \Rightarrow [\{f(y, m, d)\} \times \{\phi_{u1}\lambda_{u1}, \phi_{u2}\lambda_{u2}, \phi_{u3}\lambda_{u3} \dots \phi_{u516}\lambda_{u516}\}].\{T_{\min}, T_{\max}, T_{dry}, R, H\}\end{aligned}\tag{6}$$

Finally, the Weather Interpolated Data representation:

$$\iint ([\tau \times \{Location_{s_i} | 1 \leq i \leq 42\}].\mathcal{D}) \Rightarrow [\tau \times \{Location_{u_i} | 1 \leq i \leq 516\}].\mathcal{D}_{itp}\tag{7}$$

$$\begin{aligned}\iint ([\{f(y, m, d)\} \times \{\phi_{s1}\lambda_{s1}, \phi_{s2}\lambda_{s2}, \dots \phi_{s42}\lambda_{s42}\}].\{T_{\min}, T_{\max}, T_{dry}, R, H\}) \\ \Rightarrow [\{f(y, m, d)\} \times \{\phi_{u1}\lambda_{u1}, \phi_{u2}\lambda_{u2}, \phi_{u3}\lambda_{u3} \dots \phi_{u516}\lambda_{u516}\}].\{T_{\min}, T_{\max}, T_{dry}, R, H\}_{itp}\end{aligned}\tag{8}$$

Where, $\iint(\dots)$ is a function of Interpolation Model, τ is time_id, $Location_s$ represents location of station, $Location_u$ represents Location of upazilla (sub-district), $\mathcal{D}$ denoted as Meteorological variables, $\mathcal{D}_{itp}$ denoted as Interpolated meteorological variables.

5.1.4 Interpolation

The interpolation method, as shown in Fig. 5, was developed using multiple machine learning (ML) models to interpolate weather features. The interpolation algorithm (Algorithm 1) utilizes seven ML models described in Section 6. The interpolated data have been integrated with NCDW to enhance analytical insights, strengthening nationwide disease prevention, preparedness, and response.

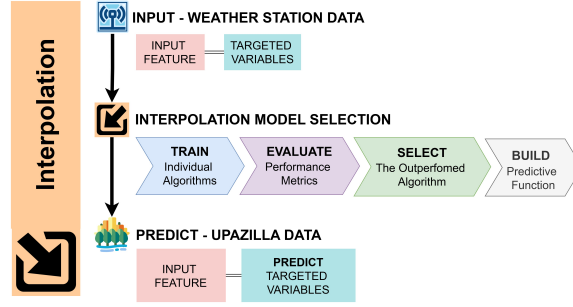


Fig. 5 A three-stage prominent spatial interpolation for estimating upazilla-level weather data from stations

Algorithm 1: Weather Data Interpolation

Input: Pre-processed Weather, weather stations and upazilla (sub-district) Location Data

Output: Upazilla (sub-district)-based Interpolated Weather Data

begin

Step 1: Load Pre-processed Weather, Weather Stations and Upazilla (sub-district) Location Data

Step 2: Define Input and Target Features:

→ **Inputs:** $\{y, m, d\} \times \{\phi_{s1}\lambda_{s1}, \phi_{s2}\lambda_{s2}, \phi_{s3}\lambda_{s3} \dots \phi_{s42}\lambda_{s42}\}$

→ **Targets:** $\{T_{\min}, T_{\max}, T_{dry}, R, H\}$

Step 3: Split Data into Train and Test sets

Step 4: Model Training

Step 5: Generate $\{y, m, d\}$ - wise $\{\phi_{u1}\lambda_{u1}, \phi_{u2}\lambda_{u2}, \phi_{u3}\lambda_{u3} \dots \phi_{u516}\lambda_{u516}\} \cdot \{T_{\min}, T_{\max}, T_{dry}, R, H\}_{itp}$

Step 6: Store the final data as upazilla (sub-district)- based Interpolated Weather Data

end

5.2 Clinical Data Stream

Real-time clinical data such as patient disease descriptions, laboratory results, and disease transmission information have been collected and stored from numerous hospitals and diagnostic centers nationwide in existing developed NCDW infrastructure [24, 26].

5.3 Analytical Infrastructure

The architecture incorporates an analytical dashboard that enables multi-dimensional data analysis. Each disease data mart is connected to an Online Analytical Processing (OLAP) system, facilitating complex queries and trend analysis. Initially, in this research, Dengue data marts are developed to facilitate targeted analysis and demonstrate the system functionality.

5.3.1 Dengue Data Mart

The Climate-NCDW facilitated nationwide clinical records that integrated with weather factors as described in Section 4. To construct the dengue datamart, we created a duplicate structure of Climate-NCDW. Following that, we selected the dengue record from FACT_CLIMATE_NCDW and fetched the copied corresponding dimensions entries from

Climate-NCDW into the replica. Algorithm 2 is used for dengue data mart creation.

Algorithm 2: NCDW to Dengue Data Mart Generation

Input: Climate-National Clinical Data Warehouse (Climate-NCDW)

Output: Dengue Data Mart

begin

Step 1: Load Climate-NCDW Fact Table (*FACT_CLIMATE_NCDW*) and Dimension Tables

Step 2: Copy Climate-NCDW structure and set name as Dengue Data Mart

Step 3: Identify and Replicate Test Records associated with dengue-
 For each record of *FACT_CLIMATE_NCDW*:

if *test_name* with “Dengue” and *test_result* $\in \{ 'Positive' \mid '+ve' \mid '+' \mid \dots \}$ **then**
 | Copy the record into the FACT.TABLE of Dengue Data Mart
 | Copy the related *Climate_NCDW* Dimensional tuples to Dengue Data Mart
 | dimensional tables

end

else

 | Continue

end

end

An End-to-End Pipeline (Climate-NCDW $\rightarrow$ Analytical Dashboard) as shown in Fig. 6 describes the data flow of healthcare data from the Climate-NCDW to a specialized analytical dataset used for dengue outbreak detection, monitoring, and research.

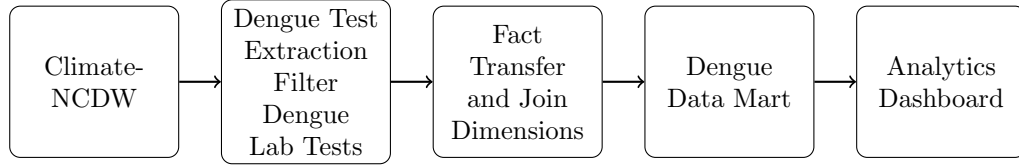


Fig. 6 An end-to-end pipeline for constructing a Dengue Data Mart from Climate-NCDW

5.3.2 Diarrhea and Other Diseases Data Marts

Diarrhea and other diseases data mart creation and operational procedure was performed using the similar Algorithm 2 by changing fact parameters and following similar workflows, such as follows:

Diarrhea data mart: *test_name* with “E. coli outcrop” and *test_result* $\in \{ 'Positive' \mid '+ve' \mid '+' \mid \dots \}$

Malaria data mart: *test_name* with “RDT” and “PCR” and *test_result* $\in \{ 'Positive' \mid '+ve' \mid '+' \mid \dots \}$

Pneumonia data mart: *test_name* with “Chest X-ray: Lung consolidation” and “Sputum Culture”, and *test_result* $\in \{ 'Positive' \mid '+ve' \mid '+' \mid \dots \}$

5.4 Analytical Dashboard

To provide meaningful insights, the architecture provides an analytical dashboard that enables multi-dimensional data analytics, incorporating Online Analytical Processing (OLAP). The analytical dashboard supports the analytical applications as follows:

- **Analysis & Reporting:** Generates real-time health reports for policymakers and health officials.
- **DSS (Decision Support System) Applications:** Supports predictive modeling for disease outbreaks.
- **Dataset Generation:** Provides structured datasets for further epidemiological studies and research.
- **Data Mining:** Enables advanced analytics to identify correlations between climate variations and disease occurrences for policy makers.

This framework facilitates real-time disease monitoring and information regarding public health interventions by leveraging a weather interpolation system, structured data marts, and OLAP-based

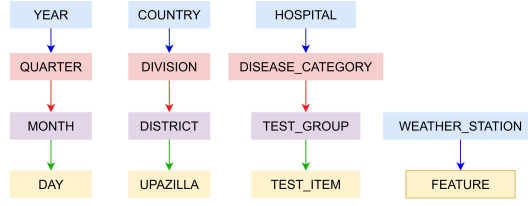


Fig. 7 CUBE Hierarchy in Weather-Disease Analysis

analytics infrastructures. Fig. 7 hierarchy shows a multidimensional schema for analyzing climate, healthcare, and disease outbreaks, which enables data aggregation and drill-down for insights such as tracking disease cases over time and across regions, along with under varying weather conditions through supporting cube and roll-up operations.

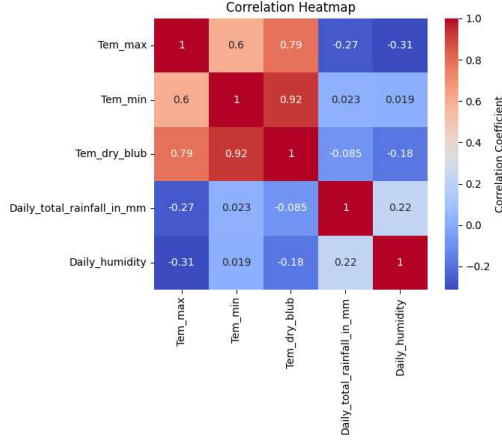


Fig. 8 Correlation Heatmap of Weather Data

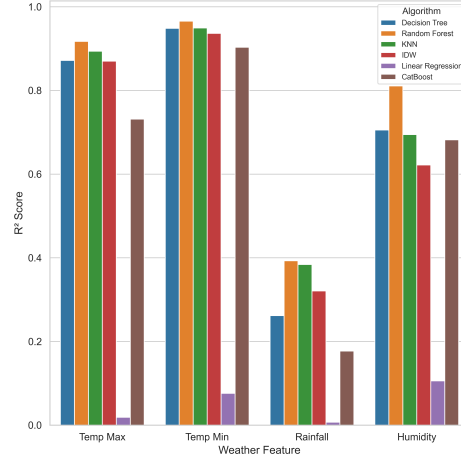


Fig. 9 Model Comparison by R^2 Score

6 Model Selection for Interpolation

Initially, the interpolation process has been conducted with the following models from existing research work as described in Section 3, along with several recommended algorithms regarding this research to determine how accurately those models will perform on weather data: Decision Tree, Random Forest Regressor, K-Nearest Neighbors (KNN), Inverse Distance Weighting (IDW), Linear Regression, CatBoost. The model exhibiting the highest accuracy is selected to construct a reliable interpolation model capable of precisely interpolating weather data across all upazilla (sub-districts) in Bangladesh. Since `tem_dry_blub` is highly correlated with `tem_max` and `tem_min`, as shown in Fig. 8, in terms of avoiding overfitting, `tem_dry_blub` has to be ignored to ensure topmost accuracy along with the hyperparameter tuning process.

The model performance comparison, as illustrated in Table 3, found that the Random Forest Regressor performs better on all weather features, showing significant variations in predictive performance across different climate features as shown in Fig. 9. Each algorithm was tested using multiple metrics, with underscoring on R^2 scores and Mean Squared Error (MSE). The Random Forest algorithm demonstrated superior performance for `Temp_max`, achieving a test accuracy of 91.73% and an R^2 score of 0.9173, with the lowest MSE (1.2525) among all models. Similarly, for `Temp_min`, Random Forest outperformed other algorithms with a test accuracy of 96.58% and an R^2 score of 0.9658, further supported by the minimal MSE (0.9931) as shown in Table 3. Notably, where all algorithms struggled with rainfall prediction, as shown in Fig. 9, the Random Forest algorithm shows a moderate performance in rainfall prediction and shows the most consistent and robust predictive capabilities across other weather features.

Table 3 Model Performance Comparison for Weather Feature Prediction

Model	Feature	Initial Accuracy (%)		Best Accuracy (%)		Hyperparameters	R ²	MSE
		Train	Test	Train	Test			
Decision Tree	Temp Max	99.37	85.34	93.71	87.19	max_depth=None, min_leaf=4	0.8719	1.9409
	Temp Min	99.61	93.98	97.48	94.85	max_depth=None, min_leaf=4	0.9485	1.4948
	Rainfall	99.25	0.00	58.56	26.19	max_depth=20, min_leaf=4	0.2619	224.1258
	Humidity	99.33	66.72	85.02	70.55	max_depth=None, min_leaf=4	0.7055	21.4499
Random Forest	Temp Max	98.33	91.65	98.35	91.73	n_est=200, max_depth=30	0.9173	1.2525
	Temp Min	99.19	96.58	99.19	96.58	n_est=200, max_depth=30	0.9658	0.9931
	Rainfall	90.51	36.53	75.25	39.29	n_est=200, max_depth=30	0.3929	184.3544
	Humidity	96.73	80.98	96.79	81.09	n_est=200, max_depth=30	0.8109	13.7712
KNN	Temp Max	92.50	88.10	99.33	89.40	n_neighbors=7	0.8940	1.6066
	Temp Min	96.17	94.25	99.56	94.93	n_neighbors=5	0.9493	1.4722
	Rainfall	57.53	32.41	99.25	38.41	n_neighbors=13	0.3841	187.0072
	Humidity	77.24	64.72	99.28	69.47	n_neighbors=7	0.6947	22.2350
IDW	Temp Max	99.37	47.36	99.37	87.00	$p = 4$	0.8700	1.9697
	Temp Min	99.61	57.19	99.61	93.64	$p = 4$	0.9364	1.8468
	Rainfall	99.25	16.77	99.25	32.07	$p = 3$	0.3207	206.2593
	Humidity	99.33	35.88	99.33	62.20	$p = 3.5$	0.6220	27.5338
Linear Regression	Temp Max	1.80	1.88	1.80	1.88	N/A	0.0188	14.8680
	Temp Min	8.04	7.61	8.04	7.61	N/A	0.0761	26.8217
	Rainfall	0.75	0.70	0.75	0.70	N/A	0.0070	301.5279
	Humidity	10.78	10.60	10.78	10.60	N/A	0.1060	65.1223
CatBoost	Temp Max	84.26	83.09	73.69	73.16	iter=200, depth=6, lr=0.05	0.7316	4.0664
	Temp Min	94.53	94.21	90.60	90.34	iter=200, depth=6, lr=0.05	0.9034	2.8057
	Rainfall	35.36	27.26	18.87	17.71	iter=200, depth=6, lr=0.05	0.1771	249.8562
	Humidity	69.68	68.22	69.68	68.22	iter=200, depth=6, lr=0.05	0.6822	23.1453

7 DATA

The weather data has been collected from the Bangladesh Meteorological Department (BMD) for this research. This organization, affiliated with the Ministry of Defense of the Government of Bangladesh, has 42 weather stations all over Bangladesh and responsible for maintaining the network of surface and upper air observatories, radar and satellite stations, agro-meteorological observatories, geomagnetic and seismological observatories, and the meteorological telecommunication system of Bangladesh [49]. This study collected daily meteorological data from 42 stations from JAN 2019 to OCT 2024, including 7 meteorological elements: maximum temperature (*Tem_max*), minimum temperature (*Tem_min*), daily air temperature (*Tem_dry_blu*), average air temperature in the month (*Average_DBT_in_Month*), daily total rainfall (*Daily_total_rainfall_in_mm*), daily humidity (*Daily_humidity*), and average humidity in the month (*Average_Humidity_in_Month*).

District and Division lists of Bangladesh, along with latitude and longitude data, have been collected from the Space Research and Remote Sensing Organization (Sparrso). Sparrso is a premier government agency under the Ministry of Defence in Bangladesh. It operates as an autonomous organization and aims to develop peaceful uses of space science and remote sensing technology throughout the country [50].

Clinical data have been collected from the listed hospital with the consent of the Directorate General of Health Services (DGHS) office, Govt. of Bangladesh. The DGHS plays a key role in providing nationwide administrative leadership and technical support to the medical colleges and the Ministry of Health and Family Welfare with the planning, management, and execution of national health policies and programs [51].

7.1 Ethical Consideration

The ethical clearance from the Bangladesh Medical Research Council (BMRC) to collect anonymous secondary data (Ref: BMRC/NREC/2019-2022/342) was obtained while the National Clinical Data Warehouse (NCDW) was being developed. Key-based Secured Record Linkage (KSRL) algorithm used to generate (PIK) to ensure anonymity of personal information, and a private-key-based encryption protocol was followed to prevent data leakage while transferring data [24].

8 RESULT AND DISCUSSION

Table 4 Summary Statistics of Weather Data

Var	Cnt	Mean	Min	25%	50%	75%	Max
Tem_max	82732	31.23	13.7	28.9	32.0	34.0	43.8
Tem_min	82732	21.53	5.0	17.0	23.4	26.0	30.9
Tem_dry	82732	25.73	9.5	22.4	27.2	29.3	35.2
Avg_DBT	82732	25.71	13.7	22.3	27.7	29.1	31.6
Rain (mm)	82732	5.92	0.0	0.0	0.0	2.0	378.0
Humidity	82732	79.33	32.0	75.0	80.0	85.0	100.0
Avg_Hum	82732	78.85	55.0	75.0	80.0	84.0	93.0

8.1 Exploratory Analysis: Climate Bangladesh

The weather dataset consists of 82,732 records of weather data from 42 weather stations across the nation. Key variables include daily maximum and minimum temperatures (Tem_max, Tem_min), rainfall (Daily_total_rainfall_in_mm), humidity (Daily_humidity), and geographic coordinates of stations (Latitude_Minutes, Longitude_Minutes). The data spans from 2019 to 2024, providing a comprehensive view of weather patterns over time. The detailed summary statistics of weather data have been enumerated in Table 4.

According to the weather dataset, Fig. 10(a) reveals distinctive seasonal rainfall patterns, with peak precipitation occurring during June-August. Humidity inversely correlates with temperature in March-April (dropping to 68-70%). Still, it positively correlates with rainfall from May to September (rising to 85-87%), with 2021 recording the highest peak humidity (87.2% in August), as shown in Fig. 10(b). The data shows that 2019 experienced the most extreme rainfall spike (19.5mm in July), while 2023 peaked (19mm) in August, demonstrating significant year-to-year variability in monsoon timing. A clear and detailed overview in Fig. 10(c). A consistent cyclical temperature pattern from 2019-2024, with maximum temperatures of 35-36°C occurring during summer months and minimum temperatures of 13-14°C during winter, creating an annual temperature range of approximately 22°C, has been shown in Fig. 10(d).

8.2 Statistical Validation of Interpolated Data

We used Pearson Correlation Coefficient statistical approach to validate interpolated weather values [52, 53]. The correlation analysis between interpolated and ground weather station data showed significant linear agreement across all variables as shown in Table 5, and additionally confirms that the interpolation preserved more than 90% of the observed data variance, indicating minimal information loss.

As shown in Fig. 11, the Pearson Correlation Analysis yielded very high coefficients ($r = 0.91-0.99$)

Table 5 Pearson Correlation (r) and Coefficient of Determination (R^2) for Station vs Interpolated Weather Variables

Variable	r	R^2
Daily maximum temperature (Tem_max)	0.986	0.972
Daily minimum temperature (Tem_min)	0.995	0.989
Daily total rainfall (mm)	0.910	0.828
Daily humidity	0.970	0.940

Note: All correlations are significant at $p < 0.001$.

with p -values < 0.001 , indicating that all relationships are statistically significant, which validates that the interpolated weather dataset is highly consistent with observed station data, ensuring the

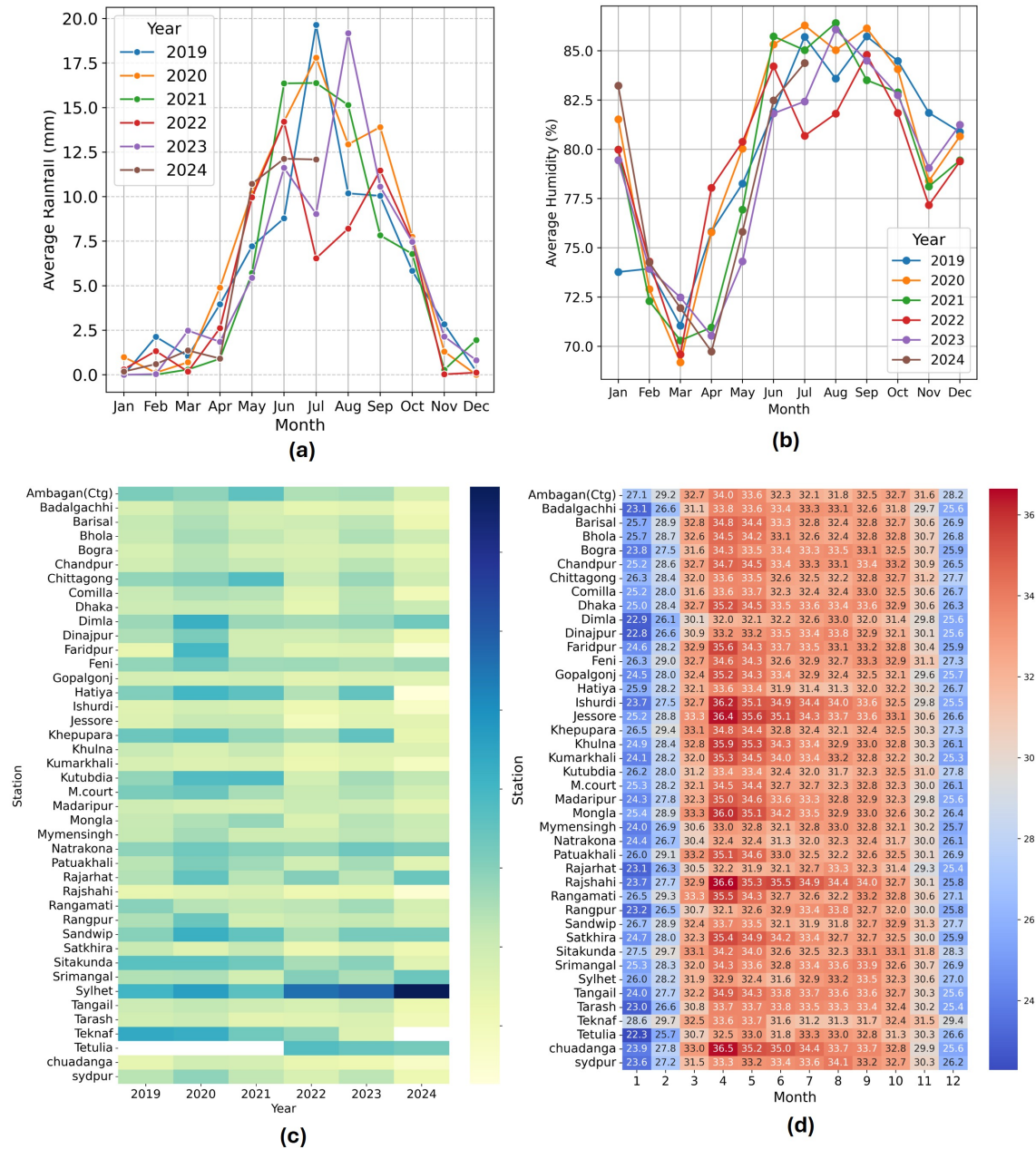


Fig. 10 Weather Variations in Bangladesh: (a) Yearwise Average Monthly Rainfall, (b) Yearwise Average Monthly Humidity, (c) Heatmap of Yearwise Average Rainfall, (d) Monthly Average Max Temperature (Stations)

reliability of the interpolation process. For instance, based on interpolated weather data, districts that are in proximity to weather stations exhibiting temperature behavior as shown in Fig. 12(a), approximately illustrate the expected outcome as station-wise temperature as shown in Fig 12(b). Overall, the interpolated dataset demonstrates excellent precision to ground-station records and is suitable for further climate-associated applications in preference to the NCDW.

8.3 Integration of Weather and Clinical Data Analysis

According to Section 4, the Climate-NCDW data, as shown in Fig. 13, shows that rainfall appears to be the most significant environmental factor associated with potential dengue outbreaks, followed by humidity and temperature, as shown in Fig. 14(a). Additionally, based on Fig. 14(b), rainfall has the strongest immediate correlation with dengue cases (0.87 at the 0-month lag), while temperature shows a delayed effect with increasing correlation over time (reaching 0.4 at the 3-month lag) when humidity maintains a moderate correlation across all lag periods. This suggests that tracking rainfall

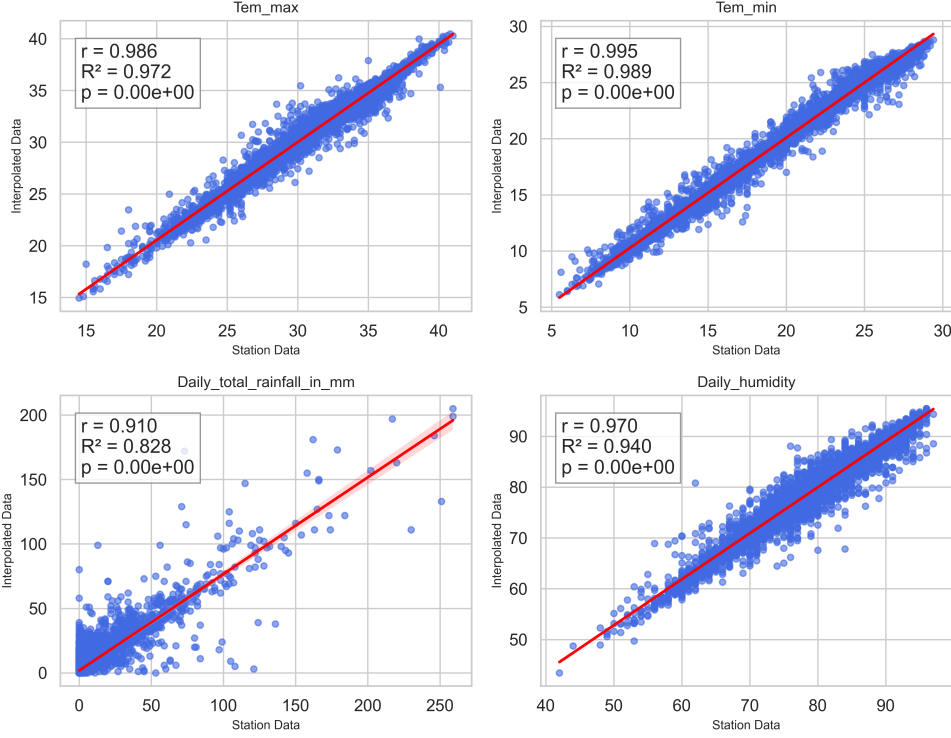


Fig. 11 Interpolated vs Station Weather Data Correlation Analysis

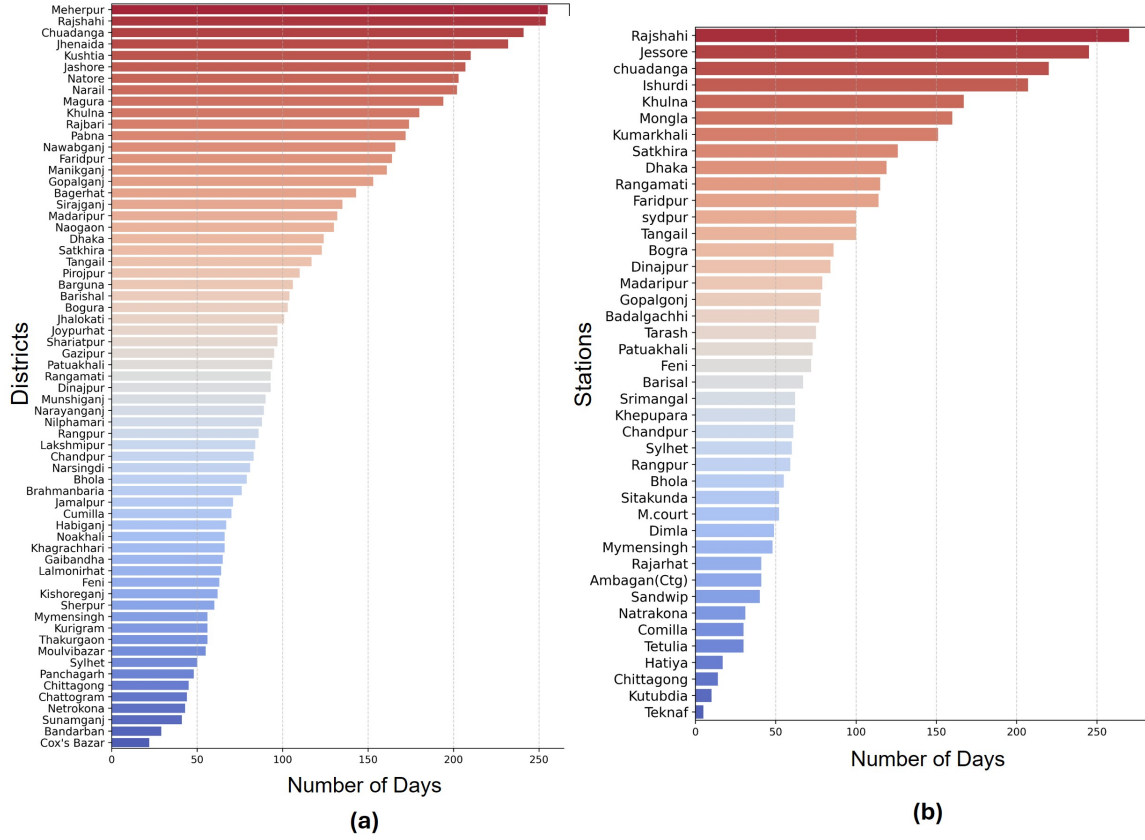


Fig. 12 Total Number of Extreme Temperature Days in 2019-2024: (a) District-wise (b) Weather Station-wise

may aid in predicting dengue and other environmental factors associated with dengue transmission, as highlighted also in [54].

Time_id	Day	Month	Year	District	Tem_max	Tem_min	Daily_humidity	Daily_total_rainfall_in.mm	Gender	Age	TESTRESULT_NO	Result
20190101	1	1	2019	Chittagong	28.9	14.3	72.1	0.2	Female	32	1190008636800	Negative
20190101	1	1	2019	Dhaka	27.0	12.4	67.1	1.1	Female	75	3190015075793	Negative
20190102	2	1	2019	Chittagong	28.6	14.5	73.0	0.0	Female	50	2190008640101	Negative
20190102	2	1	2019	Chittagong	28.6	14.5	73.0	0.0	Female	50	2190008640102	Negative
20190103	3	1	2019	Chittagong	28.1	14.1	73.3	1.2	Female	20	1190008642537	Negative
...	...	...	...	...	...	...	...	...	...	...	...	...
20211231	31	12	2021	Dhaka	26.5	15.6	70.1	0.0	Male	30	3210030122259	Negative
20211231	31	12	2021	Dhaka	26.5	15.6	70.1	0.0	Male	14	2210030126537	Negative
20211231	31	12	2021	Dhaka	26.5	15.6	70.1	0.0	Female	23	1210030122962	Negative
20211231	31	12	2021	Dhaka	26.5	15.6	70.1	0.0	Male	4	1210030129927	Negative
20211231	31	12	2021	Dhaka	26.5	15.6	70.1	0.0	Male	47	1210030120861	Negative

Fig. 13 The Climate-NCDW Data Sample of Dengue Disease

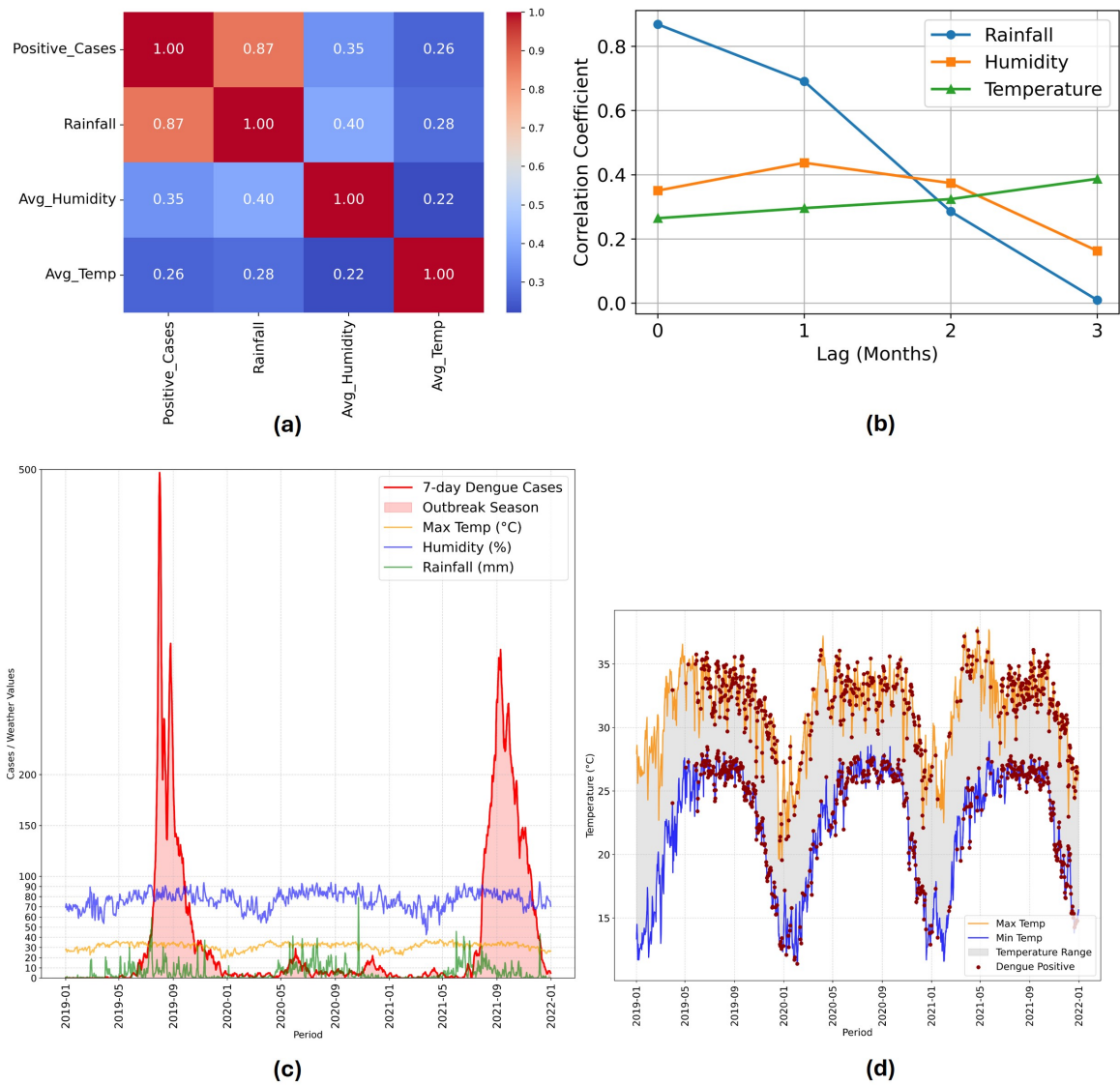


Fig. 14 Weather and Clinical Data (Dengue) Co-relation Analysis: (a) Correlation Between Dengue Cases and Climate Factors, (b) Lagged Correlation Between Climate Factors and Dengue Cases, (c) Dengue Outbreak with Weather Factors, and (d) Temperature Trend with Dengue Positive Cases

The seasonal pattern of dengue outbreaks and the strong correlations with environmental factors are evident. As shown in Fig. 14(c), the orange peaks portray distinct dengue cases, indicating a seasonal outbreak. The outbreak periods show the cyclical recurrences of high-risk windows, indicating an ideal example of the observed overlapping association between dengue transmission and weather

events - having the ability to conduct climate-based monitoring can predict potential outbreak occurrence and effectively manage risk. A clear seasonal pattern of maximum and minimum temperatures has been discovered in Fig. 14(d), which reveals a predictable clustering trend of dengue positive cases during the periods of higher temperatures, indicating a significant environmental factor influencing the assumption of dengue outbreak[37].

9 CONCLUSION AND FUTURE WORKS

This research aims to provide insights into the prevalence of climate-sensitive diseases and climate factors in Bangladesh by examining the likelihood of infections, affected areas, and the potential number of infected individuals (based on further development). Therefore, an interpolation system has been developed by leveraging the top-most performed interpolating algorithm to interpolate weather metrics across all upazillas (sub-districts) in Bangladesh, which required a big-data storage facility. Furthermore, the Climate-NCDW architecture has been established, which facilitates a big storage system for interpolated weather data and clinical data to execute integration for advanced analytics along with individual disease data marts. This will enable proactive preparedness among health sector stakeholders, such as the Ministry of Health, hospitals, doctors, researchers, and other institutions, to take necessary precautions in advance; expected to ensure significant savings in time and resources in Bangladesh's healthcare sector, thereby contributing substantially to the advancement of healthcare standards in the country.

It is noteworthy that the behavior and characteristics of diseases and infections evolve due to climate change [55, 56]. Therefore, continuous research and development will be necessary in the future to establish additional disease data marts as required all over Bangladesh. To enhance system functionality for generating actionable information, analyses, results, and insights from real-time weather and clinical data streams, collaborative support from the Bangladesh Meteorological Department (BMD) and the Directorate General of Health Services (DG-Health) will be essential to ensure effectiveness. The Climate-NCDW System will be enriched by integrating more weather features, such as wind speed with direction, atmospheric pressure, and cloudiness, along with various clinical factors that are able to estimate the effects on our health behavior and lifestyle factors. Furthermore, this system can be used for developing a nation-wide central healthcare system with proper consent from its stakeholders, which uplifted the national healthcare infrastructure to ensure more responsive and effective public health interventions in long-term.

Abbreviations

NCDW	National Clinical Data Warehouse
OLAP	Online Analytical Processing
RMSE	Root Mean Square Error
MAE	Mean Absolute Error
LSTM	Long Short-Term Memory
WHO	World Health Organization
UNICEF	United Nations Children's Fund
UNDP	United Nations Development Programme
SPARRSO	Space Research and Remote Sensing Organization
DGHS	Directorate General of Health Services
eSRD-Lab	e-Systems Research and Development Lab
BUET	Bangladesh University of Engineering and Technology

Acknowledgments

The authors acknowledge the Department of Computer Science & Engineering, Bangladesh University of Engineering and Technology, Department of Electrical and Computer Engineering, North South University, Dhaka, Bangladesh, and Ubicomp Lab, Department of Computer Science, Marquette University, for their support.

Declarations

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The weather data were purchased from the Bangladesh Meteorological Department; the Clinical data from the Bangladesh Medical Research Council (BMRC), the interpolated weather data, and integrated clinical and weather data from the NCDW during the current study are not publicly available at the time of submission; these could be available from the corresponding author on reasonable request.

Funding

This work was funded by the ICT Division of the Ministry of Posts, Telecommunications, & Information Technology of the Government of Bangladesh under Grant Code 1280101-120008431-3631108

Consent for publication

The authors hereby confirm their agreement to grant consent for the publication of this work.

References

- [1] Islam, S., Haque, C.E., Hossain, S., Hanesiak, J.: Climate variability, dengue vector abundance and dengue fever cases in dhaka, bangladesh: A time-series study. *Atmosphere* **12**(7), 905 (2021) <https://doi.org/10.3390/atmos12070905>
- [2] Hossain, S.: Generalized linear regression model to determine the threshold effects of climate variables on dengue fever: A case study on bangladesh. *Canadian Journal of Infectious Diseases and Medical Microbiology* **2023**, 2131801–12 (2023) <https://doi.org/10.1155/2023/2131801>
- [3] Grembi, J.A., Nguyen, A.T., Riviere, M., Heitmann, G.B., Patil, A., Athni, T.S., Djajadi, S., Ercumen, A., Lin, A., Crider, Y., Mertens, A., Karim, M.A., Islam, M.O., Miah, R., Famida, S.L., Hossen, M.S., Mutsuddi, P., Ali, S., Rahman, M.Z., Hussain, Z., Shoab, A.K., Haque, R., Rahman, M., Unicomb, L., Luby, S.P., Arnold, B.F., Bennett, A., Benjamin-Chung, J.: Influence of hydrometeorological risk factors on child diarrhea and enteropathogens in rural bangladesh. *PLOS Neglected Tropical Diseases* **18**(5), 0012157 (2024) <https://doi.org/10.1371/journal.pntd.0012157>
- [4] Nguyen, A.T., Grembi, J.A., Riviere, M., Heitmann, G.B., Hutson, W.D., Athni, T.S., Patil, A., Ercumen, A., Lin, A., Crider, Y., Mertens, A., Unicomb, L., Rahman, M., Luby, S.P., Arnold, B.F., Benjamin-Chung, J.: Influence of temperature and precipitation on the effectiveness of water, sanitation, and handwashing interventions against childhood diarrheal disease in rural bangladesh: A reanalysis of the wash benefits bangladesh trial. *Environmental Health Perspectives* **132**(4), 047006 (2024) <https://doi.org/10.1289/EHP13807>
- [5] Sabrina Prome, S., Basak, T., Islam Plabon, T., Khan, R.: Prediction of dengue cases in bangladesh using explainable machine learning approach. In: *Proceedings of the 7th International Conference on Inventive Computation Technologies (ICICT 2024)*, pp. 425–429. IEEE, ??? (2024). <https://doi.org/10.1109/JOCTORIES.2024.18344882> . <https://ieeexplore.ieee.org/document/10123456>
- [6] Emeto, T.I., Adegboye, O.A., Rumi, R.A., Khan, M.-U.I., Adegboye, M., Khan, W.A., Rahman, M., Streatfield, P.K., Rahman, K.M.: Disparities in risks of malaria associated with climatic variability among women, children and elderly in the chittagong hill tracts of bangladesh. *International Journal of Environmental Research and Public Health* **17**(24), 9469 (2020) <https://doi.org/10.3390/ijerph17249469>

- [7] Sharif, N., Dey, S.K.: Impact of population density and weather on covid-19 pandemic and sars-cov-2 mutation frequency in bangladesh. *Epidemiology and Infection* **149**, 16 (2021) <https://doi.org/10.1017/S0950268821000029> . Received: 16 September 2020; Revised: 15 December 2020; Accepted: 31 December 2020
- [8] Arefin, M.A., Nabi, M.N., Islam, M.T., Islam, M.S.: Influences of weather-related parameters on the spread of covid-19 pandemic – the scenario of bangladesh. *Urban Climate* **38**, 100903 (2021) <https://doi.org/10.1016/j.uclim.2021.100903>
- [9] World Health Organization: Operational Guide: Early Warning and Response System (EWARS) for Dengue Outbreaks. World Health Organization, Geneva (2017). Licence: CC BY-NC-SA 3.0 IGO. <https://apps.who.int/iris/handle/10665/254668>
- [10] Husain, M., Rahman, M., Alamgir, A., Uzzaman, M.S., Flora, M.S.: Disease surveillance system of bangladesh: Combating public health emergencies. *Online Journal of Public Health Informatics* **11**(1), 334 (2019) <https://doi.org/10.5210/ojphi.v11i1.9815>
- [11] Sarker, F., Chowdhury, M.H., Ratul, I.J., Islam, S., Mamun, K.A.: An interactive national digital surveillance system to fight against COVID-19 in bangladesh. *Frontiers in Digital Health* **5**, 1059446 (2023) <https://doi.org/10.3389/fdgth.2023.1059446>
- [12] Haider, N., Hasan, M.N., Onyango, J., Billah, M., Khan, S., Papakonstantinou, D., Paudyal, P., Asaduzzaman, M.: Global dengue epidemic worsens with record 14 million cases and 9000 deaths reported in 2024. *International Journal of Infectious Diseases* **158**, 107940 (2025) <https://doi.org/10.1016/j.ijid.2025.107940>
- [13] Hossain, M., Rakib, M.S.I., Hasan, M.M., Powshi, S.N., Islam, E., Islam, N.N.: The 2023 dengue outbreak in bangladesh: An epidemiological update. *Health Science Reports* **8**(5), 70852 (2025) <https://doi.org/10.1002/hsr2.70852>
- [14] Diseases, G., Collaborators, I.: Global incidence, prevalence, years lived with disability (ylds), disability-adjusted life-years (dalys), and healthy life expectancy (hale) for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990–2021: a systematic analysis for the global burden of disease study 2021. *The Lancet* **403**, 2133–2161 (2024) [https://doi.org/10.1016/S0140-6736\(24\)00757-8](https://doi.org/10.1016/S0140-6736(24)00757-8)
- [15] Dattani, S., Spooner, F., Ritchie, H., Roser, M.: Diarrheal diseases. *Our World in Data* (2023). <https://ourworldindata.org/diarrheal-diseases>
- [16] Khan, A.I., Islam, M.T., Amin, M.A., Khan, Z.H., Qadri, F.: Outbreak of diarrheal diseases causes mortality in different geographical locations of bangladesh during the 2021 covid-19 era. *Frontiers in Public Health* **11**, 1103518 (2023)
- [17] World Health Organization: World Malaria Report 2024: Addressing Inequity in the Global Malaria Response. World Health Organization, Geneva (2024). Licence: CC BY-NC-SA 3.0 IGO. <https://iris.who.int/>
- [18] Al Amin, M.: Bangladesh is gunning for zero malaria deaths by 2027. VaccinesWork, Gavi, the Vaccine Alliance. Accessed: 2 April 2026 (2024). <https://www.gavi.org/vaccineswork/bangladesh-gunning-zero-malaria-deaths-2027>
- [19] GBD 2023 Lower Respiratory Infections and Antimicrobial Resistance Collaborators: Global burden of lower respiratory infections and aetiologies, 1990–2023: a systematic analysis for the global burden of disease study 2023. *The Lancet Infectious Diseases* **26**(4), 343–361 (2026) [https://doi.org/10.1016/S1473-3099\(25\)00689-9](https://doi.org/10.1016/S1473-3099(25)00689-9) . Epub 2025 Dec 15
- [20] International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b): Childhood Pneumonia: Are We Doing Enough? Event report, icddr,b, Dhaka, Bangladesh. Accessed: 3 April 2026 (2023). <https://www.icddr.org/news/childhood-pneumonia-are-we-doing-enough>

- [21] Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., Hsu, K.-L.: A review of global precipitation data sets: Data sources, estimation, and intercomparisons. *Reviews of Geophysics* **56**(1), 79–107 (2018) <https://doi.org/10.1002/2017RG000574>
- [22] Gómez, J.L., Pastoriza, F.T., Álvarez, E.G., Oller, P.E.: Comparison between geostatistical interpolation and numerical weather model predictions for meteorological conditions mapping. *Infrastructures* **5**(15) (2020) <https://doi.org/10.3390/infrastructures5020015>
- [23] Khan, S.I., Hoque, A.S.M.L.: Health data integration with secured record linkage: A practical solution for bangladesh and other developing countries. In: 2017 International Conference on Networking Systems and Security (NSysS), pp. 1–6 (2017). <https://doi.org/10.1109/NSysS.2017.7885811> . IEEE
- [24] Mia, M.R., Hoque, A.S.M.L., Khan, S.I., Ahamed, S.I.: A privacy-preserving national clinical data warehouse: Architecture and analysis. *Smart Health* **23**, 100238 (2022) <https://doi.org/10.1016/j.smhl.2021.100238>
- [25] Ahmmed, S., Mondal, M.R.H., Mia, M.R., Adibuzzaman, M., Hoque, A.S.M.L., Ahamed, S.I.: A novel approach for standardizing clinical laboratory categorical test results using machine learning and string distance similarity. *Heliyon* **9**(11), 21523 (2023) <https://doi.org/10.1016/j.heliyon.2023.e21523>
- [26] Hoque, A.S.M.L., Mia, M.R., Abdullah, M.S., Islam, M.J., Nath, B.C.D., Rahman, M.T., Ahamed, S.I.: BlockPRLS: Blockchain-Based Patient Record Linkage System for Big Data Analytics . In: 2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC), pp. 877–886. IEEE Computer Society, Los Alamitos, CA, USA (2024). <https://doi.org/10.1109/COMPSAC61105.2024.00121> . <https://doi.ieeecomputersociety.org/10.1109/COMPSAC61105.2024.00121>
- [27] Soumma, S.B., Shahriar, F., Mahi, U.N., Abrar, M.H., Fahad, M.A.R., Hoque, A.L.: Design and Implementation of a Scalable Clinical Data Warehouse for Resource-Constrained Healthcare Systems (2025). <https://github.com/shovito66/ncdw>
- [28] Wang, J.: Data interpolation methods with the UNet-based model for weather forecast. *International Journal of Data Science and Analytics* (2024) <https://doi.org/10.1007/s41060-024-00611-z>
- [29] Li, C., Ren, X., Zhao, G.: Machine-learning-based imputation method for filling missing values in ground meteorological observation data. *Algorithms* **16**(9), 422 (2023) <https://doi.org/10.3390/a16090422>
- [30] Li, C., Ren, X., Zhao, G.: Machine-learning-based imputation method for filling missing values in ground meteorological observation data. *Algorithms* **16**(9) (2023) <https://doi.org/10.3390/a16090422>
- [31] López Gómez, J., Troncoso Pastoriza, F., Granada Álvarez, E., Eguía Oller, P.: Comparison between geostatistical interpolation and numerical weather model predictions for meteorological conditions mapping. *Infrastructures* **5**(2), 15 (2020) <https://doi.org/10.3390/infrastructures5020015>
- [32] Amirfakhrian, M., Samavati, F.: Weather daily data approximation using point adaptive ellipsoidal neighborhood in scattered data interpolation methods. *Applied Mathematics and Computation* **392**, 125717 (2021) <https://doi.org/10.1016/j.amc.2020.125717>
- [33] Chae, S., Kwon, S., Lee, D.: Predicting infectious disease using deep learning and big data. *International Journal of Environmental Research and Public Health* **15**(8), 1596 (2018) <https://doi.org/10.3390/ijerph15081596>
- [34] Heo, J.-H., Ryu, G.-H., Jang, J.-D.: Optimal interpolation of precipitable water using low earth orbit and numerical weather prediction data. *Remote Sensing* **10**(3), 436 (2018) <https://doi.org/10.3390/rs10030436>

- [35] Ruzanski, E., Chandrasekar, V.: Weather radar data interpolation using a kernel-based Lagrangian nowcasting technique. *IEEE Transactions on Geoscience and Remote Sensing* **53**(6), 3073–3083 (2015) <https://doi.org/10.1109/TGRS.2014.2368076>
- [36] Kamal, A.S.M.M., Al-Montakim, M.N., Hasan, M.A., Mitu, M.M.P., Gazi, M.Y., Uddin, M.M., Mia, M.B.: Relationship between urban environmental components and dengue prevalence in dhaka city—an approach of spatial analysis of satellite remote sensing, hydro-climatic, and census dengue data. *International Journal of Environmental Research and Public Health* **20**(5), 3858 (2023) <https://doi.org/10.3390/ijerph20053858>
- [37] Naish, S., Dale, P., Mackenzie, J.S., McBride, J., Mengersen, K., Tong, S.: Climate change and dengue: a critical and systematic review of quantitative modelling approaches. *BMC Infectious Diseases* **14**(167), 1–14 (2014) <https://doi.org/10.1186/1471-2334-14-167>
- [38] Emeto, T.I., Adegboye, O.A., Rumi, R.A., Khan, M.-U.I., Adegboye, M., Khan, W.A., Rahman, M., Streatfield, P.K., Rahman, K.M.: Disparities in risks of malaria associated with climatic variability among women, children and elderly in the chittagong hill tracts of bangladesh. *International Journal of Environmental Research and Public Health* **17**, 9469 (2020) <https://doi.org/10.3390/ijerph17249469>
- [39] Rahman, A., Kuddus, M.A.: Cost-effective modeling of the transmission dynamics of malaria: A case study in bangladesh. *Communications in Statistics: Case Studies, Data Analysis and Applications* (2020) <https://doi.org/10.1080/23737484.2020.1731724>
- [40] Chowdhury, F.R., Ibrahim, Q.S.U., Bari, M.S., Alam, M.M.J., Dunachie, S.J., Rodriguez-Morales, A.J., Patwary, M.I.: The association between temperature, rainfall, and humidity with common climate-sensitive infectious diseases in bangladesh. *PLOS ONE* **13**(6), 0199579 (2018) <https://doi.org/10.1371/journal.pone.0199579>
- [41] Diouf, I., Fonseca, B.R., Caminade, C., Thiaw, W.M., Deme, A., Morse, A.P., Ndione, J.-A., Gaye, A.T., Diaw, A., Ndiaye, M.K.N.: Climate variability and malaria over west africa. *American Journal of Tropical Medicine and Hygiene* **102**(5), 1037–1047 (2020) <https://doi.org/10.4269/ajtmh.19-0062>
- [42] Wu, J., Yunus, M., Ali, M., Escamilla, V., Emch, M.: Influences of heatwave, rainfall, and tree cover on cholera in Bangladesh. *Environment International* **120**, 304–311 (2018) <https://doi.org/10.1016/j.envint.2018.08.012>
- [43] Shackleton, D., Memon, F.A., Nichols, G., Phalkey, R., Chen, A.S.: Mechanisms of cholera transmission via environment in India and Bangladesh: State of the science review. *Reviews on Environmental Health* **39**, 313–329 (2023) <https://doi.org/10.1515/reveh-2022-0201>
- [44] Islam, M.T., Clemens, J.D., Qadri, F.: Cholera control and prevention in Bangladesh: An evaluation of the situation and solutions. *The Journal of Infectious Diseases* **218**, 171–172 (2018) <https://doi.org/10.1093/infdis/jiy470>
- [45] Gao, Q., Liu, Z., Xiang, J., Zhang, Y., Tong, M.X., Wang, S., Zhang, Y., Liu, Q., Jiang, B., Bi, P.: Impact of temperature and rainfall on typhoid/paratyphoid fever in taizhou, china: Effect estimation and vulnerable group identification. *American Journal of Tropical Medicine and Hygiene* **106**(2), 532–542 (2022) <https://doi.org/10.4269/ajtmh.20-1457>
- [46] Tadesse, B.T., Khanam, F., Ahmmed, F., Im, J., Islam, M.T., Kim, D.R., Kang, S.S.Y., Liu, X., Chowdhury, F., Ahmed, T., Aziz, A.B., Hoque, M., Park, J., Pak, G., Zaman, K., Khan, A.I., Pollard, A.J., Kim, J.H., Marks, F., Qadri, F., Clemens, J.D.: Prevention of typhoid by vi conjugate vaccine and achievable improvements in household water, sanitation, and hygiene: Evidence from a cluster-randomized trial in dhaka, bangladesh. *Clinical Infectious Diseases* **75**(10), 1681–1687 (2022) <https://doi.org/10.1093/cid/ciac289>
- [47] Haque, F., Lampe, F.C., Hajat, S., Stavrianaki, K., Hasan, S.M.T., Faruque, A.S.G., Ahmed, T., Jubayer, S., Kelman, I.: Impacts of climate change on diarrhoeal disease hospitalisations: How

does the global warming targets of 1.5–2°C affect Dhaka, Bangladesh? PLOS Neglected Tropical Diseases (2024) <https://doi.org/10.1371/journal.pntd.0012139>

- [48] Haque, F., Lampe, F., Hajat, S., Stavrianaki, K., Faruque, A.S.G., Jubayer, S., Kelman, I., Ahmed, T.: Effects of diurnal temperature range on diarrhea in the subtropical megacity of dhaka, bangladesh. *The Journal of Climate Change and Health* **17**, 100305 (2024) <https://doi.org/10.1016/j.joclim.2024.100305> . Open Access under CC BY-NC-ND license
- [49] United Nations ESCAP: Bangladesh Improves Disaster Early Warning System with ESCAP Support. United Nations Economic and Social Commission for Asia and the Pacific (ESCAP). Online; accessed 20-Nov-2024 (2013). <http://www.unescap.org/features/bangladesh-improves-disaster-early-warning-system-with-ESCAP-support>
- [50] Bangladesh Asiatic Society: SPARRSO. Accessed: 2024-12-03 (2024). <https://en.banglapedia.org/index.php?title=SPARRSO>
- [51] Directorate General of Health Services (DGHS): Who Are We. [Accessed: 2024-05-15] (2023). <https://old.dghs.gov.bd/index.php/en/component/content/article/109-english-root/about-us/436-who-are-we>
- [52] Liu, C.: Spatially-informed interpolation for reconstructing lake area time series using semantic neighborhood correlation. *Scientific Reports* **15**(24787) (2025) <https://doi.org/10.1038/s41598-025-09410-3>
- [53] Benesty, J., Chen, J., Huang, Y., Cohen, I.: *Pearson Correlation Coefficient*, pp. 1–4. Springer, Berlin, Heidelberg (2009). https://doi.org/10.1007/978-3-642-00296-0_5 . https://doi.org/10.1007/978-3-642-00296-0_5
- [54] Caldwell, J.M., LaBeaud, A.D., Lambin, E.F., Stewart-Ibarra, A.M., Ndenga, B.A., Mutuku, F.M., Krystosik, A.R., Beltrán Ayala, E., Anyamba, A., Borbor-Cordova, M.J., Damoah, R., Grossi-Soyster, E.N., Heras Heras, F., Ngugi, H.N., Ryan, S.J., Shah, M.M., Sippy, R., Mordecai, E.A.: Climate predicts geographic and temporal variation in mosquito-borne disease dynamics on two continents. *Nature Communications* **12**(1233) (2021) <https://doi.org/10.1038/s41467-021-21496-7>
- [55] Vuurst, P., Escobar, L.E.: Climate change and infectious disease: a review of evidence and research trends. *Infectious Diseases of Poverty* **12**(51), 1–10 (2023) <https://doi.org/10.1186/s40249-023-01102-2>
- [56] Harley, D.: Climate change and infectious diseases. *Pathology* **43** (2011) [https://doi.org/10.1016/S0031-3025\(16\)33089-6](https://doi.org/10.1016/S0031-3025(16)33089-6)